Factors influencing strategies for ‘treatment paradox’ correction in a time-dependent Cox model

Zhenwei Yang

OUTLINE

1. Background
2. Data generating mechanism
3. Correction techniques
4. Performance measurement
5. Results

1.

Background

PROBLEM WITH "UNTREATED RISK"

- Prognostic models aim to produce individual risk without getting intervention

- However, treatment is common in observational studies

- High proportion of treatment use in validation dataset would lead to overestimation of survival probability and affect model performance

- Plenty of studies for binary logistic regression models, but hardly in survival context

RESEARCH GOALS

- To conduct a **simulation study** to investigate the **treatment paradox** in **survival analysis**. Specifically, we aim to compare the **performance of different methods** to model treatment in prediction models when using data with time-to-event outcomes.

2.

Data generating mechanism

DATA GENERATING MECHANISM (DGM)

$$P \sim Unif(0,1), \; U \sim Unif(0,1)$$

$$E(Y_t = 1 | P, U, A_t) = \frac{exp(b_0 + b_1 P + b_2 U + b_3 A)}{1 + \; exp(b_0 + b_1 P + b_2 U + b_3 A)}$$

$$E(A_t = 1 | P, U) = \frac{exp(a_0 + a_1 P + a_2 U)}{1 + exp(a_0 + a_1 P + a_2 U)}$$

- a0: the intercept linked to instantaneous probability of treatment. [***seq(-8,-4,1)***]
- b3: the effect of treatment (A) on outcome (Y). [***seq(-2,0,0.2)***]
- b1: the effect of observed predictor (P) on outcome (Y). [***seq(0.2,2,0.2)***]
- b0: fixed to be -4.5 (but we also try flexible b0 value = -4.25 – 0.5*b1 to fix the instantaneous probability of outcome to be exp(-4)/(1+exp(-4)))
- a1, a2, b2: fixed to be 0.5

3.

Correction techniques

Six strategies

Strategies

1. Ignorance strategy
2. Exclusion strategy
3. Competing risk strategy
4. “crude” correction strategy
5. Treatment as time-varying covariate modeling strategy (IPTW strategy)
6. Treatment as time-varying covariate censoring strategy (IPCW strategy)

Predicted survival probability

1. $\Pr(T > t.hor|P)$
2. $\Pr(T > t.hor|P, A_{t.hor} = 0)$
3. $\Pr(\min(T, V) > t.hor|P)$
4. $\Pr(T > t.hor|P) * \Pr(V > t.hor|P)$

Note: T is the time outcome occurs; V is the time treatment occurs; t.hor is the time point to estimate the survival probability; P is the observed predictor; U is the unobserved predictor.

4.

Performance measurement

Performance measurement

Four methods

1. Brier score: $\frac{1}{N}\sum_{i=1}^{N}(O_i - E_i)^2$
2. O:E ratio: $\frac{\sum observed\ probability\ of\ survival}{\sum predicted\ probability\ of\ survival}$
3. Calibration slope: the slope when observed probability of survival is plotted against predicted probability of survival.
4. C index: the area under the ROC curve

NOTE

The observed survival is represented by $Pr(T > t,\ \ A_t = 0|\ P\ , U)$ obtained from DGM. Since DGM is known, this is the true "untreated" survival probability. Instead observed probability is contaminated by the treatment use and bias the results

5.

Results

Brier score

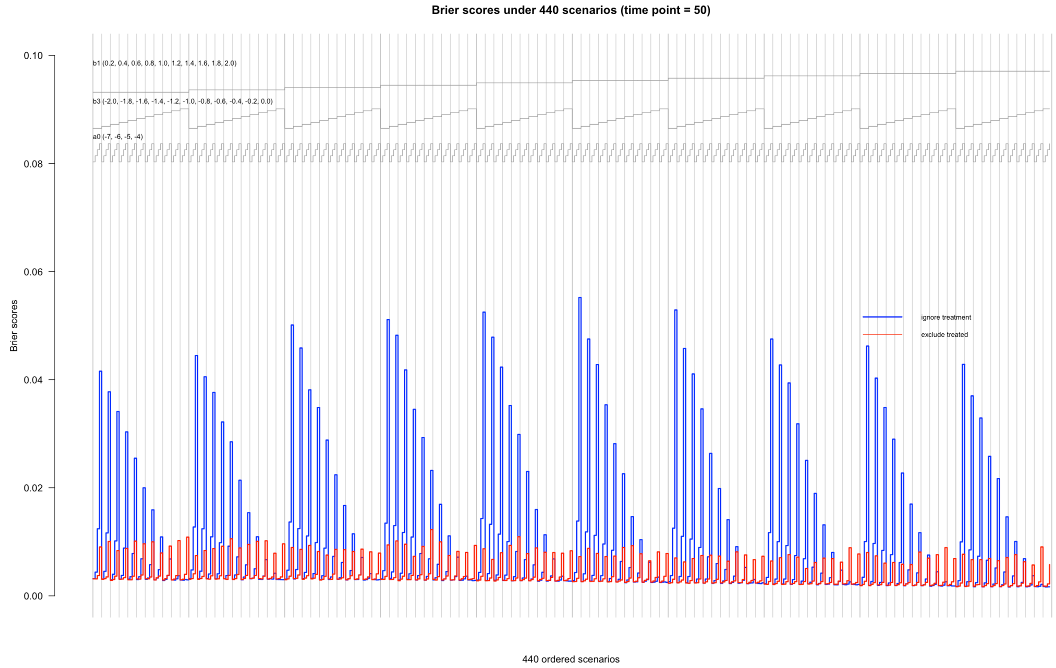


Figure 1 Nested loop plot for Brier scores of ignorance strategy (blue line) and exclusion strategy (red line) under 440 scenarios

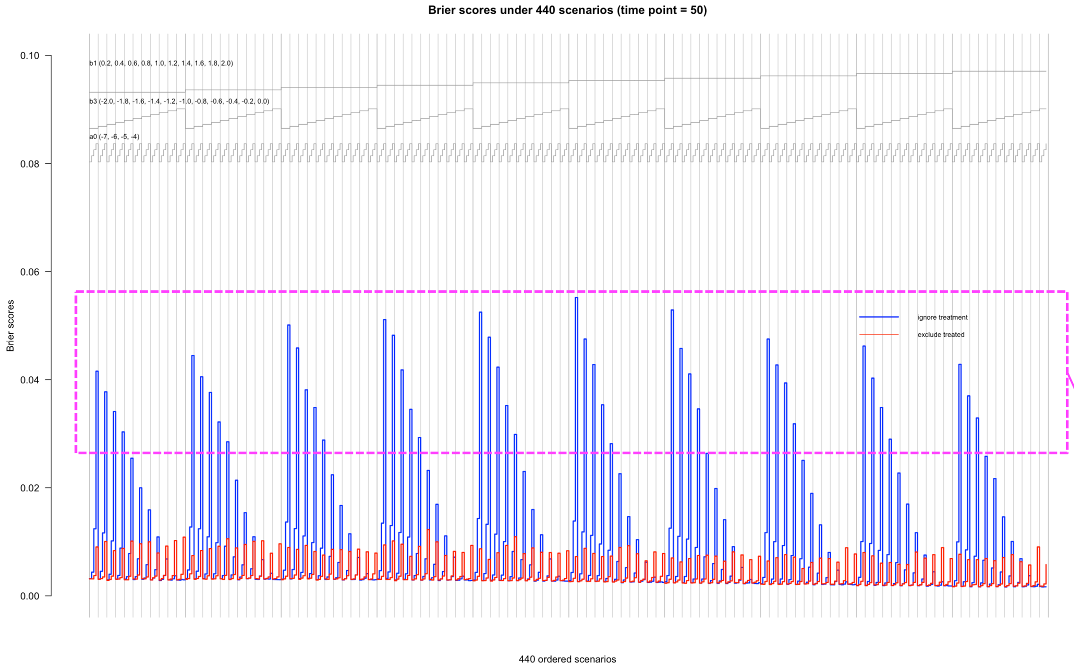
Ignorance strategy

- Larger a0 leads to more treatment use and ultimately the increase of survival probability, then larger Brier score
- Larger b3 means weaker effect from treatment, causing the estimating survival probability closer to the reference

