BMC Medical Research Methodology

RESEARCH ARTICLE Open Access

Constructing treatment selection rules based on an estimated treatment effect function: different approaches to take stochastic uncertainty into account have a substantial effect on performance

Maren Eckert[1*] and Werner Vach[1,2]

Abstract

Background: Today we are often interested in the predictive value of a continuous marker with respect to the expected difference in outcome between a new treatment and a standard treatment. We can investigate this in a randomized control trial, allowing us to assess interactions between treatment and marker and to construct a treatment selection rule. A first step is often to estimate the treatment effect as a function of the marker value. A variety of approaches have been suggested for the second step to define explicitly the rule to select the treatment, varying in the way to take uncertainty into account. Little is known about the merits of the different approaches.

Methods: Four construction principles for the second step are compared. They are based on the root of the estimated function, on confidence intervals for the root, or on pointwise or simultaneous confidence bands. All of them have been used implicitly or explicitly in the literature. As performance characteristics we consider the probability to select at least some patients, the probability to classify patients with and without a benefit correctly, and the gain in expected outcome at the population level. These characteristics are investigated in a simulation study.

Results: As to be expected confidence interval/band based approaches reduce the risk to select patients who do not benefit from the new treatment, but they tend to overlook patients who can benefit. Simply using positivity of the estimated treatment effect function for selection implies often a larger gain in expected outcome.

Conclusions: The use of 95% confidence intervals/bands in constructing treatment selection rules is a rather conservative approach. There is a need for better construction principles for treatment selection rules aiming to maximize the gain in expected outcome at the population level. Choosing a confidence level of 80% may be a first step in this direction.

Keywords: Interaction models, Pointwise confidence band, Root, Simultaneous confidence band, Treatment selection

*Correspondence: eckert@imbi.uni-freiburg.de
[1]Institute of Medical Biometry and Statistics, Section of Health Care Research and Rehabilitation Research, Faculty of Medicine and Medical Center - University of Freiburg, Hebelstr. 11, 79104 Freiburg, Germany
Full list of author information is available at the end of the article

Background

Today we are often confronted with the task to investigate the predictive value of a continuous marker with respect to the expected difference in outcome between a new treatment and a standard treatment. A randomized controlled trial (RCT) can (and should) be used for such an investigation. It does not only allow to demonstrate an interaction between treatment choice and the marker, but also to construct a treatment selection rule. Such a rule aims at identifying those patients who can expect to benefit from the new treatment. It is a function of the marker value and hence can be applied also to future patients outside of the trial.

Several statistical methods have been proposed in the literature to construct treatment selection rules. Many of them are based on estimating the treatment effect $\theta(x)$ as a continuous function of the biomarker value x. Both parametric [1–3] as well as semi- or nonparametric approaches [4–6] can be found. However, although estimating $\theta(x)$ is a valuable step, it does not automatically provide a rule to determine those biomarker values with $\theta(x) > 0$; it remains the question whether and how to take stochastic uncertainty of $\hat{\theta}(x)$ into account.

Confidence bands have been considered by several authors to describe the uncertainty in $\hat{\theta}(x)$. Pointwise bands (e.g. [5]) and simultaneous confidence bands (e.g. [4]) as well as both together (e.g. [7, 8]) have been suggested. Mackey and Bengtsson, Riddell et al. [1, 3] suggest to construct a confidence interval for the root of $\theta(x)$ (with respect to 0 or another threshold), and similarly [2] suggest to compute horizontal confidence intervals. In contrast, some authors (e.g. [6]) only present a raw estimate of $\theta(x)$. However, all these authors do not explicitly address the question how to move from a (graphical) illustration of uncertainty to a concrete rule.

In recent years, there are some papers addressing the question more explicitly. Baker and Bonetti [9] as well as [10] suggest to check where the lower bound of the simultaneous confidence interval of the estimated subgroup treatment effect is positive. The former uses a confidence level of 95% and the latter one of 99%. In an overview about the construction of treatment selection rules [11] also consider pointwise and simultaneous confidence bands and rules based on comparing the lower bound with 0 or another given threshold.

In summary, we would like to argue that all authors directly or implicitly suggest to use one of the following types of treatment selection rules: If only the estimate $\hat{\theta}(x)$ is (graphically) presented, in future all patients with $\hat{\theta}(x) > 0$ should receive the new treatment. If pointwise or simultaneous confidence bands for the treatment effect are also shown, all covariate values x with positive values of the lower bound should define the treatment selection rule. If a confidence interval for the root of $\theta(x)$ is given, only x-values outside of this interval satisfying also $\hat{\theta}(x) > 0$ define the patients to be selected for the new treatment. We focus in this paper on the threshold 0 for the treatment effect, but our considerations are also applicable for any other threshold.

It is the purpose of this paper to give some insights into the performance of these principles to construct treatment selection rules. We are interested in differences in the impact for future patients outside of the trial when following the various principles. As potential impact we consider the correct identification of patients who do or do not benefit from the new treatment and the change in outcome at the population level.

