Impact of interval-censored data on comparative time-to-event endpoints:

a simulation study applied to patient-reported outcomes in oncology

Joel Sims, Emma Martin, Mike Greenwood, Rachael Lawrance
Adelphi Values Ltd, Bollington, United Kingdom

Background

> Patient-reported outcome (PRO) time-to-event endpoints feature as key endpoints in **SISAQOL-IMI recommendations**

> In oncology clinical trials, a research question may be to evaluate whether patients in the treatment arm report **delayed PRO worsening**: time to deterioration (**TTD**)

> PRO assessments are commonly assessed at scheduled visits, so true event times are only known to lie between visits: **interval-censored data**.

> This interval censoring is not accounted for when using **Cox proportional hazards (PH)** or **Kaplan-Meier (KM) methods**, and its impact on treatment effect estimation is not routinely quantified

TTD Literature Review

Adelphi
ADELPHI VALUES
Patient-Centered Outcomes

Initial on-treatment assessment schedule	Number of trials (% of N = 70)
Weekly	4 (5.7%)
Every 2 weeks	3 (4.3%)
Every 3 weeks	18 (25.7%)
Every 4 weeks	17 (24.3%)
Every 6 weeks	10 (14.3%)
Every 8 weeks	10 (14.3%)
Every 12 weeks	6 (8.6%)
Every 16 weeks	2 (2.9%)

TTD analysis methods	
Cox PH model	60 (85.7%)
Log-rank test	45 (64.3%)
Both Cox PH and log-rank	39 (55.7%)
Competing risk model	1 (1.4%)
Not reported	3 (4.3%)

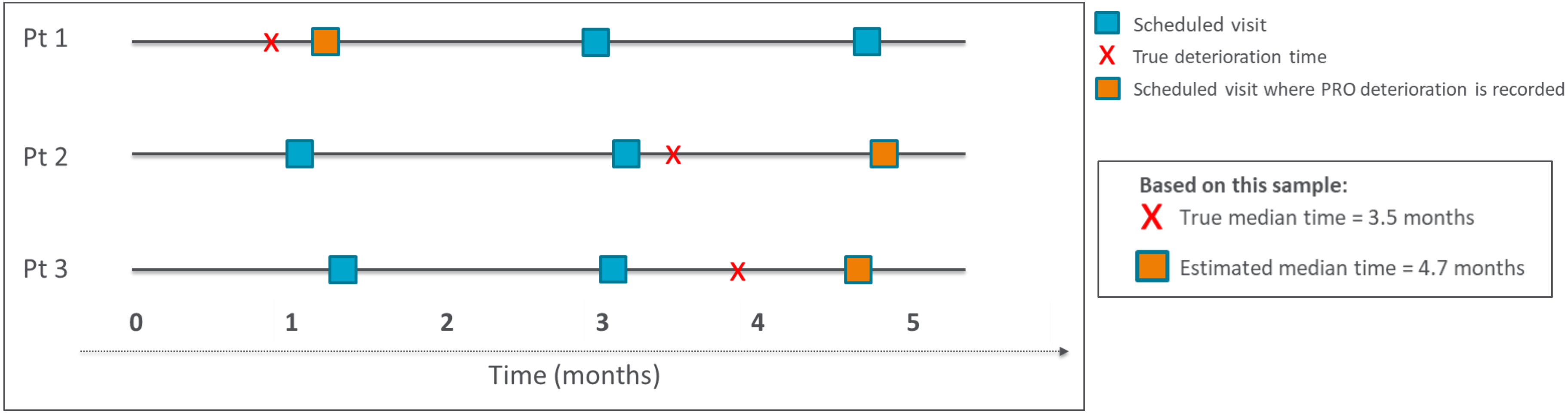
Cocks, K., King-Kallimanis, B., Sims, J. et al. Time to deterioration of patient-reported outcome endpoints in cancer clinical trials: targeted literature review and best practice recommendations. J Patient Rep Outcomes 8, 150 (2024). https://doi.org/10.1186/s41687-024-00824-7

Why is interval censoring a problem?

