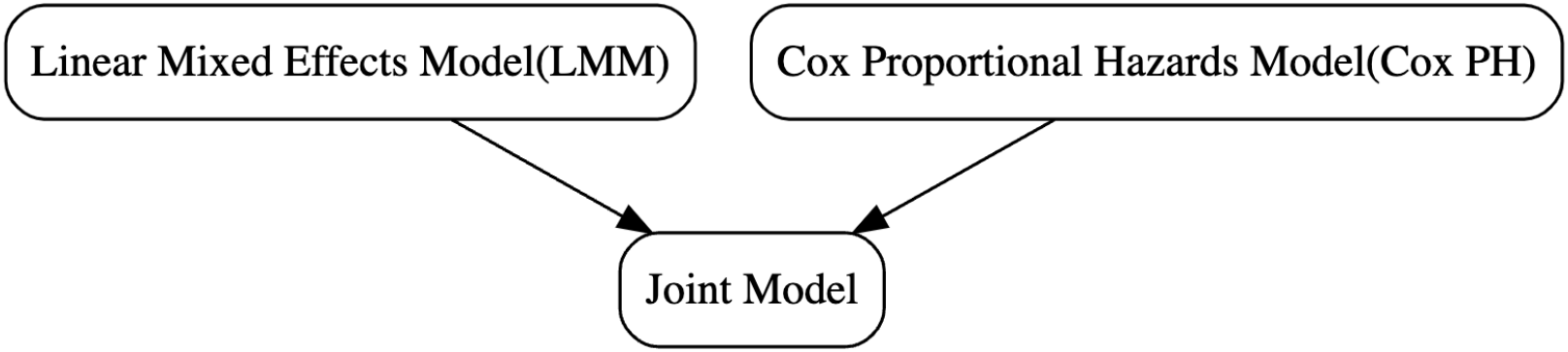
Dynamic Prediction using Simultaneous Modeling of Repeated Measures and Time-to-event data

Rebecca Quaijah

2024-11-19

Joint Modelling Overview

- Studies collect **time-to-event**(Survival time) data and **repeated measurements**, often analyzed separately.



Applications

- **Medical Studies**: Tracking CD4 counts (HIV) or tumor size (cancer) to see their impact on survival.
- **Pharmaceutical Trials**: Predicting patient response to treatment based on biomarker changes.
- **Public Health**:Monitoring creatinine to predict kidney failure or need for dialysis.

Mathematical Formulation

The joint model links repeated measures to time-to-event risk:

$$Y_{ij} = Y_i^*(t_{ij}) + Z_{ij} = X_{ij}\beta + W_{ij}B_i + Z_{ij} \qquad \textbf{(LMM)}$$

$$\lambda_i(t \mid K_i,\ Y_i^*(t)) = \lambda_0(t)\exp\left(K_i\gamma_1 + Y_i^*(t)\gamma_2\right) \qquad \textbf{(Cox PH)}$$

