Joint Modelling For Longitudinal and Survival Data

Why R? Turkey 2021

Ipek Guler

[1]Leuven Biostatistics and Statistical Bioinformatics Centre (L-BioStat)
[2]Center for Molecular and Vascular Biology, Department of Cardiovascular Sciences, KU Leuven, Leuven, Belgium

April 21, 2021

Outline

Why Joint Modelling?

To study jointly the following models:

- Longitudinal models of repeated measurements, affected by informative censoring/drop-out
- Incorporation of these longitudinal covariates in to the survival process

Table of Contents

Statistical Background

- To model longitudinal and survival processes together, there exist various methods in the literature. Such as: Extended-Cox Models, Two-stage approach, Joint Modelling Approaches.
- However, existing approaches usually deal with one single longitunal biomarker with survival process.
- We'll focus on two different approaches:
  - Two-stage (TS) approach
  - Joint Modelling (JM) Approach

Main idea behind the naive TS approach

Prentice, 1982; Self and Pawitan, 1992; Tsiatis et al., 1995; Bycott and Taylor, 1998; Dafni and Tsiatis, 1998; Albert and Shih, 2010; Rizopoulos, 2010

- **Stage 1**: Fit a **Linear Mixed Model** model to the longitudinal covariate data.
- **Stage 2**: Fit the **Survival Model** separately, with the random effects predictions or true/unobserved value of the covariate substituted by their estimates/predictions from the first stage.

NOTE: Separate likelihood calculation.

Joint Modelling (JM) Approach

Main idea: To link the longitudinal and survival process in the joint likelihood calculation

- **Y: Longitudinal process affected by informative censoring**
- **S: Survival process with time-varying covariates**
- **W: Latent random effect**

NOTE: Joint likelihood calculation.

Different JM approaches

There are three different ways to calculate the joint likelihood, among others;

- Selection Models

$$[\mathbf{Y},\mathbf{S},\mathbf{W}] = [\mathbf{W}][\mathbf{Y}|\mathbf{W}][\mathbf{S}|\mathbf{Y}]$$

Different JM approaches

There are three different ways to calculate the joint likelihood, among others;

- Selection Models

$$\mathbf{[Y,S,W] = [W][Y|W][S|Y]}$$

- Pattern-Mixture Models

$$\mathbf{[Y,S,W] = [W][S|W][Y|S]}$$

Different JM approaches

There are three different ways to calculate the joint likelihood, among others;

- Selection Models

$$[Y,S,W] = [W][Y|W][S|Y]$$

- Pattern-Mixture Models

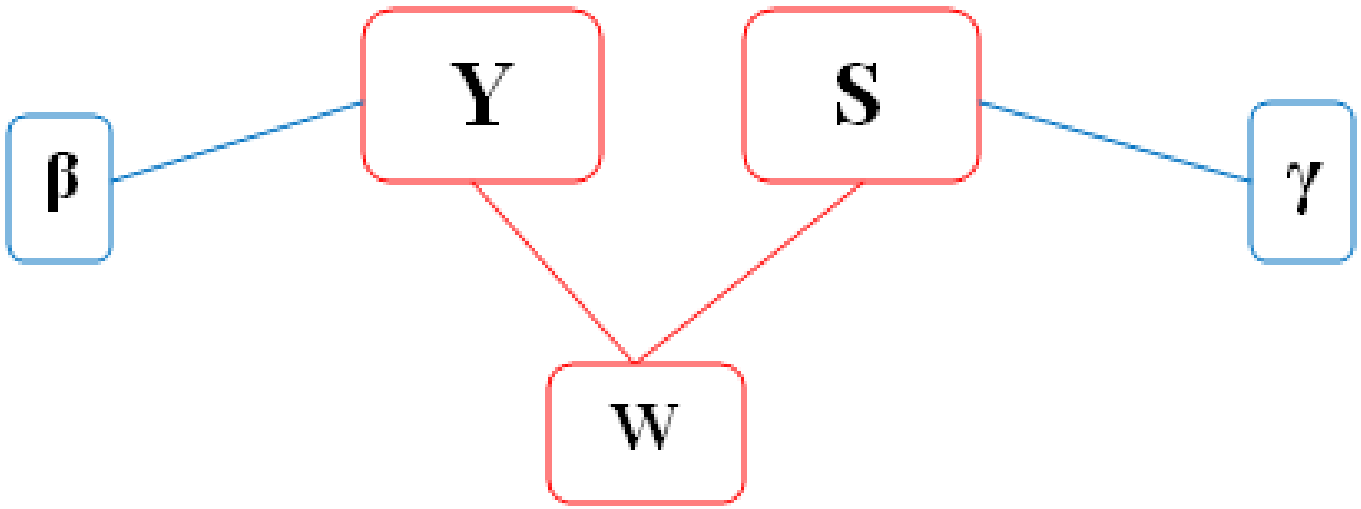
$$[Y,S,W] = [W][S|W][Y|S]$$

- **Shared Random Effect Models**

$$[Y,S,W] = [W][Y|W][S|W]$$

Shared Random Effect Models

Schluchter, 1992; Faucett and Thomas, 1996; Wulfsohn and Tsiatis, 1997; and Henderson et al., 2000



β and γ are the fix parameters of longitudinal and survival processes respectively.

Shared Random Effect Models

Longitudinal sub-model: Linear mixed effects model (Laird and Ware, 1982) *library(nlme)*

$$Y_{ij} = x_{ij}\beta + W_{1i}(t) + \epsilon_i$$

$$W_{1i}(t) = U_{0i} + U_{1i}t_ij$$

where U_{0i} and U_{1i} are individual random intercept and random slope respectively.

Shared Random Effect Models

Survival sub-model: Cox Proportional Hazard Regression, *library(survival)*

$$\lambda(t) = \lambda_0(t) \exp\left(\gamma X_i(t) + \alpha W_1 i(t)\right)$$

- α is termed the association parameter; in this case the association is based on the linear combination of the random intercept and slope predictions.

Different parameterizations

- The random effects predictions at time t (Wulfsohn1997)

$$\lambda_i(t) = \lambda_0(t)\exp(\gamma X_i + \alpha(U_{0i} + U_{1i}t_i))$$

in which α is the association between the longitudinal biomarker and the risk of death at time t with a unit change in the marker corresponding to a $\exp(\alpha)$ fold change in the risk of death.

Different parameterizations

- The true unobserved (current) value at time t (Rizopoulos2012)

$$\lambda_i(t) = \lambda_0(t)\exp(\gamma X_i + \alpha(\beta_0 + \beta_1 t_{ij} + U_{0i} + U_{1i} t_i))$$

in which α represent the association between the longitudinal biomarker and the risk of death at time t taking into account the true value of the longitudinal biomarker both with fixed and random effects predictions.

