

Supplementary material

Contents

1 Simulation settings	1
1.1 Base-case scenario	2
1.2 Deviations from base-case	2
1.3 Risk modeling	2
1.4 Approaches to individualize benefit predictions	3
2 Adaptive model selection frequencies	19
3 Discrimination and calibration for benefit	22
4 Strong relative treatment effect	26
5 Treatment interactions	26
6 Empirical illustration	31

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,

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

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

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

where

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

For treated patients, outcomes are generated from:

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

where

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

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 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 $\gamma = (\log(0.8), 1, 0)^t$.

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 $\beta = (-1.63, 0.26, \dots, 0.26)^t$ 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, quadratic and non-monotonic deviations from the assumption of constant relative effect are considered. We also consider different intensity levels of these deviations. Finally, constant absolute treatment-related harms are introduced, i.e. positive ($0.25 \times$ true average benefit), strong positive ($0.50 \times$ true average benefit) and negative ($-0.25 \times$ true average benefit; i.e. constant absolute treatment-related benefit). In case of true absent treatment effects, treatment-related harms are set to 1%, 2% and -1% for positive, strong positive and negative setting, respectively. The settings for these deviations are defined in Table S1.

1.3 Risk modeling

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

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

Individualized predictions are derived setting $t_x = 0$.

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

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

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:

$$E\{y \mid \mathbf{x}, t_x, \hat{\boldsymbol{\beta}}\} = g(lp + (\gamma_0 + \gamma_1 lp)t_x)$$

We predict absolute benefit from

$$\hat{f}_{\text{benefit}}(lp \mid \mathbf{x}, \hat{\boldsymbol{\beta}}) = g(lp) - g(\gamma_0 + (1 + \gamma_1)lp)$$

1.4.4 Restricted cubic splines

Finally, we drop the linearity assumption and predict absolute benefit using smoothing with restricted cubic splines with 3, 4, and 5 knots. More specifically, we fit the model:

$$P(y = 1 \mid lp, t_x) = g(\beta_0 + \beta_{t_x} t_x + f_{RCS}(lp) + f_{RCS}(lp) \times t_x)$$

where

$$f_{RCS}(x) = \alpha_0 + \alpha_1 h_1(x) + \alpha_2 h_2(x) + \cdots + \alpha_{k-1} h_{k-1}(x)$$

with

$$h_{j+1}(x) = (x - t_j)^3 - (x - t_{k-1})_+ \frac{t_k - t_j}{t_k - t_{k-1}} + (x - t_k)_+^3 \frac{t_{k-1} - t_j}{t_k - t_{k-1}}$$

and t_1, \dots, t_k are the selected knots. We predict absolute benefit from

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

627	high	1,063	0.75		strong-positive	-2.08	0.49	0.49	0.49	0.49	0.49	0.49	0.49	0.49	0.49	-0.084	2.035	0.210	0	0.079	0.040
628	high	1,063	0.75		negative	-2.08	0.49	0.49	0.49	0.49	0.49	0.49	0.49	0.49	0.49	-0.084	2.035	0.210	0	0.079	0.099
629	high	1,063	0.65		absent	-1.63	0.26	0.26	0.26	0.26	0.26	0.26	0.26	0.26	0.26	0.786	2.762	0.321	0	0.089	0.089
630	high	1,063	0.65		moderate-positive	-1.63	0.26	0.26	0.26	0.26	0.26	0.26	0.26	0.26	0.26	0.786	2.762	0.321	0	0.089	0.067
631	high	1,063	0.65		strong-positive	-1.63	0.26	0.26	0.26	0.26	0.26	0.26	0.26	0.26	0.26	0.786	2.762	0.321	0	0.089	0.044
632	high	1,063	0.65		negative	-1.63	0.26	0.26	0.26	0.26	0.26	0.26	0.26	0.26	0.26	0.786	2.762	0.321	0	0.089	0.111
633	high	1,063	0.85		absent	-2.70	0.82	0.82	0.82	0.82	0.82	0.82	0.82	0.82	0.82	-0.621	1.566	0.138	0	0.069	0.069
634	high	1,063	0.85		moderate-positive	-2.70	0.82	0.82	0.82	0.82	0.82	0.82	0.82	0.82	0.82	-0.621	1.566	0.138	0	0.069	0.052
635	high	1,063	0.85		strong-positive	-2.70	0.82	0.82	0.82	0.82	0.82	0.82	0.82	0.82	0.82	-0.621	1.566	0.138	0	0.069	0.034
636	high	1,063	0.85		negative	-2.70	0.82	0.82	0.82	0.82	0.82	0.82	0.82	0.82	0.82	-0.621	1.566	0.138	0	0.069	0.086
637	high	17,000	0.75		absent	-2.08	0.49	0.49	0.49	0.49	0.49	0.49	0.49	0.49	0.49	-0.084	2.035	0.210	0	0.079	0.079
638	high	17,000	0.75		moderate-positive	-2.08	0.49	0.49	0.49	0.49	0.49	0.49	0.49	0.49	0.49	-0.084	2.035	0.210	0	0.079	0.059
639	high	17,000	0.75		strong-positive	-2.08	0.49	0.49	0.49	0.49	0.49	0.49	0.49	0.49	0.49	-0.084	2.035	0.210	0	0.079	0.040
640	high	17,000	0.75		negative	-2.08	0.49	0.49	0.49	0.49	0.49	0.49	0.49	0.49	0.49	-0.084	2.035	0.210	0	0.079	0.099
641	high	17,000	0.65		absent	-1.63	0.26	0.26	0.26	0.26	0.26	0.26	0.26	0.26	0.26	0.786	2.762	0.321	0	0.089	0.089
642	high	17,000	0.65		moderate-positive	-1.63	0.26	0.26	0.26	0.26	0.26	0.26	0.26	0.26	0.26	0.786	2.762	0.321	0	0.089	0.067
643	high	17,000	0.65		strong-positive	-1.63	0.26	0.26	0.26	0.26	0.26	0.26	0.26	0.26	0.26	0.786	2.762	0.321	0	0.089	0.044
644	high	17,000	0.65		negative	-1.63	0.26	0.26	0.26	0.26	0.26	0.26	0.26	0.26	0.26	0.786	2.762	0.321	0	0.089	0.111
645	high	17,000	0.85		absent	-2.70	0.82	0.82	0.82	0.82	0.82	0.82	0.82	0.82	0.82	-0.621	1.566	0.138	0	0.069	0.069
646	high	17,000	0.85		moderate-positive	-2.70	0.82	0.82	0.82	0.82	0.82	0.82	0.82	0.82	0.82	-0.621	1.566	0.138	0	0.069	0.052
647	high	17,000	0.85		strong-positive	-2.70	0.82	0.82	0.82	0.82	0.82	0.82	0.82	0.82	0.82	-0.621	1.566	0.138	0	0.069	0.034
648	high	17,000	0.85		negative	-2.70	0.82	0.82	0.82	0.82	0.82	0.82	0.82	0.82	0.82	-0.621	1.566	0.138	0	0.069	0.086

