

PHÂN TÍCH SỐNG SỐT (SURVIVAL ANALYSIS)

Phân tích sống sót (PTSS) hoặc phân tích sự kiện khi nhà nghiên cứu muốn tìm hiểu ảnh hưởng đến các biến kết cục (biến phụ thuộc) mang tính thời gian.

Khi so sánh 2 phương pháp điều trị cho các bệnh có tần số tử vong cao như bệnh AIDS, các bệnh ung thư... Nếu mô hình phân tích như phân tích hồi qui logistic, chỉ để ý đến biến kết cục (sống/chết hoặc khỏi bệnh/không khỏi bệnh) mà không quan tâm đến yếu tố thời gian thì đôi khi không tìm thấy sự khác biệt giữa 2 phương pháp điều trị vì tỉ lệ tử vong gần như nhau, nhưng thời gian dẫn đến tử vong ở 2 nhóm có thể khác nhau. Một ví dụ khác khi so sánh 2 loại kháng sinh điều trị bệnh thương hàn, tỉ lệ khỏi bệnh của 2 loại kháng sinh có thể như nhau nhưng thời gian cắt sốt của 2 nhóm có thể khác nhau, vì vậy chúng ta phải sử dụng mô hình PTSS thì mới thấy sự khác biệt này. Như vậy mô hình nghiên cứu mô tả kết cục là biến nhị phân (sống/chết-hết sốt/còn sốt) tuy quan trọng nhưng không chính xác.

Một ưu điểm của PTSS là xử lý được các trường hợp đối tượng nghiên cứu bỏ cuộc giữa chừng (như mất dấu theo dõi, ngưng điều trị do tác dụng phụ của thuốc hoặc tử vong do bệnh lý khác...).

Trong mô hình phân tích này các đối tượng còn sống kể cả đối tượng bỏ cuộc được gọi là **censored** hoặc sự kiện chưa xảy ra. Các đối tượng tử vong hoặc hết sốt (ví dụ trong nghiên cứu bệnh thương hàn) được gọi là **events** hoặc sự kiện đã kết thúc.

Phép ước tính thường được dùng để PTSS được gọi là **ước tính Kaplan-Meier**. Phép ước tính này giúp ta tính được xác suất sống sót tích lũy tại các mốc thời gian khác nhau (sẽ minh họa trong các ví dụ sau). Nếu muốn so sánh sự khác biệt giữa 2 nhóm điều trị, dùng kiểm định log-rank bằng cách so sánh 2 hàm xác suất tích lũy của 2 nhóm.

Sau đây là 1 ví dụ minh họa:

So sánh 2 phác đồ điều trị (1 và 2) cho bệnh nhân AIDS, mỗi nhóm gồm 30 bệnh nhân được theo dõi trong 2 năm, nhà nghiên cứu muốn tìm hiểu sự khác biệt giữa 2 loại phác đồ điều trị thuốc kháng retrovirus (ARV) khác nhau. Các số liệu thu thập trong bảng 1.

Bảng 1. Số liệu 60 bệnh nhân AIDS sau 2 năm nghiên cứu

MABN	NHOM	CANNANG	CD4	THOIGIAN	KETCUC
BN1	1	45	30	3	1
BN2*	1	52	140	16	2
BN3	1	51	150	24	0
BN4	1	46	60	6	1
BN5	1	57	100	6	1
BN6	1	52	20	24	0
BN7	1	54	40	24	0
BN8	1	47	20	3	1
BN9	1	56	100	24	0
BN10	1	65	50	24	0
BN11	1	54	80	8	1
BN12	1	60	30	24	0
BN13	1	48	100	24	0
BN14	1	50	120	24	0
BN15	1	45	20	3	1
BN16	1	53	230	24	0
BN17	1	46	30	3	1
BN18	1	49	120	24	0
BN19	1	38	100	9	1
BN20	1	56	230	24	0
BN21	1	48	50	4	1
BN22	1	65	120	24	0
BN23	1	67	200	24	0
BN24	1	50	40	9	1
BN25	1	52	120	12	1
BN26	1	54	350	24	0
BN27	1	43	30	3	1
BN28	1	55	200	24	0
BN29	1	56	190	24	0
BN30	1	46	80	12	1
BN31	2	57	170	24	0
BN32**	2	55	180	17	2
BN33	2	44	100	15	1
BN34	2	52	40	8	1
BN35	2	58	160	24	0
BN36	2	49	250	24	0
BN37	2	56	110	14	1
BN38	2	65	230	24	0
BN39	2	67	180	24	0
BN40	2	54	40	14	1
BN41	2	55	50	10	1
BN42	2	54	60	24	0
BN43	2	50	300	24	0
BN44	2	51	60	22	1
BN45	2	46	120	24	0
BN46	2	47	200	24	0

