Cox regression and time-dependent model

醫學研究部生統小組

陳俊朋

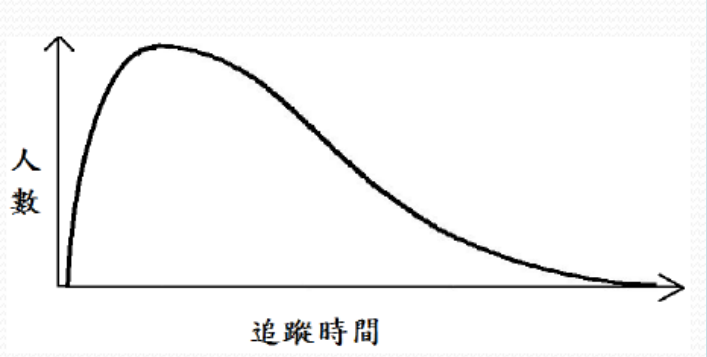
2025/5/13

統計方法

- 無母數分析
  - Kaplan-Meier
  - Log-rank test
  - Cox Proportional Hazards Model
    - Time-varying covariates

Survival Analysis

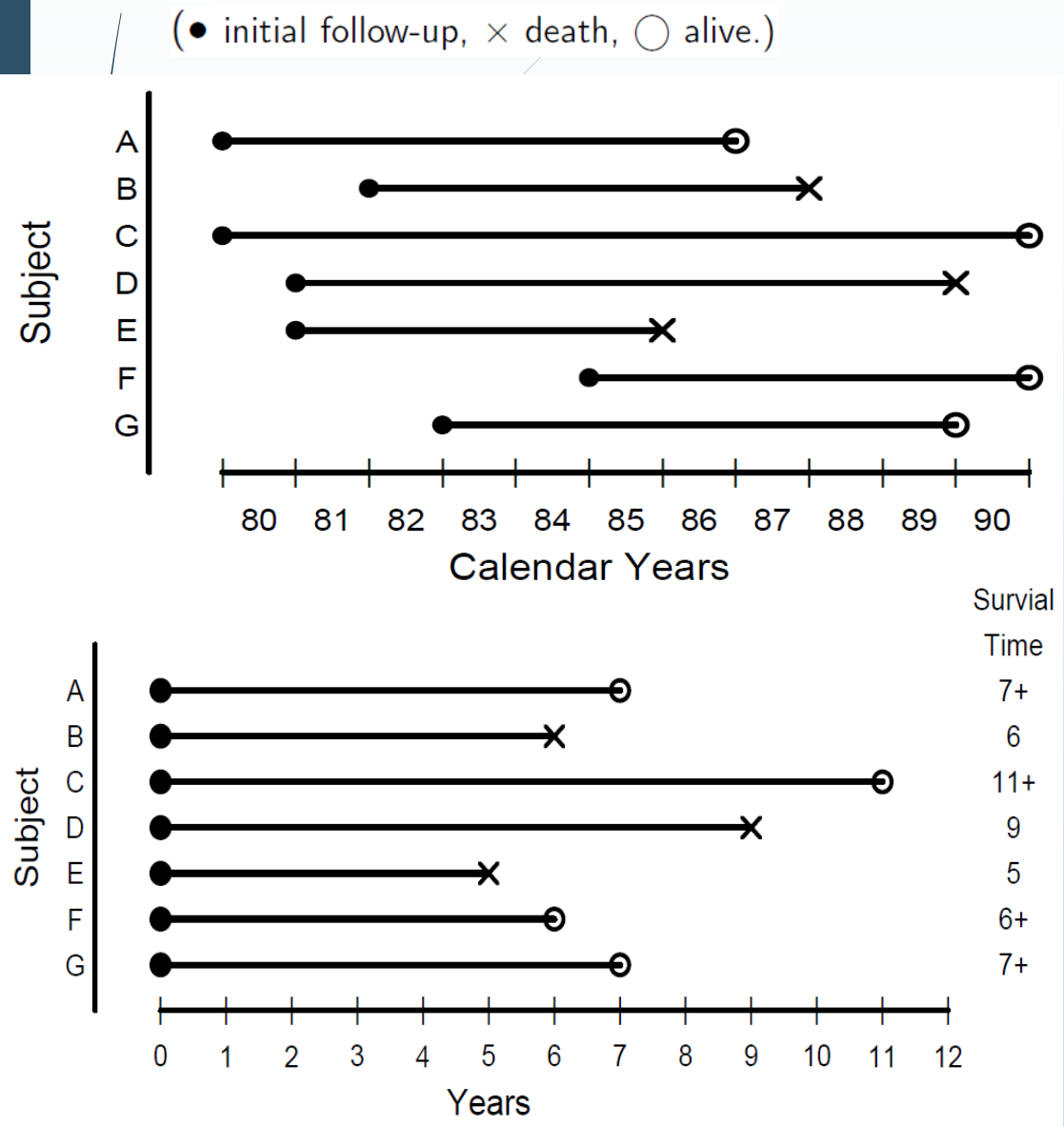
- 將『時間』變項列入分析的統計方法
- 從一時間點至事件(event)發生的時間(time to event)，稱為survival time
  - Start time- 研究起始時間/確診癌症時間
  - End time- 疾病發生時間/死亡時間
- 資料特性：資料通常不是常態分配



資料分類

- 完整資料(complete data)
  - 觀察期間提供事件發生的時間點
- 設限資料(censored data)
  - 觀察期間失去聯絡或觀察結束時仍未發生事件
- 競爭死因資料(competing cause of deaths)
  - 觀察期間死於其他死因之時間

Survival time



- Survival time= 事件發生時間 – 追蹤的起始時間(需大於0)

- 個案BDE是完整資料；個案ACFG為設限資料

林建甫(2008) 存活分析。雙葉書廊

存活資料建檔

基本資料　存活狀態　起訖時間　分析整理

ID	Sex	Age	Tx	event (death)	Start_date	Death_date	follow_date	End_date	follow_year
Case1	0	45	0	0	2012/5/8		2016/7/15	2016/7/15	4.19
Case2	1	55	0	1	2013/12/9	2015/1/16		2015/1/16	1.10
Case3	0	56	0	0	2012/5/10		2014/3/12	2014/3/12	1.84
Case4	1	51	1	1	2014/9/11	2016/7/18		2016/7/18	1.85
Case5	0	62	1	1	2012/8/12	2016/12/19		2016/12/19	4.35
Case6	1	70	0	1	2018/7/13	2019/7/20		2019/7/20	1.02
Case7	1	58	0	0	2015/5/14		2016/8/21	2016/8/21	1.27
Case8	0	66	0	0	2018/8/15		2019/7/22	2019/7/22	0.93
Case9	0	60	1	0	2019/5/16		2019/7/23	2019/7/23	0.19
Case10	0	63	1	0	2017/10/17		2018/7/24	2018/7/24	0.77
Case11	0	66	1	0	2016/3/20		2017/7/25	2017/7/25	1.35
Case12	0	69	1	0	2014/8/22		2016/7/26	2016/7/26	1.93

- 日期格式統一西元年
- 可分組比較存活曲線差異，如Sex (F vs M), Age (<60 vs >=60)
- 事件(event): 0=存活; 1=死亡
- 未死亡個案End_date，研究者需設定結束時間(建議可用最後看診日期)