ADELPHI VALUES

Patient-Centered Outcomes

> **Fixed-visit imputation** – assigning the event time to the visit at which it is first detected



> Other methods have been suggested:

- **Mid-point imputation** – event time taken to be the mid-point of the interval
- Other more sophisticated methods such as **Turnbull's estimator** (survival estimation) and **Finkelstein's Generalization of Cox Regression** (for effect estimation)

Objectives

We simulated various realistic scenarios to investigate the following:

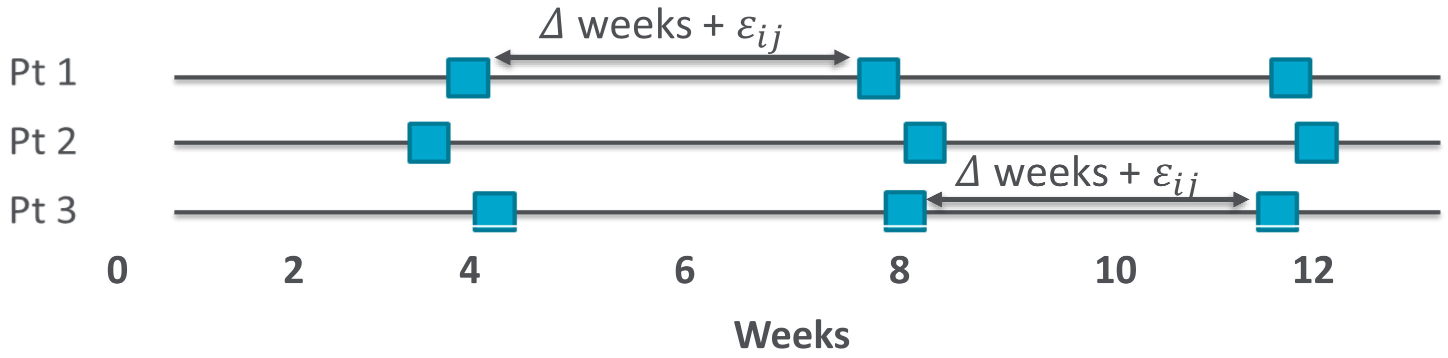
1. How does interval-censored data impact on **bias** for PRO TTD endpoints? Assuming **assessment frequency same** between arms.
2. What is the impact of **reduced/altered** assessment frequency following occurrence of intercurrent events (ICE)?

We focus on the **bias** of **Cox proportional hazards regression model** for estimating the hazard ratio (HR).

