Comparison of Statistical Methods for Assessing Vaccine Efficacy in COVID-19 Vaccine Efficacy Trials

Alex Luedtke

ID-PRISM Center Seminar – May 27, 2020

Acknowledgements: Fred Hutch and UW Statistics/Biostatistics Faculty, Staff, Students Working on COVID-19 VE Trials

- **Fred Hutch/UW HVTN+HPTN+IDCRC:** Many individuals and growing; e.g. Lindsay Carpp, Tom Fleming, Yunda Huang, Holly Janes, Michal Juraska, **Peter Gilbert**
- **NIAID Biostatistics:** Dean Follmann, Martha Nason
- NIH has launched the COVID-19 Prevention Network (Corey, Mascola, Fauci, Collins, *Science*, May 2020)
- WHO soon launching Solidarity Trial for COVID-19 Vaccines
- **Collaborators on covariate adjustment for COVID-19 studies:** Xiudi Li (UW Biostats), Incheoul Chung (UW Stats), Tzu-Jung Huang (UW Stats), Sijia [Lucy] Li (UW Biostats), Marco Carone (UW Biostats), David Benkeser (Emory), Michael Rosenblum (JHU), Iván Díaz (Weill-Cornell), Dan Scharfstein (JHU), Jodi Segal (JHU)

Disclaimer: Everything that I share today reflects my views, and doesn't necessarily reflect those of any of the individuals or organizations acknowledged on the previous slide.

Table of Contents

Basic setup for COVID-19 vaccine efficacy trials

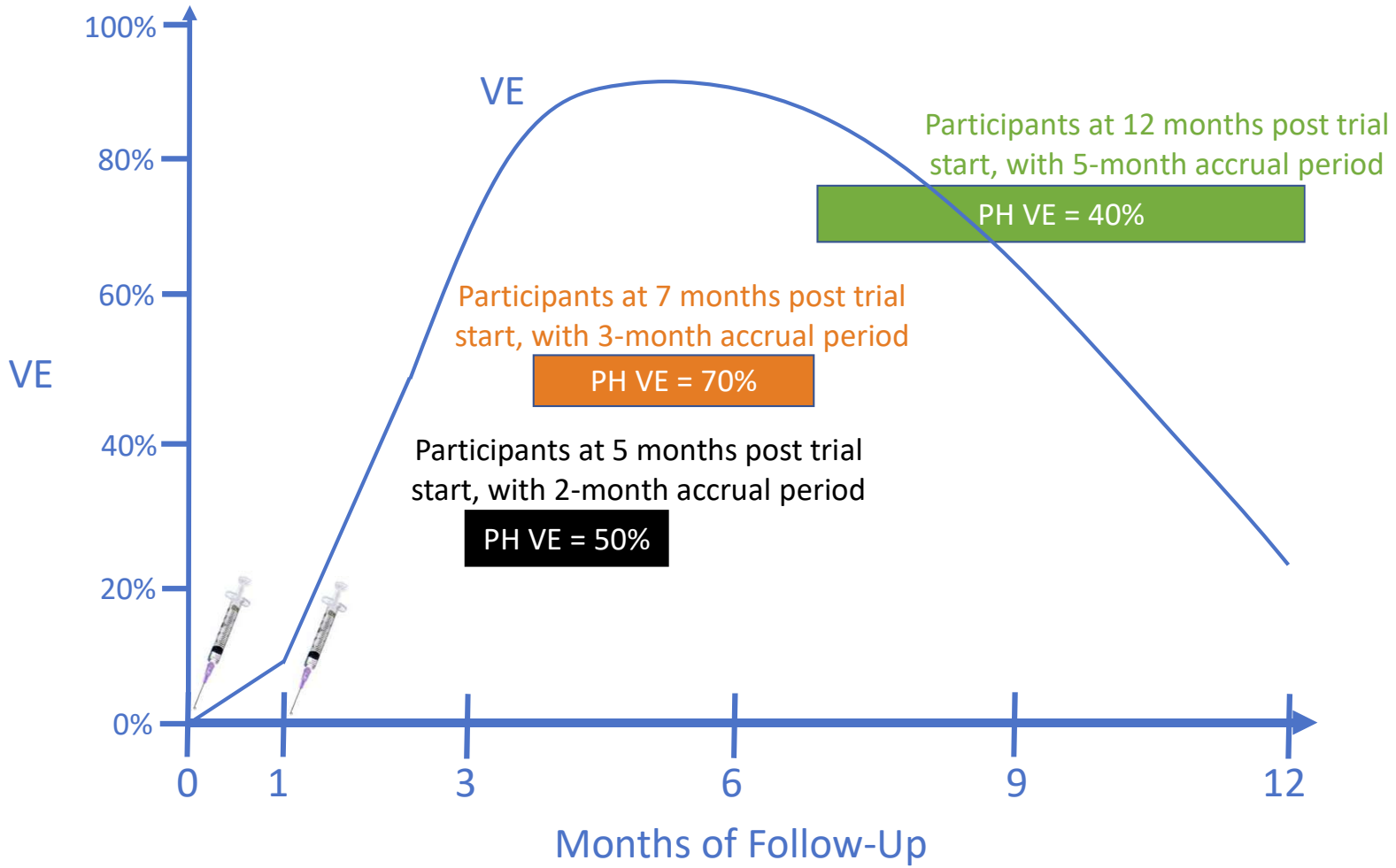
- Trials will enroll quickly ($\approx$ 3 months)
- Analyses will be conducted as quickly as possible
  - Event-driven designs, with interim analyses planned to test the null of vaccine efficacy (VE) $\leq$ 20% or VE $\leq$ 30%
- In each trial, it's likely that these interim analyses and the primary analysis for efficacy will be based on estimates of the Cox proportional hazards VE

$$1 - \beta,$$

where β is the hazard ratio between the vaccine and placebo groups

Proportional hazards VE in the presence of ramping/waning efficacy

- Inference for the proportional hazards VE assumes that β is constant over time, which, as Peter discussed last week, may not hold when there is ramping or waning VE



Cumulative incidence VE

- The trials specify secondary analyses that assess time-varying VE and durability using the cumulative incidence VE (VE hereafter) through time t for a range of times t.
- The VE is defined as

$$1 - \frac{P(T \leq t \mid A = 1)}{P(T \leq t \mid A = 0)},$$

  i.e., one minus the ratio of cumulative incidences of the primary endpoint (COVID-19 disease diagnosis) by time t in the vaccine arm ($A = 1$) and placebo arm ($A = 0$).
- The VE also writes as

$$\frac{S(t \mid 1) - S(t \mid 0)}{1 - S(t \mid 0)},$$

  where $S(t \mid a) := P(T > t \mid A = a)$ is the survival function in vaccine arm a.
- This suggests estimating the VE using an **arm-specific Kaplan-Meier estimator**.

Table of Contents

Condition for the validity of the arm-specific Kaplan-Meier estimator

- The validity (consistency, asymptotic normality) of the arm-specific Kaplan Meier estimator of $S(\cdot \mid a)$ relies on the condition that

$$T \perp\!\!\!\perp C \mid A = a,$$

that is, that the time of COVID-19 disease (relative to enrollment) is independent of right-censoring time (relative to enrollment) conditionally on vaccine arm $A = a$.

  - The validity for estimating $S(t \mid a)$ at a particular t requires the weaker condition that

$$\min\{T, t\} \perp\!\!\!\perp \min\{C, t\} \mid A = a,$$

and also requires that $P(C \geq t \mid A = a) > 0$.

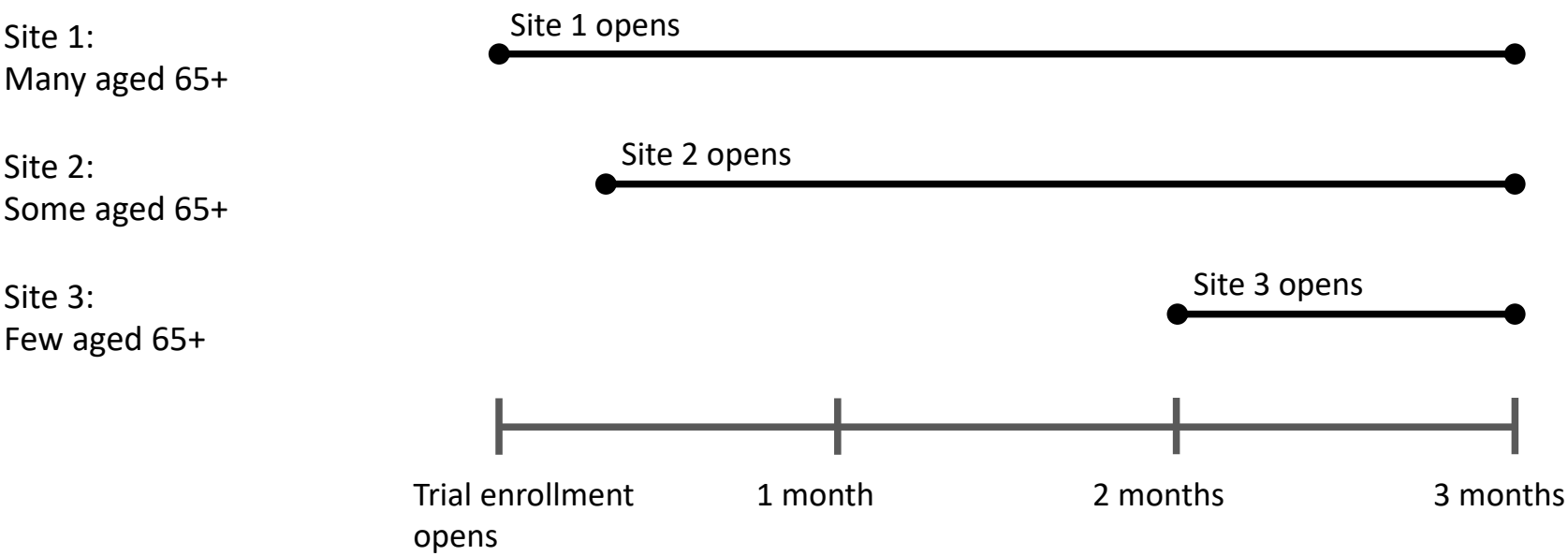
Arm-specific Kaplan-Meier estimator

- During an outbreak, we anticipate high participant retention ($\approx$ 98% over 12 months),[1] and therefore expect that most censoring events will be administrative.
  - Because trials will enroll quickly ($\approx$ 3 months), properly accounting for (potentially informative) censoring will be most important in early analyses.
- So, concerns related to the assumption that $T \perp\!\!\!\perp C \mid A$ are most likely to arise due to enrollment patterns.