BN47	2	48	230	24	0
BN48	2	50	180	24	0
BN49	2	53	150	24	0
BN50	2	57	100	24	0
BN51	2	56	110	24	0
BN52	2	58	120	24	0
BN53	2	60	160	24	0
BN54	2	46	120	24	0
BN55	2	48	200	24	0
BN56	2	52	230	24	0
BN57	2	55	130	24	0
BN58	2	46	140	18	1
BN59	2	56	150	24	0
BN60	2	53	160	24	0

Số liệu gồm 6 cột:

Cột 1: mã bệnh nhân [lưu ý: bệnh nhân số 2 (BN2*) bỏ cuộc sau 16 tháng điều trị, bệnh nhân 32 (BN32**) chết sau 17 tháng vì nhồi máu cơ tim, không liên quan đến AIDS.

Cột (nhóm điều trị) với 2 loại phác đồ khác nhau 1 và 2.

Cột 3: Cân nặng bệnh nhân (kg)

Cột 4: lượng tế bào CD4/mm³

Cột 5: thời gian điều trị (tính bằng tháng)

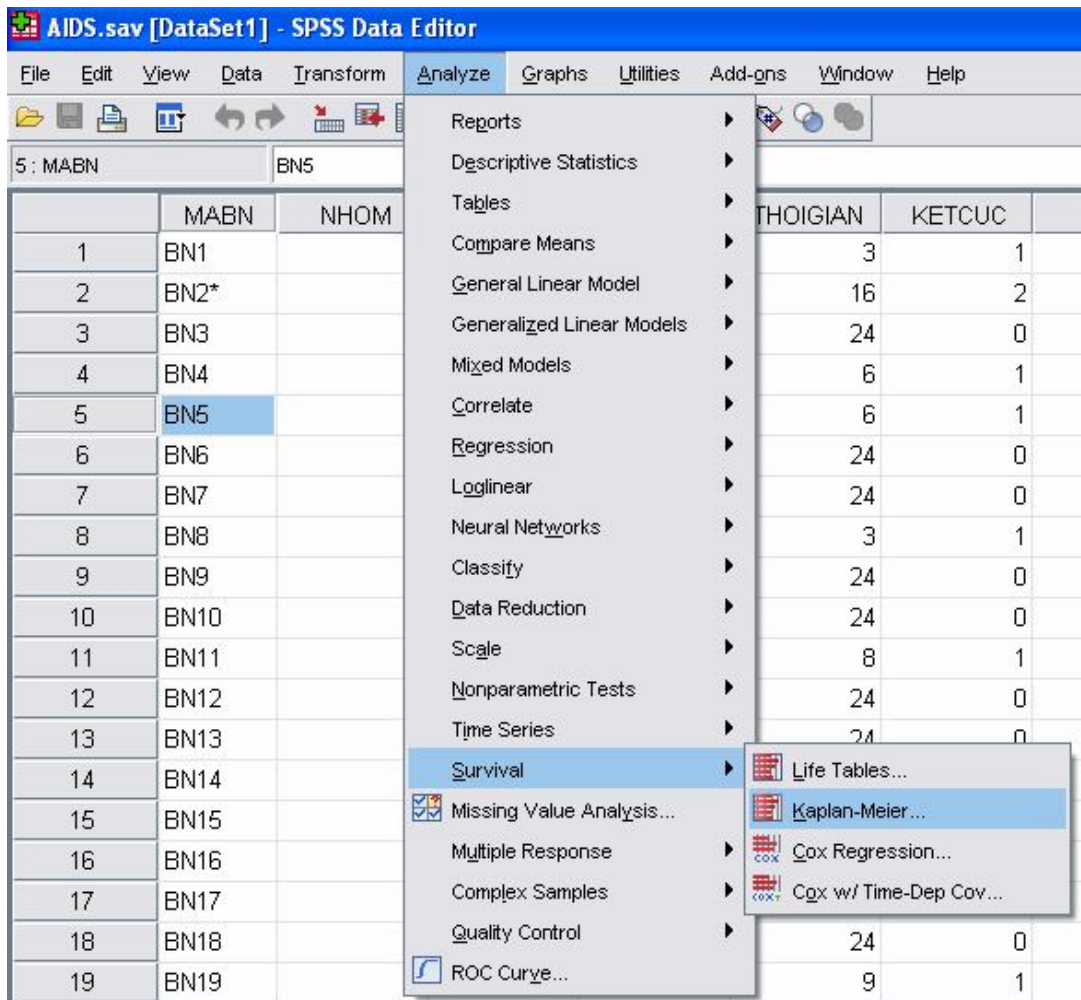
Cột 6: kết cục gồm 0: sống (censored) 1: tử vong (events); 2: bỏ cuộc (censored)

