Analyzing Survival Data with Competing Risks Using R

Jinheum Kim
Univ Suwon
2022.3.23

OUTLINE

경쟁위험자료란?

누적발생함수의 추정

비례위험 모형

좌절단 자료

시간변동 공변량 자료

다변량 생존자료의 분류

자료 유형	순서 여부	
	Yes	No
동일한 유형	재발사건자료	군집자료
다른 유형	준경쟁위험자료	경쟁위험자료

Melanoma data

- Data on patients after operation for malignant melanoma collected at Odense University Hospital
- cause: 0: alive, 1: dead from melanoma, 2: dead from other cause
- time: survival time
- ulcer: ulceration (0: absent, 1: present)
- thickness: tumour thickness (mm)
- sex: 0: female, 1: male
- age

available at library(MASS)

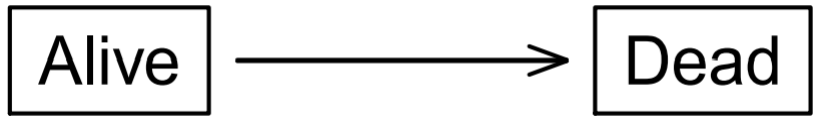
Hospital acquired pneumonia (HAP) in ICU data

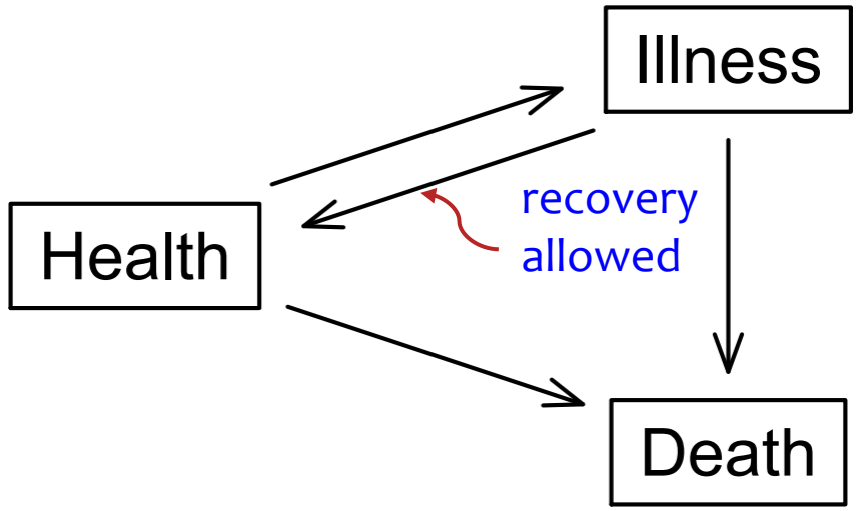
- Assess the impact of HAP on ICU mortality
- id
- start: start of the observation time
- stop: failure time
- status: censoring status
- event: event type (2: death in ICU, 3: discharge alive)
- pneu: nosocomial pneumonia indicator (time-varying)
- adm.cens.exit: administrative censoring times
- age
- sex

available at library(kmi)

Multi-state model

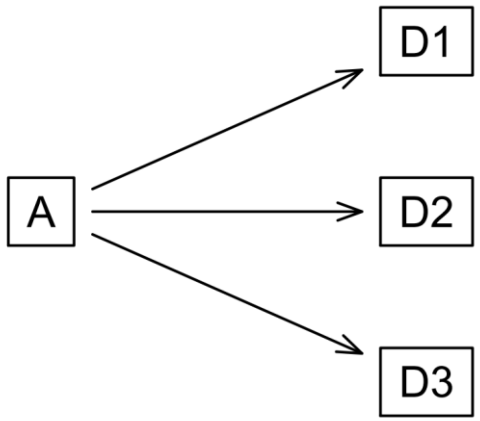
- $K(\geq 2)$개의 상태를 가진 모형
- 생존자료 모형: 두 가지 상태(0: alive, 1: dead)를 가진 모형
  - 1: absorbing (흡수) state
- Illness-death model: 3 states (0: health, 1: illness, 2: death)





Competing risks model

- absorbing state가 여러 개인 multi-state model
- 한 사건의 발생은 다른 사건의 발생을 중도절단시킴
- Remarks
  - 어떤 사건의 발생이 다른 사건의 발생을 중도절단시키지만, 그 역은 성립하지 않는 경쟁위험을 준경쟁위험 (semi-competing risks) 이라고 함

