



Joint Models for Longitudinal & Time-to-Event Data with R

Jinheum Kim
Univ Suwon
2022.9.28

생존자료에 관한 결합모형

모형 확장

예측: 이벤트 발생 확률

○○○ 평가: ROC 곡선, AUC

OUTLINE

이런 질문에 답을 ...

- ▶ 결합모형은 언제 필요한가?
- ▶ 시간-가변 공변량을 가진 Cox 비례위험모형과 다른 점은 무엇인가?
- ▶ 종단자료와 생존자료를 연결하는 방법은 다양한가?
- ▶ 이벤트를 두 개 이상 가진 생존자료와 결합도 가능한가?
- ▶ ...

Mayo Clinic Primary Biliary Cirrhosis (PBC) 자료

- ▶ **id** patients identifier (312)
- ▶ **years** number of years between registration and the **earlier** of **death**, **transplantation**, or **study analysis time**
- ▶ **year** number of years between **enrollment** and **this visit date**
- ▶ **status** a factor with levels **alive (143)**, **transplanted (29)** and **dead (140)**
- ▶ **status2** 1 denoting if the patient was **dead**, and 0 if the patient was **alive** or **transplanted**
- ▶ **drug** a factor with levels **placebo (154)** and **D-penicil (디-페니실린) (158)**
- ▶ **serBilir** serum bilirubin in mg/dl
- ▶ **hepatomegaly (간비대)** a factor with levels **No** and **Yes**
- ▶ **spiders (혈관 기형)** a factor with levels **No** and **Yes**
- ▶ age, sex, ascites, edema, serChol, albumin, alkaline, SGOT, platelets, prothrombin, histologic

질문: 디-페니실린은 효과적인가? 혈청 빌리루빈은 어떻게 영향을 미치는가?

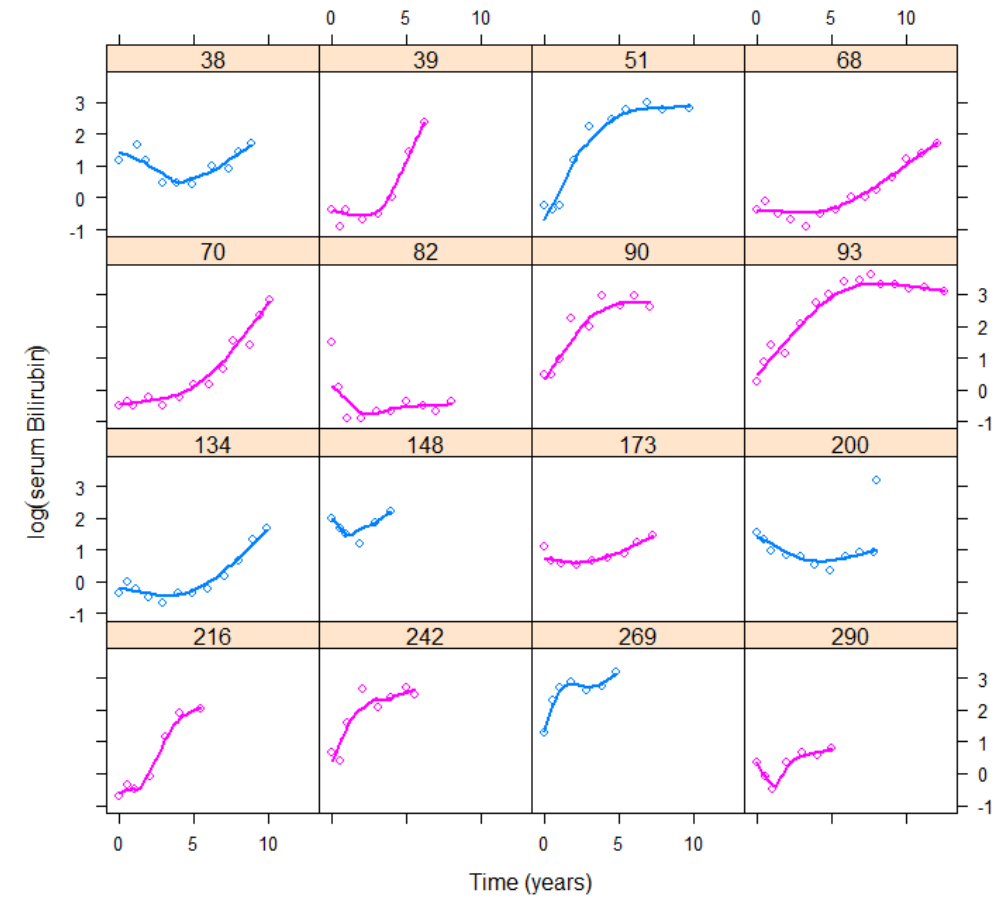
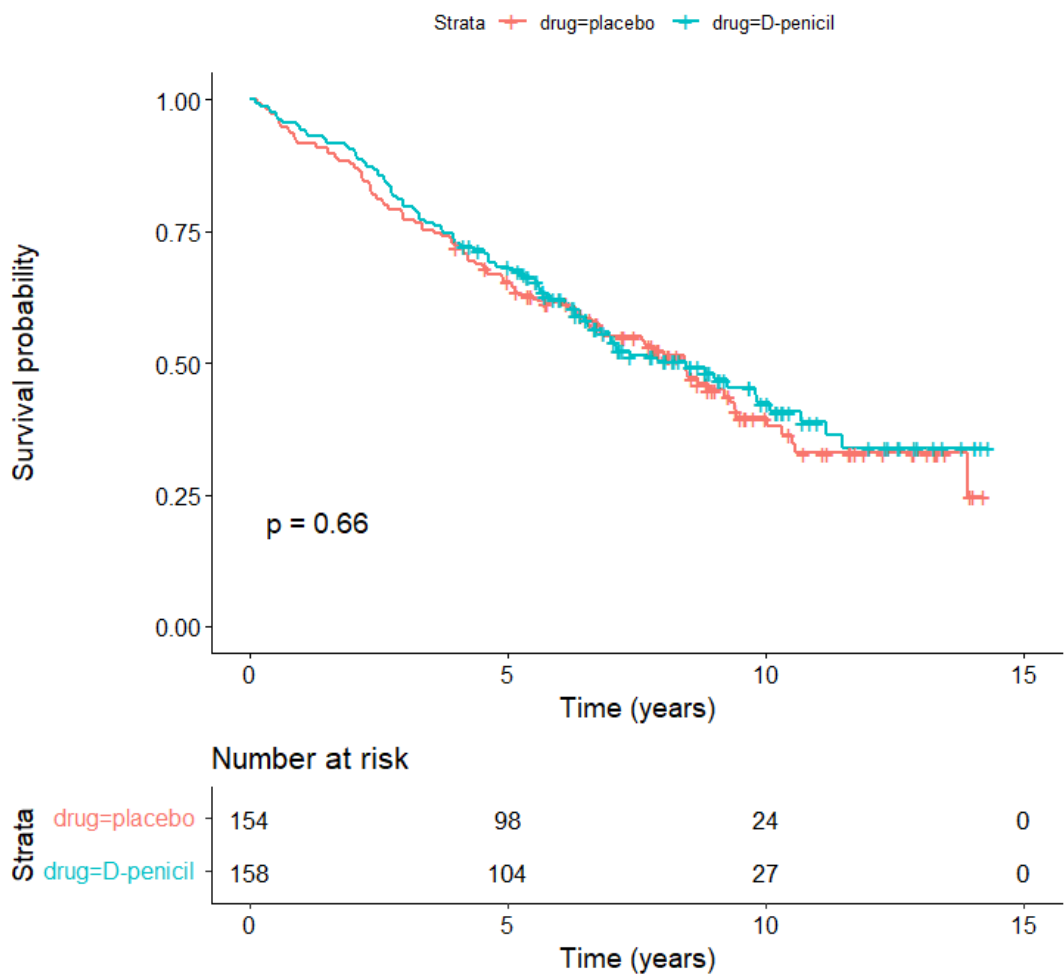
Dataset 비표: pbc2, pbc2.id, pbc

| | id | years | status | drug | year | hepatomegaly | spiders | serBilir | status2 |
|----|----|----------|--------|-----------|-----------|--------------|---------|----------|---------|
| 12 | 3 | 2.770781 | dead | D-penicil | 0.0000000 | No | No | 1.4 | 1 |
| 13 | 3 | 2.770781 | dead | D-penicil | 0.4818749 | Yes | Yes | 1.1 | 1 |
| 14 | 3 | 2.770781 | dead | D-penicil | 0.9966050 | No | No | 1.5 | 1 |
| 15 | 3 | 2.770781 | dead | D-penicil | 2.0342788 | No | Yes | 1.8 | 1 |

| | id | years | status | drug | year | hepatomegaly | spiders | serBilir | status2 |
|---|----|----------|--------|-----------|------|--------------|---------|----------|---------|
| 3 | 3 | 2.770781 | dead | D-penicil | 0 | No | No | 1.4 | 1 |