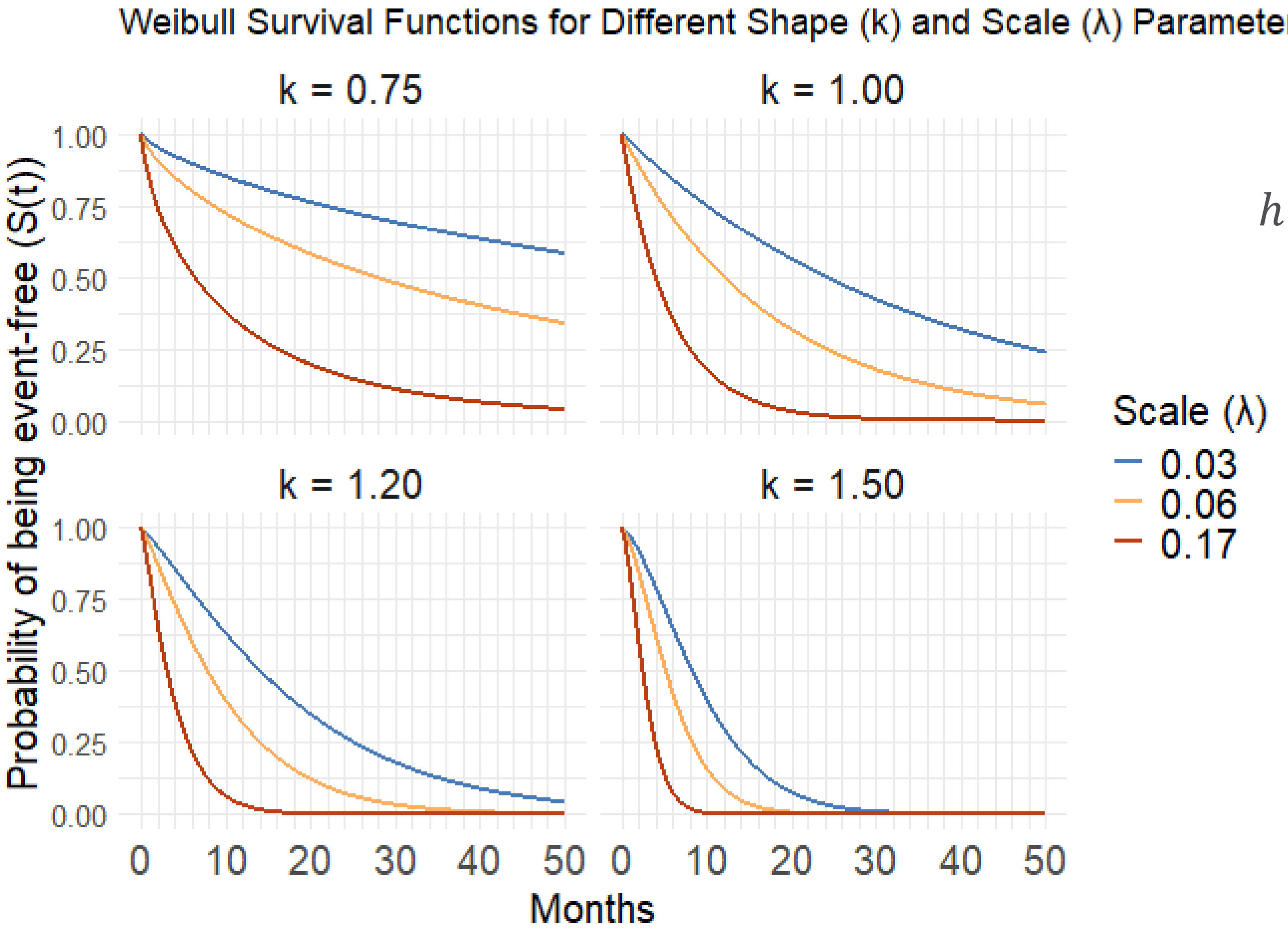
Simulation methods: study design

ADELPHI VALUES
Patient-Centered Outcomes

- 1:1 randomization ratio for two treatment groups, $Trx = \{0,1\}$
  - Sample size (n) = 400
- Visits specified according to an assessment schedule with **fixed interval length** Δ varied: $\boldsymbol{\Delta = \{4, 8, 12, 16\}}$ **weeks**
  - Small **scheduling variability** for patient i and interval j: $\boldsymbol{\Delta + \varepsilon_{ij}}$,
  - Assume interval length **non-informative** and all assessments completed



Simulation methods: true event times



- Assume **Cox PH** model

$$h(t) = h_0(t)\exp(\boldsymbol{\beta_{Trx}} \times \boldsymbol{Trx})$$

- Simulate true event times based on **Weibull** distribution
- Multiple **baseline hazard shapes** assumed to mimic real PRO deterioration event time distributions

