Real-World Dynamic Risk Assessment of Oral Corticosteroid Use: A Bayesian Time-Varying Survival Analysis

Zhengfan Wang, Aaron Crowley, Craig S. Parzynski, Bret Zeldow
GENESIS RESEARCH GROUP

Introduction

Background and Motivation

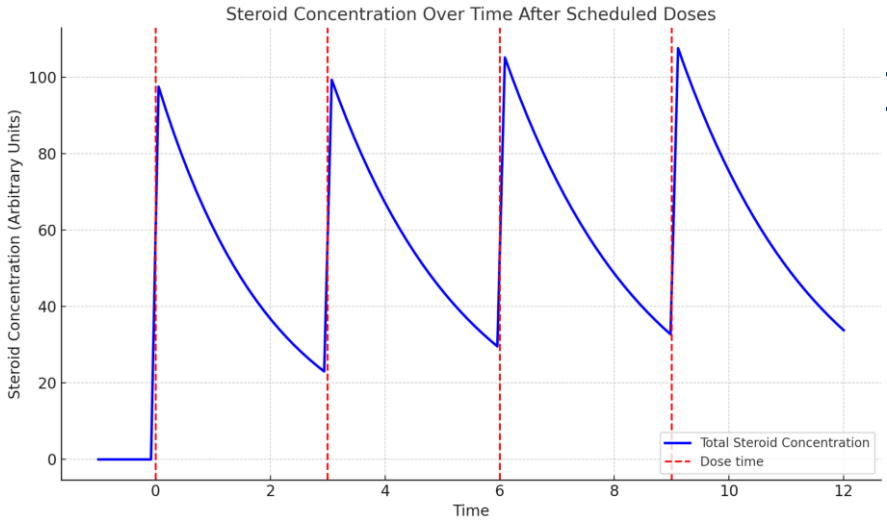
Background

- Oral corticosteroids (OCS) are widely prescribed for inflammatory and autoimmune conditions
- Prolonged OCS use is associated with serious side effects (e.g., cardiovascular events, osteoporosis, infections).
- The real-life impact of cumulative OCS use, which may differ from controlled clinical trial settings.
- OCS use is tracked in large real-world cohorts using electronic health records (EHRs) and claims data.

Motivation

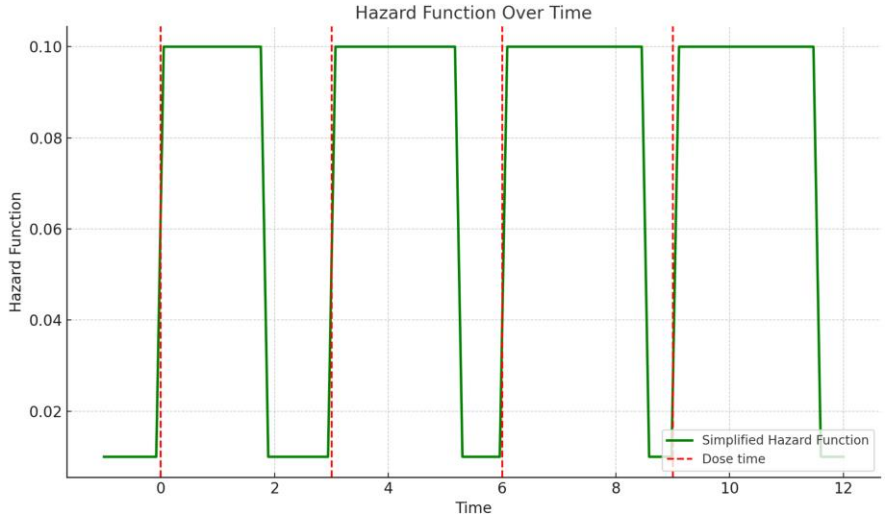
- In real-world populations, the risk of adverse events from OCS use varies over time, making it essential to capture dynamic changes in hazard rates, which are influenced by real-world treatment patterns, adherence, and co-existing conditions.
- Using real-world data allows us to track cumulative OCS exposure and better reflect the diversity of patient experiences, such as comorbidities and varying doses, that influence risk.

