Mitigating Measurement Error in Real-World Indirect Treatment Comparisons with Time-To-Event Outcomes: An Approximate Bayesian Computation Approach

Arcturis

Steven Soutar[1], Joseph E, O'Reilly[1], Jamie Wallis[1], and Lewis Carpenter[1]
[1]: Arcturis Data, Building One, Oxford Technology Park, Technology Drive, Kidlington, OX5 1GN UK | steven.soutar@arcturisdata.com

Introduction

- Real-world Indirect treatment comparisons (ITCs) allow for the generation of relative-efficacy evidence in settings where implementation of a randomised controlled trial is infeasible [1,2].
- However, real-world ITCs with time-to-event outcomes, such as presence of disease progression, are susceptible to measurement error bias.
- Sources of measurement error arise through differences between arms in assessment timing – inducing assessment time bias (ATB) – or through error-prone assessment – inducing outcome misclassification bias (OMB) [3, 4].
- Current methods for addressing measurement error are often limited to exclusively one source of error and lack generalisability to simultaneously address other forms of bias, e.g. confounding.
- Comprehensive approaches to bias mitigation requires statistical methods capable of complex model specification without compromising inferential tractability.
- Approximate Bayesian computation (ABC) has been proposed for tractable bias-mitigated inference in the presence of complex biases [5, 6].

Methods

Approximate Bayesian Computation Framework

- Our proposed ABC framework assumes an ITC with gold-standard assessment in the treatment arm and error-prone assessment in the control arm.
- We assume error-prone control assessment manifests itself as a patient-specific delay in detection, $\tau_i \sim \mathrm{Exp}(\mu)$, affecting a proportion φ of patients (Figure 1).
- Our proposed framework implements an ABC-rejection sampler for bias-mitigation with a user-specified survival model f (Figure 2).

Figure 1: Error-prone control assessment.

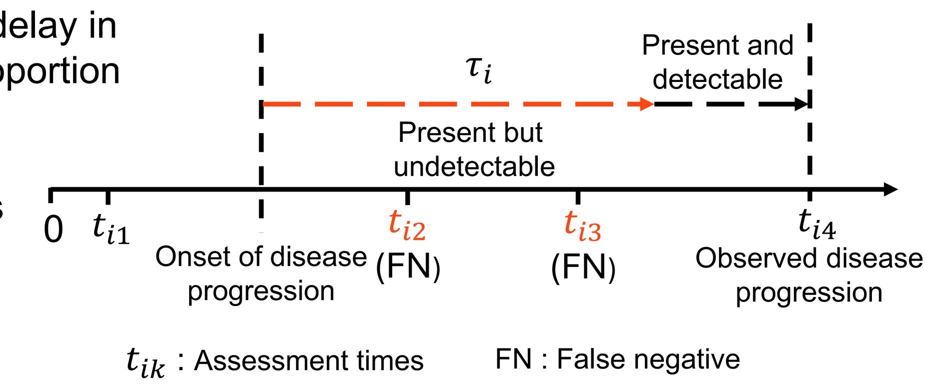


Figure 2: Posterior sampling of bias-mitigated effect estimates through ABC sampling.

ABC-Rejection Sampler

Step 1	Step 2	Step 3	Step 4
Propose treatment effect: $\boldsymbol{\theta}' \sim p_\theta$ Propose ME parameters: $(\mu', \varphi') \sim p_{\mu,\varphi}$	Simulate underlying survival: $\boldsymbol{y}' \mid \boldsymbol{\theta}' \sim f$ $\tau_i \sim \mathrm{Exp}(\mu')$ $D_i \sim \mathrm{Bernoulli}(\varphi')$ $(D_i = 1 \Rightarrow \text{delayed})$	Simulate measurement error: Mismeasured survival $\tilde{\boldsymbol{y}}'$	Compare summary statistics between observed $\boldsymbol{y}$ and simulated datasets $\tilde{\boldsymbol{y}}'$, and accept $(\boldsymbol{\theta}', \mu', \varphi')$ if: $d = \lvert S(\boldsymbol{y}), S(\tilde{\boldsymbol{y}}')\rvert \leq \epsilon$

- ATB is simulated in step 3 by rounding survival to the nearest assessment time. Comparison of datasets in step 4 is performed using a 4-dimensional summary statistic (see below).
- Iteration of steps 1-4 outputs a posterior sample for $\boldsymbol{\theta}$ which is uncontaminated by measurement error and hence represents a bias-mitigated effect estimate.