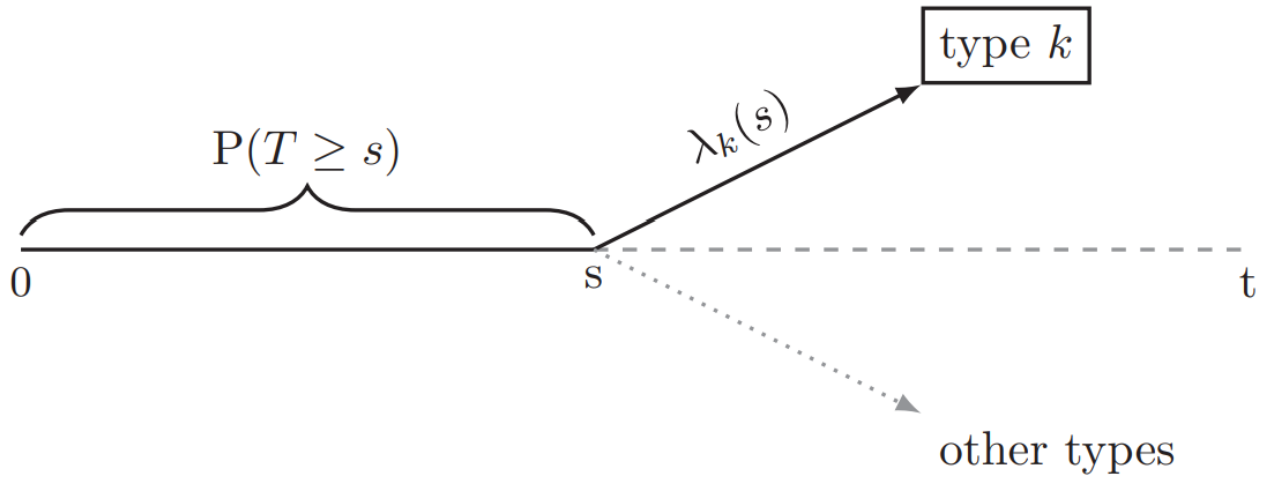


Cause-specific hazard (CSH)

- $(T = \min(T_1, T_2), e \in \{1,2\})$
  - T_k: k −번째 원인에 의한 사건 발생 시간
- $\lambda_k(t) = \lim_{\Delta t \downarrow 0} \frac{P(t \le T < t+\Delta t,\ e=k \mid T \ge t)}{\Delta t}, k = 1, 2$: CSH
- $\lambda(t) = \sum_{k=1}^{2} \lambda_k(t) = \lim_{\Delta t \downarrow 0} \frac{P(t \le T < t+\Delta t \mid T \ge t)}{h}$: overall hazard
- $\Lambda_k(t) = \int_0^t \lambda_k(s) ds$: cumulative CSH (CCSH)

Cumulative incidence function (CIF)

- $F_k(t) = P(T \leq t, e = k) = \int_0^t S(s-)\lambda_k(s)ds$: CIF. t 이내에 k –번째 원인에 의해 사건이 발생할 확률
  - $S(t) = P(T > t) = \exp\left\{-\int_0^t \lambda(s)ds\right\}$: overall survival function



CIF

- Remarks

  refer to #8

  - $\lambda_k(t) = \frac{F_k'(t)}{S(t)} \Rightarrow F_k'(t) = S(t)\lambda_k(t)$
  - $F_1(t) + F_2(t) = 1 - S(t), \forall t$
  - $F_k(\infty) = P(e = k) < 1$ 이므로 F_k를 “subdistribution function”이라고 함

Observed data

- $\{(x_1, e_1\delta_1), \ldots, (x_n, e_n\delta_n)\}$
  - $x_i = \min(t_i, c_i), \delta_i = I(t_i \leq c_i), e_i \in \{1,2\}$
  - c_i: 중도절단 시간 (only known under administrative censoring)
- $t_{(1)} < \cdots < t_{(r)}$: unique ordered uncensored times
- d_{ik}: # of patients failing from cause k at $t_{(i)}$
- $d_i\left(= \sum_{k=1}^{2} d_{ik}\right)$: total # of patients failing from any cause at $t_{(i)}$
- n_i: # of patients at risk at $t_{(i)}$

Nelson-Aalen (NA) estimator

- $\hat{\lambda}_k\left(t_{(i)}\right) = \frac{d_{ik}}{n_i}$: estimated CSH
- $\hat{\Lambda}_k(t) = \sum_{t_{(i)} \le t} \hat{\lambda}_k\left(t_{(i)}\right)$: NA estimator

Aalen-Johansen (AJ) estimator

- $\hat{F}_k(t) = \sum_{m:t_{(m)} \le t} \hat{S}(t_{(m)} -) \frac{d_{mk}}{n_m}$: estimated CIF
- $\hat{S}(t_{(m)} -) = \sum_{l:t_{(l)} < t_{(m)}} \left(1 - \frac{d_l}{n_l}\right)$: PL estimator of the probability of being free of any event just prior to $t_{(m)}$

Remarks on CIF estimation

- 만약 competing risks를 중도절단으로 간주한다면 $F_1(t)$를 과대 추정함
  - $S_1(t) = \exp\left\{-\int_0^t \lambda_1(s)ds\right\}$: probability of being free from cause 1 just prior to t
  - $S_1(t) = \exp\left\{-\int_0^t \lambda_1(s)ds\right\} \geq \exp\left\{-\int_0^t \left(\lambda_1(s) + \lambda_2(s)\right)ds\right\} = S(t)$

    $\Rightarrow \int_0^t S_1(s-)\lambda_1(s)ds \geq \int_0^t S(s-)\lambda_1(s)ds = F_1(t)$

