Comparing Approaches to Time-Dependent Covariates in Survival Analysis

UNIVERSITY OF TORONTO
DALLA LANA SCHOOL OF PUBLIC HEALTH

Vanessa Yuan Liao[1], Jasper Zhongyuan Zhang[1,2], Bill Xu[5], Camellia Zakaria[1, 3, 4]

[1]Dalla Lana School of Public Health, Biostatistics Division University of Toronto, Toronto, ON, Canada.
[2]Child Health Evaluative Sciences, Peter Gilgan Centre for Research and Learning, The Hospital for Sick Children, Toronto, ON, Canada.
[3]Dalla Lana School of Public Health, Institute of Health Policy, Management, and Evaluation, University of Toronto, Toronto, ON, Canada.
[4]Faculty of Information, University of Toronto, Toronto, ON, Canada.
[5]The Creative School, RTA School of Media, Toronto Metropolitan University, Toronto, ON, Canada

Background

Time Dependent Covariates (TDCs)

- Survival outcomes are often influenced by covariates that change over time
- These need to be accounted for correctly to make valid predictions on patient survival.

Extended Cox Model

- Traditional models like the extended Cox model handle TDCs naturally but rely on strict assumptions about underlying data structure (Therneau & Grambsch, 2000).

ML Methods & Landmarking

- Machine learning methods offer greater flexibility but are typically designed for static data.
- Landmarking takes snapshots of the data at given time points to create multiple "pseudo-baseline" datasets (van Houwelingen, 2007).
- This allows static ML models to be applied to time-dependent covariates by treating each landmark time as a new prediction task.

Objectives

- Evaluate how well machine learning models handle time-dependent covariates when adapted through landmarking.
- Compare the predictive performance of:
  - Extended Cox proportional hazards model
  - Random Survival Forests (RSF)
  - Gradient Boosted Survival Trees (GBST)
- Assess model performance using Harrell's C-index across varying data-generating conditions in a controlled simulation study.

Methods

Data Generating Mechanism (DGM)

We simulated longitudinal survival data for $n = 500$ individuals, each followed for up to 5 time units. Covariates included:

- Baseline: $age \sim N(60, 10)$
  $x_1 \sim N(0, 2)$
- Time-dependent:
  $x_{td1}(t) = \sin(2\pi(t + \delta_i)) + \epsilon_i$, where:
  - δ_i is the individual specific noise
  - $\epsilon_i \sim N(0, 1)$ if high noise is enabled
- Additional Z noise parameters, optional
- 3200 datasets were generated from a full factorial design (4 factors × 200 replicates).
- Event times were generated via inverse cumulative hazard sampling
- Censoring times were drawn uniformly between 2.5 and 5

Table 1. Simulation factors augmentable during data generation step.

Factor	Levels
Nonlinearity	Linear (x_1) /Nonlinear $(x_1{}^2)$
Noise in $x_{td1}(t)$	Low/High
Interaction	Absent/Present $(x_1 \cdot x_{td1}(t))$
Nuisance parameters	0/10 unrelated baseline covariates