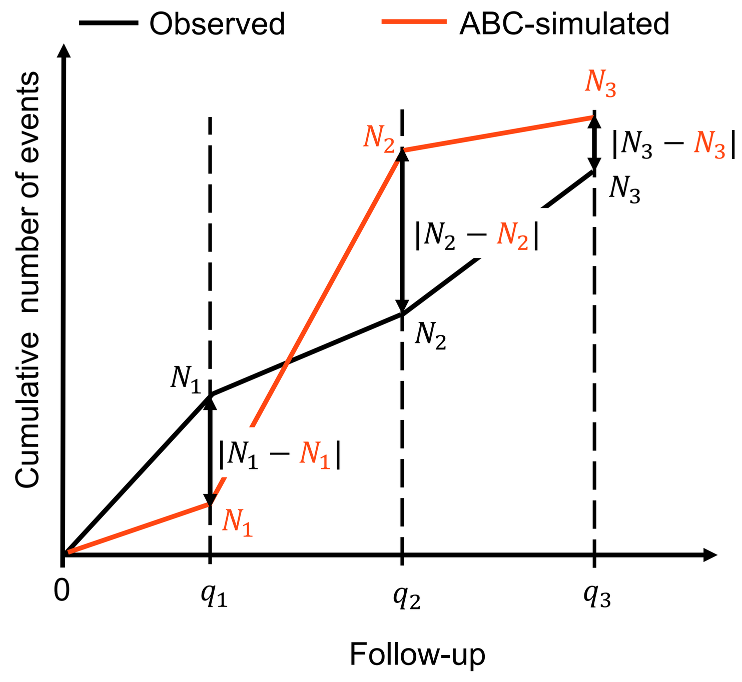
Summary Statistics

- ABC-simulated treatment and control arms are compared with their observed counterpart.
- Cumulative number of events are computed up to user-specified quantiles (q_1, q_2, q_3), e.g. 25th, 50th, 75th (Figure 3).
- When combined with the number of censored patients, N_c, vectors (N_1, N_2, N_3, N_c) are compared between datasets using the L1 norm:

$$d = |N_1 - \tilde{N}_1| + |N_2 - \tilde{N}_2| + |N_3 - \tilde{N}_3| + |N_c - \tilde{N}_c|.$$

- Inverse probability of treatment (IPT) - weights can be applied to summary statistics to adjust for confounding.

Figure 3: Comparison between datasets.



Simulation Study Assessment

- Two-arm confounded survival data was simulated under Eqs. (1) and (2) with measurement error, where the effect measure is the hazard ratio (HR).
- 12 scenarios were considered through varying $\beta_1 = \log(\mathrm{HR})$, average length of delay $(1/\mu)$, and control assessment cadence (3 or 6 months).
- Treatment assessment cadence, φ, and $\exp(\beta_0)$, were fixed to 1 month, 0.3, and 0.11, respectively.
- 200 data sets were simulated under each scenario and ABC estimates compared with naive Markov chain Monte Carlo (MCMC) estimates which ignored measurement error and confounding.
- IPT-weighting of observed summary statistics was implemented to adjust for confounding by u_i.
- 2,000 posterior samples were extracted when applying either method.

Survival Model

$$y_i \sim \mathrm{Exp}(\lambda_i), \quad i = 1, \dots, 400, \qquad (1)$$

$$\lambda_i = \exp(\beta_0 + \beta_1 z_i + 0.26 u_i),$$

$$u_i \sim \mathrm{Bernoulli}(0.5), \quad z_i = 0, 1,$$

$$\Pr(z_i = 1 \mid u_i) = \mathrm{logit}^{-1}(-1 + 2u_i). \qquad (2)$$

Prior Specification

$$\beta_1 \sim N(0, 1),$$

$$\exp(\beta_0) \sim U(0.058, 0.693),$$

$$\mu \sim U(0.1, 1),$$

$$\varphi \sim U(0.2, 0.5).$$

Objectives

- Develop an approximate Bayesian computation framework for measurement error mitigation in real-world indirect treatment comparisons with time-to-event outcomes.
- Assess the proposed ABC framework's ability for bias-mitigated estimation of relative treatment effects.
- Implement a simulation study which performs this assessment under varying degrees of measurement error.