Melanoma data

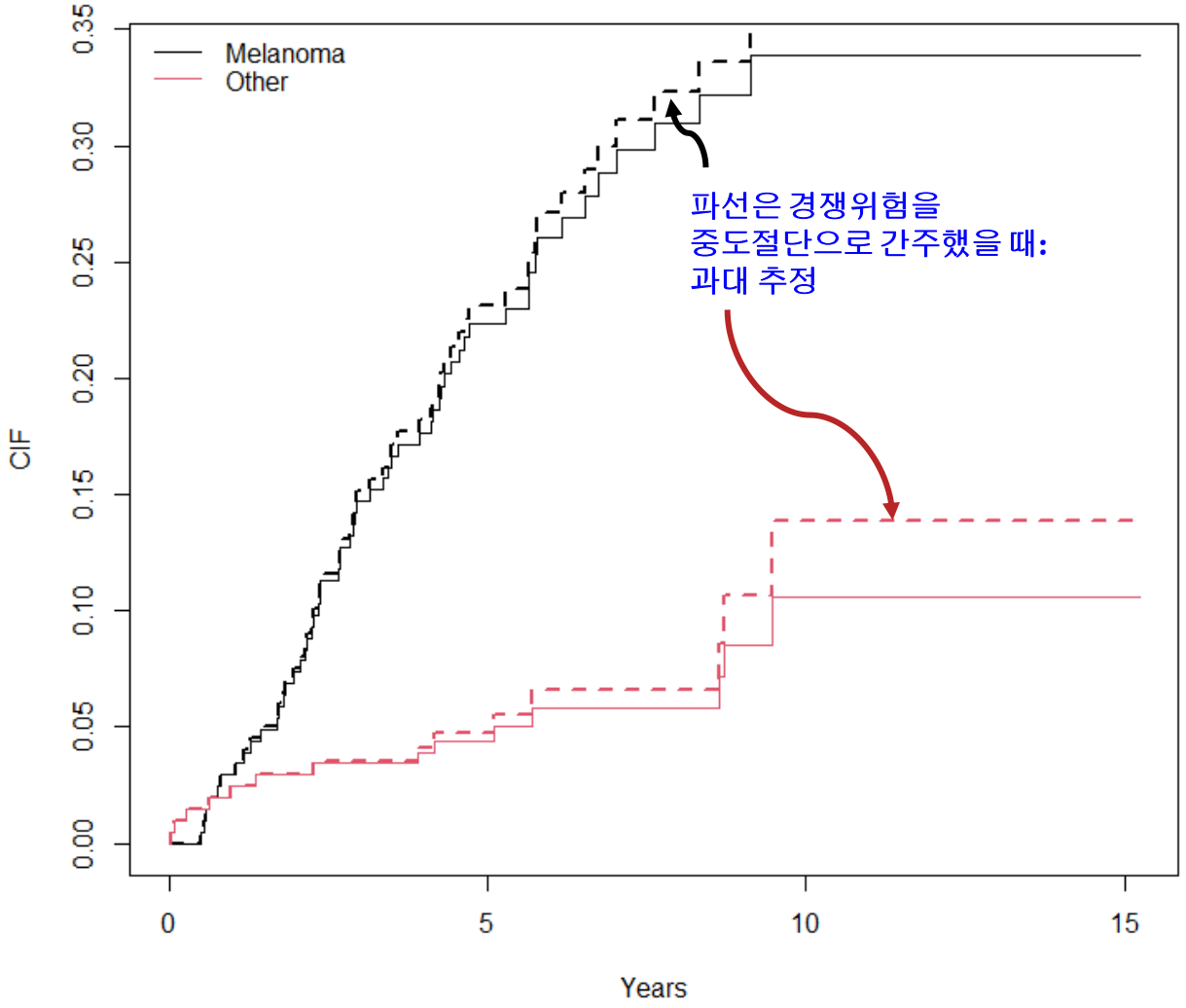
- 피부암 환자 256명을 대상으로 사망 원인을 조사:
- 중도절단: 134 (0), 피부암: 57 (1), 다른 원인: 14 (2)

