

Parametric Event-Prediction For Time-to-Event Data in Clinical Trials

Jingyuan Chen, Weijie Zhang, Chenglin Ye

Genentech

March 2023, Orlando, US

Overview

- Objectives and Incentives
- Methodologies
- Case Studies and Simulation Studies
- Events projection app
- Summary

Objectives and Incentives

- **Goal:** to predict a date when a desired number of events will happen in an **ongoing** trial; interim or final analysis (FA) clinical cutoff date (CCOD).
- **Motivations and business impact**
 - predict timeline for ongoing trials; optimization of resource allocation and logistics arrangements; filing planning.
 - inform activities such as database lock, top-line review, regulatory submission, manufacture preparation, publication, and product launch.
 - help support portfolio review and prioritization by senior management.
- **Time-to-event data:**
 - patients recruited and have already developed the event
 - patients recruited but not yet developed the event
 - potential future recruitment of patients

Methodologies

- *Conditional approach*
 - Based on Kaspar Rufibach's R package `eventTrack`.
 - **Conditional on** patients who were still event-free and on patients yet to be enrolled.
- *Marginal approach*
 - Based on a **marginal probability** of when a certain proportion of events is achieved.

The Conditional Approach ([Github](#))

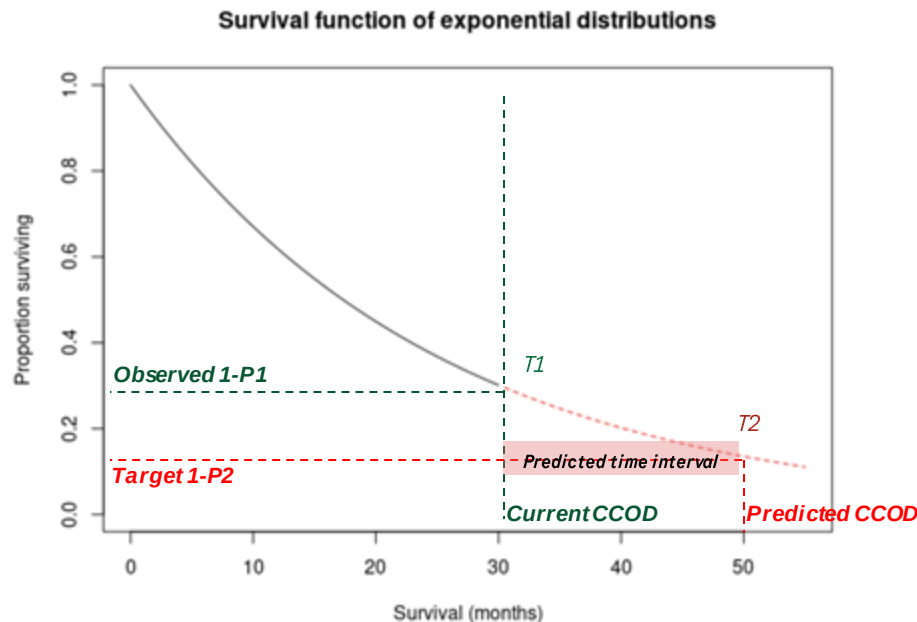
$$\begin{aligned}
 m(t, S) &= \sum_{h \in \mathcal{I}_1} 1 + \sum_{j \in \mathcal{I}_2} P(\text{event in } (t_j, T] | \text{no event up to } t_j) + \sum_{i \in \mathcal{I}_3} P(\text{event in } (a_i, T]) \\
 &= n + \sum_{j \in \mathcal{I}_2} \frac{S(t_j) - S(T)}{S(t_j)} 1\{T > t_j\} + \sum_{i \in \mathcal{I}_3} (1 - S(T - a_i)) 1\{T > a_i\}.
 \end{aligned}$$

- $m(t, S)$: No. of events at time t given survival function S .
- $\mathcal{I}_1$: patients who already had the event, n .
- $\mathcal{I}_2$: patients who are censored at t_j .
- $\mathcal{I}_3$: patients to be recruited at a_i .

WORKFLOW

1. Estimate parameters based on existing data up to current date of data cut.
2. Construct parametric survival function S .
3. Specify future recruitment (could be zero).
4. Use the equation to estimate future events and dates.