Exclusion strategy

- Sensitive to a0
- Seems stable

Brier score



The pattern caused by b1 can be explained by the growth of probability of outcome

Figure 1 Nested loop plot for Brier scores of ignorance strategy (blue line) and exclusion strategy (red line) under 440 scenarios

Brier score

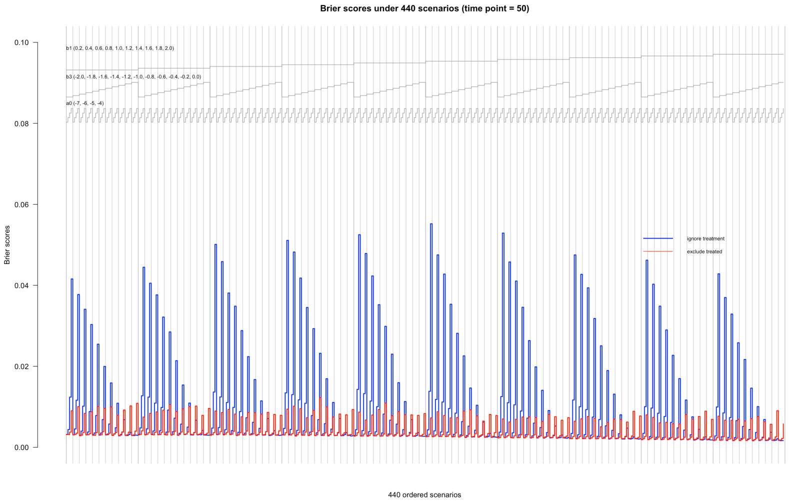


Figure 1 Nested loop plot for Brier scores of ignorance strategy (blue line) and exclusion strategy (red line) under 440 scenarios

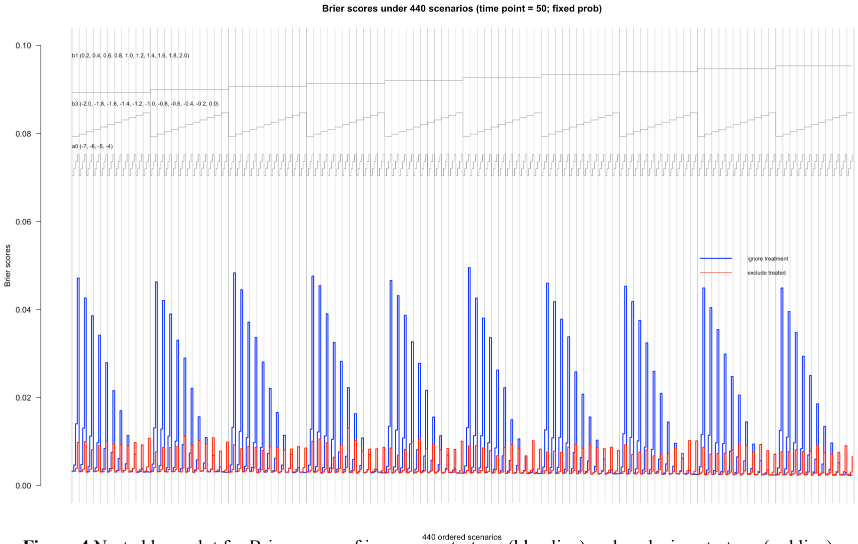


Figure 4 Nested loop plot for Brier scores of ignorance strategy (blue line) and exclusion strategy (red line) under 440 scenarios of fixed outcome probability

Brier score

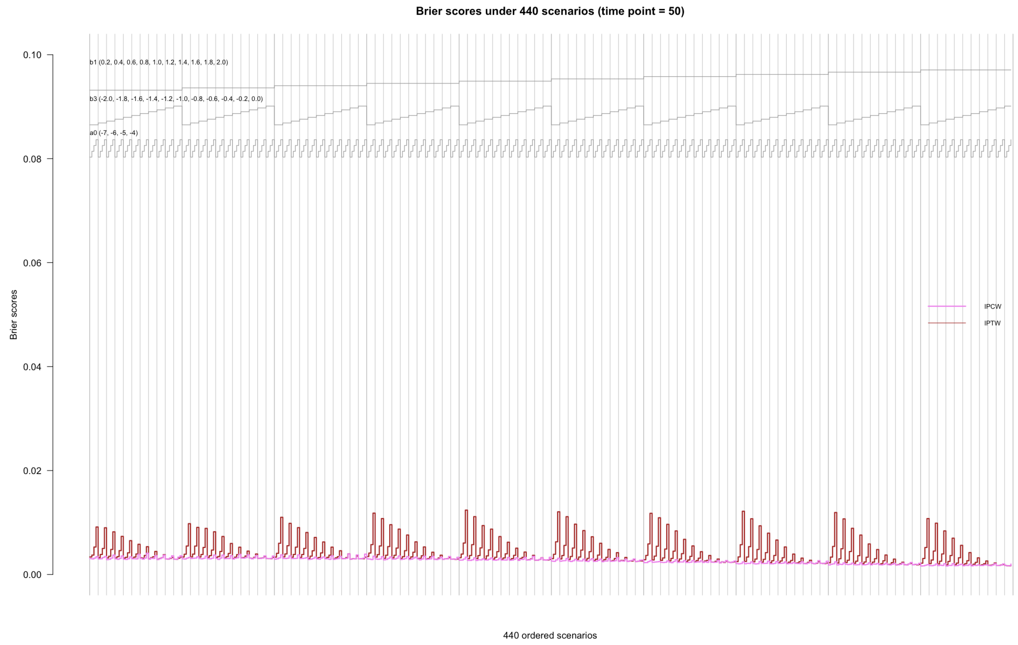


Figure 2 Nested loop plot for Brier scores of IPCW strategy (pink line) and IPTW strategy (brown line) under 440 scenarios

IPTW strategy

- Same pattern as ignorance strategy but more stable

IPCW strategy

- Most stable strategy
- Not sensitive to any parameter specified

Brier score

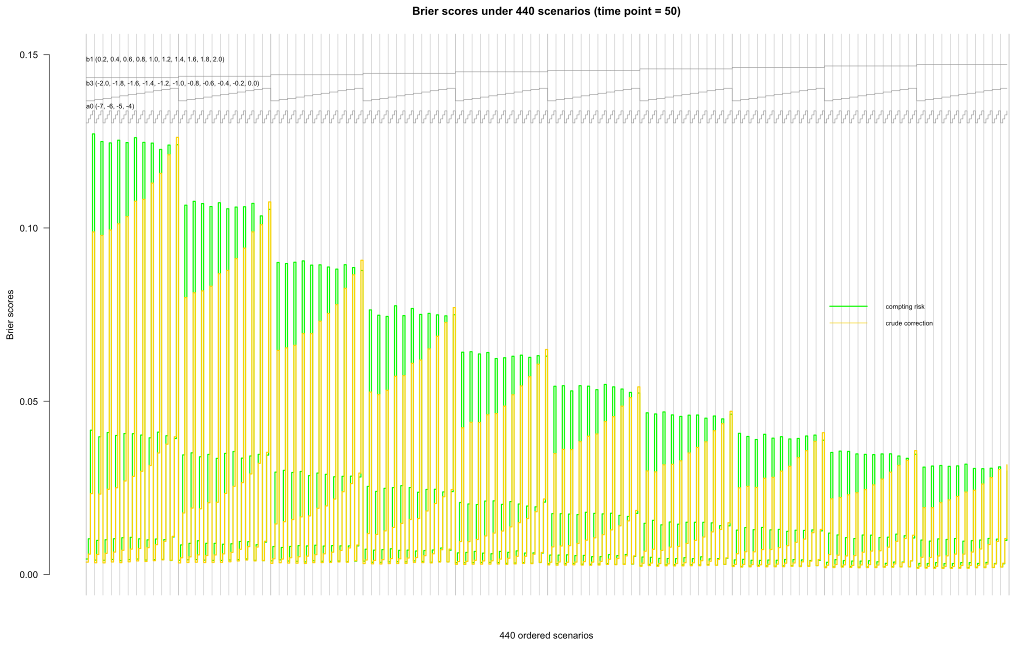


Figure 3 Nested loop plot for Brier scores of competing risk strategy (green line) and "crude" correction strategy (yellow line) under 440 scenarios