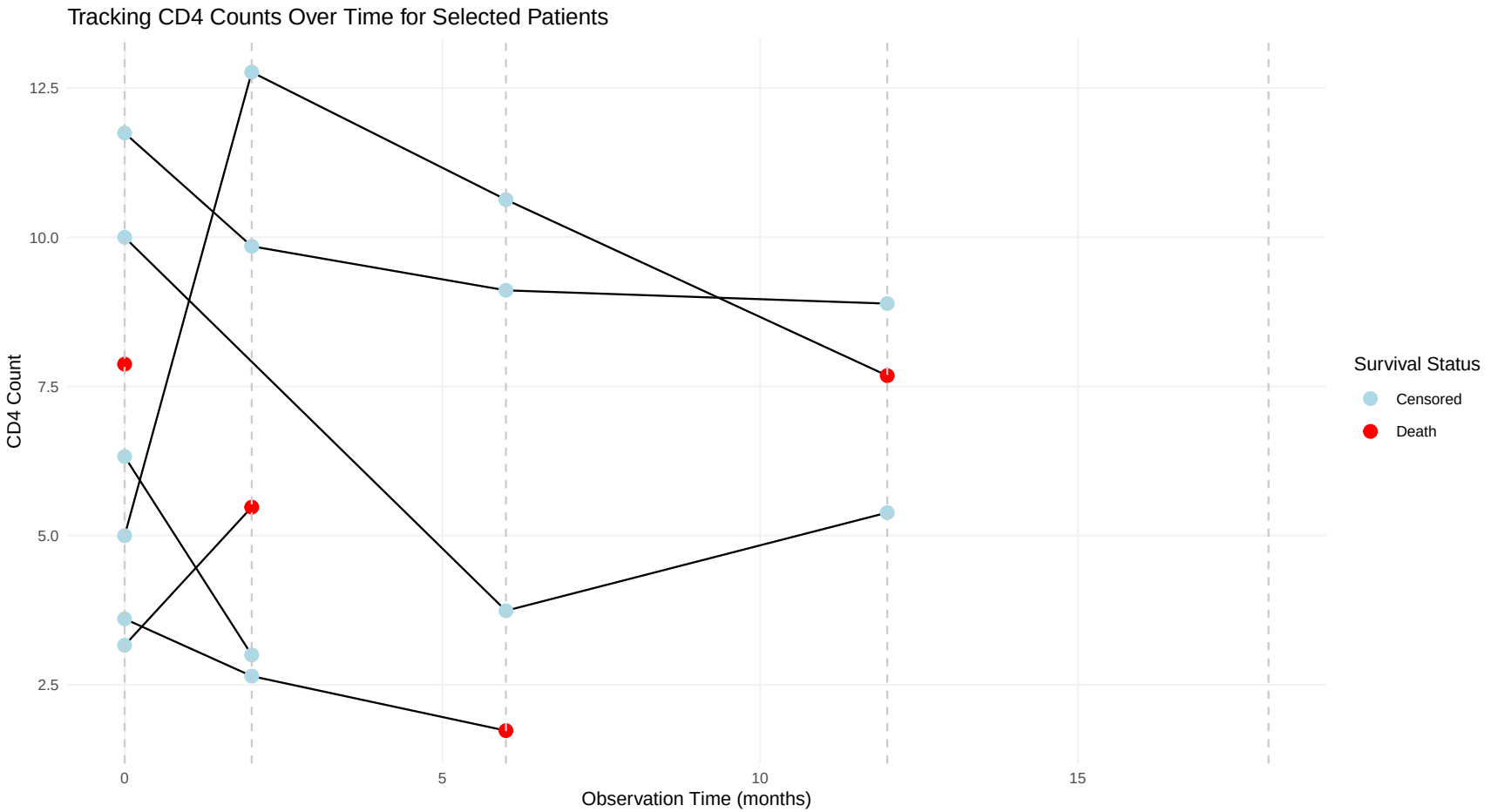
Difference between AIDS and PBC2 data

Feature	AIDS Dataset	PBC2 Dataset
Disease	HIV/AIDS	Primary biliary cirrhosis (PBC)
Longitudinal Biomarker	CD4 counts	Serum bilirubin
Number of Patients	467	466
Treatment Groups	Randomized to didanosine (ddI) or zalcitabine (ddC)	Randomized to D-penicillamine or placebo
Follow-Up Intervals	0, 2, 6, 12, 18 months	Varying time interval
Study Design	Randomized controlled trial focusing on alternative treatments	Longitudinal observational study for disease progression
Modeling Focus	Drug efficacy and survival based on CD4 levels	Disease progression and survival based on serum biomarkers
Outcome of Interest	Time to treatment failure or death	Time to death or transplant
Model	Survival	Competing Risk

Example: AIDS Data

- A longitudinal study was conducted on 467 HIV-infected patients who had previously failed or were intolerant to zidovudine therapy.
- The study compared the efficacy and safety of two antiretroviral drugs: didanosine (ddI) and zalcitabine (ddC).
- Patients were randomly assigned to either drug, with CD4 counts measured at baseline, and at 2, 6, 12, and 18 months thereafter.

Visual Representation



Joint Model: Step 1

```
fitLME <- lme(sqrt(CD4) ~ obstime + drug + obstime:drug,
              random = ~ obstime | patient, data = aids)
```

- Fits a **linear mixed-effects (LME)** model to the longitudinal data (i.e., CD4 counts).
- **Response**: sqrt(CD4) - Square root of CD4 count.
- **Fixed Effects**: obstime (time) and obstime:drug (time by treatment interaction).
- **Random Effects**: Patient-specific intercept and slope ($\sim$ obstime | patient).

Step 2

```
fitSURV <- coxph(Surv(Time, death) ~ drug,
                 data = aids.id, x = TRUE)
```

- Fits a **Cox proportional hazards model** to the survival data.
- **Formula: Surv(Time, death) ~ drug** - Models survival time based on drug treatment.
- **x = TRUE** - Saves model matrix for later use in joint modeling.

Step 3

```
fit.JM <- jointModel(fitLME, fitSURV,
    timeVar = "obstime", method = "piecewise-PH-GH")
```

- **fitLME**: Longitudinal model (Step 1).
- **fitSURV**: Survival model (Step 2).
- **timeVar** = "obstime": Time variable for longitudinal measurements.
- **method = "piecewise-PH-GH"**: Uses a piecewise proportional hazards approach with **Gauss-Hermite quadrature**(e.g combine patient data over time with survival outcomes to make accurate predictions about health risks.)

Estimated Longitudinal Submodel

Term	Estimate	Std.Err	z_value	p_value
(Intercept)	2.5416	0.0514	49.4825	<0.0001
obstime	-0.0419	0.0046	-9.0361	<0.0001
drugddl	0.0320	0.0751	0.4258	0.6703
obstime:drugddl	0.0047	0.0066	0.7092	0.4782

Estimated Survival Submodel

Value	Estimate	Std..Err	z.value	p.value
drugddI	0.3511	0.1537	2.2839	0.0224
Assoct	-1.1016	0.1180	-9.3388	$<$0.0001
log(xi.1)	-1.6489	0.2498	-6.6000	
log(xi.2)	-1.3393	0.2394	-5.5940	
log(xi.3)	-1.0231	0.2861	-3.5758	
log(xi.4)	-1.5802	0.3736	-4.2299	
log(xi.5)	-1.4722	0.3500	-4.2069	
log(xi.6)	-1.4383	0.4283	-3.3584	
log(xi.7)	-1.4780	0.5455	-2.7094	

Dynamic Prediction

- Dynamic predictions are continuously updated risk estimates that incorporate the most recent patient data over time.
- Survival probabilities can be updated each time a new measurement is available for the longitudinal outcome.

Advantages of Dynamic Predictions

- It gives a personalized health check based on each patient's own history.
- Enhances clinical decision-making by leveraging up-to-date patient information.
- Useful for monitoring disease progression and adapting treatment plans.

