Supplementary material

1 Simulation settings

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

$$x_1, \dots, x_4 \sim N(0, 1)$$
$$x_5, \dots, x_8 \sim B(1, 0.2)$$

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

$$P(y \mid \boldsymbol{x}, t_x = 0) = g(\beta_0 + \beta_1 x_1 + \cdots + \beta_8 x_8) = g(lp_0), \tag{1}$$

where

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

For treated patients, outcomes are generated from:

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

where

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

1.1 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 $\boldsymbol{\beta}$ is such that the above prediction model has an AUC of 0.75.

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

1.2 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 $\boldsymbol{\beta} = \big(-1.63, 0.26, \dots, 0.26\big)^t$ for $AUC = 0.65$ and $\boldsymbol{\beta} = \big(-2.7, 0.82, \dots, 0.82\big)^t$ 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 Table 1 and result in the effects of Figure 1.

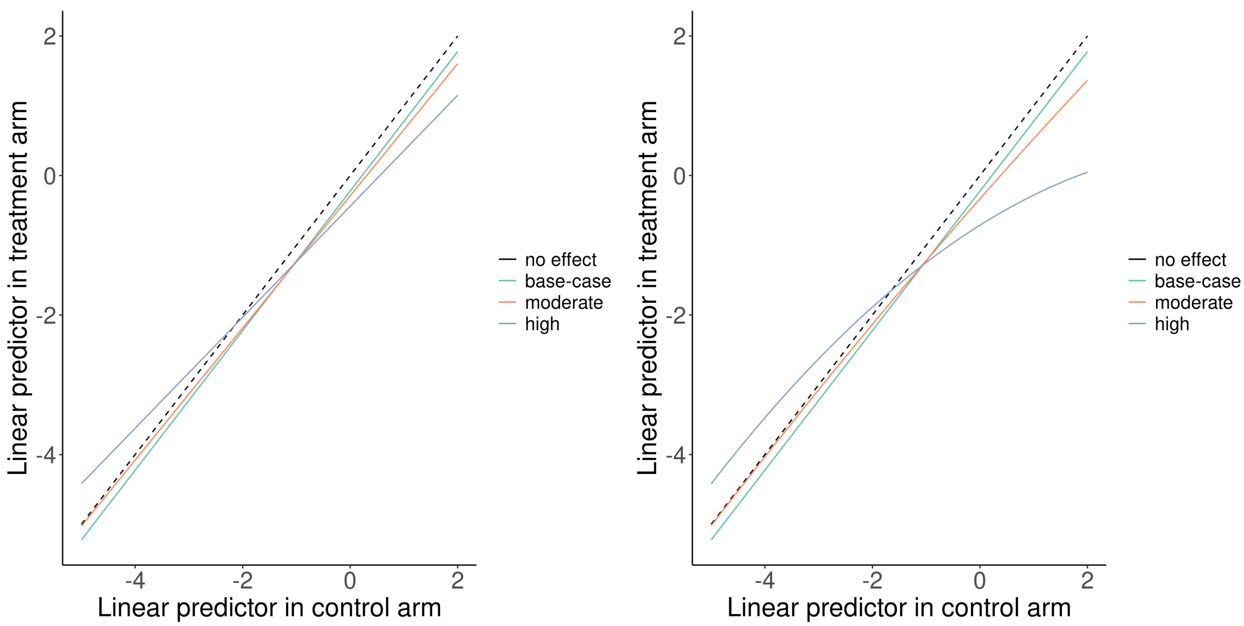

Figure 1: Linear and quadratic deviations from the base-case scenario of constant relative effect (OR=0.8)