2 Adaptive model selection frequencies

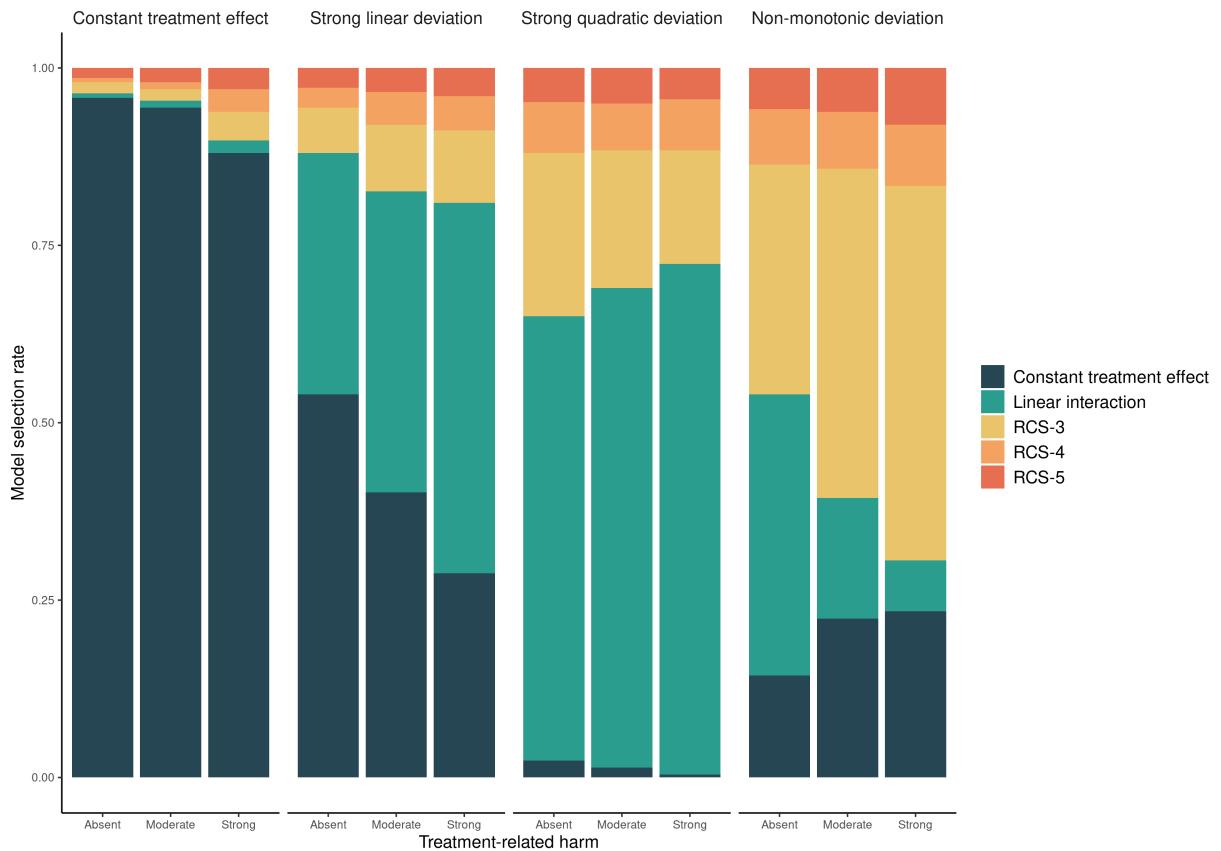


Figure S1: Model selection frequencies of the adaptive approach based on Akaike's Information Criterion across 500 replications. The scenario with the true constant relative treatment effect (first panel) had a true prediction AUC of 0.75 and sample size of 4,250.

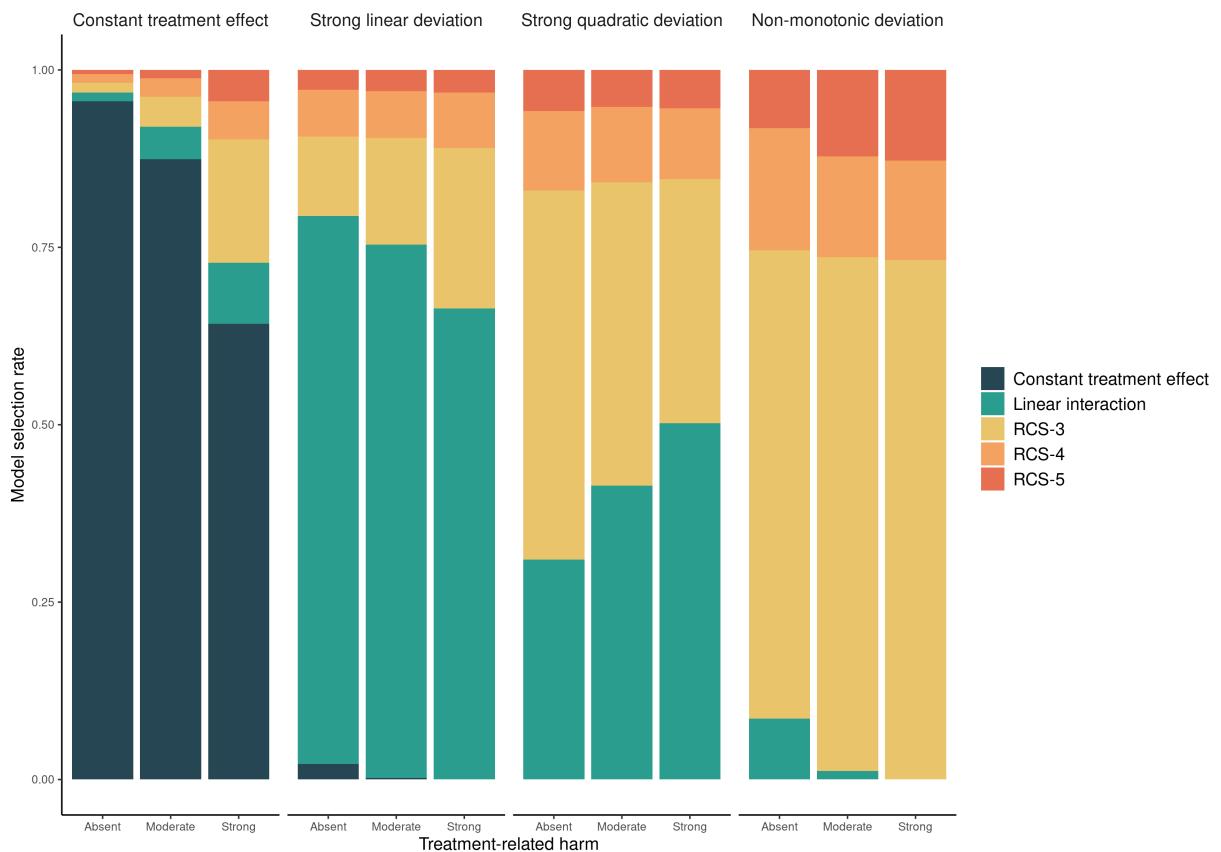


Figure S2: Model selection frequencies of the adaptive approach based on Akaike's Information Criterion across 500 replications. Sample size is 17,000 rather than 4,250 in Figure S1

