Time to event data analysis

Ammarin Thakkinstian Ph.D.

Section for Epidemiology and Biostatistics
Ramathibodi Hospital, Mahidol University
Email: raatk@mahidol.ac.th

Outline of talk

Time to event analysis

- When should it be applied
  - Type of outcome
  - Study design
- Type of data & outcome
  - Single record single event with censoring
  - Multiple records single event with censoring (Time-varying covariates)
  - Multiple records with multiple events
- Model building
  - Model selection
  - Assumption checking & Goodness of fit
- Interpretation
- Report
- Sample size estimation

Section for Clinical Epidemiology & Biostatistics
Faculty of Medicine Ramathibodi Hospital

Type of outcome

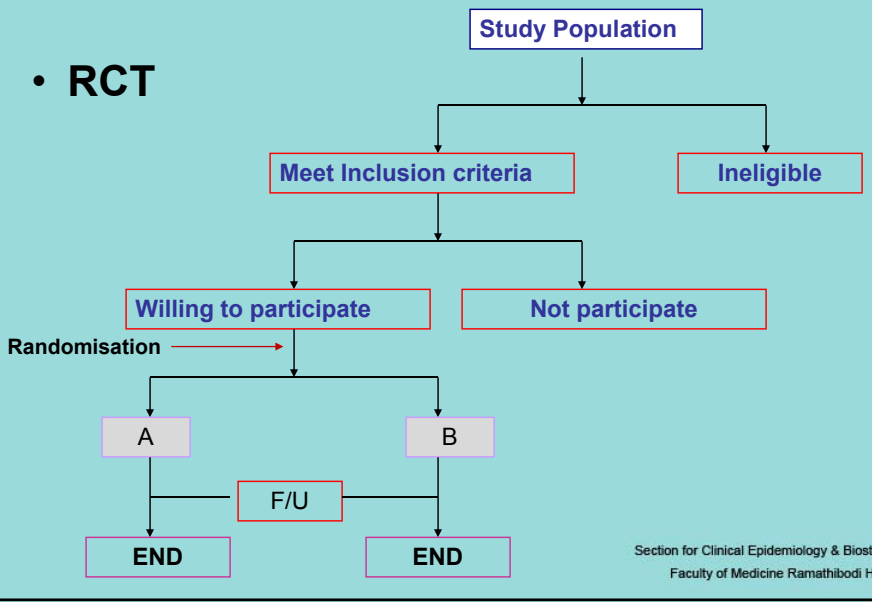
- Time to interested event
  - Event + time
- Time at diagnosis of colon cancer to death
- Infection to cure
- Cancer remission to recurrence
- ART to success (viral copies < 50)
- Thymectomy to remission (to recurrence)
- Renal transplant to graft rejection
- Time since diagnosis of hypertension to diagnosis stroke
- Time from the first episode of stroke to the second attack

Why can't be

- Linear Regression: time
  - All subjects must have events
- Logistic regression:
  - Should not have loss to FU, withdraw, terminate study before end

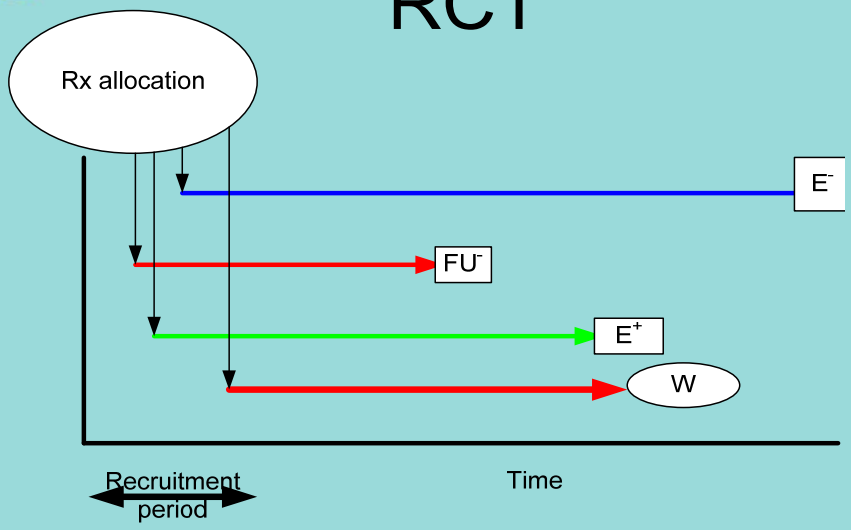
Study Design

- **RCT**

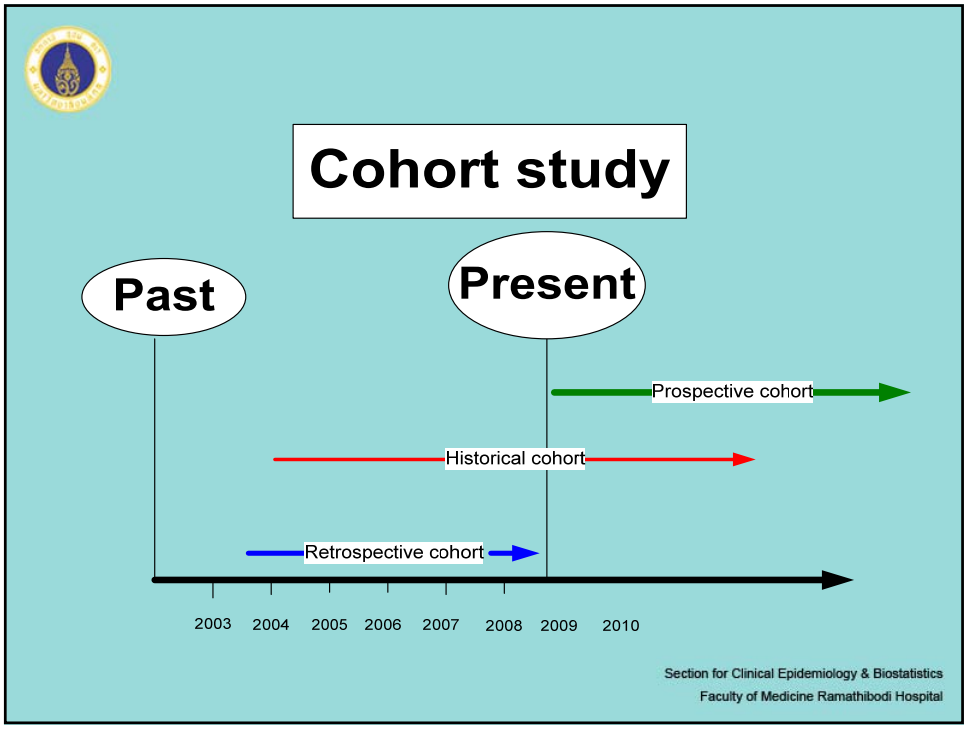


Section for Clinical Epidemiology & Biostatistics
Faculty of Medicine Ramathibodi Hospital

RCT



Section for Clinical Epidemiology & Biostatistics
Faculty of Medicine Ramathibodi Hospital



