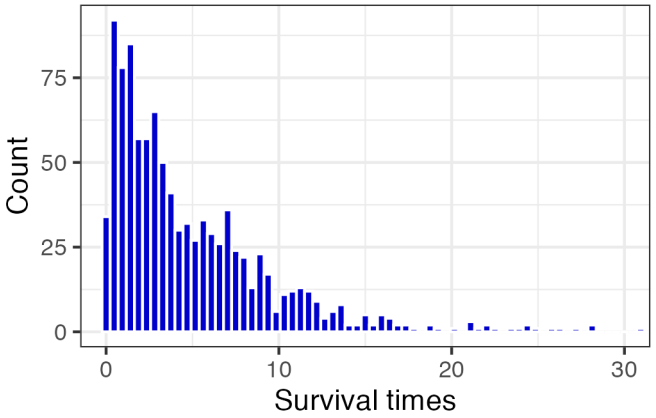
Statistical inference with left-truncated and right-censored survival data

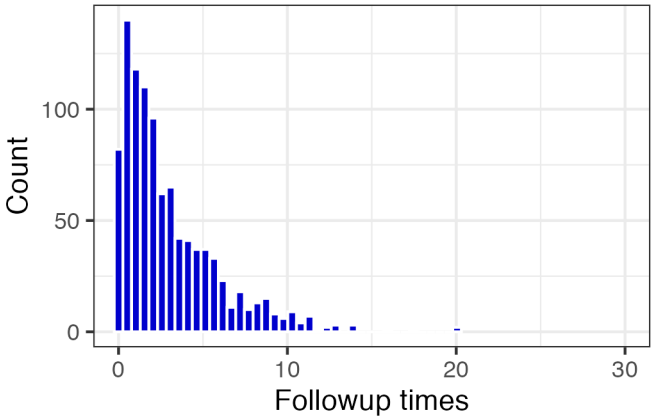
Devin Incerti | Roche & Genentech
August 25, 2021 | 37th ICPE

What is left-truncated and right-censored survival data?

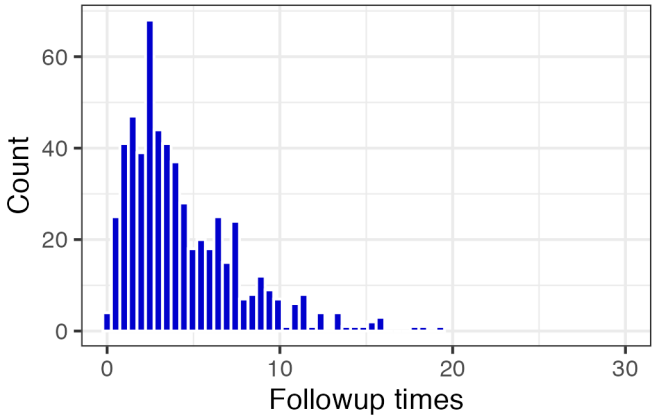
No censoring, no truncation
(N = 1000)



Right censoring, no truncation
(N = 1000)

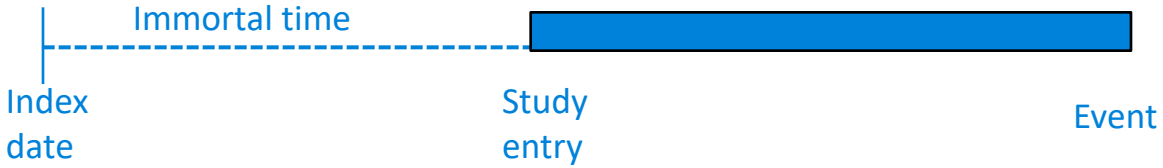


Right censoring, left truncation
(N = 602)



Immortal time bias

Left truncation can cause immortal time bias since the cohort only consists of patients that survive until the entry date



- Left truncation is distinct from other causes of immortal time bias in which the entire cohort is observed such as the measurement of a covariate after index date
- Since some patients are not in the sample, it precludes use of certain methods
  - Inverse probability weighting
  - Time varying covariates
  - Landmark analysis

Modifying the risk set

- When the data is left-truncated survival analyses can be performed by modifying the risk set so that patients are only “at risk” following their study entry date (rather than their index date)
- This results in start/stop data and can be analyzed[1] with a “counting process” formulation