$$h_0(t) = \boldsymbol{\lambda k} t^{\boldsymbol{k-1}}$$

Scenario 1 – on-treatment assessments at same frequency between arms

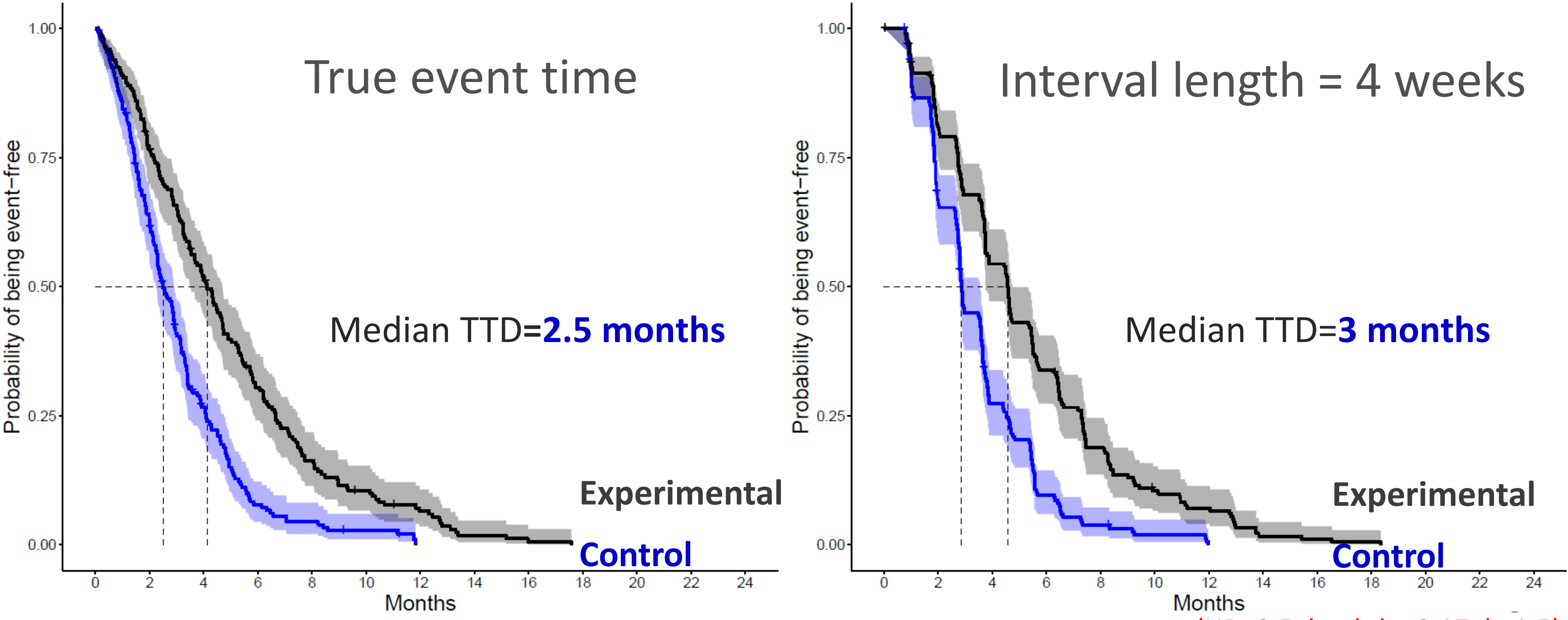
Simulation methods

- **Fixed-visit imputation** – assign event time to the visit at which it is first detected
- Assume **no informative censoring**
  - Only fixed data cut off: $C_{1_i} \sim U(0{,}100\ months)$
- Vary the true **HR = {0.5, 0.75, 1}** favouring experimental arm
- HR<1 indicates delayed deterioration in experimental arm

Scenario 1 – on-treatment assessments at same frequency between arms

Adelphi
ADELPHI VALUES
Patient-Centered Outcomes

Simulation results



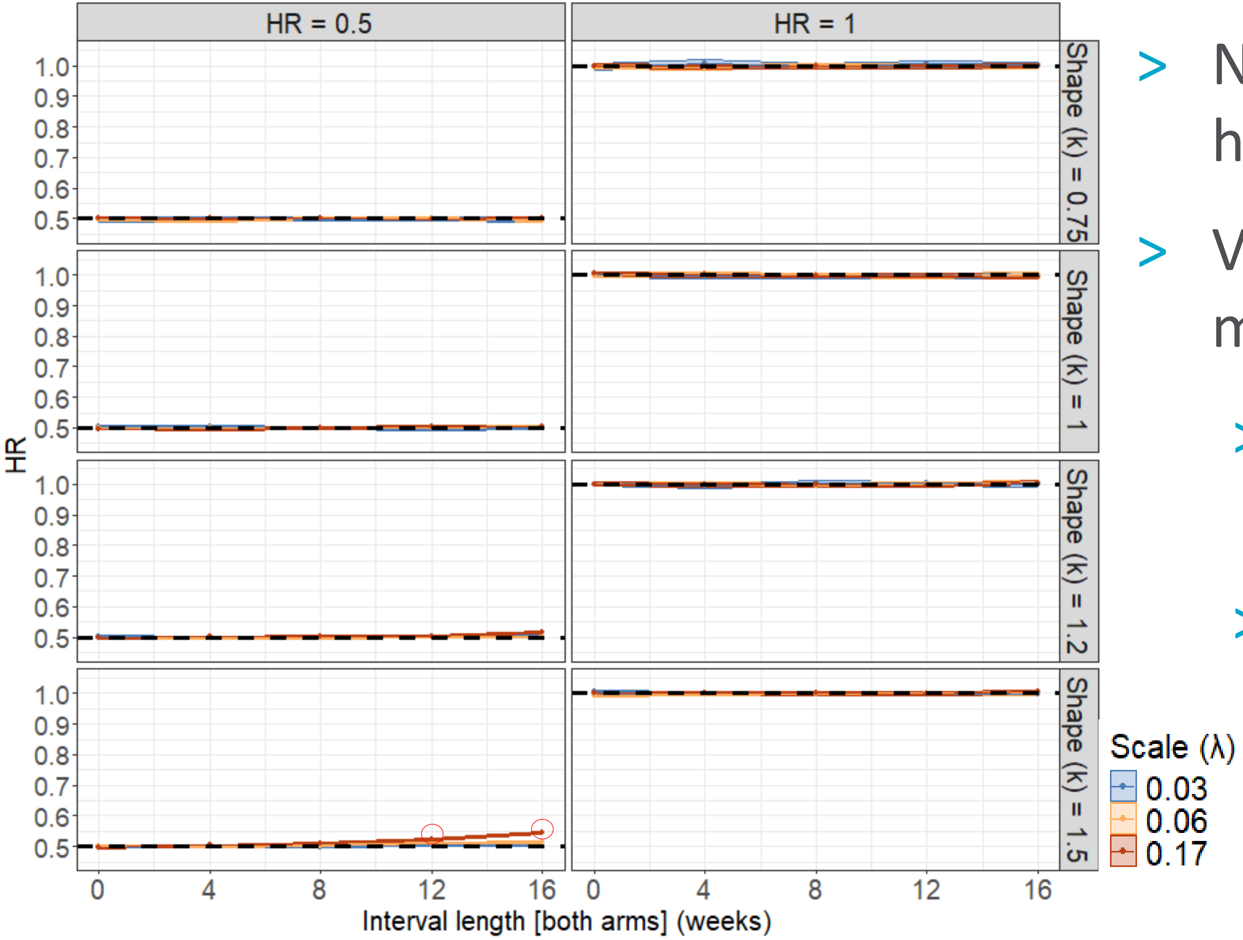
(HR=0.5, lambda=0.17, k=1.5)