Results

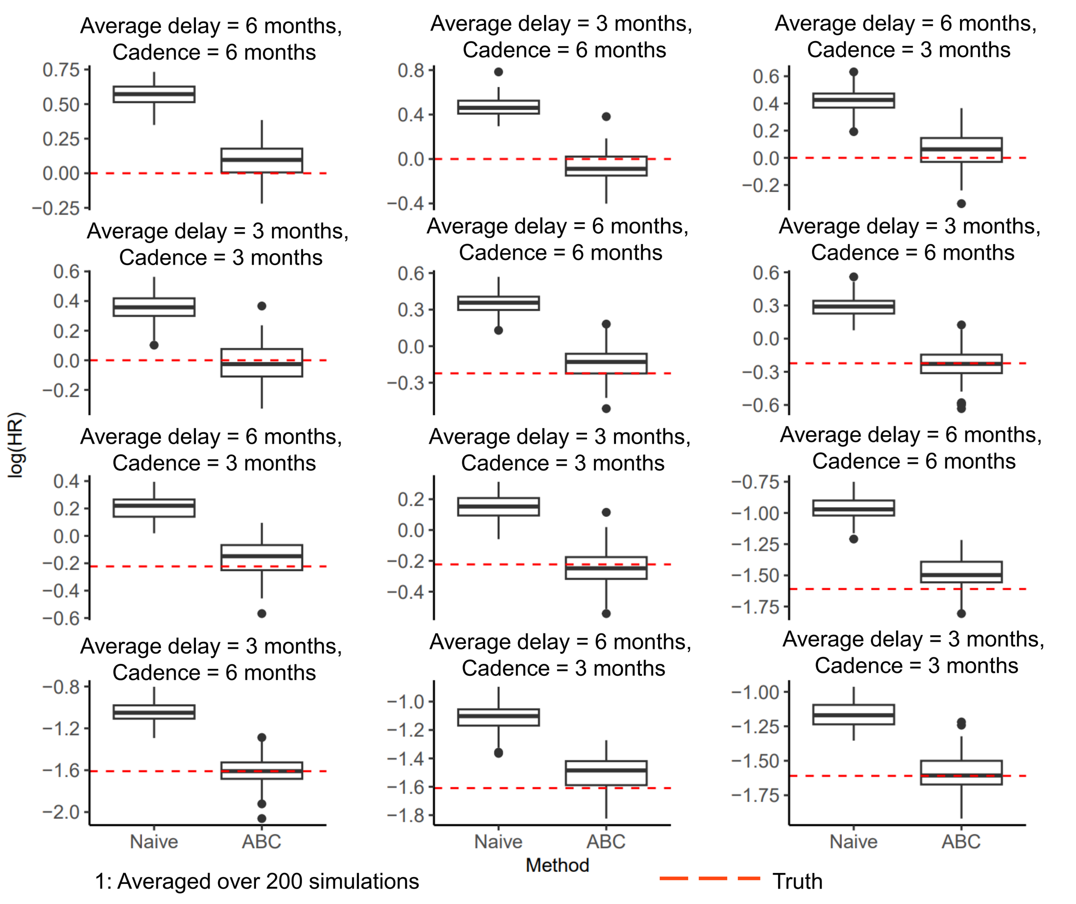
Table 1: Comparison of log(HR) posterior estimates between the ABC framework and naive MCMC.

Simulation scenario[1]		Bias[2,3]		Posterior SD[3,4]		95% credible interval coverage rate	
log(HR)	$1/\mu$[5]	ABC	Naive	ABC	Naive	ABC	Naive
0	3	-0.028	0.354	0.170	0.101	0.98	0.04
	6	0.058	0.429	0.163	0.100	0.98	0.01
log(0.8)	3	-0.019	0.373	0.164	0.100	0.98	0.02
	6	0.062	0.433	0.157	0.101	0.97	0.00
log(0.2)	3	0.016	0.411	0.157	0.113	0.95	0.00
	6	0.106	0.499	0.152	0.114	0.97	0.00

1: Control cadence = 3 months, 2: Bias defined as estimate – truth, 3: Averaged over 200 simulations, 4: Standard deviation, 5: Average delay in detection.

- With $\epsilon = 45$, ABC acceptance rates ranged from 0.16% – 0.35%.
- In Table 1, ABC bias ranged from -0.028 – 0.106 compared to 0.354 – 0.499 under naive MCMC.
- As the severity of measurement error increased, ABC performance decreased (Figure 4, top left) but was still superior compared to naive MCMC.

