

Protocol

1 Introduction

Risk-based HTE analyses typically result in risk stratum specific estimates of treatment effects. Different approaches to individualize treatment effects have been proposed, but it is unclear which approach performs best under different circumstances. We aim to compare different approaches to estimating individualized treatment effects.

For all patients we observe covariates x_1, \dots, x_8 , of which 4 are continuous and 4 are binary. More specifically,

$$\begin{aligned}x_1, \dots, x_4 &\sim N(0, 1) \\ x_5, \dots, x_8 &\sim B(1, 0.2)\end{aligned}$$

We first, generate the binary outcomes y for the untreated patients ($t_x = 0$), based on

$$P(y \mid \mathbf{x}, t_x = 0) = g(\beta_0 + \beta_1 x_1 + \dots + \beta_8 x_8) = g(lp_0), \quad (1)$$

where

$$g(x) = \frac{e^x}{1 + e^x}$$

For treated patients, outcomes are generated from:

$$P(y \mid \mathbf{x}, t_x = 1) = g(lp_1) \quad (2)$$

where

$$lp_1 = \gamma_2(lp_0 - c)^2 + \gamma_1(lp_0 - c) + \gamma_0$$

2 Settings

2.1 Base-case scenario

The base-case scenario assumes a constant odds ratio of 0.8 in favor of treatment. The simulated datasets are of size $n = 4250$, where treatment is allocated at random using a 50/50 split (80% power for the detection of an unadjusted OR of 0.8, assuming an event rate of 20% in the untreated arm). Outcome incidence in the untreated population is set at 20%. For the development of the prediction model we use the model defined in (1) including a constant treatment effect. When doing predictions, t_x is set to 0. The value of the true β is such that the above prediction model has an AUC of 0.75.

The previously defined targets are achieved when $\beta = (-2.08, 0.49, \dots, 0.49)^t$. For the derivations in the treatment arm we use $\gamma = (\log(0.8), 1, 0)^t$.

2.2 Deviations from base-case

We deviate from the base-case scenario in two ways. First, we alter the overall target settings of sample size, overall treatment effect and prediction model AUC. In a second stage, we consider settings that violate the assumption of a constant relative treatment effect, using a model-based approach.

For the first part, we consider:

- Sample size:
 - $n = 1064$
 - $n = 17000$
- Overall treatment effect:
 - $OR = 0.5$
 - $OR = 1$
- Prediction performance:
 - $AUC = 0.65$
 - $AUC = 0.85$

We set the true risk model coefficients to be $\beta = (-1.63, 0.26, \dots, 0.26)^t$ for $AUC = 0.65$ and $\beta = (-2.7, 0.82, \dots, 0.82)^t$ for $AUC = 0.85$. In both cases, β_0 is selected so that an event rate of 20% is maintained in the control arm.

For the second part linear and quadratic deviations from the assumption of constant relative effect are considered. We also consider different intensity levels of these deviations. The settings for these deviations are defined in supplementary Table ?? and result in the effects of Figure 1.