Start/stop data

```
   entry_time followup_time event
1: 0.03605339     1.6245143     1
2: 0.16755345     0.3870667     1
3: 3.15631714     4.6869822     0
4: 0.09188061     0.2115054     1
5: 0.92435944     1.9391820     0
```

Kaplan-Meier estimator

```
survfit(Surv(entry_time, followup_time, event) ~ 1)
```

Cox model

```
coxph(Surv(entry_time, followup_time, event) ~ treat)
```

[1]Example analysis using the `survival` package from R

The quasi-independence assumption

- Modification of the risk set will only result in unbiased estimates of survivor functions (Kaplan-Meier estimator) or regression coefficients (e.g., from a Cox model) if the quasi-independence[1] assumption is satisfied
- Quasi-independence requires entry times and event times to be independent in the "observable region"
  - The "observable region" is the region where patients are not left-truncated (i.e., where entry times are less than event times)
- In a regression context, we only need to assume conditional quasi-independence; that is, quasi-independence within covariate strata[2]

[1] Tsai (1990)
[2] Gross & Lai (1996)

Testing the quasi-independence assumption

- The conditional Kendall's Tau test[1] is perhaps the most common in a non-regression context, although alternative tests exist[2]
- A simple approach is to test for dependence with a Cox model, which can easily incorporate covariates

```
coxph(Surv(entry_time, followup_time, event) ~ entry_time)
```

A significant `entry_time` coefficient suggests that there may be dependent truncation

[1] Tsai (1990)

[2] See Jones & Crawley (1992) and a review by Martin & Betensky (2005)

Agenda

1. Estimation of hazard ratios (HRs) with the Cox model
2. Prediction with penalized Cox models

ESTIMATION OF HAZARD RATIOS WITH THE COX MODEL

A simulation study

Data generating process

- Time-to-event (T), study entry (E), and right censoring (C) drawn from proportional hazards models

$h(c) = h_0(c)$
$h(e) = h_0(e)exp[x\gamma]$
$h(t) = h_0(t)exp[x\beta + \rho log(e)]$

$\gamma \neq 0 \rightarrow$ covariates associated with entry time
$\rho \neq 0 \rightarrow$ quasi-dependence

- Left-truncated and right-censored (LRTC) survival data is generated where only followup time *Y* is observed

$Y = min(T, C)$
Right-censored if $C < T$
Left-truncated if $E > Y$

Settings

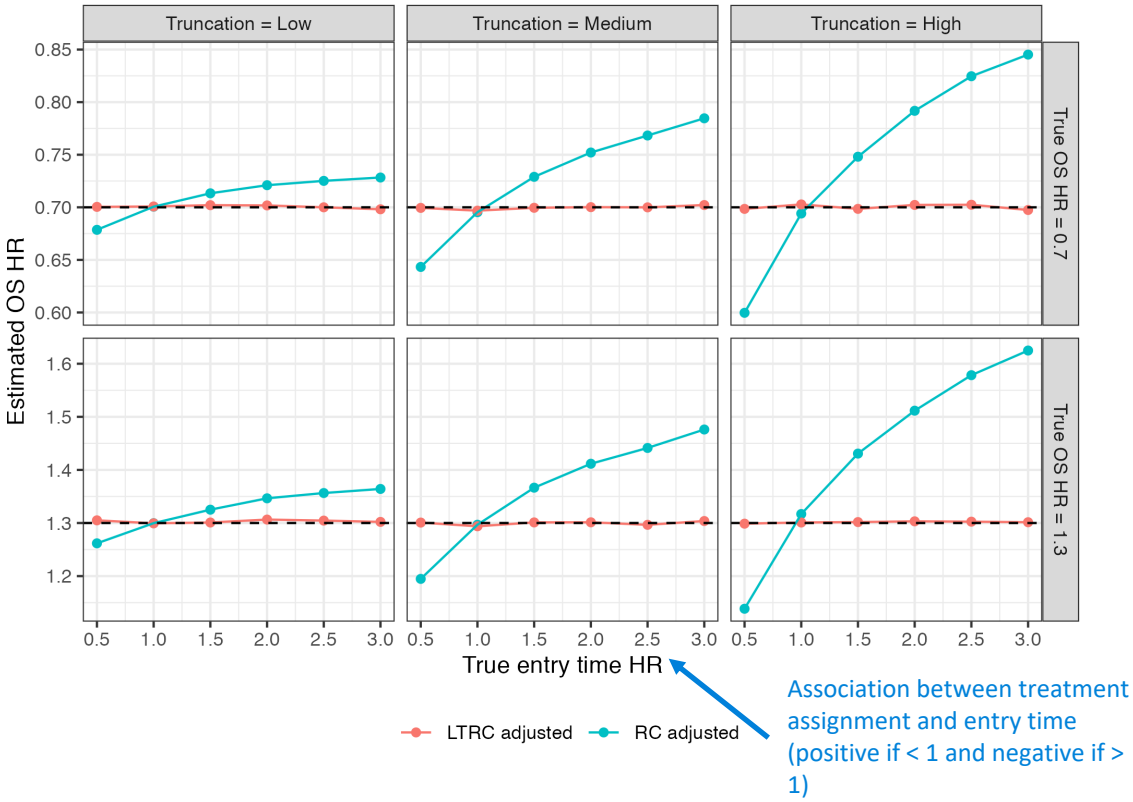
- Conceptualize time-to-event as overall survival (OS)
- Baseline hazards, h_0, exponentially distributed
- One covariate for treatment assignment (1 = treat, 0 = control)
- Vary association between treatment assignment and entry time
- Three truncation scenarios: low (~5% of patients truncated), medium (~15 % truncated), and high (~30% truncated)