Prediction in Joint Models

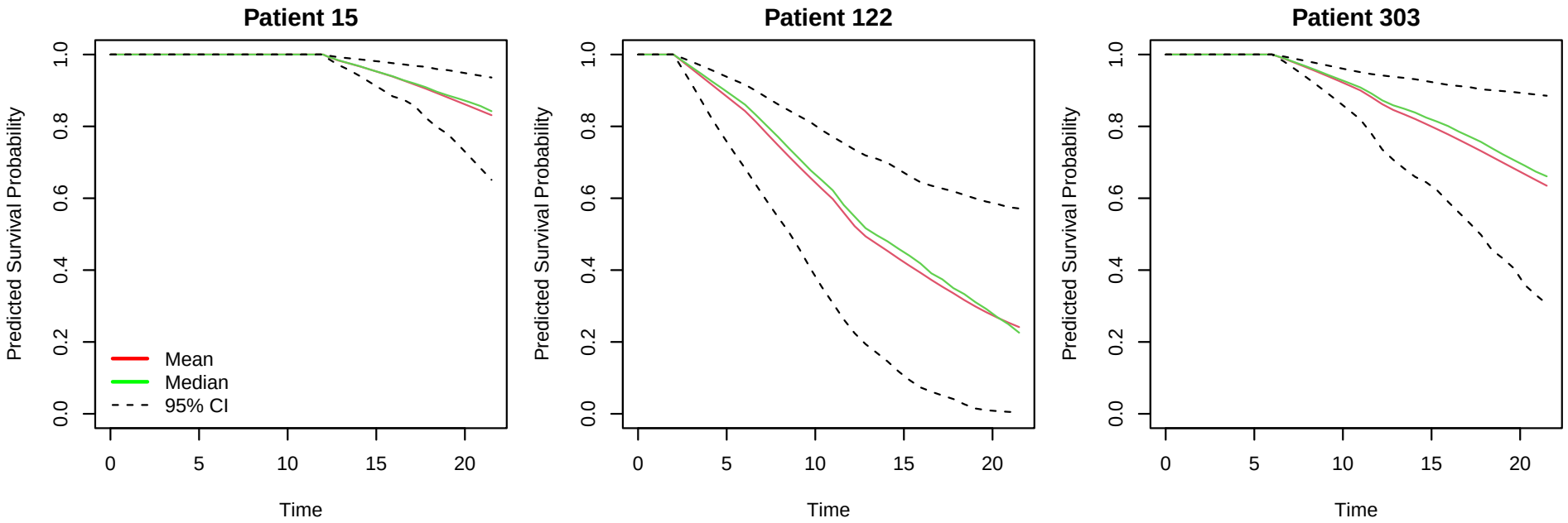
- The function **survfit** computes the conditional probability of surviving later times than the last observed time for which a longitudinal measurement was available.
- **object**: The fitted joint model used for generating survival predictions.
- **newdata**: Data frame with new patient data to dynamically update survival probabilities.
- **last.time**: The latest observation time for adjusting predictions based on recent data.
- **simulate**: Boolean indicating whether to use Monte Carlo simulations for accounting for uncertainty.
- **M**: Number of simulations to improve prediction accuracy when simulate = TRUE.

Evaluating Predictions

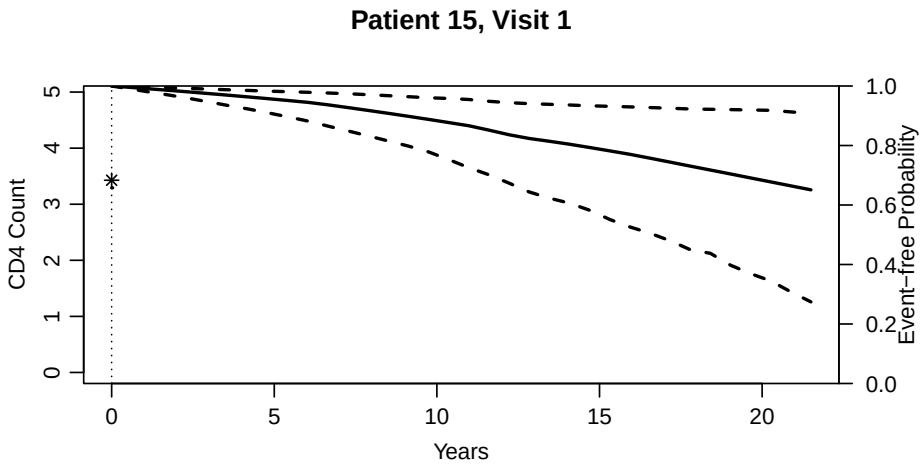
- **Time dependent AUC**: Measures the joint model's ability to distinguish between patients who will experience the event **before a specific time t**(e.g., death, relapse) and those who will survive beyond that time.
  - Higher values (close to 1) is equal to better discrimination
- **Brier Score**: It measures how close the predicted probabilities are to the actual binary outcomes (i.e., whether an event happens or not)..
  - Lower values is equal to more accurate predictions
- **Bootstrap Validation** A resampling technique to test the robustness of the model's predictions for survival outcomes.
  - Identifies overfitting by comparing model performance on resampled data vs. the original dataset.

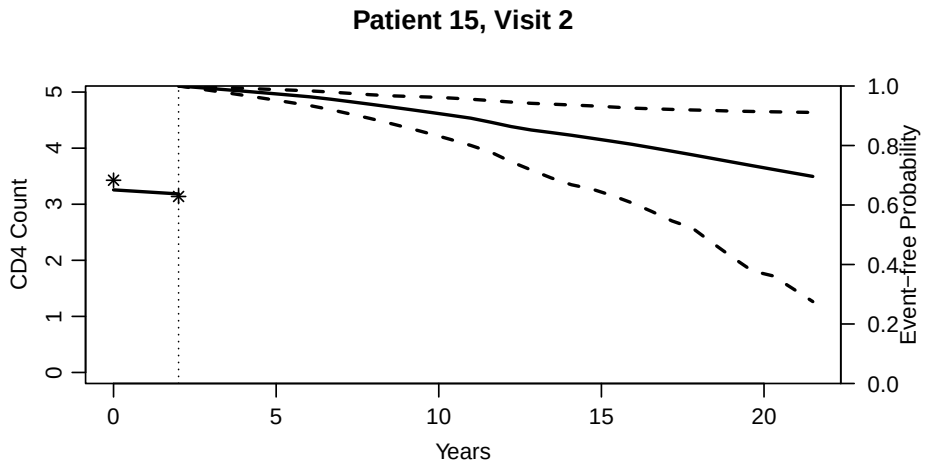
Predicted Probabilities of Survival for Three Patients