---

[1] S. Browne, T. Carter, R. Eckes, G. Grandits, M. Johnson, I. Moore, and L. McNay. A review of strategies used to retain participants in clinical research during an infectious disease outbreak: The PREVAIL I Ebola vaccine trial experience.
Contemporary clinical trials communications, 11:50–54, 2018

Example: informative censoring due to staggered site entry times



- Suppose that VE is lower in those aged 65+, so that VE is lowest at site 1 and highest at site 3.
- Suppose an analysis to evaluate VE through 6 months post-enrollment is conducted 6.5 months after the trial opens.
- Because the site with the highest VE has no follow-up beyond Month 4.5, the arm-specific Kaplan-Meier VE estimator is biased towards the null.
  - Specifically, biased for the VE over 6 months for the population enrolled in the trial (which includes a lower proportion of people aged 65+ than are represented in those sites with follow-up past 4.5 months).

Potential strategy: stratify by joint predictors of T and C

- Suppose that there is a discrete baseline characteristic X that takes on finitely many values (e.g., sex, age category) and such that, for all levels x,

$$T \perp\!\!\!\perp C \mid A = a, X = x,$$
$$P(C \geq t \mid A = a, X = x) > 0.$$

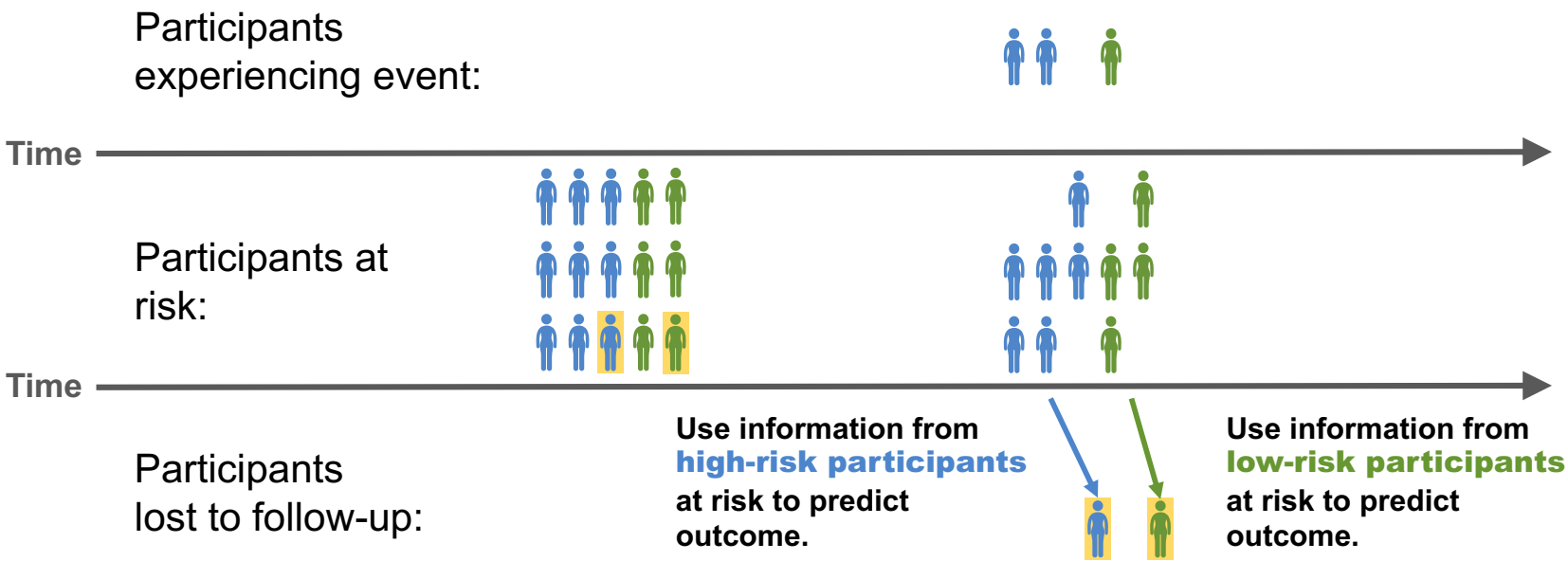
- Then, by our earlier discussion, a valid estimator of $S(t \mid a, x) := P(T > t \mid A = a, X = x)$ is given by **fitting a Kaplan-Meier estimator $\widehat{S}(\cdot \mid a, x)$ among all participants with $(A, X) = (a, x)$.**
- Also, by the law of total expectation,

$$S(t \mid a) = \sum_x S(t \mid a, x)\, p(X = x | A = a).$$

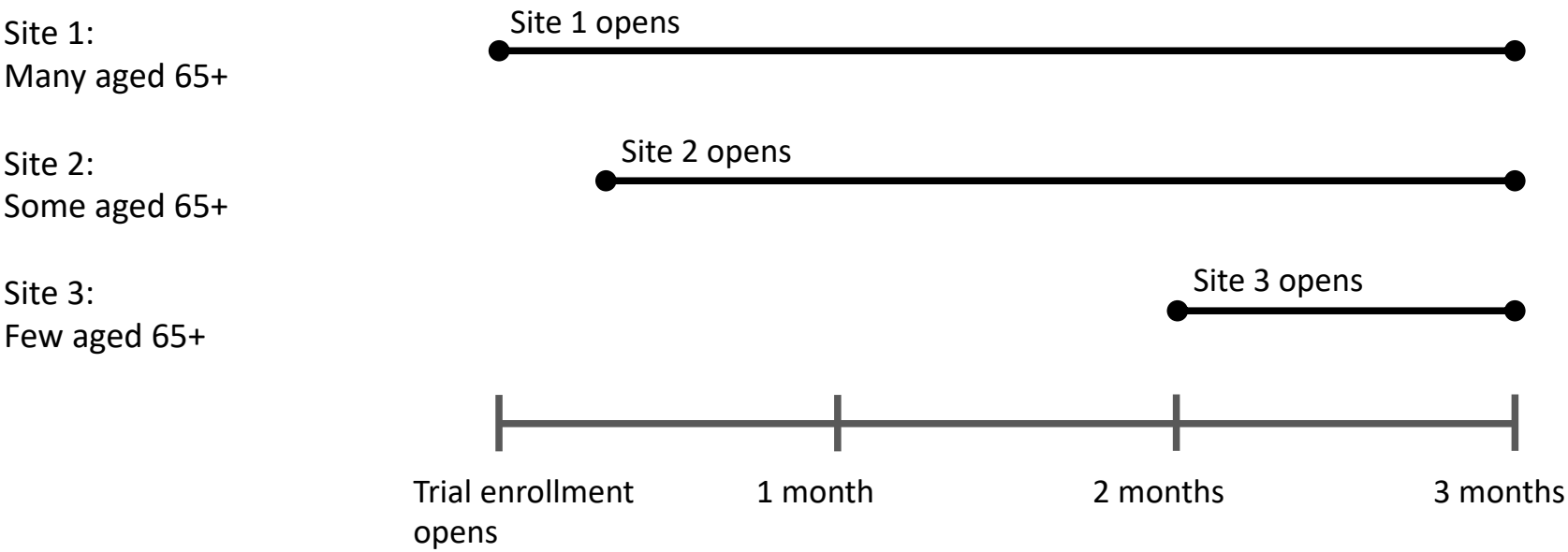
And so, $S(t \mid a)$ can be estimated by

$$\sum_x \widehat{S}(t \mid a, x)\, \underbrace{\widehat{p}(X = x | A = a)}_{\text{empirical in arm } a}.$$

Another view of this covariate-adjusted estimator



Recall example: informative censoring due to staggered site entry times



- VE(0-6) is estimated 6.5 months after the trial opens
- In this scenario, if $X =$ site, then adjusted estimator will fail since $P(C \geq 4.5 \,|\, A = a, X = x) = 0$ for site 3!
- **Take-away message:** exercise caution when using site as an adjustment covariate for estimating cumulative incidence VE.

Can we benefit from our knowledge of the fact that A is randomized?

- Recall that

$$S(t \mid a) = \sum_x S(t \mid a, x)\, p(X = x | A = a).$$

- So far, we've proposed estimating this quantity with

$$\sum_x \widehat{S}(t \mid a, x)\, \underbrace{\widehat{p}(X = x | A = a)}_{\text{empirical in arm } a}. \tag{1}$$

- But, if we know that $A \perp\!\!\!\perp X$, then we can improve our estimate of $p(X = x|A = a)$ by using observations from all participants (rather than just those in arm a). This suggests using the estimator:

$$\sum_x \widehat{S}(t \mid a, x)\, \underbrace{\widehat{p}(X = x)}_{\text{empirical}}.$$

- Can show that this estimator generally has **lower variance** than does the estimator in (1).

What should we do when X takes many values or is continuous?

- Zeng (2004) suggested a simple approach for adjusting for baseline covariates when they are continuous,[2] or even high-dimensional.
  - Here we present a stratified variant of this estimator for $S(t|a)$, with a fixed.