Counting process 포맷

| | id | hepatomegaly | status2 | tstart | tstop | cEvent | hepato | serBili | spider |
|----|----|--------------|---------|-----------|-----------|--------|--------|---------|--------|
| 12 | 3 | No | 1 | 0.0000000 | 0.4818749 | 0 | No | 1.4 | No |
| 13 | 3 | No | 1 | 0.4818749 | 0.9966050 | 0 | Yes | 1.1 | Yes |
| 14 | 3 | No | 1 | 0.9966050 | 2.0342788 | 0 | No | 1.5 | No |
| 15 | 3 | No | 1 | 2.0342788 | 2.7707809 | 1 | No | 1.8 | Yes |



- 기점, 6개월 후, 1년 후, 년 1회 측정: 총 1945회 측정. 환자 1인당 평균 6.2회
- 파란색: 위약 분홍색: 디-페니실린
- 혈청빌리루빈: 비선형적인 경향도 보임

시간-가변 공변량: 외적

- ▶ 외적공변량이란? 이벤트에 영향을 주지만 이벤트 발생 이후부터는 이벤트에 영향을 받지 않는 공변량
- ▶ PBC: drug, age, sex 등

시간-가변 공변량: 내적

- ▶ 바이오 마커 대부분은 내적공변량
 - ▶ PBC: **serBilir, spiders, hepatomegaly**, ascites, edema, serChol, albumin, alkaline, SGOT, platelets, prothrombin, histologic
- ▶ **측정오차**를 포함
- ▶ 위험집합에 있을 때만 관측 가능
- ▶ 이벤트 발생 여부에 직접적으로 연관

자료구조

- ▶ Data: $D_n = \{T_i, \delta_i, y_i; i = 1, 2, \dots, n\}$,
- ▶ $T_i = \min(T_i^*, C_i)$, T_i^* : 이벤트 시간, C_i : 중도절단 시간
- ▶ $\delta_i = I(T_i^* \leq C_i)$,
- ▶ $y_i = (y_{i1}, y_{i2}, \dots, y_{in_i})'$: $n_i \times 1$ 종단자료값 벡터
 - ▶ $y_{il} = y_i(t_{il})$: 시점 t_{il} ($l = 1, 2, \dots, n_i$)에서 관측한 i -번째 개체의 종단자료값

두 모형: 선형혼합모형

- ▶ 연속형(정규) 바이오 마커
- ▶ $y_i(t) = m_i(t) + \epsilon_i(t) = x_i'(t)\beta + z_i'(t)b_i + \epsilon_i(t)$,
 - ▶ $m_i(t)$: i -번째 개체의 참 종단 값 → 관측할 수 없으며 대신 (측정)오차가 포함된 값만 관측 가능
 - ▶ $x_i(t)$: 외적공변량 벡터
 - ▶ $z_i(t)$: t 의 다항 차수 벡터. 예: $(1, t), (1, t, t^2)$ 등
 - ▶ $\epsilon_i(t) \sim iid N(0, \sigma^2)$: 측정오차
 - ▶ $b_i \sim N(0, D)$: 랜덤효과
 - ▶ b_i 와 $\epsilon_i(t)$ 는 서로 독립

두 모형: Cox 비례위험모형

▶ $h_i(t|w_i) = h_0(t) \exp(\gamma_i' w_i),$

▶ w_i : (기저) 공변량 벡터 (예: 성별, 치료 여부 등) → 시간 - 가변일 수도! 즉 외적공변량 벡터 $w_i(t)$

▶ $h_0(t)$: 비모수적 혹은 **모수적** (예: 와이블분포, 조각지수분포, B-스플라인) 가정

모수적 기저위험함수

- ▶ 와이블분포 : $h_0(t) = \sigma t^{\sigma-1}$
- ▶ 계단함수 :
 - ▶ $0 = v_0 < v_1 < v_2 < \dots < v_Q < v_{Q+1} = \infty$: 노드
 - ▶ $\log h_0(t) = \sum_{q=1}^{Q+1} \xi_q I(v_{q-1} < t \leq v_q)$ → 서로 다른 이벤트를 시점을 노드로 정하면 비모수적 방법
- ▶ B-스플라인:
 - ▶ $\log h_0(t) = \kappa_0 + \sum_{d=1}^{m+1} \kappa_d B_d(t)$
 - ▶ $B_d(t)$: 3차 B-스플라인 기저함수
 - ▶ m : 내부 노드 개수

결합 모형

- ▶ $h_i(t|\tilde{m}_i(t), w_i) = h_0(t) \exp(\gamma' w_i + \alpha m_i(t))$,
 - ▶ $\tilde{m}_i(t) = \{m_i(s), 0 \leq s < t\}$: t 시점 전까지 종단자료값 프로세스
- ▶ α : t 시점에서 바이오 마커의 값과 이벤트 발생 위험 간의 연관 모수 → 관심 있는 가설 $H_0: \alpha = 0$
- ▶ “현재 값” 연관 모형이라고 부름!

추정: 최대우도방법

- ▶ 모수: $\theta = (\theta'_y, \theta'_t, \theta'_b)'$
 - ▶ $\theta_y = (\beta', \sigma^2)'$: 종단자료에 관한 모수
 - ▶ $\theta_t = (\gamma', \alpha, \theta'_{h_0})'$: 이벤트에 관한 모수
 - ▶ $\theta_b = \text{vec}(D)$: 랜덤효과에 관한 모수
- ▶ 가정:
 - ▶ $P(T_i, \delta_i, y_i | b_i; \theta) = P(T_i, \delta_i | b_i; \theta)P(y_i | b_i; \theta)$
 - ▶ $P(y_i | b_i; \theta) = \prod_l P(y_i(t_{il}) | b_i; \theta)$
 - ▶ 중도절단 시간과 방문 시간이 서로 독립

로그우도

- ▶ $\log P(T_i, \delta_i, y_i; \theta) = \log \int P(T_i, \delta_i, y_i, b_i; \theta) db_i$
 $= \log \int P(T_i, \delta_i | b_i; \theta_t, \beta) P(y_i | b_i; \theta_y) P(b_i; \theta_b) db_i$
- ▶ $P(T_i, \delta_i | b_i; \theta_t, \beta) = h_i(T_i | \tilde{m}_i(T_i), w_i; \theta_t, \beta)^{\delta_i} S_i(T_i | \tilde{m}_i(T_i), w_i; \theta_t, \beta)$
- ▶ $S_i(T_i | \tilde{m}_i(T_i), w_i; \theta_t, \beta) = \exp\left(-\int_0^{T_i} h_0(s) \exp(\gamma' w_i + \alpha m_i(s)) ds\right)$
- ▶ $l(\theta) = \sum_i \log P(T_i, \delta_i, y_i; \theta)$: (관측 자료) 로그 우도
- ▶ $l(\theta)$ 를 최대로 하는 해를 직접 구하기 쉽지 않음 → EM 알고리즘 적용
(“ b_i ”를 결측값으로 간주)

EM 알고리즘

- ▶ $\sum_i \log P(T_i, \delta_i, y_i, b_i; \theta)$: (완전자료) 로그 우도
- ▶ $Q(\theta|\theta^{(s)}) = \sum_i E[\log P(T_i, \delta_i, y_i, b_i; \theta) | T_i, \delta_i, y_i; \theta^{(s)}]$
 $= \sum_i \int \log P(T_i, \delta_i, y_i, b_i; \theta) P(b_i | T_i, \delta_i, y_i; \theta^{(s)}) db_i$
: (완전자료) 로그우도의 조건부기댓값 → E-step
- ▶ MLE: $l(\theta)$ 대신 $Q(\theta|\theta^{(s)})$ 를 최대로 하는 해로 정의 →
M-step
 - ▶ $\theta^{(s+1)} = \operatorname{argmax}_{\theta} Q(\theta|\theta^{(s)})$