Methodological characteristics of survival study

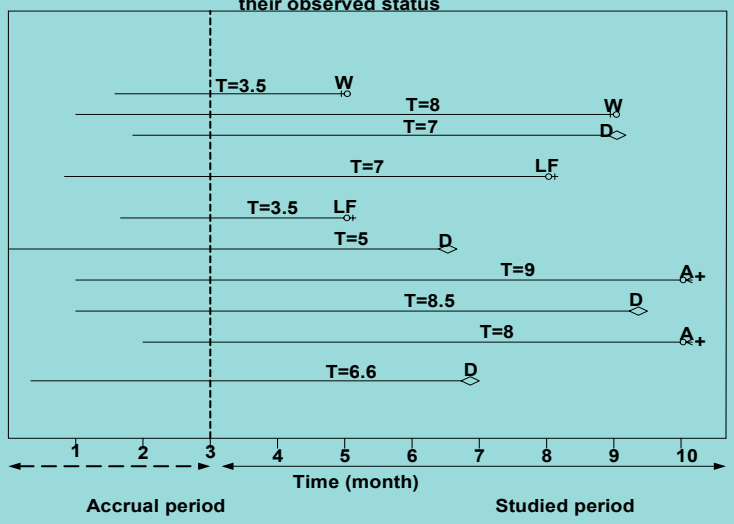
- Subjects enroll to the study at different time and stay with the study with different durations
- Different events occur on each subject
- Start date
  - Date at diagnose,
  - Date at receiving treatment
  - Date of operation, etc.

Section for Clinical Epidemiology & Biostatistics
Faculty of Medicine Ramathibodi Hospital

* Patient's status
  - Death if it is the final outcome
  - Recurrence, disease free
  - Infection, non-infection
  - Remission, non-remission
  - Loss to follow up
  - Withdraw
* Patients will be censored on that date if they are
  - loss to follow up
  - Withdraw/drop out
  - Alive/disease-free at the end

- End date
  - If death is the final outcome,
    - death date if patients die or
    - date at the end of study if patients still survive
  - If recurrence is interested outcome,
    - date at finding disease recurrence or
    - date at the endof study if still disease-free
  - date at the last meeting if loss to follow-up or withdraw or terminate study

Figure 1. Patients enrolled to follow-up study & their observed status



D=death, A=alive, LF = lost follow-up, W=withdraw

Section for Clinical Epidemiology & Biostatistics
Faculty of Medicine Ramathibodi Hospital

Estimation of survival probability

- Kaplan-Meier
  - Distinct time
- Life table
  - Interval time

Section for Clinical Epidemiology & Biostatistics
Faculty of Medicine Ramathibodi Hospital

Survival and Hazard Function

From Collete 1994

Let T is a random variable associated with survival time. Suppose that a random variable T has a proability distributi on with underlying probabilit y density funciton f(t).

The distributi on funciton of failure T is given by

$F(t) = P(T < t) = \int_0^t f(u)d(u)$,

represents probabilit y that survival time is less than time t

Let S(t) is survival function, which is defined as probabilit y that survival time is greater than or equal t

$S(t) = P(T \geq t) = 1 - F(t)$

Equivalent ly for given time t, S(t) is a proportion of individual who have not yet failed

$$\hat{S}(t) = \frac{\text{Number of individual s who survive at t or longer}}{\text{Total number of individual s}}$$

Let t_i = observed time = $t_1, t_2, t_3, \ldots t_j$

Sort n survival times $t_1, t_2, \ldots, t_n$ in ascending order such that

$$t_1 \leq t_2 \leq \ldots \leq t_n$$

Let d_j = no. subjects die at $t_j = 1,2,\ldots. d_j$

then the survivorsh ip function at t_j can be estimated as

$$\hat{S}(t_j) = \frac{n_j - d_j}{n_j} = 1 - \frac{d_j}{n_j}$$

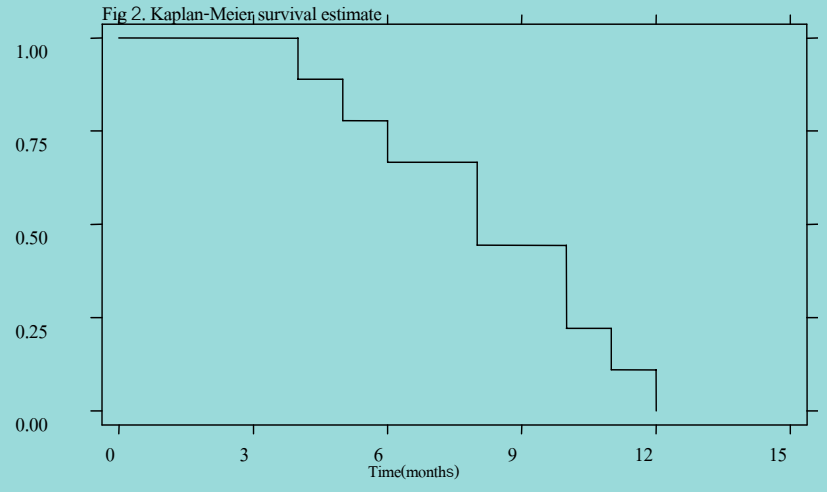
Section for Clinical Epidemiology & Biostatistics
Faculty of Medicine Ramathibodi Hospital

Example 1.

T_j	dj	$S(t)$
4	1	9/10=.9
5	2	8/10=.8
6	3	7/10=.7
8	4	4/10=.4
8	5	4/10=.4
8	6	4/10=.4
10	7	2/10=.3
10	8	2/10=.3
11	9	1/10=.1
12	10	0/10=0

Section for Clinical Epidemiology & Biostatistics
Faculty of Medicine Ramathibodi Hospital

Fig 2. Kaplan-Meier survival estimate



Section for Clinical Epidemiology & Biostatistics
Faculty of Medicine Ramathibodi Hospital

Kaplan-Meier with censoring data

$$S(t) = \prod_{j=1}^{k} s(t_j)$$

$$= s(t_j) x s(t_{j-1}) x s(t_{j-2}) x \ldots s(t_k)$$

$$\hat{s}(t) = \frac{n_j - d_j}{n_j} = 1 - \frac{d_j}{n_j}$$

$$\hat{S}(t) = \prod_{1}^{j} \left(1 - \frac{d_j}{n_j}\right)$$

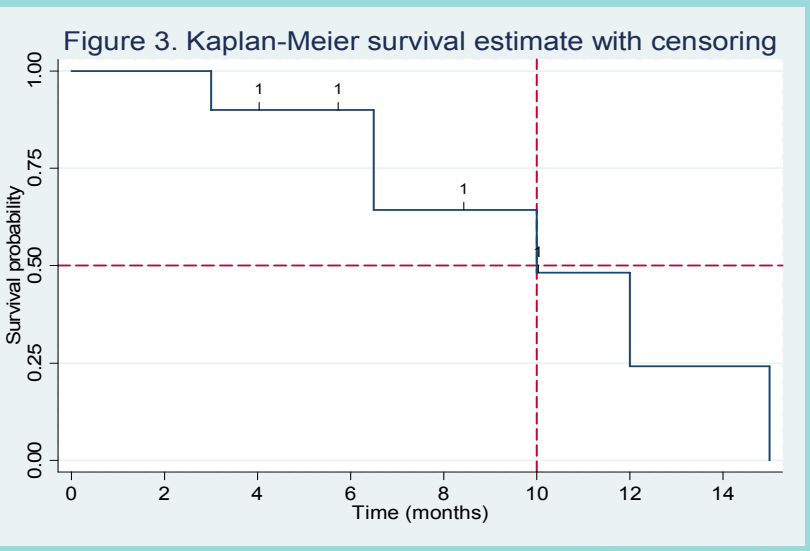
for $t_k \le t \le t_{k+1}$

Section for Clinical Epidemiology & Biostatistics
Faculty of Medicine Ramathibodi Hospital

Example 2.