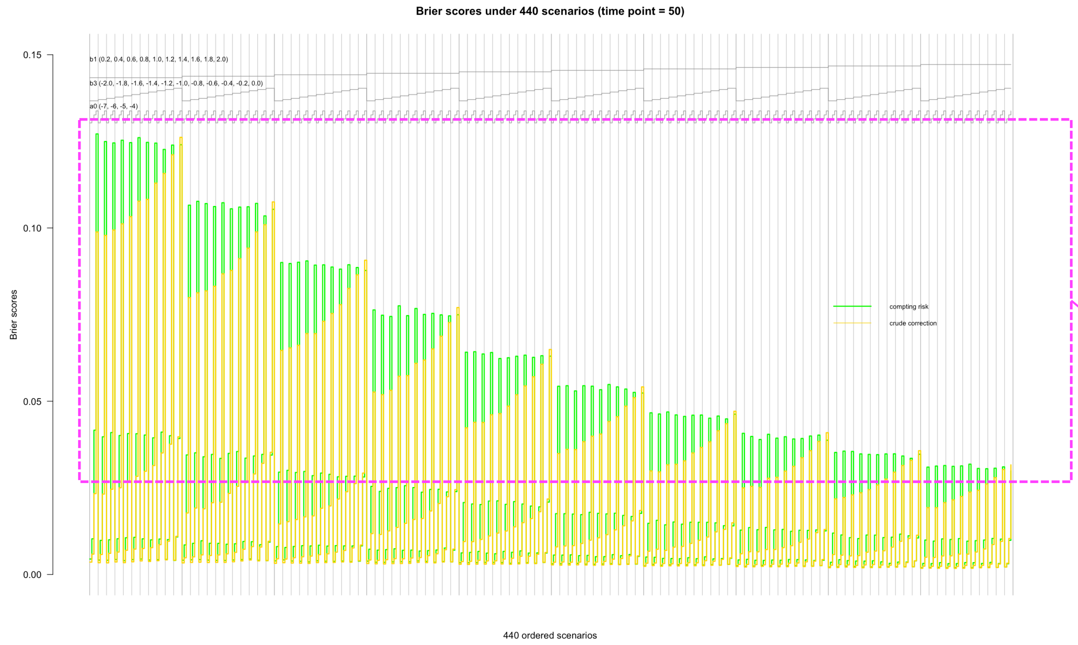
Competing risk strategy

- Larger a0 leads to more treatment use and ultimately the increase of survival probability, but the main effect is dominated by the decrease of the probability of not getting treatment, thus the increase of Brier score
- Not sensitive to b3

"crude" correction strategy

- Larger a0 leads to decrease of not getting treatment, thus higher Brier score
- When treatment effect is independent of outcome, results from competing risk and "crude" correction will be close

Brier score



The pattern caused by b1 can be explained by the growth of probability of outcome

Figure 3 Nested loop plot for Brier scores of competing risk strategy (green line) and "crude" correction strategy (yellow line) under 440 scenarios

Brier score

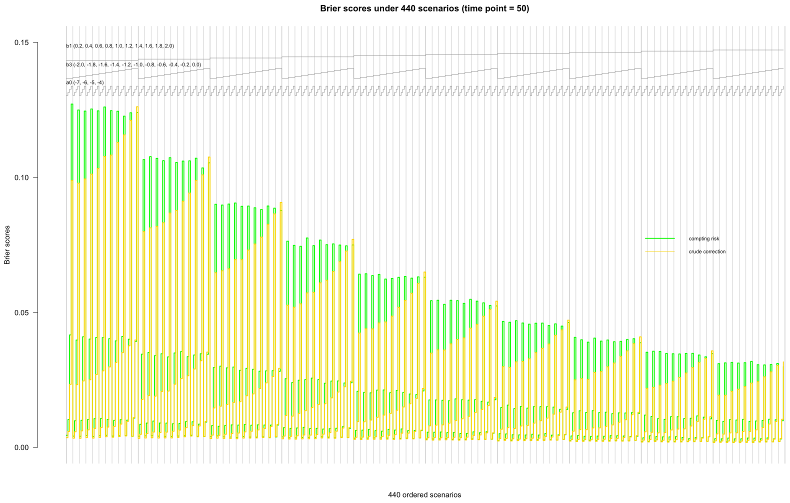


Figure 3 Nested loop plot for Brier scores of competing risk strategy (green line) and “crude” correction strategy (yellow line) under 440 scenarios

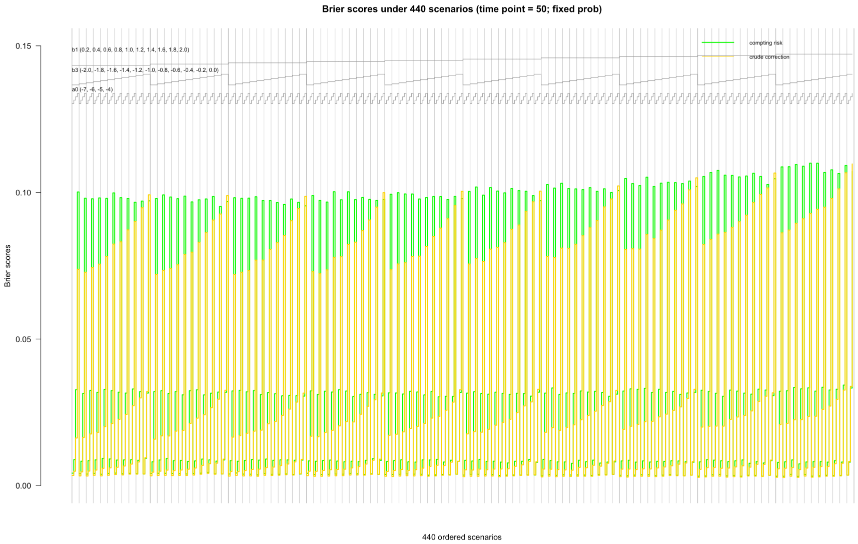


Figure 5 Nested loop plot for Brier scores of competing risk strategy (green line) and “crude” correction strategy (yellow line) under 440 scenarios of fixed outcome probability