OCS concertation and hazard function over time



Threshold effect:

- **The risk increases once the concertation passes threshold**
- **Effect diminishes until the next OCS**

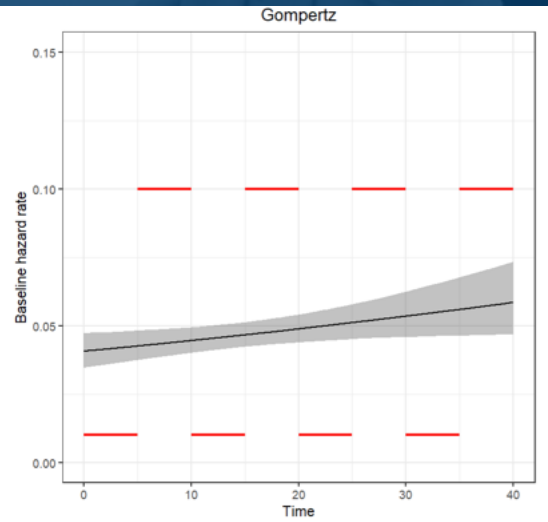


The hazard function depends on whether the cumulative dose passes the threshold right after administration.

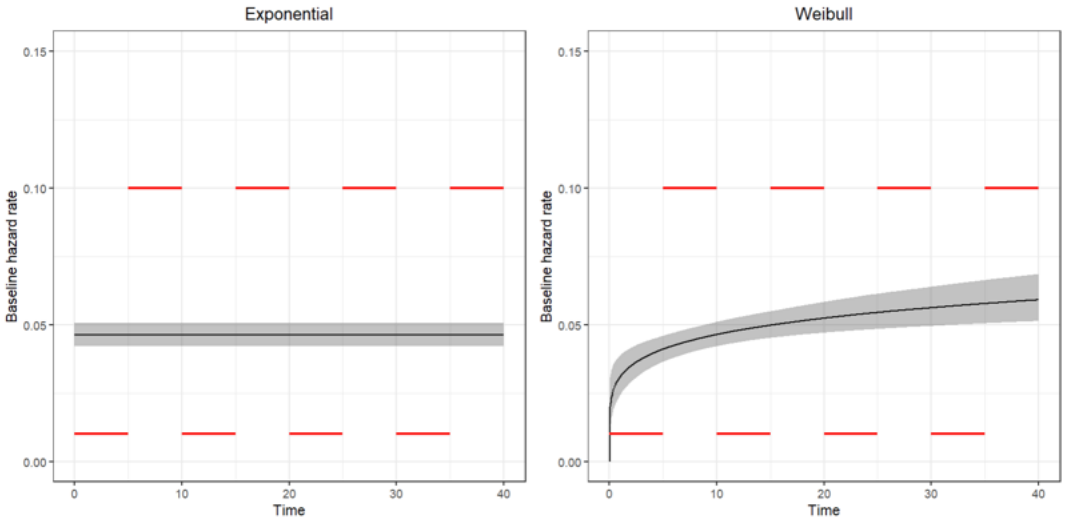
Abbreviations: VSC, Value and Scientific Communications

Why Bayesian?

- Understanding the impact of cumulative OCS exposure on adverse events is critical for patient safety.
- Traditional cox models often assume constant effect, and most parametric distributions does not flexible enough. Because OCS dosage and effect change over time, making it challenging to estimate accurately.



- Bayes Model
  - Bayesian models provide a flexible framework to incorporate prior knowledge, handle complex relationships, and accommodate time-varying effects.
  - Allows estimation of uncertainty around model parameters, offering robust inference in the presence of varying cumulative exposure.