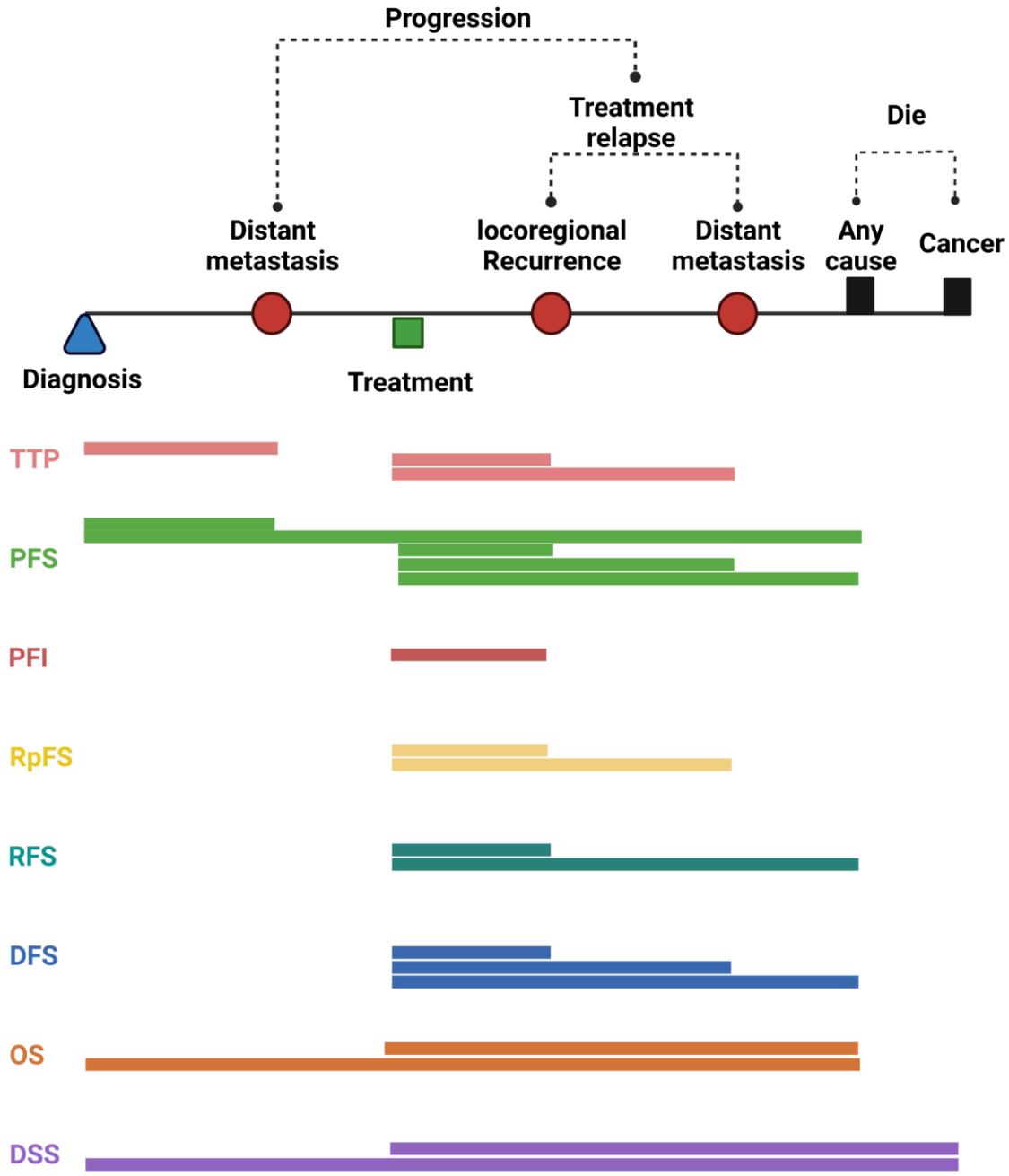
Definition of Time to Event Variables

指標	事件	設限
1.無疾病進展存活期 (progression free survival, PFS)	疾病進展或死亡	無疾病進展
2.無疾病存活期 (disease free survival, DFS)	疾病復發或死亡	無疾病復發
3.治療至疾病進展時間 (time to progression, TTP)	疾病進展	無疾病進展、無疾病狀態死亡(如車禍或不明原因)
4.治療至治療失敗時間 (time to treatment failure, TTF)	疾病進展、疾病復發、藥物副作用、死亡	無疾病進展或復發
5.總存活時間 (overall survival, OS)	死亡	存活
6. 疾病特定存活期 (disease specific survival, DSS)	由疾病本身所導致的死亡(車禍或其他原因要歸為設限資料)	存活、因其他原因死亡

莊其穆(2011) 臨床醫師如何執行第二期臨床試驗。臺灣醫界54卷7期

Definition for the Assessment of Time-to-event Endpoints in Cancer

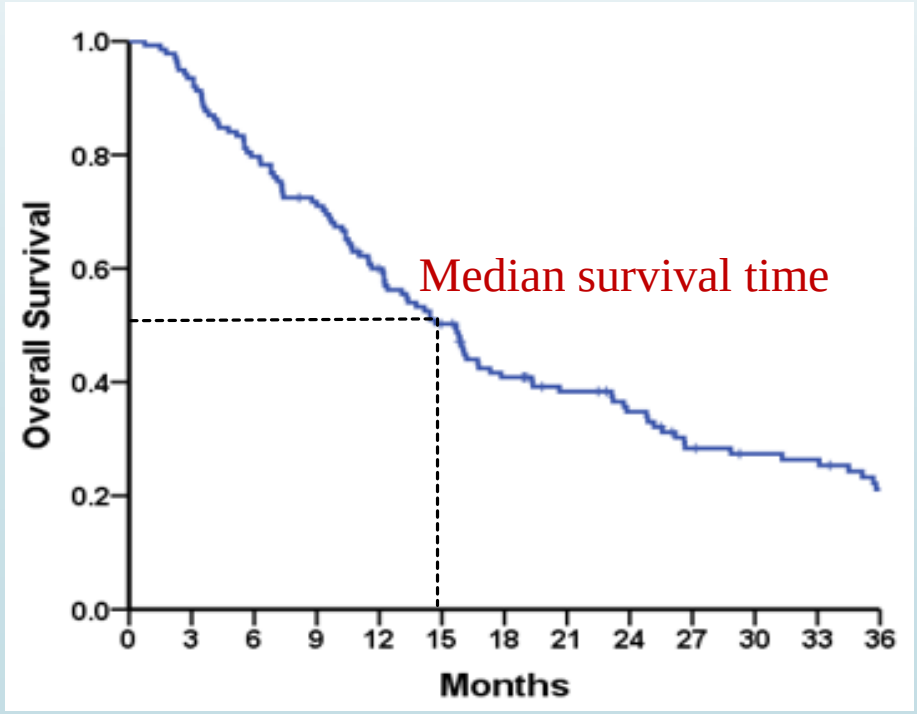
Time to progress	The time from diagnosis/treatment to the progression of tumor (in any aspect).
Progression-free survival	The time from diagnosis/treatment to the progression of tumor (in any aspect) or death (for any cause).
Progression-free interval	The time from treatment to locoregional recurrence.
Relapse-free survival	The time from treatment to the relapse (local, regional, distant).
Recurrence-free survival	The time from treatment to locoregional recurrence or death (for any cause)
Disease-free survival	The time from treatment to locoregional recurrence, metastasis or death (for any reason)
Overall survival	The time from diagnosis/treatment to death (for any reason).
Disease-specific survival	The time from diagnosis/treatment to death (cancer-specific).

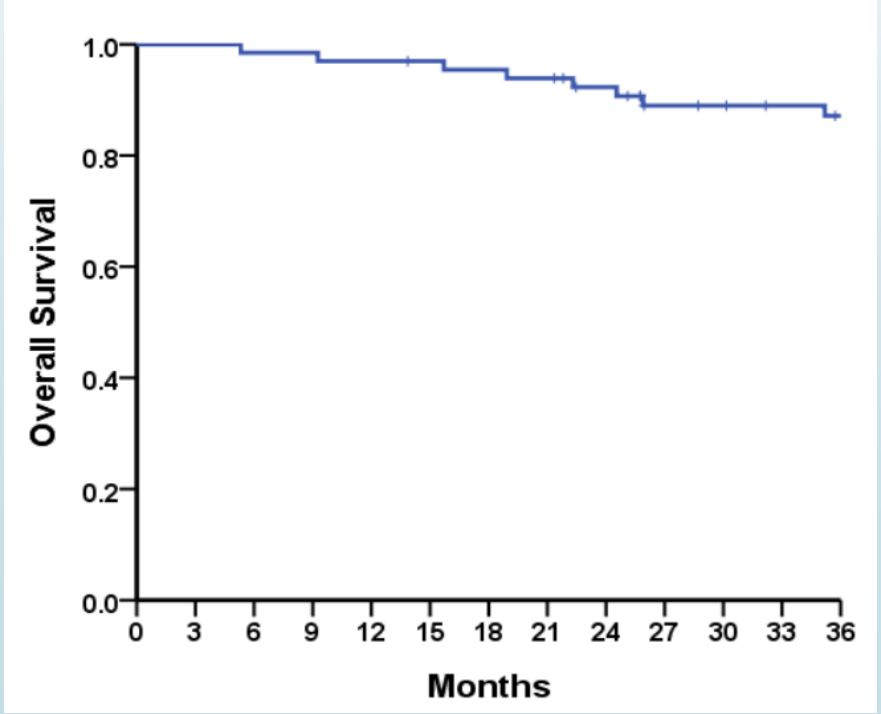


Toraih, Eman A., et al. "A miRNA-based prognostic model to trace thyroid cancer recurrence." Cancers 14.17 (2022): 4128.

Kaplan-Meier

- 常用來估計存活曲線的方法，此方法用每一個事件發生時間點及設限點來設定區間
- 可估計存活中位數及不同時間點的存活率





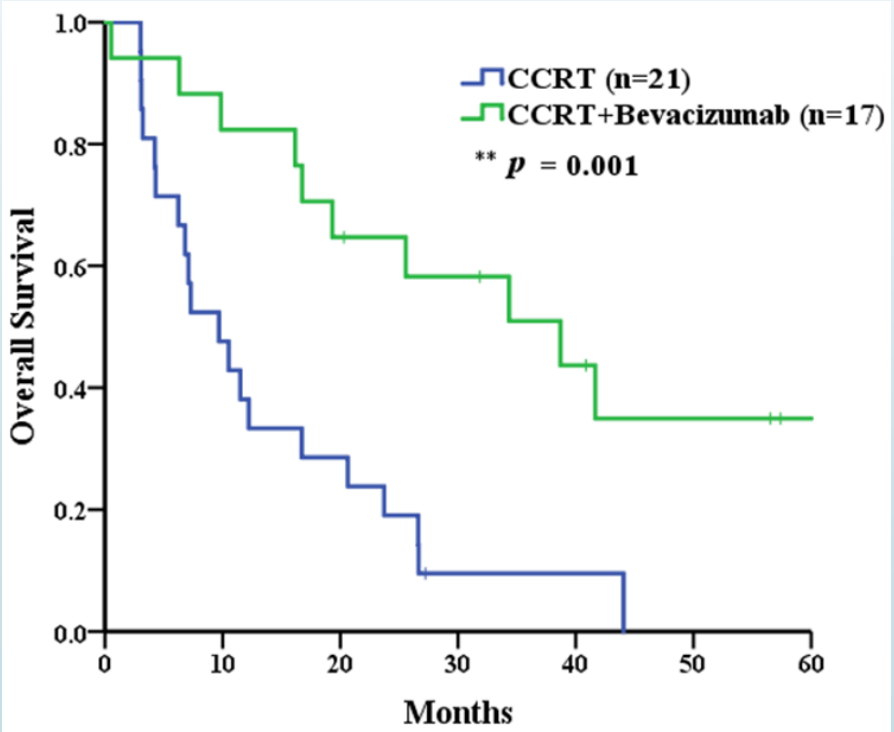
Kaplan-Meier Estimates

ID	fu_time (month)	status
1	2	0
2	2	0
3	2	1
4	3	0
5	3	0
6	4	1
7	4	0
8	5	0
9	6	1
10	6	0