***Lưu ý: bỏ cuộc vẫn coi như censored**

1. Cách phân tích xác suất sống sót tích lũy Kaplan-Meier trong SPSS
(phiên bản 16.0).

Nhập số liệu vào SPSS

Vào menu: **Analyze>Survival> Kaplan-Meier...** như hình sau

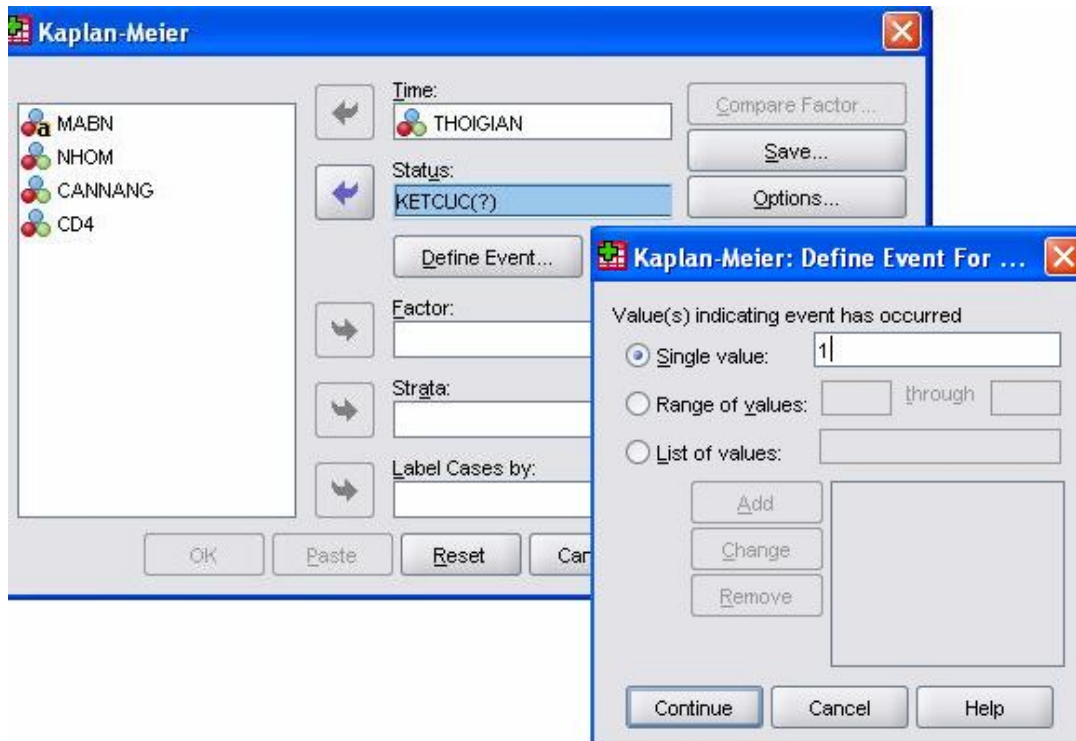


Mở màn hình Kaplan Meier, nhấp biến THOIGIAN vào ô Time

Biến KETCUC vào ô Status

Sau đó nhấn nút Define Event

Vào màn hình Define Event chọn Single value bằng 1 (là tử vong trong nghiên cứu này)



Nhấp Continue và nhấp OK sẽ có kết quả như sau:

Kết quả:

➔ **Kaplan-Meier**

Case Processing Summary

Total N	N of Events	Censored	
		N	Percent
60	20	40	66.7%

Total N=60: 2 nhóm có 60 BN

N of Events=20: số ca chết (sự kiện xảy ra)