Scenario 1 – on-treatment assessments at same frequency between arms

Simulation results – Bias (HR)

- > Log(HR), β'_{Trx}, averaged over multiple simulations (n_{sim} =**1000**)
  - > Shaded areas represent Monte-Carlo 95% CI ('simulation error')
- > Dotted line **- - -**, marks **true HR** (HR=1)
  - > Deviation from this line indicates **bias** of the Cox PH model

Scenario 1 – on-treatment assessments at same frequency between arms



- No bias across almost all hazard distributions
- Very minimal bias in the most extreme scenario
  - Bias is in direction of the **null**
  - Bias increases as true treatment difference increases

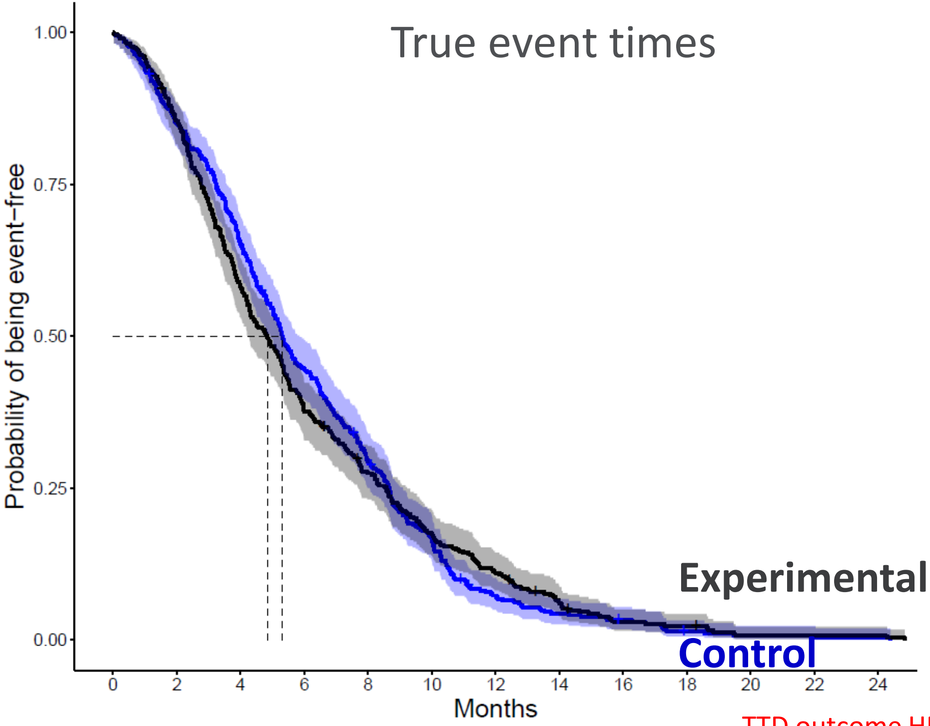
Scenario 2 – assessment frequency reduced to every 12 weeks **post-ICE**

