Joint Modeling of Longitudinal and Survival Outcomes

Anteneh Tesema and Yebelay Berehan

Ethiopian Public Health Institute (EPHI)

National Data Management Center for Health (NDMC)

August 24-27, 2023

Joint Models for Longitudinal and Time-to-Event Data

- What is Joint Modelling?
- a joint modeling approach to analyze two types of outcomes often observed in longitudinal studies:
  - A set of longitudinal response measurements.
  - The time to an event of interest, such as default, death, etc.
- Traditionally, these two outcomes have been analyzed separately:
  - Using a mixed effects model for the longitudinal response.
  - A survival model for the time-to-event.
- However, in this section, we will explore how these outcomes can be analyzed jointly.

Overview of survival or time to event

- Survival analysis is a set of statistical techniques designed for analyzing data where the outcome variable is the time until an event occurs.
- This event time is often referred to as failure time, survival time, or event time.
- Survival time signifies the time from a specific starting point (e.g., treatment initiation) to a particular endpoint (time-to-event).
- Time, Time Origin, Time Scale, Event

In survival analysis, the definition of an individual's failure time requires three elements:

1. **Time Origin**: The starting point for measuring time.
2. **Time Scale**: The units used for measuring time (e.g., years, months, days).
3. **Event**: The specific occurrence of interest (e.g., death, disease incidence, default).

- For biomedical applications, this could involve events like death or disease incidence.
- In fields like credit scoring, it might be default, and in engineering, it could be component failure.

When considering multiple events, such as various causes of death, the problem can involve recurrent events or competing risks.

Goals of Survival Analysis

The primary objectives of survival analysis include:

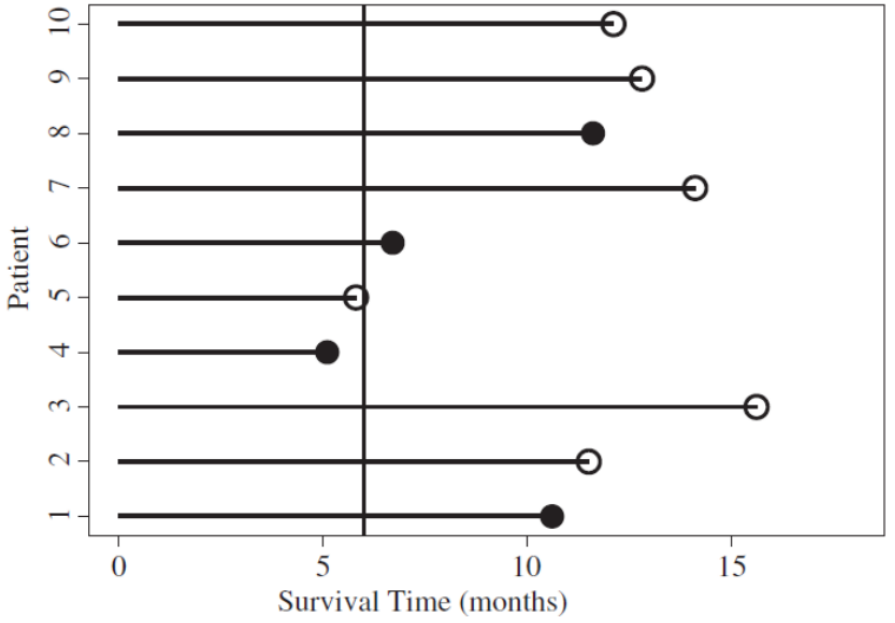
1. Estimating Time-to-Event: estimate the time it takes for an event to occur for a group of individuals.

2. Comparing Time-to-Event between two or more groups.

3. Assessing Covariate Relationships: to assess how covariates relate to the time-to-event.

Censoring

- The distinguishing feature of survival analysis is that it incorporates a phenomenon called censoring.
- Censoring occurs when we have some information about individual survival time, but we don’t know the time exactly.



Kaplan-Meier (KM) Curves:

- a graphical representation of the estimated survival probability over time.
- visualize how the survival probability changes as time progresses.
- typically stratified by different groups, allowing comparisons between these groups.