OE ratio

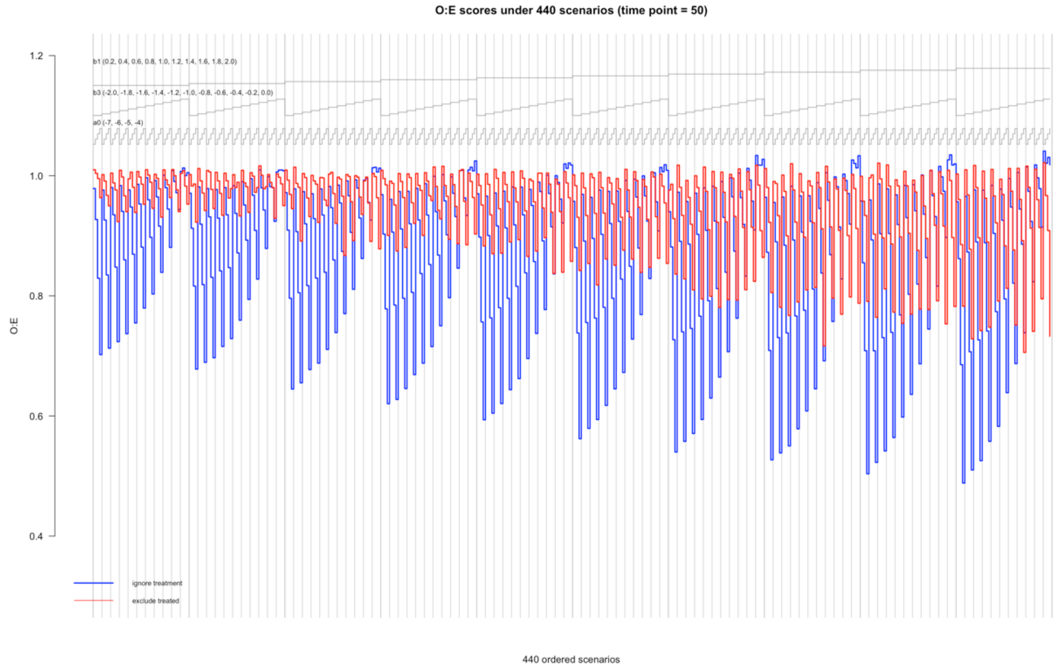


Figure 6 Nested loop plot for OE ratio of ignorance strategy (blue line) and exclusion strategy (red line) under 440 scenarios

Ignorance strategy

- Larger a0 leads to more treatment use and ultimately the increase of survival probability, then lower OE ratio
- Larger b3 means weaker effect from treatment, causing the estimating survival probability closer to the reference

Exclusion strategy

- Sensitive to a0

OE ratio

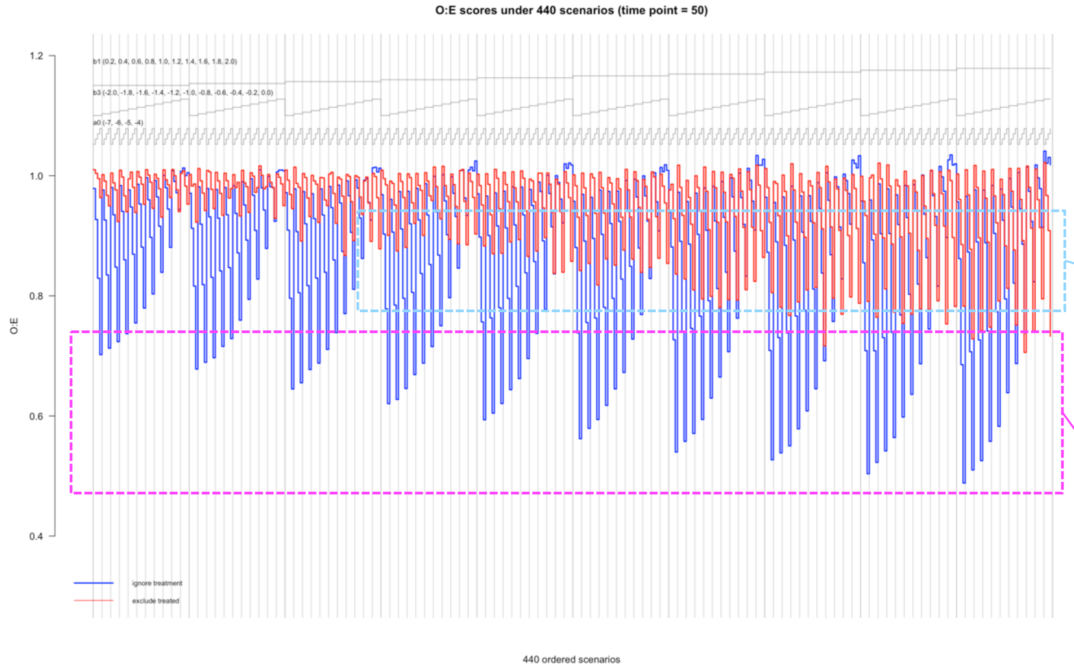


Figure 6 Nested loop plot for OE ratio of ignorance strategy (blue line) and exclusion strategy (red line) under 440 scenarios

- The pattern caused by b1 can be partially explained by the growth of probability of outcome;
- Instability is also caused by the underestimation of relation between P and outcome due to selection process

The pattern caused by b1 can be explained by the growth of probability of outcome

OE ratio

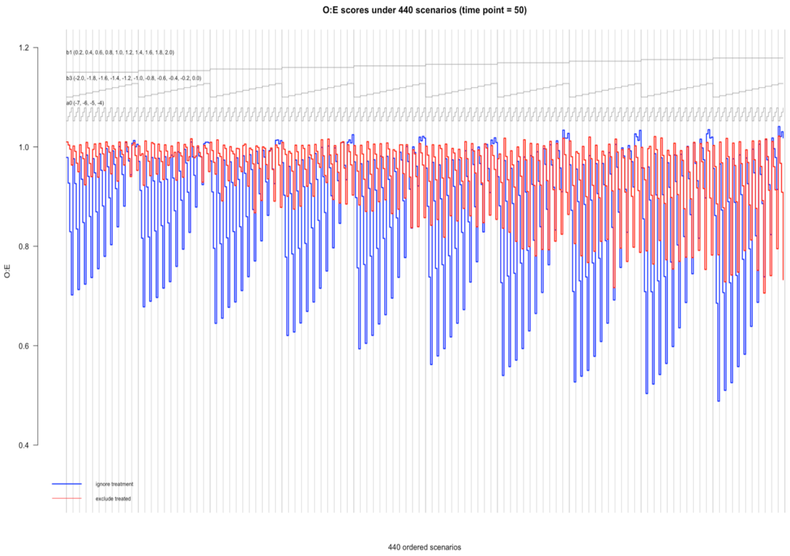


Figure 6 Nested loop plot for OE ratio of ignorance strategy (blue line) and exclusion strategy (red line) under 440 scenarios

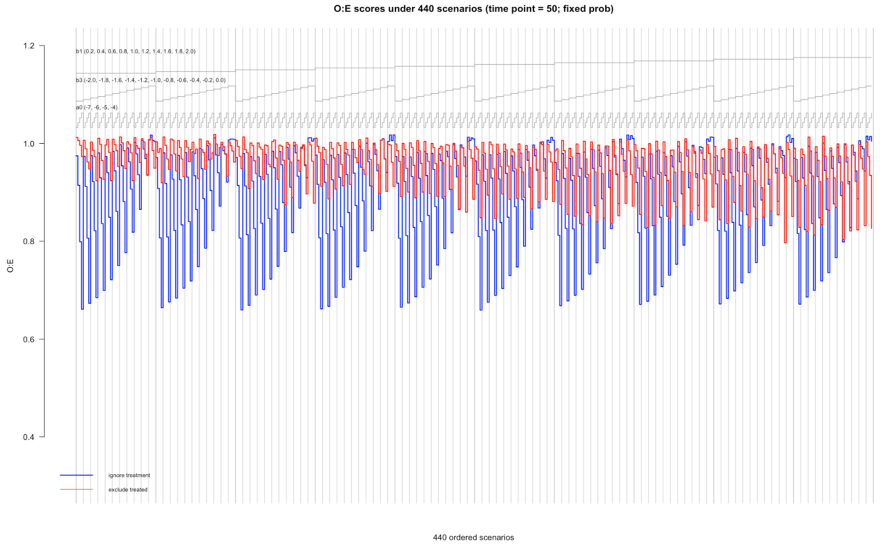


Figure 9 Nested loop plot for OE ratio of ignorance strategy (blue line) and exclusion strategy (red line) under 440 scenarios of fixed outcome probability