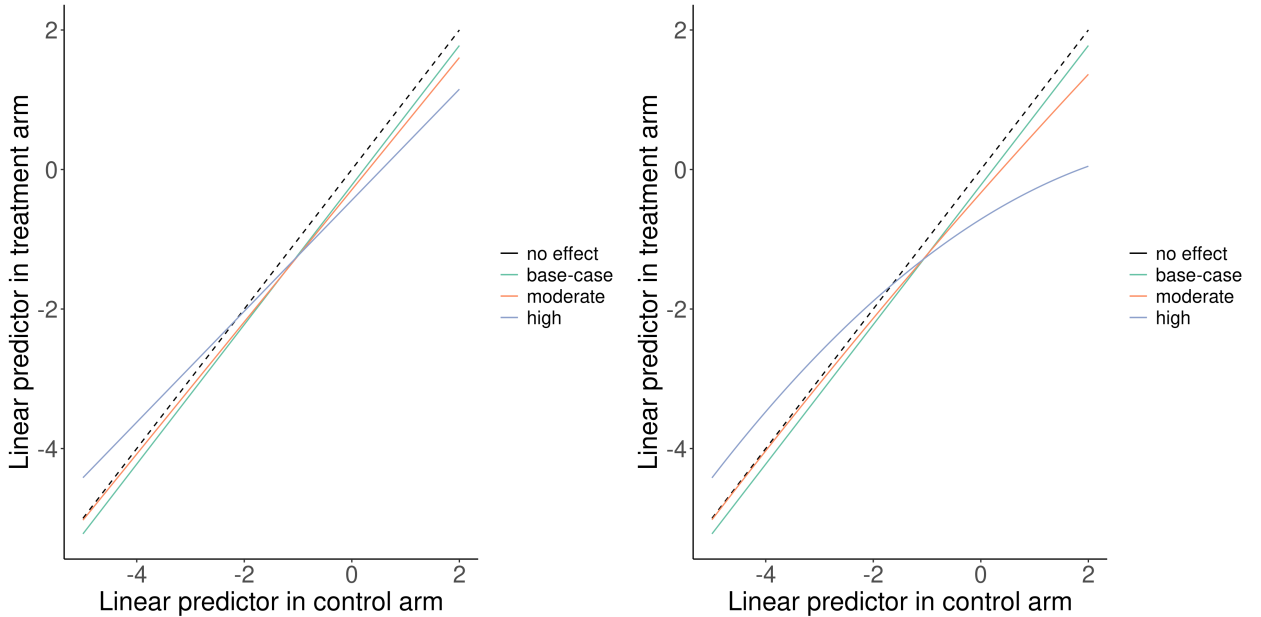


Figure 1: Linear and quadratic deviations from the base-case scenario of constant relative effect ($OR=0.8$)

In Figure 2 the absolute benefits observed based on different settings are presented. The base-case scenario is also presented as a reference.

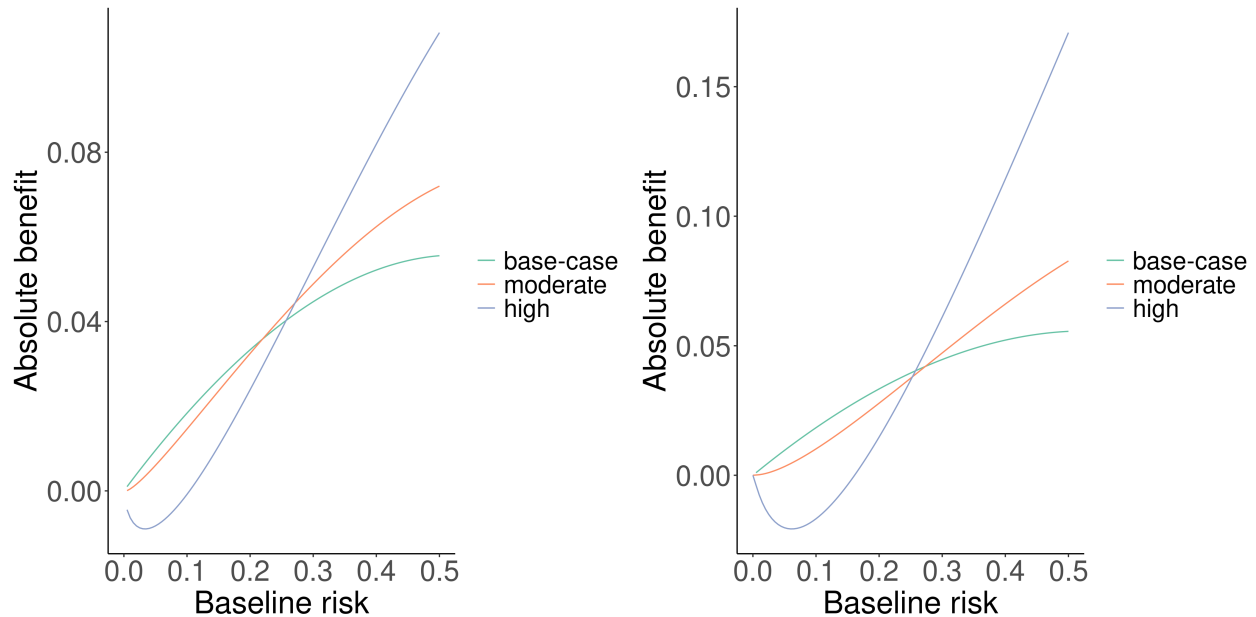


Figure 2: Linear and quadratic deviations from the base-case scenario of constant relative effect (OR=0.8)