The Marginal Approach



P1: Observed event proportion till current CCOD

P2: Desired event proportion

WORKFLOW

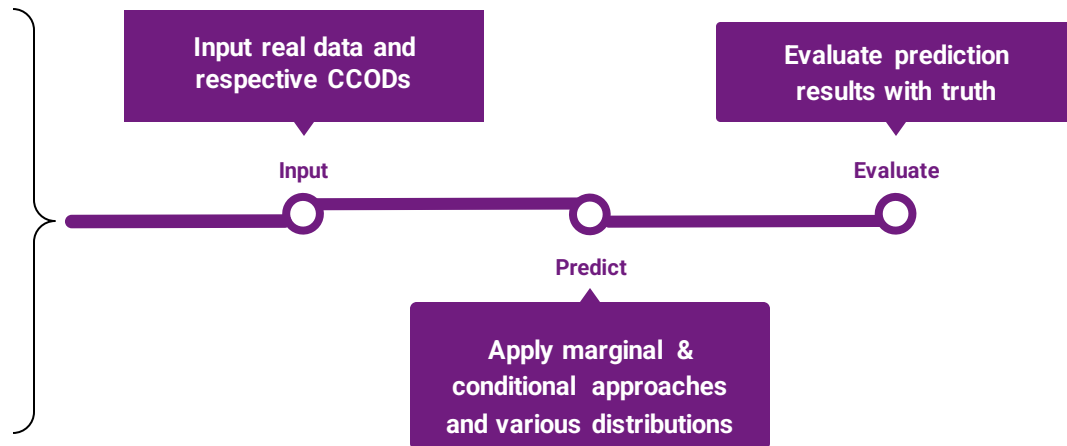
1. Estimate parameters based on existing data before current CCOD.
2. Estimate parametric survival functions and density functions.
3. Calculate current CCOD event rate **P1** and target event rate **P2**
4. Estimate **time interval** between current CCOD and future CCOD via $F^{-1}(\mathbf{P2}) - F^{-1}(\mathbf{P1})$
5. Obtain predicted future CCOD with desired number of events via
 $\mathbf{Current\ CCOD} + \mathbf{Predicted\ time\ interval}$