Different parameterizations

- Time-dependent slopes including both current value and the slope of the trajectory at time t (Ye2008b)

$$\lambda_i(t) = \lambda_0(t) \exp(\gamma X_i + \alpha_1 \omega_i(t) + \alpha_2 \omega_i'(t))$$

Parameter α_1 has the same interpretation as in true value and α_2 represents for patients having the same level of the true longitudinal biomarker at time t, the log hazard ratio for a unit increase in the current slope of the longitudinal trajectory. This parameterization could capture situations in which, at a specific time point, two patients show similar true marker levels, but they may differ in the rate of change of the marker.

Different parameterizations

- Cumulative effect including the whole area under the trajectory (Rizopoulos2012)

$$\lambda_i(t) = \lambda_0(t)\exp(\gamma X_i + \alpha_1(\int_0^t (\beta_0 + \beta_1 s_i + U_{0i} + U_{1i} s_i)ds))$$

With this parameterization, α_1 is the association between the whole history (area under the trajectory) of the longitudinal biomarker and survival.

Shared Random Effect Models

Joint Likelihood Calculation

$$p(T_i, \delta_i, y_i | W_i; \theta) = p(T_i, \delta_i | W_i; \theta) p(y_i | W_i; \theta)$$

$$p(y_i | W_i; \theta) = \prod_j p y_i(tij) | W_i; \theta$$

Log-likelihood contribution:

$$\log p(T_i, \delta_i, y_i; \theta) = \log \int p(T_i, \delta_i, y_i | W_i; \theta) dW_i$$

$$= log \int p(T_i, \delta_i | W_i; \theta_t, \beta)[\prod p(y_i | W_i; \theta_y)] p(W_i; \theta_W) dW_i$$

where T_i is the time to survival and δ_i is the censoring indicator.

Table of Contents

Extensions on Joint Modelling for longitudinal and survival data

- Multiple longitudinal data
- Dynamic predictions for time-to-event data
- Competing risks on time-to-event data
- Non-linear longitudinal trends
- Non-linear effects in Cox proportional hazard model

Extensions on Joint Modelling for longitudinal and survival data

- Multiple longitudinal data
  Bayesian models and two-stage approaches. *JMbayes* uses IS (Importance Sampling (IS)) two-stage approach to correct the full likelihood. *JMbayes2*
- Dynamic predictions for time-to-event data
  JM, *JMbayes* packages
- Competing risks on time-to-event data
  JMbayes2 package
- Non-linear longitudinal trends
  JM, Two-stage modelling approaches using *mgcv*
- Non-linear effects in Cox proportional hazard model
  Two-stage modelling approaches using *mgcv*

Table of Contents

Motivating Database: Peritoneal Dialysis Program

The Peritoneal Dialysis Programme includes 160 patients at the Unit of the Nephrology Department (Hospital Geral de Santo António, Centro Hospital do Porto) followed up between years: October 1999 and February 2013.

- Baseline characteristics: age, gender
- The event that forced the patient to abandon the treatment program: time to kidney failure with % 66.88 censored times
- The median of follow-up time was 27.4 months (IQR: 12.8-49.0 months)
- Longitudinal biomarkers: Albumin and Calcium measurements among others...

Motivating Database: Peritoneal Dialysis Program

Main Objective: Association between longitudinal Albumin and Calcium measurements and their effects on survival.

- Low albumin level is usually associated with kidney failure
- Calcium levels of the blood may drop when the kidney fails

Motivating Database: Peritoneal Dialysis Program

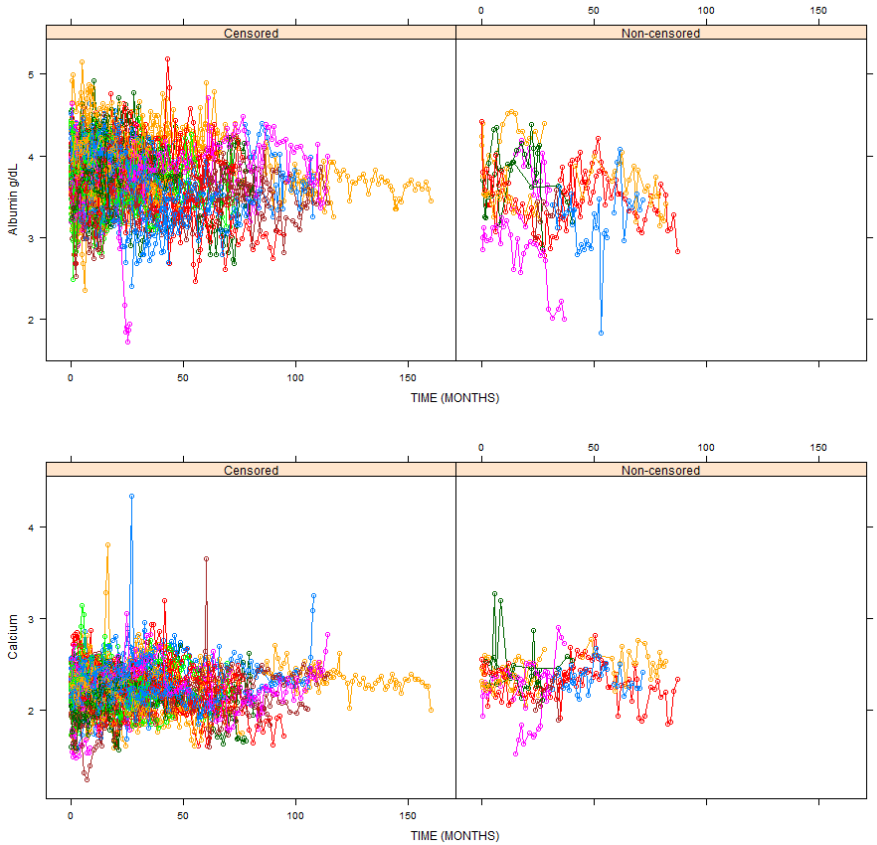


Figure: Subject specific trajectories of Albumin and Calcium levels for censored and non-censored times

Motivating Database : OLT Data

Motivating Database 2: Institutional clinical database, adult patients who underwent OLT in the Hospital Clínico Universitario de Santiago, between July 1994 and July 2011.

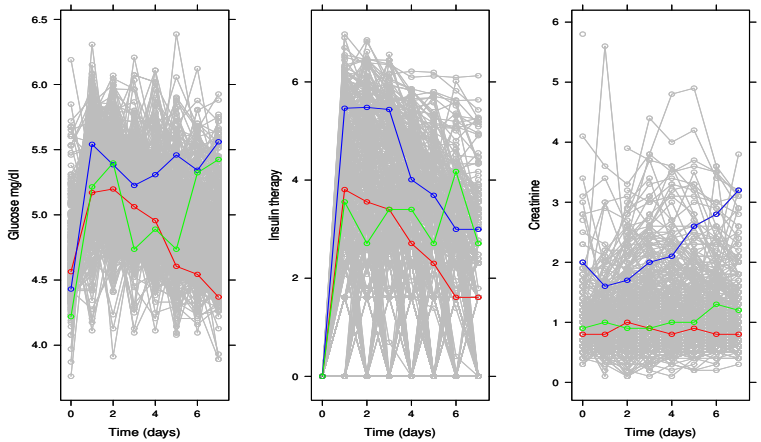


Figure: Glucose, Insulin and Creatinine levels of patients. Subject specific trajectories of 3 patients for each measurements are highlighted in colors.