OE ratio

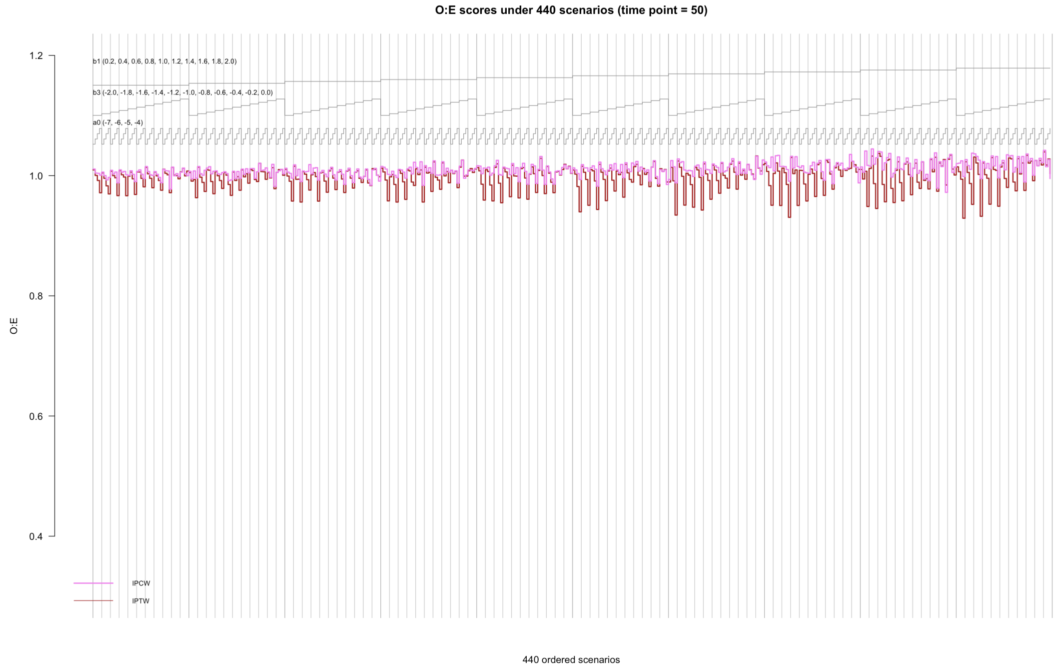


Figure 7 Nested loop plot for OE ratio of IPCW strategy (pink line) and IPTW strategy (brown line) under 440 scenarios

IPTW strategy

- Only sensitive to a0
- Predicted probability is still slightly Larger than reference probability but more stable

IPCW strategy

- Most stable strategy
- No obvious pattern observed

OE ratio

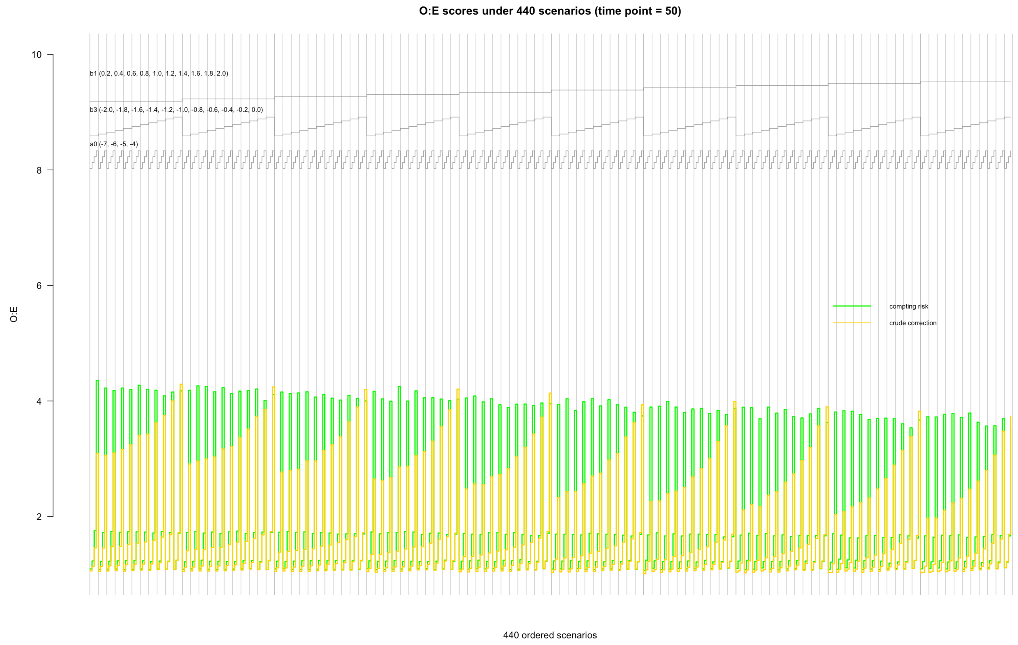


Figure 8 Nested loop plot for OE ratio of competing risk strategy (green line) and "crude" correction strategy (yellow line) under 440 scenarios

Competing risk strategy

- Larger a0 leads to more treatment use and ultimately the increase of survival probability (however, the probability of not getting treatment will decrease)
- b3 and b1 do not affect that much

"crude" correction strategy

- Larger a0 leads to decrease of not getting treatment and much smaller predicted probability, thus higher OE ratio
- When treatment effect is independent of outcome, results from competing risk and "crude" correction will be close

Calibration Slope

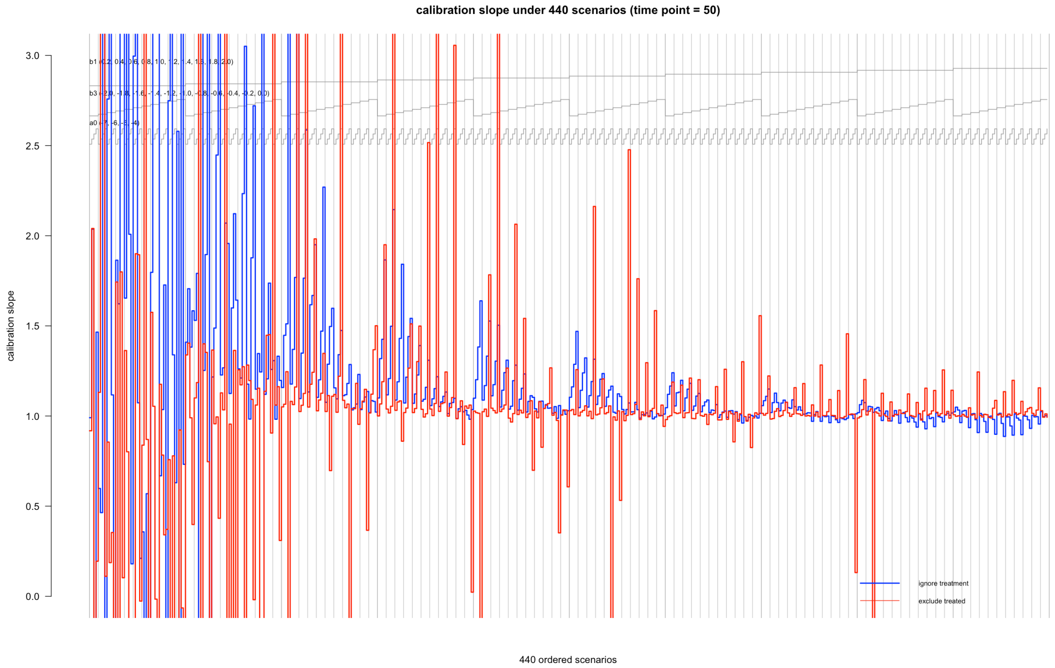


Figure 10 Nested loop plot for calibration slopes of ignorance strategy (blue line) and exclusion strategy (red line) under 440 scenarios