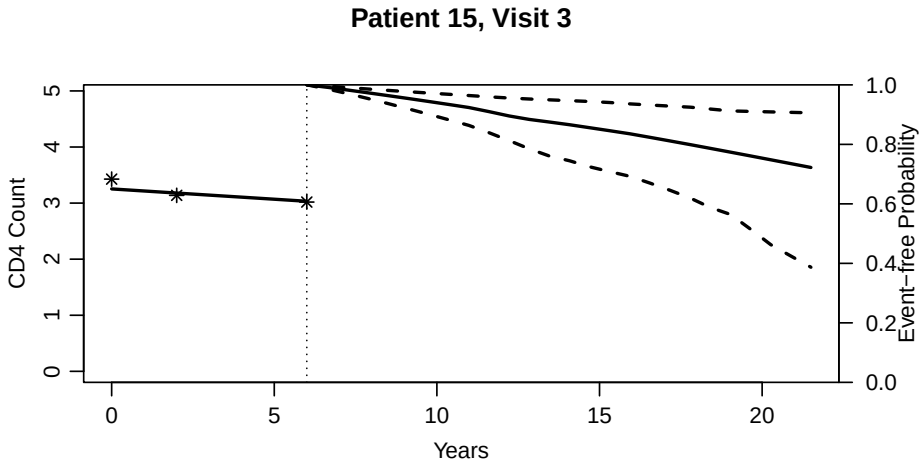
```
predSurv <- JM::survfitJM(fit.JM, newdata = ND,
idVar = "patient", timeVar = "obstime", last.time = "Time")
```