Piecewise constant baseline hazard model

Notation of Piecewise Constant Baseline Hazard model

Divide the time axis into K intervals: $[0, \tau_1), [\tau_1, \tau_2), \ldots, [\tau_K, \infty]$.

- The **baseline hazard function** for time $t \in [\tau_{k-1}, \tau_k]$ is
$$h_0(t) = \lambda_k$$
- The **cumulative hazard function** $H_0(t)$ up to time t is the sum of the hazard across the intervals
$$H_0(t) = \sum_{k=1}^{j-1} \lambda_k(\tau_k - \tau_{k-1}) + \lambda_j(t - \tau_{j-1})$$
- The **survival function** $S(t)$, which gives the probability of surviving beyond time t, is related to the cumulative hazard by:
$$S(t) = \exp(-H_0(t))$$

Likelihood of the Piecewise Constant Hazard Model

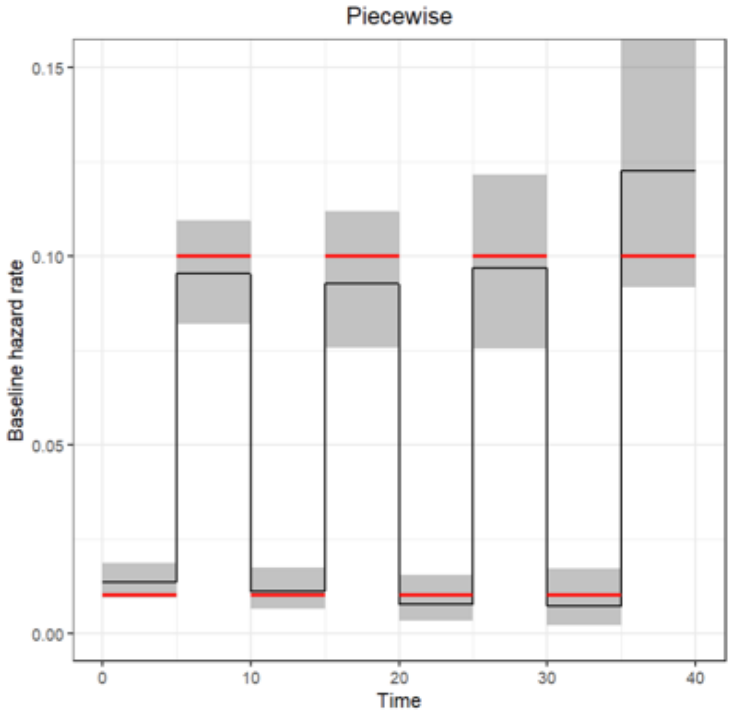
The **Likelihood** for the piecewise constant hazard model is

$$L_i = \left(\prod_{k=1}^{j-1} \exp(-\lambda_k(\tau_k - \tau_{k-1})) \right) \exp\left(-\lambda_j(t_i - \tau_{j-1})\right) \lambda_j^{\delta_i}$$

where **brown term** represents that survival probability up to τ_{j-1}, the **blue term** is the probability of surviving until time t_i within the its interval, and the **black term** accounts for the probability of an event occurring at t_i if it is not censored.

A More Flexible Hazard Function is Needed

- The Piecewise constant hazard model is the most effective in estimating the true hazard rate due to its ability to model changes at specific intervals, (Exponential, Weibull, Gompertz) fail to capture the periodic nature of the hazard.
- But in the real world, the situation would be more complicated.
  - OCS dosing schedules fluctuate.
  - The effect of OCS on the hazard rate may not be constant over time.
  - The effect of OCS also depends on the covariates.
  - The result is very sensitive to the setting of the time intervals.
- A more flexible/smoother estimation is necessary.