Log-Rank Test:

- The Log-Rank test is statistical test used to compare the survival distributions of two or more groups.
- It assesses whether there are significant differences in survival times between the groups.
- The test is based on comparing the observed number of events and expected number of events under the null hypothesis of equal survival distributions.

Pros and Cons of the Kaplan-Meier Estimator

Pros:

- It is commonly used to describe survival.
- It is commonly used to compare two study populations.
- It provides an intuitive graphical presentation of survival data.

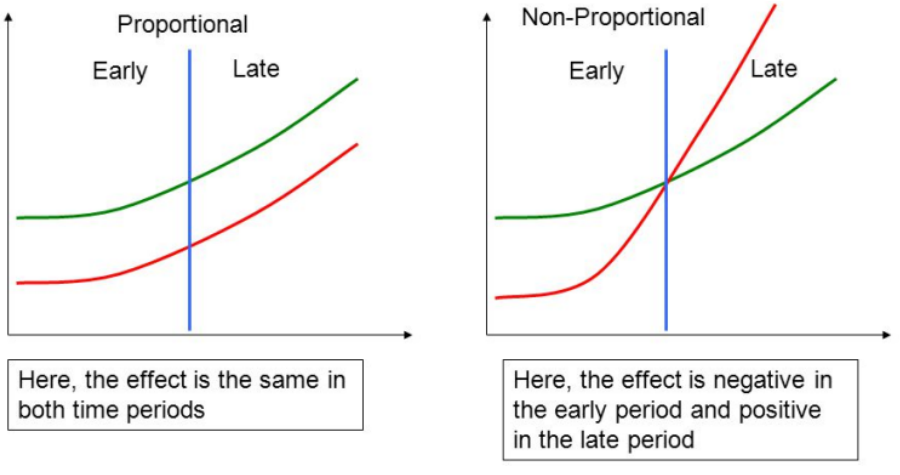
Cons:

- It is mainly descriptive in nature.
- It does not control for covariates or other factors that may influence survival.
- It cannot accommodate time-dependent variables in its basic form.

The Cox Proportional Hazard Model

- The Cox proportional hazard model provides the following benefits:
- Adjusts for multiple risk factors simultaneously.
- Allows quantitative (continuous) risk factors, helping to limit the number of strata.
- Provides estimates and confidence intervals of how the risk changes across the strata and across unit increases in quantitative variables.
- Can handle data sets with right censoring, staggered entry, etc.; so long as we have adequate data at each time point.

- The proportional hazard function has the form:
- $h(t) = h_0(t)e^{\beta_1 x_1 + \ldots + \beta_p x_p}$ Where h_0 is the baseline hazard rate, i.e., $x_1 = 0, x_2 = 0,$ etc.
- Note that the ratio of 2 hazard functions does not depend on t.



Time-Varying Covariates

- Often, there is interest in the association between a time-varying covariate and the risk of an event.
  - Treatment changes with time (e.g., dose)
  - Time-dependent exposure (e.g., smoking, diet)
  - Markers of disease or patient condition (e.g., blood pressure, PSA levels)

Example: PBC Study

- In the PBC study, we explore if longitudinal bilirubin measurements are associated with the hazard of death.

Time-Varying Covariates

- To address our questions of interest, we must formulate a model that connects:
  - Serum bilirubin levels
  - Time-to-death

Association with Baseline Marker Levels

- The connection between baseline marker levels and the risk of death can be assessed using standard statistical methods, such as Cox regression.

Study of Time-Varying Covariates

- When examining time-varying covariates, more careful consideration is essential.

Types of Time-Varying Covariates

- There are two types of time-varying covariates (Kalbfleisch & Prentice, 2002):
  - **External (aka exogenous):** The value of the covariate at time point t is not affected by the occurrence of an event at time point u, with t > u.
  - **Internal (aka endogenous):** The covariate is not External.
- Example: External vs. Internal
- This concept can be challenging to grasp, so let's clarify with an example...