Figure 4: Comparison of posterior mean estimates[1] of the log(HR) between the ABC framework and naive MCMC.



Conclusions and future work

- Our proposed ABC framework provides a unified approach for successful bias-mitigated estimation of relative treatment effects under both measurement error and confounding
- In practice, careful calibration would be required to ensure successful application of the ABC framework, with external information regarding measurement error dynamics required to inform the magnitude of delay and its prevalence.
- These dynamics can be characterised either through assessment of a subset of data for which both gold-standard and error-prone assessment are observed, or through expert clinical elicitation.
- Future analyses will apply the ABC framework to empirical data and consider application of the ABC approach for bias-mitigation to other forms of measurement error and bias, e.g. selection bias arising from left truncation.

References

1. Hashmi M, Rassen J, Schneeweiss S. Single-arm oncology trials and the nature of external controls arms. Journal of Comparative Effectiveness Research. 2021;10:1053–66.
2. Sola-Morales O, Curtis LH, Heidt J, Walsh L, Casso D, Oliveria S, et al. Effectively Leveraging RWD for External Controls: A Systematic Literature Review of Regulatory and HTA Decisions. Clinical Pharmacology & Therapeutics. 2023;114:325–55.
3. Adamson BJS, Ma X, Griffith SD, Sweeney EM, Sarkar S, Bourla AB. Differential frequency in imaging-based outcome measurement: Bias in real-world oncology comparative-effectiveness studies. Pharmacoepidemiology and Drug Safety. 2022;31:46–54.
4. Ackerman B, Gan RW, Meyer CS, Wang JR, Zhang Y, Hayden J, et al. Measurement error and bias in real-world oncology endpoints when constructing external control arms. Front Drug Saf Regul. 2024;4.
5. Götte H, Kirchner M, Kieser M. Adjustment for exploratory cut-off selection in randomized clinical trials with survival endpoint. Biometrical Journal. 2020;62:627–42.
6. Götte H, Kirchner M, Sailer MO, Kieser M. Simulation-based adjustment after exploratory biomarker subgroup selection in phase II. Statistics in Medicine. 2017;36:2378–90.

Introduction

- Real-world Indirect treatment comparisons (ITCs) allow for the generation of relative-efficacy evidence in settings where implementation of a randomised controlled trial is infeasible [1,2].
- However, real-world ITCs with time-to-event outcomes, such as presence of disease progression, are susceptible to measurement error bias.
- Sources of measurement error arise through differences between arms in assessment timing – inducing assessment time bias (ATB) – or through error-prone assessment – inducing outcome misclassification bias (OMB) [3, 4].
- Current methods for addressing measurement error are often limited to exclusively one source of error and lack generalisability to simultaneously address other forms of bias, e.g. confounding.
- Comprehensive approaches to bias mitigation requires statistical methods capable of complex model specification without compromising inferential tractability.
- Approximate Bayesian computation (ABC) has been proposed for tractable bias-mitigated inference in the presence of complex biases [5, 6].

Objectives

- Develop an approximate Bayesian computation framework for measurement error mitigation in real-world indirect treatment comparisons with time-to-event outcomes.
- Assess the proposed ABC framework's ability for bias-mitigated estimation of relative treatment effects.
- Implement a simulation study which performs this assessment under varying degrees of measurement error.

Next

Methods

Approximate Bayesian Computation Framework

- Our proposed ABC framework assumes an ITC with gold-standard assessment in the treatment arm and error-prone assessment in the control arm.
- We assume error-prone control assessment manifests itself as a patient-specific delay in detection, $\tau_i \sim \text{Exp}(\mu)$, affecting a proportion φ of patients (Figure 1).
- Our proposed framework implements an ABC-rejection sampler for bias-mitigation with a user-specified survival model f (Figure 2).

Figure 1: Error-prone control assessment.