Methods

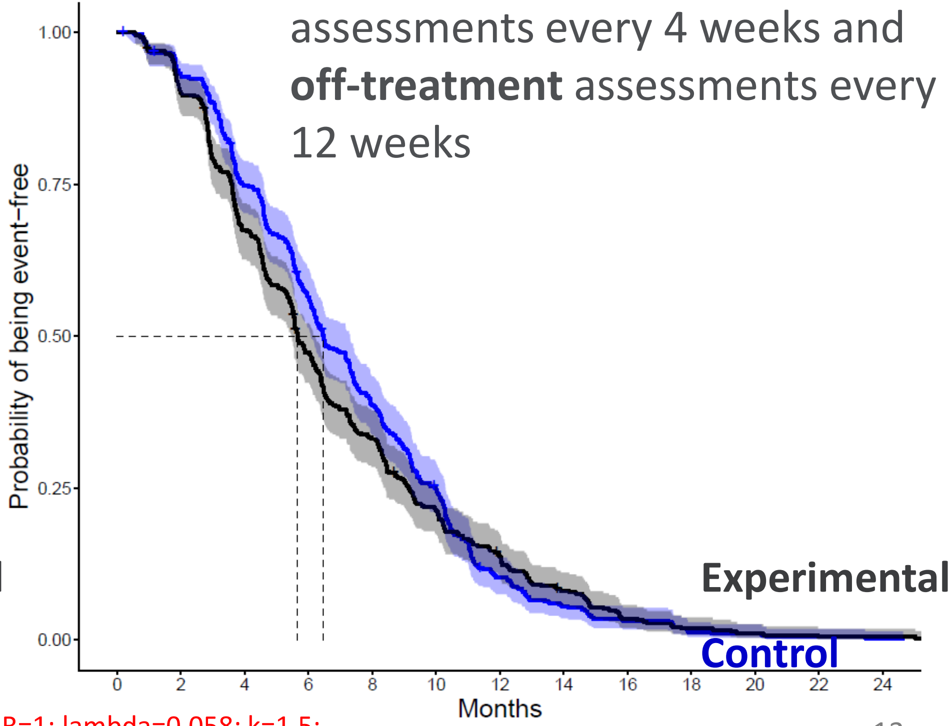
- Assume **no informative censoring**
- Introduce **intercurrent event (ICE)** → occurrence of which triggers change in PRO assessment schedule i.e. 'off-treatment schedule'
  - Assume 'off treatment' collection fixed to be every 12 weeks **indefinitely**
- Simulate ICE times from **exponential distribution**
- Vary rate of ICE (lambda) and **ICE HR = {0.5, 0.75, 1}**
- Repeat simulations with increasing correlation between PRO TTD and ICE times (dependence structure introduced using **Gaussian copula**)

Scenario 2 – assessment frequency reduced to every 12 weeks **post-ICE**

Simulation results



Symmetric **on-treatment** assessments every 4 weeks and **off-treatment** assessments every 12 weeks