Case studies using real data

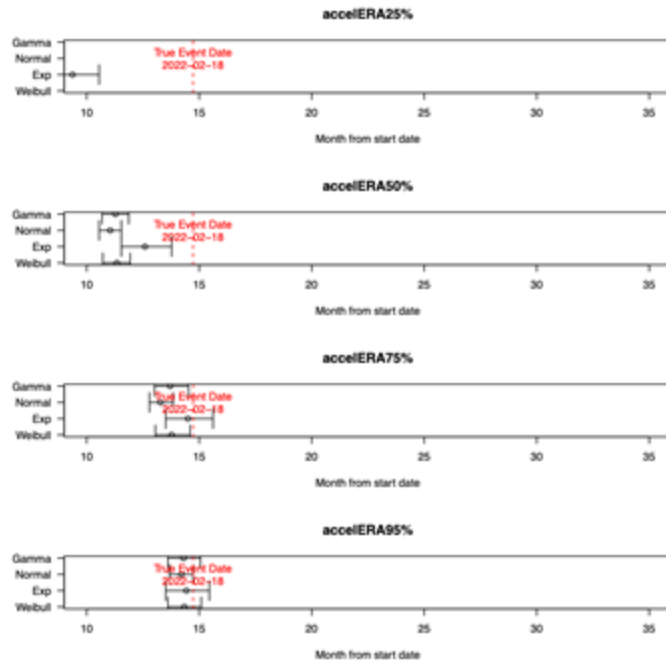
Data Sources



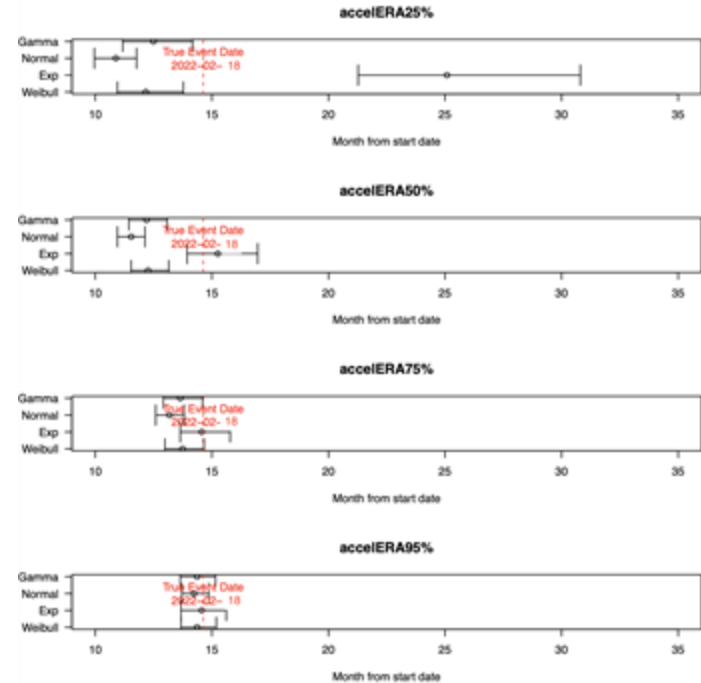
Study approach



Conditional Approach:

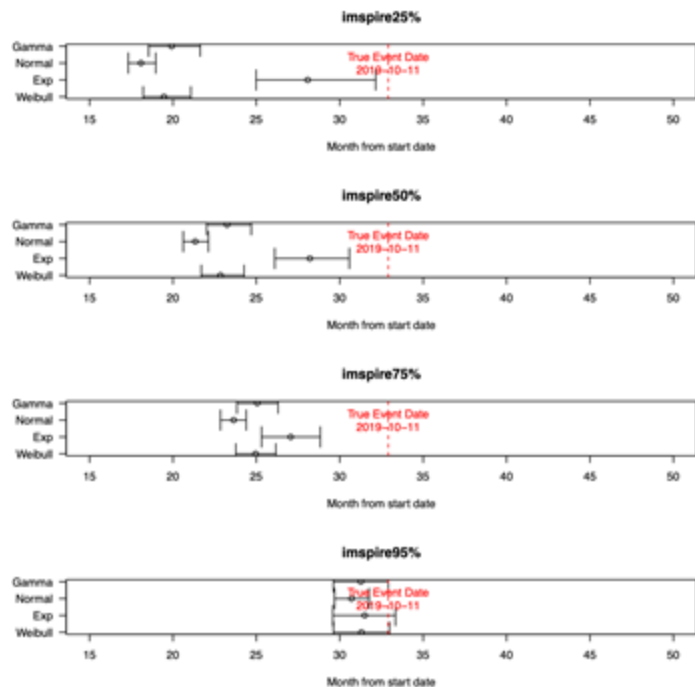


Marginal Approach:

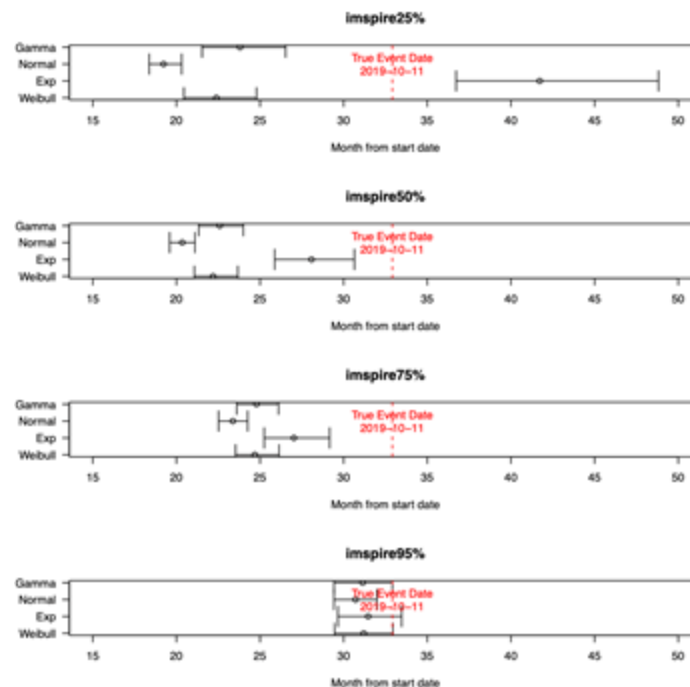


- A short-term enrollment study (~1yr, N=303, Events=181);
- Exponential gives a good prediction in this situation;

Conditional Approach:

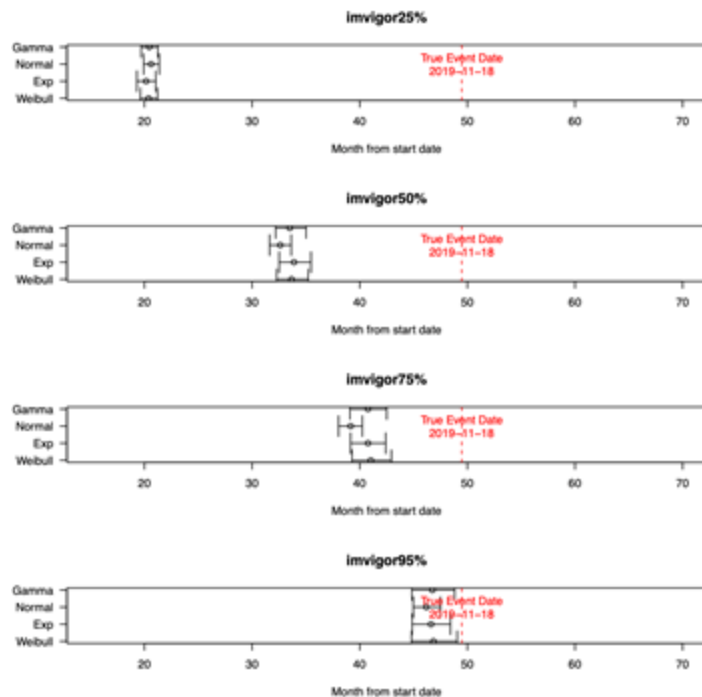


Marginal Approach:

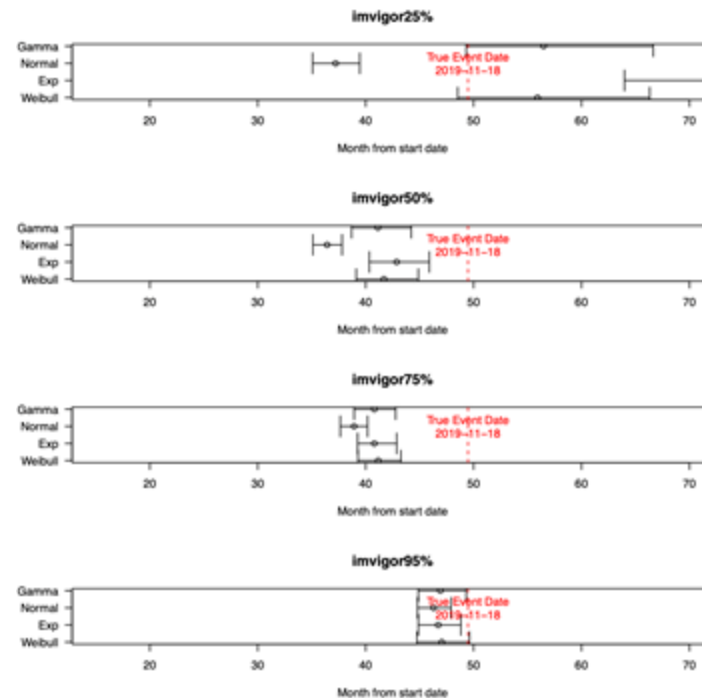


- A study with moderate sample size (N=518, Events=327, ~1.5 yrs).
- The two approaches yield similar results with 75% data and beyond.

Conditional Approach:



Marginal Approach:



- A study with large sample size (N=809, Events=420), long time of recruitment (~3 yrs)
- At early stage (~25%), marginal shows better predictions than conditional.
- Immature data provided poor prediction