SE

- ▶ $S(\theta) = \frac{\partial Q(\theta|\theta^{(s)})}{\partial \theta}$: 스코어 벡터
- ▶ $I(\theta) = -\frac{\partial S(\theta)}{\partial \theta'}$: 관찰 자료 정보 행렬
- ▶ $\text{vâr}(\hat{\theta}) = I(\hat{\theta})^{-1}$: $\hat{\theta}$ (MLE)의 분산-공분산 행렬의 추정값

PBC

- ▶ 종단자료모형: $m_i(t) = \beta_0 + \beta_1 \times \text{drug}_i + \beta_2 t + \beta_3 t^2 + \beta_4 (t \times \text{drug}_i) + \beta_5 (t^2 \times \text{drug}_i) + b_{i0} + b_{i1} t + b_{i2} t^2$
- ▶ 결합모형: $h_i(t) = h_0(t) \exp(\gamma_1 \times \text{drug}_i + \gamma_2 \times \text{hepatomegaly}_i + \alpha \times m_i(t))$

↑
내적공변량이지만
기저값으로 고정

| 기저 분포 | 로그 우도 | AIC | BIC |
|-----------|----------------|--------|--------|
| 와이블분포 | -1859.4 | 3754.9 | 3822.3 |
| 조각지수분포 | -1858.1 | 3762.2 | 3848.2 |
| B-스플라인 | -1855.8 | 3761.6 | 3855.2 |
| 콕스 (비모수적) | -2395.6 | 4823.2 | 4883.1 |

```
Call:
jointModel(lmeObject = lmeFit.pbc, survObject = coxFit.pbc, timeVar = "year",
  method = "weibull-PH-aGH")
```

Data Descriptives:

Longitudinal Process Event Process
 Number of Observations: 1945 Number of Events: 169 (54.2%)
 Number of Groups: 312

Joint Model Summary:

Longitudinal Process: Linear mixed-effects model
 Event Process: Weibull relative risk model
 Parameterization: Time-dependent

| log.Lik | AIC | BIC |
|----------|----------|----------|
| -1859.45 | 3754.899 | 3822.273 |

Variance Components:

| | StdDev | Corr |
|-------------|--------|----------------|
| (Intercept) | 0.9974 | (Intr) year |
| year | 0.3084 | 0.1918 |
| I(year^2) | 0.0251 | 0.0279 -0.8561 |
| Residual | 0.3026 | |

Coefficients:

Longitudinal Process

| | Value | Std.Err | z-value | p-value |
|-------------------------|---------|---------|---------|---------|
| (Intercept) | 0.5805 | 0.0739 | 7.8550 | <0.0001 |
| drugD-penicil | -0.1398 | 0.1055 | -1.3246 | 0.1853 |
| year | 0.1654 | 0.0311 | 5.3201 | <0.0001 |
| I(year^2) | 0.0031 | 0.0032 | 0.9639 | 0.3351 |
| drugD-penicil:year | 0.0021 | 0.0428 | 0.0498 | 0.9603 |
| drugD-penicil:I(year^2) | -0.0021 | 0.0043 | -0.4787 | 0.6321 |

Event Process

| | Value | Std.Err | z-value | p-value |
|-----------------|---------|---------|----------|---------|
| (Intercept) | -5.0626 | 0.3258 | -15.5406 | <0.0001 |
| drugD-penicil | 0.0705 | 0.1661 | 0.4248 | 0.6710 |
| hepatomegalyYes | 0.6870 | 0.1802 | 3.8122 | 0.0001 |
| Assoct | 1.2806 | 0.0917 | 13.9604 | <0.0001 |
| log(shape) | 0.1639 | 0.0736 | 2.2274 | 0.0259 |



Event Process

| | Value | Std.Err | z-value | p-value |
|-----------------|---------|---------|----------|---------|
| drugD-penicil | 0.0711 | 0.1666 | 0.4268 | 0.66 |
| hepatomegalyYes | 0.7015 | 0.1813 | 3.8688 | 0.00 |
| Assoct | 1.2794 | 0.0917 | 13.9474 | <0.00 |
| log(xi.1) | -5.0070 | 0.3038 | -16.4790 | |
| log(xi.2) | -4.6472 | 0.3035 | -15.3124 | |
| log(xi.3) | -4.6410 | 0.3164 | -14.6680 | |
| log(xi.4) | -4.4701 | 0.3404 | -13.1337 | |
| log(xi.5) | -4.5040 | 0.3404 | -13.2332 | |
| log(xi.6) | -4.3045 | 0.3688 | -11.6724 | |
| log(xi.7) | -4.8382 | 0.4933 | -9.8070 | |

Event Process

| | Value | Std.Err | z-value | p-value |
|-----------------|---------|---------|----------|---------|
| drugD-penicil | 0.0668 | 0.1667 | 0.4006 | 0.6887 |
| hepatomegalyYes | 0.7001 | 0.1817 | 3.8536 | 0.0001 |
| Assoct | 1.2738 | 0.0914 | 13.9331 | <0.0001 |
| bs1 | -4.6942 | 0.5177 | -9.0676 | <0.0001 |
| bs2 | -5.5616 | 0.6161 | -9.0276 | <0.0001 |
| bs3 | -4.2414 | 0.5931 | -7.1507 | <0.0001 |
| bs4 | -5.0253 | 0.4900 | -10.2563 | <0.0001 |
| bs5 | -4.0881 | 0.4937 | -8.2812 | <0.0001 |
| bs6 | -5.0842 | 0.5873 | -8.6569 | <0.0001 |
| bs7 | -2.3617 | 1.1040 | -2.1393 | 0.0324 |
| bs8 | -8.7391 | 2.2218 | -3.9333 | 0.0001 |
| bs9 | -2.9099 | 1.9146 | -1.5199 | 0.1285 |



PBC: Cox 모형과 비교

- ▶ $h_i(t) = h_0(t) \exp(\gamma_1 \times \text{drug}_i + \gamma_2 \times \text{hepatomegaly}_i + \alpha y_i(t))$ → 방문 사이에 혈청 빌리루빈의 로그값은 변하지 않는다고 가정. **외적공변량**으로 간주 (counting process 포맷)

```
call:
coxph(formula = surv(tstart, tstop, cEvent) ~ drug + hepatomegaly +
      log(serBili), data = pbc)
```

| | coef | exp(coef) | se(coef) | z | p |
|-----------------|---------|-----------|----------|--------|----------|
| drugD-penicil | 0.11216 | 1.11869 | 0.17258 | 0.650 | 0.515767 |
| hepatomegalyYes | 0.72716 | 2.06919 | 0.19082 | 3.811 | 0.000139 |
| log(serBili) | 1.29276 | 3.64283 | 0.08834 | 14.635 | < 2e-16 |

```
Likelihood ratio test=302.2 on 3 df, p=< 2.2e-16
n= 1945, number of events= 140
```

모형 확장: $\alpha m_i(t)$

- ▶ “상호작용” 연관: $\alpha_1 m_i(t) + \alpha_2 (w_i \times m_i(t))$. w_i : 부분그룹 공변량 → interFact = 옵션 이용
- ▶ “지연” 연관: $\alpha m_i \{\max(t - c, 0)\}$, c : 지연된 시간 → lag = 옵션 이용
- ▶ “현재 값 + 변화율” 연관: $\alpha_1 m_i(t) + \alpha_2 m_i'(t)$ → derivForm = 옵션 이용
- ▶ “누적” 연관: $\alpha \int_0^t m_i(s) ds$ → derivForm = 옵션 이용
- ▶ “공유 랜덤효과” 연관: $\alpha' b_i$ (시간 독립 → 간단. 하지만 다변량 랜덤효과의 해석은 난해)

PBC

외적공변량과 내적공변량의 상호작용이 이벤트와 연관

| Event Process | Value | Std.Err | z-value | p-value |
|-------------------------------|---------------|---------------|---------------|---------------|
| drugD-penicil | 0.0891 | 0.1677 | 0.5310 | 0.5954 |
| hepatomegalyYes | 0.2302 | 0.4115 | 0.5595 | 0.5758 |
| Assoct | 1.1572 | 0.1287 | 8.9897 | <0.0001 |
| Assoct:hepatomegalyYes | 0.2349 | 0.1851 | 1.2689 | 0.2045 |
| log(xi.1) | -4.7967 | 0.3300 | -14.5333 | |
| log(xi.2) | -4.4299 | 0.3329 | -13.3078 | |
| log(xi.3) | -4.3988 | 0.3546 | -12.4063 | |
| log(xi.4) | -4.2184 | 0.3797 | -11.1092 | |
| log(xi.5) | -4.2434 | 0.3827 | -11.0876 | |
| log(xi.6) | -4.0574 | 0.4021 | -10.0902 | |
| log(xi.7) | -4.6046 | 0.5149 | -8.9427 | |