Table of Contents

Software

R packages for Joint Models

- *joineR* by Pete Philipson, Ines Sousa , Peter J. Diggle,Paula Williamson
  Allows one longitudinal and survival data with random effects parameterization only.
- *joineRML* by Graeme L. Hickey, Pete Philipson, Andrea Jorgensen, Ruwanthi Kolamunnage-Dona, Paula Williamson, Dimitris Rizopoulos, Alessandro Gasparini
  Allows joint models for multiple longitudinal and survival data with Monte Carlo Expectation-Maximisation.

Software

R packages for Joint Models

- *JM* by Dimitris Rizopoulos
  Allows one longitudinal and survival data with different parametrizations.
- *JMbayes* by Dimitris Rizopoulos
  Directly implements the MCMC, allows for categorical longitudinal data as well, allows for general transformation functions
- *JMbayes2* by Dimitris Rizopoulos
  Multiple longitudinal outcomes of mixed type (continuous/categorical) and multiple event times (competing risks and multi-state processes) are accommodated.

Example in JMbayes (Rizopoulos example)

```
install.packages("JMbayes")
install.packages("rjags")
library(JMbayes)
library(rjags)

MixedModelFit2 <- mvglmer(list(log(serBilir) ~
year + (year | id),

spiders ~ year + (1 | id)), data = pbc2,
families = list(gaussian, binomial))

JMFit2 <- mvJointModelBayes(MixedModelFit2, CoxFit,
timeVar = "year")
```

Example in JMbayes2, (Rizopoulos example)

```
install.packages("JMbayes2")
# Cox model for the composite event
pbc2.id$status2 <- as.numeric(pbc2.id$status != 'alive')
    CoxFit <- coxph(Surv(years, status2) ~ sex,
     data = pbc2.id)

# a linear mixed model for log serum bilirubin
fm1 <- lme(log(serBilir) ~ year * sex,
data = pbc2, random = ~ year | id)

# a linear mixed model for the prothrombin time
fm2 <- lme(prothrombin ~ year * sex,
data = pbc2, random = ~ year | id)
```

Example in JMbayes2

```
# a mixed effects logistic regression model for ascites
fm3 <- mixed_model(ascites ~ year + sex, data = pbc2,
random = ~ year | id, family = binomial())

# the joint model that links all sub-models
jointFit <- jm(CoxFit, list(fm1, fm2, fm3),
time_var = "year",
n_iter = 12000L, n_burnin = 2000L, n_thin = 5L)
summary(jointFit)
```

Example in JMbayes2

```
library("JMbayes")

MixedModelFit <- mvglmer(list(prothrombin ~
 year * sex + (year | id),
hepatomegaly ~ year * sex + (year | id)),
data = pbc2,
families = list(gaussian, binomial))


pbc2.id$event <- as.numeric(pbc2.id$status != "alive")
CoxFit <- coxph(Surv(years, event) ~ drug + age,
 data = pbc2.id, model = TRUE)
```

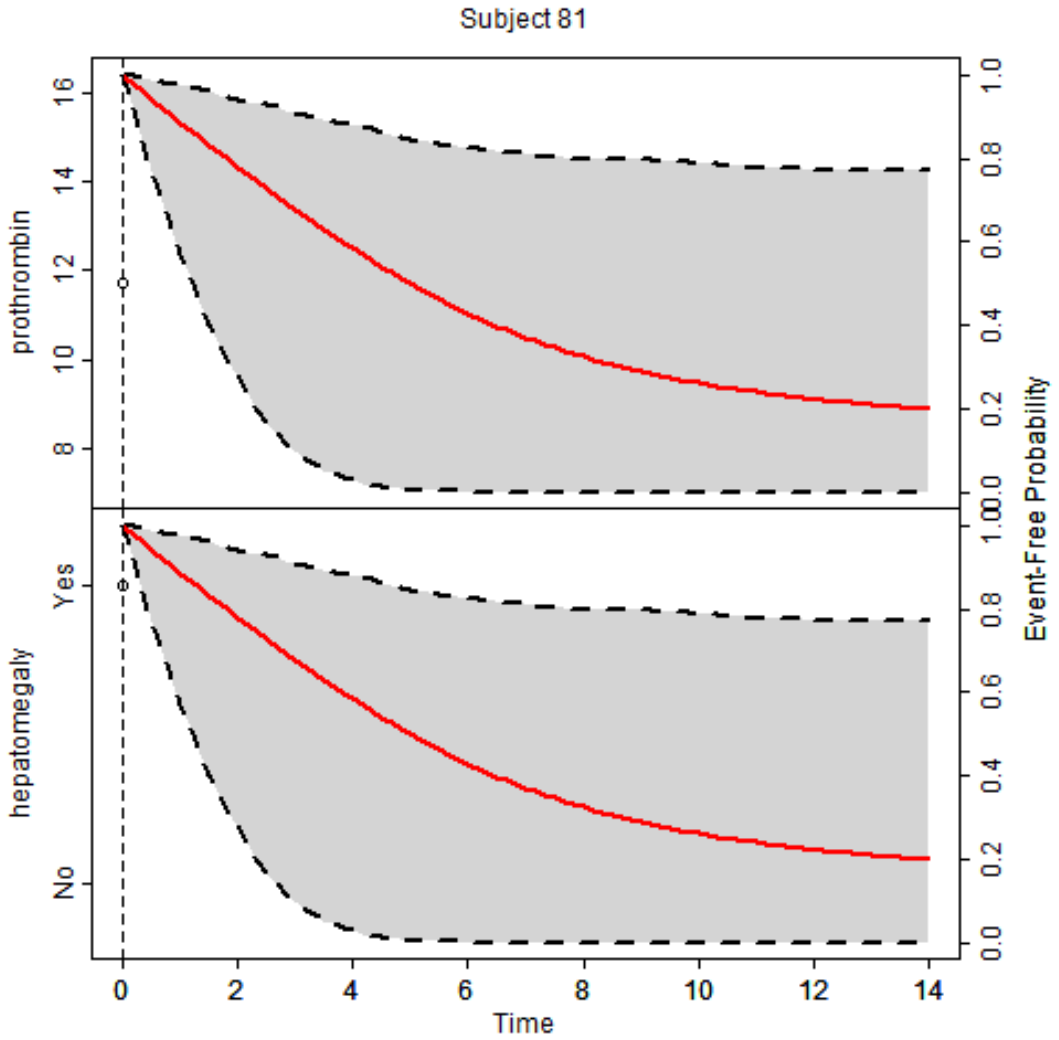
Example in JMbayes2

```
ND <- pbc2[pbc2$id == 81, ]

sprobs <- survfitJM(JMFit, ND)
sprobs

plot(sprobs, split = c(2, 1), surv_in_all = TRUE,
lty_lines_CI = 3, col_lines = "blue",
 col_fill_CI = "pink2",
main = "Patient 81", ylab = c("Prothro", "Hepa"),
col_points = "red", pch_points = 16,
cex_xlab = 0.8, cex_ylab = 0.8, cex_zlab = 0.8,
cex_main = 0.8, cex_axis = 0.7)
```