Methods

Notation

To compare these principles we introduce some basic notations. Let X be the continuous covariate representing the biomarker value. Let Y be a continuous outcome and T the treatment indicator, randomized with a 50 percent chance to 0 or 1, and indicating a treatment with the standard or the new treatment, respectively. The treatment effect $\theta(x)$ is defined as the difference between the expected outcomes:

$$\theta(x) := E(Y \mid X = x, T = 1) - E(Y \mid X = x, T = 0)$$

We assume that higher values of Y represent a higher treatment success. Thus, a positive treatment effect characterizes superiority of the new treatment.

A treatment selection rule can be regarded as the choice of a subset C of all possible values of X. Patients with covariate values in C should receive the new treatment instead of the standard treatment in future. A construction method is an algorithm to transform the data $(Y_i, X_i, T_i)_{i=1,\ldots,n}$ observed in an RCT into a set C. Since the result of a construction method depends on random data, we consider it as a set-valued random variable $\mathcal{C}$. We can study the performance of the construction method by considering the distribution of $\mathcal{C}$.

Performance characteristics

We start by defining quality measures for a single set C. Since this set C determines the treatment selection for future patients, we introduce a new random variable X^* denoting the biomarker value for future patients. We consider three quality measures:

$$\begin{aligned} \text{Sensitivity} &:= P(X^* \in C \mid \theta(X^*) \geq 0) \\ \text{Specificity} &:= P(X^* \notin C \mid \theta(X^*) < 0) \\ \text{Overall gain} &:= E(\theta(X^*)\mathbb{1}_{X^* \in C}) \end{aligned}$$

Sensitivity and specificity focus on the correct classification of patients by the treatment selection rule. Sensitivity measures the ability to select those patients who

can expect to benefit from the new treatment. Specificity measures the ability to avoid recommending the new treatment to patients who cannot benefit from it. The overall gain is a summary measure taking into account also the magnitude of the treatment effect. It represents the change in the average outcome (i.e. in $E(Y)$), when we apply the proposed treatment selection rule in future, i.e. patients with $x^* \notin C$ receive the standard treatment and patients with $x^* \in C$ receive the new treatment. It takes into account that $\theta(x^*)$ may be actually negative for some patients selected by the rule. The gain can be also seen as one specific way to balance between sensitivity and specificity, or – to be precise – between true positive and false positive decisions. A patient with $\theta(x) > 0$ correctly selected to receive the new treatment gets a weight equal to his or her individual benefit. A patient with $\theta(x) < 0$ incorrectly selected to receive the new treatment gets a weight equal to his or her individual, negative benefit. All patients selected for standard treatment get a weight of 0.

We have chosen these three measures, as they cover important characteristics. The different construction principles mentioned in the introduction can be regarded as attempts to control the specificity at the price of a reduced sensitivity. The overall gain measures the success of obtaining a sufficient balance in the sense that a low specificity decreases the overall gain by including too many patients with a negative $\theta(x^*)$, and a low sensitivity decreases the overall gain by excluding too many patients with a positive $\theta(x^*)$. However, it takes also into account that it is most favourable to include patients with large positive values of $\theta(x^*)$ and least favourable to include patients with large negative values of $\theta(x^*)$. Measures similar to the overall gain have been considered in the literature, but mainly with respect to the optimal rule $C = \{x \mid \theta(x) \geq 0\}$ as a measure of the benefit we can expect from a new biomarker. See [2] and the references given there. In the presentation of the results we will also indicate the maximal possible overall gain as a benchmark, defined as $E(\theta(X^*)\mathbb{1}_{\theta(X^*)\geq 0})$.

To describe the performance of a construction method for treatment selection rules, we study the distribution of these three quality measures when applied to $\mathcal{C}$ under the assumption that X^* follows the same distribution as X. In this paper we will only consider the mean of this distribution, i.e. the expected sensitivity, the expected specificity, and the expected overall gain. In the context of comparing different subgroup analysis strategies, the expected overall gain has also been considered by [12].

Construction principles for treatment selection rules

As mentioned above, we will consider four different construction principles for the treatment selection rule. All of them are based on the assumption that we have some statistical method providing us with an estimate $\hat{\theta}(x)$. Three principles assume that we can also perform certain types of statistical inference in order to construct pointwise or simultaneous confidence bands of the treatment effect or confidence intervals for the roots of $\theta(x)$. In the sequel, let $l_p(x)$ and $l_s(x)$ denote the value of the lower bound of a 95 percent pointwise and simultaneous confidence band, respectively. Let $\mathrm{CI}(x_r)$ denote a confidence interval around any root x_r, i.e. $x_r \in \hat{\theta}^{-1}(0) = \{x \mid \hat{\theta}(x) = 0\}$. Then, the construction principles can be described like shown in Table 1.

There is a close conceptual relation between the two principles POI and CIR. Both aim at excluding marker values x for which $\theta(x) = 0$ is "likely". POI tries to identify these values by considering the uncertainty in $\hat{\theta}(x)$. CIR tries to identify these values by considering the uncertainty in determining the root(s) of $\theta(.)$. (There can be several roots when $\theta(.)$ is chosen as a non-linear function, resulting in the somewhat technical definition shown above). Moreover, there is a direct mathematical relation. If a pointwise $1-\gamma$ confidence band for $\theta(.)$ is given, we can interpret it not only vertically, but also horizontally in the following sense: If for a given θ_t we consider all values of x such that (θ_t, x) is within the confidence band, then these values define a $1-\gamma$ confidence interval for $\theta^{-1}(\theta_t)$. A proof is outlined in Additional file 1.