Study Design Overview

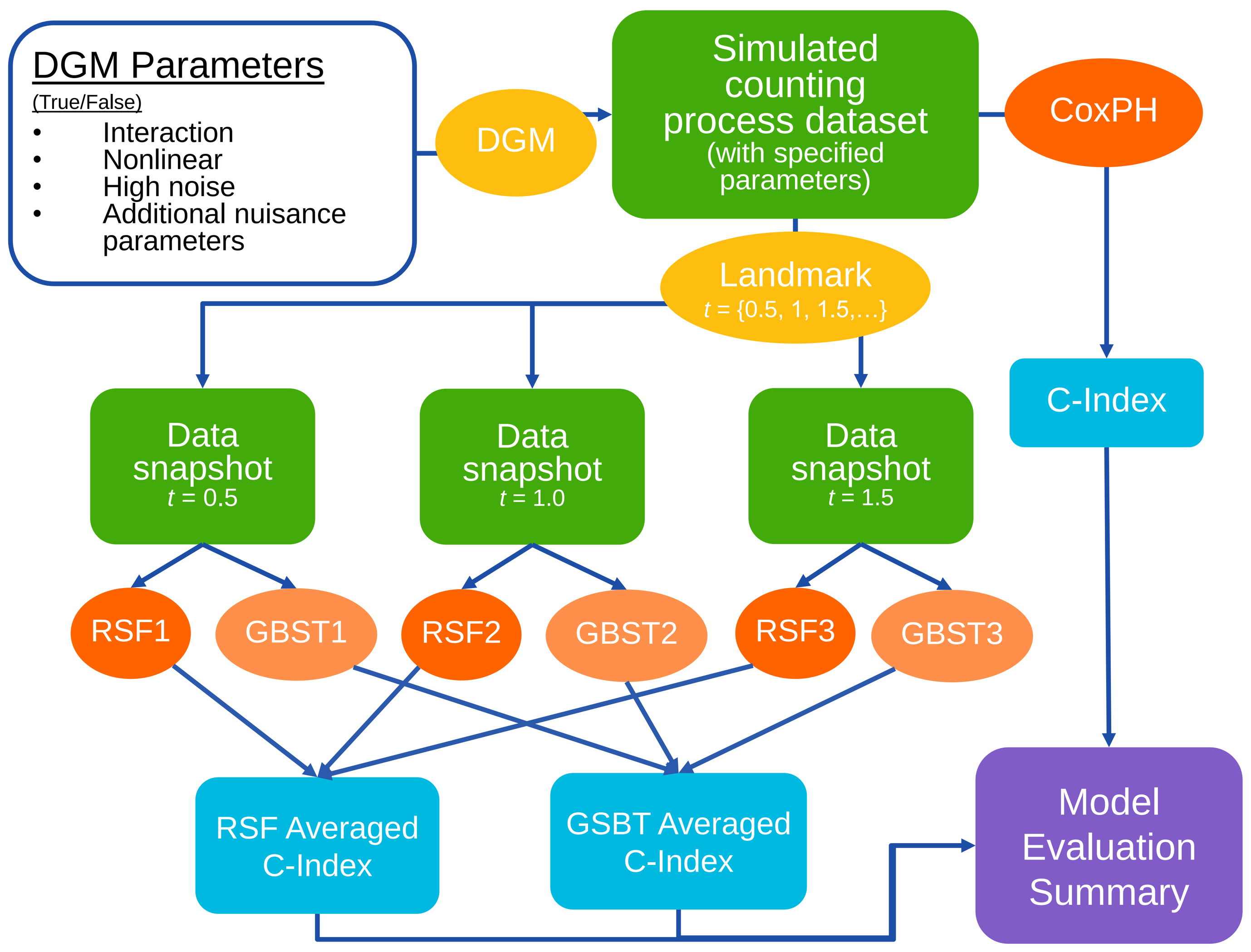

Figure 1. Flowchart illustrating the full study pipeline from data simulation to model evaluation. Note that only three landmarking times were depicted for clarity.

TDC Handling & Landmarking

- Snapshots were created at t = 0.5, 1.0, 1.5, …, 4.5 including only individuals still at risk. Prediction time was 1.0 unit.
- Separate RSF and GBST models were fit at each time point to predict survival beyond the landmark.

Models

- Extended Cox

  Fit using counting process notation (start–stop format). The hazard is modeled as:

  $$h_i(t) = h_0(t)\exp(\beta^T x_i(t))$$

- Random Survival Forests

  An ensemble of decision trees using bootstrapped samples and log-rank splitting. Captures nonlinearity and interactions automatically.

- Gradient Boosted Survival Trees

  Sequentially builds trees to minimize a loss function related to survival (C-index). Offers high predictive accuracy.

Table 2. Comparison of models used.