- The dimension of the baseline covariate is reduced by fitting two standard Cox models based on participants from arm a:
  1. Using T **as the event** and covariate X, yielding coefficient $\hat{\beta}_a$.
  2. Using C **as the event** and covariates X, yielding coefficient $\hat{\gamma}_a$.

- Define a new 2-dimensional covariate $Z_i := (Z_{i1}, Z_{i2})$, where
  1. Z_{i1} is the empirical quartile into which $\hat{\beta}_a^\top X_i$ falls.
  2. Z_{i2} is the empirical quartile into which $\hat{\gamma}_a^\top X_i$ falls.

- Fit the already-discussed covariate-adjusted Kaplan-Meier estimator, stratifying on Z instead of X.

---

[2] D. Zeng. Estimating marginal survival function by adjusting for dependent censoring using many covariates.
The Annals of Statistics, 32(4):1533–1555, 2004

Properties of estimator

- **Doubly robust:** consistent and asymptotically normal if one of the two Cox models is correct and Z is defined as an increasingly fine discretization of $\hat{\beta}_a^\top X_i$ and $\hat{\gamma}_a^\top X_i$.

- **Wald-type confidence intervals:** can be constructed using standard techniques.

  - In the simulations, we develop intervals for the log relative risk $\log P(T \leq t|A = 1)/P(T \leq t|A = 0)$, which can then be transformed to yield intervals for the VE.

Table of Contents

Objectives

- Conducted a simulation study to (roughly) mimic the setting of upcoming COVID-19 VE trials.
- Main goal will be to study the relative performance of covariate-unadjusted and covariate-adjusted estimators of cumulative incidence VE across a range of settings, varying:
  - Level of VE
  - How predictive the covariates are of:
    - arm-specific outcomes
    - conditional vaccine efficacy
    - censoring time
  - Timing of analysis (relative to time that the trial opened)
- In each setting, I'll try to provide a heuristic explanation for the relative performance of the adjusted and unadjusted approaches.

Basic setup: incidence and VE

- Uniform enrollment of 20k participants (10k vaccine, 10k placebo) over 3 months
- 2% loss to follow-up uniformly over 12 months
- 6-month incidence of 1% in the placebo arm
- Cumulative incidence VE over 6 months is either:
  - Null: 30%
  - Alternative: 60%
- Except where otherwise specified, we focus on analysis conducted 6 months after initiation of trial (on average 4.5 months follow-up per participant)
  - Expect ≈ 128 events under null, ≈ 105 events under alternative

Basic setup: covariates

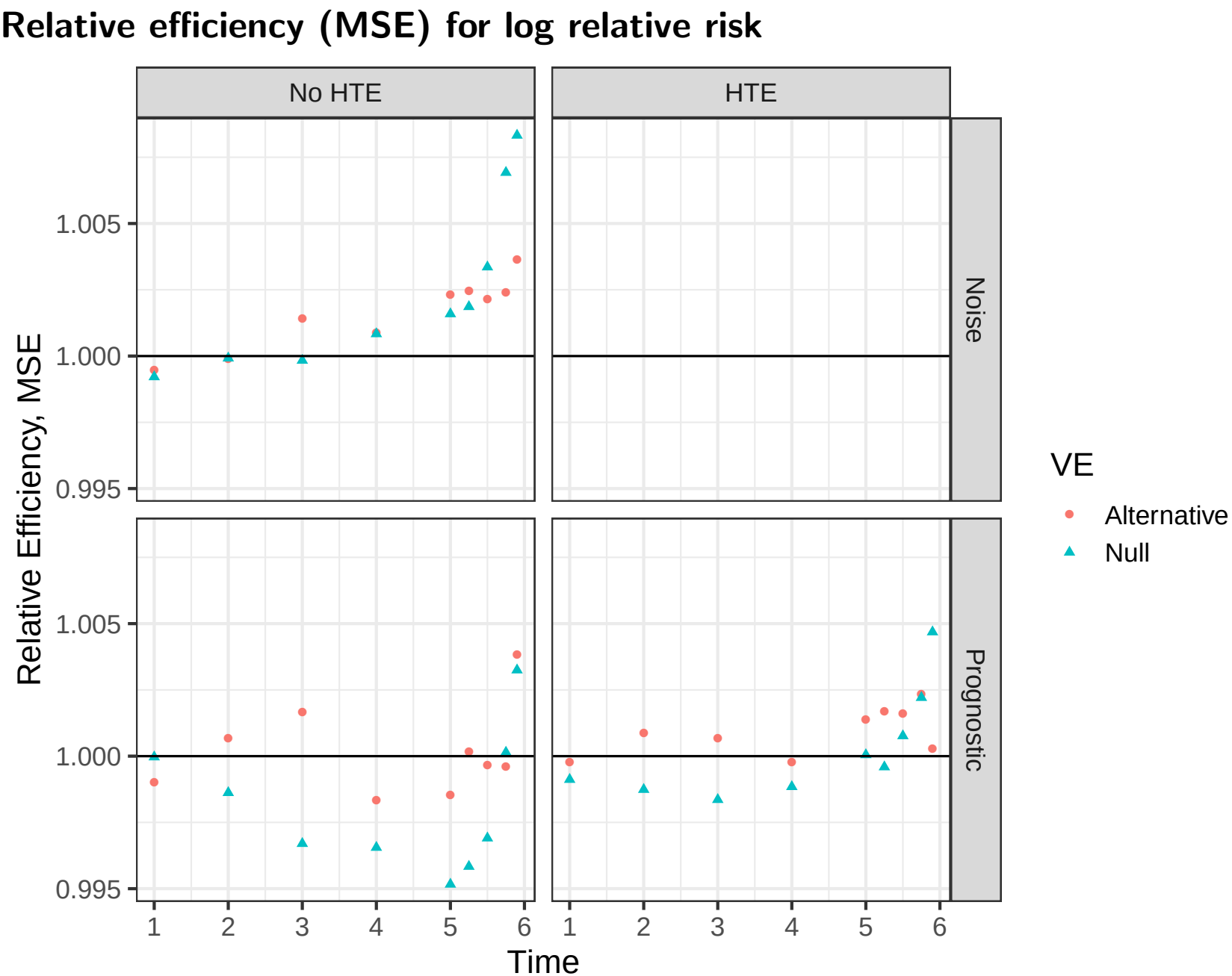
- 8 baseline covariates $X_1, \ldots, X_8$, iid $\mathrm{Unif}(-1, 1)$
- $X_1, \ldots, X_4$ are always noise, independent of both time to COVID-19 disease and censoring time
- $X_5, \ldots, X_8$ play differing roles depending on the simulation
  - When they are predictive of T or C, they are predictive through $W := \sum_{j=5}^{8} X_j$
  - In each vaccine arm, T is simulated according to a Cox model with the dependence on W chosen so that the hazard satisfies:

    No HTE: $\lambda(t|A = 0, w) \propto \exp(W)$ $\lambda(t|A = 1, w) \propto \exp(W)$,

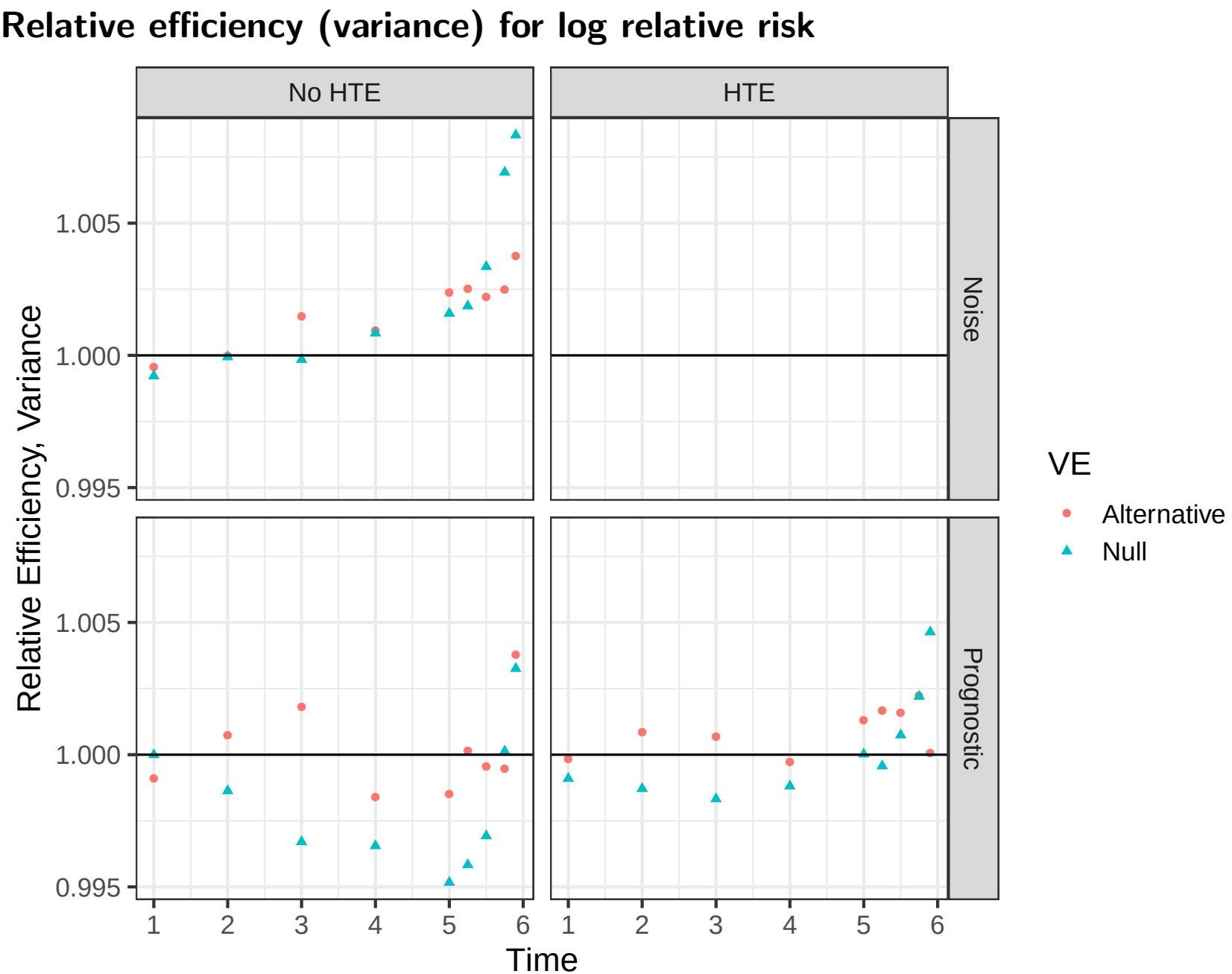
    HTE: $\lambda(t|A = 0, w) \propto 1$ $\lambda(t|A = 1, w) \propto \exp(W)$.