We will nevertheless consider POI and CIR as different approaches, as there are a variety of methods to obtain confidence intervals for $\theta^{-1}(0)$. In particular we will consider a simple application of the delta rule to obtain standard errors of $\theta^{-1}(0)$, as it has been also used in [1].

Design of simulation study

In the general set up of the simulation study we generate a random variable $X \in [0,1]$ representing the biomarker. T is generated as a Bernoulli random variable with a probability of 0.5. The continuous outcome Y follows a normal error model: $Y = \alpha(X) + \theta(X)T + \varepsilon$, where $\varepsilon \sim N(0,1)$. As the error variance is fixed to one, the value of $\theta(x)$ can be interpreted roughly as an effect size. We chose to investigate three shapes for the treatment effect function $\theta(x)$, a linear, a concave and a convex shape, see Fig. 1. Within each shape we have a scaling parameter

Table 1 Construction principles and the corresponding treatment selection rules

Construction principle	Treatment selection rule
estimator (EST)	$\mathcal{C}_{\mathrm{EST}} := \{x \mid \hat{\theta}(x) \geq 0\}$
95 percent pointwise confidence band (POI)	$\mathcal{C}_{\mathrm{POI}} := \{x \mid l_p(x) \geq 0\}$
95 percent simultaneous confidence band (SIM)	$\mathcal{C}_{\mathrm{SIM}} := \{x \mid l_s(x) \geq 0\}$
95 percent confidence interval of all roots (CIR)	$\mathcal{C}_{\mathrm{CIR}} := \{x \mid \hat{\theta}(x) \geq 0\} \setminus \bigcup_{x_r \in \theta^{-1}(0)} \mathrm{CI}(x_r)$

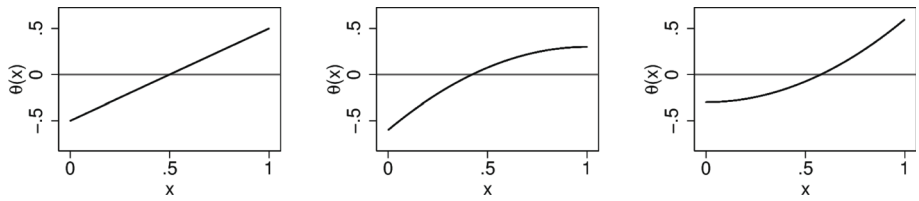

Fig. 1 Three shapes for $\theta(x)$ with $\beta = 1$. **a** $\theta(x) = \beta(x - 0.5)$ **b** $\theta(x) = \beta(0.3 - 0.9(x-1)^2)$ **c** $\theta(x) = \beta(-0.3 + 0.9x^2)$

β reflecting the steepness of the function. For the linear case we chose to investigate two different distributions of the biomarker, $X \sim \mathcal{U}(0,1)$ or $X \sim \mathcal{T}(0,1,1/3)$, while we only look at a uniformly distributed biomarker for the other two shapes. Here $\mathcal{T}(a,b,c)$ denotes a triangular distribution on the interval (a,b) with a mode in c. We do not consider the case of a normally distributed X, as the theory behind the methods we use to construct simultaneous confidence bands applies only to bounded intervals. Thus, in total we are investigating four scenarios summarized in Table 2. Without loss of generality we will assume $\alpha(x) = 0$ in generating the data. This is justified if we assume that the analysis models used are correctly specified with respect to $\alpha(x)$, such that the estimates for $\theta(x)$ are invariant under the transformations $Y' = Y + \alpha(X)$.

In estimating $\theta(x)$ we use linear regression assuming a linear or a quadratic model for $\alpha(X)$ and $\theta(X)$:

General analysis model:	$Y = \alpha(X) + \theta_\beta(X)T$
Linear analysis model:	$\alpha(X) = \alpha_0 + \alpha_1 X$
	$\theta_\beta(X) = \beta_0 + \beta_1 X$
Quadratic analysis model:	$\alpha(X) = \alpha_0 + \alpha_1 X + \alpha_2 X^2$
	$\theta_\beta(X) = \beta_0 + \beta_1 X + \beta_2 X^2$

We will focus on using the "correct" analysis model, i.e. we apply the quadratic analysis model if $\theta(x)$ is concave or convex, and the linear model otherwise. The mathematics for building the pointwise and simultaneous confidence bands and the confidence intervals for the roots are outlined in Additional file 2. Candidate sets are constructed as described above for each of the four principles. However, this step is only performed in case of a significant interaction test, i.e. if H_0:$\beta_1 = 0$ or H_0:$\beta_1 = \beta_2 = 0$, respectively, could be rejected at the 5 percent level. In case of no significance all candidate sets are empty, i.e. $\mathcal{C} = \emptyset$.