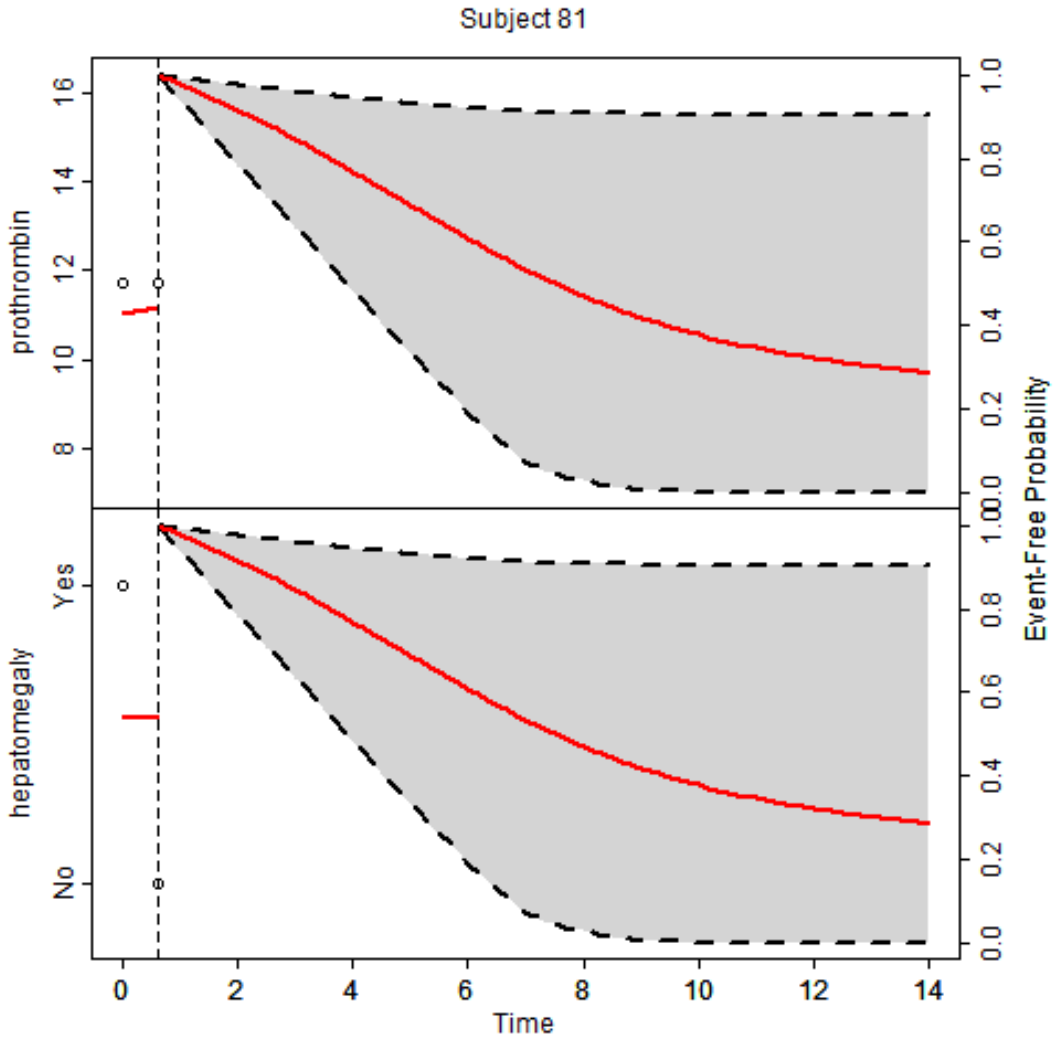
Example in JMbayes2

```
N <- nrow(ND)
dyn_sprobs <- vector("list", N)

for (i in seq_len(N)) {
dyn_sprobs[[i]] <- survfitJM(JMFit, ND[1:i, ],
survTimes = seq(0, 14, length.out = 85))
plot(dyn_sprobs[[i]], split = c(2, 1),
surv_in_all = TRUE)
}
```