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

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

1.3 Risk modeling

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

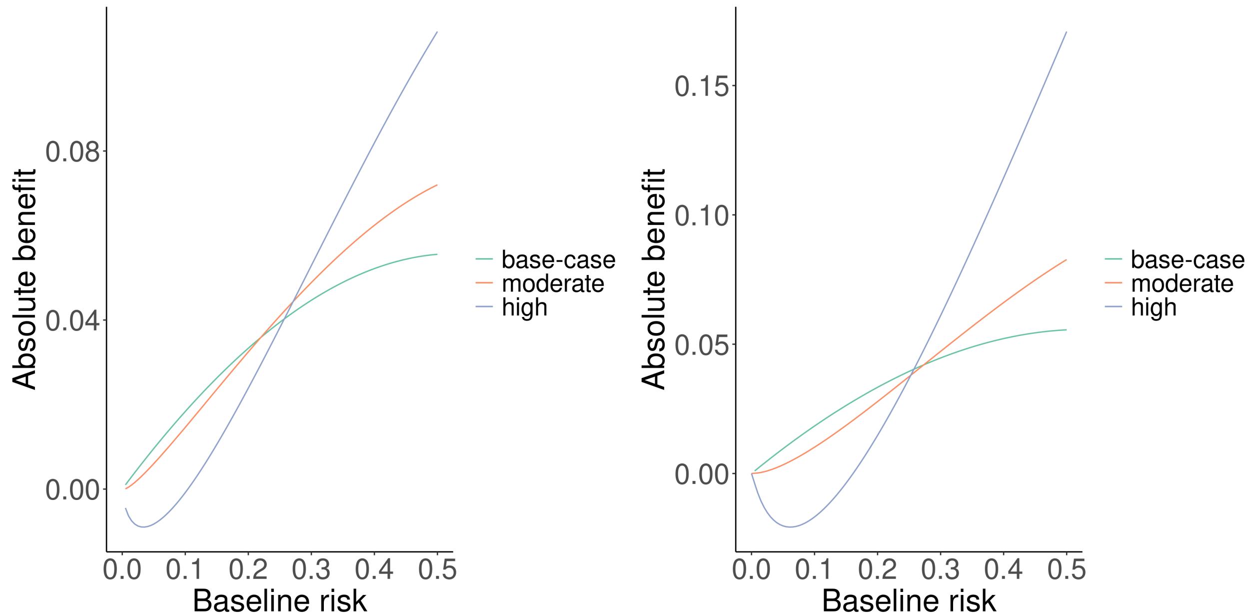

Figure 2: Linear and quadratic deviations from the base-case scenario of constant relative effect (OR=0.8)

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

Individualized predictions are derived setting $t_x = 0$.

1.4 Approaches to individualize benefit predictions

1.4.1 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.

1.4.2 Constant treatment effect

Assuming a constant relative treatment effect, fit the adjusted model in (1.3). Then, an estimate of absolute benefit can be derived from

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

1.4.3 Linear interaction

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

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

We predict absolute benefit from

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

1.4.4 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:

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

Table 1:

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
Constant treatment effect																
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
Linear deviation																
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	-0.29	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	-0.35	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	-0.37	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	-0.44	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	-0.54	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	-0.50	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	-0.44	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	-0.54	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	-0.50	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	-0.44	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	-0.54	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	-0.50	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	-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	-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	-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	-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	-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	-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	-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	-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	-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	-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	-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	-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	-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	-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	-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	-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	-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	-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

2 Results in scenarios with interactions

When we considered a set of 4 true linear treatment-covariate interactions the model containing a linear interaction with the prognostic index had the lowest median RMSE. We observed an increasing trend in prediction errors with increasing interaction intensity (Figure 3). The model with restricted cubic spline smoothing (3 knots) had very comparable performance to the linear interaction model. Increasing the flexibility of the smooth methods resulted in increasing median RMSE. These results may be explained by the fact that the interactions considered were linear, thus favoring the linear interaction model. More flexible approaches may be better suited for higher-order treatment-covariate interactions. Finally, the adaptive approach had adequate performance under all scenarios, resembling the performance of the best-performing approach every time.