Two methods evaluated

Name	Risk set adjustment	Data for estimation
RC adjusted	No	Sample with Y > E
LTRC adjusted	Yes	

Bias in Cox models under quasi-independence

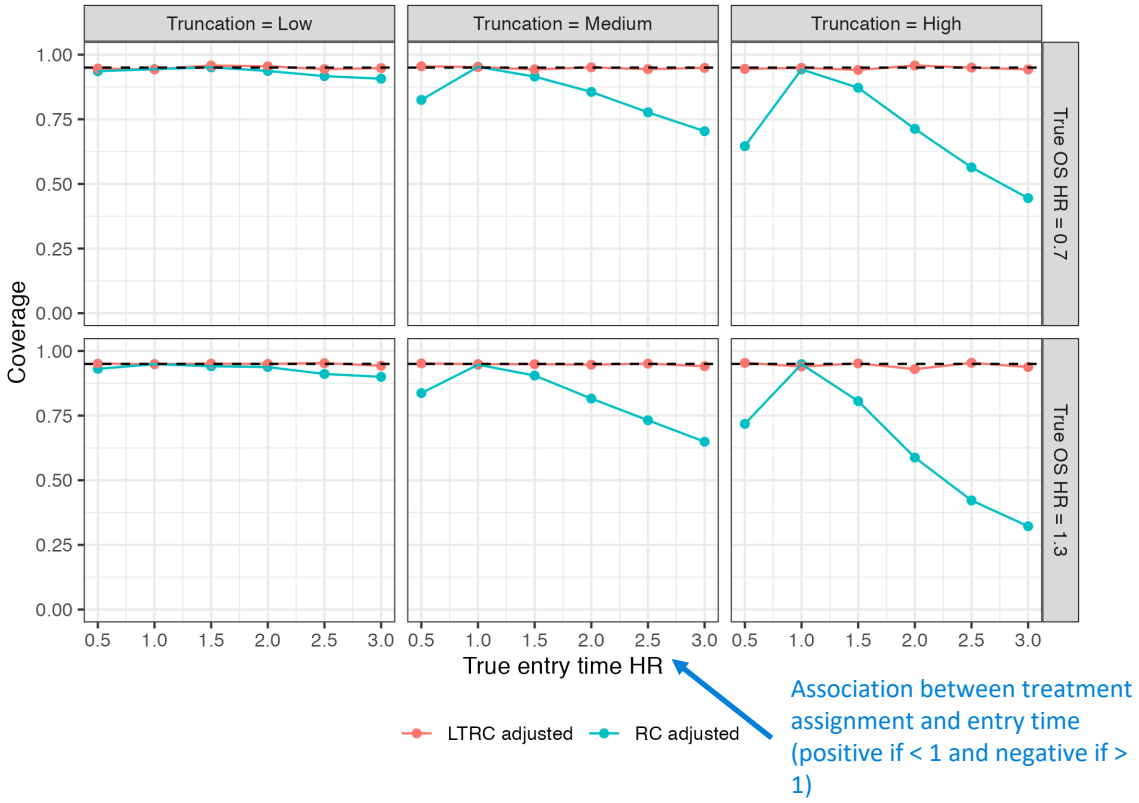
- OS HRs in models that *adjust the risk set* (LTRC adjusted) are unbiased
- OS HRs in models that *do not adjust* the risk set (RC adjusted) are biased
  - The bias is increasing in (i) the proportion of patients that are truncated and (ii) the magnitude of the association between treatment assignment and entry time (true entry time HR)
  - The direction of the bias does not depend on the direction of the true OS HR