Table 2 Characteristics of the investigated scenarios

Scenario	Shape of $\theta(x)$	Distribution of X
1	Linear: $\theta(x) = \beta(x - 0.5)$	$\mathcal{U}(0,1)$
2	Linear: $\theta(x) = \beta(x - 0.5)$	$\mathcal{T}(0,1,1/3)$
3	Concave: $\theta(x) = \beta(0.3 - 0.9(x-1)^2)$	$\mathcal{U}(0,1)$
4	Convex: $\theta(x) = \beta(-0.3 + 0.9x^2)$	$\mathcal{U}(0,1)$

In addition to the performance characteristics expected sensitivity, expected specificity, and expected overall gain, we also consider $P(\mathcal{C} \neq \emptyset)$, i.e. the probability to select at least some patients for the new treatment. We refer to this probability as the power, as it reflects the chance to get a "positive" result from the investigation of interest. It will also allow to judge the relevance of a chosen β value. The numerical computation of the performance characteristics is outlined in Additional file 3.

The sample size for a single trial was chosen in order to obtain for a clinically relevant situation a power of at least 90 percent with the most conservative method (i.e. SIM) in scenario 1. The relevant situation is characterized by one quarter of the patients to have a treatment effect above 0.2, corresponding to the choice $\beta = 0.8$. The calculations resulted in a sample size of 1500, which we used for all scenarios. The number of repetitions in the simulation study was set to 2500, allowing to estimate a power of 90 percent with a standard error of 0.6 percent.

All calculations were performed using `Stata 13`. We used the available built-in procedures for generating random numbers, performing linear regression, construction of pointwise confidence bands (`lincom`) and application of the delta rule (`nlcom`). The calculation of the simultaneous confidence intervals were performed with self-written `Stata` programs and self-written functions in `Mata`, a programming language integrated in `Stata`. Source code for reproducing the results of the simulation can be viewed as Additional file 4 which also includes the data sets produced by the simulation.

Results

Scenario 1

In this scenario we consider the case of a linear true treatment effect $\theta(x)$ and X being uniformly distributed. We can observe distinct differences between all four construction principles (Fig. 2). As expected EST has the highest power while SIM, as the most conservative method, has the lowest power. As β increases so does power, sensitivity and overall gain for all construction methods. In contrast, specificity is rather constant with a level of about 95 percent for EST and levels close to 100 percent for

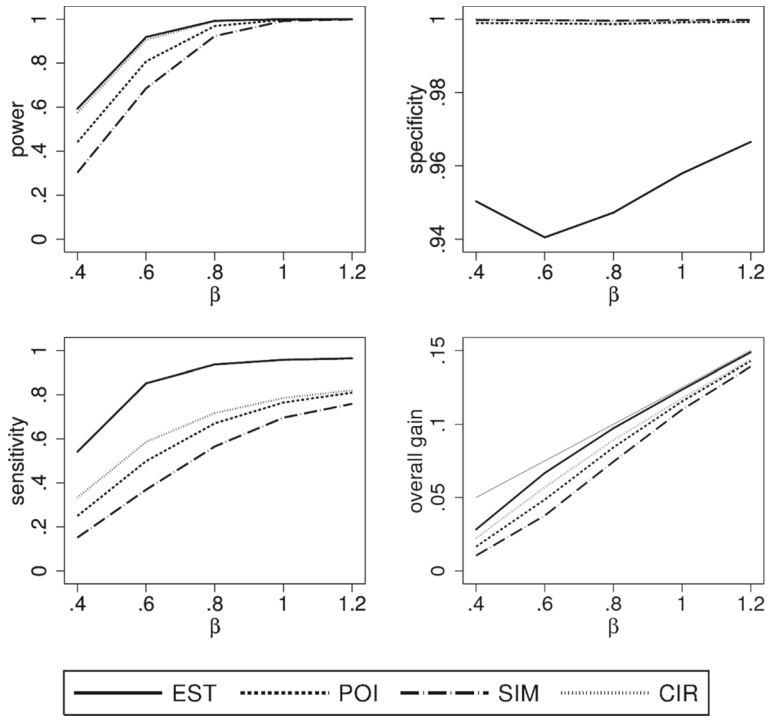

Fig. 2 Simulation results of the performance characteristics for all four construction principles as function of β. Shown is scenario 1, i.e. $\theta(x)$ linear, $X \sim \mathcal{U}(0,1)$ using a linear model for analysis. For the overall gain, the thin grey line indicates the maximally possible overall gain

the other three methods. Sensitivity of POI, SIM, CIR is smaller compared to EST. SIM, being the most conservative method, evidently has the lowest value, while the most liberal method, EST, has the highest value. Looking at the overall gain and hence balancing the opposite trends for sensitivity and specificity, EST performed best and comes close to the maximal possible gain for $\beta \geq 0.8$. Using a confidence band or confidence interval to lower the number of patients incorrectly selected for the new treatment reduces the overall gain by a small amount.