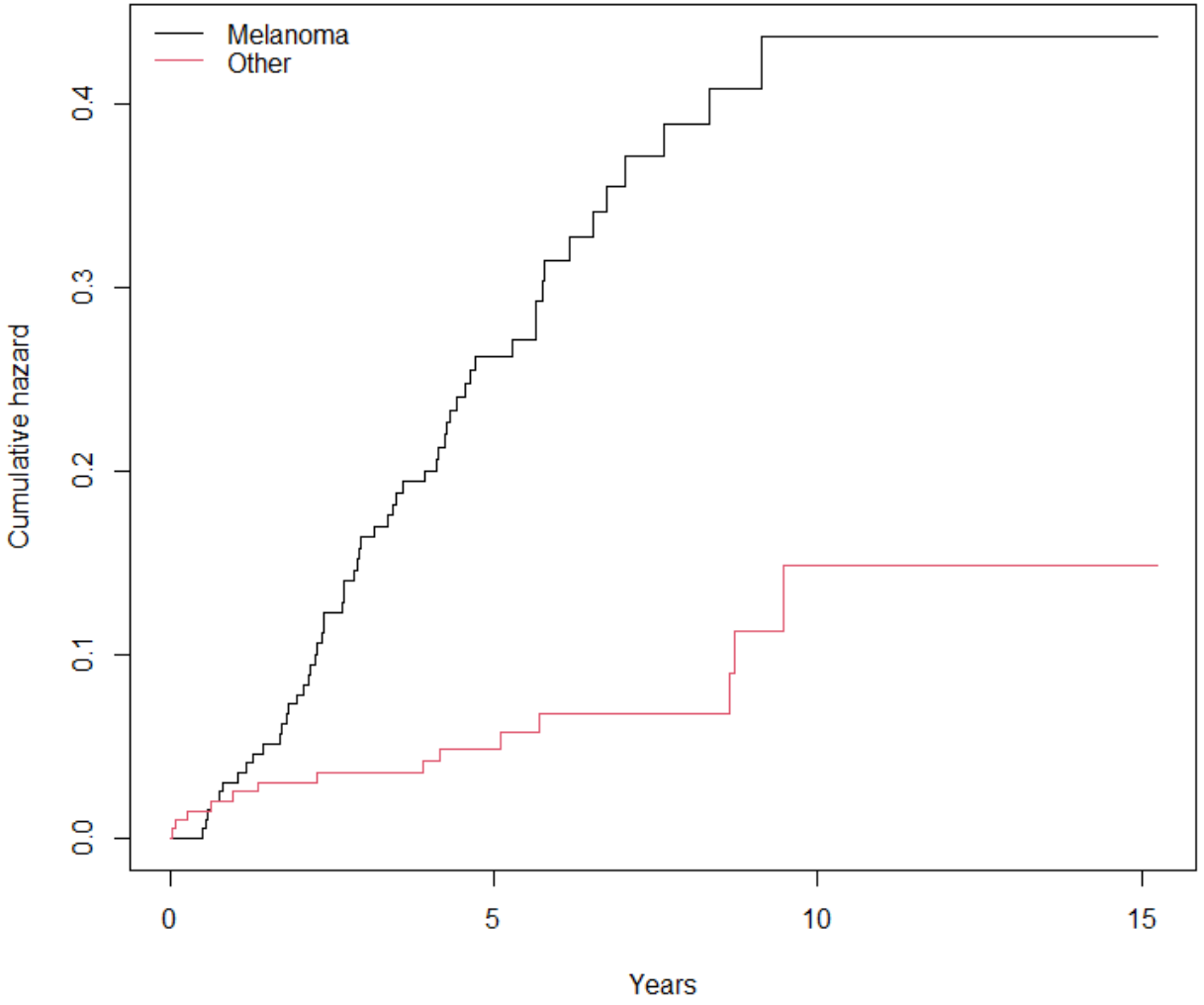
```
> Melanoma[inx, c(1, 3:4, 6:8)]
        time  sex age thickness   ulcer cause
3 0.09582478 male  41      1.34  absent     0
5 0.50650240 male  52     12.08 present     1
1 0.02737851 male  76      6.76 present     2
```

Using R

- For NA: survfit{survival}, mvna{mvna}, …
- For AJ: survfit{survival}, cuminc{cmprsk}, etmCIF{etm}, Cuminc{mstate}, …

Estimated CCSH & CIF





Subdistribution hazard (SDH)

- $\lambda_k^*(t) \equiv -\left(\log\left(1-F_k(t)\right)\right)' = \frac{F_k'(t)}{1-F_k(t)}$: SDH. k −번째 subdistribution function에 대응하는 위험함수로 정의

refer to #10

- Remarks
  - $\lambda_k^*(t) = \lim_{\Delta t \downarrow 0} \frac{P(t \le T_k < t+\Delta t | T \ge t \cup (T \le t, e \ne k))}{\Delta t}$: SDH의 다른 정의

refer to #8

  - 아직 사건이 발생하지 않은 환자 뿐만 아니라 이미 경쟁위험으로 사망한 환자도 위험 집합에 포함됨

Subdistribution hazard (SDH)

- $\Lambda_k^*(t) = \int_0^t \lambda_k^*(s)ds$: cumulative SDH (CSDH)
- $F_k(t) = 1 - \exp\left\{-\int_0^t \lambda_k^*(s)ds\right\}$: CIF의 다른 표현
- Remarks
  - $P(e=k) = F_k(\infty) = 1 - \exp\left(-\Lambda_k^*(\infty)\right)$
  - $P(e=1) + P(e=2) = 1 \Leftrightarrow \exp\left(-\Lambda_1^*(\infty)\right) + \exp\left(-\Lambda_2^*(\infty)\right) = 1$
  - SDHs cannot be independently modeled!