Notes: OS = overall survival. The horizontal line is value of true OS HR. The true entry time HR is exp(γ) and represents the association between the treatment assignment covariate and entry time

Coverage of 95% confidence intervals in Cox models under quasi-independence

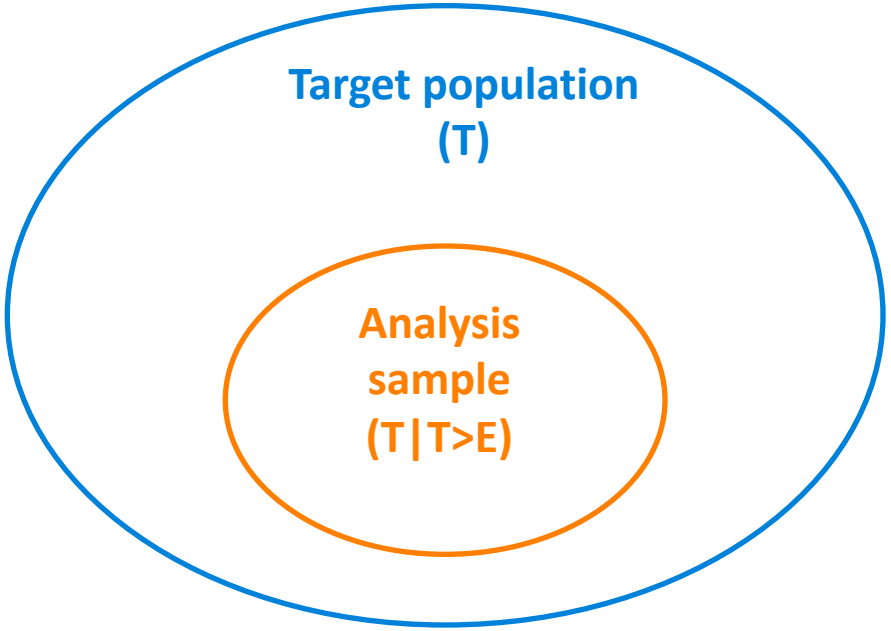
- Coverage of 95% confidence intervals in models that *adjust the risk set* (LTRC adjusted) are at the 95% nominal probability
- Coverage of 95% confidence intervals in models that *do not adjust* the risk set (RC adjusted) is too low.
  - Undercoverage is increasing in (i) the proportion of patients that are truncated and (ii) the magnitude of the association between treatment assignment and entry time (true entry time HR)
  - The coverage probabilities do not depend on the direction of the true OS HR



Notes: The horizontal line the desired coverage of the 95% confidence interval. The true entry time HR is exp(γ) and represents the association between the treatment assignment covariate and entry time

Conditional vs. marginal hazard ratios

Only observe a subset of the target population



A conditional HR is estimated when fitting the Cox model on the truncated truncated analysis sample in that it does not depend on x when conditioning on a given patient's covariate profile X=x[1]

The marginal HR is the HR among patients in the entire target population and is arguably more relevant

Under **quasi-independence**, a marginal HR can be estimated by using covariates to weight the analysis sample to look like the target population (i.e., by using IPTW-ATT weights)[2,3]