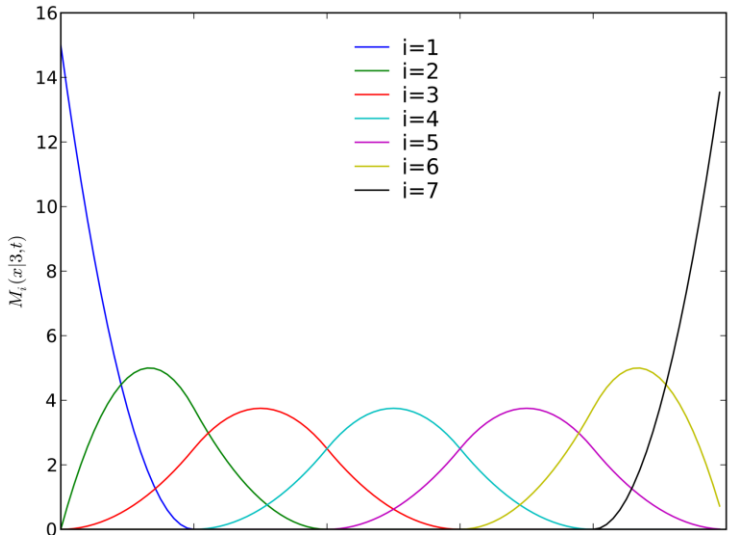
Bayesian spline hazard model with time-varying effect

Introduction to Spline in Survival Analysis

- Splines are flexible functions used to model relationship between variables
- They are piecewise polynomial functions joined together smoothly at specific points called knots.
- Splines are useful when relationship between variables is non-linear, but the exact form of the curve is unknown. It allows to model more complex, time-varying hazard functions, capturing changes that other parametric models (Exponential, Weibull, or Gompertz) might miss.

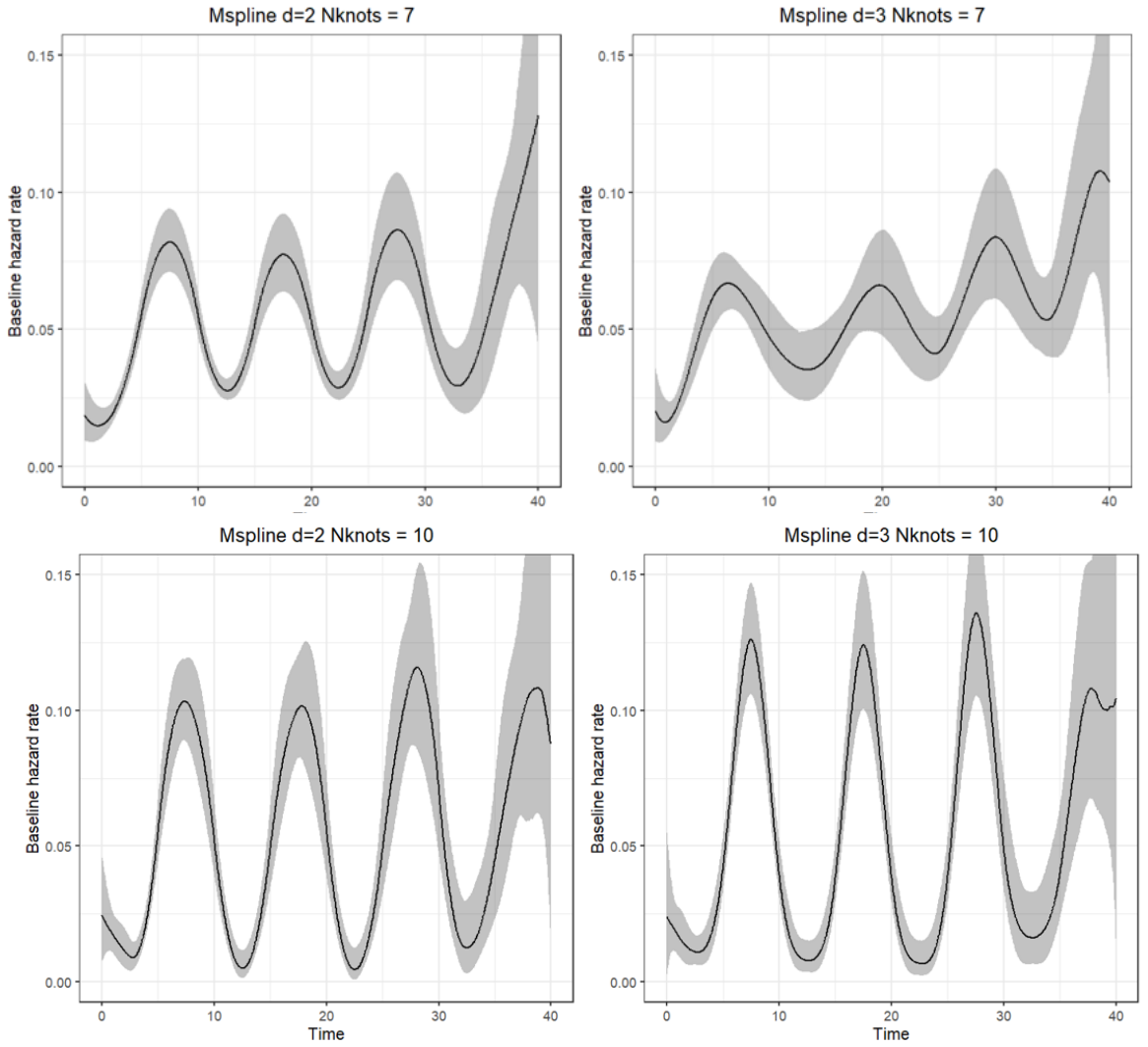
The hazard function with a M-spline basis for the baseline hazard can be written as:

$$\log h_0(t) = \sum_{j=1}^{J} \gamma_j M_j(t; \delta, k)$$



Introduction to Spline in Survival Analysis

- **Splines** provide a **smooth and continuous estimation** of the hazard function, avoiding the jumps that occur at the "knots" (boundaries) of a piecewise constant model.
- Splines reduce the risk of model misspecification due to arbitrary knot placement, offering a more **data-driven** approach to modeling hazard rates.
- Splines are better suited for **dynamic and non-linear risk patterns**, making them useful when the effect of OCS on the hazard rate fluctuates in complex ways over time.



Incorporating Covariates into the Spline-Based Hazard Model

Covariates like age, OCS dose, or comorbidities may affect the hazard rate differently at different times, and using Splines allows us to model these effects flexibly.

The hazard function $h(t)$ can be modeled flexibly using splines and linear components for covariates.

$$h_i(t) = h_0(t) \times \exp(\beta_1(t)x_{1i} + \beta_2 x_{2i} + \cdots)$$

- $h_0(t)$ is the baseline hazard function, which can be modeled with parametric distributions (e.g., exp, Weibull, Gompertz), or piecewise constant hazard function, or M-Spline.
- $\beta_1(t)x_{i1}$ is the time-varying effect of the covariates, where $\beta(t)$ can change over time. The time varying effect can be modeled using splines as well.

$$\beta_1(t) = \gamma_0 + \sum_{j=1}^{J} \gamma_j M_j(t)$$

- β_2 is the fixed effect

Simulation

Simulate survival data for 1000 individuals.

- Covariates:
  - Treatment group: a binary indicator where 50% of individuals are randomly assigned to treatment and the rest to the control group.
  - Age: simulated from a normal distribution with a mean of 60 and a standard deviation of 10.
  - Gender: a binary variable representing gender, with 50% of sample being male