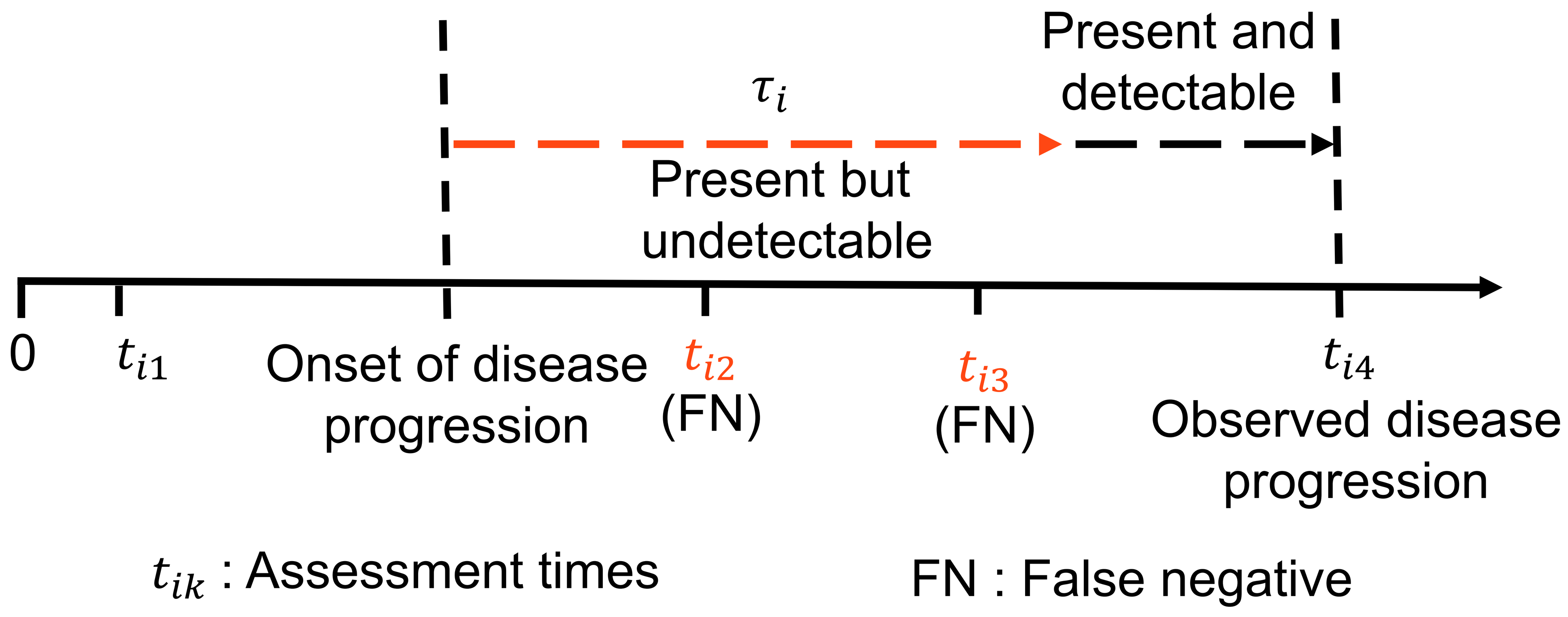


Figure 2: Posterior sampling of bias-mitigated effect estimates through ABC sampling.

ABC-Rejection Sampler

Step 1

Propose treatment effect:

$\boldsymbol{\theta}' \sim p_\theta$

Propose ME parameters:

$(\mu', \varphi') \sim p_{\mu,\varphi}$

Step 2

Simulate underlying survival:

$\boldsymbol{y}' \mid \boldsymbol{\theta}' \sim f$

$\tau_i \sim \text{Exp}(\mu')$

$D_i \sim \text{Bernoulli}(\varphi')$

$(D_i = 1 \Rightarrow$ delayed)

Step 3

Simulate measurement error:

Mismeasured survival $\tilde{\boldsymbol{y}}'$

Step 4

Compare summary statistics between observed $\boldsymbol{y}$ and simulated datasets $\tilde{\boldsymbol{y}}'$, and accept $(\boldsymbol{\theta}', \mu', \varphi')$ if:

$d = |S(\boldsymbol{y}), S(\tilde{\boldsymbol{y}}')| \leq \epsilon$

- ATB is simulated in step 3 by rounding survival to the nearest assessment time. Comparison of datasets in step 4 is performed using a 4-dimensional summary statistic (see next page).
- Iteration of steps 1-4 outputs a posterior sample for $\boldsymbol{\theta}$ which is uncontaminated by measurement error and hence represents a bias-mitigated effect estimate.