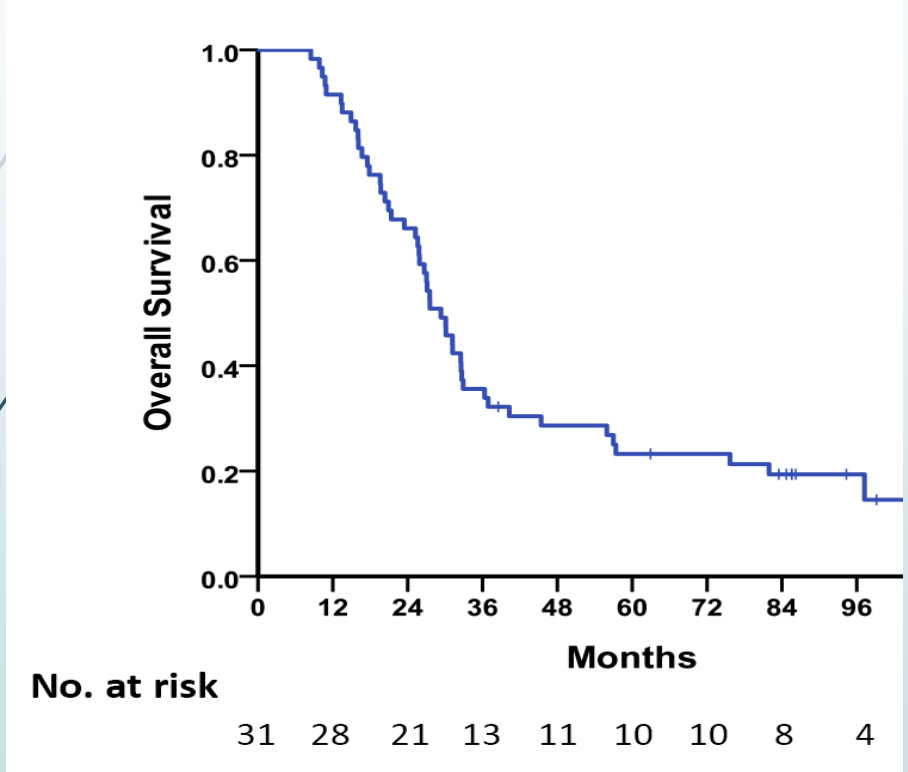
month	Number at risk	Number of events	Number of censored	Conditional Probability	Survival Function
1	10	0	0	10/10 = 1.00	1.00
2	10	1	2	9/10 = 0.90	0.90*1.00 = 0.90
3	7	0	2	7/7 = 1.00	1.00*0.90 = 0.90
4	5	1	1	4/5 = 0.80	0.80*0.90 = 0.72
5	3	0	1	3/3 = 1.00	1.00*0.72 = 0.72
6	2	1	1	1/2 = 0.50	0.50*0.72 = 0.36

Log-rank Test

- Kaplan-Meier法僅能了解不同組別的存活曲線分佈
- 比較組別間是否差異
  - H0: 兩條存活曲線相同
  - H1: 兩條存活曲線不相同



Presentation of Survival Curve



Survival Table

	Time	Status	Cumulative Proportion Surviving at the Time		N of Cumulative Events	N of Remaining Cases
			Estimate	Std. Error		
1	8.467	Death	.968	.032	1	30
12M	9.833	Death	.935	.044	2	29
	10.767	Death	.903	.053	3	28
4	14.900	Death	.871	.060	4	27
5	15.667	Death	.839	.066	5	26
6	16.667	Death	.806	.071	6	25
7	17.533	Death	.774	.075	7	24
8	17.833	Death	.742	.079	8	23
24M	20.933	Death	.710	.082	9	22
	23.467	Death	.677	.084	10	21
11	25.233	Death	.645	.086	11	20
12	25.800	Death	.613	.087	12	19
13	27.033	Death	.581	.089	13	18
14	27.533	Death	.548	.089	14	17
15	27.567	Death	.516	.090	15	16
16	29.333	Death	.484	.090	16	15
36M	30.133	Death	.452	.089	17	14
	32.500	Death	.419	.089	18	13
48M	40.300	Death	.387	.087	19	12
	45.400	Death	.355	.086	20	11
60M	55.933	Death	.323	.084	21	10
	75.667	Death	.290	.082	22	9
84M	81.933	Death	.258	.079	23	8
	85.567	Survival	.	.	23	7
25	85.600	Survival	.	.	23	6
96M	86.233	Survival	.	.	23	5
	94.333	Survival	.	.	23	4
28	97.233	Death	.194	.081	24	3
29	99.167	Survival	.	.	24	2
30	104.167	Survival	.	.	24	1
31	105.367	Survival	.	.	24	0

Presentation of Survival Curve

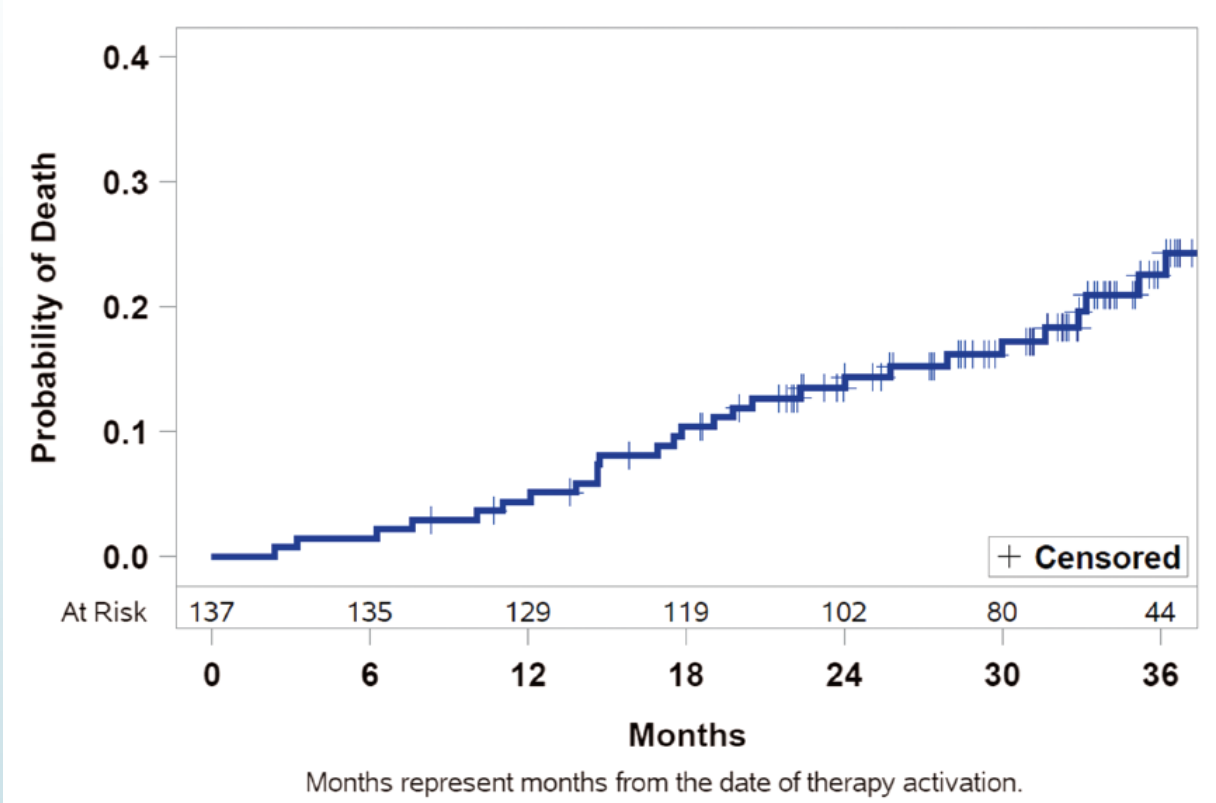


Figure 5. Kaplan–Meier curve of mortality. Kaplan–Meier curve showing estimated mortality through 36 months of active therapy using the pooled treatment and former control groups. Patients who did not die were censored at last contact if they did not reach 36 months of active therapy.

Long-term efficacy and safety of phrenic nerve stimulation for the treatment of central sleep apnea. *Sleep 42.11 (2019): zsz158*.

Cox Proportional Hazards Model

- 評估多個變數對存活(Time to event)的影響
- 可使用類別/連續變數呈現危險因子，並估算出這些危險因子對outcome的影響

$$\log\frac{h(t)}{h_0(t)} = \beta_1 x_1 + \beta_2 x_2 + \cdots + \beta_k x_k$$

Hazard ratio (HR)	Explanation
HR=1	Two groups have the same survival experience
HR>1	Survival is better in the control group
HR<1	Survival is better in the treatment group

Kaplan-Meier Example

- 比較在第三期子宮內膜癌術後治療，Sandwich組和CT alone兩組治療預後的差異

frontiers | Frontiers in Oncology

TYPE Original Research
PUBLISHED 23 September 2022
DOI 10.3389/fonc.2022.946113

IF: 4.7 Rank: Q2

Check for updates

OPEN ACCESS

EDITED BY
Shannon Neville Westin,
University of Texas MD Anderson Cancer Center, United States

REVIEWED BY
Melissa Young,
Yale University, United States
Katherine Fuh,
Washington University in St. Louis, United States

*CORRESPONDENCE
Chien-Hsing Lu
chlu@vghtc.gov.tw

Outcomes of "sandwich" chemoradiotherapy compared with chemotherapy alone for the adjuvant treatment of FIGO stage III endometrial cancer

Shao-Jing Wang[1], Lily Wang[2], Lou Sun[1], Yu-Hsiang Shih[1], Shih-Tien Hsu[1,3,4], Chin-Ku Liu[1], Sheau-Feng Hwang[1,5] and Chien-Hsing Lu[1,6*]

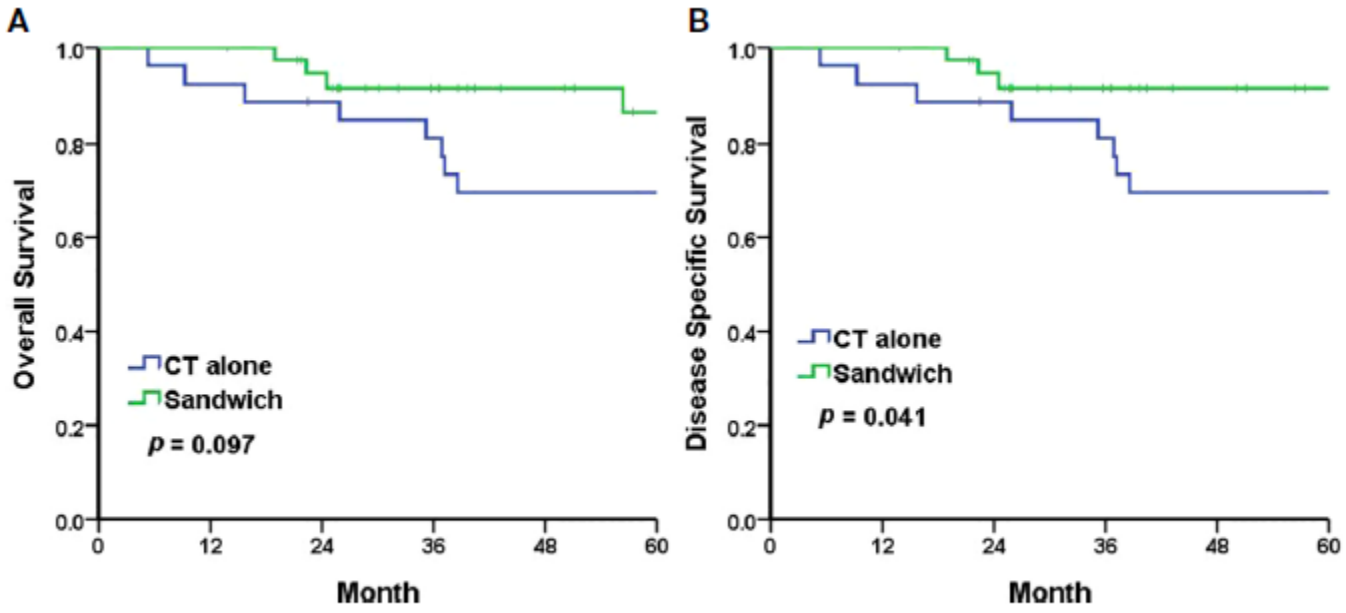


FIGURE 2
The Kaplan–Meier survival curves for 5-year overall survival **(A)** and 5-year disease-specific survival **(B)**. CT chemotherapy.