Scenario 2

When changing the distribution of X to be triangular with mode at 1/3 there are less patients with a positive treatment effect. Power is lower in this situation (Fig. 3), as $\hat{\theta}(x)$ is more variable and confidence intervals for true positive effects are larger due to fewer observations. Specificity behaves similar as in scenario 1 but sensitivity and overall gain are considerably lower. Furthermore, there are bigger differences between the construction principles. For larger values of β, the loss in sensitivity is substantially greater when going from a liberal method to a more conservative one. A distinct loss can also be seen in the overall gain. For example, for $\beta = 0.8$ more than half of the overall gain is lost when using SIM instead of EST and more than one third when using POI instead of EST. In contrast, the overall gain in EST is only about 15 percent below the maximal possible gain.

Scenario 3

Figure 4 shows the results for this scenario with a uniformly distributed X and a concave true treatment effect. The results for power and specificity are similar to the first scenario but the specificity of EST is now slightly below 95 percent. On the other hand, there is a substantial loss in sensitivity and overall gain when comparing POI, SIM, and CIR with EST. This is probably due to the fact that the positive values of the treatment effect $\theta(x)$ are closer to zero than in the linear case (cf. Fig. 1). However, it still holds that the overall gain of EST is close to the maximal possible gain if $\beta \geq 0.8$.

Scenario 4

The last scenario considers a convex true treatment effect and a uniform distribution of X. The results shown in Fig. 5 look similar to the first scenario with a linear true treatment effect. The loss in sensitivity and overall gain is minor when choosing a more conservative method instead of EST, especially when compared to the last two scenarios. This can be explained by large positive values of $\theta(x)$ for the majority of patients with $\theta(x) \geq 0$.

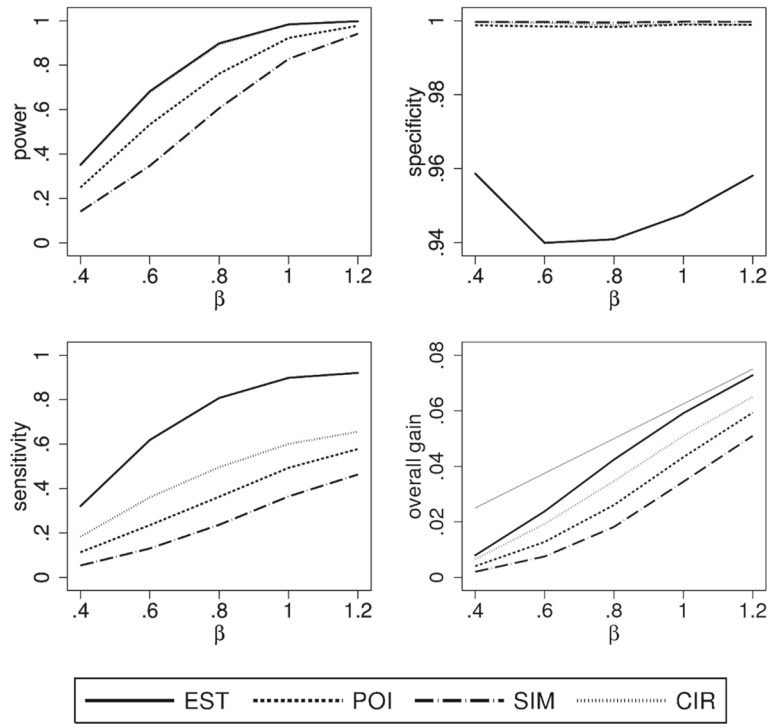

Fig. 3 Simulation results of the performance characteristics for all four construction principles as function of β. Shown is scenario 2, i.e. $\theta(x)$ linear, $X \sim \mathcal{T}(0, 1, 1/3)$ using a linear model for analysis. For the overall gain, the thin grey line indicates the maximally possible overall gain

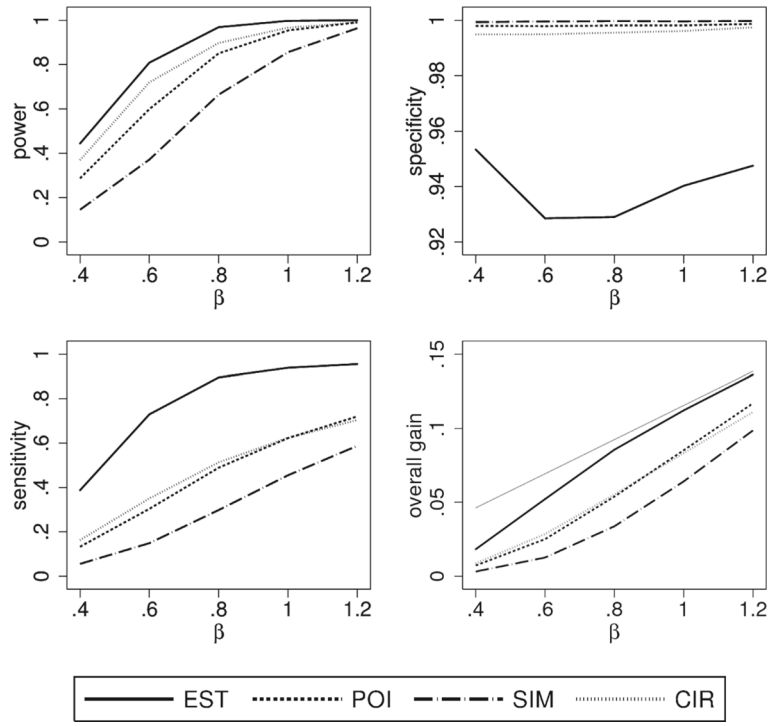

Fig. 4 Simulation results of the performance characteristics for all four construction principles as function of β. Shown is scenario 3, i.e. $\theta(x)$ concave, $X \sim \mathcal{U}(0, 1)$ using a quadratic model for analysis. For the overall gain, the thin grey line indicates the maximally possible overall gain