Model	Type	Handles TDCs	Nonlinearity & Interactions	Interpretability
CoxPH	Semi-parametric	Directly	Requires manual specification	High
RSF	Tree ensemble	Landmarking	Automatic	Moderate
GBST	Boosted trees	Landmarking	Automatic	Low

Evaluation

- Predictive performance assessed using Harrell's C-index
- For RSF and GBST, C-index was computed at each landmark
- Monte Carlo Error (MCE) was used to quantify precision across 200 replicates per setting

Results

Overall Performance

Table 3. Average C-index and Monte Carlo error (MCE) for each model

Model	Avg C-Index	MCE
CoxPH	0.798 ± 0.0631	0.0011
GBST	0.972 ± 0.0241	0.0004
RSF	0.933 ± 0.0308	0.0005

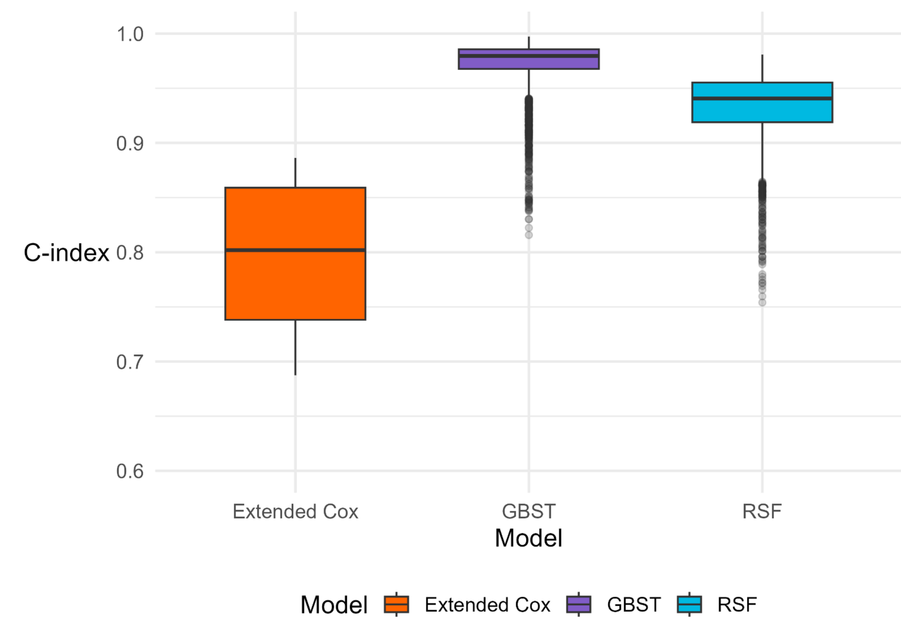

Figure 2. Distribution of Harrell's C-index across each dataset by model.

- CoxPH consistently underperformed compared to ML models across all conditions.
- ML performance declined at later landmark times, with many datasets skipped at t = 4.0 and t = 4.5 due to insufficient follow-up.