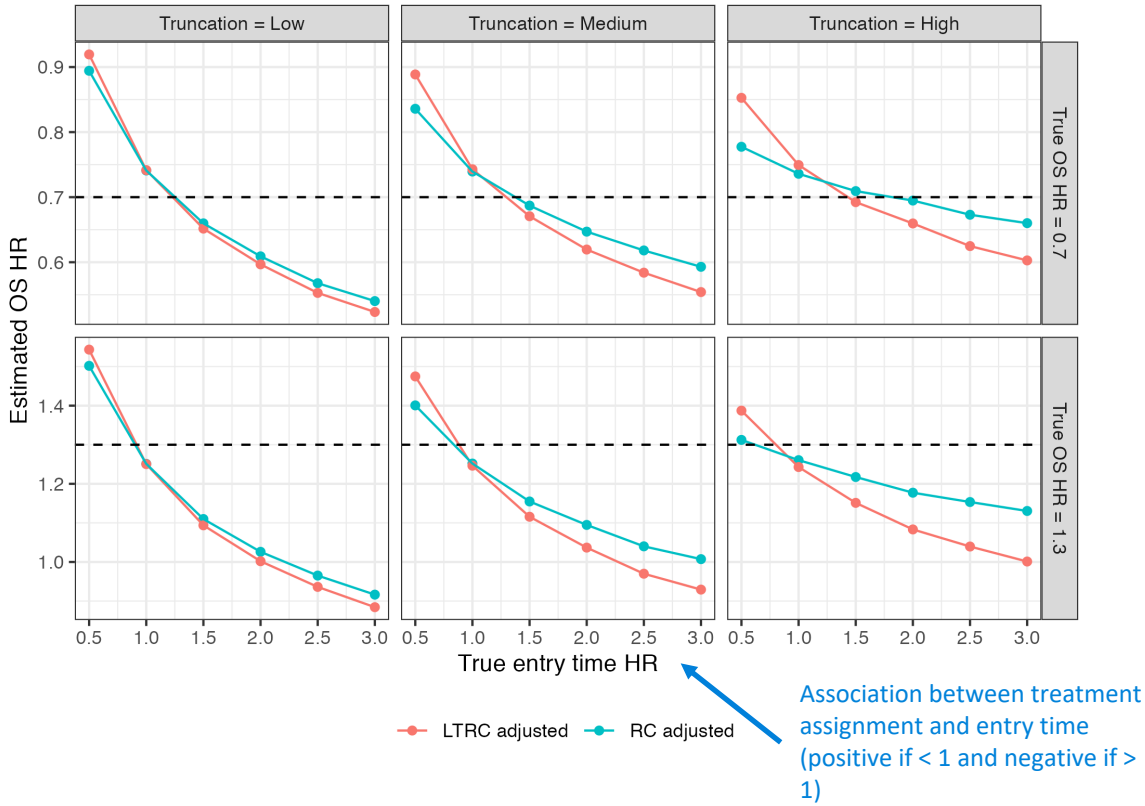
[1]Daniel et al. (2020)

[2]Sondhi (2021)

[3]*IPTW-ATT = inverse probability of treatment weights permitting estimation of the average treatment effect on the treated*

Bias with dependent truncation

- We introduce dependent truncation by setting $\rho = log(1.5)$; i.e., an OS HR of 1.5 for the log of entry time
- In this case, the OS HR (for treatment assignment) is biased even when modifying the risk set
- The magnitude of bias can be large and depends the relationship between treatment assignment and entry time (true entry time HR)
- Although not shown here, the bias is also increasing in ρ (i.e., the strength of the relationship between entry and event times)



Notes: OS = overall survival. The horizontal line is value of true OS HR. The true entry time HR is exp(γ) and represents the association between the treatment assignment covariate and entry time.