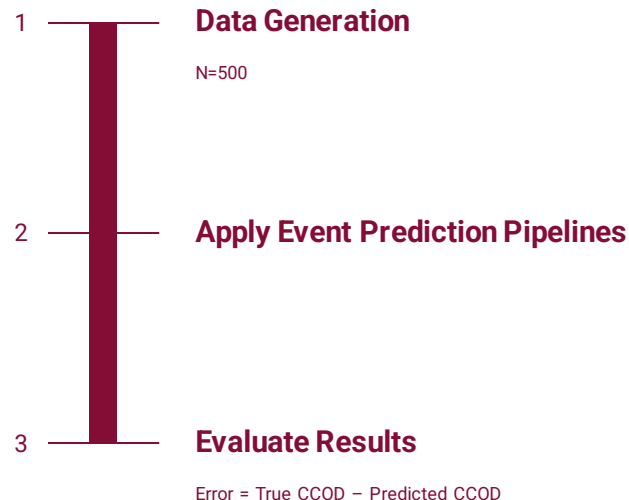
Messages from Case Studies

- We recommend using data with **75% events** for reasonable prediction.
- In the case of accelERA, immature data (50% events data) provided good predictions. Additional studies may be needed to better understand the feasibility of early predictions, in which case 80% or wider CIs may be considered.
- Studies with a big sample size and a long-term recruitment, marginal prediction shows better performance at early time when data not mature (before enrollment finished).
- Studies with short-term enrollment, prediction from exponential approach shows better performances.
- **As data maturity increases, both marginal and conditional provide similar results.**

Simulation studies

Data Generation

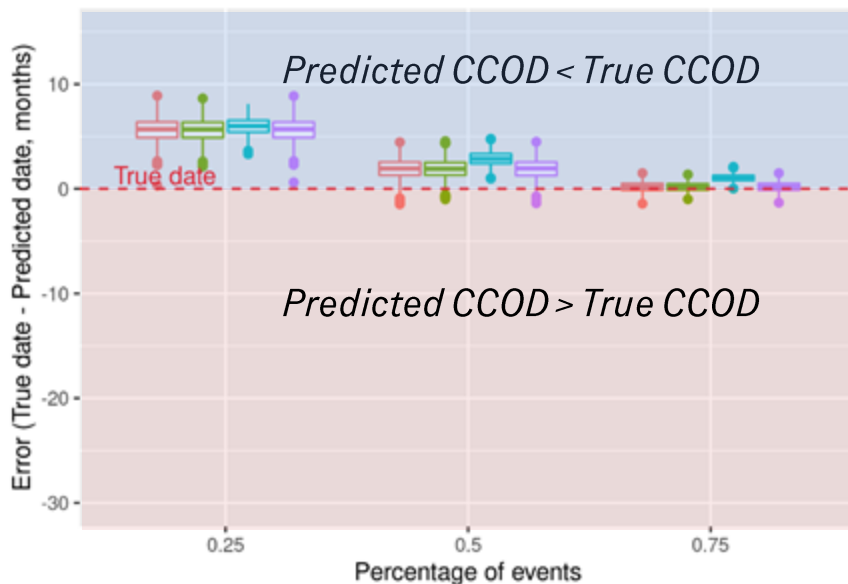
1. From real data (*acceleRA*, *IMvigor*, and *IMspire*) obtain accrual for each month and median survival of each arm.
2. Simulate accrual of each month assuming a uniform rate.
3. Simulate survival time using an **exponential distribution** with median survival months for each arm.
4. Set CCODs when the desired number of events happens.
5. Censor/process data using the CCODs.