Finally, we consider 3 additional scenarios of interaction of individual covariates with treatment. These scenarios include a 4 weak interactions ($OR_{t_x=1}/OR_{t_x=0} = 0.82$), 4 strong interactions ($OR_{t_x=1}/OR_{t_x=0} = 0.61$), and 2 weak and 2 strong interactions (Table 2).

3 Risk modeling

Merging treatment arms, we develop prediction models including a constant relative treatment effect:

$$E\{y \mid x, t_x\} = P(y \mid x, t_x) = g(\beta_0 + \beta_1 x_1 + \dots + \beta_8 x_8 + \gamma t_x) \quad (3)$$

Individualized predictions are derived setting $t_x = 0$.

4 Approaches to individualize benefit predictions

4.1 Risk stratification

Derive a prediction model using the same approach as above and divide the population in equally sized risk-based subgroups. Estimate subgroup-specific absolute benefit from the observed absolute differences. Subject-specific benefit predictions are made by attributing to individuals their corresponding subgroup-specific estimate.

4.2 Constant treatment effect

Assuming a constant relative treatment effect, fit the adjusted model in (3). Then, an estimate of absolute benefit can be derived from

$$\hat{f}_{\text{benefit}}(lp \mid \mathbf{x}, \hat{\beta}) = g(lp) - g(lp + \hat{\gamma})$$

4.3 Linear interaction

The assumption of constant relative treatment effect is relaxed modeling a linear interaction of treatment with the risk linear predictor:

$$E\{y \mid \mathbf{x}, t_x, \hat{\beta}\} = g(lp + (\delta_0 + \delta_1 lp)t_x)$$

We predict absolute benefit from

$$\hat{f}_{\text{benefit}}(lp \mid \mathbf{x}, \hat{\beta}) = g(lp) - g(\delta_0 + (1 + \delta_1)lp)$$

4.4 Non-linear interaction

Finally, we drop the linearity assumption and predict absolute benefit by taking the difference between smooth fits, separately derived in each treatment arm:

$$f_{\text{benefit}}(lp \mid \mathbf{x}, \hat{\beta}) = \hat{f}_{\text{smooth}}(lp \mid \mathbf{x}, \hat{\beta}, t_x = 0) - \hat{f}_{\text{smooth}}(lp \mid \mathbf{x}, \hat{\beta}, t_x = 1)$$

We consider three different approaches to smoothing:

- Loess
- Restricted cubic splines
- Local likelihood

5 Evaluation

Assuming that $\tau(\mathbf{x}) = E\{y \mid \mathbf{x}, t_x = 0\} - E\{y \mid \mathbf{x}, t_x = 1\}$ is the true benefit for each patient and $\hat{\tau}(\mathbf{x})$ is the estimated benefit from a method under study, the ideal loss function to use for the considered methods would be the unobservable root mean squared error $E\{(\hat{\tau} - \tau)^2 \mid \mathbf{x}\}$. However, in simulations the true patient-level benefit is available. Therefore, we will estimate the RMSE from

$$\text{RMSE} = \frac{1}{n} \sum_{i=1}^n (\tau(\mathbf{x}_i) - \hat{\tau}(\mathbf{x}_i))^2$$

We also compare the predictive performance of the methods under study. We assess discrimination using the c-for-benefit statistic described in [1]. Patients in each treatment arm are ranked based on their predicted benefit and then are matched 1:1, dropping patients in the larger treatment arm without a pair. We define *observed* pair-specific treatment benefit as the difference of observed outcomes between the untreated and the treated patient of each pair. Pair-specific *predicted* benefit is defined as the average of predicted benefits within each pair. Then, c-for-benefit is defined as the probability that from two randomly chosen matched patient pairs with unequal observed benefit, the pair with greater observed benefit also has a higher predicted benefit.