CSH와 SDH의 관계

- $\lambda_1(t) = \left(1+\frac{F_2(t)}{S(t)}\right)\lambda_1^*(t)$
  - $1 - F_1(t) = 1 - P(T \le t, e = 1) = P(T > t) + P(T \le t, e = 2)$
    $= S(t) + F_2(t)$
- SDH는 CSH보다 weighted down 되고, 가중값은 시간에 따라 변하고 competing events에도 의존함

Estimation: SDH & CIF

- n_i^*: estimated # at risk at $t_{(i)} \Rightarrow n_i \leq n_i^*$
- $\hat{\lambda}_k^*(t_{(i)}) = \frac{d_{ik}}{n_i^*}$: estimated SDH
- $\hat{F}_k(t) = 1 - \prod_{i:t_{(i)} \leq t}\left(1 - \hat{\lambda}_k^*(t_{(i)})\right)$: estimated CIF (PL-form)

Contribution of individual l to n_i^*

- still at risk at $t_{(i)}$: 1
- competing event observed at $t_{(l)}$ before $t_{(i)}$:

$$P\{C \geq t_{(i)} | C \geq t_{(l)}\}$$

  - $\Gamma(t) = P(C > t)$: censoring time distribution
  - $\hat{\Gamma}(t)$: Kaplan-Meier estimate of $\Gamma(t)$ by reversing the role of T and C
  - estimated as $\frac{\hat{\Gamma}(t_{(i)}-)}{\hat{\Gamma}(t_{(l)}-)}$

Using R

- For SDH: survfit{survival} along with crprep{mstate} counting process format

Estimated weights

```
> maxT
[1] 9.138946
> EX
    pid       time status2
3     3 0.09582478       0
5     5 0.50650240       1
114 114 5.70841889       2
> WT
     pid   Tstart      Tstop status weight.cens count failcode
117    3 0.000000 0.09582478      0   1.0000000     1        1
176    5 0.000000 0.50650240      1   1.0000000     1        1
528  114 0.000000 5.70841889      2   1.0000000     1        1
529  114 5.708419 5.75770021      2   0.9890110     2        1
530  114 5.757700 5.77138946      2   0.9778985     3        1
531  114 5.771389 6.17659138      2   0.8992170     4        1
532  114 6.176591 6.53798768      2   0.8536870     5        1
533  114 6.537988 6.75427789      2   0.7960055     6        1
534  114 6.754278 7.02258727      2   0.7374757     7        1
535  114 7.022587 7.61670089      2   0.6780018     8        1
536  114 7.616701 8.32854209      2   0.6295731     9        1
537  114 8.328542 9.13894593      2   0.4519325    10        1
> ipcw
[1] 0.4519325
```

pid=114 환자의 사망 이후 피부암으로 사망한 시점:

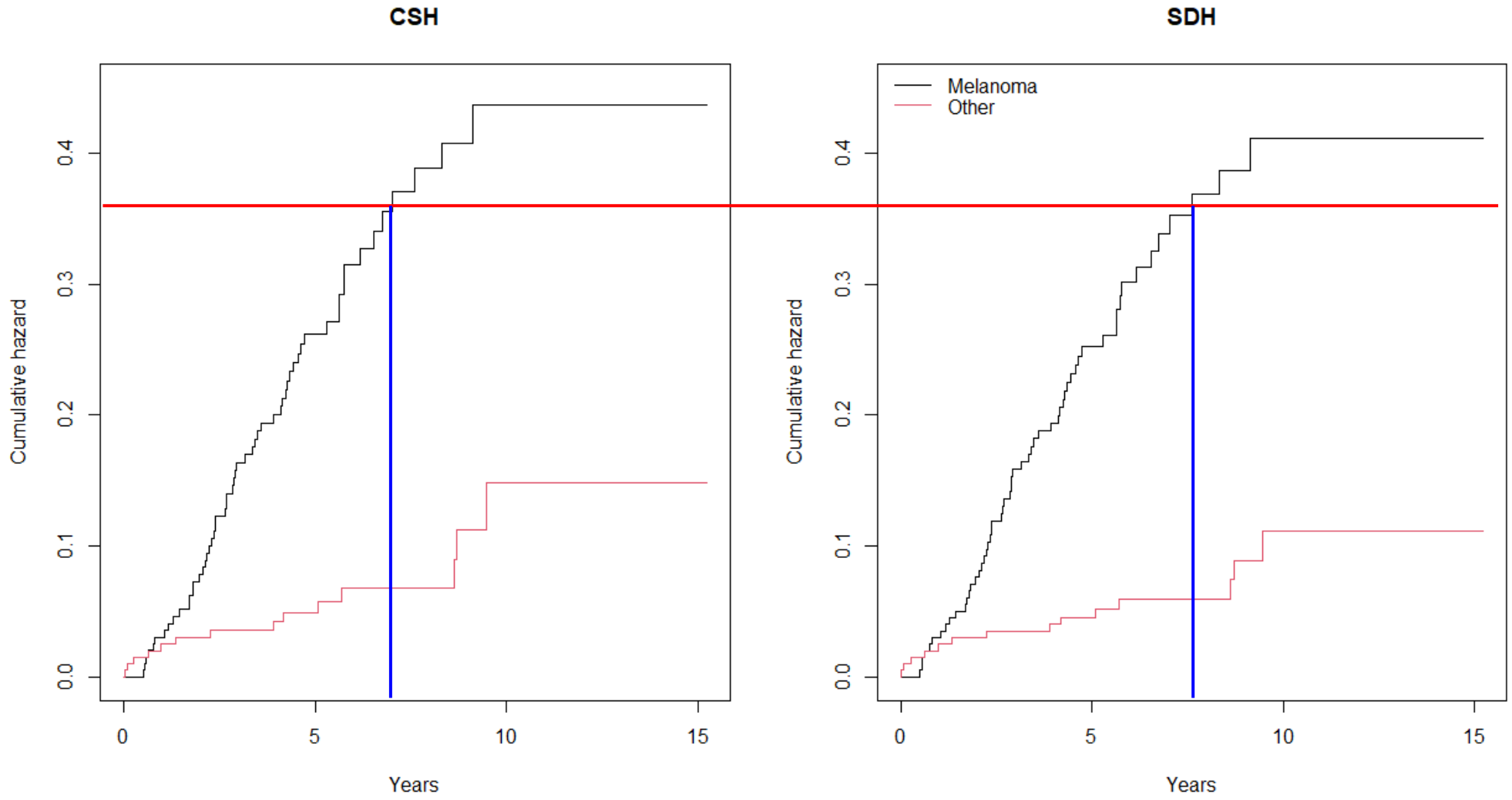
```
[1] 5.757700 5.771389 6.176591 6.537988 6.754278 7.022587 7.616701 8.328542 9.138946
```