    The proportionality constants are selected to attain the incidence/VEs described on the preceding slide.
  - Enrollment time (relative to time at which the first participant enrolled) is chosen in such a way that enrollment is uniform (marginal over covariates) and has either
    - Negative Correlation with W: enrollment time and W have correlation -0.2. When W also predicts T, C **and** T **are negatively correlated** (**Both-Neg**).
    - Positive Correlation with W: enrollment time and W have correlation $+0.2$. When W also predicts T, C **and** T **are positively correlated** (**Both-Pos**).

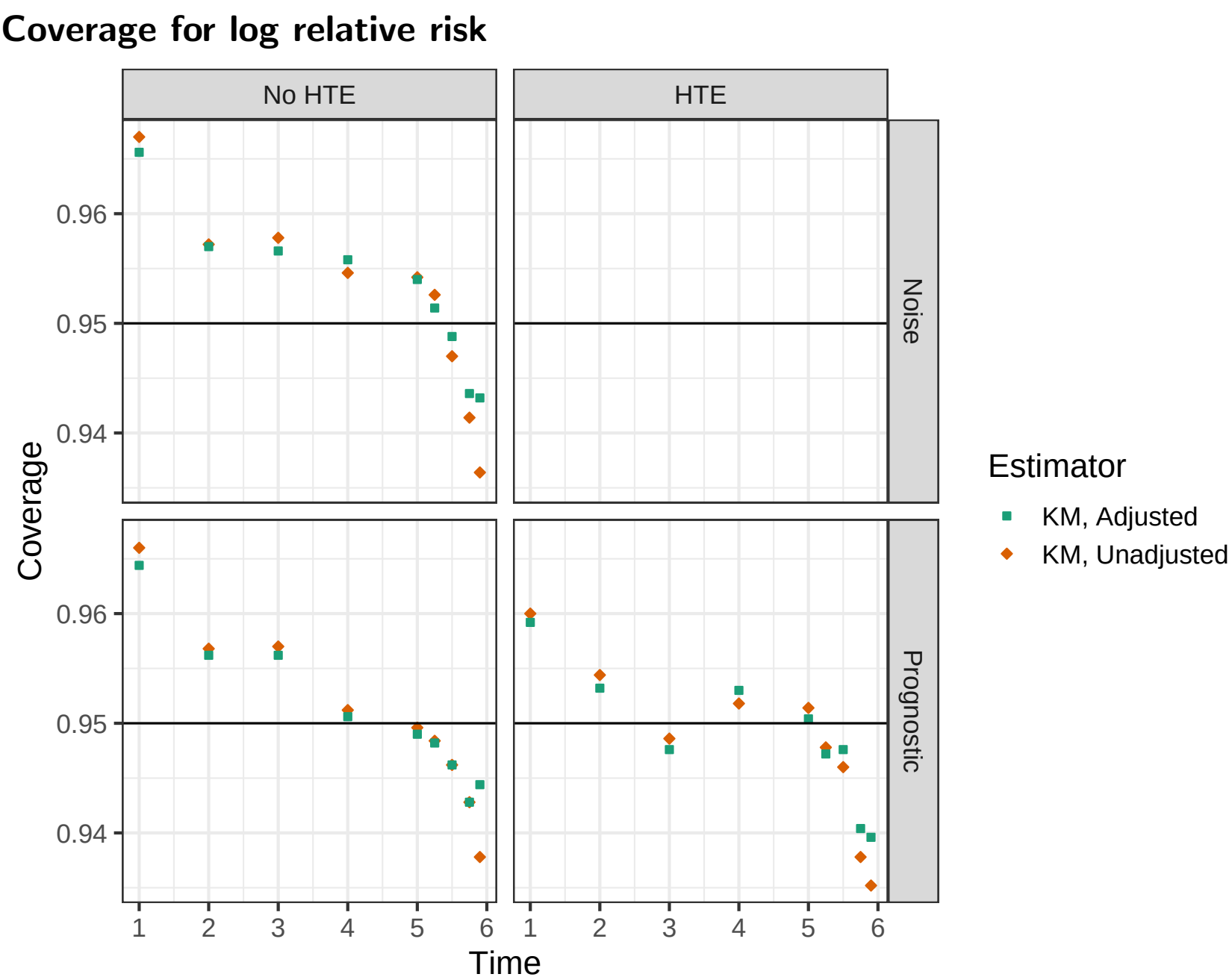
Case 1: covariates not predictive of censoring time



Case 1: covariates not predictive of censoring time



Case 1: covariates not predictive of censoring time



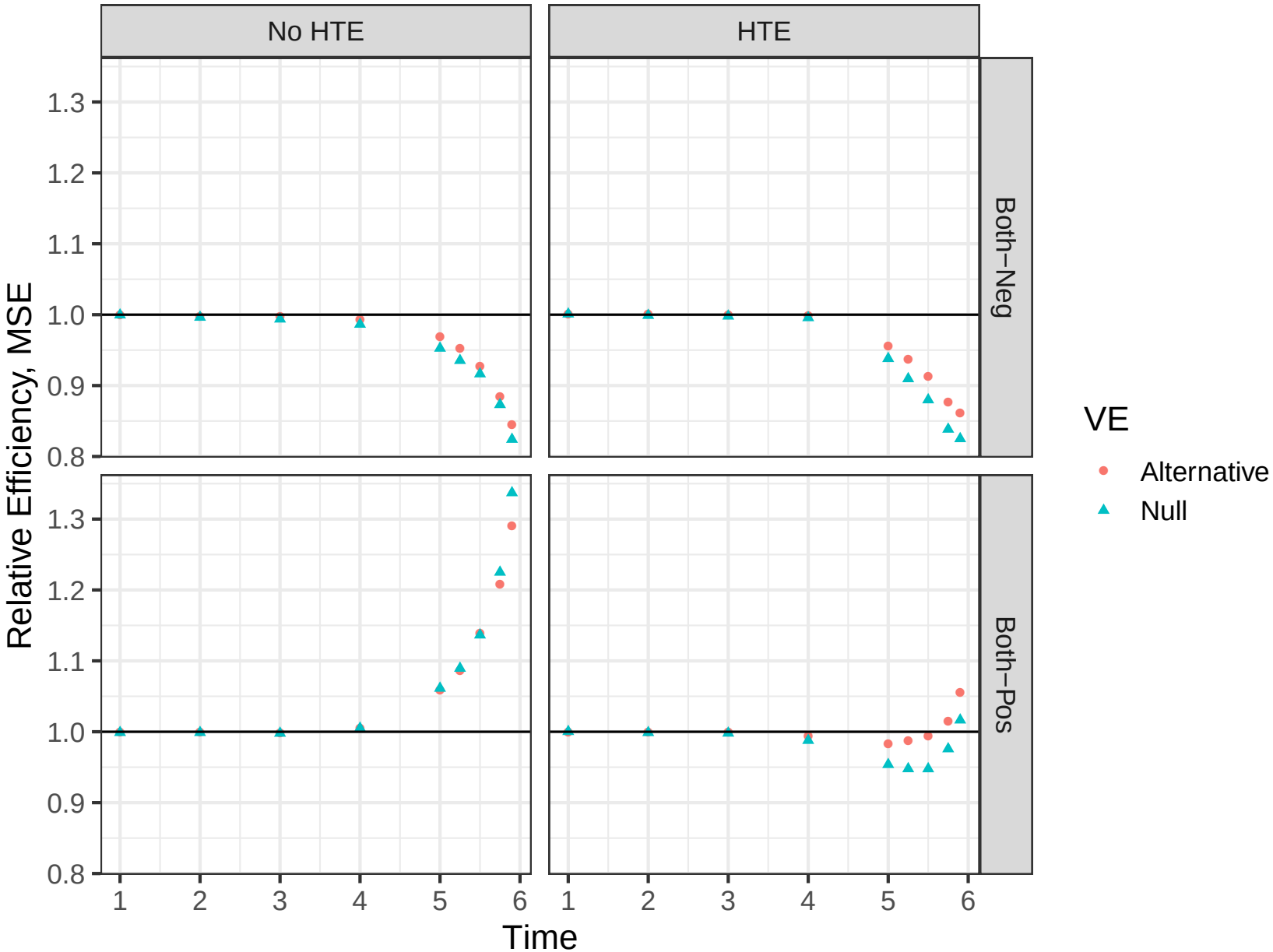
Case 1: covariates not predictive of censoring time

- **Take-away message:** When covariates are not predictive of censoring time, covariate adjustment does not appear to meaningfully improve precision in this rare event setting.