We evaluate calibration in a similar manner, using the integrated calibration index (ICI) for benefit [2]. After creating pairs based on predicted benefit, observed benefits are regressed on the predicted benefits using a locally weighted scatterplot smoother (loess). The ICI is the area between the loess fit and the diagonal, with values closer to 0 indicating better calibration.

In all cases, the evaluation metrics are calculated in a large simulated super-population of 500,000 patients.

6 References

- 1 Klaveren D van, Steyerberg EW, Serruys PW *et al.* The proposed “concordance-statistic for benefit” provided a useful metric when modeling heterogeneous treatment effects. *Journal of Clinical Epidemiology* 2018;**94**:59–68. doi:10.1016/j.jclinepi.2017.10.021
- 2 Austin PC, Steyerberg EW. The integrated calibration index (ICI) and related metrics for quantifying the calibration of logistic regression models. *Statistics in Medicine* 2019;**38**:4051–65. doi:10.1002/sim.8281

7 Supplement

Analysis ID				Baseline risk									True treatment effect			
Scenario	Effect	N	AUC	b0	b1	b2	b3	b4	b5	b6	b7	b8	g0	g1	g2	c
1	absent	4,250	75	-2.08	0.49	0.49	0.49	0.49	0.49	0.49	0.49	0.49	0.00	1.00	0.00	0
2	absent	4,250	65	-1.63	0.26	0.26	0.26	0.26	0.26	0.26	0.26	0.26	0.00	1.00	0.00	0
3	absent	4,250	85	-2.70	0.82	0.82	0.82	0.82	0.82	0.82	0.82	0.82	0.00	1.00	0.00	0
4	absent	1,064	75	-2.08	0.49	0.49	0.49	0.49	0.49	0.49	0.49	0.49	0.00	1.00	0.00	0
5	absent	1,064	65	-1.63	0.26	0.26	0.26	0.26	0.26	0.26	0.26	0.26	0.00	1.00	0.00	0
6	absent	1,064	85	-2.70	0.82	0.82	0.82	0.82	0.82	0.82	0.82	0.82	0.00	1.00	0.00	0
7	absent	17,000	75	-2.08	0.49	0.49	0.49	0.49	0.49	0.49	0.49	0.49	0.00	1.00	0.00	0
8	absent	17,000	65	-1.63	0.26	0.26	0.26	0.26	0.26	0.26	0.26	0.26	0.00	1.00	0.00	0
9	absent	17,000	85	-2.70	0.82	0.82	0.82	0.82	0.82	0.82	0.82	0.82	0.00	1.00	0.00	0
<i>Constant treatment effect</i>																
10	moderate	4,250	75	-2.08	0.49	0.49	0.49	0.49	0.49	0.49	0.49	0.49	-0.22	1.00	0.00	0
11	moderate	4,250	65	-1.63	0.26	0.26	0.26	0.26	0.26	0.26	0.26	0.26	-0.22	1.00	0.00	0
12	moderate	4,250	85	-2.70	0.82	0.82	0.82	0.82	0.82	0.82	0.82	0.82	-0.22	1.00	0.00	0
13	moderate	1,064	75	-2.08	0.49	0.49	0.49	0.49	0.49	0.49	0.49	0.49	-0.22	1.00	0.00	0
14	moderate	1,064	65	-1.63	0.26	0.26	0.26	0.26	0.26	0.26	0.26	0.26	-0.22	1.00	0.00	0
15	moderate	1,064	85	-2.70	0.82	0.82	0.82	0.82	0.82	0.82	0.82	0.82	-0.22	1.00	0.00	0
16	moderate	17,000	75	-2.08	0.49	0.49	0.49	0.49	0.49	0.49	0.49	0.49	-0.22	1.00	0.00	0
17	moderate	17,000	65	-1.63	0.26	0.26	0.26	0.26	0.26	0.26	0.26	0.26	-0.22	1.00	0.00	0
18	moderate	17,000	85	-2.70	0.82	0.82	0.82	0.82	0.82	0.82	0.82	0.82	-0.22	1.00	0.00	0
19	high	4,250	75	-2.08	0.49	0.49	0.49	0.49	0.49	0.49	0.49	0.49	-0.69	1.00	0.00	0
20	high	4,250	65	-1.63	0.26	0.26	0.26	0.26	0.26	0.26	0.26	0.26	-0.69	1.00	0.00	0
21	high	4,250	85	-2.70	0.82	0.82	0.82	0.82	0.82	0.82	0.82	0.82	-0.69	1.00	0.00	0
22	high	1,064	75	-2.08	0.49	0.49	0.49	0.49	0.49	0.49	0.49	0.49	-0.69	1.00	0.00	0
23	high	1,064	65	-1.63	0.26	0.26	0.26	0.26	0.26	0.26	0.26	0.26	-0.69	1.00	0.00	0
24	high	1,064	85	-2.70	0.82	0.82	0.82	0.82	0.82	0.82	0.82	0.82	-0.69	1.00	0.00	0
25	high	17,000	75	-2.08	0.49	0.49	0.49	0.49	0.49	0.49	0.49	0.49	-0.69	1.00	0.00	0
26	high	17,000	65	-1.63	0.26	0.26	0.26	0.26	0.26	0.26	0.26	0.26	-0.69	1.00	0.00	0
27	high	17,000	85	-2.70	0.82	0.82	0.82	0.82	0.82	0.82	0.82	0.82	-0.69	1.00	0.00	0
<i>Linear deviation</i>																
28	moderate	4,250	75	-2.08	0.49	0.49	0.49	0.49	0.49	0.49	0.49	0.49	-0.29	0.95	0.00	0
29	moderate	4,250	65	-1.63	0.26	0.26	0.26	0.26	0.26	0.26	0.26	0.26	-0.35	0.93	0.00	0
30	moderate	4,250	85	-2.70	0.82	0.82	0.82	0.82	0.82	0.82	0.82	0.82	-0.37	0.93	0.00	0
31	moderate	1,064	75	-2.08	0.49	0.49	0.49	0.49	0.49	0.49	0.49	0.49	-0.29	0.95	0.00	0
32	moderate	1,064	65	-1.63	0.26	0.26	0.26	0.26	0.26	0.26	0.26	0.26	-0.35	0.93	0.00	0
33	moderate	1,064	85	-2.70	0.82	0.82	0.82	0.82	0.82	0.82	0.82	0.82	-0.37	0.93	0.00	0