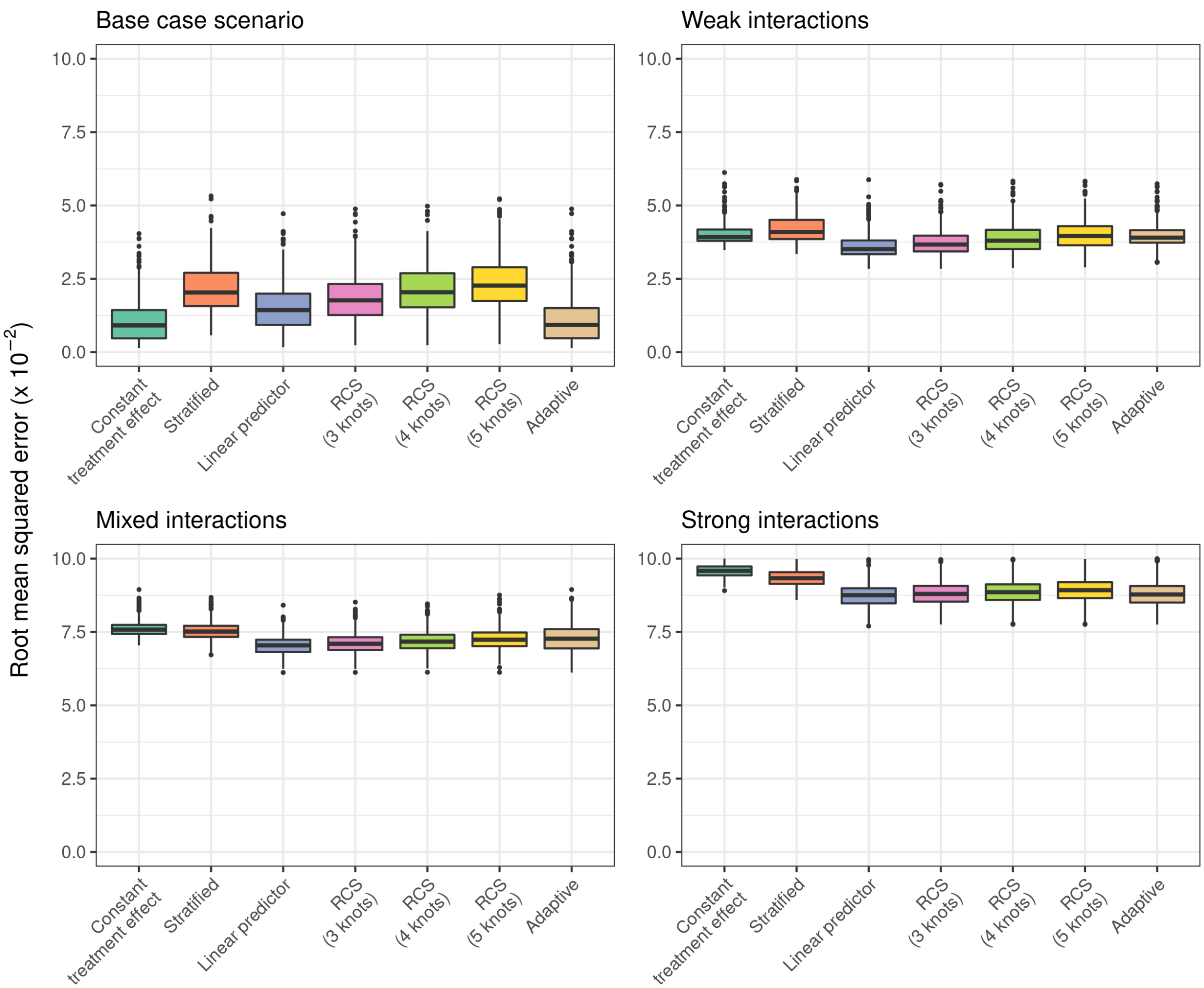

Figure 3: Linear and quadratic deviations from the base-case scenario of constant relative effect (OR=0.8)

The constant treatment effects model, the linear interaction model and the model with RCS smoothing (3 knots) had the highest c-for-benefit across all scenarios (Figrue 4). RCS smoothing with 4 or 5 knots did not improve performance. On the contrary, we observed an increasing trend in c-for-benefit variability, as was the case in the main text. The adaptive apprach again had statisfactory performance.