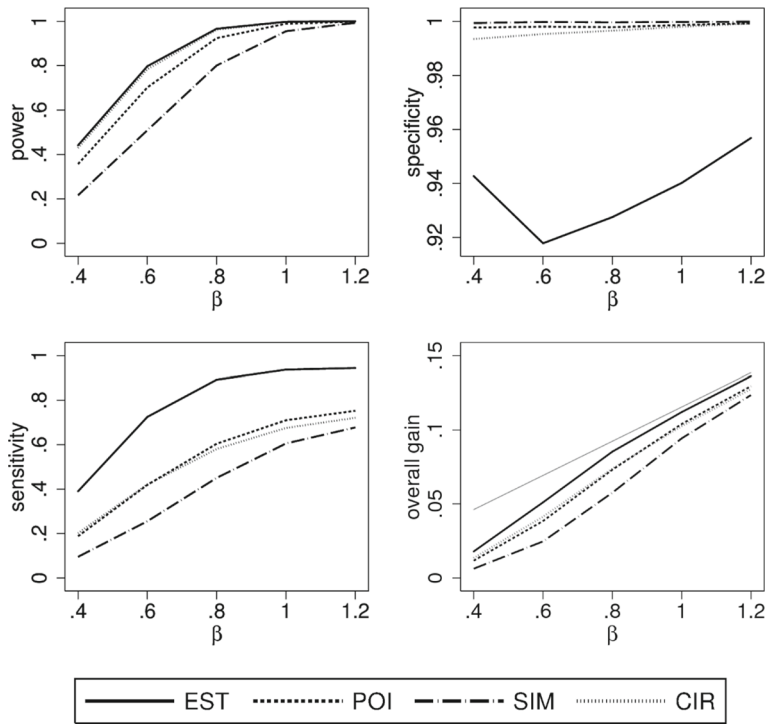

Fig. 5 Simulation results of the performance characteristics for all four construction principles as function of β. Shown is scenario 4, i.e. $\theta(x)$ convex, $X \sim \mathcal{U}(0,1)$ using a quadratic model for analysis. For the overall gain, the thin grey line indicates the maximally possible overall gain

Further results

When choosing the quadratic model for analysis in scenario 3 and 4, there may be a concern that the interaction test has little power as we test for a difference in two parameters. As we expect a monotone treatment effect it can be justified to use here also the interaction test based on the linear model. We investigated also this alternative, but the results were very similar. There may also be a concern that our results presented so far are too optimistic, as the model used to analyse the data coincides always with the true model. In Additional file 5 we present further results for misspecified models. They support the results presented so far.

Finally, we should mention that the performance characteristics between CIR and POI partially differed – in particular when using the linear analysis model – although POI can be also interpreted as a CIR approach. This indicates that using the delta method may not be very adequate. Indeed, in the linear analysis model the root is a ratio (cf. Additional file 2).

Discussion

Summary of results

The results of our simulation study indicate that using confidence bands for $\theta(x)$ or confidence intervals for $\theta^{-1}(0)$ to construct treatment selection rules are rather conservative approaches when compared to selecting just those patients with a positive treatment effect estimate. They allow to move the rate of incorrect selections in patients not benefiting from the new treatment from about 5 percent to nearly 0 percent. But we have to pay the price to overlook a substantial fraction of patients who could benefit from the new treatment. Consequently, we often obtain a substantially lower overall gain than it would be possible when just requiring positive treatment effect estimates. Actually, this simple approach allows often to approach the maximally possible gain.

Outlook

The step from modelling treatment effects as a function of a covariate to explicit construction of treatment selection rules has not yet been addressed systematically in the literature. The results of our simulation study suggest that requiring the lower bound of a 95 percent confidence interval for $\theta(x)$ to be above 0 is a very strict rule. At first sight such a rule may make sense, as in deciding whether to select patients with the biomarker value x for the new treatment, we control the probability of a type I error in these patients: If patients with this value do not benefit on average from the new treatment, the probability to select the new treatment is limited to 2.5 percent. This sounds similar to the traditional rationale in RCTs. However, in

traditional RCTs we make a decision for a large patient population. Now we make a decision for a very small patient population, namely those with a specific covariate value. So it might not be surprising that the probability of a type II error, namely to overlook the benefit from the new treatment for this small population, is actually rather large.

Such considerations may suggest to allow higher type-I error rates in order to decrease the type II error rate and hence to improve the overall gain. In Fig. 6 we consider specificity and overall gain as a function of the (1-γ)-level of the confidence bands / the confidence interval in the case of $\beta = 0.8$. We can observe a distinct increase of the overall gain when lowering $(1 - \gamma)$ from 0.95 to values around 0.8, but only a moderate decrease in specificity, keeping it at levels above 0.98 for all construction principles. This holds for all four scenarios and actually also for all values of $\beta \in \{.4, .6, .8, 1, 1.2\}$, see Additional file 6.