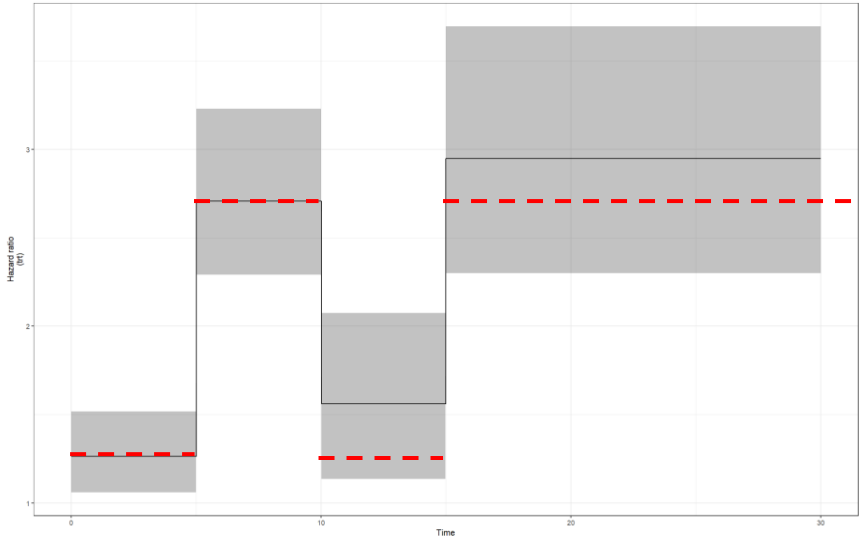
The data generating hazard function $h(t)$ given the covariates is:

$$h_i(t) = h_0(t) \times \exp(\beta_{trt}(t)trt_i + \beta_{age}age_i + \beta_{female}female_i)$$

- $h_0(t) = 0.01$
- $\beta_{trt}(t) = 1$ if $t \in [5,10]$ or $t \in [15, \infty)$, and $\beta_1(t) = 0.2$, otherwise.
- $\beta_{age} = 0.03$, $\beta_{female} = 0.15$
- The maximum follow-up is 30.

Results

	Elpd diff	Se diff
SHtve	0	0
COXPH	-33.8	8.2



- The **expected log pointwise predictive density (elpd)** measures the predictive accuracy of a model, with higher values indicating better fit.
  - TVE model is better than Coxph model by 33.8 elpd units.
  - -33.8 is more than 2 * 8.2 = 16.4, which indicates that the difference in predictive accuracy is significant different.