Outcomes of" sandwich" chemoradiotherapy compared with chemotherapy alone for the adjuvant treatment of FIGO stage III endometrial cancer. *Frontiers in oncology, 12 (2022), 946113-946113.*

Cox Regression Example

5-year disease-specific survival

	Univariate			Multivariable		
	HR	95% CI	*p*-value	HR	95% CI	*p*-value
Age group						
<60	Reference					
≥60	1.34	(0.36-5.07)	0.665			
FIGO stage						
IIIA and IIIB and IIIC1	Reference					
IIIC2	1.45	(0.44-4.75)	0.541			
Histology grading						
Grades 1 and 2	Reference			Reference		
Grade 3	8.70	(1.11-68.01)	0.039*	9.16	(1.17-71.70)	0.035*
Treatment						
CT alone	Reference			Reference		
Sandwich	0.27	(0.07-1.04)	0.056	0.23	(0.06-0.87)	0.030*
LVSI						
Absent	Reference					
Present	3.82	(0.49-29.89)	0.201			
Deep myometrial invasion						
Absent	Reference			Reference		
Present	7.85	(1.00-61.34)	0.050	9.44	(1.20-74.15)	0.033*

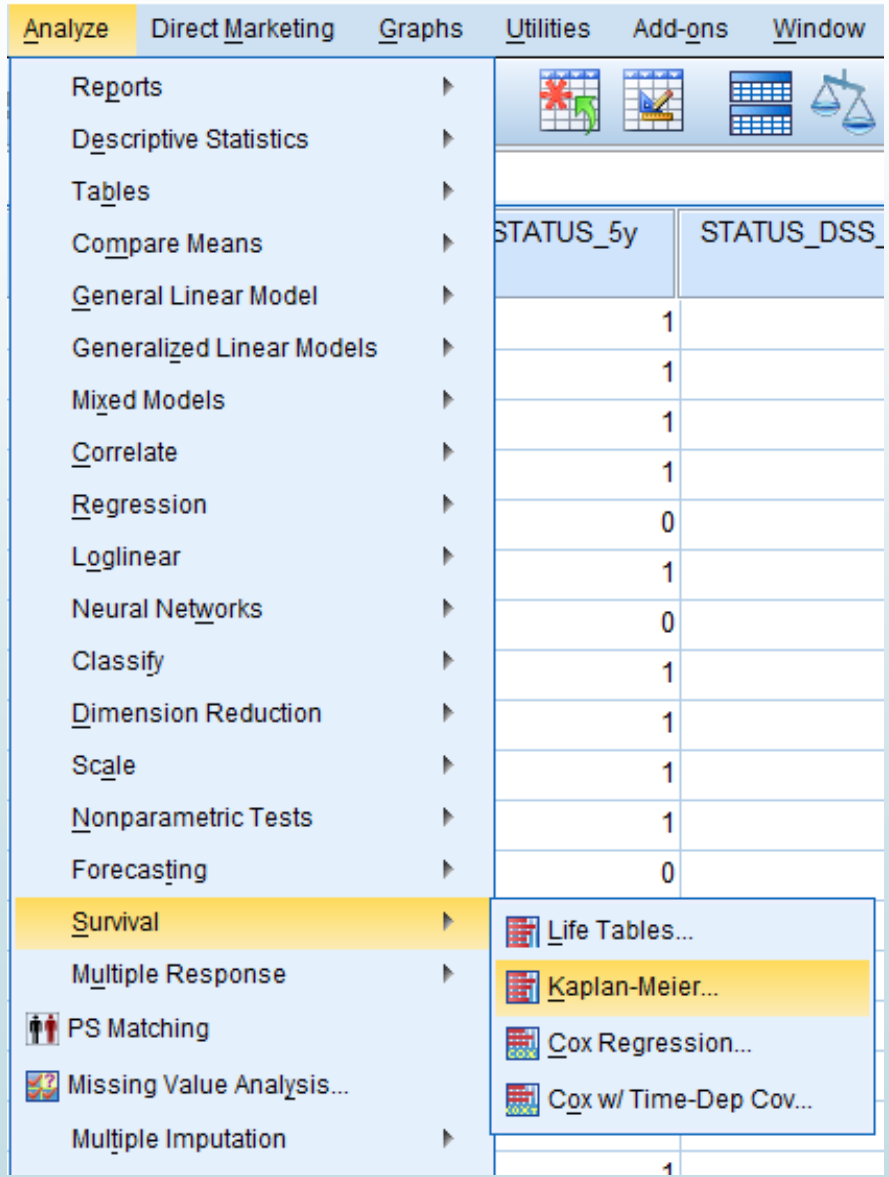
Outcomes of" sandwich" chemoradiotherapy compared with chemotherapy alone for the adjuvant treatment of FIGO stage III endometrial cancer. *Frontiers in oncology, 12 (2022), 946113-946113.*

KM and Cox Regression SPSS dataset

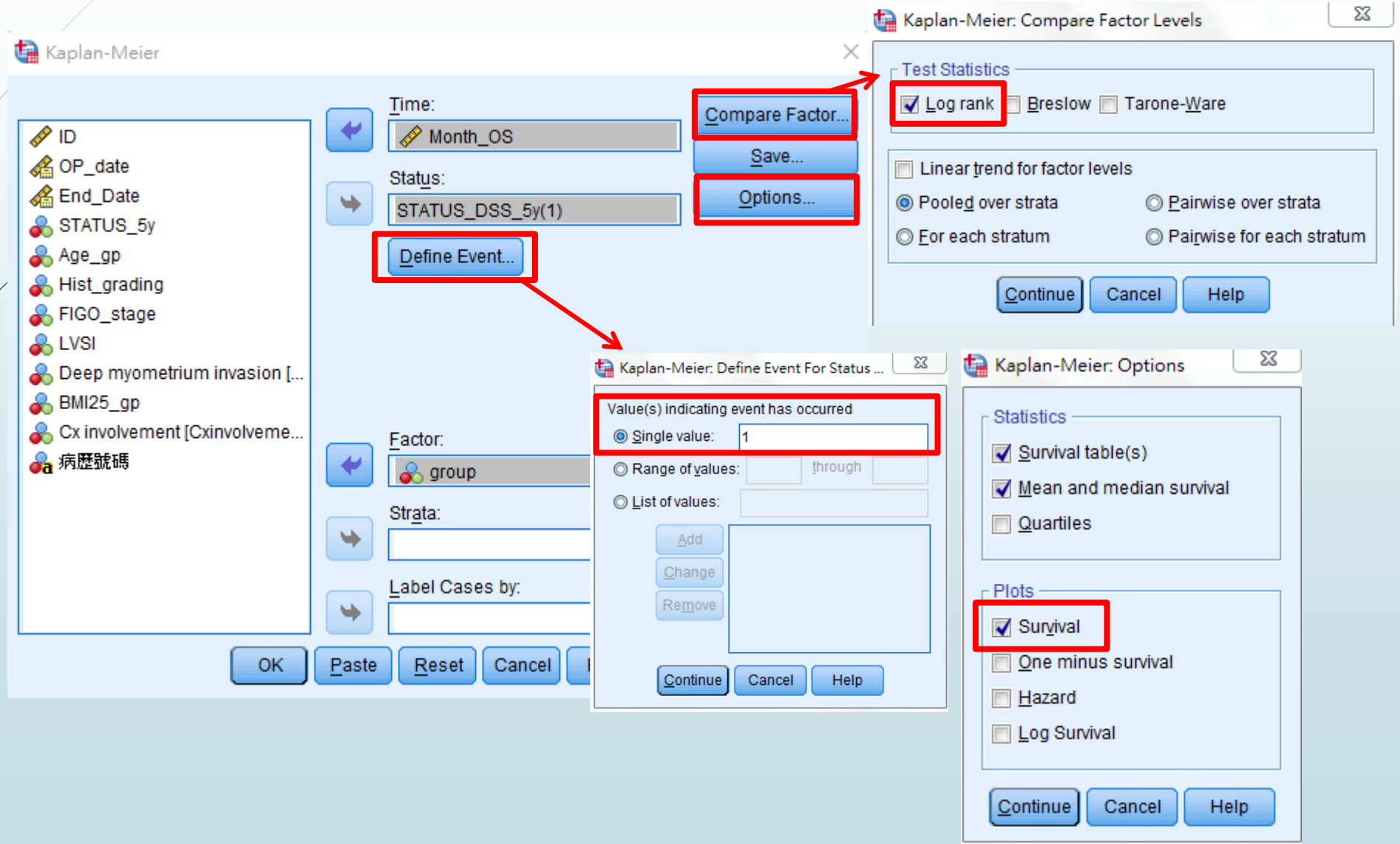
ID	group	OP_date	End_Date	STATUS_5y	STATUS_DSS_5y	Month_OS	Age_gp	Hist_grading
1	1	2016/04/20	2021/09/09	1	0	64.66	0	2
2	1	2012/04/27	2020/02/20	1	0	93.80	1	1
3	1	2015/12/30	2021/01/19	1	0	60.68	0	1
4	0	2010/11/12	2016/04/02	1	0	64.66	0	2
5	1	2012/11/21	2014/10/01	0	1	22.31	0	2
6	1	2013/10/08	2020/02/12	1	0	76.16	0	1
7	0	2006/03/29	2008/05/25	0	1	25.89	0	1
8	1	2012/06/19	2020/12/23	1	0	102.14	1	2
9	1	2018/10/29	2021/10/20	1	0	35.71	1	1
10	1	2019/08/27	2021/10/18	1	0	25.72	0	2
11	0	2010/02/05	2017/11/24	1	0	93.60	0	2
12	0	2011/02/24	2014/05/14	0	1	38.60	0	2
13	1	2017/08/25	2021/01/05	1	0	40.38	1	1
14	1	2012/04/20	2015/07/10	1	0	38.64	0	2
15	0	2005/10/06	2015/02/18	1	0	112.43	0	2
16	1	2018/06/25	2021/10/14	1	0	39.66	0	2
17	1	2015/04/08	2021/09/28	1	0	77.70	1	1
18	1	2014/10/17	2021/10/27	1	0	84.34	0	1
19	0	2012/08/03	2013/05/12	0	1	9.26	0	2
20	1	2013/01/14	2017/09/26	0	0	56.38	0	2