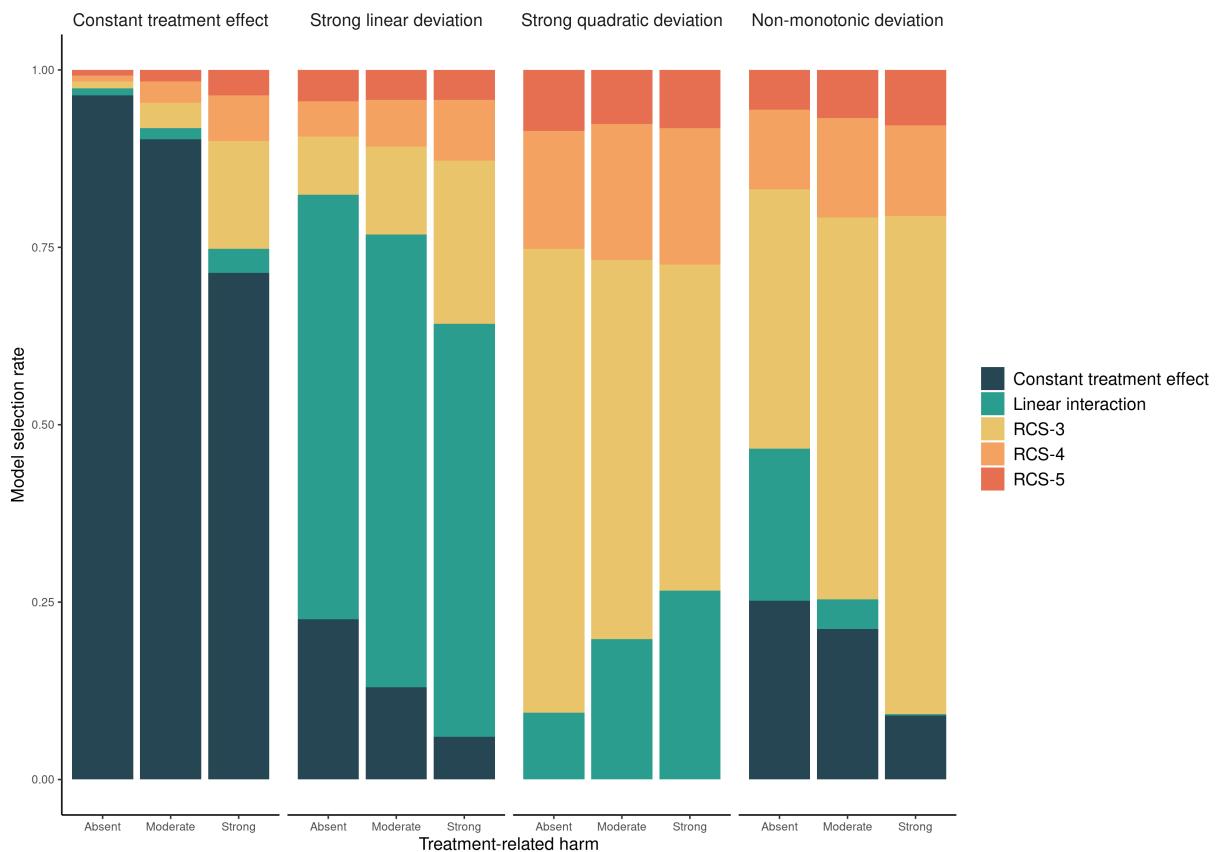


Figure S3: Model selection frequencies of the adaptive approach based on Akaike's Information Criterion across 500 replications. AUC is 0.85 rather than 0.75 in Figure S1

3 Discrimination and calibration for benefit

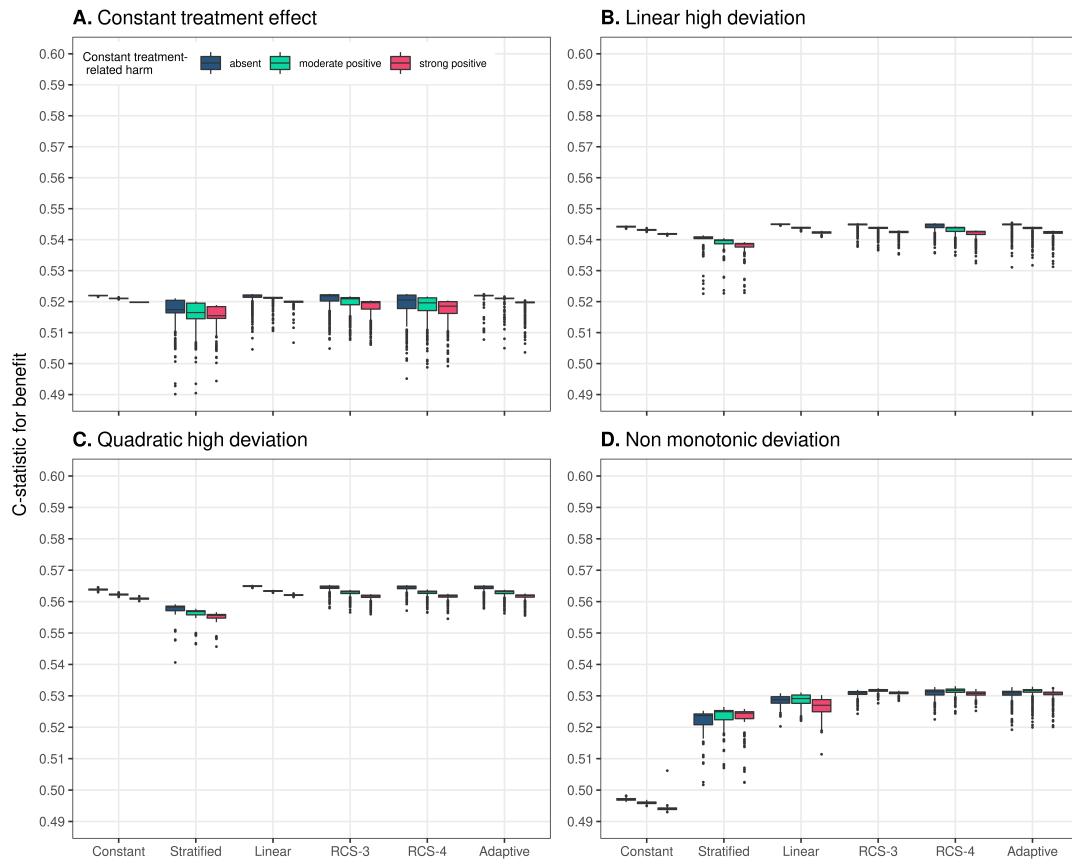


Figure S4: Discrimination for benefit of the considered methods across 500 replications calculated in a simulated sample of size 500,000. True prediction AUC of 0.75 and sample size of 17,000