혈청 빌리루빈의 로그값이 1-unit 증가함에 따라
이벤트 위험은 **hepatomegaly**가
없는 환자: $\exp(1.157) = 3.18$ (2.47, 4.09)
있는 환자: $\exp(1.392) = 4.02$ (3.11, 5.20) 배 증가

내적공변량의 현재 값과 기울기가 이벤트와 연관

| Event Process | Value | Std.Err | z-value | p-value |
|-----------------|---------------|---------------|---------------|-------------------|
| drugD-penicil | 0.1042 | 0.1787 | 0.5830 | 0.5599 |
| hepatomegalyYes | 0.6231 | 0.1899 | 3.2807 | 0.0010 |
| Assoct | 1.1936 | 0.1043 | 11.4400 | <0.0001 |
| Assoct.s | 2.5427 | 0.5935 | 4.2843 | <0.0001 |
| log(xi.1) | -5.7783 | 0.4082 | -14.1568 | |
| log(xi.2) | -5.2096 | 0.3630 | -14.3499 | |
| log(xi.3) | -5.1178 | 0.3641 | -14.0575 | |
| log(xi.4) | -4.9155 | 0.3840 | -12.8020 | |
| log(xi.5) | -4.9297 | 0.3774 | -13.0636 | |
| log(xi.6) | -4.8116 | 0.4208 | -11.4343 | |
| log(xi.7) | -4.9376 | 0.5116 | -9.6521 | |

혈청 빌리루빈의 로그값이 동일한 두 환자에 대해
혈청 빌리루빈의 로그값의 기울기가 1-unit 증가함에
따라 이벤트 로그위험은 **2.54** (1.38, 3.71) 만큼 증가

모형확장: $\gamma' w_i$

$$\begin{aligned} \blacktriangleright S_i(t | \tilde{m}_i(t), w_i) &= \exp \left(- \int_0^t h_0(s) \exp(\gamma' w_i(s) + \alpha m_i(s)) ds \right) \\ &= \exp \left(- \sum_{q=1}^{Q_i} \int_{\Omega_{iq}} h_0(s) \exp(\gamma' w_{iq} + \alpha m_i(s)) ds \right) \end{aligned}$$

- ▶ $w_i(t)$: 시간-가변 외적공변량
- ▶ $\{\Omega_{iq}, q = 1, 2, \dots, Q_i\}$: $w_i(t)$ 의 값이 일정한 시구간들의 집합

PBC

기저 분포: B-스플라인만 가능

| Event Process | Value | Std.Err | z-value | p-value |
|--------------------------|---------|---------|---------|---------|
| drugD-penicil | -0.0818 | 0.2099 | -0.3898 | 0.6967 |
| spidersYes | 0.2985 | 0.2408 | 1.2394 | 0.2152 |
| drugD-penicil:spidersYes | 0.1678 | 0.3424 | 0.4900 | 0.6241 |
| Assoct | 1.2738 | 0.0899 | 14.1688 | <0.0001 |
| bs1 | -4.2821 | 0.5010 | -8.5466 | <0.0001 |
| bs2 | -5.0865 | 0.5936 | -8.5685 | <0.0001 |
| bs3 | -3.8568 | 0.5761 | -6.6949 | <0.0001 |
| bs4 | -4.6765 | 0.4784 | -9.7755 | <0.0001 |
| bs5 | -3.8088 | 0.4850 | -7.8533 | <0.0001 |
| bs6 | -4.7223 | 0.5843 | -8.0819 | <0.0001 |
| bs7 | -2.2097 | 1.1056 | -1.9986 | 0.0457 |
| bs8 | -7.8001 | 2.1703 | -3.5939 | 0.0003 |
| bs9 | -3.2650 | 2.1701 | -1.5045 | 0.1324 |

내적 공변량 혈철 빌리루빈을 외적 공변량으로 처리하면 ...

```
Call:
coxph(formula = Surv(tstart, tstop, cEvent) ~ drug * spiders +
      log(serBilir), data = pbc)

              coef exp(coef) se(coef)      z      p
drugD-penicil  0.003747  1.003754 0.197032  0.019  0.985
spidersYes     0.261475  1.298845 0.226944  1.152  0.249
log(serBilir)  1.014405  2.757721 0.087376 11.610 <2e-16
drugD-penicil:spidersYes 0.011571  1.011639 0.321580  0.036  0.971

Likelihood ratio test=155.6 on 4 df, p=< 2.2e-16
n= 1945, number of events= 169
```

→ 혈철 빌리루빈의 로그값이 이벤트에 미치는 영향이 과소추정 됐을 수도!

모형확장: $h_0(t)$

- ▶ $h_{ik}(t|M_i(t), w_i) = h_{0k}(t) \exp(\gamma' w_i + \alpha m_i(t))$
 - ▶ $h_{0k}(t)$: k -번째 층의 기저 위험함수
- ▶ $h_{ik}(t|M_i(t), w_i) = h_{0k}(t) \exp(\gamma'_k w_i + \alpha_k m_i(t))$: 기저함수 뿐만 아니라 회귀계수도 층별로 추정

PBC

| Event Process | Value | Std.Err | z-value | p-value |
|---------------|----------|---------|---------|---------|
| drugD-penicil | 0.0571 | 0.1680 | 0.3402 | 0.7337 |
| Assoct | 1.3062 | 0.0949 | 13.7682 | <0.0001 |
| bs1(No) | -3.6190 | 0.7856 | -4.6068 | <0.0001 |
| bs2(No) | -5.7029 | 1.1682 | -4.8818 | <0.0001 |
| bs3(No) | -4.1819 | 1.0626 | -3.9354 | 0.0001 |
| bs4(No) | -5.4619 | 0.8216 | -6.6474 | <0.0001 |
| bs5(No) | -4.5514 | 0.7352 | -6.1909 | <0.0001 |
| bs6(No) | -4.3125 | 0.7663 | -5.6274 | <0.0001 |
| bs7(No) | -3.3626 | 1.5589 | -2.1571 | 0.0310 |
| bs8(No) | -10.0357 | 3.3180 | -3.0246 | 0.0025 |
| bs9(No) | -0.6216 | 1.8042 | -0.3445 | 0.7304 |
| bs1(Yes) | -4.3681 | 0.5807 | -7.5217 | <0.0001 |
| bs2(Yes) | -4.7811 | 0.6696 | -7.1407 | <0.0001 |
| bs3(Yes) | -3.6668 | 0.6605 | -5.5512 | <0.0001 |
| bs4(Yes) | -4.2535 | 0.5658 | -7.5171 | <0.0001 |
| bs5(Yes) | -3.2423 | 0.6151 | -5.2709 | <0.0001 |
| bs6(Yes) | -5.3363 | 0.9190 | -5.8069 | <0.0001 |
| bs7(Yes) | -1.2174 | 1.6719 | -0.7281 | 0.4665 |
| bs8(Yes) | -5.9690 | 3.3670 | -1.7728 | 0.0763 |
| bs9(Yes) | -7.4341 | 7.1605 | -1.0382 | 0.2992 |

Wald Test for Stratification Factors

$\chi^2 = 23.4929$, $df = 9$, $p\text{-value} = 0.0052$

hepatomegaly 여부에 따라 기저분포가 다름

| Event Process | Value | Std.Err | z-value | p-value |
|-------------------------------|---------|---------|---------|---------|
| drugD-penicil | -0.0852 | 0.2866 | -0.2974 | 0.7662 |
| drugD-penicil:hepatomegalyYes | 0.2251 | 0.3526 | 0.6383 | 0.5233 |
| Assoct | 1.1945 | 0.1327 | 9.0003 | <0.0001 |
| Assoct:hepatomegalyYes | 0.1868 | 0.1869 | 0.9997 | 0.3174 |
| bs1(No) | -3.4326 | 0.8080 | -4.2484 | <0.0001 |
| bs2(No) | -5.5074 | 1.1902 | -4.6274 | <0.0001 |
| bs3(No) | -3.9271 | 1.0843 | -3.6217 | 0.0003 |
| bs4(No) | -5.1285 | 0.8417 | -6.0927 | <0.0001 |
| bs5(No) | -4.1506 | 0.7601 | -5.4603 | <0.0001 |
| bs6(No) | -3.9497 | 0.7745 | -5.0995 | <0.0001 |
| bs7(No) | -3.0390 | 1.5770 | -1.9270 | 0.0540 |
| bs8(No) | -9.6079 | 3.3373 | -2.8789 | 0.0040 |
| bs9(No) | -0.5535 | 1.9135 | -0.2893 | 0.7724 |
| bs1(Yes) | -4.5913 | 0.6277 | -7.3143 | <0.0001 |
| bs2(Yes) | -5.0165 | 0.7104 | -7.0616 | <0.0001 |
| bs3(Yes) | -3.9365 | 0.7112 | -5.5353 | <0.0001 |
| bs4(Yes) | -4.3773 | 0.5961 | -7.3432 | <0.0001 |
| bs5(Yes) | -3.3734 | 0.6411 | -5.2618 | <0.0001 |
| bs6(Yes) | -5.4248 | 0.9225 | -5.8808 | <0.0001 |
| bs7(Yes) | -1.5521 | 1.6975 | -0.9143 | 0.3605 |
| bs8(Yes) | -6.0397 | 3.3582 | -1.7985 | 0.0721 |
| bs9(Yes) | -7.5421 | 6.9472 | -1.0856 | 0.2776 |