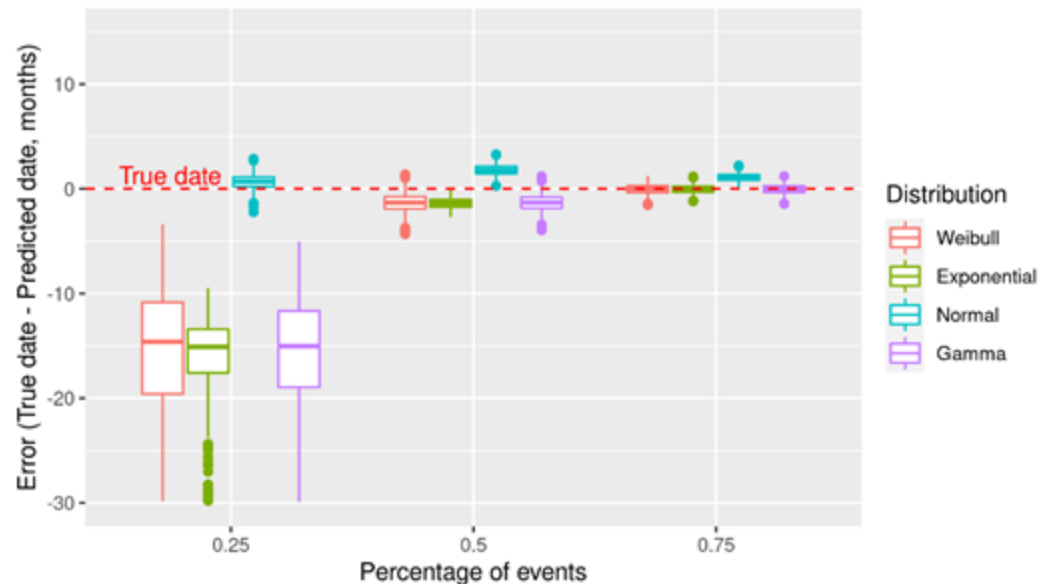
accelERA

N=313; Events=181

accelera Simulation (Conditional)



accelera Simulation (Marginal)

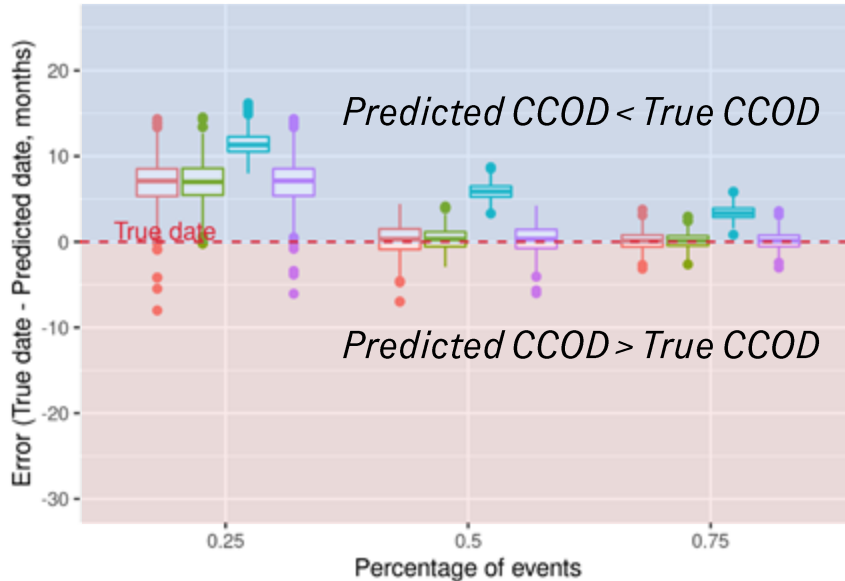


- 50% and 75% gave good results.
- Normal distribution in marginal models performed well with 25% data. Also gave predicted CCODs < Truth.

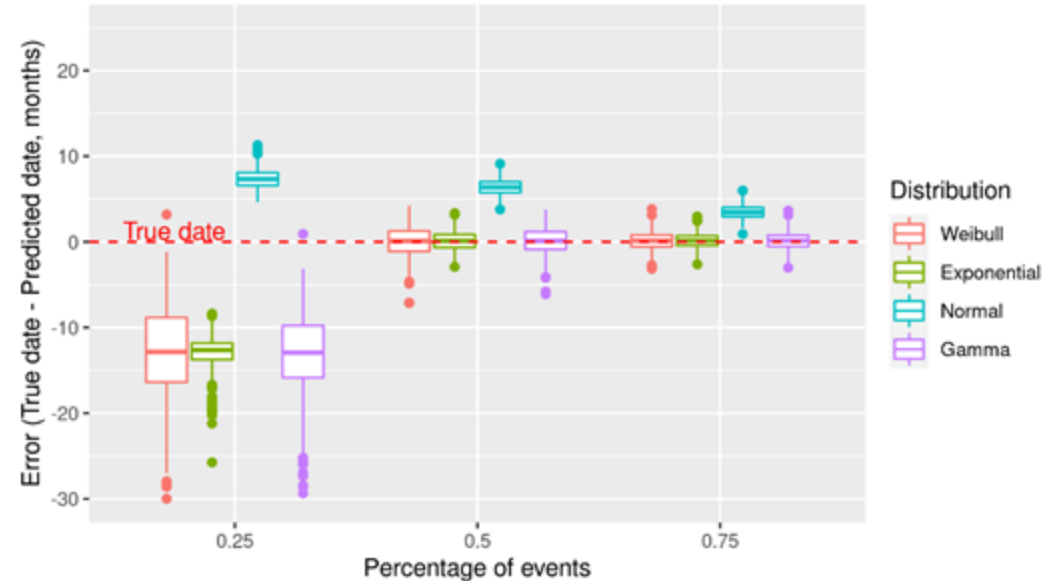
IMpsire

N=518; Events=327

impsire Simulation (Conditional)



impsire Simulation (Marginal)