Performance Under Varying Data Conditions

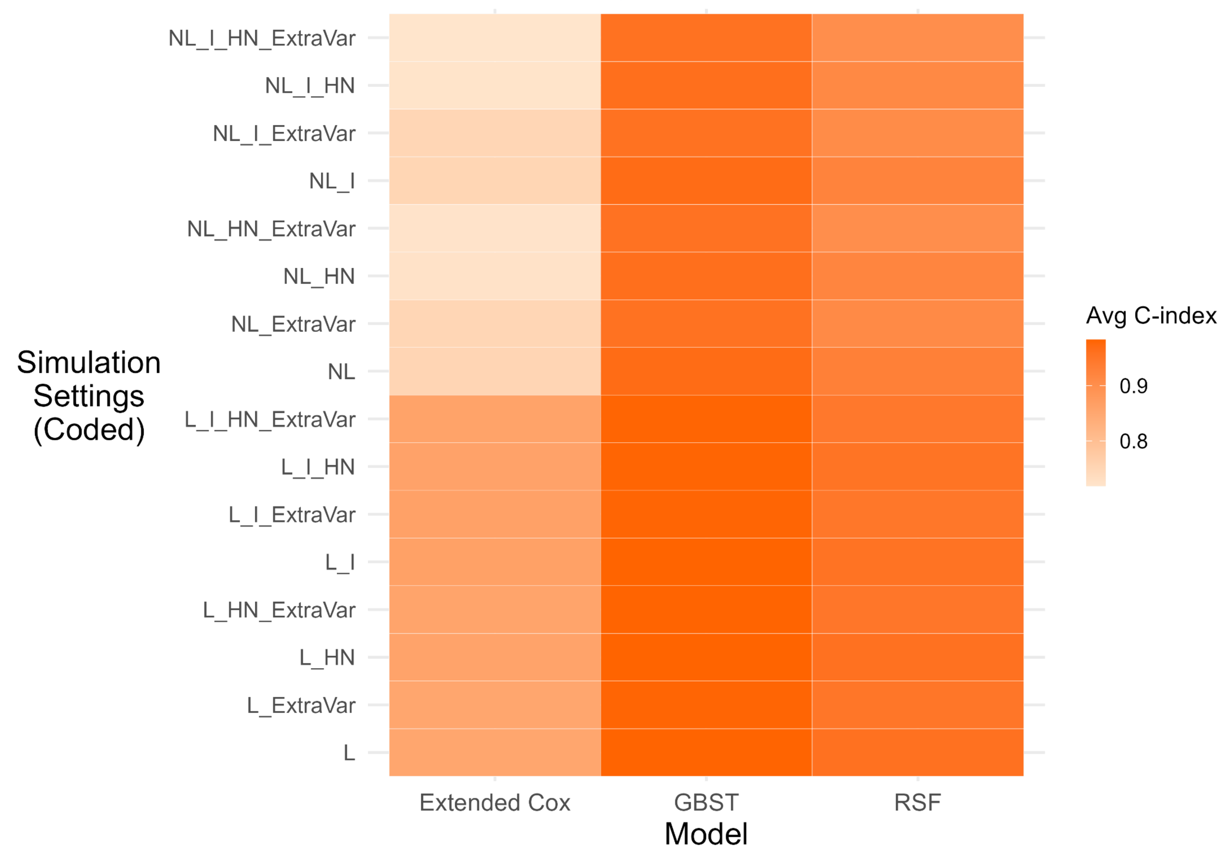

Figure 3. Average Harrell's C-index for each model across 16 factorial simulation settings, varying in nonlinearity (NL), interaction (I), noise (HN), and presence of extra covariates..

- Model performance remained stable for RSF and GBST, regardless of added complexity.
- CoxPH performance degraded in scenarios with nonlinearity and high noise.

Conclusion

- **CoxPH** is interpretable but degrades with nonlinearity and noise.
- **RSF** is robust and stable across all conditions with low MCE.
- **GBST** offers high accuracy but at higher computational cost.
- **Landmarking** enables flexible use of ML models with TDCs, but care must be taken at later timepoints due to sample attrition.

References

Ishwaran, H., Kogalur, U. B., Blackstone, E. H., & Lauer, M. S. (2008). Random survival forests. *The Annals of Applied Statistics*, *2*(3). https://doi.org/10.1214/08-aoas169

Morris, T. P., White, I. R., & Crowther, M. J. (2019). Using simulation studies to evaluate statistical methods. *Statistics in Medicine*, *38*(11), 2074–2102. https://doi.org/10.1002/sim.8086

Tanner, K. T., Sharples, L. D., Daniel, R. M., & Keogh, R. H. (2020). Dynamic survival prediction combining landmarking with a machine learning ensemble: Methodology and empirical comparison. *Journal of the Royal Statistical Society Series A: Statistics in Society*, *184*(1), 3–30. https://doi.org/10.1111/rssa.12611

Therneau, T. M., & Grambsch, P. M. (2013). Multiple Events per Subject. In *Modeling Survival Data: Extending the Cox Model* (pp. 169–229). essay, Springer.

Van Houwelingen, H. C. (2006). Dynamic prediction by landmarking in event history analysis. *Scandinavian Journal of Statistics*, *34*(1), 70–85. https://doi.org/10.1111/j.1467- 9469.2006.00529.x