**처치 여부가 이벤트에 미치는 영향이나
혈청 빌리루빈의 로그값이 이벤트에 미치는 영향이
hepatomegaly 여부에 따라 다르지 않음**

모형 확장: 경쟁위험

- ▶ $T_i = \min(T_1^*, T_2^*, \dots, T_K^*, C_i), \delta_i \in \{0, 1, 2, \dots, K\}$
- ▶ $h_{ik}(t | \tilde{m}_i(t), w_i) = h_{0k}(t) \exp(\gamma'_k w_i + \alpha_k m_i(t)),$
 $k = 1, 2, \dots, K \rightarrow$ “원인별” 위험함수
- ▶ $P(T_i, \delta_i | b_i; \theta_t, \beta)$
 $= \prod_{k=1}^K [h_{0k}(T_i) \exp(\gamma'_k w_i + \alpha_k m_i(T_i))]^{I(\delta_i=k)}$
 $\times \exp\left(-\sum_{k=1}^K \int_0^{T_i} h_{0k}(s) \exp(\gamma'_k w_i + \alpha_k m_i(s)) ds\right)$

PBC

- ▶ 생존자료 long 포맷 변환: crLong 함수 이용
- ▶ 경쟁위험을 층(CR)으로 정의

| | id | years | status | status2 | | id | years | CR | status | status2 |
|----|----|-----------|--------------|---------|---|------|-------|-----------|--------------|---------|
| 1 | 1 | 1.095170 | dead | 1 | → | 1 | 1 | 1.095170 | dead | 1 |
| 3 | 2 | 14.152338 | alive | 0 | | 1.1 | 1 | 1.095170 | transplanted | 0 |
| 23 | 5 | 4.120578 | transplanted | 1 | | 3 | 2 | 14.152338 | dead | 0 |
| | | | | | | 3.1 | 2 | 14.152338 | transplanted | 0 |
| | | | | | | 23 | 5 | 4.120578 | dead | 0 |
| | | | | | | 23.1 | 5 | 4.120578 | transplanted | 1 |

PBC

“현재 값” 연관

| Event Process | Value | Std.Err | z-value | p-value |
|----------------------|---------------|---------------|---------------|---------------|
| drugD-penicil | -0.3929 | 0.3874 | -1.0142 | 0.3105 |
| age | -0.0859 | 0.0245 | -3.5076 | 0.0005 |
| drugD-penicil:CRdead | 0.3719 | 0.4240 | 0.8772 | 0.3804 |
| age:CRdead | 0.1524 | 0.0259 | 5.8815 | <0.0001 |
| Assoct | 1.0223 | 0.1920 | 5.3230 | <0.0001 |
| Assoct:CRdead | 0.4477 | 0.2185 | 2.0492 | 0.0404 |
| bs1(transplanted) | -4.7910 | 4.2890 | -1.1170 | 0.2640 |
| bs2(transplanted) | -3.9373 | 2.4255 | -1.6233 | 0.1045 |
| bs3(transplanted) | 0.1662 | 1.7477 | 0.0951 | 0.9242 |
| bs4(transplanted) | -2.3629 | 1.4417 | -1.6390 | 0.1012 |
| bs5(transplanted) | 0.1909 | 1.4245 | 0.1340 | 0.8934 |
| bs6(transplanted) | -2.2313 | 1.8368 | -1.2147 | 0.2245 |
| bs7(transplanted) | -4.2120 | 5.5095 | -0.7645 | 0.4446 |
| bs8(transplanted) | -4.1632 | 10.7076 | -0.3888 | 0.6974 |
| bs9(transplanted) | -3.7348 | 13.6952 | -0.2727 | 0.7851 |
| bs1(dead) | -8.0858 | 0.7652 | -10.5665 | <0.0001 |
| bs2(dead) | -8.7640 | 0.8356 | -10.4882 | <0.0001 |
| bs3(dead) | -7.9724 | 0.8542 | -9.3336 | <0.0001 |
| bs4(dead) | -8.3966 | 0.7484 | -11.2188 | <0.0001 |
| bs5(dead) | -8.1438 | 0.7916 | -10.2884 | <0.0001 |
| bs6(dead) | -8.2087 | 0.8001 | -10.2591 | <0.0001 |
| bs7(dead) | -6.1988 | 1.2686 | -4.8862 | <0.0001 |
| bs8(dead) | -9.7725 | 2.1035 | -4.6458 | <0.0001 |
| bs9(dead) | -6.9024 | 2.2717 | -3.0384 | 0.0024 |

혈청 빌리루빈의 로그값이 1-unit 증가함에 따라
 이식 위험: $\exp(1.02) = 2.78 (1.91, 4.05)$
 사망 위험: $\exp(1.47) = 4.35 (3.51, 5.40)$ 배 증가

“현재 값 + 기울기” 연관

| Event Process | Value | Std.Err | z-value | p-value |
|------------------------|---------------|---------------|---------------|---------------|
| drugD-penicil | -0.3226 | 0.3946 | -0.8175 | 0.4137 |
| age | -0.0861 | 0.0247 | -3.4889 | 0.0005 |
| drugD-penicil:CRdead | 0.2953 | 0.4319 | 0.6838 | 0.4941 |
| age:CRdead | 0.1488 | 0.0262 | 5.6832 | <0.0001 |
| Assoct | 0.9397 | 0.2127 | 4.4172 | <0.0001 |
| Assoct:CRdead | 0.4383 | 0.2409 | 1.8189 | 0.0689 |
| Assoct.s | 1.8563 | 1.6997 | 1.0922 | 0.2748 |
| Assoct.s:CRdead | 0.6658 | 1.8147 | 0.3669 | 0.7137 |
| bs1(transplanted) | -5.1606 | 4.0548 | -1.2727 | 0.2031 |
| bs2(transplanted) | -4.2498 | 2.4647 | -1.7243 | 0.0847 |
| bs3(transplanted) | -0.4222 | 1.7978 | -0.2348 | 0.8143 |
| bs4(transplanted) | -2.6866 | 1.4947 | -1.7975 | 0.0723 |
| bs5(transplanted) | -0.0716 | 1.4489 | -0.0494 | 0.9606 |
| bs6(transplanted) | -2.4460 | 1.8287 | -1.3376 | 0.1810 |
| bs7(transplanted) | -4.2892 | 5.3560 | -0.8008 | 0.4232 |
| bs8(transplanted) | -4.3283 | 10.8923 | -0.3974 | 0.6911 |
| bs9(transplanted) | -3.9369 | 14.8621 | -0.2649 | 0.7911 |
| bs1(dead) | -8.8341 | 0.8982 | -9.8358 | <0.0001 |
| bs2(dead) | -9.3948 | 0.9110 | -10.3126 | <0.0001 |
| bs3(dead) | -8.4760 | 0.9193 | -9.2205 | <0.0001 |
| bs4(dead) | -8.5633 | 0.7917 | -10.8160 | <0.0001 |
| bs5(dead) | -8.4384 | 0.8407 | -10.0380 | <0.0001 |
| bs6(dead) | -8.4259 | 0.8510 | -9.9007 | <0.0001 |
| bs7(dead) | -6.5109 | 1.3538 | -4.8092 | <0.0001 |
| bs8(dead) | -9.6438 | 2.2693 | -4.2497 | <0.0001 |
| bs9(dead) | -6.9295 | 2.4329 | -2.8482 | 0.0044 |