Example: Asthma Study

- Let's consider a study on asthma, specifically focusing on the time until an asthma attack for a group of patients.
- We have two time-varying covariates:
  - **Pollution levels**
  - **A biomarker for asthma**

Pollution Levels and Biomarker

- Suppose a patient had an asthma attack at a certain time point, denoted as u.
- For the time-varying covariates:
  - **Pollution levels:** The pollution levels at a time point t > u will not be affected by the fact that the patient had an attack at u. (External)
  - **Biomarker:** The biomarker level at a time point t > u may be affected by the fact that the patient had an attack at u. (Internal)

Distinguishing Covariate Types

- It's crucial to differentiate between these two types of time-varying covariates, as the type of covariate determines the appropriate analysis.
- In our motivating examples, all time-varying covariates are Biomarkers. These are always endogenous covariates:
  - Measured with error (i.e., biological variation)
  - The complete history is not available
  - Existence is directly related to failure status

Extension of Cox Model for Time-Varying Covariates

- The Cox model presented earlier can be extended to handle time-varying covariates using the counting process formulation:

$$h_i(t|Y_i(t), w_i) = h_0(t)R_i(t)\exp\{\gamma^T w_i + \alpha y_i(t)\}$$

where:

- $N_i(t)$ is a counting process that tracks the number of events for subject i by time t,
- $h_i(t)$ denotes the intensity process for $N_i(t)$,
- $R_i(t)$ denotes the at-risk process (equals 1 if subject i is still at risk at time t),
- $y_i(t)$denotes the value of the time-varying covariate for subject i at time t.
- This formulation allows for the incorporation of time-varying covariates into the Cox model.

Interpretation

The formulation:

$$h_i(t|Y_i(t), w_i) = h_0(t)R_i(t)\exp\{\gamma^T w_i + \alpha y_i(t)\}$$

has the following interpretation:

- The term $\exp(\alpha)$ denotes the relative increase in the risk of an event at time t that results from a one-unit increase in $y_i(t)$ at the same time point.

Handling Time-Varying Covariates in the Extended Cox Model

The extended Cox model handles time-varying covariates as follows:

- It assumes no measurement error.
- The covariate path is represented by a step function.
- The existence of the covariate is not related to failure status.

Validity of the Extended Cox Model

- The extended Cox model is valid only for exogenous time-varying covariates.
- Treating endogenous covariates as exogenous may produce spurious results!

Joint Modeling Framework

- To account for the special features of endogenous covariates, a new class of models has been developed: **Joint Models for Longitudinal and Time-to-Event Data**.
- The intuitive idea behind these models:
    1. Use an appropriate model to describe the evolution of the covariate/marker over time for each patient.
    2. The estimated evolutions are then used in a Cox model.
- A key feature of these models is that covariate levels are not assumed constant between visits.

Notation

- Some notation:
  - T_i^*: True event time for patient i
  - T_i: Observed event time for patient i
  - δ_i: Event indicator, i.e., equals 1 for true events
  - y_i: Longitudinal covariate
- We will formulate the joint model in 3 steps – in particular, . . .

Step 1: Formulation of Joint Model

- Step 1: Let's assume that we know $m_i(t)$, i.e., the true and unobserved value of the covariate at time t.
- With this assumption, we can define a standard relative risk model:

$$h_i(t|M_i(t)) = h_0(t)\exp\{\gamma^T w_i + \alpha m_i(t)\},$$

  where:

  - $M_i(t) = \{m_i(s), 0 \leq s < t\}$ represents the longitudinal history,
  - α quantifies the association between the time-varying covariate and the risk of an event,
  - w_i represents the baseline covariates.

Step 2: Reconstructing Covariate History

- Step 2: From the observed longitudinal data $y_i(t)$, reconstruct the covariate history for each subject.
- We use a mixed effects model to achieve this (focusing on continuous covariates for now):

$$y_i(t) = m_i(t) + \epsilon_i(t)$$
$$= x_i(t)^T \beta + z_i(t)^T b_i + \epsilon_i(t),$$

  where:

  - $x_i(t)$ and β: Fixed-effects part,
  - $z_i(t)$ and b_i: Random-effects part, $b_i \sim N(0, D)$, $\epsilon_i(t) \sim N(0, \sigma^2)$.