Time	n at risk	Censor	No death	P(t)	s(t)	$\hat{S}(t)$
3	10	0	1	1/10	1-1/10	.9
4	9	1	0	0	1	.9x1
5.7	8	1	0	0	1	.9x1x1
6.5	7	0	2	2/7	1-2/7=.714	.9x1x1x.714=.643
8.4	5	1	0	0	1	.9x1x1x.714x1=.643
10	4	1	1	¼	1-1/4=.750	.9x1x1x.714x1x.75=.482
12	2	0	1	½	1-1/2=.500	.9x1x1x.714x1x.75x.500=.241
15	1	0	1	1/1	1-1/1=0	.9x1x1x.714x1x.75x.500x0=0

Section for Clinical Epidemiology & Biostatistics
Faculty of Medicine Ramathibodi Hospital



Section for Clinical Epidemiology & Biostatistics
Faculty of Medicine Ramathibodi Hospital

Variance of S(t)

Greenwood's variance

$$\text{var}\left[\hat{S}(t)\right] = \left[\hat{S}(t)\right]^2 \sum^{(t)} \frac{d_j}{n_j(n_j - d_j)} \quad \text{or}$$

$$\text{var}\left[\hat{S}(t)\right] = \frac{\left[\hat{S}(t)\right]^2\left[1-\hat{S}(t)\right]}{n(t)}$$

$$95\%CI = \hat{S}(t_j) \pm 1.96x\sqrt{\text{var}\left[\hat{S}(t_j)\right]}$$

Section for Clinical Epidemiology & Biostatistics
Faculty of Medicine Ramathibodi Hospital

Log-log variance

$$Var\left[\log\{-\log S(t)\}\right] \approx \frac{1}{\left[\log S(t)\right]^2} \sum_{j=1}^{k} \frac{d_j}{n_j(n_j - d_j)}$$

$$95\%CI\ S(t) = S(t)^{\exp \pm Z_{\alpha/2} SE\left[\log\{-\log S(t)\}\right]}$$

Estimation methods

Kaplan-Meier method

Time	Beg. Total	Fail	Net Lost	Survivor Function	Std. Error	[95% Conf.	Int.]
3	10	1	0	0.9000	0.0949	0.4730	0.9853
4	9	0	1	0.9000	0.0949	0.4730	0.9853
5.7	8	0	1	0.9000	0.0949	0.4730	0.9853
6.5	7	2	0	0.6429	0.1679	0.2447	0.8705
8.4	5	0	1	0.6429	0.1679	0.2447	0.8705
10	4	1	1	0.4821	0.1877	0.1254	0.7739
12	2	1	0	0.2411	0.1946	0.0132	0.6263
15	1	1	0	0.0000	.	.	.

Hazard function

- Hazard function h(t) is the probability that an individual die (fail) at time t, conditional on s/he's having survived before time t.
- In another word, the hazard function is the death rate for an individual surviving at time t.

$$h(t) = \frac{f(t)}{s(t)}$$

$$= -\frac{d}{dt}\{\log s(t)\}$$

$$s(t) = \exp\{-h(t)\} \text{ where}$$

$$H(t) = \int_0^t f(u)d(u)$$

$$H(t) = -\log S(t)$$

H(t) is called cummulative hazard function.

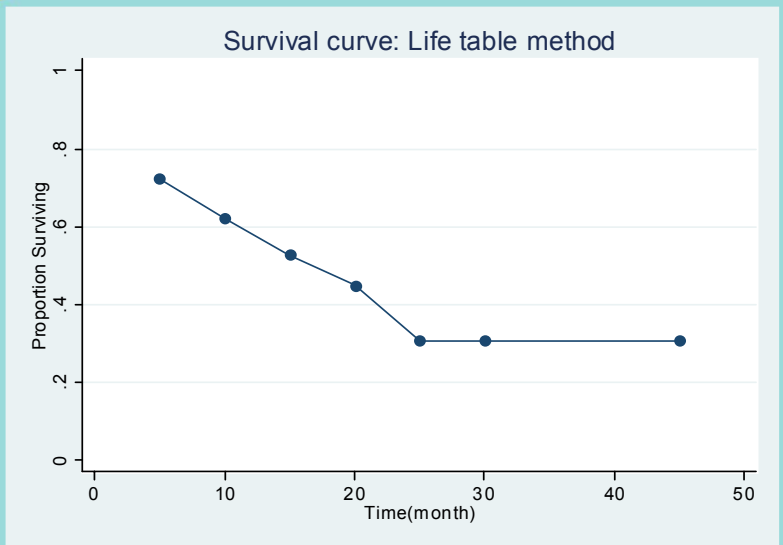
Section for Clinical Epidemiology & Biostatistics
Faculty of Medicine Ramathibodi Hospital

Estimation methods

Lift table method

```
ltable time death, interval(5)
```

Interval		Beg. Total	Deaths	Lost	Survival	Std. Error	[95% Conf.	Int.]
0	5	156	36	52	0.7231	0.0392	0.6375	0.7917
5	10	68	8	23	0.6207	0.0475	0.5203	0.7060
10	15	37	5	7	0.5281	0.0556	0.4138	0.6300
15	20	25	3	10	0.4488	0.0634	0.3225	0.5672
20	25	12	3	5	0.3071	0.0804	0.1615	0.4656
25	30	4	0	3	0.3071	0.0804	0.1615	0.4656
40	45	1	0	1	0.3071	0.0804	0.1615	0.4656



Section for Clinical Epidemiology & Biostatistics
Faculty of Medicine Ramathibodi Hospital

Type of design/databases

- Single record with single event
  - Subjects' characteristics (e.g., demographic, clinical, laboratory, and behavioral data) and interested event(s) are measured at baseline or at enrollment period
  - Only the outcome variable is considered as change over time, whereas risk/prognostic factors are constant since baseline

Type of design/databases

- Multiple records with single event
  - Studies variable
    - Some variables at baseline/enrollment are changed overtime
  - Outcome event
    - Consider only one episode
  - Time varying co-variates

Type of design/databases

- Multiple records with multiple events
  - Not only co-variables but also outcomes are changed overtime
  - Time-varying co-variates & outcomes

Single record single event

```
     +----------------------+
     | id   fail       time |
     |----------------------|
  1. |  1      1   11.96721 |
  2. |  2      2   4.131147 |
  3. |  3      2   6.491803 |
  4. |  4      1   19.14754 |
  5. |  5      4   13.01639 |
     |----------------------|
  6. |  6      3   7.803279 |
     +----------------------+
lab list
death:
           1 death
           2 alive
           3 loss
           4 withdraw
```

Section for Clinical Epidemiology & Biostatistics
Faculty of Medicine Ramathibodi Hospital

