Longitudinal Survival Analysis & Risk Stratification Using Real-World Clinical Data (Glioblastoma)

Positioning
Core expertise: longitudinal survival analysis, time-dependent modeling, and risk stratification.
Data modality: incidental (imaging-derived features from a public dataset).

Client Scenario

A clinical research team has access to longitudinal follow-up data from patients with glioblastoma but lacks a robust framework to transform this time-varying information into **survival-ready datasets and interpretable risk models** suitable for clinical research or trial analytics.

The goal was not methodological novelty, but generating survival outputs directly usable for clinical research workflows.

Problem

This leads to loss of time-dependent information, incorrect handling of censoring and follow-up intervals, and survival models that are difficult to interpret or deploy.

The core challenge is not the data type, but how to correctly model survival in the presence of longitudinal measurements.

Solution

I implemented a **generalizable longitudinal survival-analysis pipeline** that converts real-world follow-up data into a survival-ready analytic framework.

Key methodological components:

- Interval-based (start–stop) survival data construction
- Proper handling of censoring and time-varying covariates
- Feature reduction for model stability and interpretability
- Parsimonious Cox proportional hazards modeling
- Internal validation and clinical utility assessment

This workflow is **data-agnostic** and applicable to longitudinal omics, imaging-derived, or clinical variables.

Deliverables

1. Longitudinal survival dataset with validated time intervals
2. Compact multivariable Cox model (low-complexity, interpretable)
3. Continuous patient-level risk score
4. Time-dependent discrimination metrics (C-index over follow-up)
5. Calibration assessment at clinically relevant horizons
6. Decision-curve analysis to evaluate downstream research utility

All outputs are reproducible and can be adapted to client-specific longitudinal datasets.

Key Results

- **Model simplicity:** minimal number of predictors + age
- **Discrimination:**
  - Harrell's C-index ≈ **0.77**
  - Stable time-dependent performance across follow-up
- **Risk stratification:**
  - Clear separation of low-, intermediate-, and high-risk groups
- **Clinical relevance:**
  - Demonstrated net benefit across practical risk thresholds

Why This Matters for CROs and Clinical Research Teams

This case study demonstrates the ability to:

- Correctly model **longitudinal survival data**, independent of data modality
- Translate complex follow-up measurements into interpretable risk scores
- Support retrospective analyses, feasibility studies, and trial stratification
- Produce survival outputs that are methodologically sound and audit-ready

What This Case Study Represents

✓ Advanced longitudinal survival analysis
✓ Time-dependent modeling and validation
✓ Interpretable, low-risk statistical framework
☒ Not an imaging-focused project or black-box AI model

Typical Use Cases

Typical use cases include retrospective survival analyses with longitudinal data, trial feasibility and patient stratification studies, evaluation of time-varying biomarkers, and method development prior to external validation.

Data Source

Public longitudinal glioblastoma follow-up dataset (LUMIERE cohort)

One-sentence takeaway

This case study demonstrates my ability to transform complex longitudinal biomedical data into robust, interpretable survival models suitable for clinical research and CRO-driven studies. This workflow can be directly adapted to other longitudinal clinical, omics, or real-world datasets.