Step 3: Associating the Two Processes and Defining a Joint Distribution Model

Joint models for associating two processes are often structured as follows (Tsiatis & Davidian, Stat. Sinica, 2004):

The joint distribution is given by:

$$p(y_i, T_i, \delta_i) = Zp(y_i|b_i)h(T_i|b_i)^{\delta_i} S(T_i|b_i)p(b_i)\, db_i,$$

where:

- b_i is a vector of random effects that explains the interdependencies.
- $p(\cdot)$ represents the density function.
- $S(\cdot)$ represents the survival function.
- Z represents any normalizing constant.

This structure allows us to define a model for the joint distribution of the two processes.

Analysis of a Real Data Example Using JM

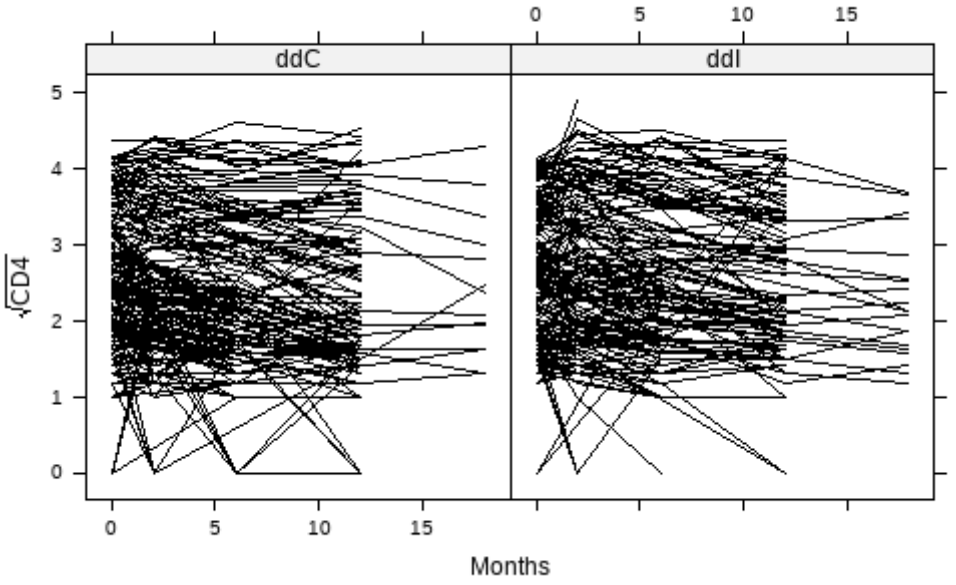
- Consider a longitudinal study on 467 HIV infected patients who had failed or were intolerant of zidovudine therapy.
- Aim: compare the efficacy and safety of two alternative antiretroviral drugs: **didanosine (ddI) and zalcitabine (ddC)**.
- Patients were randomly assigned to receive either ddI or ddC, and CD4 cell counts were recorded at study entry and at 2, 6, 12, and 18 months thereafter.
  - By the end of the study, 188 patients had died, resulting in 59.7% censoring.
- Our main research question is to test for a treatment effect on survival after adjusting for the CD4 cell count.
- "The CD4 cell count measurements are generated by patients and are only available at specific visit times.
  - This situation exemplifies a typical time-dependent covariate, measured intermittently with error."

- The longitudinal and survival information is available in the data frames `aids` and `aids.id` respectively.
- The CD4 cell counts exhibit right-skewed distribution shapes; for analysis, we work with the square root of the CD4 cell values.
- As a descriptive analysis, Figure 1 shows subject-specific longitudinal profiles and the Kaplan-Meier estimate for time-to-death.