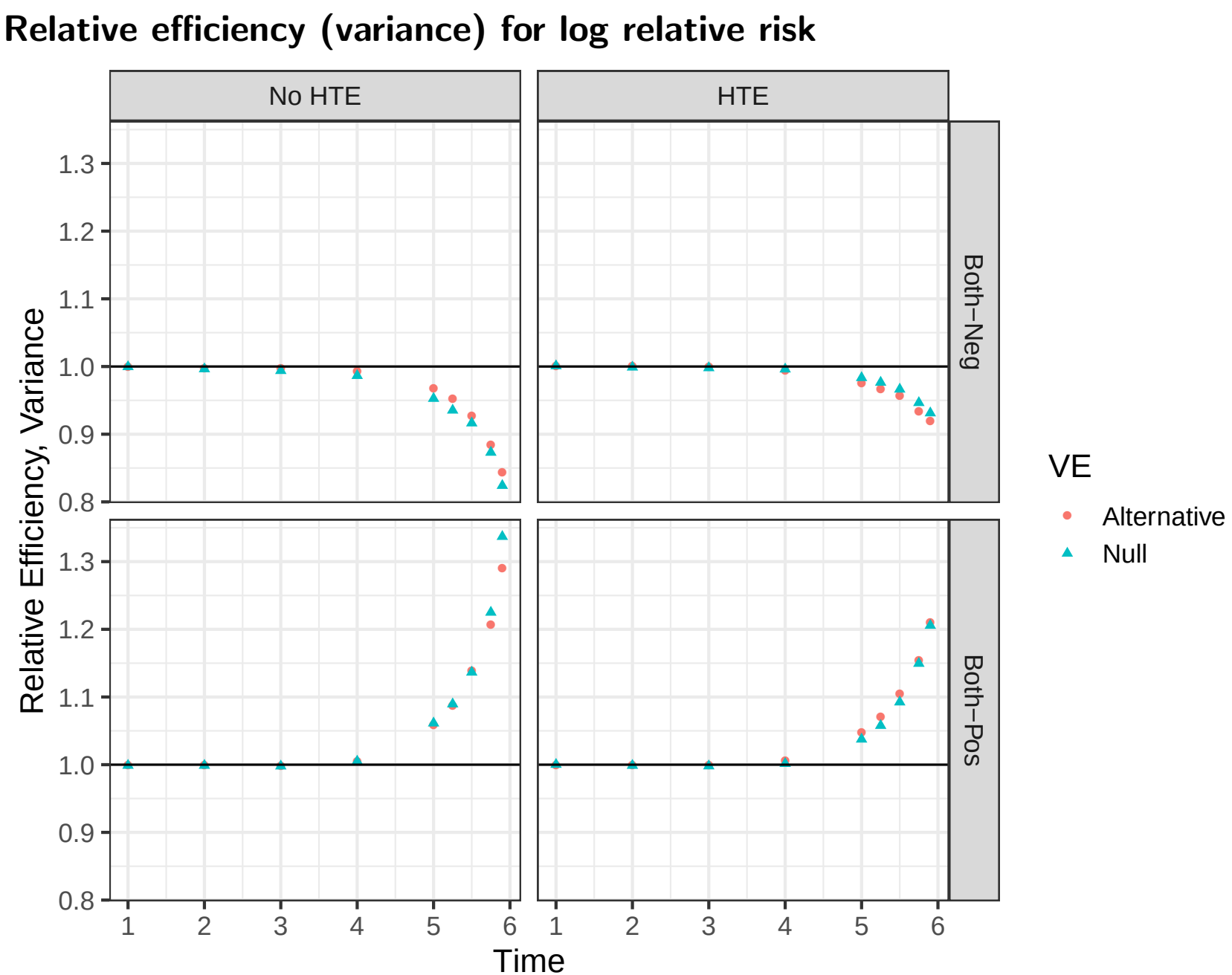
  But: it is also not harmful, and in fact leads to slight improvements in coverage.

Case 2: covariates predictive of censoring time and time to COVID-19

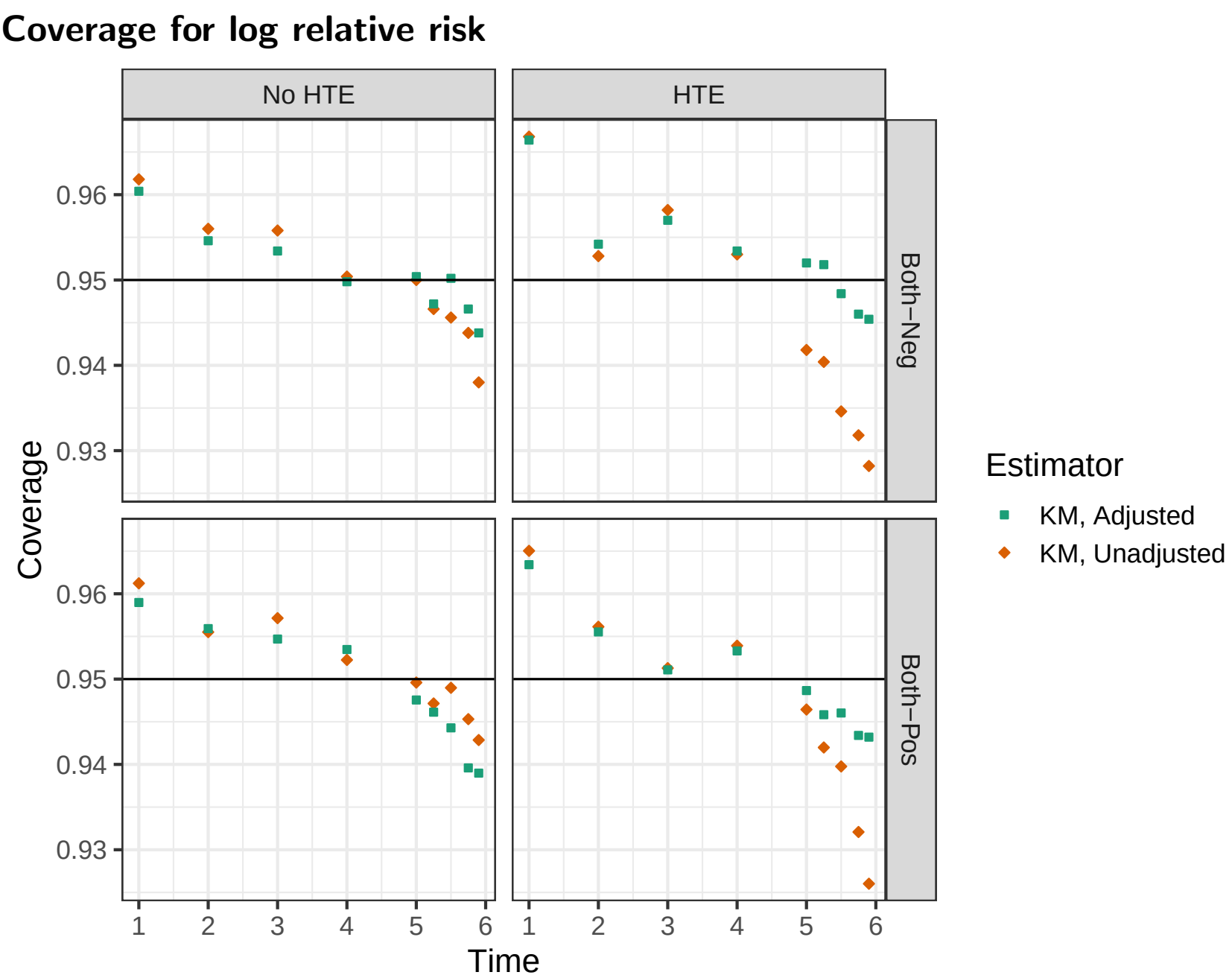
Relative efficiency (MSE) for log relative risk