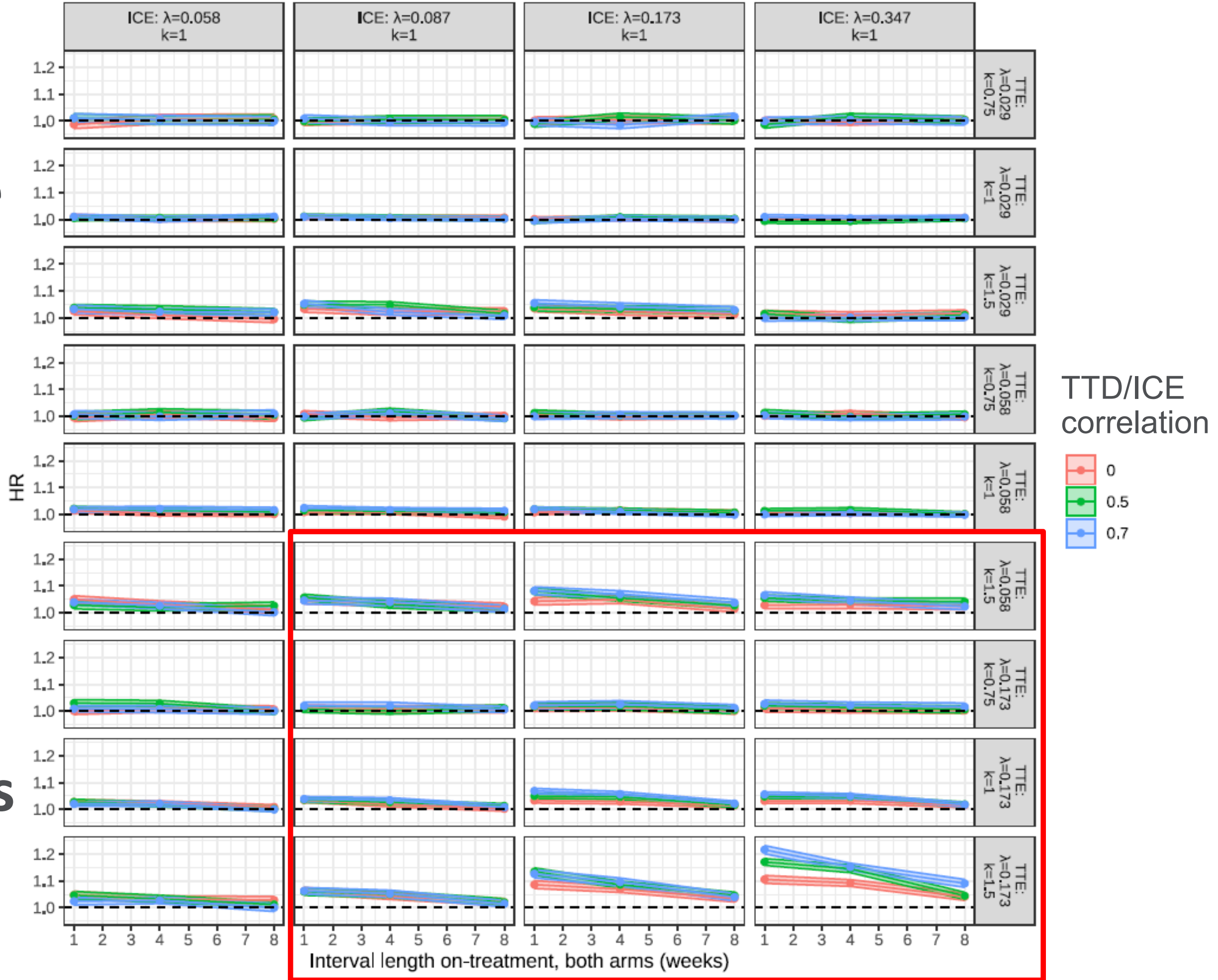


TTD outcome HR=1; lambda=0.058; k=1.5;
ICE HR=0.5; lambda=0.173

Scenario 2 – assessment frequency reduced to every 12 weeks **post-ICE**

For simplicity:

- > **Assume no difference in PRO TTD (HR=1):** black dotted line - - -
- > **Assume HR for ICE=0.5** (ICEs more frequent in control)
- > This results is a **non-negligible level of bias** in favour of control



Conclusions

- Scenario 1 shows us that Cox PH model generally robust when assessment schedule symmetric between arms, even if interval length large (16 weeks)
  - **Kaplan-Meier estimates** will still be biased even if Cox HR is not!
- Scenario 2 demonstrates asymmetry can be introduced through key design features related to post-ICE collection
  - This introduces bias favouring the arm with highest rate of transition to lower frequency assessments

Key takeaways

Adelphi
ADELPHI VALUES
Patient-Centered Outcomes

- Avoid introducing any asymmetry between arms as much as possible
- Regular PRO assessment frequency continuing **post-ICE** is the least biased approach
- Other **informative/dependent censoring** mechanisms are likely more important to address than interval censoring
  - Focus should remain on robust strategies for handling right-censored data not observed after ICEs

Adelphi
ADELPHI VALUES
Patient-Centered Outcomes

Thank you for listening

Any questions?