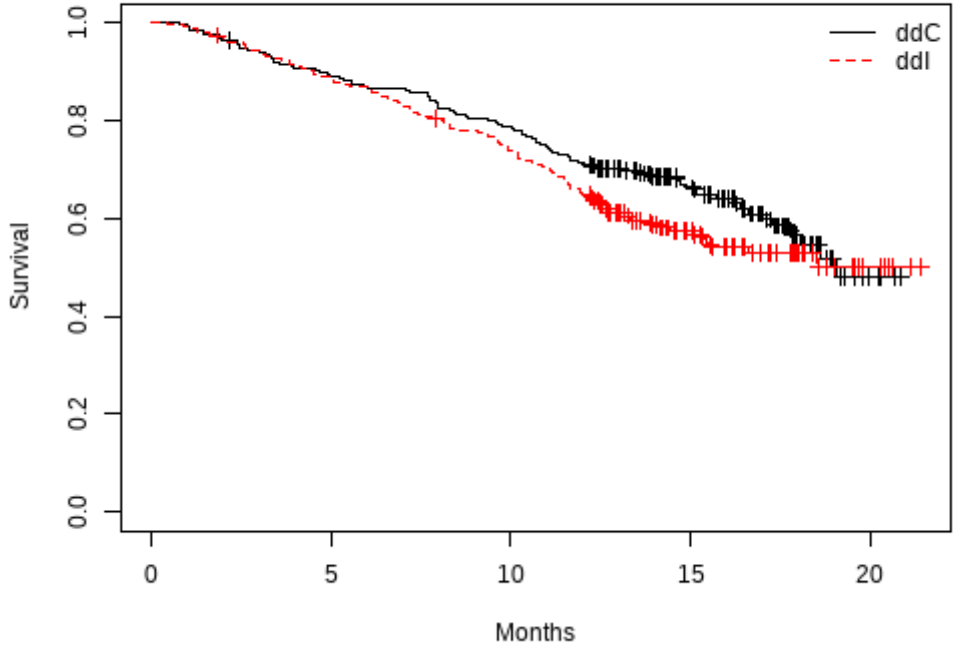
Descriptive Analysis - Longitudinal Profiles and Survival

To perform a descriptive analysis of the data, we can visualize the longitudinal profiles and survival curves using R and the `JM` and `lattice` libraries.

```
library("JM"); library("lattice")
xyplot(sqrt(CD4) ~ obstime | drug, group = patient, data = aids,
       xlab = "Months", ylab = expression(sqrt("CD4")), col = 1, type = "l")
```



```r
# Survival Curves
plot(survfit(Surv(Time, death) ~ drug, data = aids.id), conf.int = FALSE,
     mark.time = TRUE, col = c("black", "red"), lty = 1:2,
     ylab = "Survival", xlab = "Months")
legend("topright", c("ddC", "ddI"), lty = 1:2, col = c("black", "red"),
       bty = "n")
```



Observations and Initial Analysis

- We observe that both groups of patients exhibit similar variability in their longitudinal profiles.
- However, from the Kaplan-Meier estimate , it appears that the ddC group has slightly higher survival than the ddI group after six months of follow-up.
- To highlight the advantages of the joint modelling approach, we will begin with a 'naive' analysis.
- In this analysis, we ignore the special characteristics of CD4 cell counts and fit a Cox model that includes the treatment indicator and CD4 as a typical time-dependent covariate.

- We will use the standard counting process form of the Cox model to fit this analysis:

```
td.Cox <- coxph(Surv(start, stop, event) ~ drug + sqrt(CD4), data = aids)
summary(td.Cox)
```

```
## Call:
## coxph(formula = Surv(start, stop, event) ~ drug + sqrt(CD4),
##     data = aids)
##
##   n= 1405, number of events= 188
##
##               coef exp(coef) se(coef)      z Pr(>|z|)
## drugddI    0.32678   1.38650  0.14708  2.222   0.0263 *
## sqrt(CD4) -0.72302   0.48528  0.07997 -9.042   <2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
##           exp(coef) exp(-coef) lower .95 upper .95
## drugddI      1.3865     0.7212    1.0393    1.8498
## sqrt(CD4)    0.4853     2.0606    0.4149    0.5676
##
## Concordance= 0.696  (se = 0.018 )
## Likelihood ratio test= 86.14  on 2 df,   p=<2e-16
## Wald test            = 83.51  on 2 df,   p=<2e-16
## Score (logrank) test = 83.25  on 2 df,   p=<2e-16
```