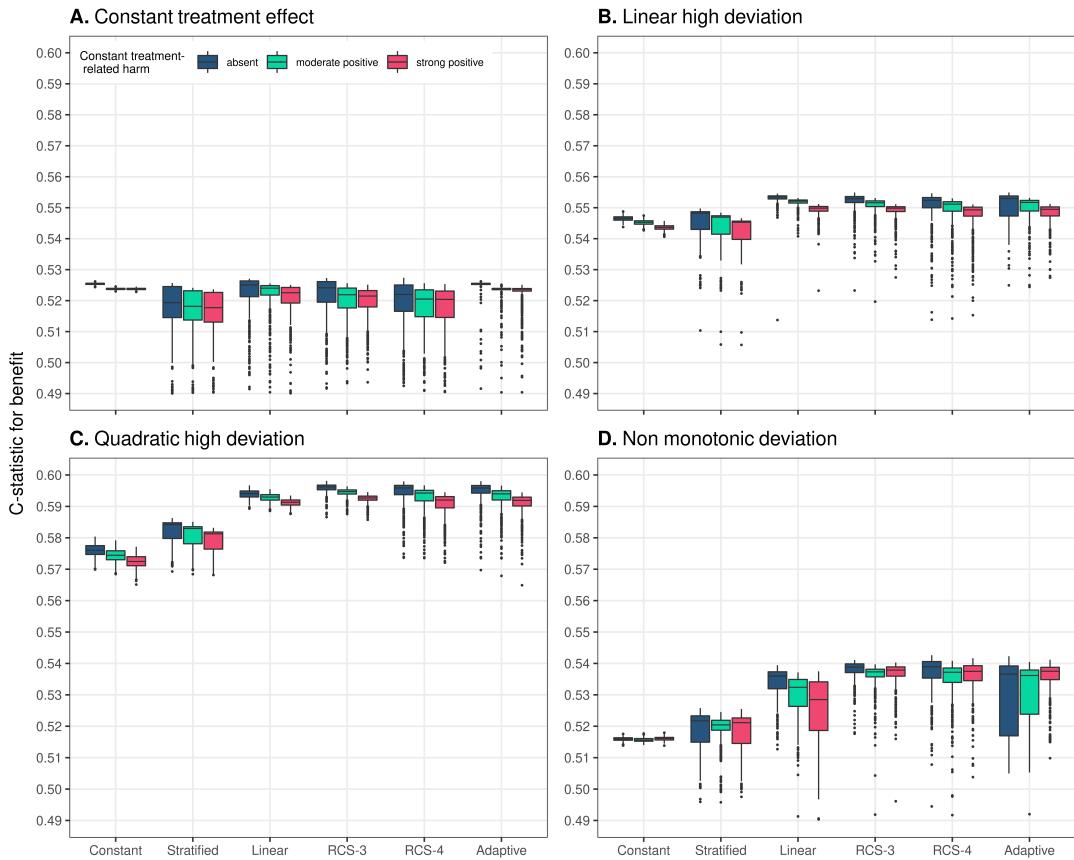


Figure S5: Discrimination for benefit of the considered methods across 500 replications calculated in a simulated sample of size 500,000. True prediction AUC of 0.85 and sample size of 4,250