Despite the very similar performance in terms of prediction errors, the linear interaction model resulted

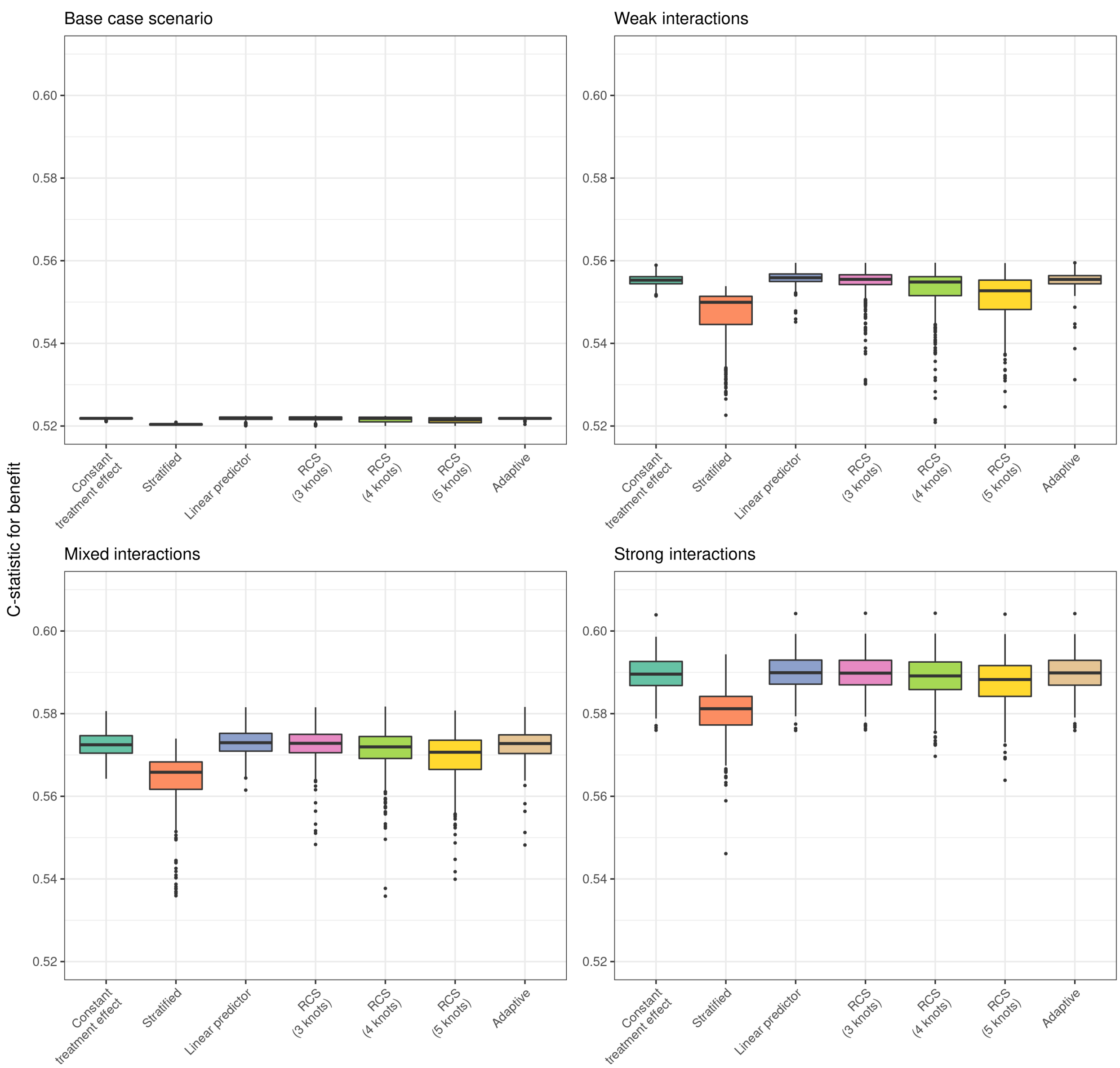

Figure 4: Linear and quadratic deviations from the base-case scenario of constant relative effect (OR=0.8)

in better-calibrated benefit predictions compared to the rest of the methods (Figure 5). The constant treatment effects model had the highest median ICI-for-benefit across all scenario settings, which became more pornounced with increasing treatment-covariate interaction intensity.