Kaplan-Meier 分析操作

分析>存活分析
>Kaplan-Meier 統計



Kaplan-Meier 分析操作



Kaplan-Meier Output

Case Processing Summary

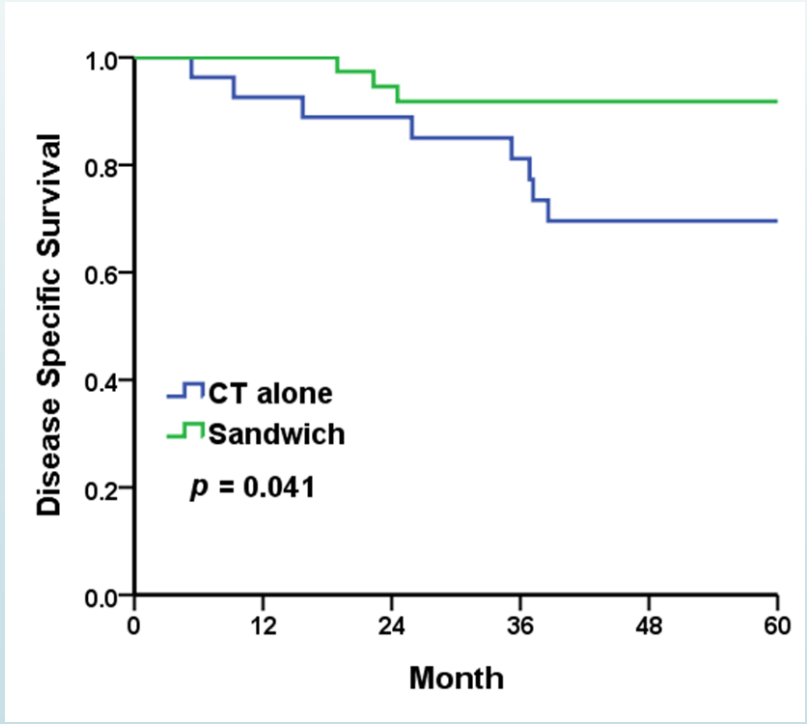
group	Total N	N of Events	Censored	
			N	Percent
CT alone	27	8	19	70.4%
Sandwich	39	3	36	92.3%
Overall	66	11	55	83.3%

Disease specific survival

	Total	DSS	Censored		Survival rate (%)			p for log rank
			n	%	1y	3y	5y	
Treatment								
CT alone	27	8	19	70.4%	92.6%	81.2%	69.6%	0.041
Sandwich	39	3	36	92.3%	100.0%	97.4%	91.8%	

Survival Table

group		Time	Status	Cumulative Proportion Surviving at the Time		N of Cumulative Events	N of Remaining Cases
				Estimate	Std. Error		
CT alone	1	5.322	DSS	.963	.036	1	26
	2	9.265	DSS	.926	.050	2	25
	3	15.704	DSS	.889	.060	3	24
	4	22.472	no-DSS	.	.	3	23
	5	25.889	DSS	.850	.069	4	22
	6	35.187	DSS	.812	.076	5	21
	7	36.862	DSS	.773	.082	6	20
	8	37.191	DSS	.734	.086	7	19
	9	38.604	DSS	.696	.090	8	18
	10	63.639	no-DSS	.	.	8	17
	11	64.526	no-DSS	.	.	8	16
	12	64.657	no-DSS	.	.	8	15
	13	72.279	no-DSS	.	.	8	14
	14	85.290	no-DSS	.	.	8	13
	15	89.823	no-DSS	.	.	8	12



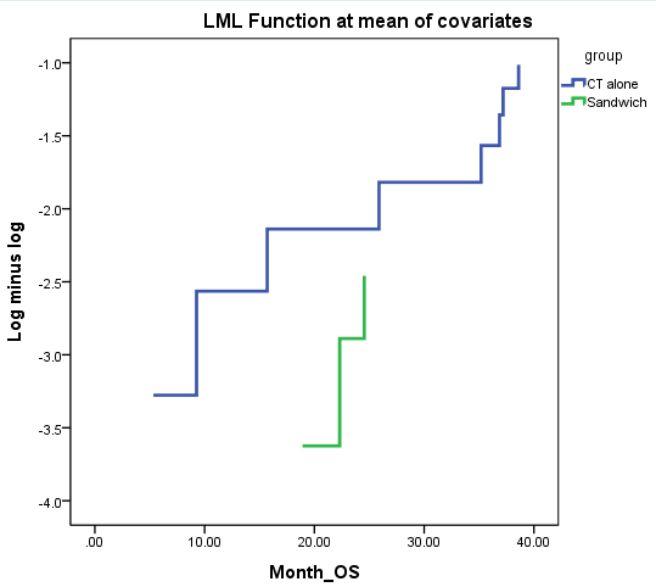
Overall Comparisons

	Chi-Square	df	Sig.
Log Rank (Mantel-Cox)	4.169	1	.041

Test of equality of survival distributions for the different levels of group.