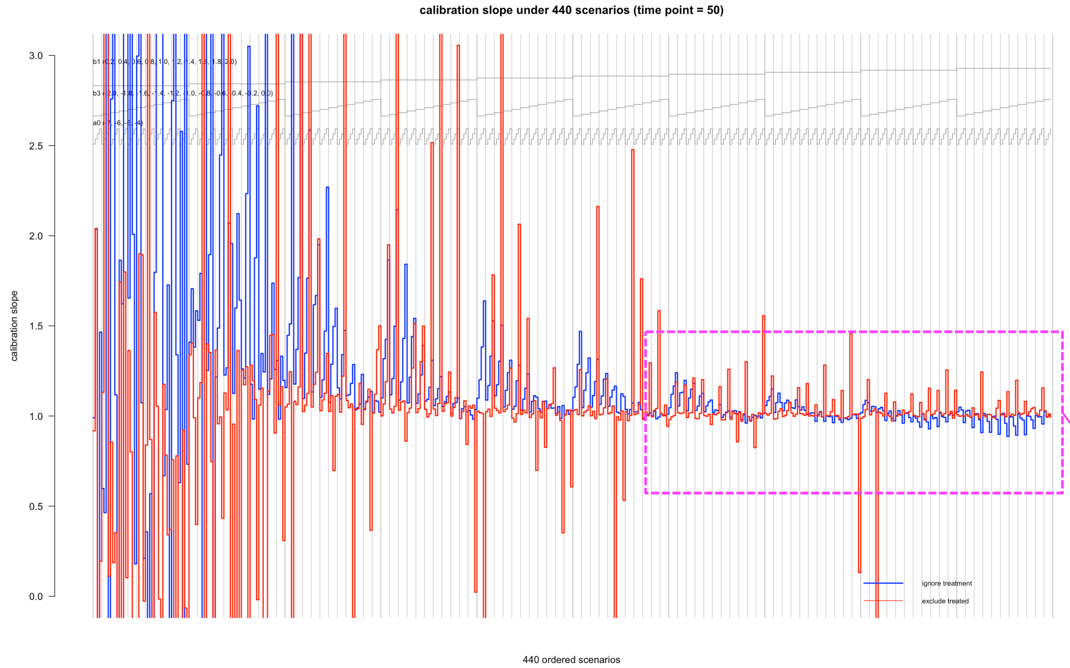
Ignorance strategy

- Larger a0 leads to more treatment use and ultimately the increase of survival probability, thus higher calibration slope
- Larger b3 means weaker effect from treatment, causing the estimating survival probability closer to the reference
- Larger b1 might lead to overfitting

Exclusion strategy

- No obvious pattern is observed

Calibration Slope



Overfitting is caused by the growth of probability of outcome (extreme cases)

Figure 10 Nested loop plot for calibration slopes of ignorance strategy (blue line) and exclusion strategy (red line) under 440 scenarios

Calibration Slope

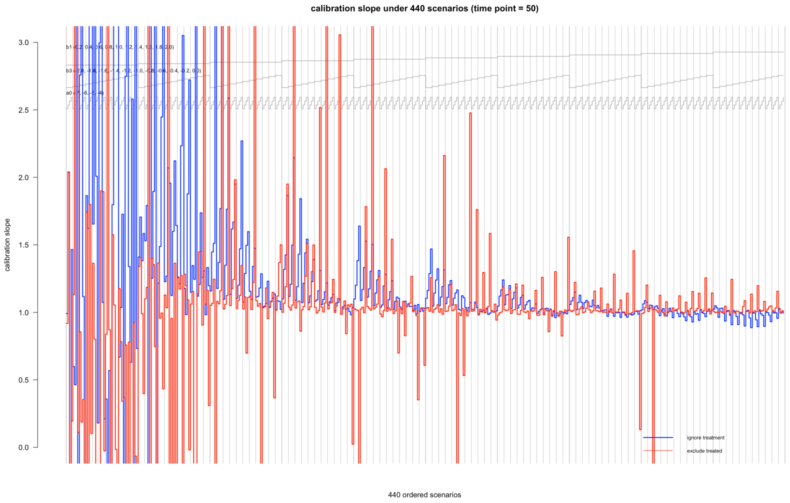


Figure 10 Nested loop plot for calibration slopes of ignorance strategy (blue line) and exclusion strategy (red line) under 440 scenarios

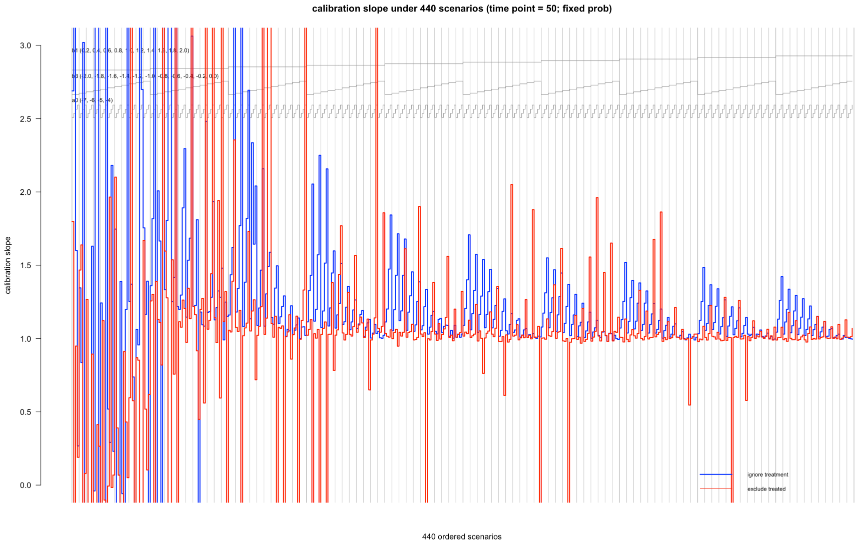


Figure 13 Nested loop plot for calibration slopes of ignorance strategy (blue line) and exclusion strategy (red line) under 440 scenarios of fixed outcome probability

Calibration slope

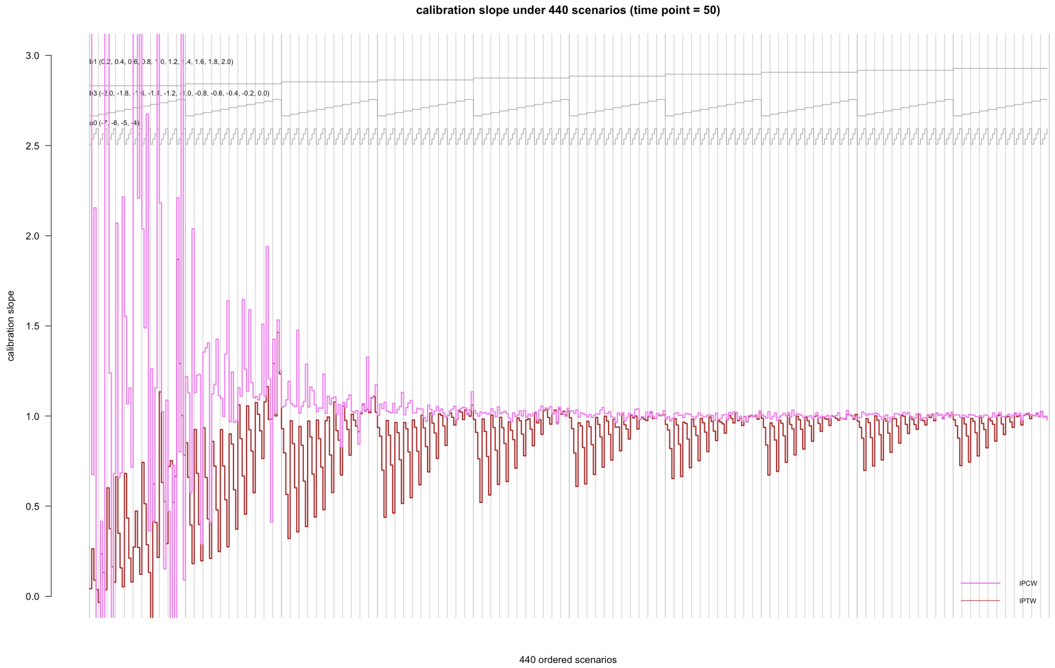


Figure 11 Nested loop plot for calibration slopes of IPCW strategy (pink line) and IPTW strategy (brown line) under 440 scenarios