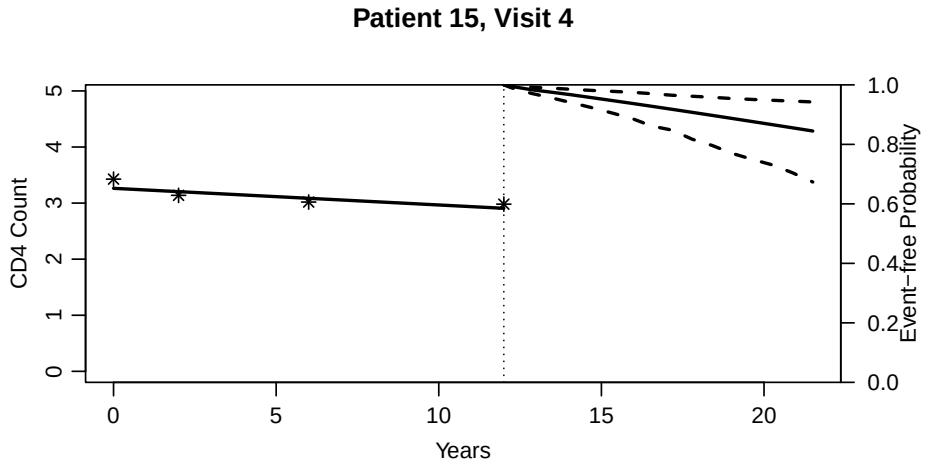


Predicted Survival Probability for Patient 15





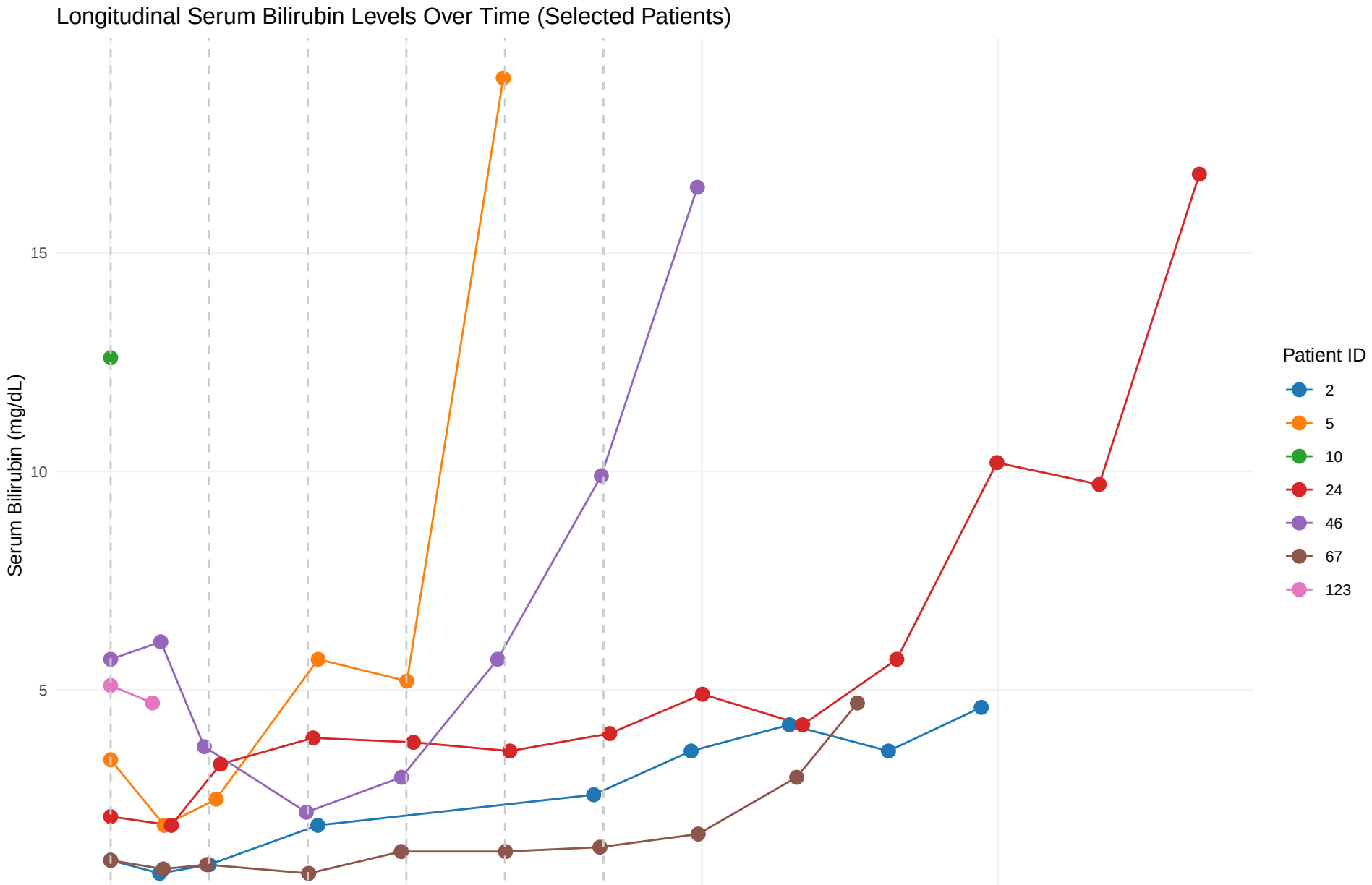




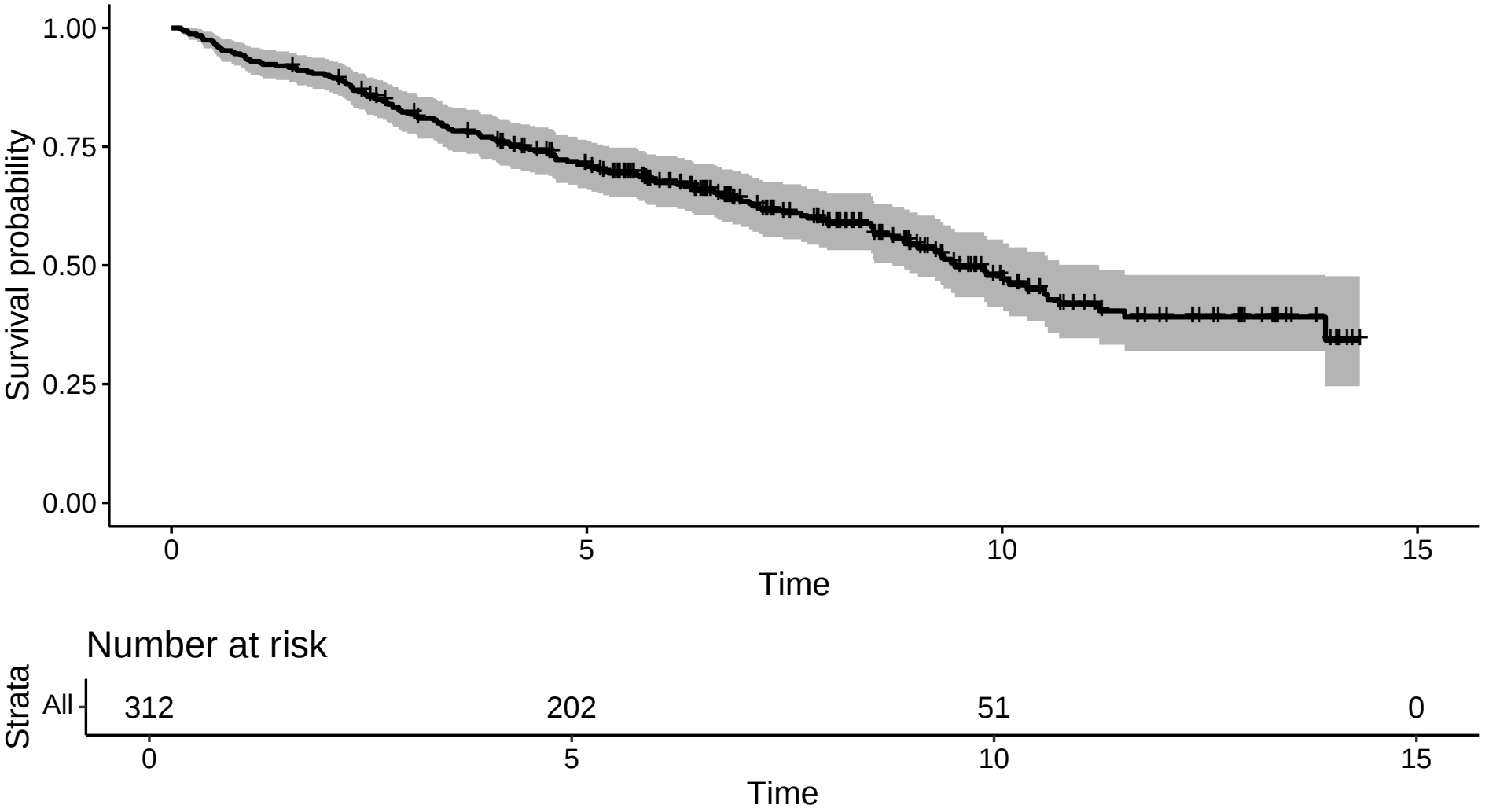
Example Data:PBC2

- The PBC2 dataset includes data on 466 patients with primary biliary cirrhosis collected over a decade, focusing on serum bilirubin levels as a key biomarker.
- Patients were randomized to receive either D-penicillamine (312 patients) or a placebo (154 patients), with a total of 1,945 measurements collected.