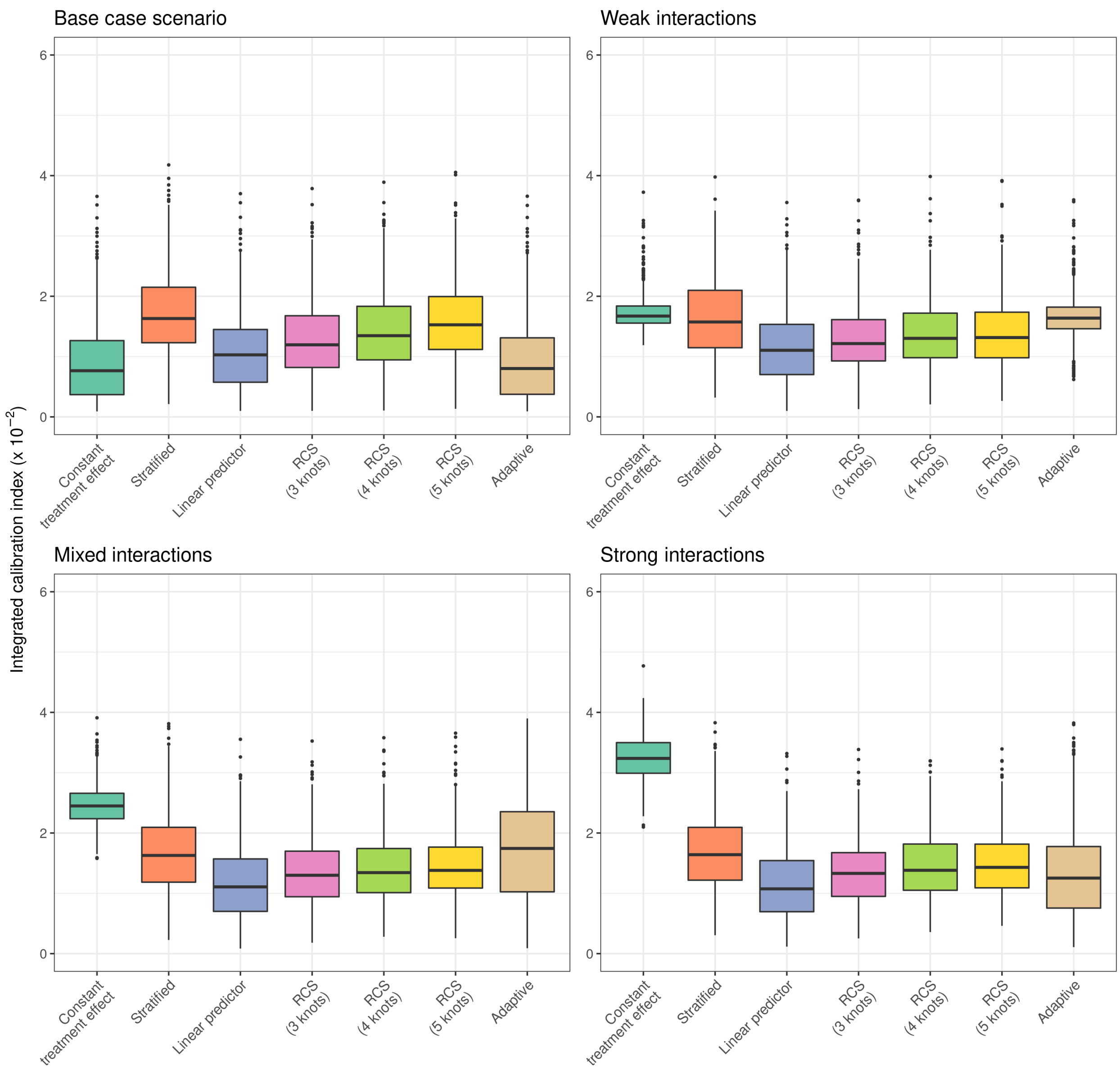

Figure 5: Linear and quadratic deviations from the base-case scenario of constant relative effect (OR=0.8)