Cox PH assumption

- 透過Cox model LML圖形判定是否符合假設
  - 交叉代表未符合
  - 平行代表符合

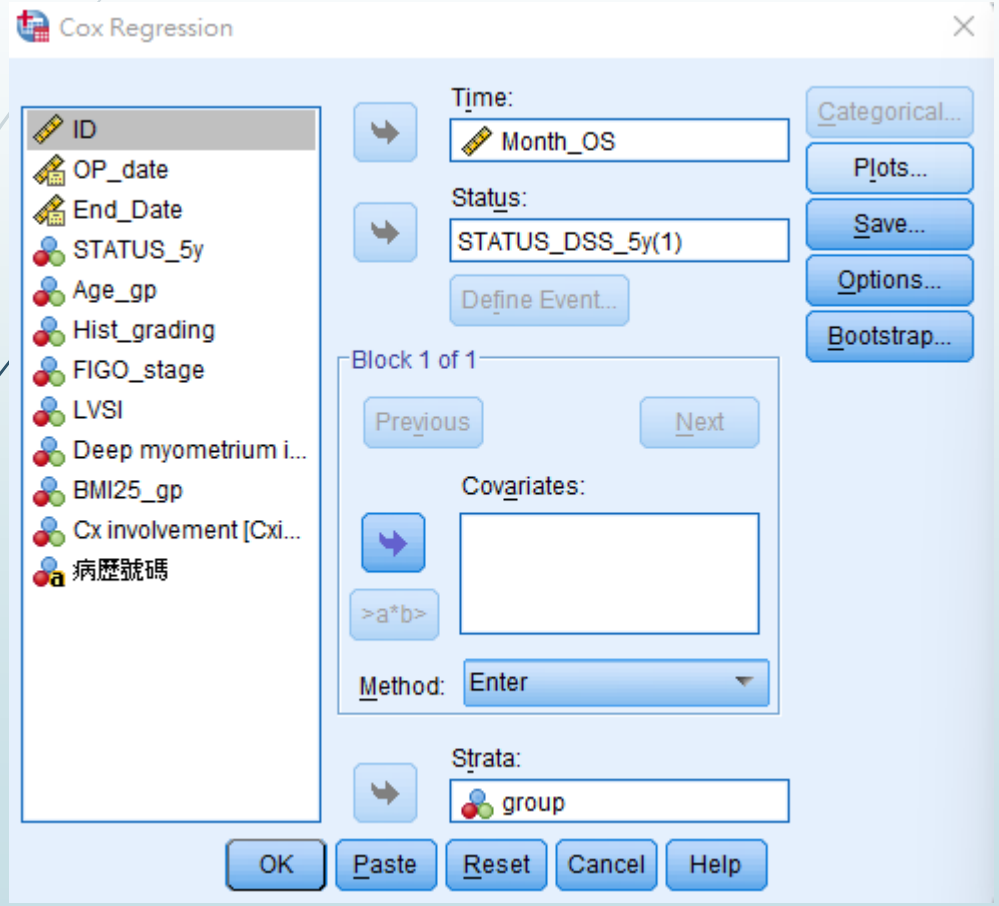


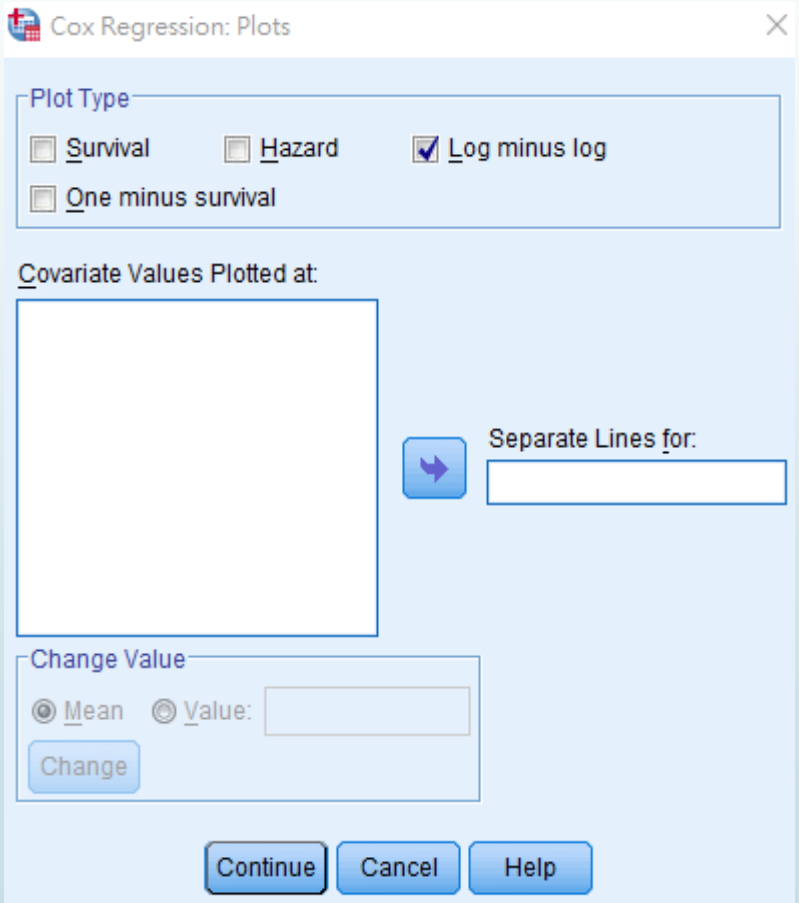
- 在Cox model加入time-dependent變數的交互作用項判定是否符合假設
  - 顯著代表未符合
  - 未顯著代表符合

Variables in the Equation

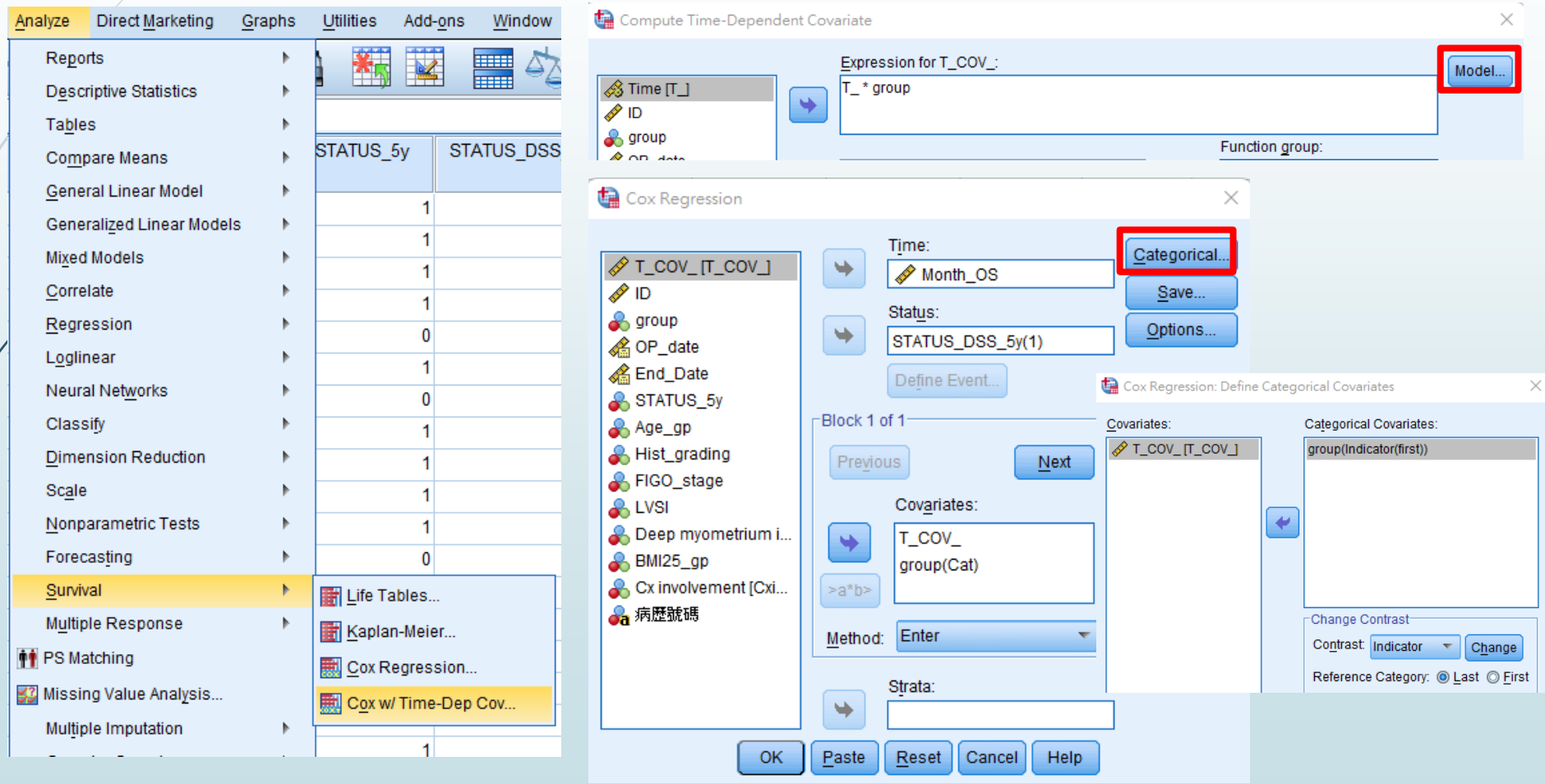
	B	SE	Wald	df	Sig.	Exp(B)
T_COV_	-.022	.062	.120	1	.729	.979
group	-.786	1.581	.247	1	.619	.456

Cox PH assumption- Log minus log (LML)