Visual Representation for Longitudinal Data



Visualizing the Survival Curve



Mixed Effects Model

- **Linearity Check**: Determine if the data follows a linear or non-linear pattern.
  - **Linear Models**:
    - Assume a straight-line relationship over time.
    - Use linear mixed-effects models to analyze simpler data patterns.
    - The **JM package** is best for these cases, with added flexibility using splines or polynomial terms.
  - **Non-Linear Models**:
    - Capture curved or complex relationships that change over time.
    - Use non-linear mixed-effects models for a better fit with complex data.
    - The **JMbayes** package is ideal, leveraging a Bayesian approach for improved accuracy and handling uncertainty.

Natural Splines

- **Natural splines** model curved (non-linear) patterns while becoming linear at the boundaries to prevent overfitting and ensure stable predictions.
- The **degrees of freedom (df)** control how flexible the spline curve is.
  - **Higher df** means more curves and flexibility, which fits the data more closely.
  - **Lower df** means fewer curves, making the fit simpler and smoother.
- **R Implementation**- ns(x, df)
  - **x**; the predictor variable
  - **df**: degrees of freedom

Data Analysis of pbc2 data

```
fit_lm <-lme(log(serBilir) ~ ns(year, 3) + age + sex + drug,
             random = list(id = pdDiag(form = ~ ns(year, 3))),
             data = pbc2)
```

- **Response Variable**:Log-transformed serum bilirubin (log(serBilir)) stabilizes variance and normalizes the data.
- **Fixed Effects**:Time modeled with a spline (ns(year, 3)) for non-linear trends. Includes patient age, sex, and drug treatment effects.
- **Random Effects**:Patient-specific variations (id) with unique intercepts and slopes. Assumes uncorrelated random effects using a diagonal structure (pdDiag).

Survival data

```
fit_surv <- coxph(Surv(years, status2) ~ age +
           sex + drug, data = pbc2.id, x = TRUE)
```

- **Survival Object**:
  - **years**: Time to event or censoring.
  - **status2**: 1 if event occurred, 0 if censored.
- **Covariates**: age, sex, and drug (treatment group).
- **x = TRUE**: Retains design matrix for further analysis.

Fitting pbc2 data:JMbayes

```
library(JMbayes)
jointFit.pbc <- jointModelBayes(fit_lm, fit_surv,
        timeVar = "year", baseHaz = "regression-splines",
          verbose = FALSE)
```

- **fit_lm**: The fitted linear mixed-effects model (for longitudinal data).
- **fit_surv** : The fitted Cox proportional hazards model (for survival data).
- **timeVar("year")**: Specifies that the variable year is used to indicate time points for the longitudinal measurements.
- **baseHaz** :"regression-splines": Uses regression splines to model the baseline hazard function to allow for more flexibility in capturing non-linear effects.
- **verbose(FALSE)** ; Suppresses output details during model fitting for a cleaner output.