Censored N=40 : gồm 38 ca còn sống và 2 ca bỏ cuộc

Survival Table

	Time	Status	Cumulative Proportion Surviving at the Time		N of Cumulative Events	N of Remaining Cases
			Estimate	Std. Error		
1	3.000	1	.	.	1	59
2	3.000	1	.	.	2	58
3	3.000	1	.	.	3	57
4	3.000	1	.	.	4	56
5	3.000	1	.917	.036	5	55
6	4.000	1	.900	.039	6	54
7	6.000	1	.	.	7	53
8	6.000	1	.867	.044	8	52
9	8.000	1	.	.	9	51
10	8.000	1	.833	.048	10	50
11	9.000	1	.	.	11	49
12	9.000	1	.800	.052	12	48
13	10.000	1	.783	.053	13	47
14	12.000	1	.	.	14	46
15	12.000	1	.750	.056	15	45
16	14.000	1	.	.	16	44
17	14.000	1	.717	.058	17	43
18	15.000	1	.700	.059	18	42
19	16.000	2	.	.	18	41
20	17.000	2	.	.	18	40
21	18.000	1	.683	.060	19	39
22	22.000	1	.665	.061	20	38
23	24.000	0	.	.	20	37
24	24.000	0	.	.	20	36
25	24.000	0	.	.	20	35
26	24.000	0	.	.	20	34
27	24.000	0	.	.	20	33
28	24.000	0	.	.	20	32
29	24.000	0	.	.	20	31
30	24.000	0	.	.	20	30
31	24.000	0	.	.	20	29

Bảng phân tích sống sót (Life table) gồm 6 cột:

Cột 1 gồm 60 ca (trong bảng mô tả 31 ca đầu tiên)

Cột 2 (time): mốc thời gian theo tháng

Cột 3 (Status) tình trạng BN 1: chết; 2: bỏ cuộc; 0: còn sống

Cột 4 bên trái (ước lượng xác suất sống sót tích lũy) được tính như sau:

Tại mốc 3 tháng (hàng 5 ở bảng trên) đã có 5 BN chết vậy xác suất chết là $5/60=0,083$ và xác suất sống sót tích lũy là: $1- 0,083=0,917$.

Tại mốc 4 tháng (hàng 6) thêm 1 BN chết : xác suất chết là $1/55= 0,018$ và xác suất còn sống: $1-0,018=0,982$ và xác suất tích lũy tại mốc 4 tháng là:

$0,982 \times 0,917 = 0,900$ (tính xác suất tích lũy bằng cách lấy xác suất còn sống đến 3 tháng nhân cho xác suất còn sống đến 4 tháng).

Tương tự tại mốc 6 tháng (hàng 8) thêm 2 BN chết: xác suất chết là $2/54 = 0,037$, xác suất sống: $1 - 0,037 = 0,963$ và xác suất tích lũy tại mốc 6 tháng: $0,963 \times 0,900 = 0,867$

.....

Cuối cùng tại thời điểm 22 tháng có tất cả 20 BN chết và xác suất tích lũy còn sống tại thời điểm này là 0,665 (66,5%)

Cột 4 bên phải là sai số chuẩn (standard error)

Cột 5: Số ca chết (sự kiện xảy ra) tích lũy

Cột 6: Số ca còn lại

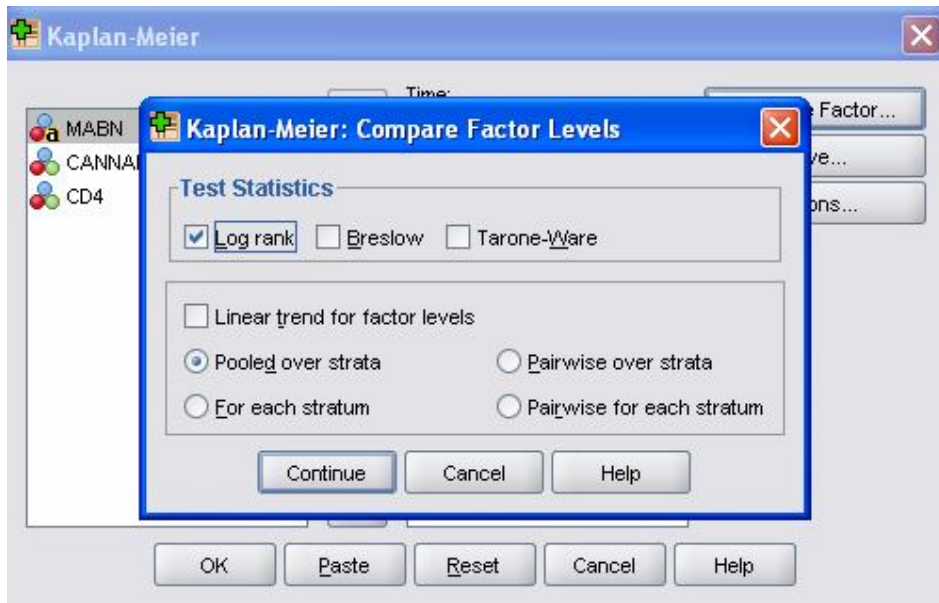
2. Kiểm định Log rank:

Bây giờ nếu muốn biết sự khác biệt giữa 2 phác đồ điều trị (1 và 2) dùng kiểm định **Log Rank**

Thực hiện các bước y như trên, tuy nhiên muốn so sánh 2 nhóm do vậy, dùng chuột nhấp biến NHOM (2 loại phác đồ) từ ô bên trái chuyển vào ô Factor như hình sau:



Sau đó nhấp Compare Factor bên góc trên trái màn hình Kaplan-Meier



Mở hộp thoại Compare Factor, sau đó đánh dấu nháy ✓ vào ô Log rank như hình trên.

Nhấp Continue, và cuối cùng nhấp OK sẽ cho kết quả sau:

Case Processing Summary

NHOM	Total N	N of Events	Censored	
			N	Percent
1	30	13	17	56.7%
2	30	7	23	76.7%
Overall	60	20	40	66.7%

Cột 1: Nhóm (2 phác đồ điều trị 1 và 2)

Cột 2: mỗi nhóm có 30 ca

Cột 3: sự kiện xảy ra (chết): nhóm 1 có 13 ca, nhóm 2 có 7 ca

Cột 4: Số ca còn sống hoặc bỏ cuộc (censored): nhóm 1 có 17 ca nhóm 2 có 23 ca

Bảng sống sót của nhóm 1 (có 30 ca, chỉ minh họa 20 ca đầu tiên):

Survival Table

NHOM	Time	Status	Cumulative Proportion Surviving at the Time		N of Cumulative Events	N of Remaining Cases	
			Estimate	Std. Error			
1	1	3.000	1	.	.	1	29
	2	3.000	1	.	.	2	28
	3	3.000	1	.	.	3	27
	4	3.000	1	.	.	4	26
	5	3.000	1	.833	.068	5	25
	6	4.000	1	.800	.073	6	24
	7	6.000	1	.	.	7	23
	8	6.000	1	.733	.081	8	22
	9	8.000	1	.700	.084	9	21
	10	9.000	1	.	.	10	20
	11	9.000	1	.633	.088	11	19
	12	12.000	1	.	.	12	18
	13	12.000	1	.567	.090	13	17
	14	16.000	2	.	.	13	16
	15	24.000	0	.	.	13	15
	16	24.000	0	.	.	13	14
	17	24.000	0	.	.	13	13
	18	24.000	0	.	.	13	12
	19	24.000	0	.	.	13	11
	20	24.000	0	.	.	13	10

Tại mốc 3 tháng (cột time hàng 5): có 5 tử vong ($5/30=0,166$) → tỉ lệ sống tích lũy: $1-0,166=0,833$.

Tại mốc 4 tháng (hàng 6): có 1 tử vong ($1/25=0,04$) → tỉ lệ sống tích lũy: $(1-0,04) \times 0,833=0,800$

.....

Tại mốc 12 tháng tổng cộng có 13 tử vong và tỉ lệ sống tích lũy là 0,567

Bảng sống sót của nhóm 2 (cũng chỉ minh họa 20 ca đầu tiên) :

2	1	8.000	1	.967	.033	1	29
	2	10.000	1	.933	.046	2	28
	3	14.000	1	.	.	3	27
	4	14.000	1	.867	.062	4	26
	5	15.000	1	.833	.068	5	25
	6	17.000	2	.	.	5	24
	7	18.000	1	.799	.074	6	23
	8	22.000	1	.764	.078	7	22
	9	24.000	0	.	.	7	21
	10	24.000	0	.	.	7	20
	11	24.000	0	.	.	7	19
	12	24.000	0	.	.	7	18
	13	24.000	0	.	.	7	17
	14	24.000	0	.	.	7	16
	15	24.000	0	.	.	7	15
	16	24.000	0	.	.	7	14
	17	24.000	0	.	.	7	13
	18	24.000	0	.	.	7	12
	19	24.000	0	.	.	7	11
	20	24.000	0	.	.	7	10
	21	24.000	0	.	.	7	9
	22	24.000	0	.	.	7	8
	23	24.000	0	.	.	7	7
	24	24.000	0	.	.	7	6
	25	24.000	0	.	.	7	5
	26	24.000	0	.	.	7	4
	27	24.000	0	.	.	7	3
	28	24.000	0	.	.	7	2
	29	24.000	0	.	.	7	1
	30	24.000	0	.	.	7	0

Tại mốc 1 tháng (cột time): có 1 tử vong ($1/30=0,033$) → tỉ lệ sống tích lũy: $1-0,033=0,967$.