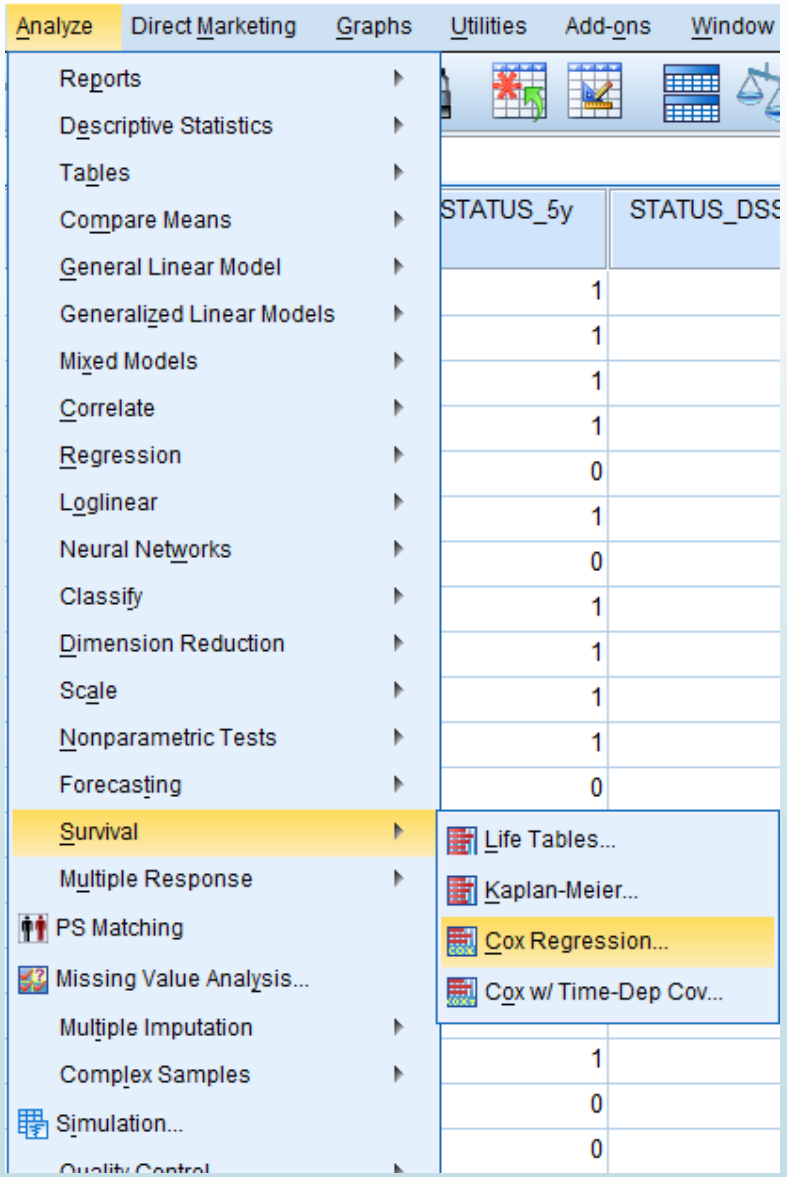


Cox PH assumption-交互作用項

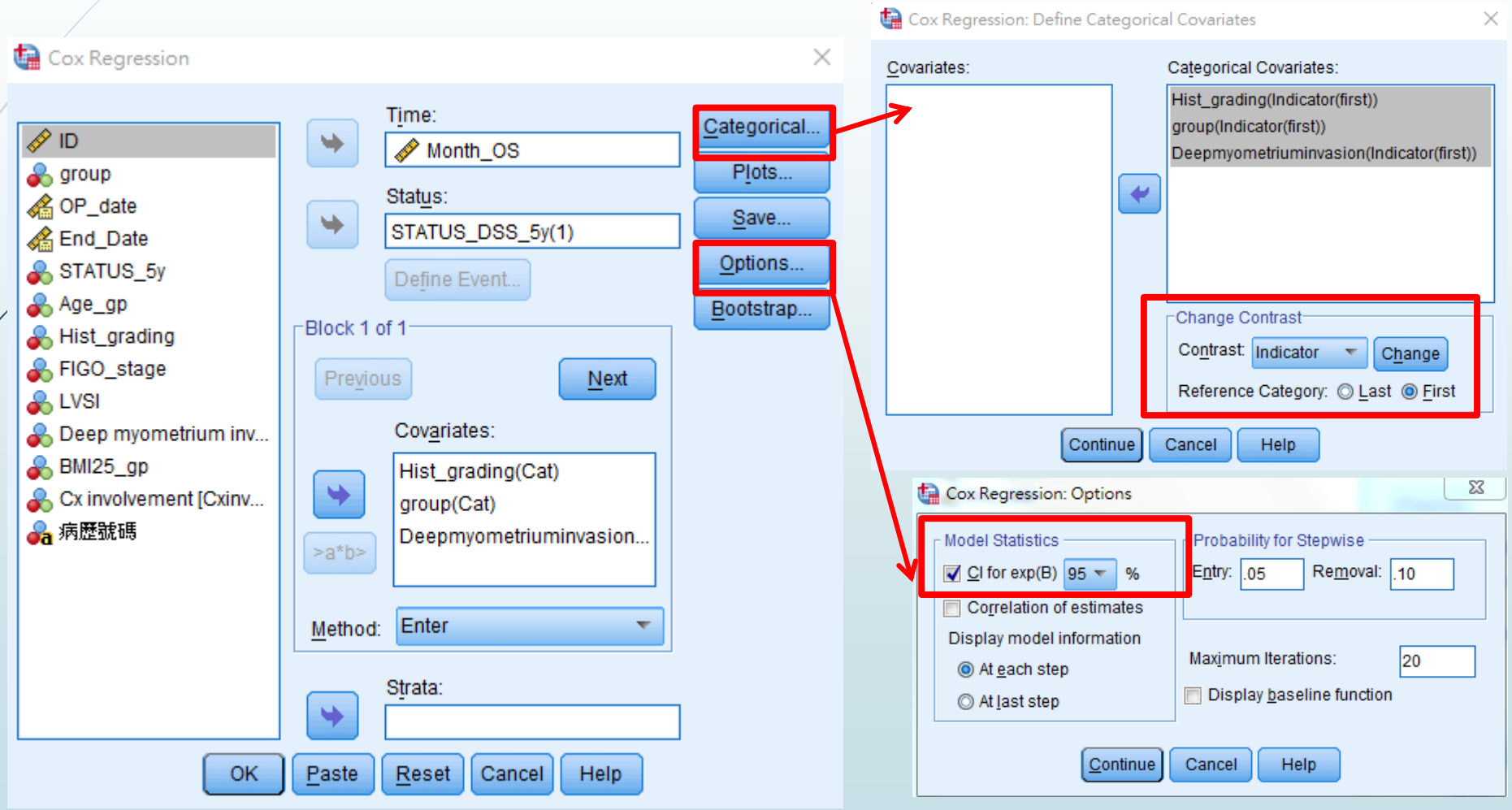


Cox model-分析操作

分析>存活分析>Cox迴歸



Cox model-分析操作



Cox model-Output

Categorical Variable Codings[a,c,d]

		Frequency	(1)
group[b]	0=CT alone	27	0
	1=Sandwich	39	1
Hist_grading[b]	1=Grade 1 and 2	29	0
	2=Grade 3	37	1
Deepmyometriuminvasion[b]	0=Absent	27	0
	1=Present	39	1

a. Category variable: group

b. Indicator Parameter Coding

c. Category variable: Hist_grading

d. Category variable: Deepmyometriuminvasion (Deep myometrium invasion)

經多變項調整後, Sandwich相較於CT alone降低0.23倍的風險死於子宮內膜癌且有統計差異(p= 0.037)

Variables in the Equation

	B	SE	Wald	df	Sig.	Exp(B)	95.0% CI for Exp(B)	
							Lower	Upper
Hist_grading	2.215	1.050	4.452	1	.035	9.161	1.170	71.701
group	-1.474	.679	4.711	1	.030	.229	.060	.867
Deepmyometriuminvasion	2.245	1.051	4.559	1	.033	9.442	1.202	74.150

Cox model with time-dependent covariates

- 稱為隨時間變動的共變數(Time-varying covariate)
- 自變項會隨著追蹤時間所變化(短時間效應)
  - 藥物劑量、血壓、年齡
  - 出血狀態、中風

$$h(t|X(t)) = h_0(t)\exp(\beta_1 X_1(t) + \beta_2 X_2(t) + ... + \beta_p X_p(t))$$

Cox model with time-dependent covariates

- 在短期內體重不足是洗腎患者死亡率的重要風險因素，但長期反而是體重過重更為明顯

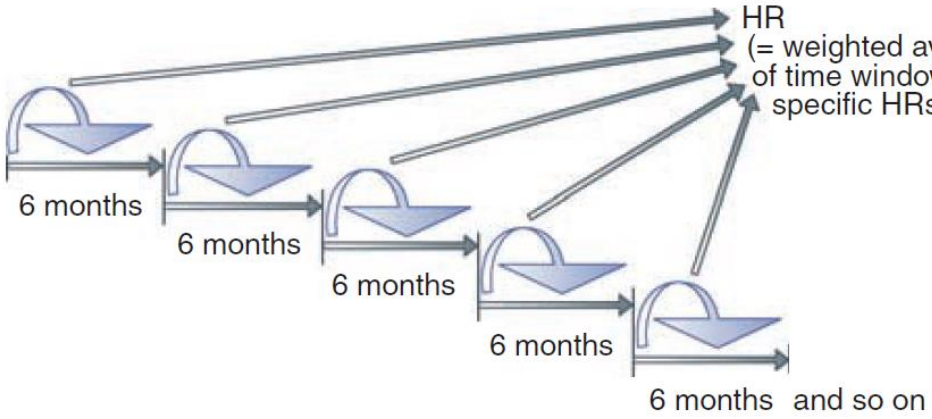
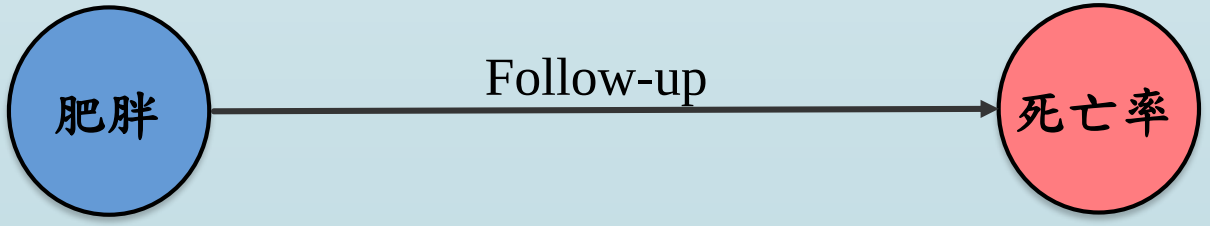


Figure 3 | Effect of time-varying risk on mortality.



Dekker, Friedo W., et al. "Survival analysis: time-dependent effects and time-varying risk factors." *Kidney international* 74.8 (2008): 994-997.

Cox model with time-dependent covariates

- External Time-Dependent Covariates (外在時間變數)
  - 這類變數不受個體影響，僅與時間有關
    - 年齡、藥品劑量
    - 空氣污染指數
    - 經濟狀況變化
- Internal Time-Dependent Covariates (內在時間變數)
  - 這類變數與個體特徵有關且隨個體狀態變化
    - 體重變化
    - 血壓變化
    - 服藥狀態

Cox model with time-dependent covariates

- 情境
  - 追蹤期間才可能會發生的變項
  - 不斷改變
- 注意事項
  - 干擾因子不可為中介變數
  - Time-dependent covariates 不可為中介變數

Time-dependent covariates-example

- 寬資料(wide format) 需轉換長資料(long format)

id	entry_time	transplant_time	death_time	status	age
1	0	50	70	1	45
2	0	30	90	1	60
3	0	NA	40	1	50
4	0	20	60	1	55
5	0	30	50	0	35
6	0	15	40	0	45
7	0	35	60	0	50
8	0	NA	40	1	55
9	0	25	50	1	65
10	0	NA	30	0	60

Time-dependent covariates-example

- 長資料(long format)

id	Start	Stop	transplant	status	age
1	0	50	0	0	45
1	50	70	1	1	45
2	0	30	0	0	60
2	30	90	1	1	60
3	0	40	0	1	50
4	0	20	0	0	55
4	20	60	1	1	55
5	0	30	0	0	35
5	30	50	1	0	35
6	0	15	0	0	45
6	15	40	1	0	45
7	0	35	0	0	50
7	35	60	1	0	50

Time-dependent covariates-R操作

安裝套件並載入

匯入原始數據

轉換資料格式

狀態設定

執行分析

	id	entry_time	transplant_time	death_time	status	age
1	1	0	50	70	1	45
2	2	0	30	90	1	60
3	3	0	NA	40	1	50
4	4	0	20	60	1	55
5	5	0	30	50	0	35
6	6	0	15	40	0	45
7	7	0	35	60	0	50
8	8	0	NA	40	1	55
9	9	0	25	50	1	65
10	10	0	NA	30	0	60

```
#安裝套件並載入
install.packages("dplyr")
install.packages("tidyr")
install.packages("survival")
library(dplyr)
library(tidyr)
Library(survival)

#原始數據(wide format)
data <- data.frame(
  id = c(1, 2, 3, 4, 5, 6, 7, 8, 9, 10),
  entry_time = c(0, 0, 0, 0, 0, 0, 0, 0, 0, 0),
  transplant_time = c(50, 30, NA, 20, 30, 15, 35, NA, 25, NA),
  death_time = c(70, 90, 40, 60, 50, 40, 60, 40, 50, 30),
  status = c(1, 1, 1, 1, 0, 0, 0, 1, 1, 0),
  age = c(45, 60, 50, 55, 35, 45, 50, 55, 65, 60)
)
```

安裝套件並載入

匯入原始數據

轉換資料格式

狀態設定

執行分析

	id	Start	Stop	transplant	age	status
1	1	0	50	0	45	0
2	1	50	70	1	45	1
3	2	0	30	0	60	0
4	2	30	90	1	60	1
5	3	0	40	0	50	1
6	4	0	20	0	55	0
7	4	20	60	1	55	1
8	5	0	30	0	35	0
9	5	30	50	1	35	0
10	6	0	15	0	45	0

```
# 轉換為長格式 (long format)
long_data <-data %>%
  mutate(has_transplant = !is.na(transplant_time)) %>%
  tidyr::pivot_longer(cols = c(transplant_time, death_time), names_to = "period", values_to = "Time") %>%
 filter(!is.na(Time)) %>%
 arrange(id, Time) %>%
 group_by(id) %>%
  mutate(Start = dplyr::lag(Time, default = first(entry_time)),  Stop = Time) %>%
  mutate(transplant = ifelse(period == "transplant_time"& has_transplant == "TRUE"| has_transplant == "FALSE" , 0, 1)) %>%
select(id, Start, Stop, transplant, age, status)

# 狀態設定只出現在最後一個時間段
long_data <- long_data %>%
  group_by(id) %>%
  mutate(status = ifelse(Stop == max(Stop), status, 0))
```