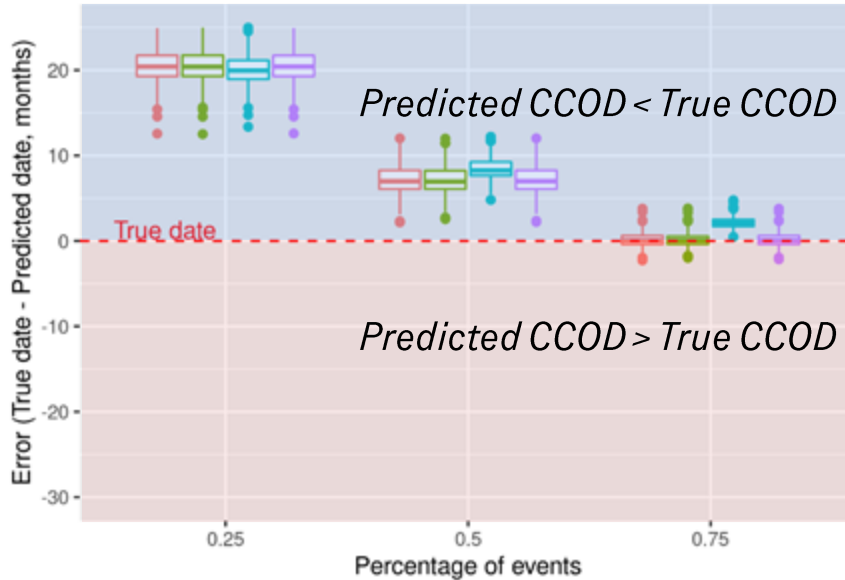


- Sample size between accelERA and IMvigor.
- 50% and 75% both gave good results.
- Normal distribution models gave predicted CCODs < true CCODs in all situations.

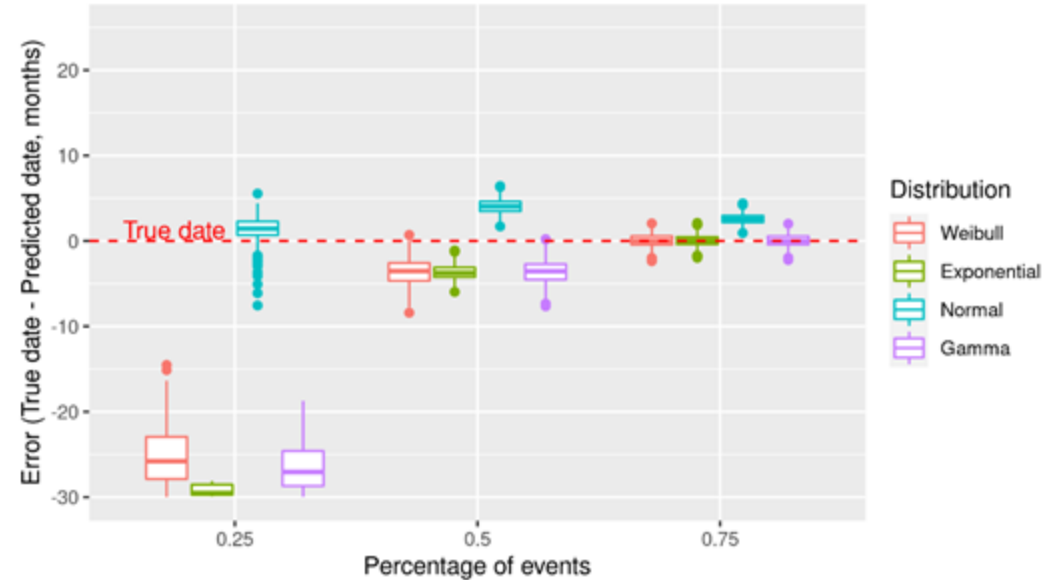
IMvigor

N=809; Events=420

imvigor Simulation (Conditional)



imvigor Simulation (Marginal)



- Only 75% gave good results.
- Normal distribution marginal models performed well with 25% data. Also gave predicted CCODs < truth in other situations.