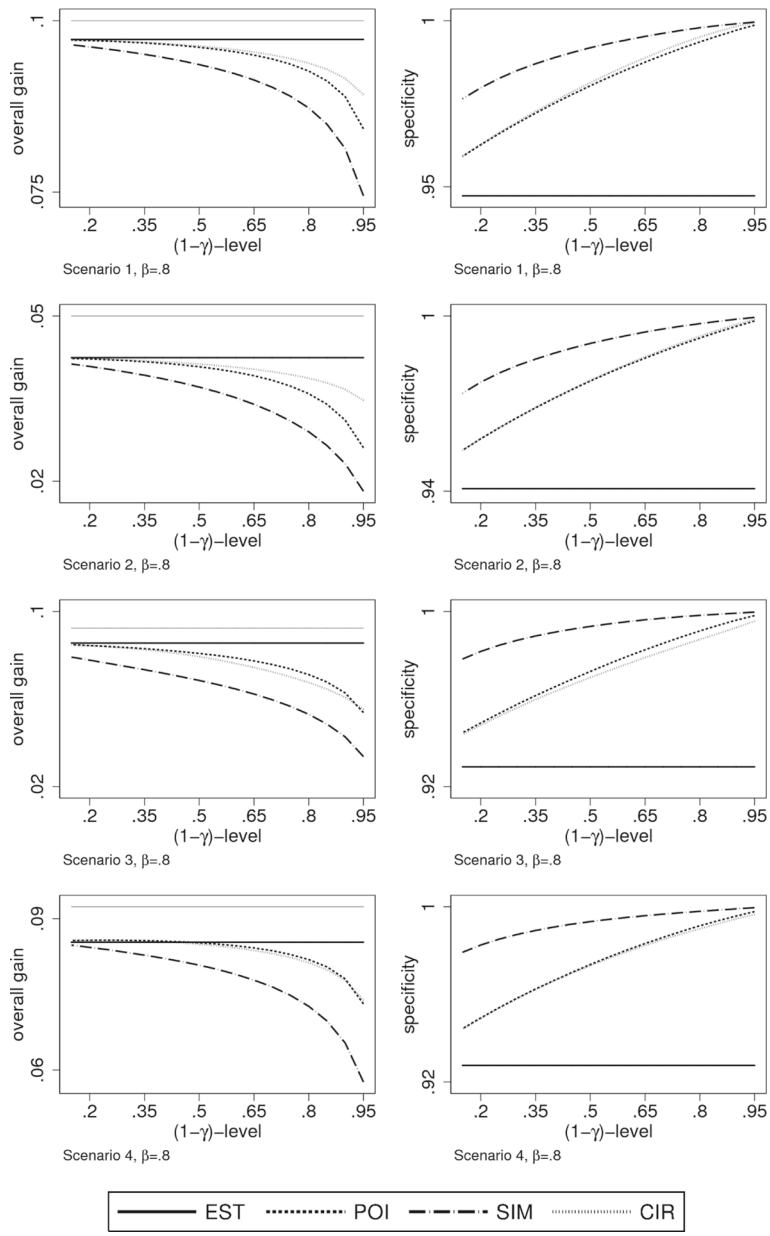

Fig. 6 Simulation results of overall gain and specificity for $\beta = 0.8$ in dependence of the $(1 - \gamma)$-level. Shown are all four scenarios individually. The first two scenarios are analysed using a linear model, the latter two using a quadratic model. For the overall gain, the thin grey line indicates the maximally possible overall gain

Limitations

Our investigation was mainly limited to the case of correctly specified models in the sense that the true model is within the class of models used in the analysis. Misspecification of the model used for the analysis has a further impact on the performance characteristics, briefly touched in Additional file 5. However, the main point we tried to make in this paper is that even in the case of a correctly specified model, there is a need to come to a consensus on how to take uncertainty in parameter estimates into account when deriving a treatment selection rule. Consequently, our focus was also on rules varying in the way to take this uncertainty into account. Further variants of the rules which may take other aspects into account were not considered. For example rules of the type $\hat{\theta}(x) > c$ for some c may aim to take the clinical relevance of the treatment effect into account. We also focused on the three specific performance characteristics sensitivity, specificity and gain, as these were sufficient to make our point. However, for a complete picture it might be necessary to take further aspects into account, for example we can define the unmet gain as the average potential benefit for patients with $\theta(x) > 0$ who are overlooked by the rule.

Future comparisons should also include methods based on selecting optimal cutpoints directly, for example those on fitting cut point models [13, 14], or using the treatment selection curve [15]. Also alternatives to simply using an interaction test as pretest [2] can have an impact on the performance. In particular such alternatives may take into account the possibility that all patients may benefit from the new treatment to a similar degree.

Conclusions

The use of 95% confidence intervals/bands in constructing treatment selection rules is a rather conservative approach. There is a need for better construction principles for treatment selection rules aiming to maximize the gain in expected outcome at the population level. Choosing a confidence level of 80% may be a first step in this direction.