Advanced Analysis - Fitting a Joint Model

- After adjusting for the square root of CD4 count in the Cox model, no strong evidence for a treatment effect is observed.
- We proceed by specifying and fitting a joint model that explicitly postulates a linear mixed-effects model for CD4 cell counts.
- Taking advantage of the randomization setup of the study, we include in the fixed-effects part of the longitudinal submodel the main effect of time and the interaction of treatment with time.
- In the random-effects design matrix, we include an intercept and a time term.
- For the survival submodel (similarly to the Cox model), we include the treatment effect as a time-independent covariate, and as a time-dependent one, the true underlying effect of CD4 cell count estimated from the longitudinal model.
- The baseline risk function is assumed piecewise constant with six knots placed at equally spaced percentiles of the observed event times.

Fitting the Joint Model

- To fit the joint model, a two-step process is followed. First, the linear mixed-effects and Cox models are fitted separately.
- The returned objects from these separate fits are then used as main arguments in the `jointModel()` function.
- Importantly, the structure of the joint model for the longitudinal and survival submodels mirrors that of the separately fitted models.
- In the survival submodel, the estimated 'true' longitudinal outcome $m_i(t)$ is incorporated into the linear predictor.
- Due to the fact that `jointModel()` extracts necessary information from these two objects, in the `coxph()` function call, we must specify `x = TRUE` to include the Cox model's design matrix in the returned object.

```r
# Separate Model Fits
fitLME <- lme(sqrt(CD4) ~ obstime + obstime:drug,
              random = ~ obstime | patient, data = aids)
summary(fitLME)
```

```
## Linear mixed-effects model fit by REML
##   Data: aids
##        AIC      BIC    logLik
##   2699.069 2735.789 -1342.535
##
## Random effects:
##  Formula: ~obstime | patient
##  Structure: General positive-definite, Log-Cholesky parametrization
##             StdDev     Corr
## (Intercept) 0.87143264 (Intr)
## obstime     0.03617033 -0.015
## Residual    0.36844785
##
## Fixed effects:  sqrt(CD4) ~ obstime + obstime:drug
##                     Value  Std.Error  DF  t-value p-value
## (Intercept)      2.5118005 0.04258901 936 58.97766  0.0000
## obstime         -0.0375070 0.00440225 936 -8.51997  0.0000
## obstime:drugddI  0.0082141 0.00632277 936  1.29912  0.1942
##  Correlation:
##                 (Intr) obstim
## obstime         -0.118
## obstime:drugddI  0.000 -0.687
##
## Standardized Within-Group Residuals:
##           Min            Q1           Med            Q3           Max
## -4.2480426451 -0.4082420037 -0.0002391742  0.4336550882  3.7150583354
##
## Number of Observations: 1405
```

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

```
## Call:
## coxph(formula = Surv(Time, death) ~ drug, data = aids.id, x = TRUE)
##
##   n= 467, number of events= 188
##
##           coef exp(coef) se(coef)     z Pr(>|z|)
## drugddI 0.2102    1.2339   0.1462 1.437    0.151
##
##         exp(coef) exp(-coef) lower .95 upper .95
## drugddI     1.234     0.8104    0.9264     1.643
##
## Concordance= 0.531  (se = 0.019 )
## Likelihood ratio test= 2.07  on 1 df,   p=0.2
## Wald test            = 2.07  on 1 df,   p=0.2
## Score (logrank) test = 2.07  on 1 df,   p=0.1
```

```r
# Joint Model Fit
fitJM <- jointModel(fitLME, fitSURV, timeVar = "obstime",
                    method = "piecewise-PH-GH")
```

```r
summary(fitJM)
```

```
Coefficients:
Longitudinal Process
                  Value Std.Err z-value p-value
(Intercept)      2.5558  0.0372 68.7961 <0.0001
obstime         -0.0423  0.0046 -9.1931 <0.0001
obstime:drugddI  0.0051  0.0065  0.7821  0.4342

Event Process
            Value 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
```