Case 2: covariates predictive of censoring time and time to COVID-19

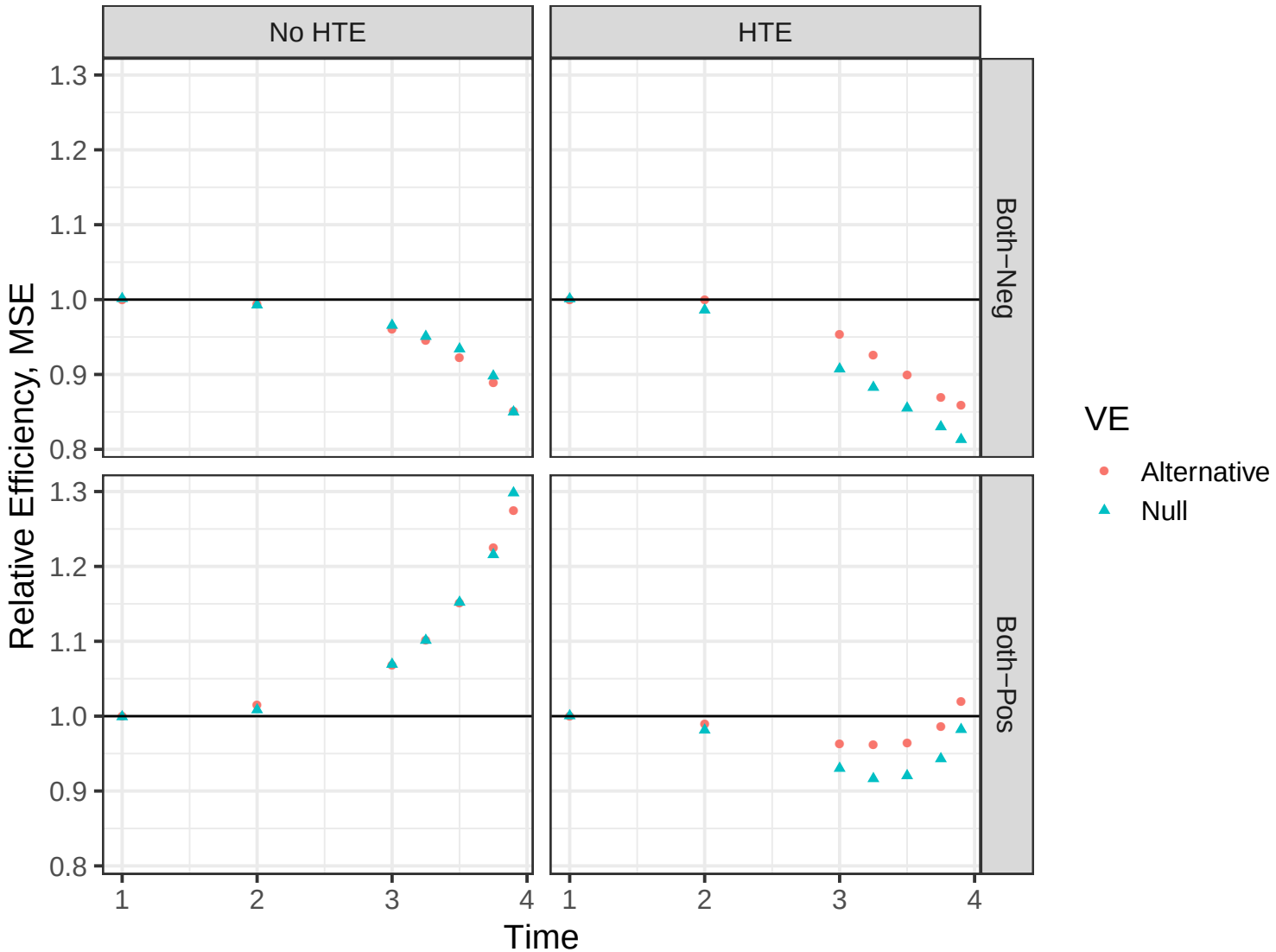


Case 2: covariates predictive of censoring time and time to COVID-19



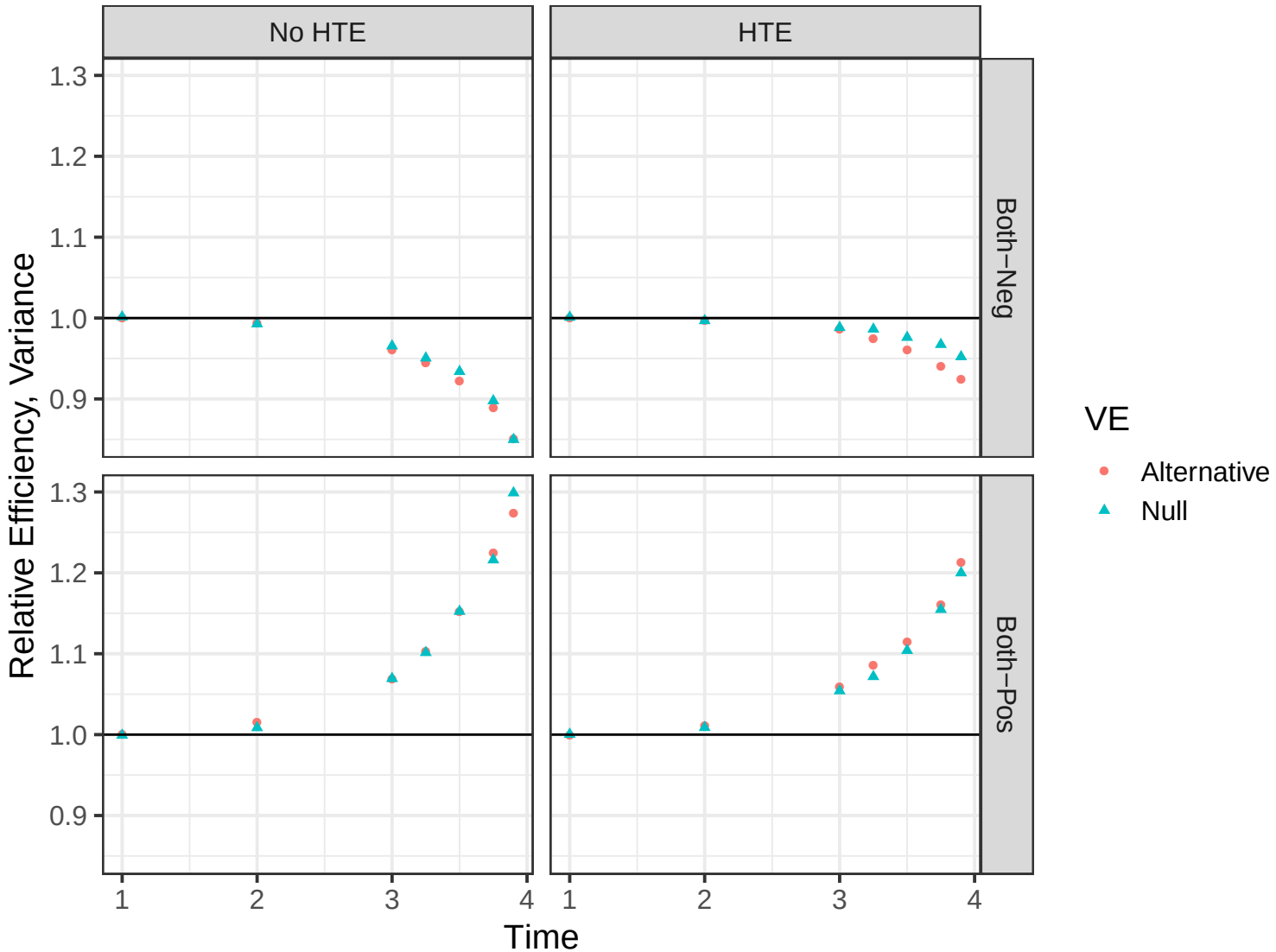
Case 2: covariates predictive of censoring time and time to COVID-19

Relative efficiency (MSE) for log relative risk, analysis conducted 4 months after trial initiation



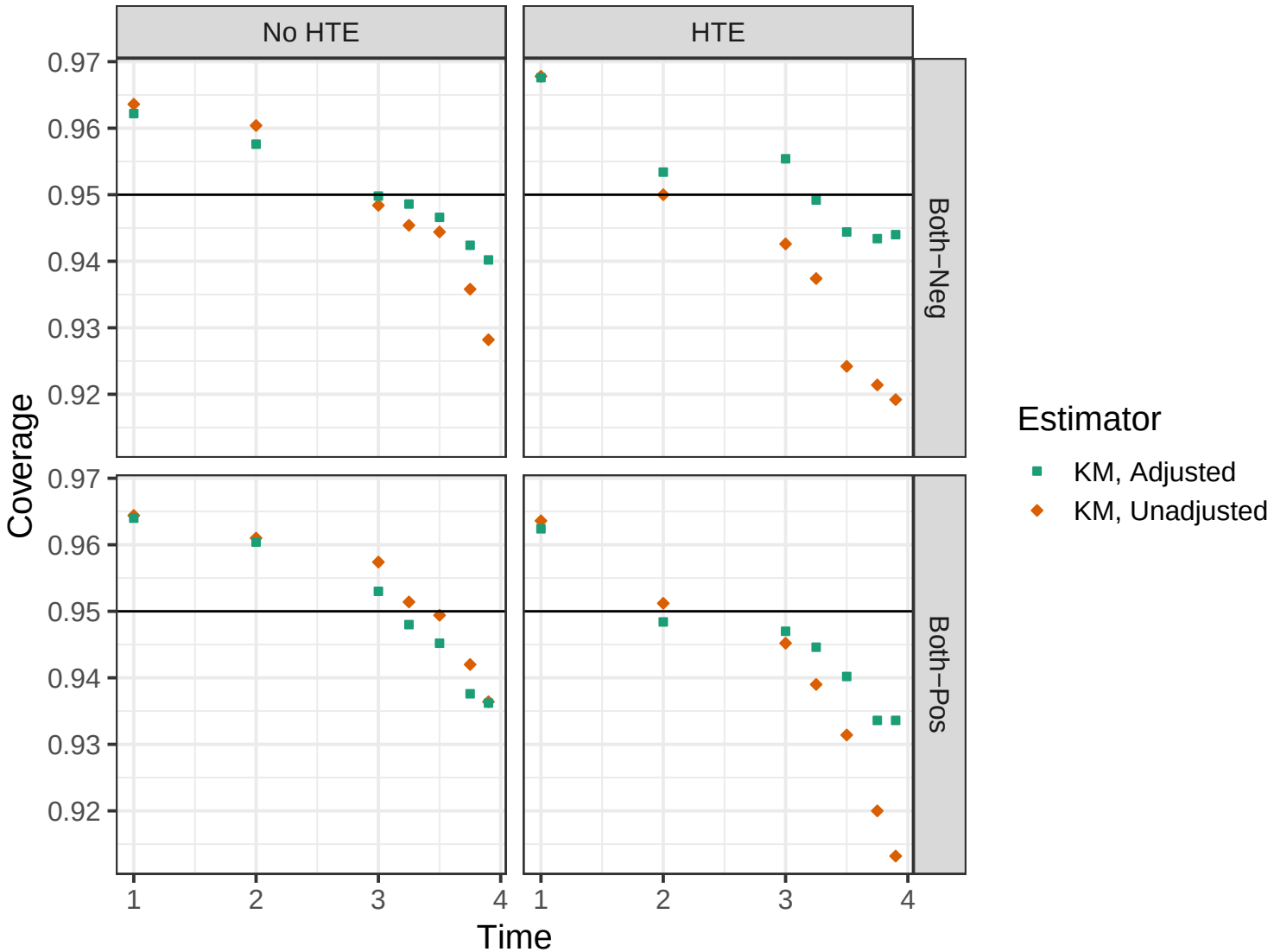
Case 2: covariates predictive of censoring time and time to COVID-19

Relative efficiency (variance) for log relative risk, analysis conducted 4 months after trial initiation



Case 2: covariates predictive of censoring time and time to COVID-19

Coverage for log relative risk, analysis conducted 4 months after trial initiation



Case 2: covariates predictive of censoring time and time to COVID-19

Take-away message 1: When covariates are predictive of both censoring time and time to COVID-19 disease, then, for large times t:

- If there is **effect modification**, adjustment can lead to meaningful improvements in coverage.
- If there is **no effect modification**, coverage is similar.
  - In results not shown, have seen that unadjusted estimators of arm-specific survival functions are still biased in this setting, but this bias essentially cancels out when estimating the relative risk.