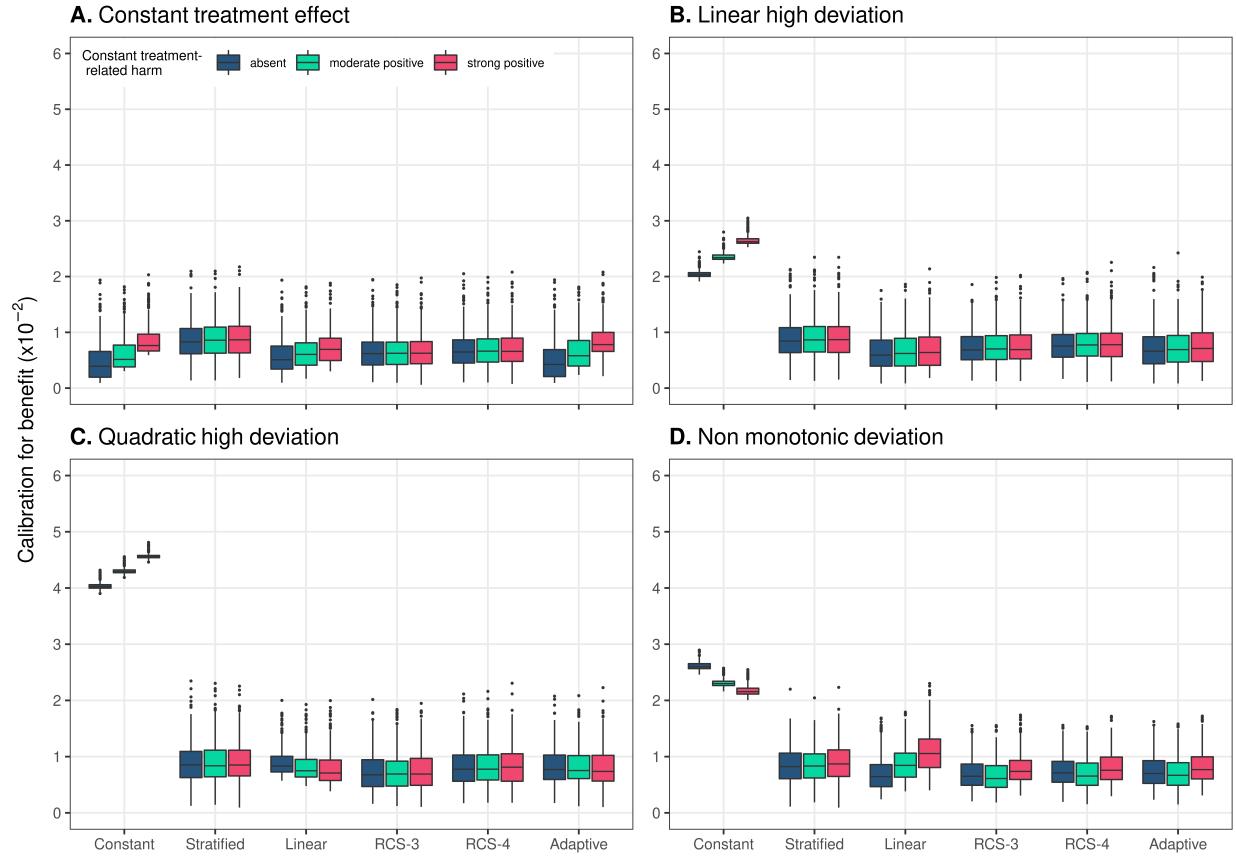


Figure S6: Calibration for benefit of the considered methods across 500 replications calculated in a simulated sample of size 500,000. True prediction AUC of 0.75 and sample size of 17,000

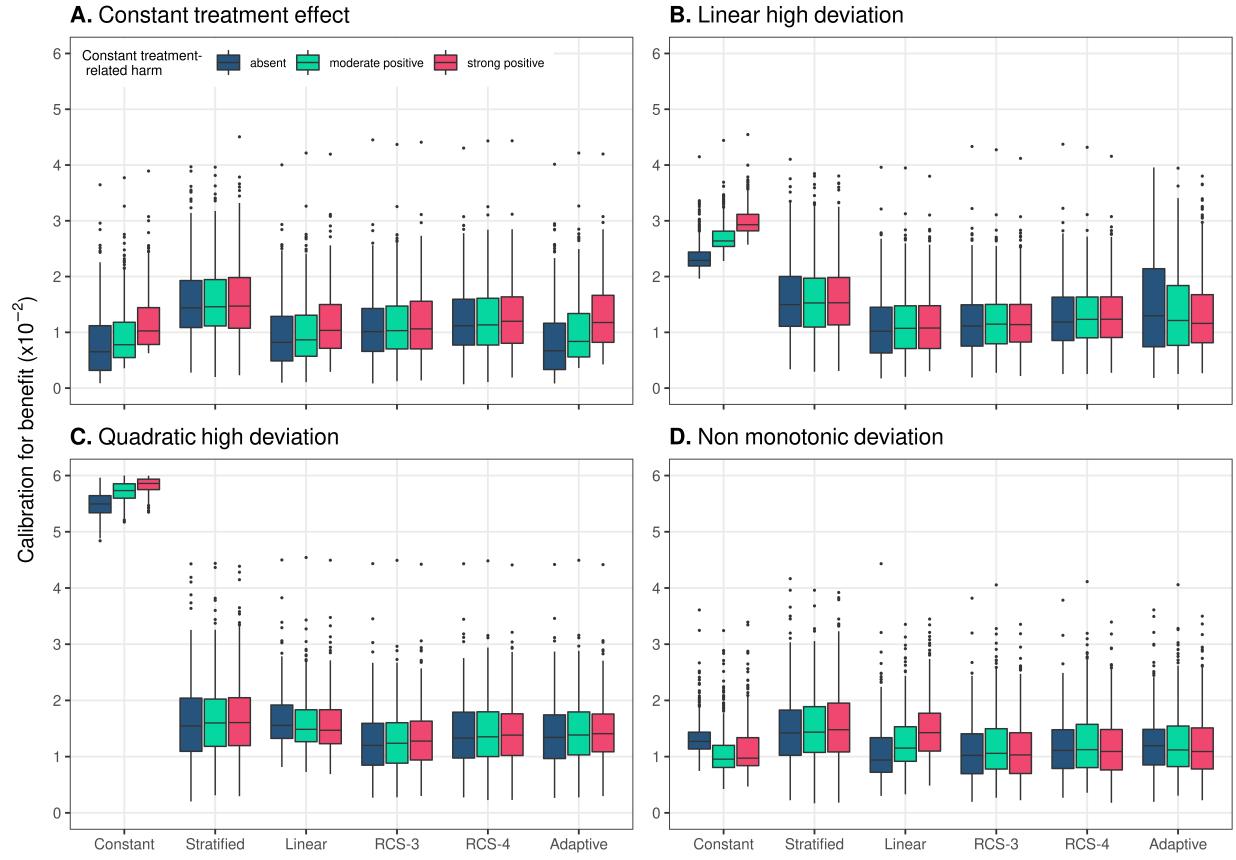


Figure S7: Calibration for benefit of the considered methods across 500 replications calculated in a simulated sample of size 500,000. True prediction AUC of 0.85 and sample size of 4,250

4 Strong relative treatment effect

Here we present the root mean squared error of the considered methods using strong constant relative treatment effect ($OR = 0.5$) as the reference. Again, the same sample size and prediction performance settings were considered along with the same settings for linear, quadratic and non-monotonic deviations from the base case scenario of constant relative treatment effects are considered. All results can be found at https://arekkas.shinyapps.io/simulation_viewer/.