安裝套件並載入

匯入原始數據

轉換資料格式

狀態設定

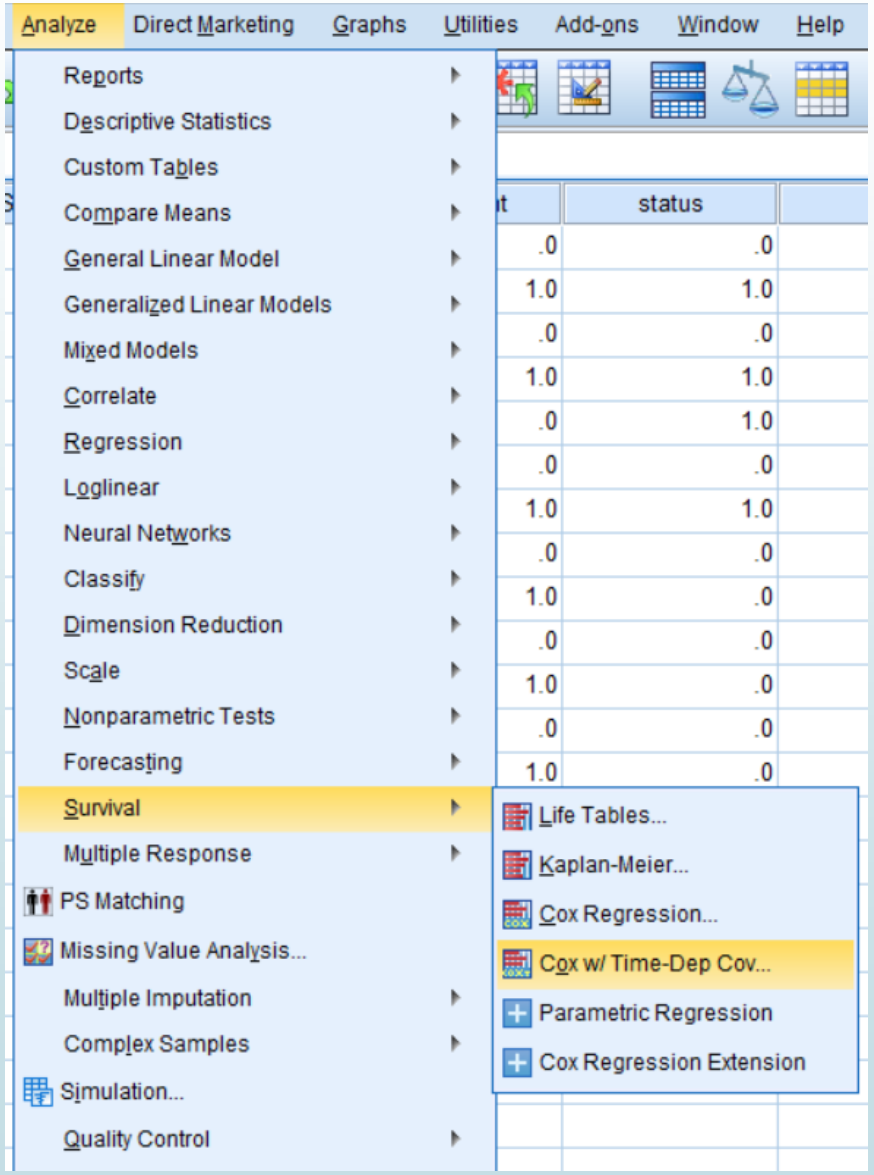
執行分析

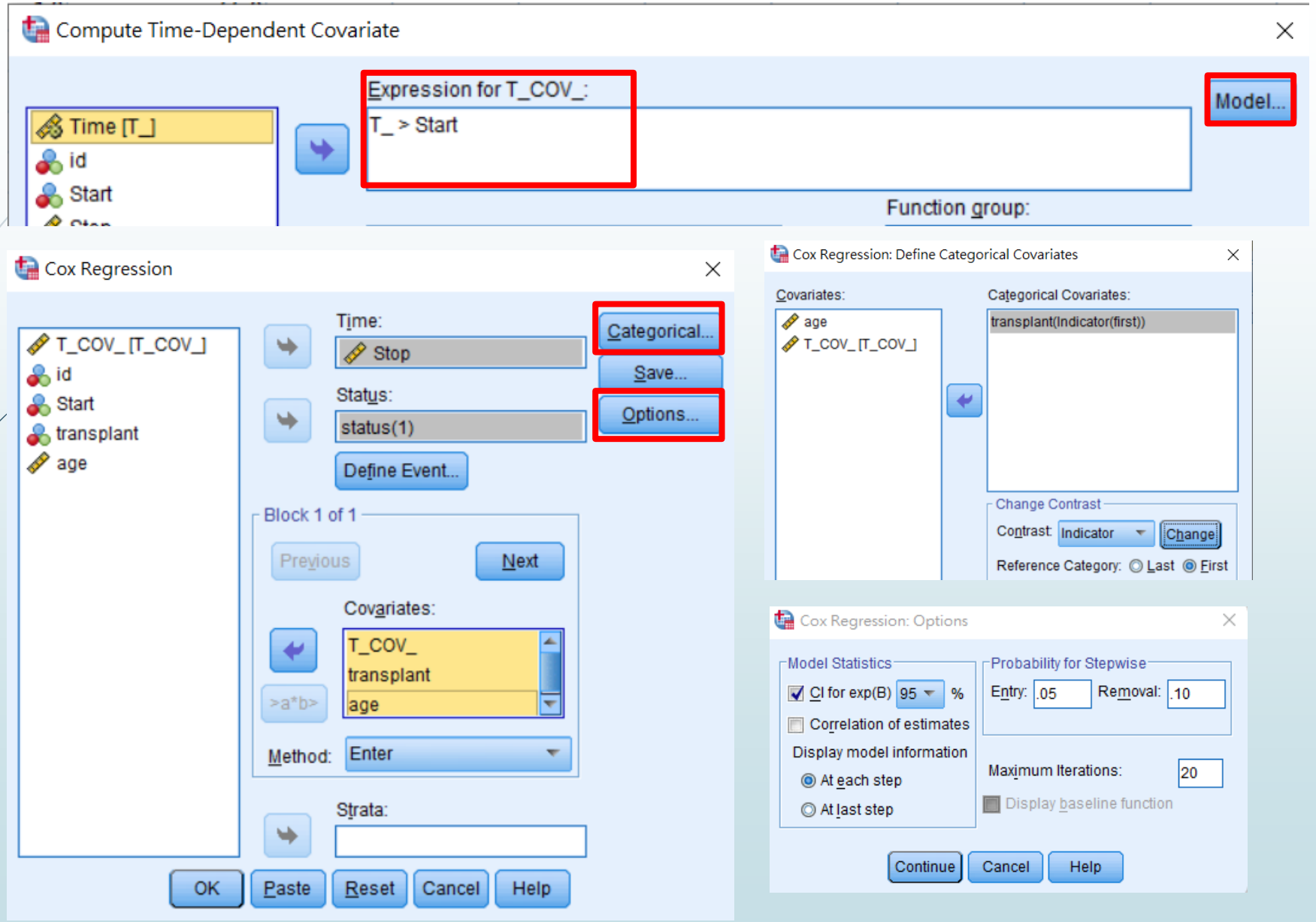
```
coxph(formula = Surv(Start, Stop, status) ~ transplant + age,
    data = long_data)

  n= 17, number of events= 6

               coef exp(coef) se(coef)      z Pr(>|z|)
transplant -2.54704   0.07831  1.54815 -1.645   0.0999 .
age         0.08891   1.09299  0.08323  1.068   0.2854
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

           exp(coef) exp(-coef) lower .95 upper .95
transplant   0.07831    12.7693  0.003767     1.628
age          1.09299     0.9149  0.928478     1.287

Concordance= 0.87  (se = 0.097 )
Likelihood ratio test= 3.73  on 2 df,   p=0.2
Wald test            = 2.74  on 2 df,   p=0.3
Score (logrank) test = 3.62  on 2 df,   p=0.2
```

```
cox_model <- coxph(Surv(Start, Stop,status)~transplant + age,
data= long_data)

summary(cox_model)
```

Time-dependent covariates-SPSS操作

分析>存活分析>Cox w/ Time-Dep Cov




Compute Time-Dependent Covariate
Time [T_]
id
Start
Expression for T_COV_:
T_ > Start
Model...
Function group:
Cox Regression
T_COV_ [T_COV_]
id
Start
transplant
age
Time:
Stop
Status:
status(1)
Define Event...
Categorical...
Save...
Options...
Block 1 of 1
Previous
Next
Covariates:
T_COV_
transplant
age
>a*b>
Method:
Enter
Strata:
OK
Paste
Reset
Cancel
Help
Cox Regression: Define Categorical Covariates
Covariates:
age
T_COV_ [T_COV_]
Categorical Covariates:
transplant(Indicator(first))
Change Contrast
Contrast:
Indicator
Change
Reference Category:
Last
First
Cox Regression: Options
Model Statistics
CI for exp(B)
95
%
Correlation of estimates
Display model information
At each step
At last step
Probability for Stepwise
Entry:
.05
Removal:
.10
Maximum Iterations:
20
Display baseline function
Continue
Cancel
Help

Time-dependent covariates-Output

Variables in the Equation

	B	SE	Wald	df	Sig.	Exp(B)	95.0% CI for Exp(B)	
							Lower	Upper
T_COV_	12.099	1537.617	.000	1	.994	179671.761	.000	.
transplant	-2.315	1.526	2.301	1	.129	.099	.005	1.966
age	.082	.082	1.000	1	.317	1.085	.925	1.273

調整年齡後，有腎移植比無腎移植的
死亡風險無統計差異(p= 0.129)

感謝您的聆聽！

Thank you !

5/13問卷