Interpreting Joint Model Results

- The main argument `timeVar` of `jointModel()` specifies the name of the time variable in the linear mixed-effects model.
  - This is vital for the computation of $m_i(t)$.
- The `summary()` method provides a detailed output, including parameter estimates, their standard errors, and asymptotic Wald tests for both the longitudinal and survival submodels.
- In the event process results, the parameter labeled **Assoct** corresponds to parameter α.
  - It measures the effect of $m_i(t)$ (in our case, the true square root CD4 cell count) on the risk of death.

The parameters x_i are (for $i = 1, \ldots, 7$) parameters for the piecewise constant baseline risk function.

- A comparison between the standard time-dependent Cox model and the joint model reveals interesting features.
- The regression coefficient for ddI is larger in magnitude in the joint model, indicating a slightly stronger treatment effect.
- A significant bias is observed for the CD4 cell count effect.
- In the time-dependent Cox model, the estimated regression coefficient is -0.72, whereas in the joint model, it's -1.10.
- For obtaining the Hazard Ratio for this variable we have to exponenciate the value exposed in the table.
- In this case the result is 0.33. According to this, one unit increse on the CD4 count cell decreases the risk 67%.

Results Summary

- Coefficients (SEs) from mixed-effects model and joint model

Variable	Mixed model	Joint model
obstime	-0.038(0.004)	-0.042(0.004)
obstime:drugddI	0.008 (0.006)	0.005 (0.007)

- Coefficients (SEs) from extended Cox model and joint model

Variable	Cox model	Joint model
drugddI	0.327(0.147)	0.351 (0.154)
cd4	-0.723(0.080)	-1.102(0.118)

Alternative Test - Likelihood Ratio Test (LRT)

- The Likelihood Ratio Test (LRT) provides an alternative to the Wald test for hypothesis testing.
- After fitting the joint model under the null hypothesis of no treatment effect in the survival submodel, we can use the `anova()` method to perform the LRT:

```
# Null Hypothesis Testing
fitSURV2 <- coxph(Surv(Time, death) ~ 1, data = aids.id, x = TRUE)
fitJM2 <- jointModel(fitLME, fitSURV2, timeVar = "obstime", method = "piecewise-PH-GH")
anova(fitJM2, fitJM) # The model under the null is the first one
```

```
##
##            AIC     BIC  log.Lik  LRT df p.value
## fitJM2 4250.53 4312.72 -2110.26
## fitJM  4247.29 4313.64 -2107.65 5.23  1  0.0222
```

- According to the `pvalue` (as with the Wald test) we arrive to the same conclusion, there exist an affect of the treatment on the risk.

- Additionally, if we want to obtain estimates of the Hazard Ratio with confidence intervals for the final model it is possible ti apply the `confint` function to the created object

```
confint(fitJM, parm = "Event")
```

```
##               2.5 %       est.     97.5 %
## drugddI  0.04979688  0.3511323  0.6524677
## Assoct  -1.33281297 -1.1016129 -0.8704128
```

```
exp(confint(fitJM, parm = "Event"))
```

```
##              2.5 %      est.    97.5 %
## drugddI 1.0510576 1.4206752 1.9202736
## Assoct  0.2637343 0.3323346 0.4187786
```

`jointModel` Arguments

- `method:` Specifies the baseline hazard function, parameterization of the relative risk model, and procedure for numerical integration.
- Available methods:
  - `weibull-PH-aGH` (default)
  - `weibull-PH-GH`
  - `weibull-AFT-aGH`
  - `weibull-AFT-GH`
  - `piecewise-PH-aGH`
  - `piecewise-PH-GH`
  - `spline-PH-aGH` (allows strata)
  - `spline-PH-GH` (allows strata)
  - `Cox-PH-aGH`
  - `Cox-PH-GH`
- `PH`: proportional hazards; `AFT`: accelerated failure time
- `GH` or `aGH`: standard or adaptive Gauss-Hermite quadrature