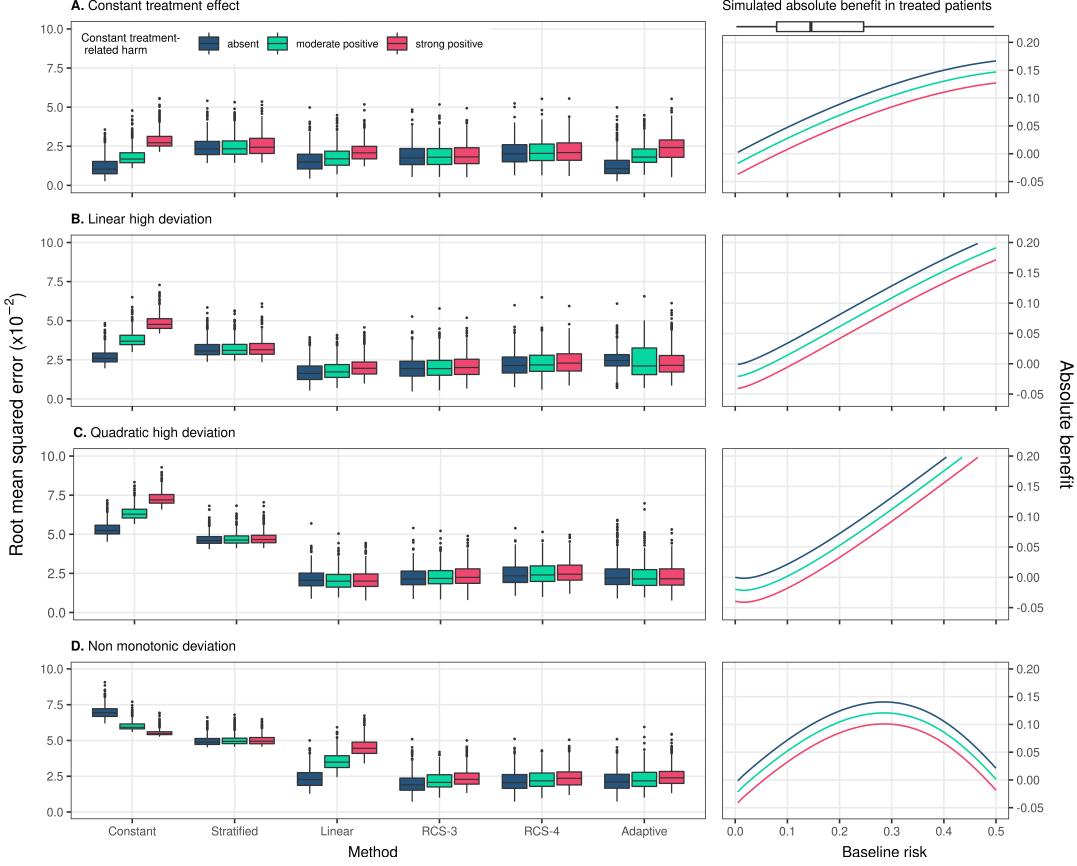


Figure S8: RMSE of the considered methods across 500 replications calculated in a simulated super-population of size 500,000. The scenario with true constant relative treatment effect (panel A) had a true prediction AUC of 0.75 and sample size of 4,250. The RMSE is also presented for strong linear (panel B), strong quadratic (panel C), and non-monotonic (panel D) deviations from constant relative treatment effects. Panels on the right side present the true relationship between baseline risk (x-axis) and absolute treatment benefit (y-axis). The 2.5, 25, 75 and 97.5 percentiles of the risk distribution are expressed in the boxplot.

5 Treatment interactions

We carried out a smaller set of simulations, in which we assumed true treatment-covariate interactions. Sample size was set to 4,250 and the AUC of the true prediction model was set to 0.75. The following scenarios were considered: 1) 4 true weak positive interactions ($OR_{t_x=1}/OR_{t_x=0} = 0.83$); 2) 4 strong positive interactions ($OR_{t_x=1}/OR_{t_x=0} = 0.61$); 3) 2 weak and 2 strong positive interactions; 4) 4 weak negative interactions ($OR_{t_x=1}/OR_{t_x=0} = 1.17$); 5) 4 strong negative interactions ($OR_{t_x=1}/OR_{t_x=0} = 1.39$); 6) 2 weak and 2 strong negative interactions; 7) combined positive and negative strong interactions. We also considered

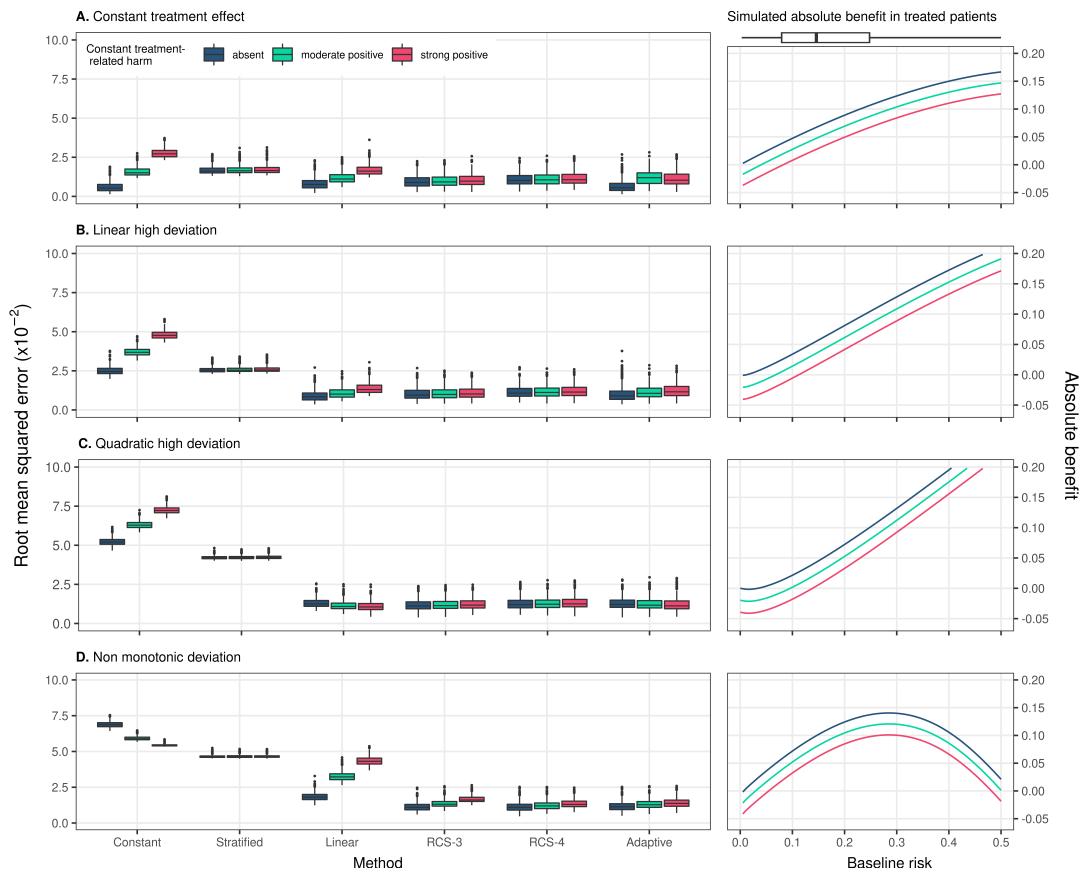


Figure S9: RMSE of the considered methods across 500 replications calculated in a simulated sample of size 500,000. Sample size is 17,000 rather than 4,250 in Figure S8.