Single record single event with date

	id	fail	date_en	date_rec
1.	1	1	01/01/99	01/01/00
2.	2	2	02/03/99	06/07/99
3.	3	2	01/12/99	16/06/00
4.	4	1	20/04/00	25/11/01
5.	5	4	10/10/99	10/11/00
6.	6	3	11/11/00	07/07/01

Example 2. Prognostic factors of AIDS patients

- Estimate survival probability at 6, 12, 24 moths
- Determine overall median survival time and median survival time (if definable) according to groups of prognostic factors
- Assess hazard risk of prognostic factors

Comparison of survival time

- **Survival times are different by**
  - **First AIDS defining illness**
    - **Crypto, Toxo, PCP, Tb, other**
  - **Age groups**
  - **Sex**
  - **Risk behaviour**
    - **Homosexual vs non-homosexual**

Statistic tests

- Wilcoxon test
- Log-rank test
- Null hypothesis

$$H_o : S_1(t) = S_2(t) = ...S_k(t) \quad \textbf{or}$$
$$H_o : H_1(t) = H_2(t) = ..H_k(t) \quad \textbf{or}$$
$$H_o : M(t_1) = M(t_2) = ...M(t_k)$$

- RQ: whether patients who had the first defining illness of AIDS as Cryptococal meningitis (Crypto) had longer survival time than patients with PCP

$$H_o : M(t_1) = M(t_2)$$

Log-rank test

$$e_{1j} = \frac{n_{1j} d_j}{n_j} \qquad e_{2j} = \frac{n_{2j} d_j}{n_j}$$

Time(month)	$d1_j$	$n1_j$	$d2_j$	$n2_j$	d_j	n_j	$e1_j$	$e2_j$
.03	1	15	0	15	1	30	1x15/30=.5	1x15/30=.50
.07	1	14	0	15	1	29	1x14/29=.48	1x15/29=.52
.1	1	13	0	15	1	28	1x13/28=.46	1x15/28=.54
.17	1	12	0	15	1	27	.44	.56
.23	1	11	0	15	1	26	.42	.57
.27	1	10	0	15	1	25	.40	.60
.50	1	9	0	15	1	24	.38	.63
3.03	1	8	0	15	1	23	.35	.65
5.93	0	7	1	15	1	22	.32	.68
9.13	1	7	0	14	1	21	.33	.67
9.80	1	6	0	14	1	20	.30	.70
13.23	0	5	1	14	1	19	.26	.74
15.83	0	5	1	13	1	18	.28	.72
Total	10		3				4.93	8.07

Section for Clinical Epidemiology & Biostatistics
Faculty of Medicine Ramathibodi Hospital

$$O_1 = \sum_{j=1}^{r} d_{1j} \quad ; \quad E_1 = \sum_{j=1}^{r} e_{1j}$$

$$O_2 = \sum_{j=1}^{r} d_{2j} \quad ; \quad E_2 = \sum_{j=1}^{r} e_{2j}$$

$$x^2 = \sum_{j=1}^{k} \frac{(O_j - E_j)^2}{E_j}$$

$$= \frac{(O_1 - E_1)^2}{E_1} + \frac{(O_2 - E_2)^2}{E_2}$$

$$= \frac{(10 - 4.93)^2}{4.93} + \frac{(3 - 8.07)^2}{8.07}$$

$$= 8.39$$

$$\mathbf{df = k - 1} = 2 - 1$$

Section for Clinical Epidemiology & Biostatistics
Faculty of Medicine Ramathibodi Hospital

Cox proportional hazard model

- Linear regression (time)
- Logistic regression (event)
- Cox (time & event)

$$H(t) = H_0(t)e^{(b_0 + b_1x_1 + b_2x_2 + ... + b_px_p)}$$

$$\log\left[\frac{H(t)}{H_0(t)}\right] = b_0 + b_1x_1 + b_2x_2 + ... + b_px_p$$

Section for Clinical Epidemiology & Biostatistics
Faculty of Medicine Ramathibodi Hospital

Simple Cox regression

$$H(t) = H_0(t)e^{b_1 x_1}$$

X = gender = 1 for male, 0 for female

For male

$$H(t_m) = H_0(t)e^{b_1}$$

For female

$$H(t_f) = H_0(t)e^{0}$$

$$\frac{H(t_m)}{H(t_f)} = \frac{H_0(t)e^{b_1}}{H_0(t)e^{0}} = e^{b_1}$$

$$HR = e^{b_1}$$

Section for Clinical Epidemiology & Biostatistics
Faculty of Medicine Ramathibodi Hospital

Dummy variables

- Dealing with categorical variables
- Codes refer to groups that subjects are belong to 1 ≠ 1, 2 ≠ 2, …
- Needs to assign one group as a reference
- Other groups are compared to the reference group
- Number of dummy variables = k-1

Section for Clinical Epidemiology & Biostatistics
Faculty of Medicine Ramathibodi Hospital

Define variable	Dummy variables				
	_Idefine_1	_Idefine_2	_Idefine_3	_Idefine_4	_Idefine_5
Crypto	1	0	0	0	0
Tb	0	1	0	0	0
PCP	0	0	1	0	0
Toxo	0	0	0	1	0
Other	0	0	0	0	1

Model selection

- How to include variables into the model
  - Stepwise
    - Backward elimination
    - Forward elimination
  - Enter
- Keep or omit variables
  - Likelihood ratio test
  - Wald test

Assumption checking

- the hazard of death at any time for individuals in one group should be proportional to the hazard of death at that time in other groups
- Known as proportional hazard or constant hazard

As for the hazard model

$$\ln h(t) = \ln h_0(t) + \beta_1 x_1$$

$$x = 1;$$

$$\ln h(t) = \ln h_0(t) + \beta_1$$

$$x = 0$$

$$\ln h(t) = \ln h_0(t)$$

$$\ln \frac{h(t / x = 1)}{h(t / x = 0)} = \ln h_0(t) + \beta_1 - \ln h_0(t)$$

$$\ln HR = \beta_1$$

Section for Clinical Epidemiology & Biostatistics
Faculty of Medicine Ramathibodi Hospital

That is β_j is not varied overtime. If it is varied, it should be a function of time as a model

$$\beta_j(t) = \beta_j + \gamma_j g_j(t)$$

Function of time {g(t)} can be any function such as actual t, ln(t), S(t), H(t), rank(t).

$$\gamma_j g_j(t) \cong E\left\{r_j^*(t)\right\}$$

$E\left\{r_j^*(t)\right\}$ = **expected value of scale Schoenfeld residuals**

Estimation of residuals

- Schoenfeld residual for the i^{th} subject on the k^{th} covariate at time j

$$\hat{r}_{ik} = c_i(x_{ik} - \bar{\hat{x}}_{w_i k})$$

which is estimator of risk set conditional on mean of covariates; r_{ik}=0 if c=0 if subject is censored

Scaled Schoenfeld residual

$$\hat{r}^* = [Var(\hat{r}_i)]^{-1}\hat{r}_i$$

$$[Var(\hat{r}_i)]^{-1}\hat{r}_i = m\hat{V}ar(\hat{\beta})$$

$$\hat{r}^* = m\hat{V}ar(\hat{\beta})\hat{r}_i$$

where m = number of failure or interested events

Section for Clinical Epidemiology & Biostatistics
Faculty of Medicine Ramathibodi Hospital