Tại mốc 2 tháng: có 1 tử vong ($1/29=0,034$) → tỉ lệ sống tích lũy: $(1-0,034) \times 0,967=0,933$

.....

Tại mốc 8 tháng tổng cộng có 7 tử vong và tỉ lệ sống tích lũy là 0,764

Thời gian sống sót trung bình của 2 nhóm được mô tả trong bảng sau:

Means and Medians for Survival Time

NHOM	Mean ^a				Median			
	Estimate	Std. Error	95% Confidence Interval		Estimate	Std. Error	95% Confidence Interval	
			Lower Bound	Upper Bound			Lower Bound	Upper Bound
1	16.300	1.656	13.054	19.546
2	21.756	.833	20.122	23.389
Overall	19.027	.992	17.083	20.970

a. Estimation is limited to the largest survival time if it is censored.

Nhóm 1: thời gian sống sót trung bình ($16,3 \pm 1,6$ tháng)

Nhóm 2: thời gian sống sót trung bình ($21,7 \pm 0,8$ tháng)

Overall Comparisons

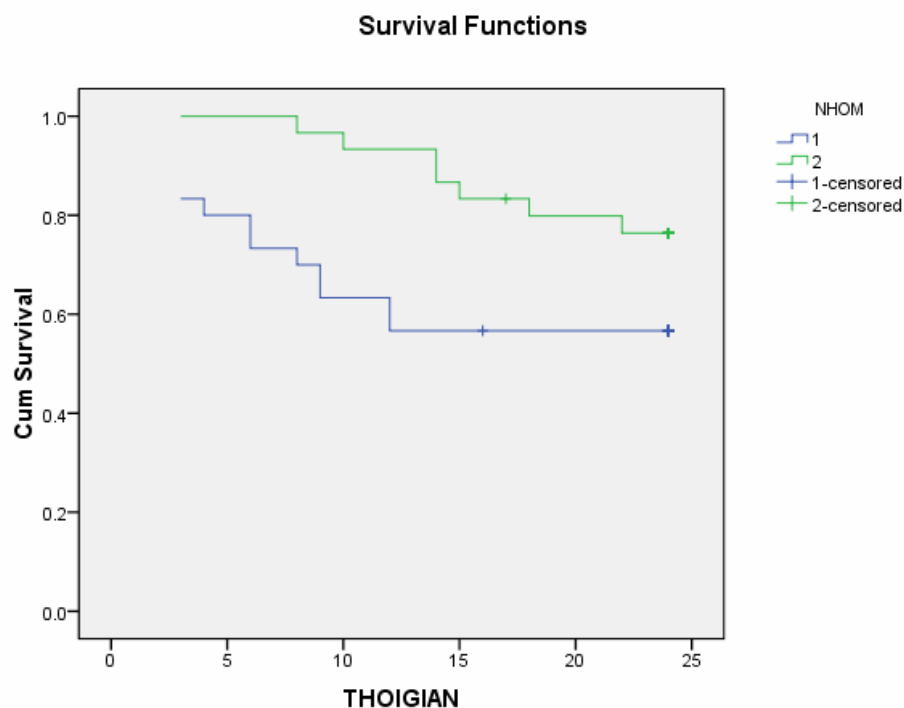
	Chi-Square	df	Sig.
Log Rank (Mantel-Cox)	3.906	1	.048

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

Kiểm định Log Rank: $\chi^2=3,906$, bậc tự do=1 và **p=0,048**

Như vậy sự khác biệt giữa 2 phác đồ điều trị có ý nghĩa thống kê, nhóm 2 có thời gian sống sót trung bình lớn hơn nhóm 1.