34	moderate	17,000	75	-2.08	0.49	0.49	0.49	0.49	0.49	0.49	0.49	0.49	0.49	-0.29	0.95	0.00	0
35	moderate	17,000	65	-1.63	0.26	0.26	0.26	0.26	0.26	0.26	0.26	0.26	0.26	-0.35	0.93	0.00	0
36	moderate	17,000	85	-2.70	0.82	0.82	0.82	0.82	0.82	0.82	0.82	0.82	0.82	-0.37	0.93	0.00	0
37	high	4,250	75	-2.08	0.49	0.49	0.49	0.49	0.49	0.49	0.49	0.49	0.49	-0.44	0.80	0.00	0
38	high	4,250	65	-1.63	0.26	0.26	0.26	0.26	0.26	0.26	0.26	0.26	0.26	-0.54	0.78	0.00	0
39	high	4,250	85	-2.70	0.82	0.82	0.82	0.82	0.82	0.82	0.82	0.82	0.82	-0.50	0.78	0.00	0
40	high	1,064	75	-2.08	0.49	0.49	0.49	0.49	0.49	0.49	0.49	0.49	0.49	-0.44	0.80	0.00	0
41	high	1,064	65	-1.63	0.26	0.26	0.26	0.26	0.26	0.26	0.26	0.26	0.26	-0.54	0.78	0.00	0
42	high	1,064	85	-2.70	0.82	0.82	0.82	0.82	0.82	0.82	0.82	0.82	0.82	-0.50	0.78	0.00	0
43	high	17,000	75	-2.08	0.49	0.49	0.49	0.49	0.49	0.49	0.49	0.49	0.49	-0.44	0.80	0.00	0
44	high	17,000	65	-1.63	0.26	0.26	0.26	0.26	0.26	0.26	0.26	0.26	0.26	-0.54	0.78	0.00	0
45	high	17,000	85	-2.70	0.82	0.82	0.82	0.82	0.82	0.82	0.82	0.82	0.82	-0.50	0.78	0.00	0