Strategies for dependent truncation

- A number of approaches have been proposed to deal with dependent truncation
  - Including entry time as a covariate in a Cox model [Mackenzie (2012)]
  - Copula [Chaieb et al. (2006)] and transformation [Chiou et al. (2019)] methods
- There are, however, a number of challenges:
  - Only valid under strong assumptions
  - Often not designed for a regression context or for estimation of marginal HRs[1]
  - Complicated to implement
- For these reasons, Sondhi et al. (2021) propose simulation based sensitivity analyses to assess the potential direction and magnitude of the bias

[1]Marginal HRs are more complex with dependent truncation because they depend on the distribution of event times in the target population (which are unobserved)

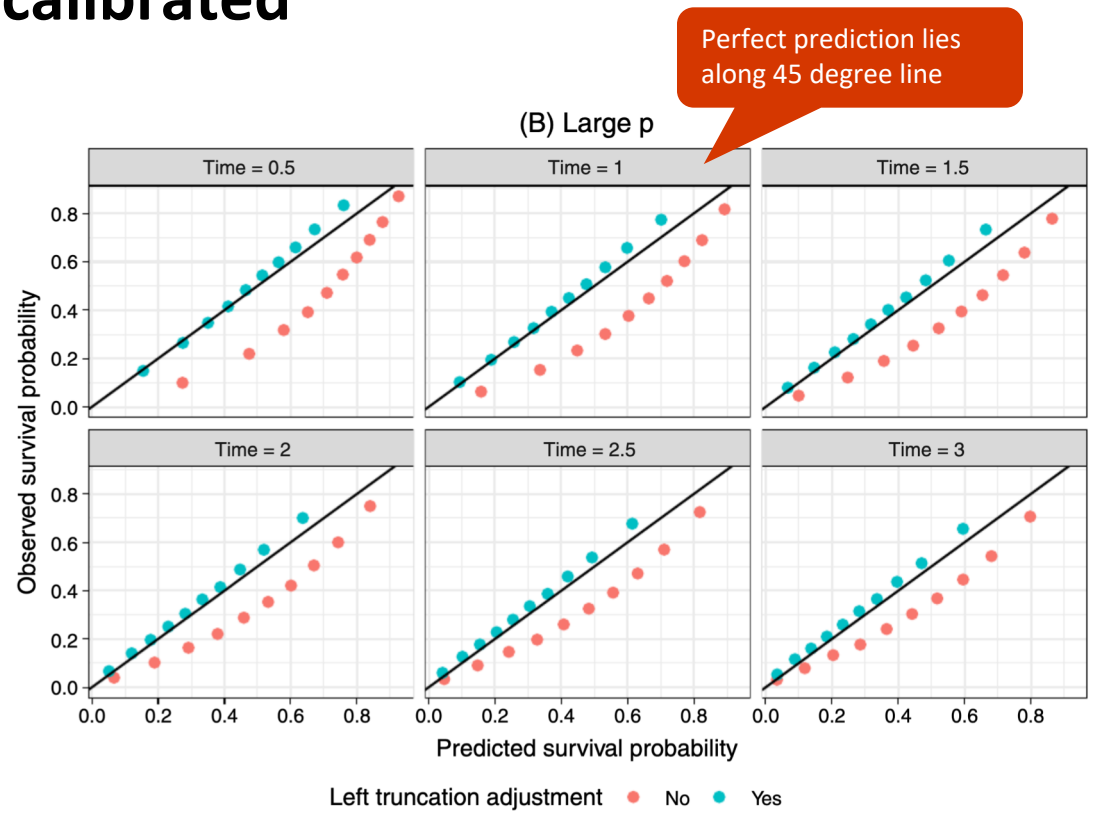
PREDICTION WITH PENALIZED COX MODELS

Model estimation and evaluation in a prediction setting

- In typical prediction problems, the objective is to predict survival probabilities as accurately as possible, not to make causal claims
- Like the standard Cox model, penalized Cox models with left-truncated and right-censored survival data can be fit by modifying the risk set
  - A similar approach can be used when machine learning methods (e.g., random forests, neural networks) are using for survival analysis
- Standard metrics for model evaluation can still be considered (e.g., prediction error curves, Brier scores, C-index), but care is needed since both the training and test sets will often be left-truncated
- A complication for both model fitting and model evaluation is that software is not usually designed for left-truncated survival data
  - Requires writing custom code

Failing to adjust the risk set results in predicted survival probabilities that are poorly calibrated

- The out-of-sample predictive performance of Cox models with lasso penalties were evaluated with a simulation study[1,2]
- A simulation allows us to estimate models on a left-truncated training dataset and evaluate models on a complete test dataset
- The "large p" model included 1,011 predictors
- While models that adjust the risk set are well calibrated, failing to adjust the risk set results in predicted survival probabilities that are too high



Notes: Unit of time is in years. A left truncation adjustment was made by modifying the risk set. Patients were divided into deciles at each time point based on their predicted survival probabilities and each point in the plot represents patients within a decile. The "Predicted survival probability" is the average of the predicted survival probabilities from the Cox model across patients within each decile and the "Observed survival probability" is the Kaplan-Meier estimate of the proportion surviving within each decile.