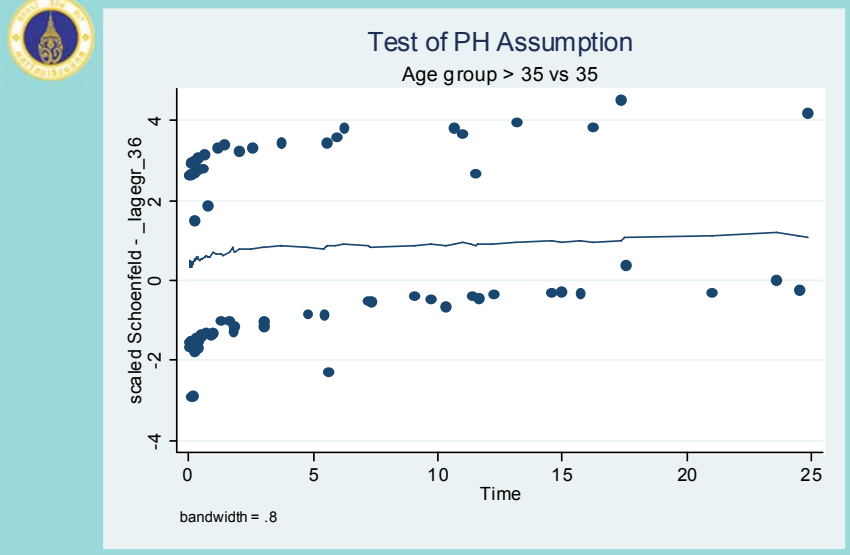
As for $\gamma_j g_j(t) \cong E\{r_j^*(t)\}$,

plotting relationship between

$r_j^*(t)$ vs ln(t)

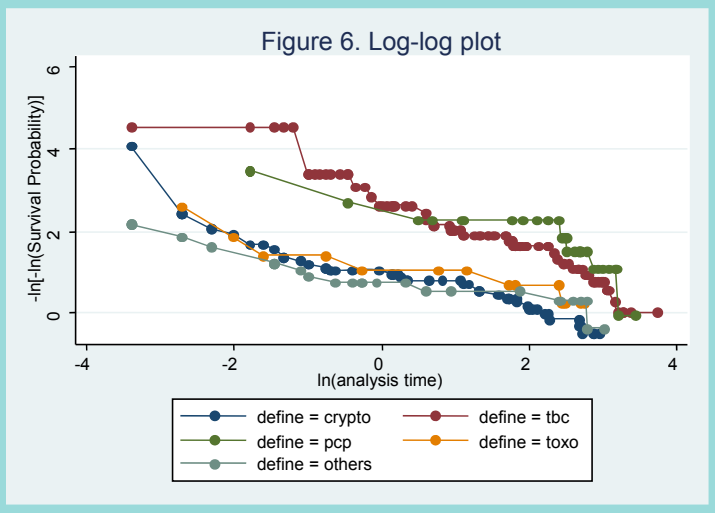
should have slope close to 0

Section for Clinical Epidemiology & Biostatistics
Faculty of Medicine Ramathibodi Hospital

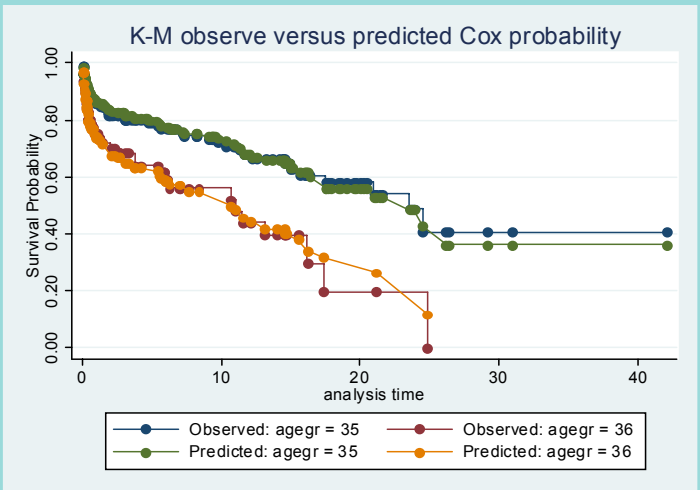


There is a little trend of correlation, thus HR should not vary over time

Log-log plot



K-M observe vs predicted Cox



Assessing goodness of fit

- This can be assessed using Cox-Snell residuals.
- This residual is calculated base on counting process of time to event regression model where the basic model is

$$N(t) = \Lambda(t, X, \beta) + M(t)$$

$$N(t) = actual\ \text{observedevents}$$
$$= c_i$$
$$\Lambda(t, X, \beta) = \text{a systematiccomponentwhichis variedby X \& t}$$
$$= \text{H(t,X,}\beta\text{)}$$

$$\hat{M}(t) = c_i - \hat{\text{H}}\text{(t,X,}\beta\text{).....Cox-snell}$$

Section for Clinical Epidemiology & Biostatistics
Faculty of Medicine Ramathibodi Hospital

Cox regression residuals

Matingale **using Breslow estiamtor & partial derivative of beta**

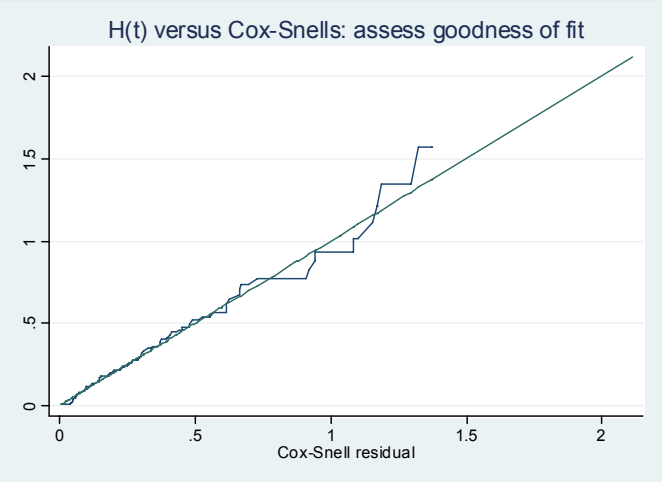
$$M_i = \sum_{i=1}^{n} x_{ik}\left[c_i - H(t_i, X, \beta)\right]$$

sum of the value of covariate times and "observed minus expected" residuals.