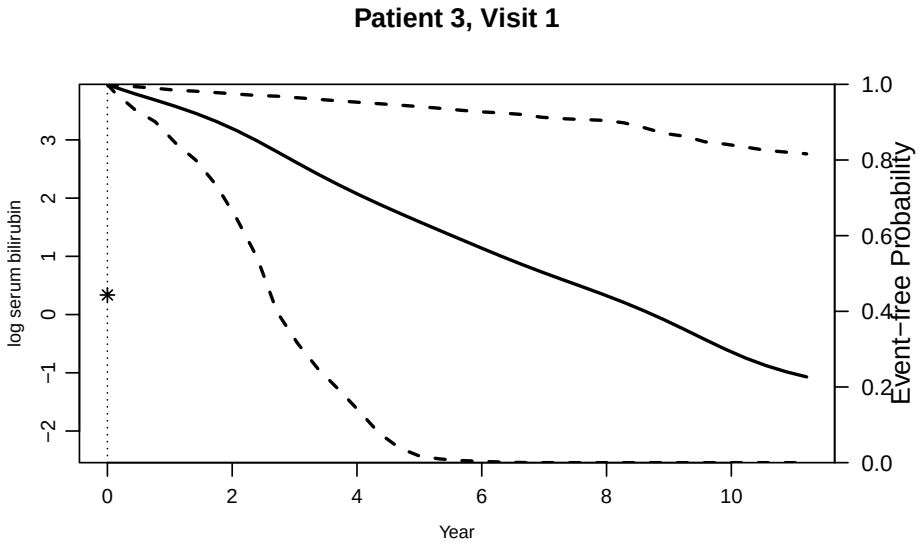
Estimated Longitudinal Submodel

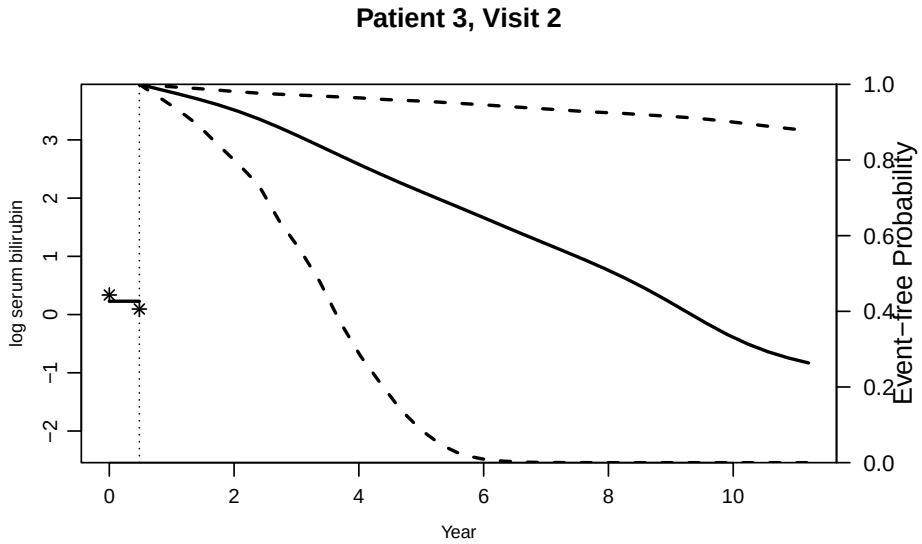
Term	Value	Std.Err	Std.Dev	2.5%	97.5%	Pvalue
(Intercept)	0.6811	0.0089	0.3520	0.0079	1.3601	0.049
ns(year, 3)1	1.0853	0.0057	0.1042	0.8799	1.2912	<0.001
ns(year, 3)2	1.3884	0.0090	0.1075	1.1935	1.6053	<0.001
ns(year, 3)3	1.1345	0.0144	0.1427	0.8532	1.4092	<0.001
age	0.0009	0.0001	0.0055	-0.0098	0.0113	0.852
sexfemale	-0.1480	0.0043	0.1777	-0.4995	0.2044	0.396
drugD-penicil	-0.1289	0.0030	0.1146	-0.3403	0.1039	0.266

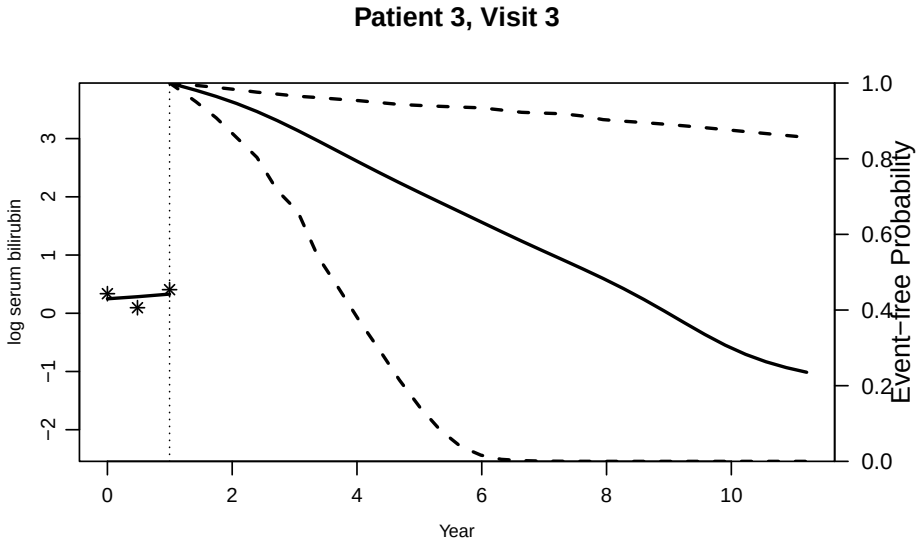
Estimated Survival Submodel

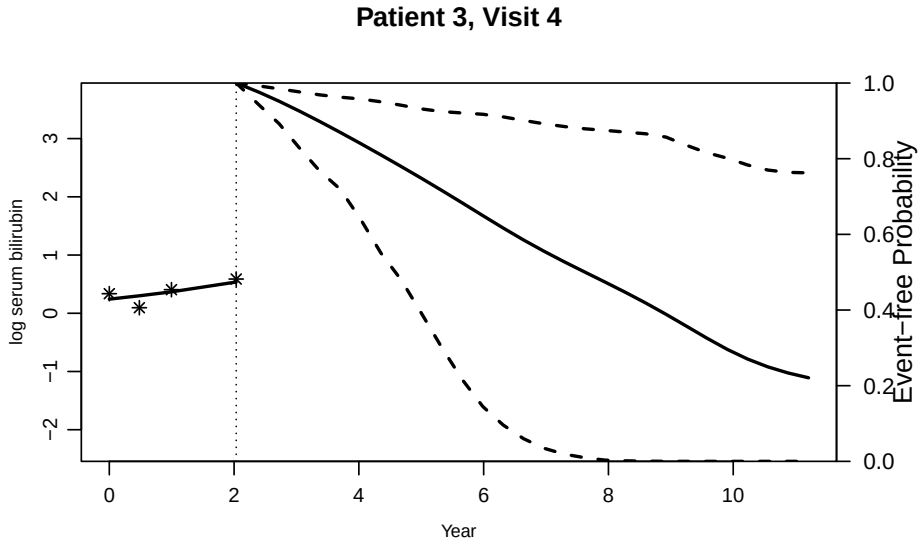
Term	Value	Std.Err	Std.Dev	2.5%	97.5%	Pvalue
age	0.0422	0.0008	0.0064	0.0307	0.0549	$<$0.001
sexfemale	-0.2518	0.0112	0.2078	-0.6375	0.1571	0.249
drugD-penicil	-0.1251	0.0090	0.1712	-0.4760	0.1927	0.484
Assoct	1.2846	0.0056	0.0930	1.1145	1.4684	$<$0.001