[1]Parameters of data generating model calibrated using advanced non-small cell lung cancer patients in a database provided jointly by Foundation Medicine and Flatiron Health that links genomic and clinical information

[2]Source: McGough et al. (2021)

Is the C-index a good measure of performance?

- The C-index is a measure of the extent to which a model can *discriminate* prognosis across patients and is commonly used to evaluate survival models
- We computed the C-index in the test set within the simulation study described on the previous slide
- As shown in the table to the right; the C-index is higher in the risk set adjusted model even though it is poorly calibrated
- The C-index may therefore be misleading with left-truncated data because biased coefficients can incorrectly result in a more discriminating model

Risk set adjustment	C-index
No	0.72
Yes	0.67

Note: C-index evaluated on test set among the observed (i.e., non-truncated patients in the simulation study

Summary

- Careful consideration of the data generating process and estimand is critical when survival data is left-truncated and right-censored
- When entry and event times are (conditionally) quasi-independent, it is sufficient to adjust the risk set so that patients are only “at risk” of an event subsequent to entry date
  - HRs in Cox models are unbiased and have correct coverage
  - Predicted survival probabilities from Cox models are well calibrated
- There is no established best practice when entry and event times are quasi-dependent; sensitivity analyses are recommended

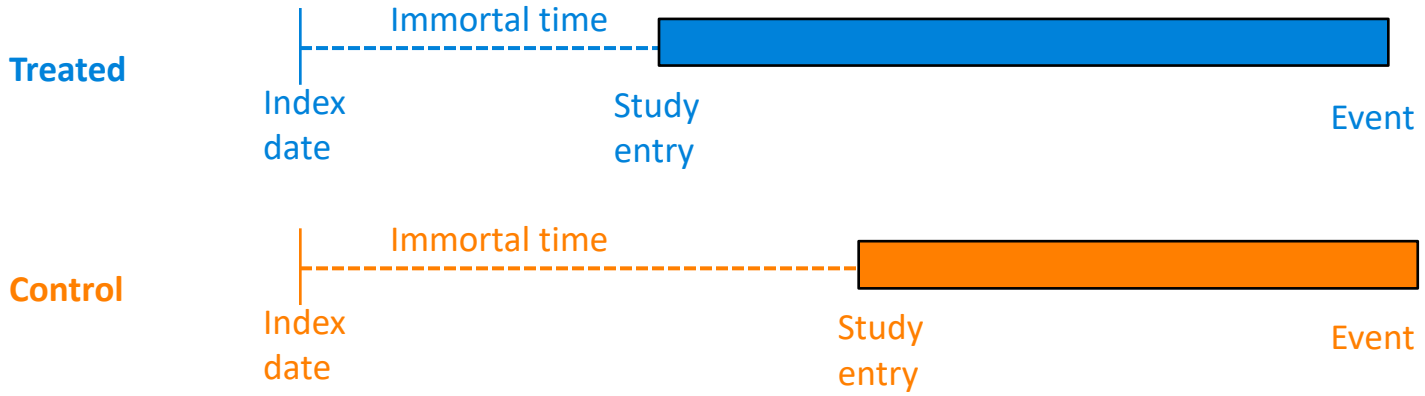
Doing now what patients need next

Characterizing bias of hazard ratios in Cox models with quasi-independence

- Consider an "entry time HR" measuring the relationship between a treatment assignment variable (1 = treated, 0 = control) and entry time

Parameter range	Interpretation	Direction of bias (without risk set adjustment)
Entry time HR > 1	Treated has shorter time to entry than control	> 0
Entry time HR < 1	Treated has longer time to entry than control	< 0

- To understand the direction of bias, consider the case where the entry time HR > 1 (results are reversed with HR < 1)



Treated group have less immortal time than control group

So survival is longer in the control group, increasing the HR

Coverage of 95% confidence intervals in Cox models with dependent truncation