위험 집합 비교: n vs. n^*

```
> Comp[seq(1,194,15),]
```

	Time	nrisk.CS	nrisk.mel.SD	nrisk.oth.SD
[1,]	0.02737851	205	205.00000	205.00000
[2,]	1.44832307	189	194.99015	198.00000
[3,]	2.64750171	174	180.99015	197.00000
[4,]	4.12320329	159	166.94021	194.09573
[5,]	4.47364819	144	152.08521	177.05698
[6,]	4.90896646	125	131.69660	160.71688
[7,]	5.31690623	109	115.66296	139.06402
[8,]	5.64544832	94	100.66296	125.05327
[9,]	6.40383299	78	84.48061	97.98169
[10,]	7.03627652	62	67.14693	84.13023
[11,]	8.62970568	47	50.43076	70.86973
[12,]	9.26762491	32	32.00000	48.46820
[13,]	10.86379192	14	14.00000	14.00000

CCSH vs. CSDH


CSH
SDH
Melanoma
Other
Cumulative hazard
0.0
0.1
0.2
0.3
0.4
0
5
10
15
Years

CIF 비교: AJ vs. FG

```
> Equiv[seq(1,194,15),]
```

	Time	CIF.mel.AJ	CIF.mel.FG	CIF.oth.AJ	CIF.oth.FG
[1,]	0.02737851	0.00000000	0.00000000	0.004878049	0.004878049
[2,]	1.44832307	0.04902198	0.04902198	0.029364888	0.029364888
[3,]	2.64750171	0.11765274	0.11765274	0.034267085	0.034267085
[4,]	4.12320329	0.18141214	0.18141214	0.039169283	0.039169283
[5,]	4.47364819	0.20689223	0.20689223	0.044197790	0.044197790
[6,]	4.90896646	0.22353960	0.22353960	0.044197790	0.044197790
[7,]	5.31690623	0.23013963	0.23013963	0.050456445	0.050456445
[8,]	5.64544832	0.24528498	0.24528498	0.050456445	0.050456445
[9,]	6.40383299	0.26936596	0.26936596	0.058111429	0.058111429
[10,]	7.03627652	0.29832966	0.29832966	0.058111429	0.058111429
[11,]	8.62970568	0.32177917	0.32177917	0.058111429	0.058111429
[12,]	9.26762491	0.33871751	0.33871751	0.085379085	0.085379085
[13,]	10.86379192	0.33871751	0.33871751	0.105947064	0.105947064

Proportional CSH models

- $\{(x_1, e_1\delta_1, z_1), \ldots, (x_n, e_n\delta_n, z_n)\}$: observed data
  - z_i: a vector of p covariates (risk factors)
  - eg: sex , treatment, age, tumour size, stage of cancer, blood measures, … (fixed)
- Proportional CSH model:

$$\lambda_k(t|z_i) = \lambda_{0k}(t)\exp(\beta_k' z_i)\,, k = 1,2$$

CSH approach

- Use the partial likelihood principle
  - $L(\beta_1, \beta_2) = \prod_{k=1}^{2} L_k(\beta_k)$,

$$L_k(\beta_k) = \prod_{i=1}^{n} \left\{ \frac{\exp\{\beta_k' z_i\}}{\sum_{j=1}^{n} I(x_j \geq x_i) \exp(\beta_k' z_j)} \right\}^{I(\delta_i e_i = k)}$$

- NPMLE: $\hat{\beta}_k$ as the solution of $\frac{\partial}{\partial \beta_k} \log L_k(\beta_k) = 0$