Dynamic Predictions:Patient 3

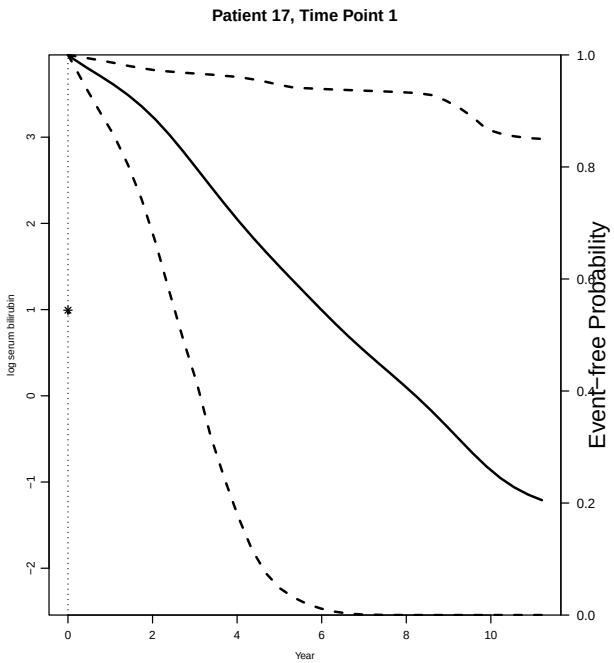


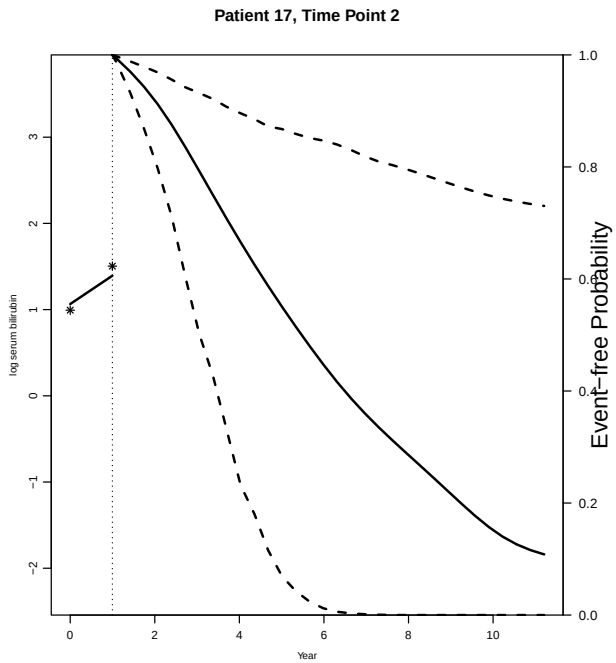


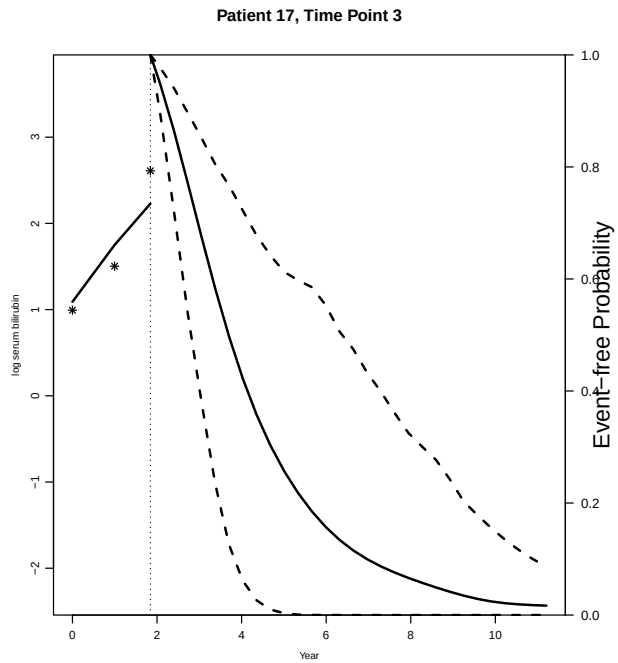




Patient 17


Patient 17, Time Point 1
log serum bilirubin
Year
Event−free Probability


Patient 17, Time Point 2
log serum bilirubin
Year
Event−free Probability


Patient 17, Time Point 3
log serum bilirubin
Year
Event−free Probability

Evaluation of Predictions: Time Dependent AUC

Description	Value
Estimated AUC	0.8372
Time Horizon (Thoriz)	7.0000
Using Information Up To (Tstart)	3.0000
Subjects Still at Risk	245.0000

Brier Score for the Joint Model

Description	Value
Brier Score	0.1243
Time Horizon (Thoriz)	7
Using Information Up To (Tstart)	3
Subjects Still at Risk	245
Loss Function	Square

Acknowledgment

- **Supervisors**: Dr. Jingjing Wu and Dr. Chel Hee Lee
- **Departmental Support**
  - Department of Mathematics and Statistics
  - Department of Critical Care Medicine

Questions

Rebecca Quaijah

Email: rebecca.quaijah@ucalgary.ca

References

- Rizopoulos, D. (2014). The R Package JMbayes for Fitting Joint Models for Longitudinal and Time-to-Event Data Using MCMC. Journal of Statistical Software, 72(7), 1-46.
- Rizopoulos, D. (2012). Joint Models for Longitudinal and Time-to-Event Data: With Applications in R. Chapman & Hall/CRC Biostatistics Series. ISBN: 978-1-4398-7884-6.
- Rizopoulos, D. (2010). JM: An R Package for the Joint Modelling of Longitudinal and Time-to-Event Data. Journal of Statistical Software, 35(9), 1-33.