IPTW strategy

- More stable with less proportion of treatment and less treatment effect
- More stable than other strategies when b1 is smaller

IPCW strategy

- Most stable strategy when b1 is larger
- No obvious pattern observed

Calibration slope

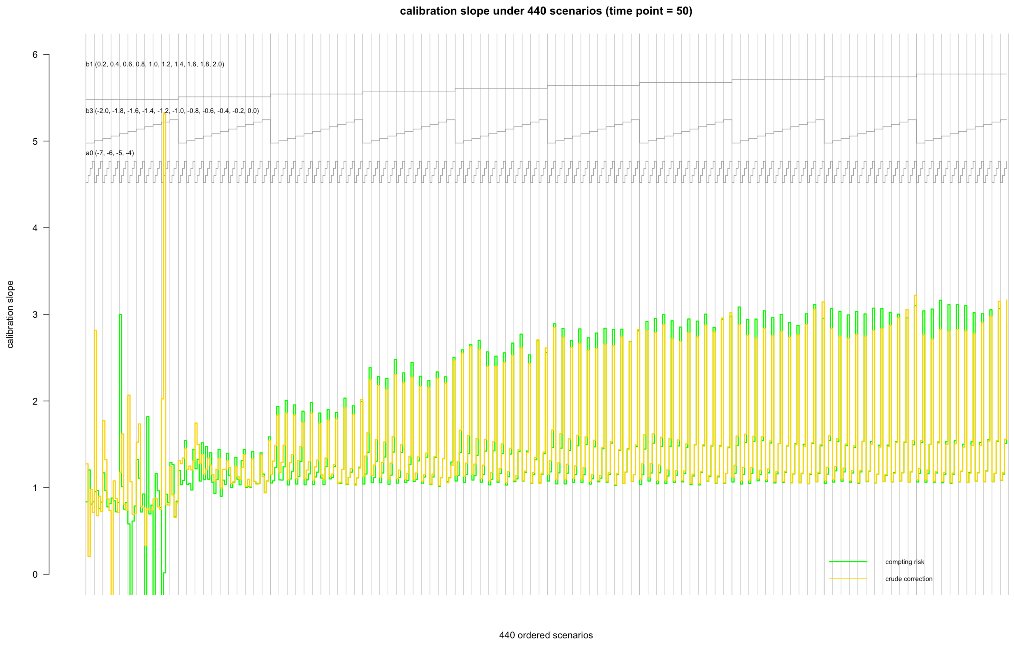


Figure 12 Nested loop plot for calibration slopes of competing risk strategy (green line) and "crude" correction strategy (yellow line) under 440 scenarios

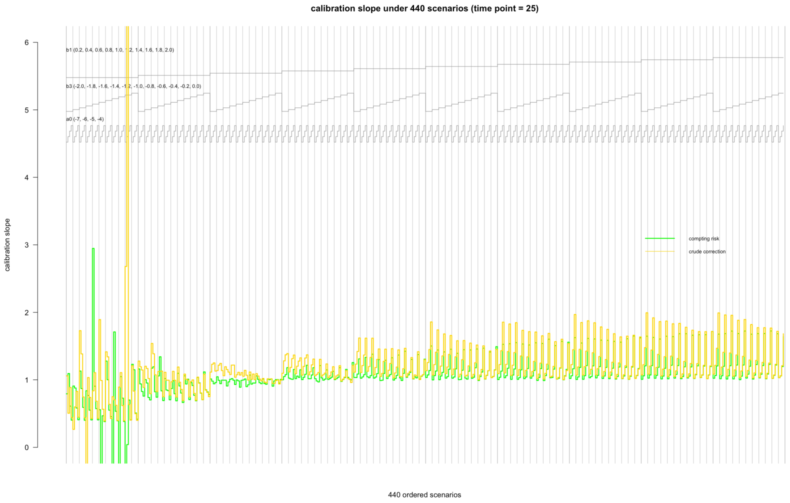
Competing risk strategy

- Larger a0 leads to more treatment use and larger deviation from 1
- b3 and b1 do not affect that much

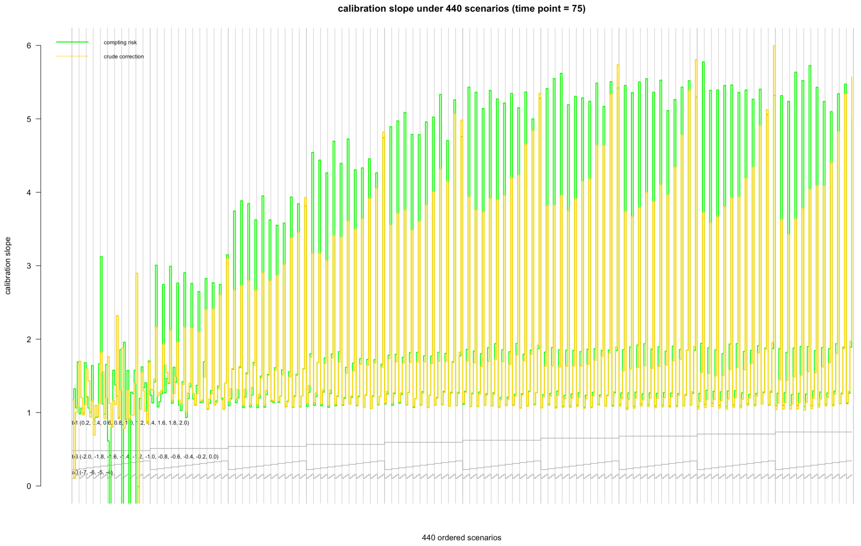
"crude" correction strategy

- Affected by a0 and b1
- Different time points will change the patterns. This is because the product is prone to changing dramatically when the probability of people not getting treatment decreases