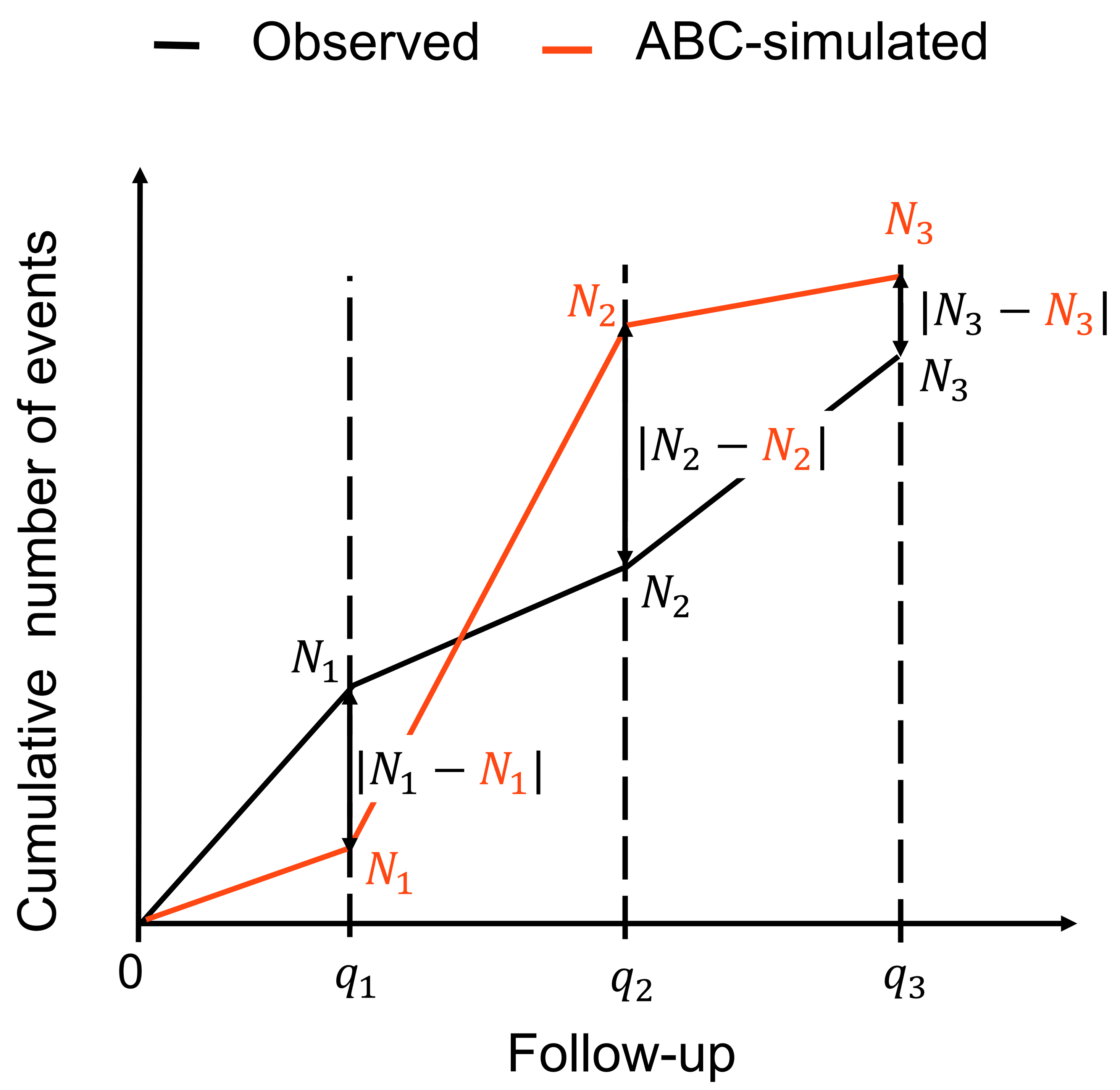
Methods

Summary Statistics

- ABC-simulated treatment and control arms are compared with their observed counterpart.
- Cumulative number of events are computed up to user-specified quantiles (q_1, q_2, q_3), e.g. 25th, 50th, 75th (Figure 3).
- When combined with the number of censored patients, N_c, vectors(N_1, N_2, N_3, N_c) are compared between datasets using the L1 norm:

$$d = |N_1 - N_1| + |N_2 - N_2| + |N_3 - N_3| + |N_c - N_c|.$$

- Inverse probability of treatment (IPT) - weights can be applied to summary statistics to adjust for confounding.