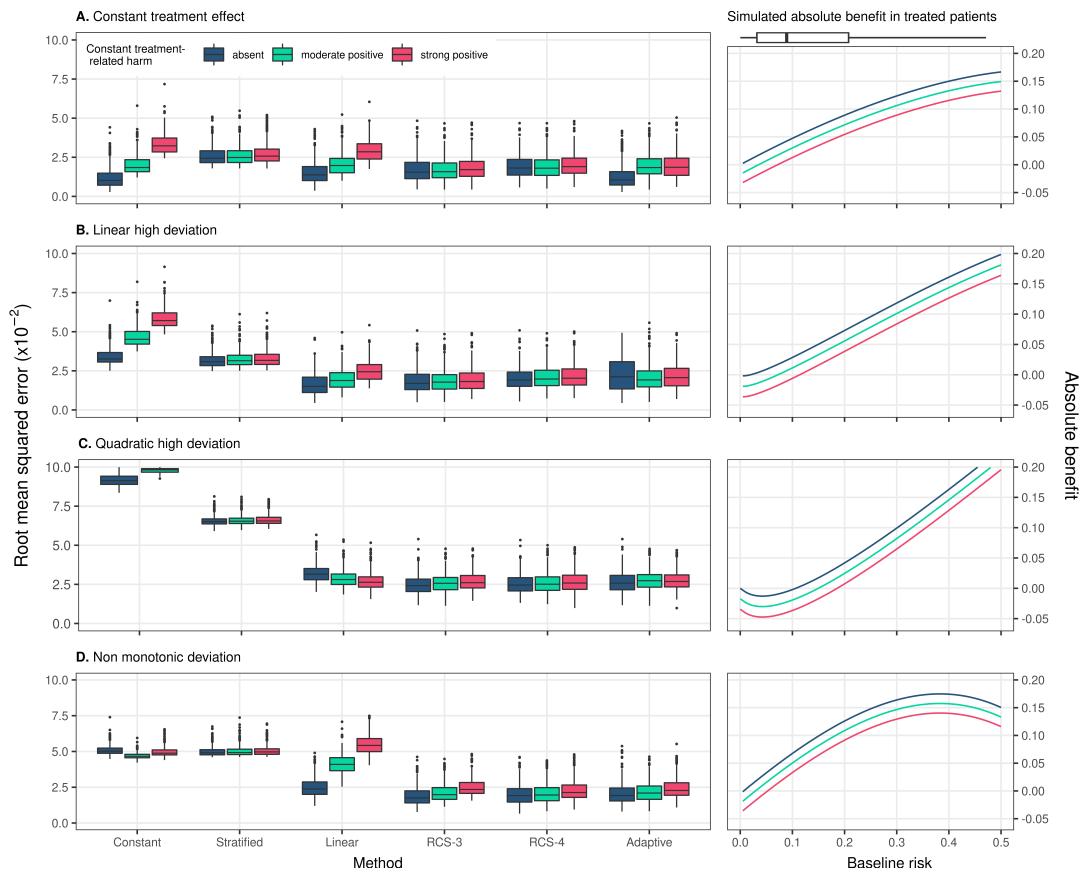


Figure S10: RMSE of the considered methods across 500 replications calculated in a simulated sample of size 500,000. AUC is 0.85 rather than in Figure S8.

constant treatment-related harms applied on the absolute scale to all treated patients. The exact settings were: 1) absent treatment-related harms; 2) moderate treatment-related harms, defined as 25% of the average true benefit of the scenario without treatment-related harms; 3) strong treatment-related harms defined as 50% of the true average benefit of the scenario without treatment-related harms; 4) negative treatment-related harms (benefit), defined as an absolute risk reduction for treated patients of 50% of the true average benefit of the scenario without treatment-related harms. The exact settings can be found in Table S2.

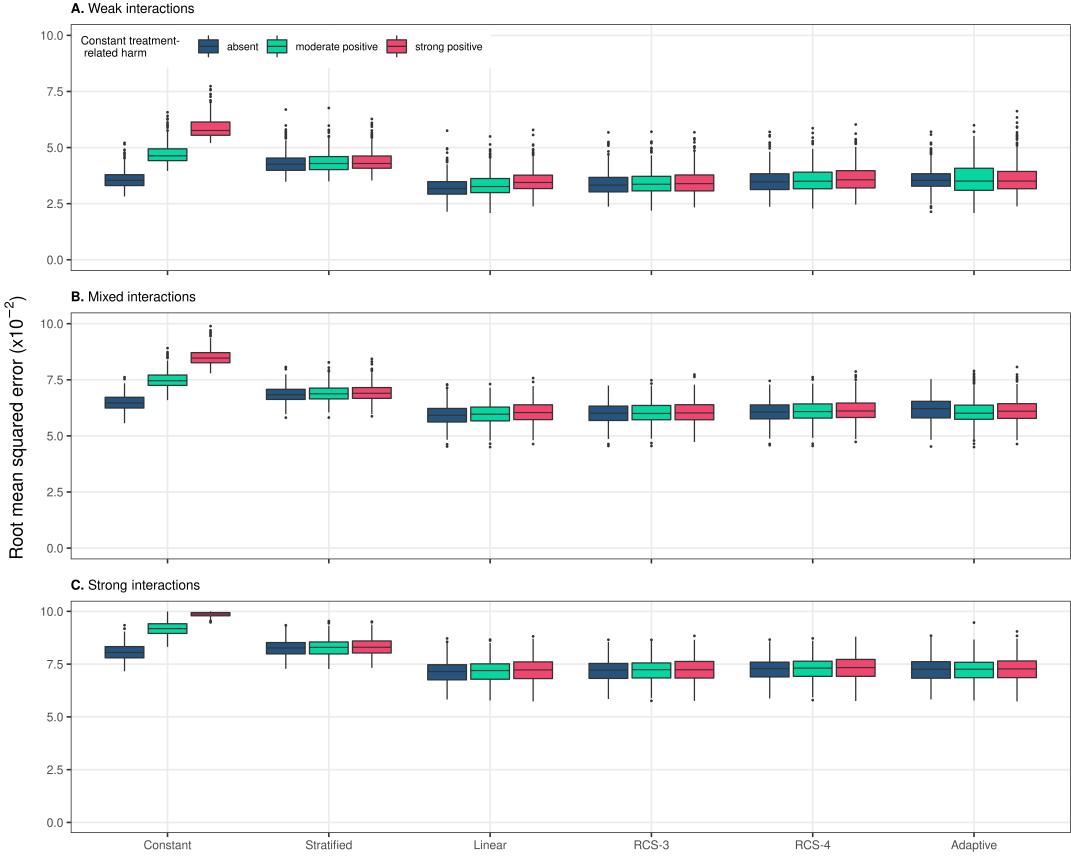


Figure S11: RMSE of the considered methods across 500 replications calculated in a simulated sample of size 500,000. AUC is 0.85 rather than in Figure S8.