This concept is the same as other linear and logistic models

Section for Clinical Epidemiology & Biostatistics
Faculty of Medicine Ramathibodi Hospital

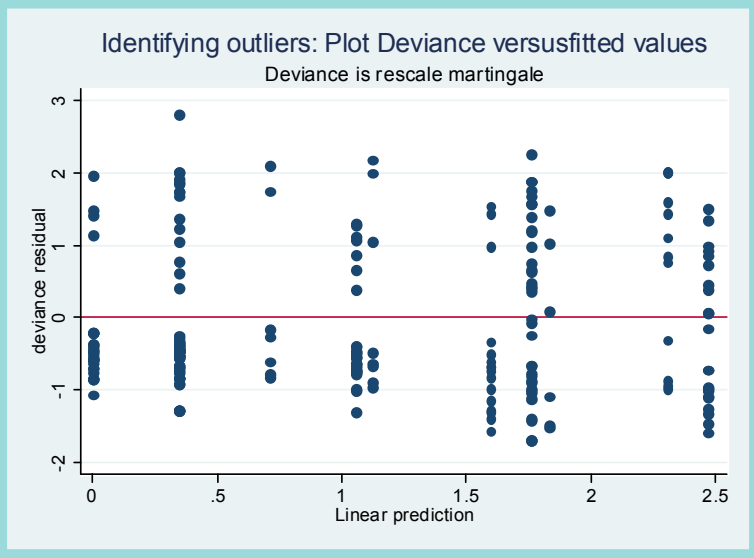
- If the Cox model fits well with data, CS residual should have a standard censored exponential with HR=1
- Plot HR versus CS should have slope =1
- Calculate HR by treating CS as time variable & original failure
- As for the final model of example 2,



Section for Clinical Epidemiology & Biostatistics
Faculty of Medicine Ramathibodi Hospital

Identifying outliers

- Plot deviance versus predicted values
- Deviance is estimated by re-scaling martingale residual and it is symmetry to 0
- The same concept as plotting standardised residual versus fitted values in linear regression



Section for Clinical Epidemiology & Biostatistics
Faculty of Medicine Ramathibodi Hospital

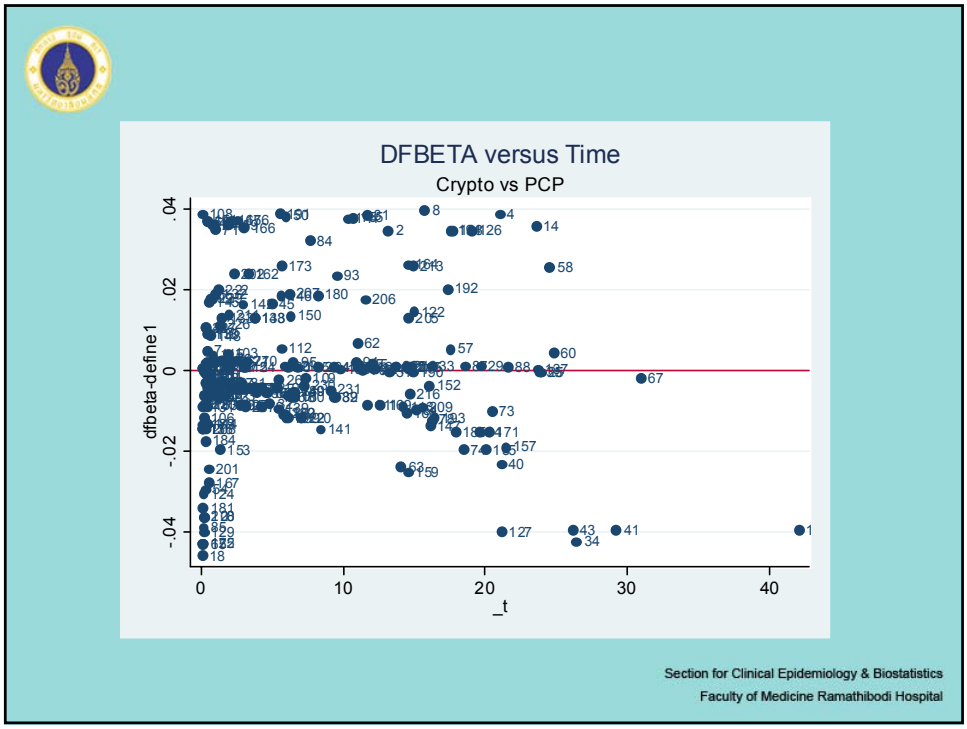
Outliers influence on estimate β_k

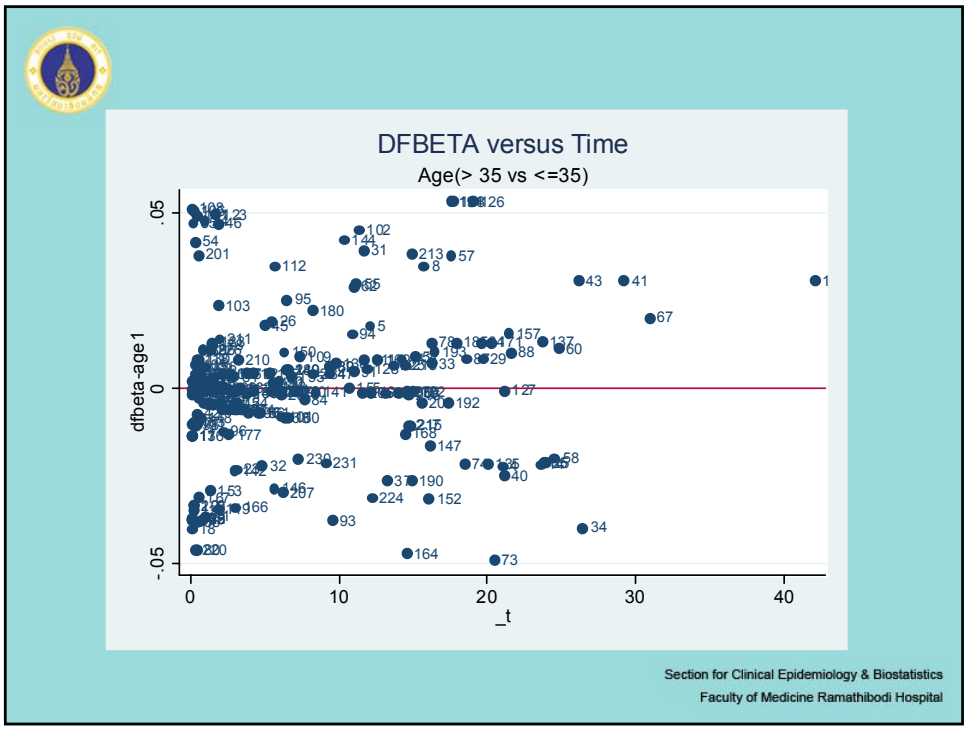
- This is the same concept as DFBETA in linear regression that the outliers influence on estimate particular coefficient k
- This can be assessed by fitting the cox hazard models with and without outliers and then compare delta change in coefficients of the two models as

$$\Delta\hat{\beta}_{ki} = \hat{\beta}_k - \hat{\beta}_{k(i)}$$

$$\Delta\hat{\beta}_{ki} = \hat{\beta}_k - \hat{\beta}_{k(i)} = \hat{Var}(\hat{\beta})\hat{L}_i$$

This $\hat{L}_i$ is called scaled (efficient) score residuals


DFBETA versus Time
Crypto vs PCP
dfbeta-define1
.04
.02
0
-.02
-.04
0
10
20
30
40
_t
Section for Clinical Epidemiology & Biostatistics
Faculty of Medicine Ramathibodi Hospital


DFBETA versus Time
Age(> 35 vs <=35)
dfbeta-age1
.05
0
-.05
0
10
20
30
40
_t
Section for Clinical Epidemiology & Biostatistics
Faculty of Medicine Ramathibodi Hospital

Time-varying covariate design

Rationale

- Taking into account possible changes overtime of covaraibles into the analysis
- should be able to explain the interested event better than considering only at baseline
- Less bias