Additional files

Additional file 1: An equivalence between pointwise confidence bands for $\theta(.)$ and $\theta^{-1}(.)$. (PDF 82 kb)

Additional file 2: Construction of the pointwise and simultaneous confidence bands and the confidence interval of the roots. (PDF 111 kb)

Additional file 3: Calculation of quality measures. (PDF 97 kb)

Additional file 4: Stata source code written for the simulation and data sets produced by simulation. (ZIP 42 kb)

Additional file 5: Results for the case of misspecified models. (PDF 236 kb)

Additional file 6: Simulation results for overall gain and specificity in dependence of the $(1-\gamma)$-level for various values of β. (PDF 492 kb)

Abbreviations

CIR: Treatment selection rule using a 95 percent confidence interval of all roots; EST: Treatment selection rule using the estimator; POI: Treatment selection rule using a 95 percent pointwise confidence band; RCT: Randomized control trial; SIM: Treatment selection rule using a 95 percent simultaneous confidence band

Acknowledgments

We are grateful to Martin Schumacher for acting as supervisor of the thesis and to Ludger Rüschendorf for acting as reviewer of the thesis.

Authors' contributions

ME and WV developed the idea for this study and defined the performance characteristics. ME performed the programming and the conduct of the simulation study. ME prepared a draft version of the manuscript. Both authors worked on the final version of the manuscript, approved it and guarantee the accuracy and integrity of the study.

Authors' information

This article is an extended version of the diploma thesis of the first author.

Funding

The article processing charge was funded by the University of Freiburg in the funding programme Open Access Publishing. The funder had no influence on the study and the presentation of results.

Availability of data and materials

All Stata code used and data sets generated by the simulation are provided as Additional file 4.

Ethics approval and consent to participate

Not applicable. The article does not report or involve the case of any animal or human data or tissue.

Consent for publication

Not applicable. The article does not contain data from any individual person.

Competing interests

The authors declare that they have no competing interests.

Author details

[1]Institute of Medical Biometry and Statistics, Section of Health Care Research and Rehabilitation Research, Faculty of Medicine and Medical Center - University of Freiburg, Hebelstr. 11, 79104 Freiburg, Germany. [2]Department of Orthopaedics and Traumatology, University Hospital Basel, Spitalstr. 21, CH-4031 Basel, Switzerland.

Received: 28 June 2018 Accepted: 15 July 2019
Published online: 01 August 2019

References

1. Mackey HM, Bengtsson T. Sample size and threshold estimation for clinical trials with predictive biomarkers. Contemp Clin Trials. 2013;36:664–72.
2. Janes H, Brown MD, Huang Y, Pepe MS. An approach to evaluating and comparing biomarkers for patient treatment selection. Int J Biostat. 2014;10(1):99–121.
3. Riddell CA, Zhao Y, Petkau J. An adaptive clinical trials procedure for a sensitive subgroup examined in the multiple sclerosis context. Stat Methods Med Res. 2016;25(4):1330–45.
4. Bonetti M, Gelber RD. Patterns of treatment effects in subsets of patients in clinical trials. Biostatistics. 2004;5:465–81.
5. Royston P, Sauerbrei W. A new approach to modelling interactions between treatment and continuous covariates in clinical trials by using fractional polynomials. Stat Med. 2004;23:2509–25.
6. Jeong JH, Costantino JP. Application of smoothing methods to evaluate treatment-prognostic factor interactions in breast cancer data. Cancer Investig. 2006;24:288–93.
7. Cai T, Tian L, Wong PH, Wei LJ. Analysis of randomized comparative clinical trial data for personalized treatment selections. Biostatistics. 2011;12:270–82.

8. Shen Y, Le LD, Wilson R, Mansmann U. Graphical presentation of patient-treatment interaction elucidated by continuous biomarkers. Methods Inf Med. 2017;56(1):13–27.
9. Baker SG, Bonetti M. Evaluating Markers for Guiding Treatment. J Natl Cancer Inst. 2016;108(9):djw101.
10. Baker SG, Kramer BS. Evaluating surrogate endpoints, prognostic markers, and predictive markers: Some simple themes. Clin Trials. 2015;12(4): 299–308.
11. Ma Y, Zhou X. Treatment selection in a randomized clinical trial via covariate-specific treatment effect curves. Stat Methods Med Res. 2017;26(1):124–41.
12. Sun H, Vach W. A framework to assess the value of subgroup analyses when the overall treatment effect is significant. J Biopharm Stat. 2016;26(3):565–78.
13. Chen BE, Jiang W, Tu D. A hierarchical bayes model for biomarker subset effects in clinical trials. Comput Stat Data Anal. 2014;71:324–34.
14. Jiang W, Freidlin B, Simon R. Biomarker-adaptive threshold design: A procedure for evaluating treatment with possible biomarker-defined subset effect. J Natl Cancer Inst. 2007;99:1036–43.
15. Song X, Pepe MS. Evaluating markers for selecting a patient's treatment. Biometrics. 2004;60:874–83.

Publisher's Note