Và nếu vẽ biểu đồ (plot) hàm sống sót ta thấy có sự khác biệt giữa 2 nhóm (đường phía trên thuộc phác đồ 2 và đường phía dưới thuộc phác đồ 1)

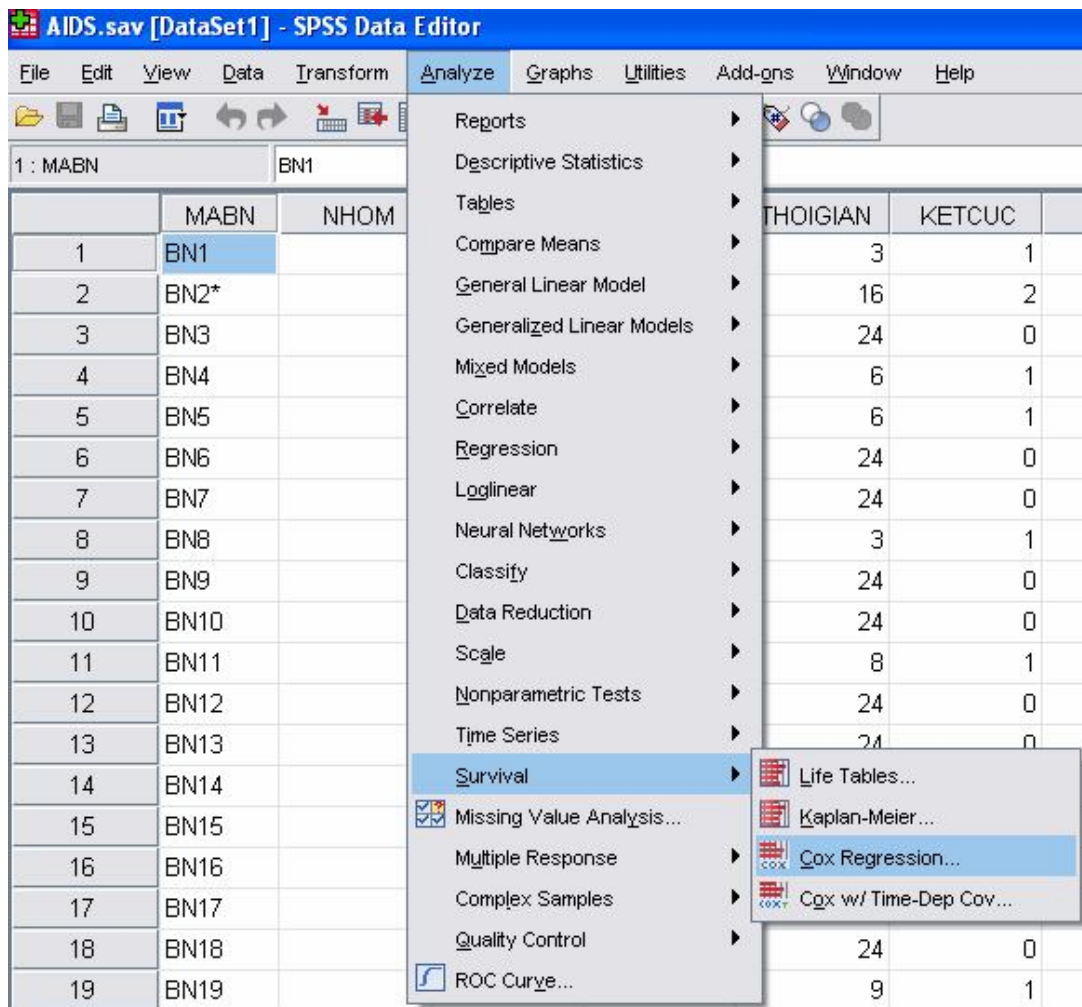


Nếu ta chỉ dừng tại đây thì ta có lẽ hài lòng vì kết luận được phác đồ điều trị 2 có hiệu quả hơn phác đồ điều trị 1 với $p < 0,05$. Tuy nhiên, theo các nghiên cứu trước đây cho thấy những bệnh nhân AIDS có cân nặng thấp hoặc trị số CD4 thấp lúc khởi đầu điều trị thì có nguy cơ tử vong cao hơn. Như vậy trước khi kết luận vấn đề cần phải hiệu chỉnh cân nặng và lượng CD4 hoặc nói cách khác cần phải phân tích đa biến.