Using R

- Use coxph{survival}
- or coxph{survival} along with msprep{mstate} stacked-long format

```
Data:
   id from to trans Tstart      Tstop       time status age  sex   ulcer
1   1    1  2     1      0 0.02737851 0.02737851      0  76 male present
2   1    1  3     2      0 0.02737851 0.02737851      1  76 male present
5   3    1  2     1      0 0.09582478 0.09582478      0  41 male  absent
6   3    1  3     2      0 0.09582478 0.09582478      0  41 male  absent
9   5    1  2     1      0 0.50650240 0.50650240      1  52 male present
10  5    1  3     2      0 0.50650240 0.50650240      0  52 male present
   thickness failcode
1       6.76 Melanoma
2       6.76    Other
5       1.34 Melanoma
6       1.34    Other
9      12.08 Melanoma
10     12.08    Other
```

CS model: two analyses

```
Call:
coxph(formula = Surv(time, cause == 1) ~ age + ulcer + thickness +
    sex, data = Melanoma)

                  coef exp(coef) se(coef)     z       p
age           0.012198  1.012273 0.008297 1.470 0.14150
ulcerpresent 1.164479  3.204253 0.309751 3.759 0.00017
thickness     0.108945  1.115101 0.037734 2.887 0.00389
sexmale       0.432817  1.541594 0.267410 1.619 0.10554
```

```
Call:
coxph(formula = Surv(time, cause == 2) ~ age + ulcer + thickness +
    sex, data = Melanoma)

                  coef exp(coef) se(coef)     z        p
age           0.07255   1.07525  0.02167 3.348 0.000813
ulcerpresent 0.10937   1.11557  0.59128 0.185 0.853256
thickness     0.04958   1.05083  0.08794 0.564 0.572891
sexmale       0.35801   1.43048  0.54859 0.653 0.514012
```

CS model: one analysis

```
Call:
coxph(formula = Surv(time, status) ~ age * strata(failcode) +
    ulcer * strata(failcode) + thickness * strata(failcode) +
    sex * strata(failcode), data = M.s)

                                        coef exp(coef)  se(coef)      z        p
age                                 0.012198  1.012273  0.008297  1.470 0.14150
ulcerpresent                        1.164479  3.204253  0.309751  3.759 0.00017
thickness                           0.108945  1.115101  0.037734  2.887 0.00389
sexmale                             0.432817  1.541594  0.267410  1.619 0.10554
age:strata(failcode)Other           0.060354  1.062212  0.023203  2.601 0.00929
strata(failcode)Other:ulcerpresent -1.055112  0.348153  0.667501 -1.581 0.11395
strata(failcode)Other:thickness    -0.059365  0.942363  0.095693 -0.620 0.53501
strata(failcode)Other:sexmale      -0.074806  0.927923  0.610293 -0.123 0.90244

Likelihood ratio test=58.28  on 8 df, p=1.015e-09
n= 410, number of events= 71
```

SDH approach

- Modeling the CIF for failure from cause 1 conditional on the covariates, $F(t;z) = P(T \leq t, e = 1|z)$ [Fine & Gray, 1999]
- Proportional SDH model: $\lambda_1^*(t|z_i) = \lambda_{01}^*(t)\exp(\gamma_1' z_i)$
- $L_1(\gamma_1) =$

$$\prod_{i=1}^{n}\left\{\frac{\exp\{\gamma_1' z_i\}}{\sum_{j=1}^{n}\left\{I(x_j \geq x_i) + I(t_j < x_i < c_j, \delta_j e_j = 2)\frac{\hat{\Gamma}(x_i-)}{\hat{\Gamma}(t_j-)}\right\}\exp(\gamma_1' z_j)}\right\}^{I(\delta_i e_i = 1)}$$

1 under administrative censoring

Using R

- Use crr{cmprsk}
- or use coxph{survival} along with crprep{mstate}

```
> crp.r[crp.r$pid %in% c(3, 162, 171), ]
     pid   Tstart      Tstop status weight.cens age    sex   ulcer thickness count failcode
45     3 0.000000 0.09582478      0   1.0000000  41   male  absent       1.34     1        1
369  162 0.000000 8.71184120      2   1.0000000  49   male  absent       6.12     1        1
370  162 8.711841 9.13894593      2   0.8139535  49   male  absent       6.12     2        1
379  171 0.000000 9.13894593      1   1.0000000  60 female present       3.22     1        1
416    3 0.000000 0.09582478      0   1.0000000  41   male  absent       1.34     1        2
909  162 0.000000 8.71184120      2   1.0000000  49   male  absent       6.12     1        2
918  171 0.000000 9.13894593      1   1.0000000  60 female present       3.22     1        2
919  171 9.138946 9.46748802      1   0.8235294  60 female present       3.22     2        2
```

FG model: two analyses

```
Call:
crr(ftime = Melanoma$time, fstatus = Melanoma$cause, cov1 = covs,
    failcode = 1, cencode = 0)

             coef exp(coef) se(coef)     z p-value
age       0.00593      1.01  0.00929 0.638  0.5200
ulcer     1.12863      3.09  0.30344 3.719  0.0002
thickness 0.08999      1.09  0.03836 2.346  0.0190
sex       0.40503      1.50  0.27558 1.470  0.1400
```

```
Call:
crr(ftime = Melanoma$time, fstatus = Melanoma$cause, cov1 = covs,
    failcode = 2, cencode = 0)

             coef exp(coef) se(coef)      z p-value
age        0.0570     1.059   0.0142  4.011 0.00006
ulcer     -0.1092     0.897   0.5866 -0.186 0.85000
thickness  0.0114     1.012   0.0851  0.135 0.89000
sex        0.2630     1.301   0.5923  0.444 0.66000
```