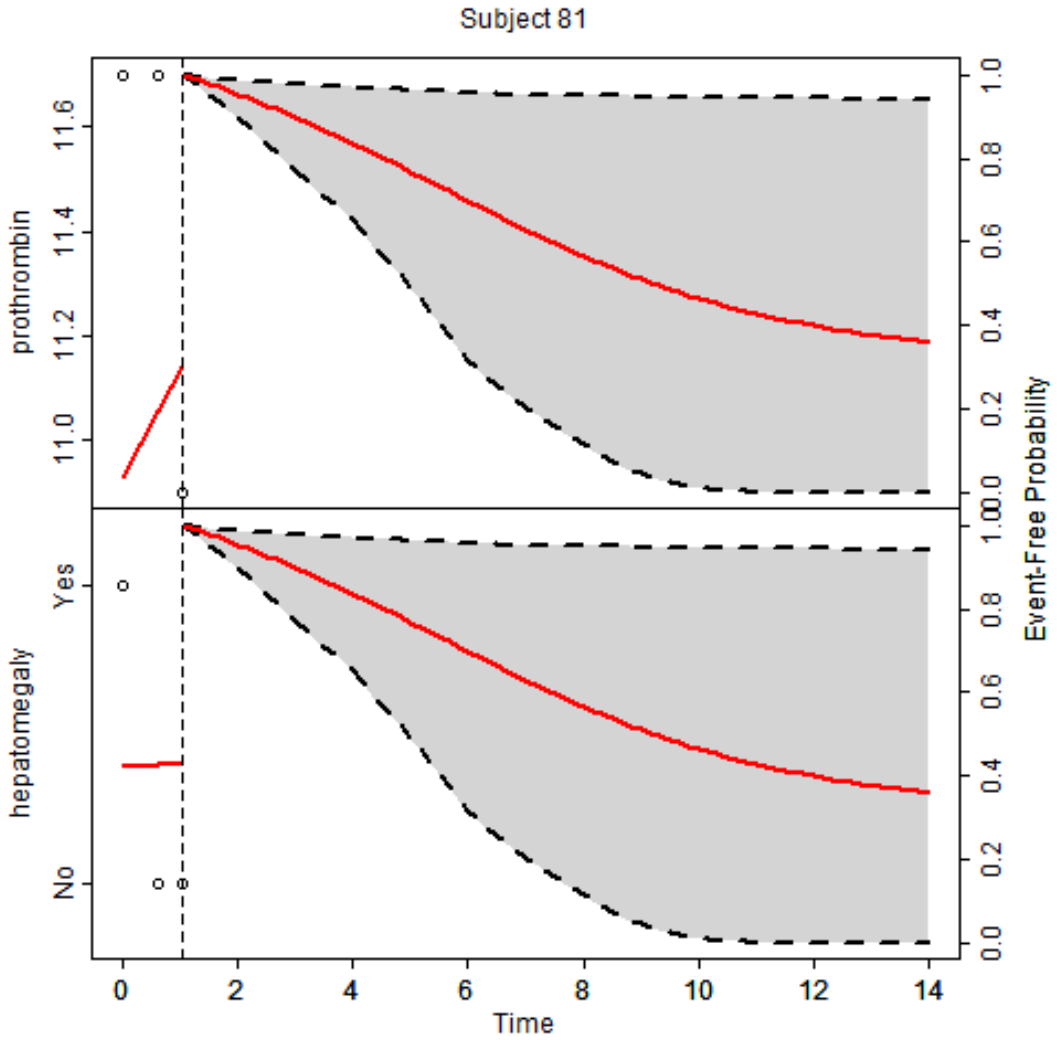
Dynamic Predictions of Survival Data



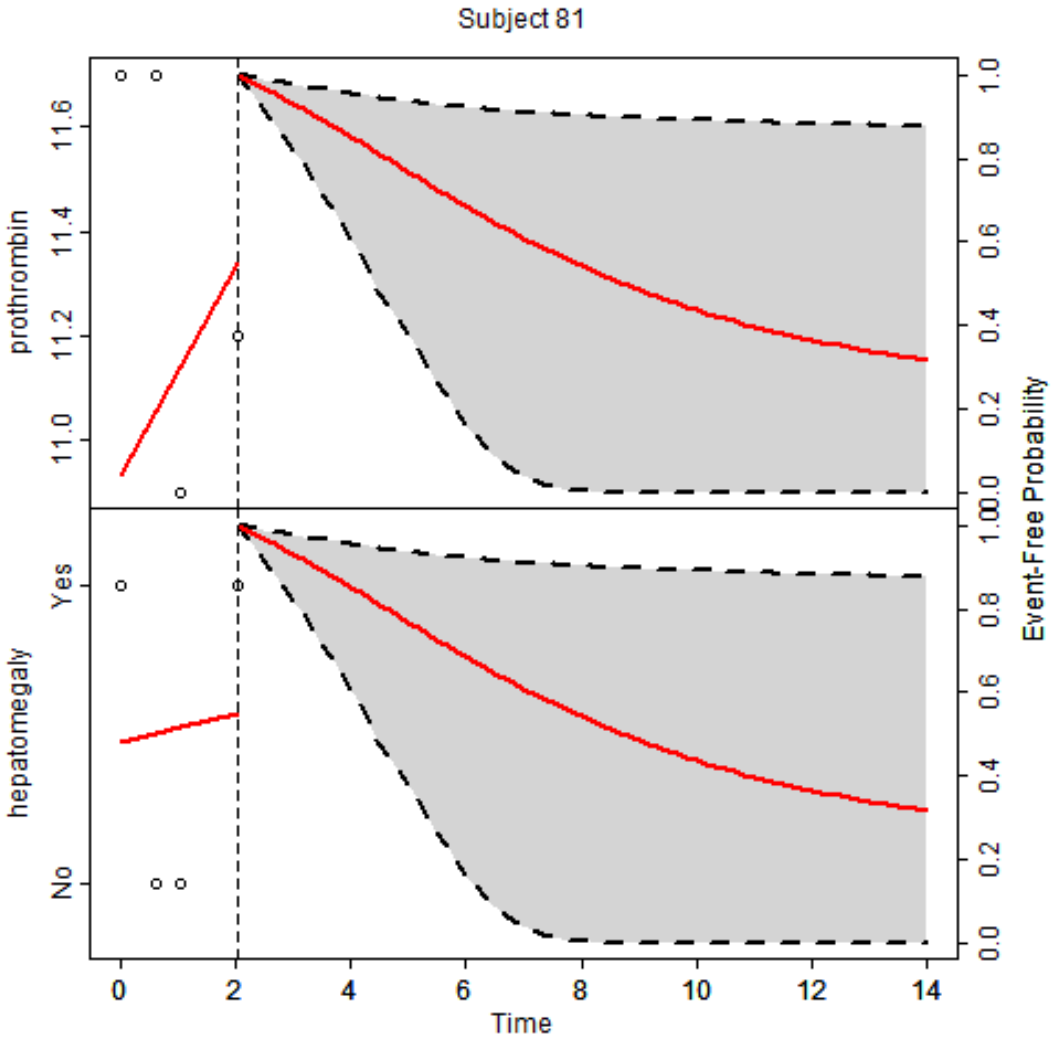
Dynamic Predictions of Survival Data



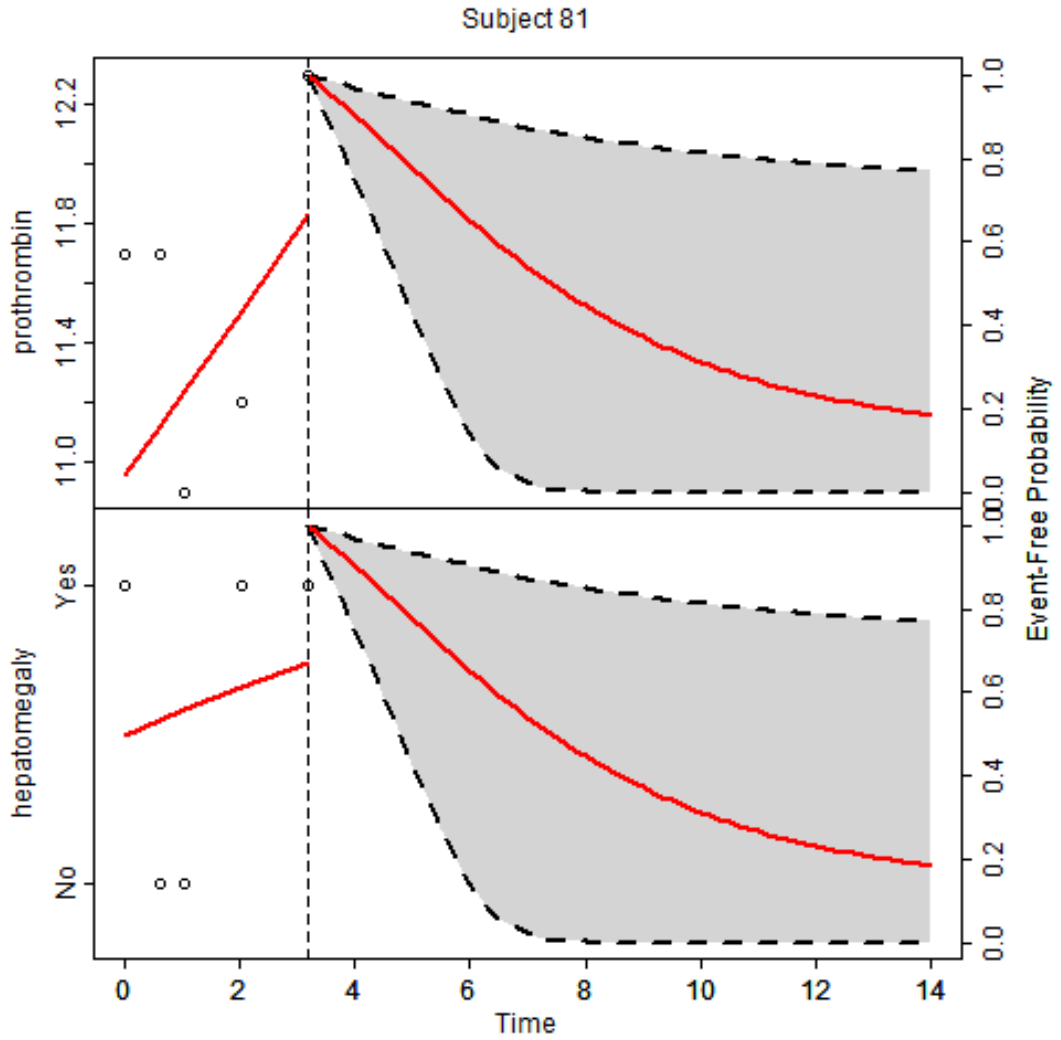
Dynamic Predictions of Survival Data



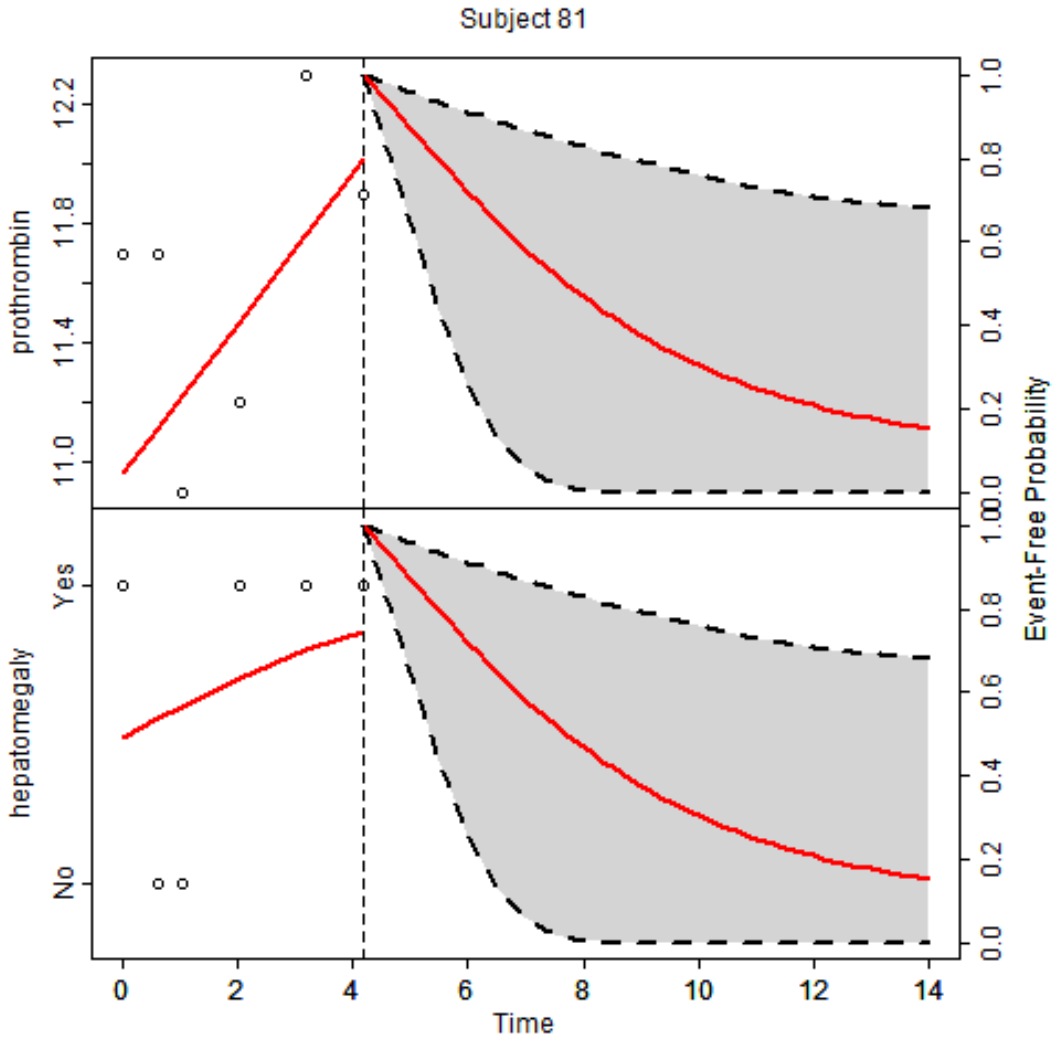
Dynamic Predictions of Survival Data



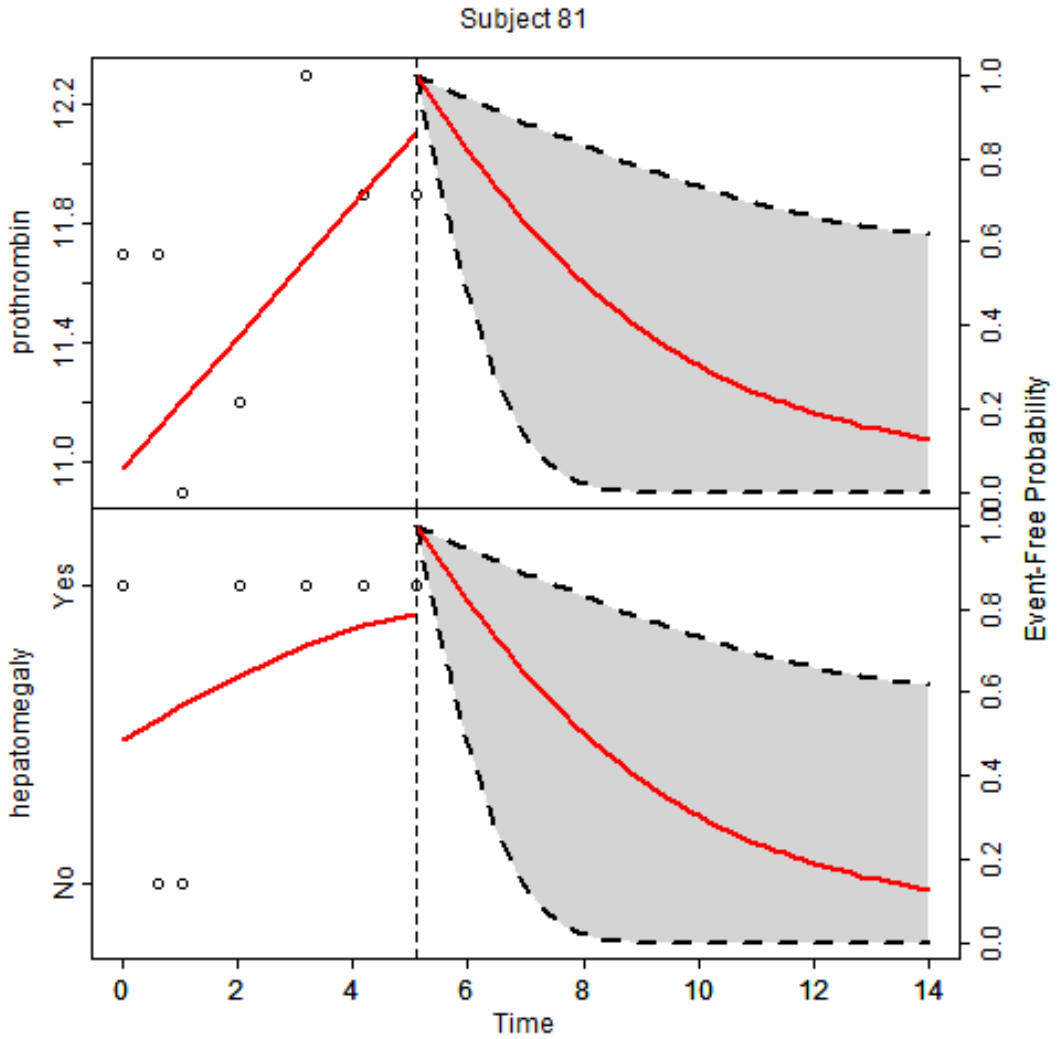
Dynamic Predictions of Survival Data