6 Empirical illustration

$$P(\text{outcome} = 1 | X = x) = \text{expit}(lp(x)), \quad (4)$$

where

$$\begin{aligned} lp(x) = & \beta_0 + \beta_1 \text{age} + \beta_2 I(\text{Killip} = II) + \beta_3 I(\text{Killip} = III) + \\ & \beta_4 I(\text{Killip} = IV) + \beta_5 \text{ceil}(\text{sysbp}, 120) + \text{lsp}(\text{pulse}, 50) + \\ & \beta_6 I(\text{pmi} = yes) + \beta_8 I(\text{miloc} = \text{Anterior}) + \\ & \beta_9 I(\text{miloc} = \text{Other}) + \gamma t_x \end{aligned} \quad (5)$$

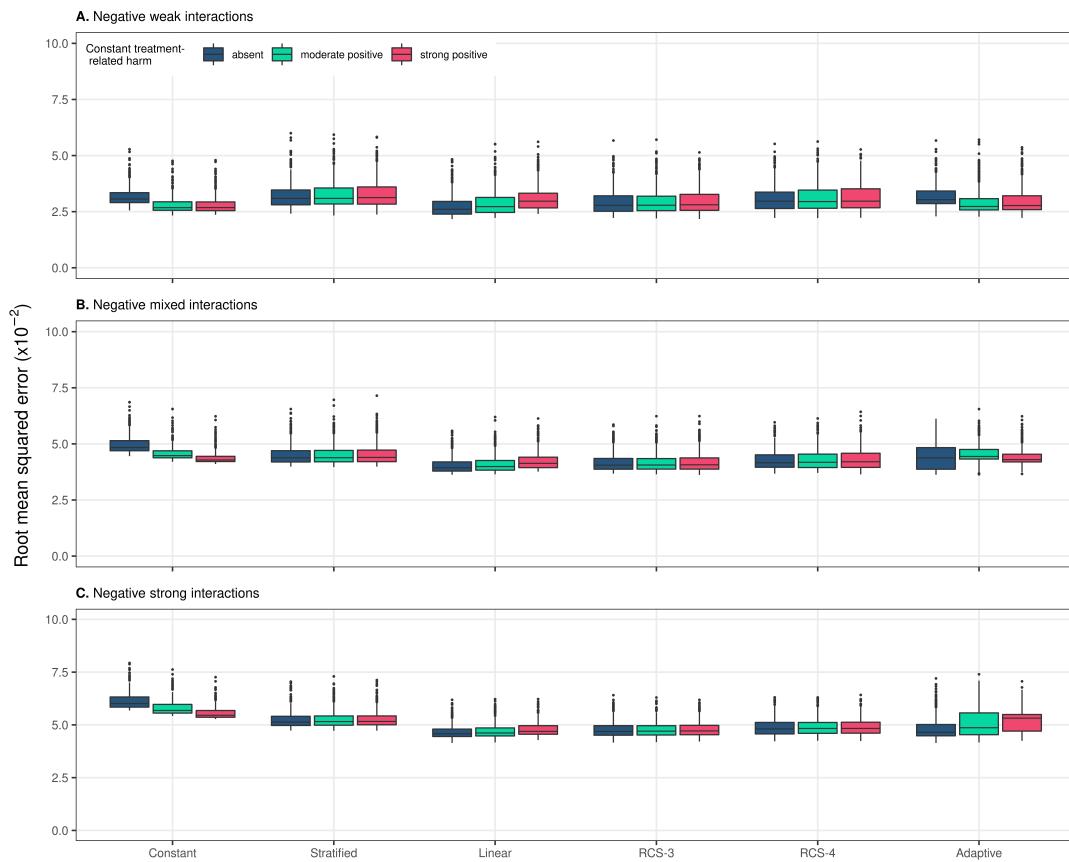


Figure S12: RMSE of the considered methods across 500 replications calculated in a simulated sample of size 500,000. AUC is 0.85 rather than in Figure S8.

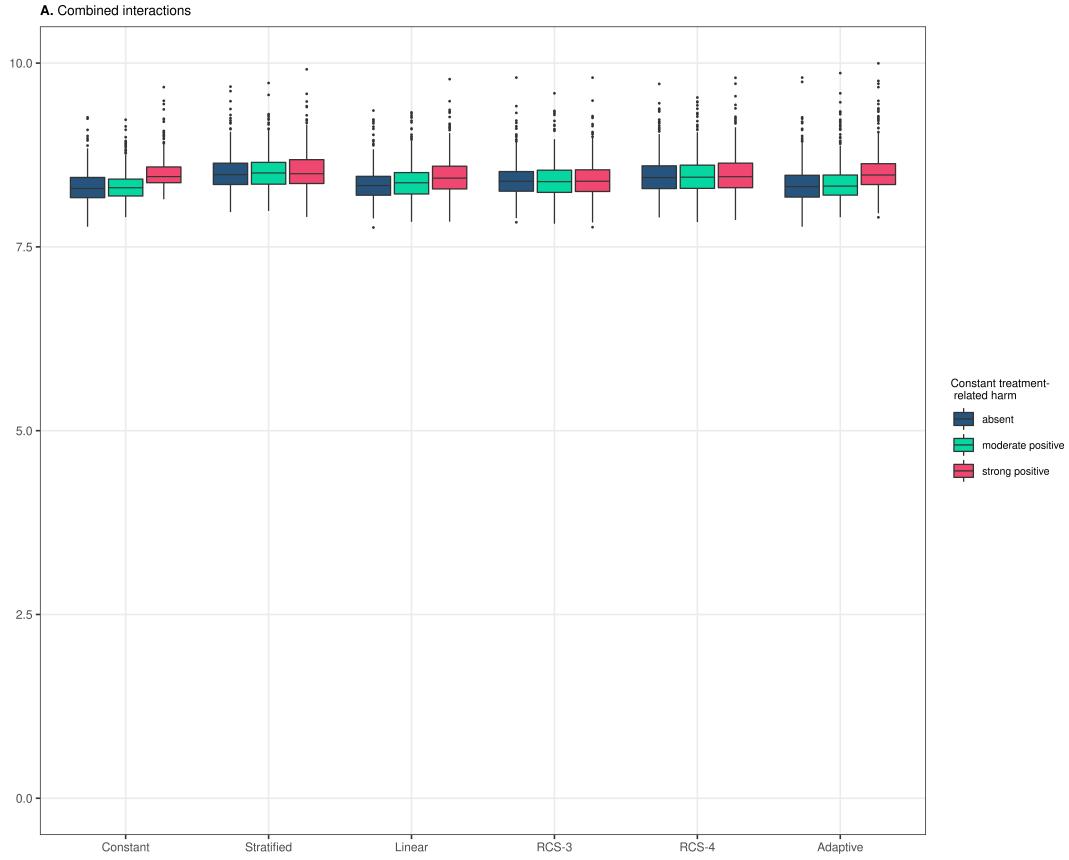


Figure S13: RMSE of the considered methods across 500 replications calculated in a simulated sample of size 500,000. AUC is 0.85 rather than in Figure S8.