FG model: one analysis

```
Call:
coxph(formula = Surv(Tstart, Tstop, status == failcode) ~ age *
    strata(failcode) + ulcer * strata(failcode) + thickness *
    strata(failcode) + sex * strata(failcode), data = crp.r,
    weights = weight.cens)

  n= 953, number of events= 71

                                              coef exp(coef)  se(coef) robust se      z Pr(>|z|)
age                                       0.005928  1.005945  0.008026  0.009028  0.657 0.511465
ulcerpresent                              1.128630  3.091418  0.313844  0.300809  3.752 0.000175 ***
thickness                                 0.089995  1.094168  0.038720  0.034614  2.600 0.009324 **
sexmale                                   0.405032  1.499350  0.268754  0.267154  1.516 0.129494
age:strata(failcode)failcode=2            0.051030  1.052354  0.020666  0.015508  3.291 0.001000 ***
strata(failcode)failcode=2:ulcerpresent -1.237810  0.290019  0.668887  0.643621 -1.923 0.054455 .
strata(failcode)failcode=2:thickness    -0.078550  0.924456  0.093643  0.087838 -0.894 0.371184
strata(failcode)failcode=2:sexmale      -0.142072  0.867559  0.610014  0.635984 -0.223 0.823232
```

Left-truncated (late entry) data

- observed: $(x_i, e_i\delta_i, z_i) \rightarrow (x_i, e_i\delta_i, z_i, l_i)$, l_i: late entry
- For CSH: $I(x_i \geq t) \rightarrow I(x_i \geq t > l_i)$
- For SDH:

$$I(x_i \geq t) + I(t_i < t < c_i, \delta_i e_i = 2) \times \frac{\Gamma(t-)}{\Gamma(t_i -)} \rightarrow$$

$$I(x_i \geq t > l_i) + I(l_i \vee t_i < t < c_i, \delta_i e_i = 2) \times \frac{\Gamma(t-)}{\Gamma(t_i-)} \times \frac{\Phi(t-)}{\Phi(t_i-)}$$

  - $\Phi(t) = P(L_i \leq t)$: estimated by reversing time

Time-varying covariates

- For CSH: coxph{survival} along with tmerge{survival} counting process format
- For SDH: cox.kmi{kmi} with imputed data set [Ruan & Gray, 2008]

HAP in ICU data

	id	start	stop	status	pneu	adm.cens.exit	age	sex	outcome
131	23371	0	4	0	0	4	63.05753	F	0
132	23371	4	29	1	1	29	63.05753	F	2
159	30236	0	6	0	0	6	38.96164	M	0
160	30236	6	26	0	1	26	38.96164	M	0
265	1002890	0	4	1	0	358	76.77918	F	3
983	1021700	0	51	0	0	51	62.48219	F	0
993	1022004	0	7	0	0	7	2.09589	F	0
994	1022004	7	15	1	1	48	2.09589	F	3
1068	2001524	0	7	1	0	7	70.94475	F	2

pneumonia	event type			total
	censored	death in the unit	discharge alive	
yes	5	21	82	108
no	16	126	1063	1205
total	21	147	1145	1313

CSH model

```
coxph(formula = Surv(start, stop, outcome == 2) ~ pneu + sex +
    age, data = icu.pneu)

           coef exp(coef)  se(coef)      z        p
pneu1 -0.087804  0.915941  0.248345 -0.354 0.723673
sexM  -0.020834  0.979382  0.169716 -0.123 0.902299
age    0.019762  1.019958  0.005312  3.720 0.000199
```

FG model via imputation

```
cox.kmi(formula = Surv(start, stop, outcome == 2) ~ pneu + sex +
    age, imp.data = imp.death)

******************
Pooled estimates:
******************
             coef exp(coef)   se(coef)         t            p
pneu1 1.06374450  2.897199 0.24071329 4.4191348 9.909827e-06
sexM  0.11047993  1.116814 0.16978135 0.6507188 5.152280e-01
age   0.01827515  1.018443 0.00508217 3.5959344 3.232297e-04
```

FG model with complete censoring

```
coxph(formula = Surv(start, adm.cens.exit, outcome == 2) ~ pneu +
    sex + age, data = icu.pneu)

          coef exp(coef) se(coef)     z        p
pneu1 1.064077  2.898162 0.240640 4.422 9.79e-06
sexM  0.112463  1.119031 0.169652 0.663 0.507390
age   0.018454  1.018625 0.005084 3.630 0.000284
```

References

- Beyersmann J, Allignol A, Schumacher M (2012). Competing risks and multistate models with R. Springer, New York.
- Geskus, RB (2015). Data analysis with competing risks and intermediate states. CRC press.

THANK YOU!!!