Simulation Study Assessment

- Two-arm confounded survival data was simulated under Eqs. (1) and (2) with measurement error, where the effect measure is the hazard ratio (HR).
- 12 scenarios were considered through varying $\beta_1 = \log(\text{HR})$, average length of delay $(1/\mu)$, and control assessment cadence (3 or 6 months).
- Treatment assessment cadence, φ, and $\exp(\beta_0)$, were fixed to 1 month, 0.3, and 0.11, respectively.
- 200 data sets were simulated under each scenario and ABC estimates compared with naive Markov chain Monte Carlo (MCMC) estimates which ignored measurement error and confounding.
- IPT-weighting of observed summary statistics was implemented to adjust for confounding by u_i.
- 2,000 posterior samples were extracted when applying either method.

Survival Model

$$y_i \sim \text{Exp}(\lambda_i), \quad i = 1, \ldots, 400, \quad (1)$$

$$\lambda_i = \exp(\beta_0 + \beta_1 z_i + 0.26 u_i),$$

$$u_i \sim \text{Bernoulli}(0.5), \quad z_i = 0, 1,$$

$$\Pr(z_i = 1 \mid u_i) = \text{logit}^{-1}(-1 + 2u_i). \quad (2)$$

Prior Specification

$$\beta_1 \sim N(0, 1),$$

$$\exp(\beta_0) \sim U(0.058, 0.693),$$

$$\mu \sim U(0.1, 1),$$

$$\varphi \sim U(0.2, 0.5).$$

Results

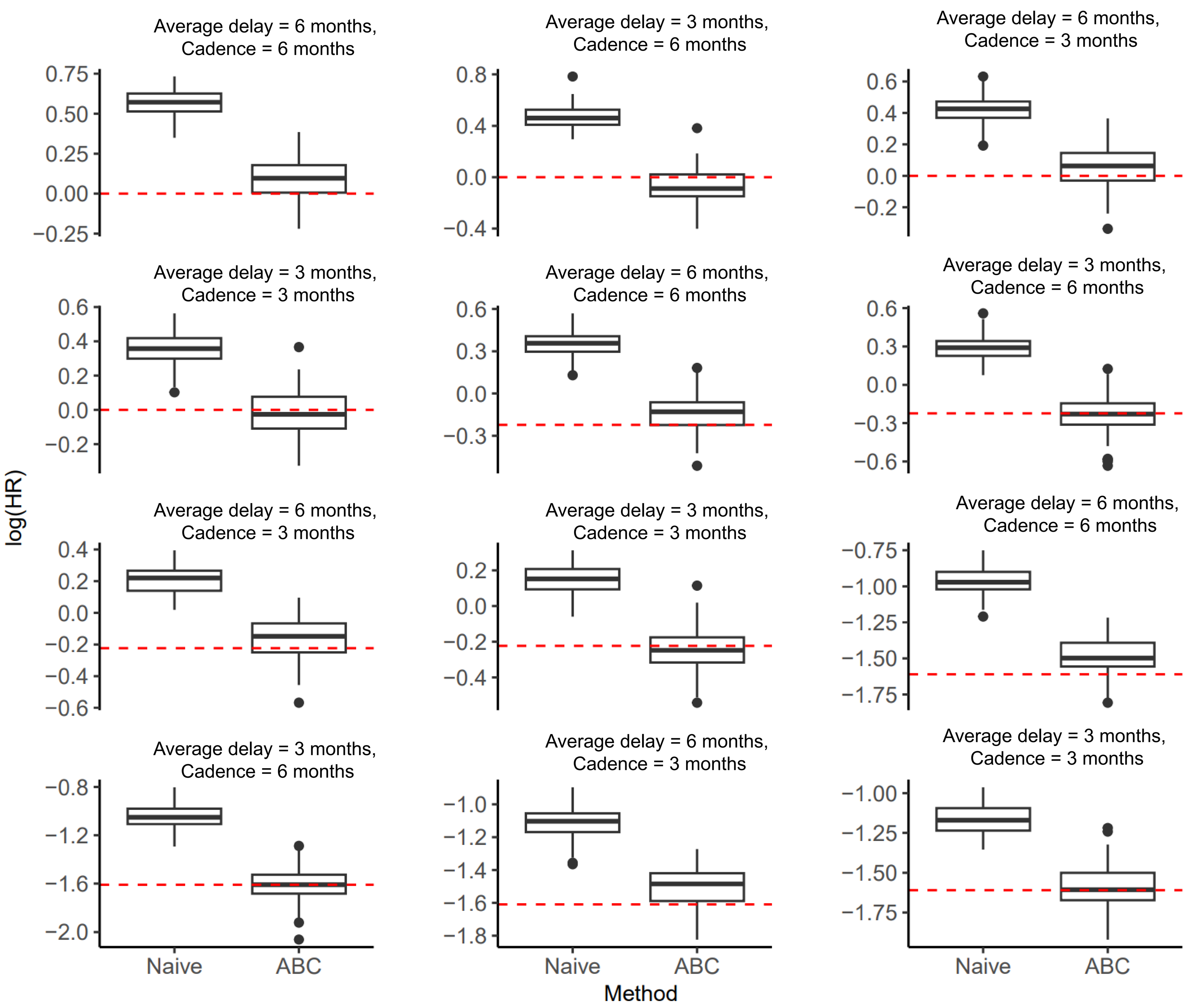
Table 1: Comparison of log(HR) posterior estimates between the ABC framework and naive MCMC.