Quadratic deviation

46	moderate	4,250	75	-2.08	0.49	0.49	0.49	0.49	0.49	0.49	0.49	0.49	0.49	-5.02	1.00	-0.01	-5
47	moderate	4,250	65	-1.63	0.26	0.26	0.26	0.26	0.26	0.26	0.26	0.26	0.26	-5.02	1.00	-0.02	-5
48	moderate	4,250	85	-2.70	0.82	0.82	0.82	0.82	0.82	0.82	0.82	0.82	0.82	-5.02	1.00	-0.02	-5
49	moderate	1,064	75	-2.08	0.49	0.49	0.49	0.49	0.49	0.49	0.49	0.49	0.49	-5.02	1.00	-0.01	-5
50	moderate	1,064	65	-1.63	0.26	0.26	0.26	0.26	0.26	0.26	0.26	0.26	0.26	-5.02	1.00	-0.02	-5
51	moderate	1,064	85	-2.70	0.82	0.82	0.82	0.82	0.82	0.82	0.82	0.82	0.82	-5.02	1.00	-0.02	-5
52	moderate	17,000	75	-2.08	0.49	0.49	0.49	0.49	0.49	0.49	0.49	0.49	0.49	-5.02	1.00	-0.01	-5
53	moderate	17,000	65	-1.63	0.26	0.26	0.26	0.26	0.26	0.26	0.26	0.26	0.26	-5.02	1.00	-0.02	-5
54	moderate	17,000	85	-2.70	0.82	0.82	0.82	0.82	0.82	0.82	0.82	0.82	0.82	-5.02	1.00	-0.02	-5
55	high	4,250	75	-2.08	0.49	0.49	0.49	0.49	0.49	0.49	0.49	0.49	0.49	-4.42	1.00	-0.05	-5
56	high	4,250	65	-1.63	0.26	0.26	0.26	0.26	0.26	0.26	0.26	0.26	0.26	-4.42	1.00	-0.06	-5
57	high	4,250	85	-2.70	0.82	0.82	0.82	0.82	0.82	0.82	0.82	0.82	0.82	-4.42	1.00	-0.06	-5
58	high	1,064	75	-2.08	0.49	0.49	0.49	0.49	0.49	0.49	0.49	0.49	0.49	-4.42	1.00	-0.05	-5
59	high	1,064	65	-1.63	0.26	0.26	0.26	0.26	0.26	0.26	0.26	0.26	0.26	-4.42	1.00	-0.06	-5
60	high	1,064	85	-2.70	0.82	0.82	0.82	0.82	0.82	0.82	0.82	0.82	0.82	-4.42	1.00	-0.06	-5
61	high	17,000	75	-2.08	0.49	0.49	0.49	0.49	0.49	0.49	0.49	0.49	0.49	-4.42	1.00	-0.05	-5
62	high	17,000	65	-1.63	0.26	0.26	0.26	0.26	0.26	0.26	0.26	0.26	0.26	-4.42	1.00	-0.06	-5
63	high	17,000	85	-2.70	0.82	0.82	0.82	0.82	0.82	0.82	0.82	0.82	0.82	-4.42	1.00	-0.06	-5

Table 2:

Analysis ID				Baseline risk										Coefficient in treatment arm			
Scenario	Effect	N	AUC	b0	b1	b2	b3	b4	b5	b6	b7	b8	g1	g2	g5	g6	
64	weak	4,250	75	-2.08	0.49	0.49	0.49	0.49	0.49	0.49	0.49	0.49	-0.19	-0.19	-0.19	-0.19	
65	strong	4,250	75	-2.08	0.49	0.49	0.49	0.49	0.49	0.49	0.49	0.49	-0.49	-0.49	-0.49	-0.49	
66	mixed	4,250	75	-2.08	0.49	0.49	0.49	0.49	0.49	0.49	0.49	0.49	-0.19	-0.49	-0.19	-0.49	