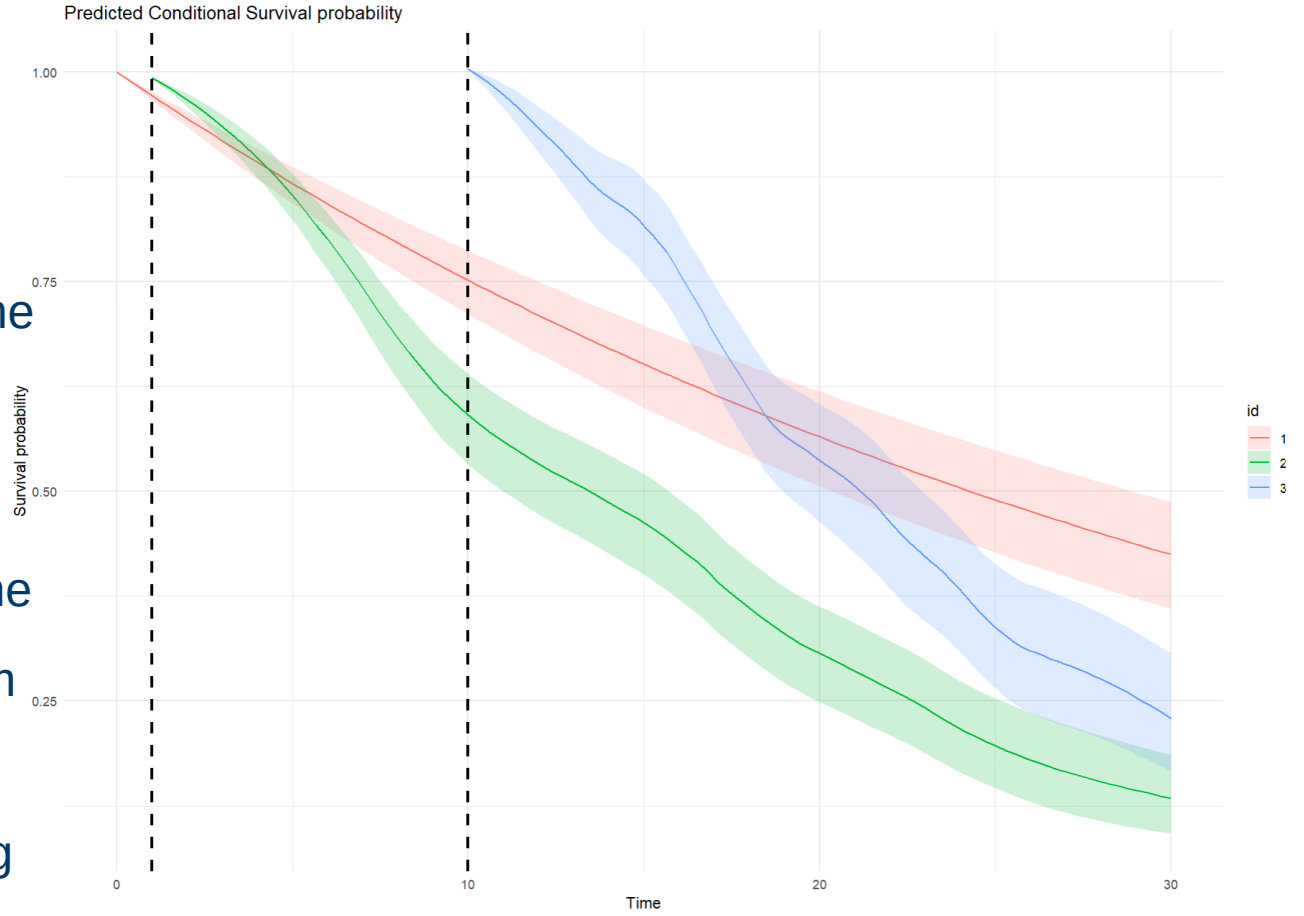
Results

Coefficient	Bias	Coverage 90%
Age	-0.00002	0.89
Gender	0.009	0.92
Trt0	-0.006	0.95
Trt1	0.005	0.93
Trt2	0.038	1
Trt3	0.019	1

- The biases are all relatively small, indicating that the model is performing well in terms of consistently estimating parameters.
- The coverage for most estimates is close to or above 0.9, indicating that the 90% credible intervals contain the true parameter values a high proportion of the time.

Prediction

- We are interested in the survival probability for patients with different duration of exposure to OCS.
- This figure illustrated the predicted conditional survival probabilities over time for three different patients.
  - Id1 is not exposure to OCS
  - Id2 received OCS for 1 unit of time
  - Id3 received OCS for 10 units of time
- Based on the predictions generated from our survival model, we can accurately estimate each patient's conditional survival probability over time, accounting for their respective exposure durations.



Summary

- In this study, we developed a **flexible Bayesian time-varying survival model** to evaluate the **impact of OCS** on patient outcomes. Utilizing real-world data and Bayesian techniques, this framework captures the **dynamic nature** of OCS effects over time and allows for a better understanding of how **cumulative OCS exposure** impacts patient survival probability.
  - The Bayesian model accommodates **time-varying effects**, reflecting the real-world changes in hazard rates associated with OCS exposure.The approach allows for better understanding of how the risk of adverse outcomes varies with the timing of OCS administration.
  - By implementing **M-splines**, we capture complex, non-linear relationships and provide a more **robust and flexible** estimation of the hazard function, compared to traditional parametric models.
  - LOOCV and WAIC(Watanabe-Akaike Information Criterion) can be used to guide model selection, e.g. knots and degree of splines.
- The proposed models can be implemented in `rstanarm`.

THANKS

Questions?