혈청 빌리루빈의 로그값이 동일한 두 환자에 대해
 혈청 빌리루빈의 로그값의 기울기가 1-unit 증가함에 따라
 이식 로그-위험: **1.86 (-1.47, 5.19)**
 사망 로그-위험: **2.52 (1.29, 3.75)**
 만큼 증가

| | AIC | BIC | log.Lik | LRT | df | p.value |
|-------------|----------------|----------------|-----------------|--------------|----------|--------------|
| valueOnly | 3866.87 | 4005.36 | -1896.43 | | | |
| both | 3852.11 | 3998.08 | -1887.05 | 18.76 | 2 | 1e-04 |

모형 확장

- ▶ 종단자료 모형의 확장
 - ▶ 범주형 종단자료
 - ▶ 이변량(혹은 다변량) 종단자료
- ▶ 생존자료 모형의 확장
 - ▶ 다중(순서 없음) 이벤트자료
 - ▶ 반복(순서 있음) 이벤트자료: 종단 이벤트 유무
 - ▶ 다중상태자료

이벤트 확률 예측

- ▶ 새로운 개체 j 에 대해

$$\pi_j(u|t) = P(T_j^* \geq u | T_j^* > t, \tilde{y}_j(t), D_n), t < u$$

를 예측하고자 함

- ▶ $\tilde{y}_j(t) = \{y_j(s); 0 \leq s < t\}$

베이지안 방법

- ▶ $\pi_j(u|t) = \int P(T_j^* \geq u | T_j^* > t, \tilde{y}_j(t), \theta) p(\theta | D_n) d\theta$
- ▶ $P(T_j^* \geq u | T_j^* > t, \tilde{y}_j(t), \theta)$
 $= \int P(T_j^* \geq u | T_j^* > t, b_j, \theta) p(b_j | T_j^* > t, \tilde{y}_j(t); \theta) db_j$
 $= \int \frac{S_j(u | \tilde{m}_j(u), b_j, \theta)}{S_j(t | \tilde{m}_j(t), b_j, \theta)} p(b_j | T_j^* > t, \tilde{y}_j(t); \theta) db_j$

몬테 칼로 방법

- ▶ step 1: $\theta^{(l)} \sim N(\hat{\theta}, \text{var}(\hat{\theta}))$ 을 생성
- ▶ step 2: $b_i^{(l)} \sim \{b_j | T_j^* > t, \tilde{y}_j(t), \theta^{(l)}\}$ 을 생성
 - ▶ 사후분포로부터 랜덤효과를 생성하기 위해 메트로폴리스-헤이스팅 알고리즘 적용
- ▶ step 3: $\pi_j^{(l)}(u|t) = \frac{s_j(u|\tilde{m}_j(u), b_j^{(l)}, \theta^{(l)})}{s_j(t|\tilde{m}_j(t), b_j^{(l)}, \theta^{(l)})}$ 을 계산
- ▶ step 4: step 1 ~ step 3을 $l (1, 2, \dots, L)$ 번 반복 →
 $\hat{\pi}_j(u|t) = \text{median} \{ \pi_j^{(l)}(u|t) \}$ 혹은 $\hat{\pi}_j(u|t) = \frac{1}{L} \sum_{l=1}^L \pi_j^{(l)}(u|t)$

PBC: #20번과 동일한 모형

▶ 종단자료모형:

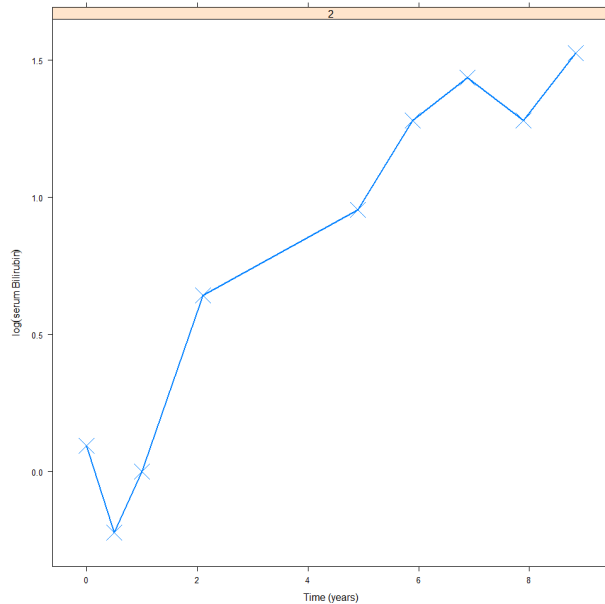
$$\text{▶ } m_i(t) = \beta_0 + \beta_1 \times \text{drug}_i + \beta_2 t + \beta_3 t^2 + \beta_4 (t \times \text{drug}_i) + \beta_5 (t^2 \times \text{drug}_i) + b_{i0} + b_{i1} t + b_{i2} t^2$$

▶ 결합모형:

$$\text{▶ } h_i(t) = h_0(t) \exp(\gamma_1 \times \text{drug}_i + \gamma_2 \times \text{hepatomegaly}_i + \alpha \times m_i(t))$$

PBC

| id | years | status | year | serBilir | drug | hepatomegaly |
|----|-------|----------|-------|-----------|---------------|--------------|
| 3 | 2 | 14.15234 | alive | 0.000000 | 1.1 D-penicil | Yes |
| 4 | 2 | 14.15234 | alive | 0.4983025 | 0.8 D-penicil | Yes |
| 5 | 2 | 14.15234 | alive | 0.9993429 | 1.0 D-penicil | Yes |
| 6 | 2 | 14.15234 | alive | 2.1027270 | 1.9 D-penicil | Yes |
| 7 | 2 | 14.15234 | alive | 4.9008871 | 2.6 D-penicil | Yes |
| 8 | 2 | 14.15234 | alive | 5.8892783 | 3.6 D-penicil | Yes |
| 9 | 2 | 14.15234 | alive | 6.8858833 | 4.2 D-penicil | Yes |
| 10 | 2 | 14.15234 | alive | 7.8907020 | 3.6 D-penicil | Yes |
| 11 | 2 | 14.15234 | alive | 8.8325485 | 4.6 D-penicil | Yes |