Calibration slope



Time point = 25



Time point = 75

C index

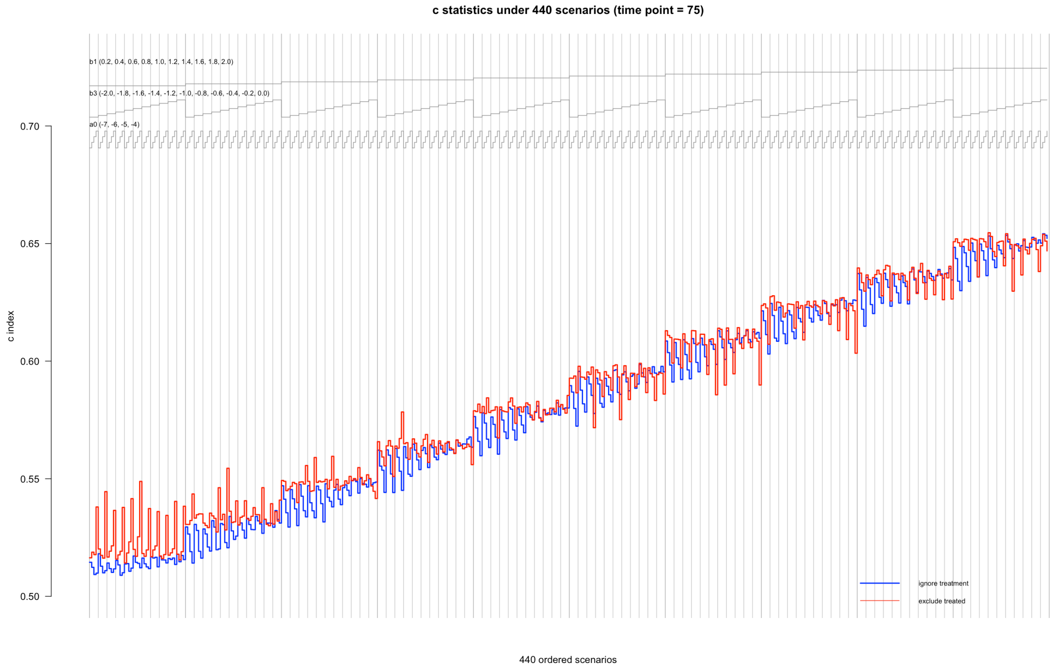


Figure 14 Nested loop plot for c index of ignorance strategy (blue line) and exclusion strategy (red line) under 440 scenarios

Ignorance strategy

- Larger a0 will slightly lower c index
- b1 shows a dominating role in discrimination

Exclusion strategy

- b1 shows a dominating role in discrimination
- When b1 is extremely small, excluding treatment can increase c index; while with the increasing b1, treatment exclusion will be affected by the selection process, thus influence c index

C index

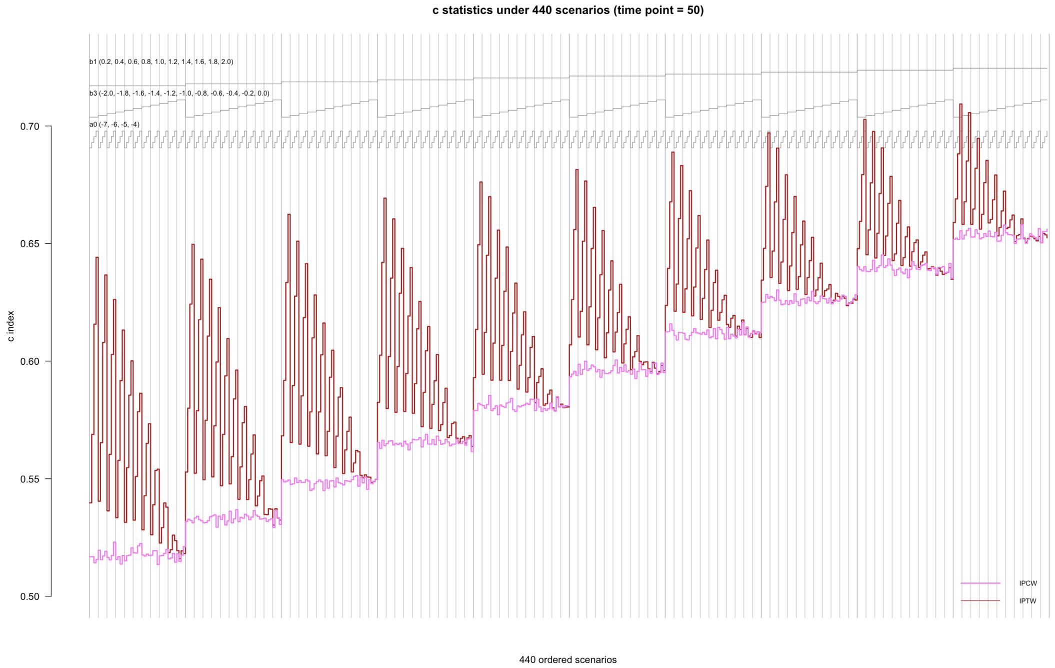


Figure 15 Nested loop plot for c index of IPCW strategy (pink line) and IPTW strategy (brown line) under 440 scenarios

IPTW strategy

- Highest c index among six strategies
- Still be affected by three parameters

IPCW strategy

- Most stable strategy but lower c index
- b1 shows a dominating role in discrimination

C index

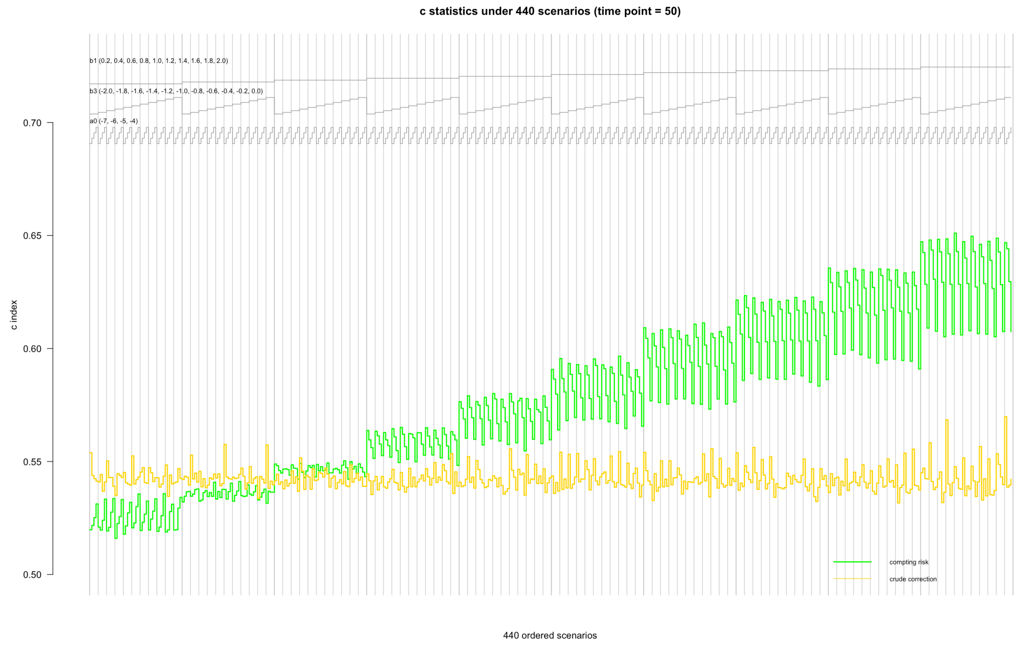


Figure 16 Nested loop plot for c index of competing risk strategy (green line) and "crude" correction strategy (yellow line) under 440 scenarios

Competing risk strategy

- b1 shows a dominating role in discrimination
- When b1 is extremely small, including competing risk can increase c index; while with the increasing b1, competing risk will take more account into treatment but is conflicted with the predictive ability of P, thus affected c index

"crude" correction strategy

- Performs worst
- When b1 grows, observed predictor will dominate the occurrence of outcome rather than treatment. In other words, there is informative censoring which leads to more biased estimates from the latter part, $\Pr(V > t.hor|P)$.

Summary

- Under most strategies, larger probability of treatment (larger a0), larger treatment effect (more negative b3) will make performance worse.
- b1 (the effect of observed predictor on outcome) only has a leading role in better discrimination but has limited ability on increasing calibration.
- IPTW and IPCW work much better than other strategies. IPTW models treatment so that it is good for discrimination; IPCW censors treatment so that it is better for calibration.
- Choosing different time points would not affect the performance of each strategy, but will be influenced by the proportion of people getting outcome
- Competing risk and “crude” correction perform worst among six strategies. The predicted probability will be more likely to become extremely low.

Thanks!