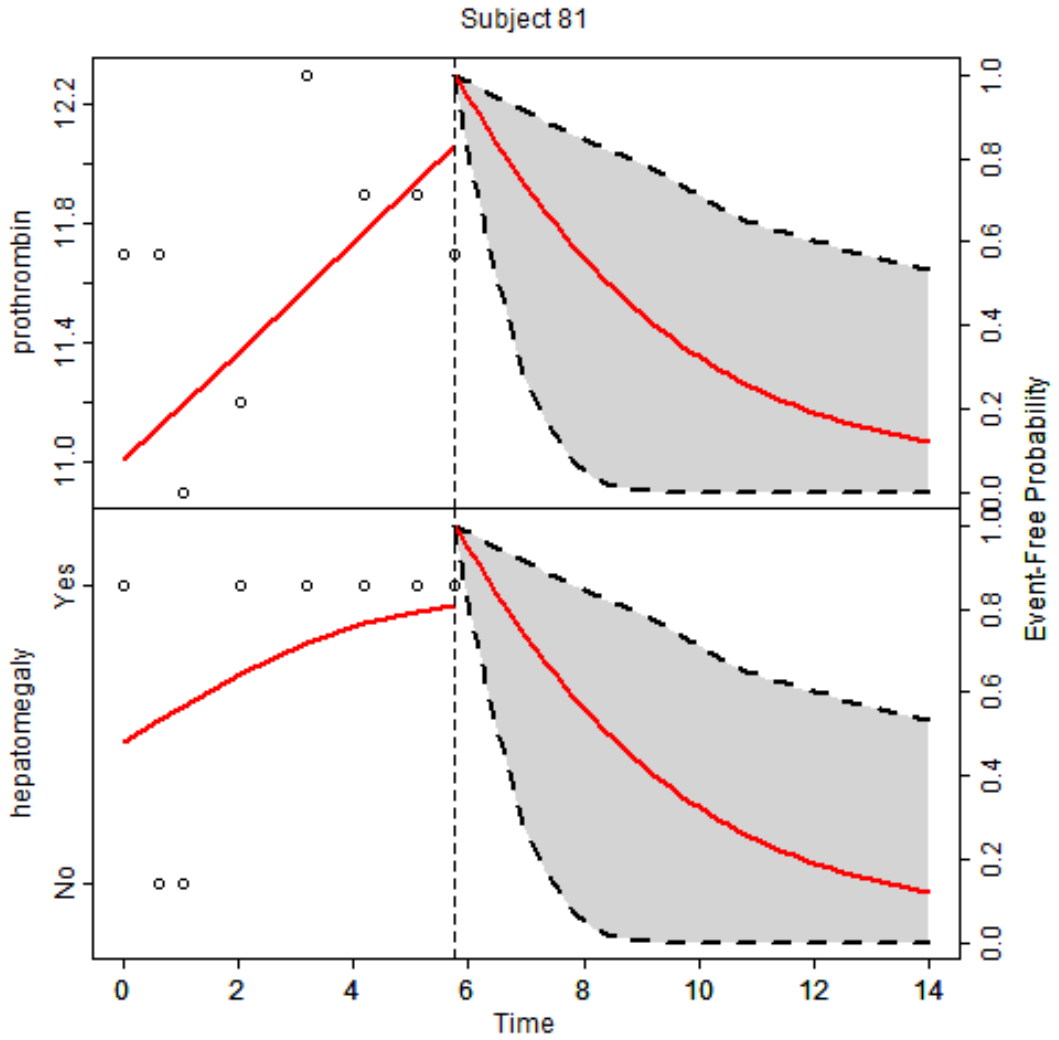
Dynamic Predictions of Survival Data



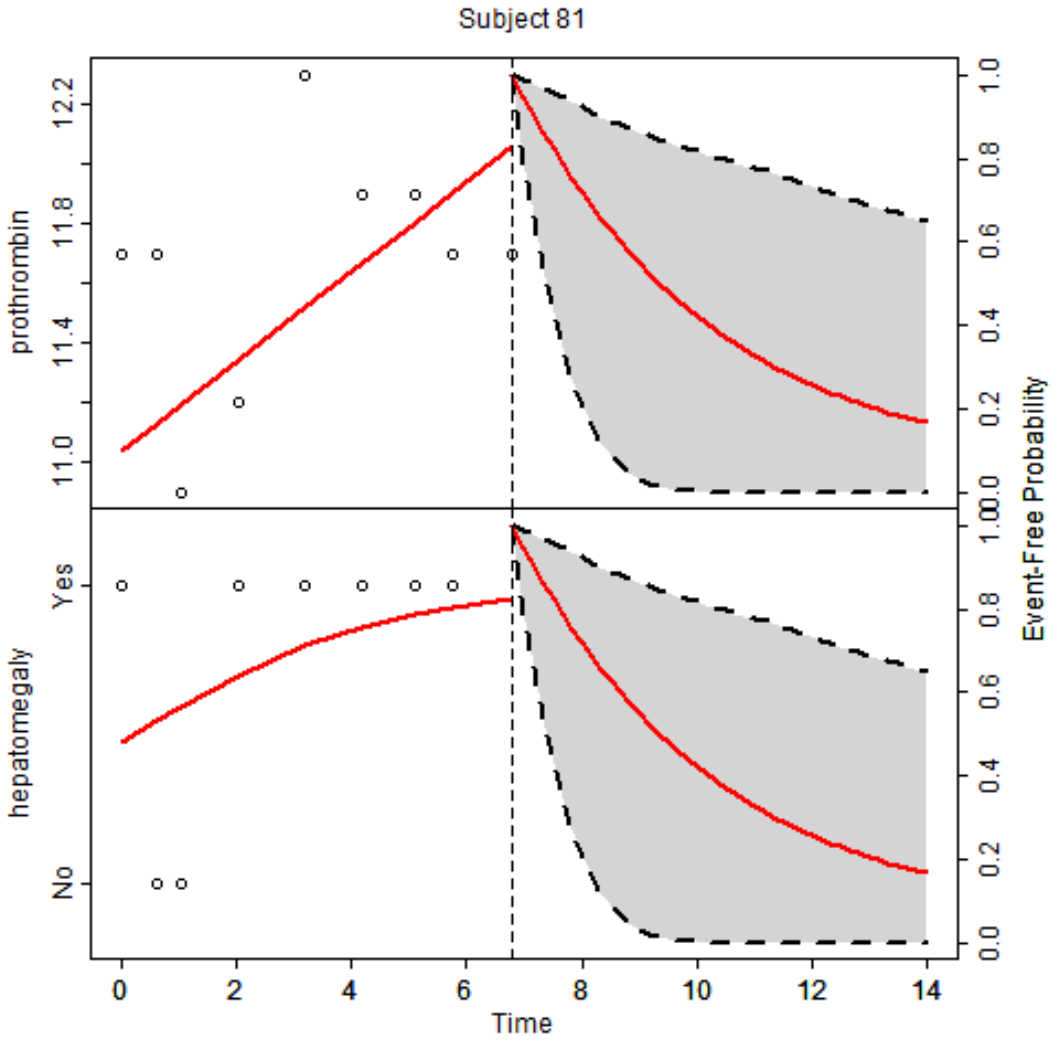
Dynamic Predictions of Survival Data



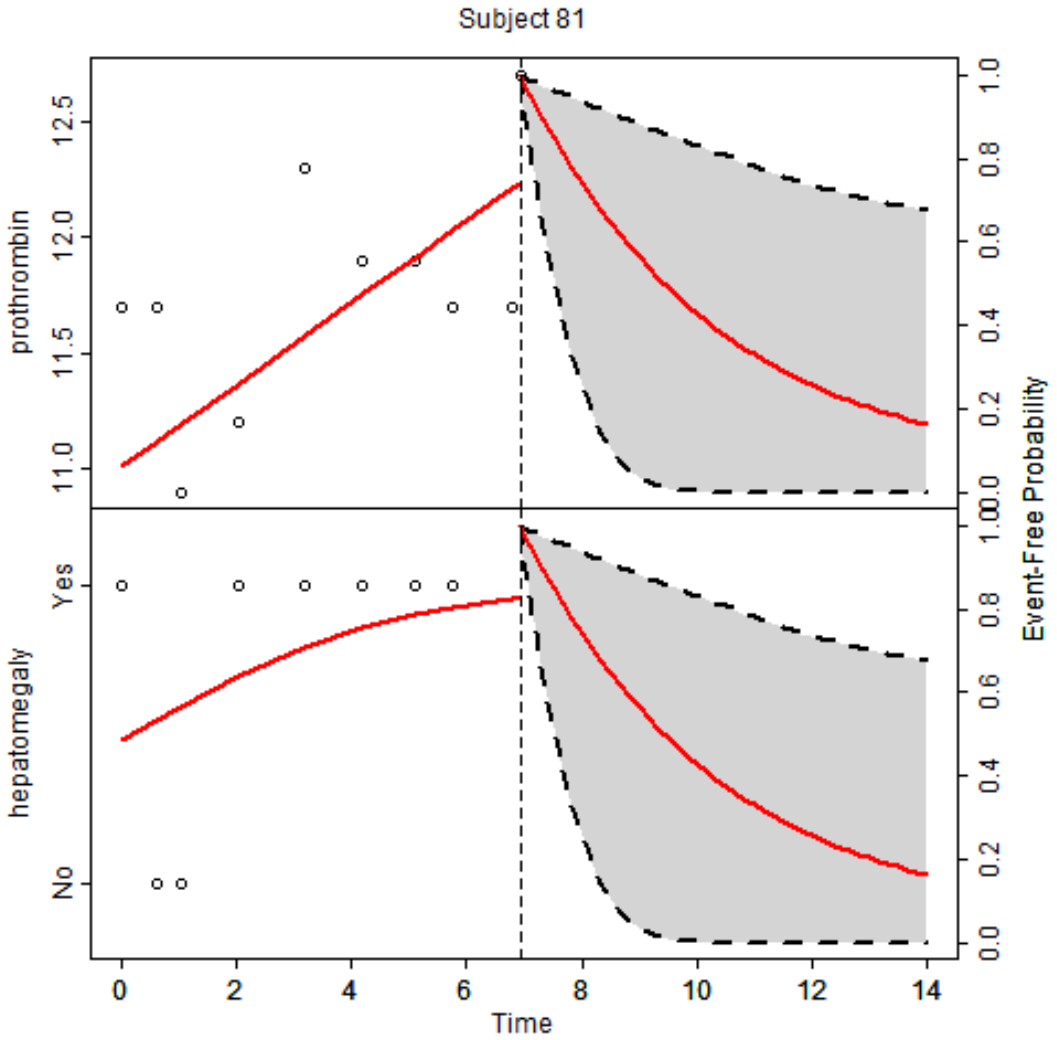
Dynamic Predictions of Survival Data



Dynamic Predictions of Survival Data



Dynamic Predictions of Survival Data



Selected References

- Tsiatis A. A., DeGruttola V., Wulfsohn. M. S. (1995) Modeling the relationship of survival to longitudinal data measured with error: applications to survival and CD4 counts in patients with AIDS. *Journal of the American Statistical Association*. vol. 90, pp. 27-37, 1995.
- Guler, I., Faes, C., Cadarso Suarez, C., Teixeira, L., Rodrigues, A. and Mendonça, D. (2017), Two stage model for multivariate longitudinal and survival data with application to nephrology research. *Biometrical Journal*, 59: 1204-1220
- 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.
- Rizopoulos D. (2011) Dynamic predictions and prospective accuracy in joint models for longitudinal and time-to-event data. *Biometrics*, 67:819?29.
- Rizopoulos, D. (2016) "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.

ipek.guler@kuleuven.be