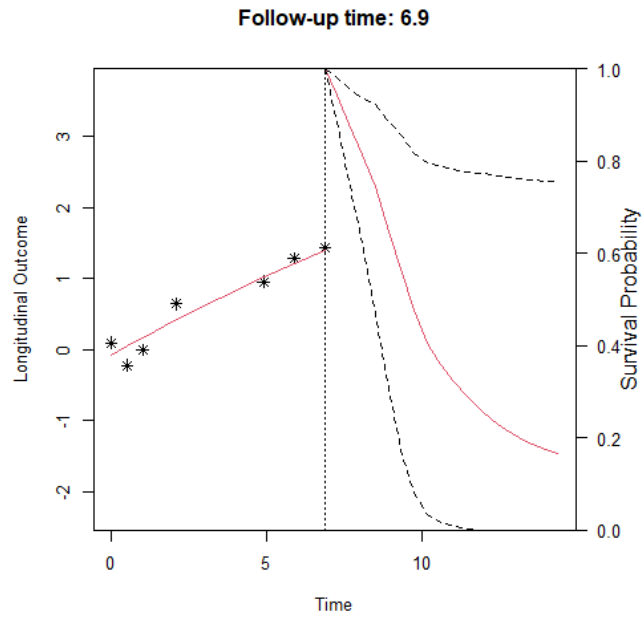
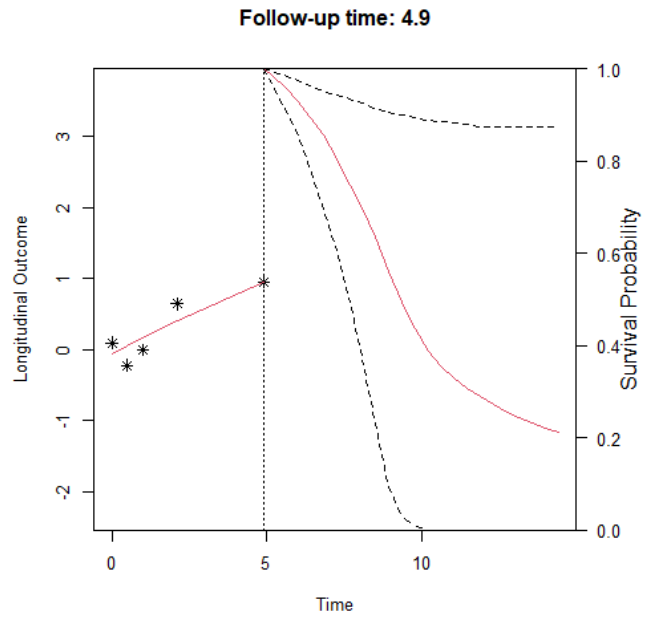
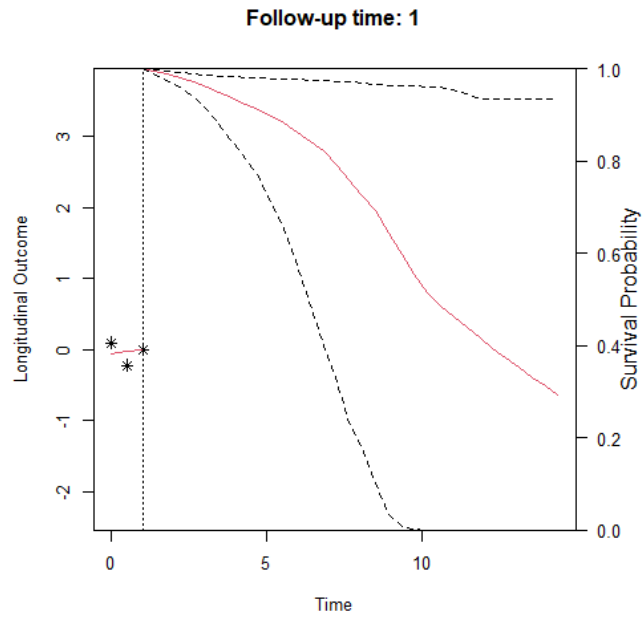
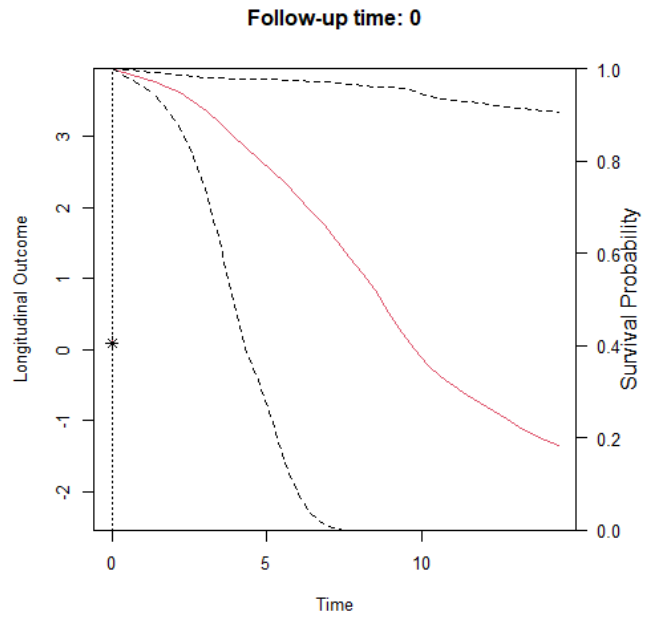
$\pi_2(u|t = 8.8325)$ 에 관한 예측

| | times | Mean | Median | Lower | Upper |
|----|---------|--------|--------|--------|--------|
| 1 | 8.8325 | 1.0000 | 1.0000 | 1.0000 | 1.0000 |
| 1 | 8.9405 | 0.9750 | 0.9774 | 0.9480 | 0.9921 |
| 2 | 9.3609 | 0.8816 | 0.8930 | 0.7558 | 0.9628 |
| 3 | 9.7813 | 0.7946 | 0.8129 | 0.5775 | 0.9363 |
| 4 | 10.2017 | 0.7146 | 0.7383 | 0.4180 | 0.9126 |
| 5 | 10.6221 | 0.6612 | 0.6831 | 0.3489 | 0.8763 |
| 6 | 11.0425 | 0.6176 | 0.6435 | 0.2982 | 0.8643 |
| 7 | 11.4629 | 0.5766 | 0.6011 | 0.2212 | 0.8557 |
| 8 | 11.8833 | 0.5385 | 0.5531 | 0.1788 | 0.8495 |
| 9 | 12.3037 | 0.5035 | 0.5130 | 0.1195 | 0.8480 |
| 10 | 12.7241 | 0.4716 | 0.4786 | 0.0692 | 0.8459 |
| 11 | 13.1445 | 0.4431 | 0.4422 | 0.0315 | 0.8379 |
| 12 | 13.5649 | 0.4178 | 0.4106 | 0.0112 | 0.8321 |
| 13 | 13.9853 | 0.3956 | 0.3862 | 0.0029 | 0.8304 |
| 14 | 14.4057 | 0.3764 | 0.3552 | 0.0008 | 0.8292 |

$\pi_2(u|t = 14.1523)$ 에 관한 예측

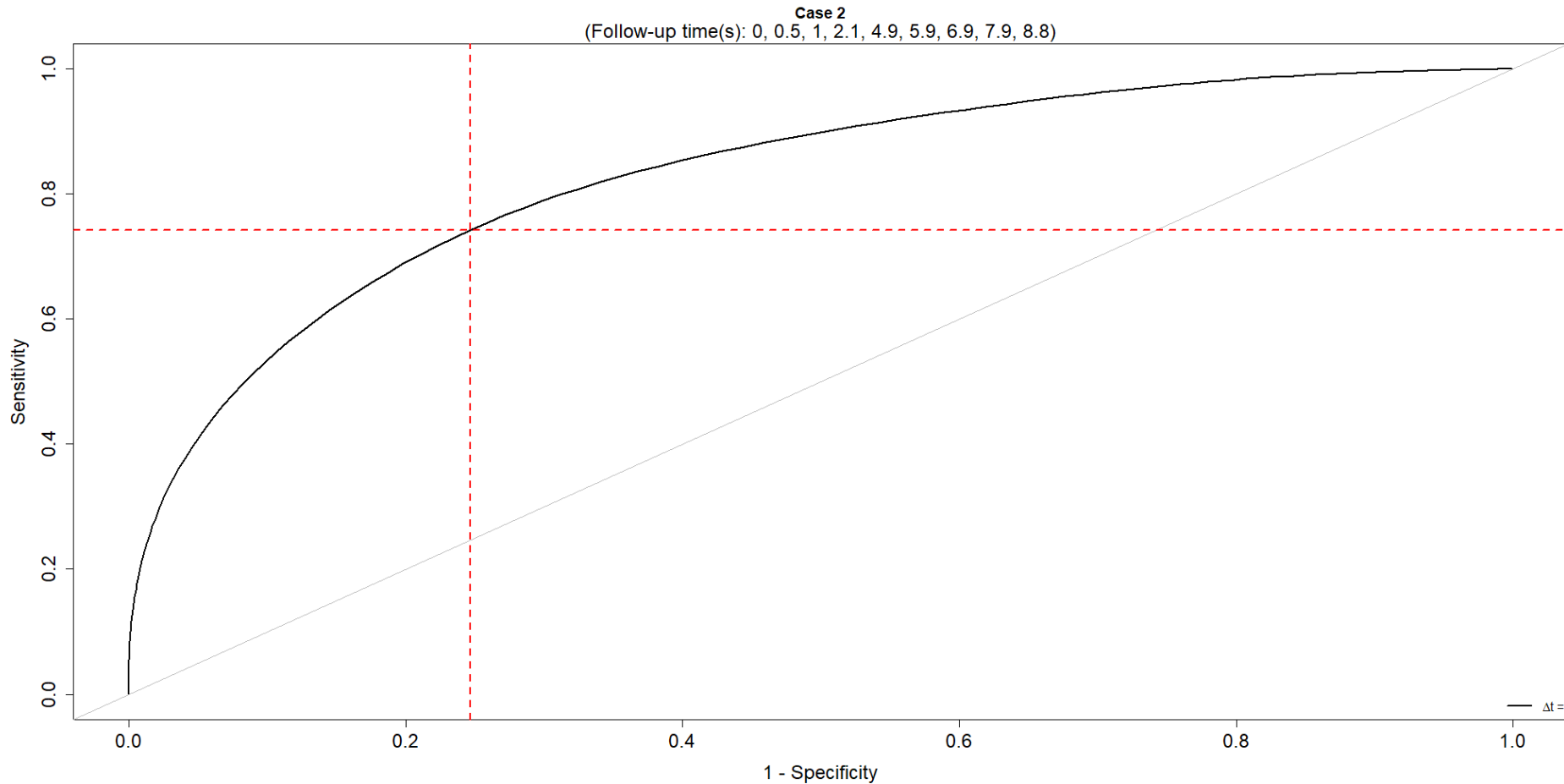
| | times | Mean | Median | Lower | Upper |
|---|---------|--------|--------|--------|--------|
| 1 | 14.1523 | 1.0000 | 1.0000 | 1.0000 | 1.0000 |
| 1 | 14.4057 | 0.9535 | 0.9815 | 0.7197 | 0.9992 |