Simulation study findings

- **Conditional** approach tends to yield **predicted CCODs < true CCODs**; predicted CCODs approach true dates as data become more mature.
- **Marginal** approach is generally **volatile** with very limited data size (e.g., 25%).
- For a few cases with **25%** data, **normal distribution** models with the **marginal approach** give surprisingly good results.
- On the other hand, **normal distribution** models also tend to give **predicted CCODs < true CCODs** compared to other three distributions – consistent with Case studies.
- The normal distribution assumes the risk of failure increases with time, which is less applicable to human subject studies.
- Limitations: simplifying reality; only one distribution.

Predict Analysis Timepoints for Clinical Trials

Welcome! The goal of this app is to predict a future **clinical cut-off date (CCOD)** with an expected number of events; i.e., when a study will reach a certain number of events.

Instructions on using the App

A typical workflow includes

- 1. Upload data. We also provide sample datasets.

Please make sure your data contains variables AGE, RACE, COUNTRY, and SEX. Your data should also contain variables indicating ID, Randomization times/start times, CNSR, and analysis date (See data upload page for details).

- 2. Exploratory data analysis. We provide functionalities to visualize several variables.
- 3. Predict CCODs. We implement two different approaches, namely Marginal and Conditional. If your study recruitment has not yet completed and you wish to input futural accrual, please use the conditional approach as the marginal approach does not support this specification.
- 4. Choose if confidence intervals should be calculated and specify the percentage accordingly. You may find more information under the prediction pages.

Methodologies

Here, we implement two different approaches to for event prediction, namely **Conditional** and **Marginal**.

Conditional:

The Conditional approach directly implements the method described by Rufibach et al. with R package [eventTrack](#). Briefly, the function `predictEvents` partitions events into 3 parts: observed events, future events conditioning on patients who were still event-free, and future events from patients yet to be enrolled.

As described by Kaspar Rufibach, the expected number of events, $m(T, S)$, at a given timepoint T , depends on T and on the survival function at which events happen. If these two quantities are known, we can predict m using the following formula:

$$\begin{aligned} m(t, S) &= \sum_{h \in I_1} 1 + \sum_{j \in I_2} P(\text{event in } (t_j, T] | \text{no event up to } t_j) + \sum_{j \in I_3} P(\text{event in } (a_i, T]) \\ &= n + \sum_{j \in I_2} \frac{S(t_j) - S(T)}{S(t_j)} 1\{T > t_j\} + \sum_{j \in I_3} (1 - S(T - a_i)) 1\{T > a_i\}. \end{aligned}$$

This formula defines the following sets of patients:

- I_1 : patients who already had the event, n .
- I_2 : patients who are censored at t_j .
- I_3 : patients to be recruited at a_i .

For more information please visit the [eventTrack](#) Github Page.

Marginal

Summary

- In practice, **Conditional and the Marginal approaches** can be used to predict future CCODs.
- **Events Projection Shiny app** has been constructed to provide event prediction and exploratory data analysis.
- **Case studies and simulation studies all support**
 - recommendation of using ~75% events for prediction, while varying depending on study sizes and recruitment speeds.
 - the closer prediction start date and true CCOD are, the better prediction results will be (even with less mature data, e.g., 50% in accelERA and IMspire).

Acknowledgement

Kaspar Rufibach (Roche colleague)

Colleagues who provides study data:

Haney Ahmed

Reference

Rufibach, Kaspar. 2022. *Event tracking: predict analysis timepoints for ongoing clinical trials*.

Rufibach, Kaspar. 2020. *EventTrack: Implements Functions and Datasets for Event Tracking*.

Bell, James. 2020. *Gestate: Generalised Survival Trial Assessment Tool Environment*. <https://CRAN.R-project.org/package=gestate>.

“Event Projection: Quantify Uncertainty and Manage Expectations of Broader Teams.” Talk at BBS seminar 18th April 2016 <http://bbs.ceb-institute.org/?p=858>.

Fang, L., and Z. Su. 2011. “A hybrid approach to predicting events in clinical trials with time-to-event outcomes.” *Contemp Clin Trials* 32 (5): 755–59.



***Doing now what patients need
next***