Case 2: covariates predictive of censoring time and time to COVID-19

Take-away message 2: When covariates are predictive of both censoring time and time to COVID-19 disease, then, for large times t:

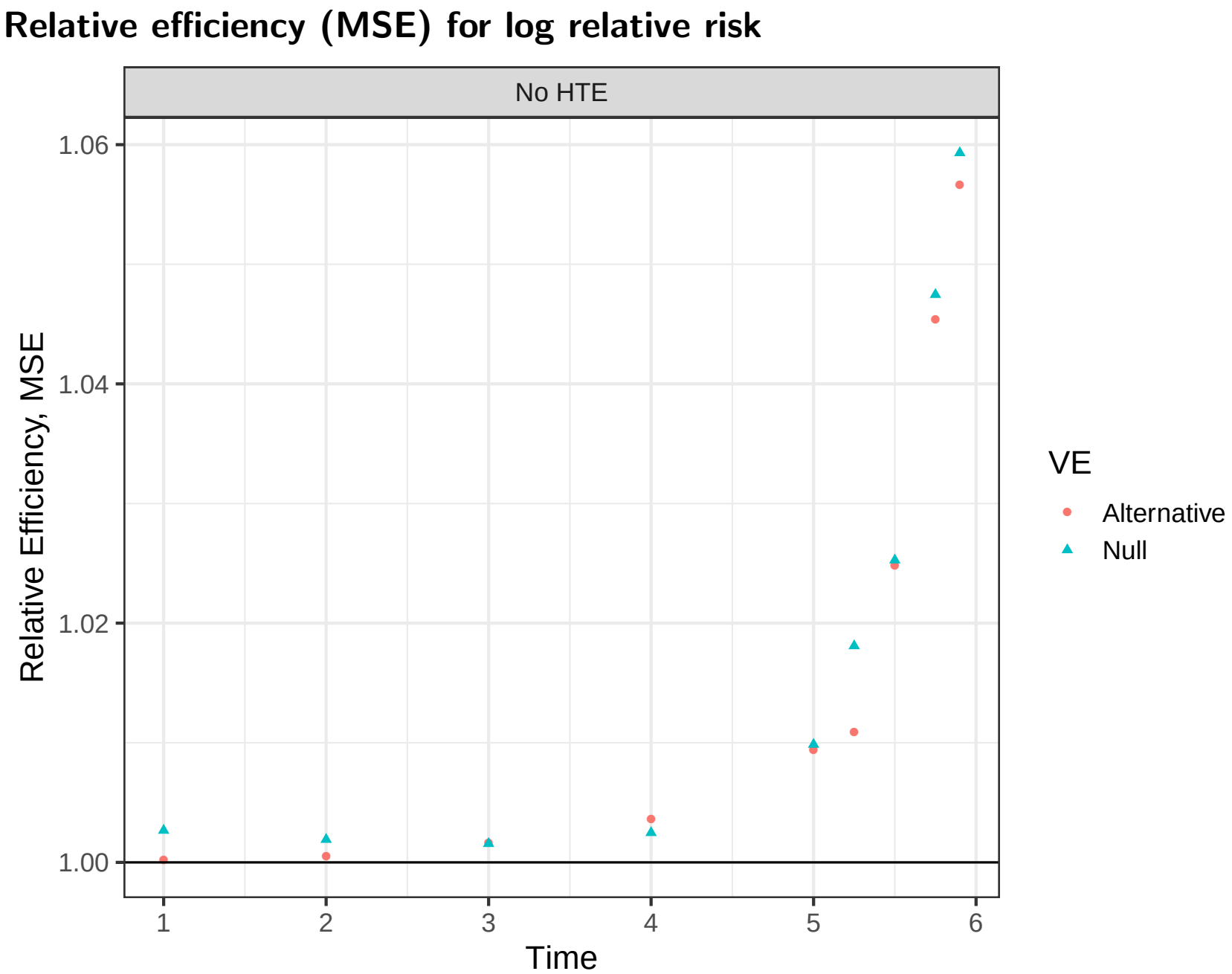
- If C and T have **opposite directions of association** with W and are negatively correlated (Both-Neg), then **adjusting for covariates increases precision**.
- If C and T have the **same direction of association** with W and are positively correlated (Both-Pos), then **adjusting for covariates reduces precision**.

Why? See next slide.

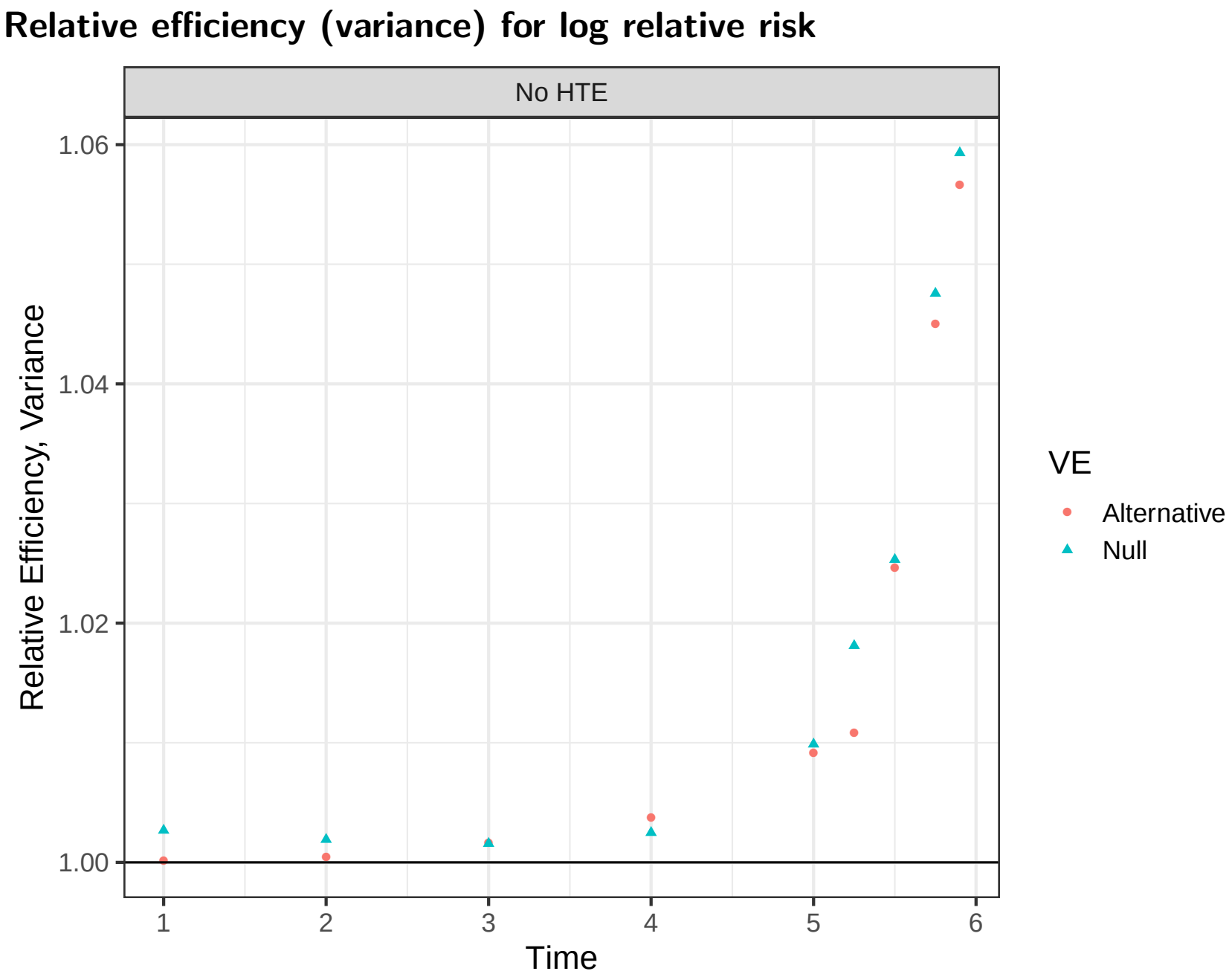
Case 2: covariates predictive of censoring time and time to COVID-19

- Consider the scenario from the previous slide in which C and T have **opposite directions of association** with W.
- Events that occur before time t will often be observed (because the censoring time will tend to be large for these participants).
- More specifically, $P(C \geq c|W)|_{c=T}$ tends to be relatively close to 1 for most events that occur before time t.
- In the analysis, it is approximately true that an event for participant i is weighted as $1/P(C \geq c|W_i)|_{c=T_i}$.
- So, the fact that $P(C \geq c|W)|_{c=T}$ is relatively close to 1 means there's little variability induced by accounting for censoring — and in fact, **adjusting for covariates increases precision** because, when a participant is censored early, it is not difficult to use W to categorize these people into low and high risk groups for experiencing the event.