동적 예측 확률:
 $t = 0, 1, 4.9, 6.9$



○○○ (시간-가변) ROC 곡선과 AUC ○○○

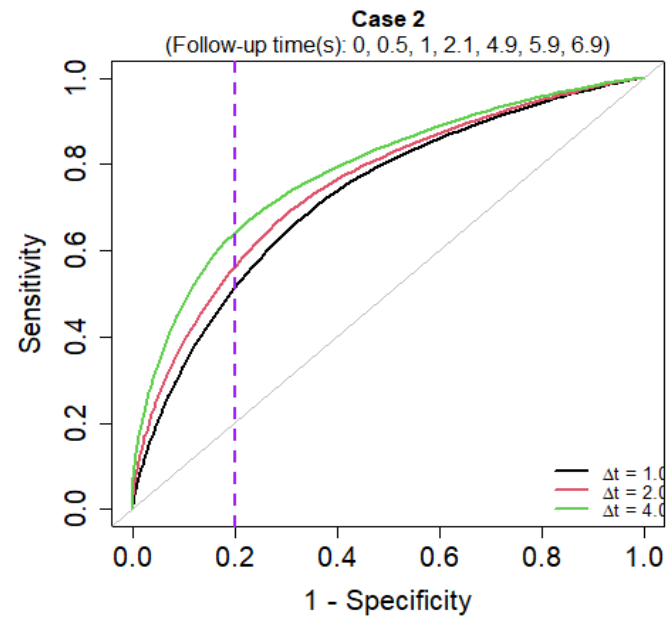
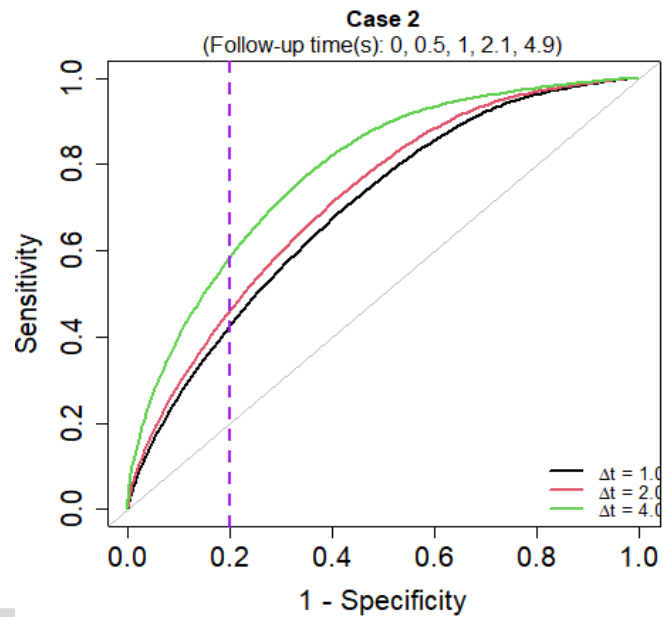
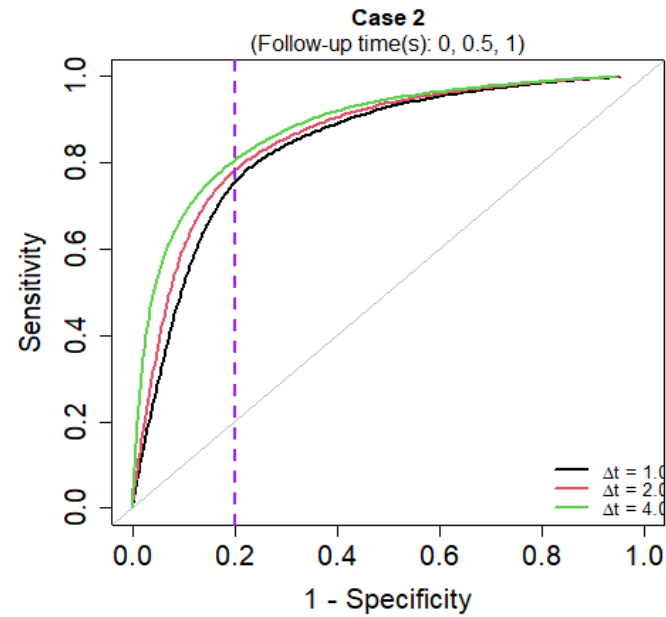
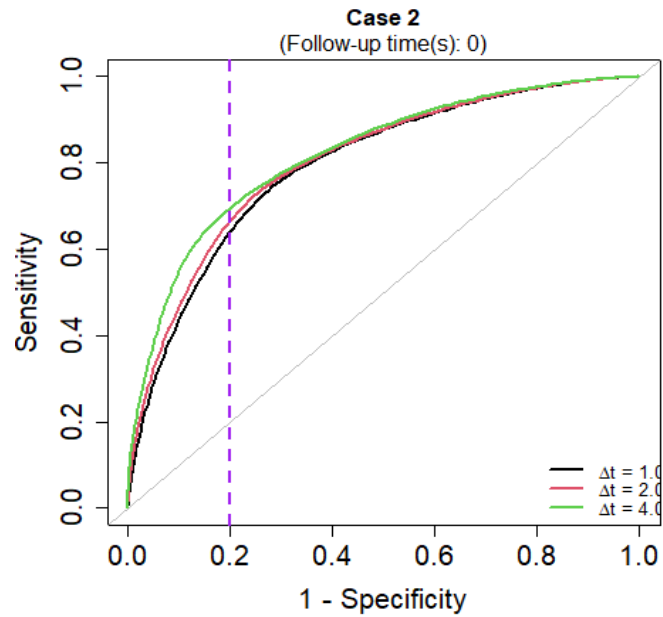
- ▶ 고정된 $t, \Delta t$ 와 R_y (바이오 마커의 표본공간)에 속하는
경계값 c 에 대해,
- ▶ 민감도: $TP_t^{\Delta t}(c) = P(y_i(t) \geq c | T_i^* > t, T_i^* \in (t, t + \Delta t]; \theta)$
- ▶ 특이도: $1 - FP_t^{\Delta t}(c) = P(y_i(t) < c | T_i^* > t, T_i^* > t + \Delta t; \theta)$
- ▶ ROC 곡선: $ROC_t^{\Delta t}(p) = TP_t^{\Delta t} \left\{ [FP_t^{\Delta t}]^{-1}(p) \right\}, 0 < p < 1$
 - ▶ $[FP_t^{\Delta t}]^{-1}(p) = \inf_c \{ FP_t^{\Delta t}(c) \leq p \}$
- ▶ AUC: $AUC_t^{\Delta t} = \int_0^1 ROC_t^{\Delta t}(p) dp \rightarrow \text{rocJM} \{JM\}$ 함수 이용



Case: 2

Recorded time(s): 0, 0.5, 1, 2.1, 4.9, 5.89, 6.89, 7.89, 8.83

| dt | t + dt | AUC | Cut |
|----|--------|--------|------|
| 4 | 12.83 | 0.8233 | 1.33 |



주요 R 패키지

- ▶ **JM**: 연속형 종단자료. ML 추정 \Leftrightarrow %JM 매크로 (**SAS**)
- ▶ **JMBayes**: 범주형 종단자료. 베이지안 추정
- ▶ **JMcmprsk**: 경쟁위험자료. 연속형/범주형 종단자료 \Leftrightarrow %SPM 매크로 (**SAS**)
- ▶ **jointRML**: 다변량 종단자료

참고문헌

▶ Books:

- ▶ Rizopoulos D (2012). Joint Models for Longitudinal and Time-to-Event Data With Applications in R. CRC press.

▶ Review papers:

- ▶ Hickey, G. et al. (2018). Joint Models of Longitudinal and Time-to-Event Data with More Than One Event Time Outcome: A Review. The International Journal of Biostatistics, 14, 1-19.
- ▶ Król, A. et al. (2017). Tutorial in Joint Modeling and Prediction: A Statistical Software for Correlated Longitudinal Outcomes, Recurrent Events and a Terminal Event. Journal of Statistical Software, 81, 1-52.



THANK YOU!!!