Simulation scenario[1]		Bias[2,3]		Posterior SD[3,4]		95% credible interval coverage rate	
$\log(\mathrm{HR})$	$1/\mu$[5]	ABC	Naive	ABC	Naive	ABC	Naive
0	3	-0.028	0.354	0.170	0.101	0.98	0.04
	6	0.058	0.429	0.163	0.100	0.98	0.01
$\log(0.8)$	3	-0.019	0.373	0.164	0.100	0.98	0.02
	6	0.062	0.433	0.157	0.101	0.97	0.00
$\log(0.2)$	3	0.016	0.411	0.157	0.113	0.95	0.00
	6	0.106	0.499	0.152	0.114	0.97	0.00

1: Control cadence = 3 months, 2: Bias defined as estimate – truth, 3: Averaged over 200 simulations, 4: Standard deviation, 5: Average delay in detection.

- With $\epsilon = 45$, ABC acceptance rates ranged from 0.16% – 0.35%.
- In Table 1, ABC bias ranged from -0.028 – 0.106 compared to 0.354 – 0.499 under naive MCMC.
- As the severity of measurement error increased, ABC performance decreased (Figure 4, top left) but was still superior compared to naive MCMC.

Figure 4: Comparison of posterior mean estimates[1] of the log(HR) between the ABC framework and naive MCMC.



1: Averaged over 200 simulations

Truth

Conclusions and future work

- Our proposed ABC framework provides a unified approach for successful bias-mitigated estimation of relative treatment effects under both measurement error and confounding
- In practice, careful calibration would be required to ensure successful application of the ABC framework, with external information regarding measurement error dynamics required to inform the magnitude of delay and its prevalence.
- These dynamics can be characterised either through assessment of a subset of data for which both gold-standard and error-prone assessment are observed, or through expert clinical elicitation.
- Future analyses will apply the ABC framework to empirical data and consider application of the ABC approach for bias-mitigation to other forms of measurement error and bias, e.g. selection bias arising from left truncation.

References

1. Hashmi M, Rassen J, Schneeweiss S. Single-arm oncology trials and the nature of external controls arms. Journal of Comparative Effectiveness Research. 2021;10:1053–66.
2. Sola-Morales O, Curtis LH, Heidt J, Walsh L, Casso D, Oliveria S, et al. Effectively Leveraging RWD for External Controls: A Systematic Literature Review of Regulatory and HTA Decisions. Clinical Pharmacology & Therapeutics. 2023;114:325–55.
3. Adamson BJS, Ma X, Griffith SD, Sweeney EM, Sarkar S, Bourla AB. Differential frequency in imaging-based outcome measurement: Bias in real-world oncology comparative-effectiveness studies. Pharmacoepidemiology and Drug Safety. 2022;31:46–54.
4. Ackerman B, Gan RW, Meyer CS, Wang JR, Zhang Y, Hayden J, et al. Measurement error and bias in real-world oncology endpoints when constructing external control arms. Front Drug Saf Regul. 2024;4.
5. Götte H, Kirchner M, Kieser M. Adjustment for exploratory cut-off selection in randomized clinical trials with survival endpoint. Biometrical Journal. 2020;62:627–42.
6. Götte H, Kirchner M, Sailer MO, Kieser M. Simulation-based adjustment after exploratory biomarker subgroup selection in phase II. Statistics in Medicine. 2017;36:2378–90.

Next