Case 3: covariates only predictive of censoring time

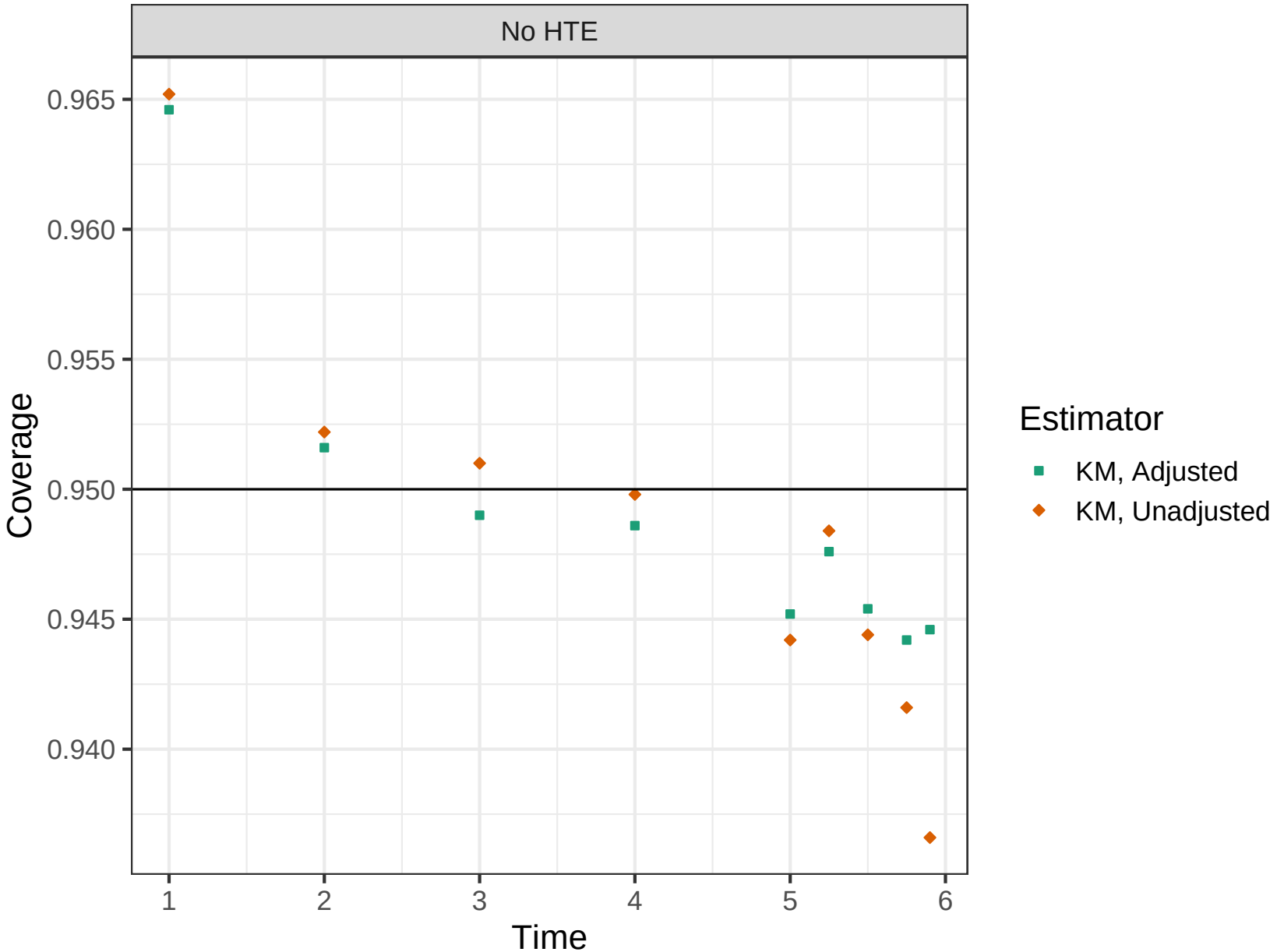


Case 3: covariates only predictive of censoring time



Case 3: covariates only predictive of censoring time

Coverage for log relative risk

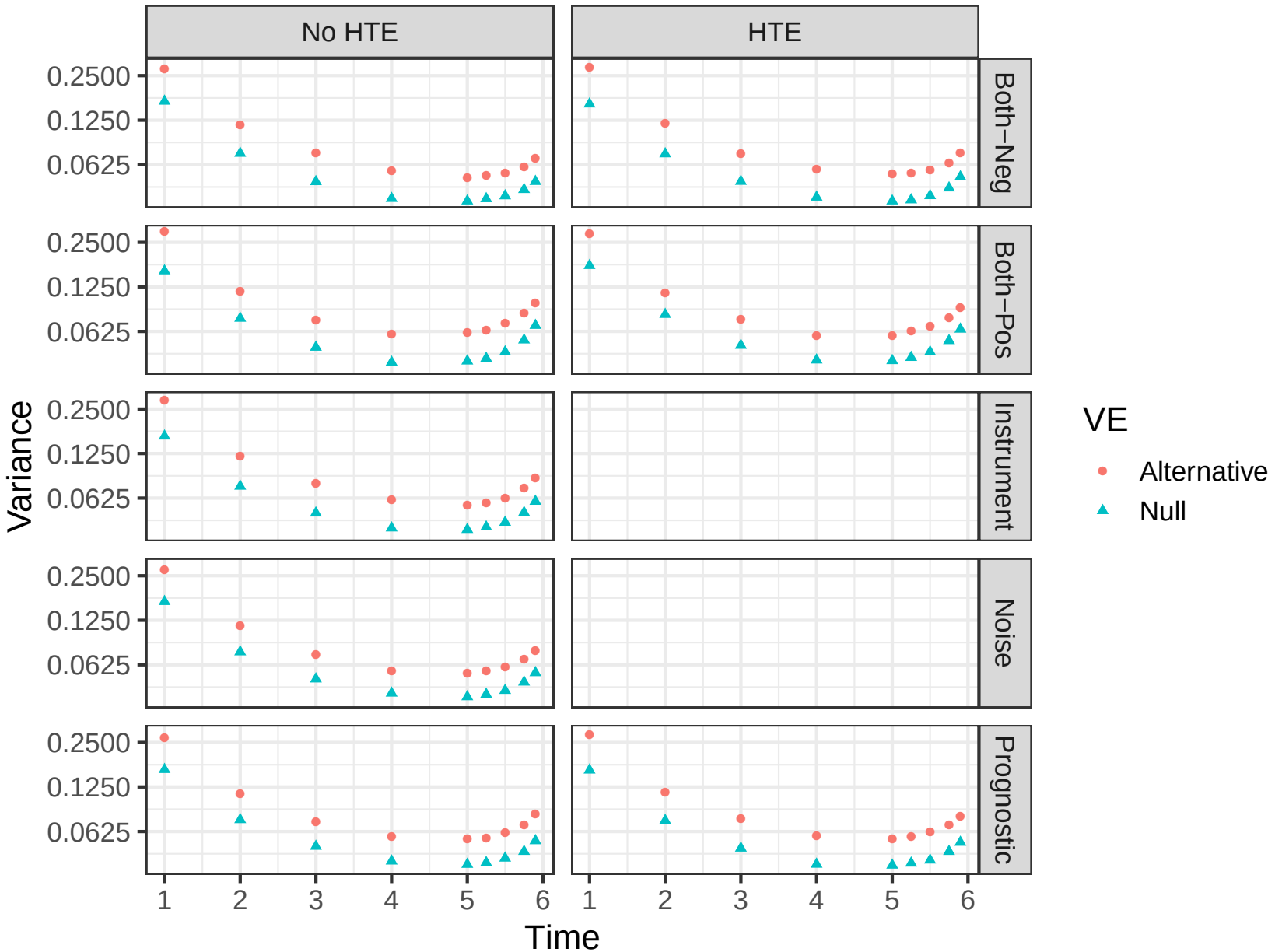


Case 3: covariates only predictive of censoring time

- **Take-away message:** When covariates are predictive only of censoring time, both the adjusted estimator and the unadjusted estimator will yield valid inferences, but the adjusted estimator will have higher variance. **Adjusting for covariates that are only predictive of censoring time should be avoided when possible.**

  But: the harm in this simulation was relatively small, amounting to an effective reduction in sample size of at most 6%.

How large can t be chosen when estimating VE(0-t), relative to maximal follow-up time?



What about an adjusted estimator for VE(0-t) that simply marginalizes a Cox model?

- We also considered the estimator that does the following to estimate each $S(t|a)$:
  1. Using data from arm a, estimate $S(t|a,x)$ with a Cox model, yielding $\widehat{S}(t|a,x)$.
  2. Estimate $S(t|a)$ as:

$$\widehat{S}(t|a) = \frac{1}{n}\sum_{i=1}^{n}\widehat{S}(t|a,X_i).$$

- This estimator performed nearly identically to the Zeng (2004) estimator across all settings except for Case 3, where it outperformed the Zeng (2004) estimator.
  - But this estimator is less robust than the Zeng (2004) estimator, as its validity relies on the correct specification of $\widehat{S}(t|a,x)$.
  - In the simulation that I've shown, $\widehat{S}(t|a,x)$ was correctly specified.

Conclusion

- The described covariate adjustment strategy is simple and well understood
- In any given trial, all that remains to specify are the covariates to be adjusted for
  - **Must be prespecified prior to unblinding!**
- Hopefully the simulations provide some insight into what considerations should go into this choice

References

S. Browne, T. Carter, R. Eckes, G. Grandits, M. Johnson, I. Moore, and L. McNay. A review of strategies used to retain participants in clinical research during an infectious disease outbreak: The PREVAIL I Ebola vaccine trial experience. *Contemporary clinical trials communications*, 11:50–54, 2018.

D. Zeng. Estimating marginal survival function by adjusting for dependent censoring using many covariates. *The Annals of Statistics*, 32(4):1533–1555, 2004.

How strong are the prognostic factors in the simulation?

- When W predicts the outcome T in arm a, the probability of infection by Month 6 is:
  - Top vs bottom 50% of participants: ≈7x larger
  - Top vs bottom 25% of participants: ≈20x larger
  - Top vs bottom 10% of participants: ≈50x larger
  - Top vs bottom 5% of participants: ≈80x larger
  - Top vs bottom 1% of participants: ≈300x larger