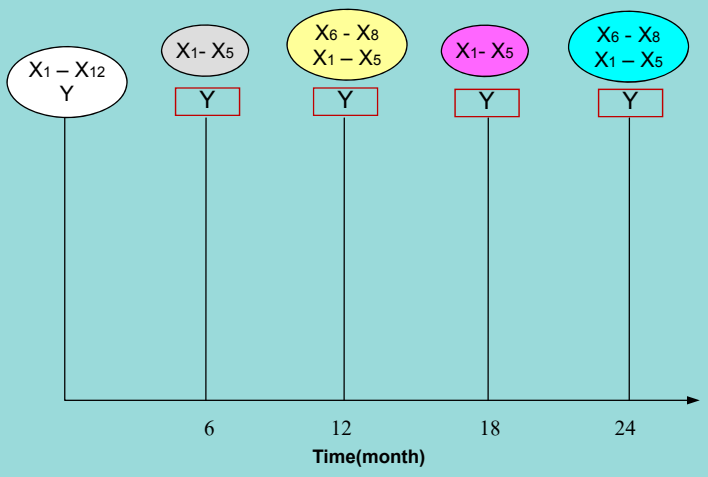
Diabetic nephropathy

- Patient may have very high HbA1C at baseline but later on after well complying with treatments, his/her HbA1C reduces to normal. Modeling only baseline for this case would account a high value of HbA1C and thus over estimate association
- Patient may not have other co-morbidities (e.g., cholesterol, uric acid, blood pressure, OA) at baseline but later on develops few co-morbidities in which influence in shortening time to Nephropathy. Again, modeling only values at baseline for this case would treat him/her as normal for these values

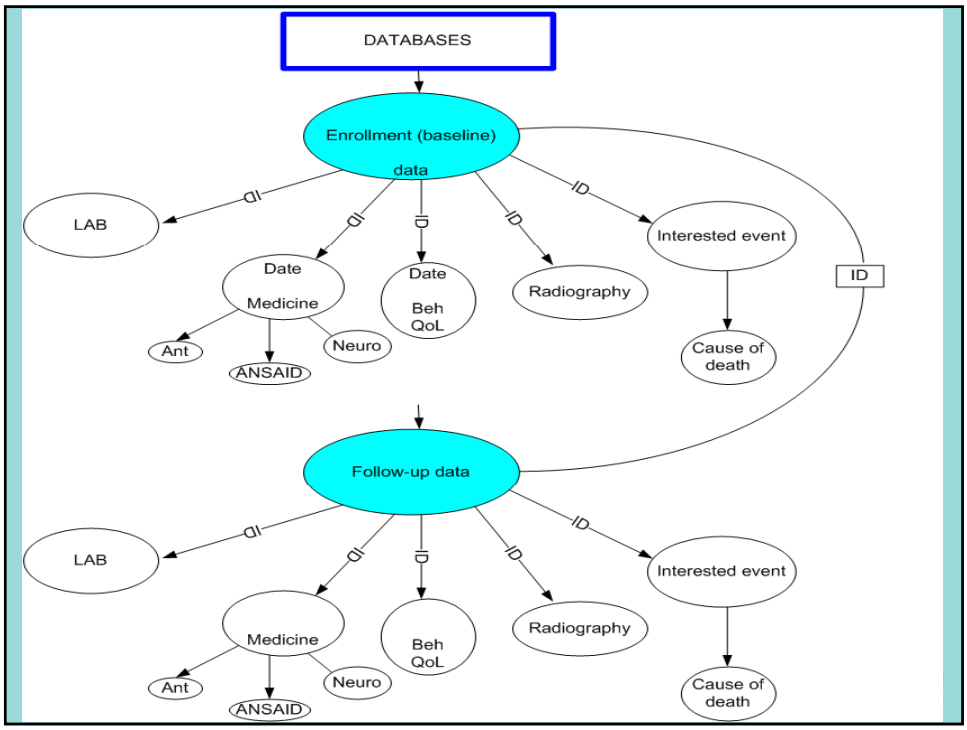
Diabetic drug & cancer

- Metformin
- Sulphonilurea
- Insulin
  - Human insulin
  - Insulin gargine

Time-varying covariate pattern



Section for Clinical Epidemiology & Biostatistics
Faculty of Medicine Ramathibodi Hospital



Cox regression equation

- Let X is a co-variable k of subject j and measured at time ti. A matrix of co-variables is

$$X_j'(t_i) = \left[x_{j1}(t_1), x_{j2}(t_2), x_{j3}(t_3), \ldots, x_{jp}(t_p)\right]$$

$$h(t, \mathbf{X}(t), \beta) = h_0(t) \exp\{\beta \mathbf{X}'(t)\}$$

$$h_{jk}(t_i) = h_0(t) \exp\left\{\sum_{k=1}^{p} \beta_k x_{jk}(t_i)\right\}$$

Example 3.

Prognosis factors of disease remission after Thymectomy in Myastenia Gravis patients

- **To assess prognostic factors of disease remission**
- **Prognostic factors**
  - **Osserman grading**
  - **Treatment**
    - **Prednisolone**
    - **Mestinon**
    - **IVIG**
    - **Azathiprine**
  - **Agegr**
  - **Duration from Dx-Operation**
  - **SEX**
- **Multiple records with single event**

```
list ID DATETHY DATEVIS  remiss in 1/20

    list ID DATETHY DATEVIS _d _t osserm1 med    in 1/20, compress

    +-------------------------------------------------------------------+
    | ID    DATETHYM     DATEVIS   _d          _t   oss~1         med |
    |-------------------------------------------------------------------|
 1. |  1   25sep2000   30sep2000    0   .16528926       2   Mest+Pred |
 2. |  1   25sep2000   19oct2000    0   .79338843       2   Mest+Pred |
 3. |  1   25sep2000   16nov2000    0   1.7190083       2   Mest+Pred |
 4. |  1   25sep2000   29mar2001    0   6.1157025       2   Mest+Pred |
 5. |  1   25sep2000   03may2001    0   7.2727273       2   Mest+Pred |
    |-------------------------------------------------------------------|
 6. |  1   25sep2000   28jun2001    0   9.1239669       2   only Pred |
 7. |  1   25sep2000   23aug2001    0   10.975207       2   only Pred |
 8. |  1   25sep2000   18oct2001    0   12.826446       2   only Pred |
 9. |  1   25sep2000   07feb2002    0   16.528926       2   Mest+Pred |
10. |  1   25sep2000   18apr2002    1   18.842975       2   only Pred |
    |-------------------------------------------------------------------|
11. |  2   29oct1996   07nov1996    0   .29752066       2   Mest+Pred |
12. |  2   29oct1996   19dec1996    0   1.6859504       2   only Pred |
13. |  2   29oct1996   30jan1997    0   3.0743802       2   only Pred |
14. |  2   29oct1996   24mar1997    0   4.8264463       2   only Pred |
15. |  2   29oct1996   05apr1997    1   5.2231405       2        none |
    |-------------------------------------------------------------------|
16. |  4   22may1997   26jun1997    1   1.1570248       3   Mest+Pred |
17. |  5   30jul1992   20aug1992    1   .69421488       2   only Mest |
18. |  6   19jan2001   05apr2001    0   2.5123967       2   Mest+Pred |
19. |  6   19jan2001   31may2001    1   4.3636364       2   only Pred |
20. |  7   18dec1990   07mar1991    0   2.6115702       2   Mest+Pred |
    +-------------------------------------------------------------------+
d: 1=remission;  0=censor
```

Section for Clinical Epidemiology & Biostatistics
Faculty of Medicine Ramathibodi Hospital

Table 1. Factors associated with disease remission after Thymectomy:Univariate analysis

Factor	Incidence rate/100/ months	Median remission time (months)	P value
Gender			
Male			
Female			
Age at thymectomy			
< 45 years			
≥ 45 years			
Duration from disease onset to thymectomy			
< 3 years			
≥ 3 years			
Osserman grade at thymectomy			
0*			
1			
2			
3 + 4			
Pathology of thymus gland			
Hyperplasia			
Atrophy			
Thymoma			
Normal			
Unknown			
Pyridostigmine			
Yes			
No			
Prednisolone			
Yes			
No			

Section for Clinical Epidemiology & Biostatistics
Faculty of Medicine Ramathibodi Hospital

Table 2 Prognostic factors of clinical remission for myasthenia gravis patients after Thymectomy: Multivariate analysis

Factors	Coefficient	SE	P value	HR	95%CI HR
Duration from disease onset to thymectomy					
< 1 years					
≥ 1 years					
Osserman grading before thymectomy					
0*					
1					
2A					
2B					
3 + 4					
Prednisolone					
Yes					
No					

Section for Clinical Epidemiology & Biostatistics
Faculty of Medicine Ramathibodi Hospital

Step i: Univariate analysis

Variables	P-value

Section for Clinical Epidemiology & Biostatistics
Faculty of Medicine Ramathibodi Hospital

Step ii) Multivariate-analysis

Step iii) interaction?

Step iv) Goodness of fit

Variables	P-value

Multiple episodes or multiple events : LDA

Outcomes:

- Unordered failure events of the same type
  - Observed in the same group e.g., family, community
    - Share genetic, eating habit, environmental factors
  - Observed in the same subject or matched sample
    - Diabetic retinopathy: Left & right eyes
  - CVD: MI, stroke

Outcome

- Unordered failure event with different types (competing risk)
  - Several events of interest
  - All subjects are at risk for all events
  - Once experiences one, still remain at risk for other events
  - Each event is randomly ordered (unordered)
  - Each occurs only once per subject
  - If k possible event, each subject will appear k times in the data set

- Diabetes: retinopathy, nephropathy, Amputation, CVD
- Cirrhosis: Death, liver transplant, varices, ascites, encephalopathy, doubling bilirubin

Section for Clinical Epidemiology & Biostatistics
Faculty of Medicine Ramathibodi Hospital

Outcomes

- **Ordered failure events**
  - **The same event but repeatedly occurs**
    - **Subject is not at risk of the 2nd event until the 1st event has occurred**
    - Lupus nephritis: Disease remission & flair
    - AIDS patients: CD4 < 400mg/dl vs >400
    - HbA1C: well controlled & poor controlled

Step of analysis

- Need to decide whether interested events are the same types
- Need to decide whether they are ordered or unordered
- Select appropriate model
- Set st data
- Use appropriate command

Models

- Anderson and Gill method
  - The same type of events but can repeatedly occur
  - All failures are equal or indistinguishable
  - The problem then reduce to analysis of time to first event, time to second event , and so on
  - Allows only one event occurs at a given time t

Model

- Marginal risk set model
  - The same as competing risk model
  - Each subject will appear k times in the data set
  - As if the failure events are unordered
  - Each event has it own stratum, thus k strata per subject
  - Analysis is stratified by strata, i.e., hazard function differs for each event

- Competing risk model
  - Different events, one is the interested event and the rest are competing risk
  - Data will record only time of first events , i.e. competing risk
  - Competing risk is not censored, it affects on survival time
  - $h(t) = h_1(t) + h_2(t)$

Model

Conditional risk set model

- Subject does not risk for the second event if the first event has not occurred, and so on
- Two variations
  - Time to each event that measured from entry time
  - Time to each event is measured from the previous event

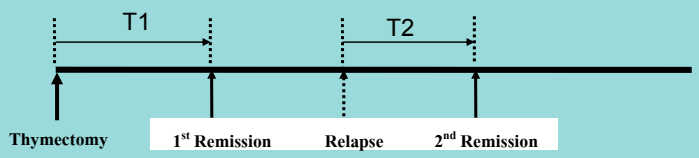
Conditional risk (cont.)

- Two approaches for analysis
  - Time from entry
    - The same as Anderson Gill concept but taking into account for number of previous failure
    - Analysis is stratified by strata
  - Time from previous failure
    - Time is not measured continuously from study entry, but the clock is set to zero after each failure event has occurred
    - Fix model (variance within subject is constant)
    - Frailty model
      - taking into account within subject's variation, i.e. random effect model

Frailty model

$$h(t_{ij}) = h_0(t)\exp(x_{ij}\beta + \zeta_i)$$

MG patients after Thymectomy



Declare st data

- `stset DATEVIS, origin(DATETHY) fail(remiss==1) id(ID) exit(time .) scale(30.25)`

```
id:  ID
     failure event:  remiss == 1
obs. time interval:  (DATEVIS[_n-1], DATEVIS]
 exit on or before:  time .
    t for analysis:  (time-origin)/30.25
            origin:  time DATETHYM
--------------------------------------------------------------------------
     4076  total obs.
        0  exclusions
--------------------------------------------------------------------------
     4076  obs. remaining, representing
      250  subjects
     1902  failures in multiple failure-per-subject data
 22831.83  total analysis time at risk, at risk from t =         0
                             earliest observed entry t =         0
                                  last observed exit t =   430.843
```

```
Stgen nf=nfailures()
stci, by(nf)
        failure _d:  remiss == 1
  analysis time _t:  (DATEVIS-origin)/30.25
            origin:  time DATETHYM
 exit on or before:  time .
                id:  ID
             |    no. of
nf           |  subjects        50%     Std. Err.     [95% Conf. Interval]
-------------+------------------------------------------------------------
           0 |       250    8.429752      .6625119      6.90909    10.4463
           1 |       225    3.702479       .286797      2.84298    5.05785
           2 |       206    6.710744      .4508909      2.08264    9.91736
           3 |       178    3.867769      .0484973      3.47107    8.89256
           4 |       155    13.65289      .1060136      8.46281    18.3802

          36 |         2     161.124             .            .          .
          37 |         2    175.4711      2.80e-06            .          .
-------------+------------------------------------------------------------
       total |       250    6.347107      .5020846      5.22314    7.27273
```


Competing risk example

Therneau and Grambsch (2000, sec. 8.4.3) analyze data from patients with primary biliary cirrhosis (PBC), a chronic liver disease characterized by progressive destruction of the bile ducts. Data were obtained from 170 patients in a randomized double-blind trial conducted at the Mayo Clinic from 1988 to 1992. The trial was for a new treatment, ursodeoxycholic acid (UDCA; Lindor et al. [1994]).

Table 1. Event codes for the `etype` variable

Event Code	Event type
0	No event (censored)
1	Death
2	Transplant
3	Histologic progression
4	Development of varices
5	Development of ascites
6	Development of encephalopathy
7	Doubling of bilirubin
8	Worsening of symptoms