3. Mô hình Cox (Cox's proportional hazards model)

Muốn phân tích đa biến hoặc hiệu chỉnh, dùng mô hình hồi qui Cox như màn hình sau:

Vào menu: **Analyze > Survival > Cox Regression**

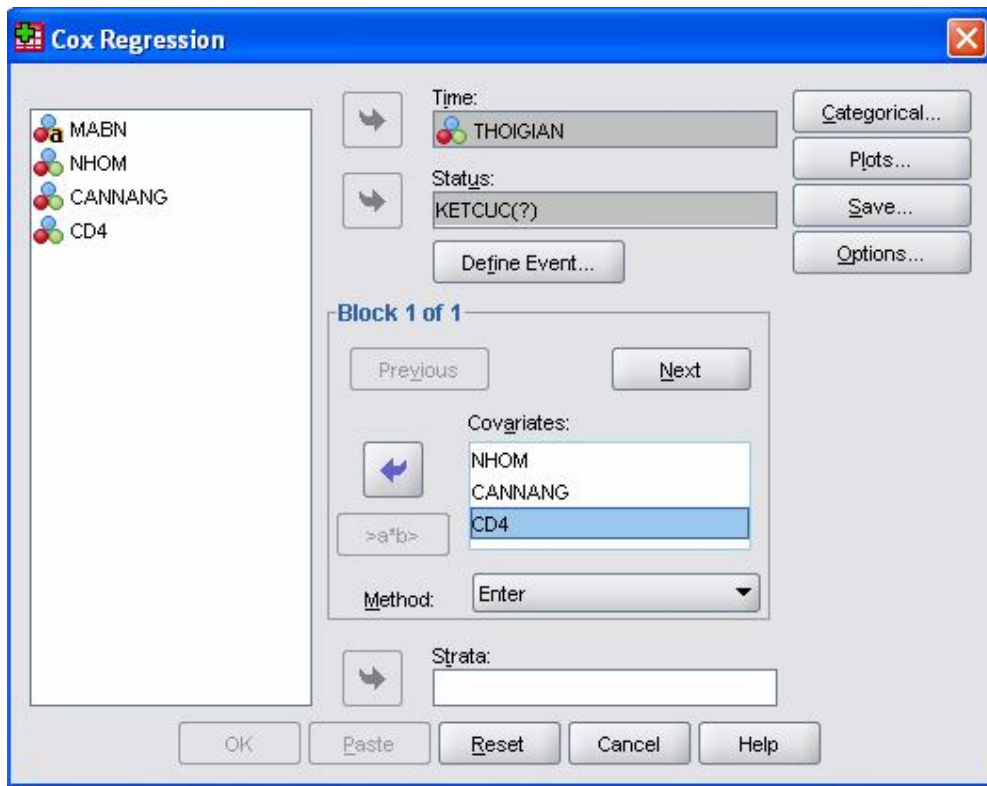


Chuyển các biến từ ô bên trái vào các ô bên phải như sau:

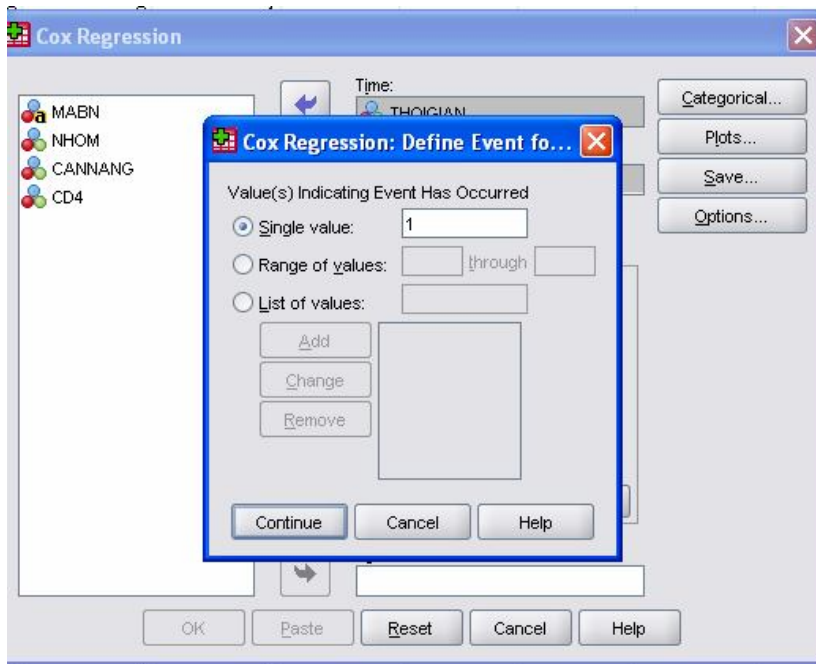
Biến thời gian vào ô Time

Biến kết cục vào ô Status

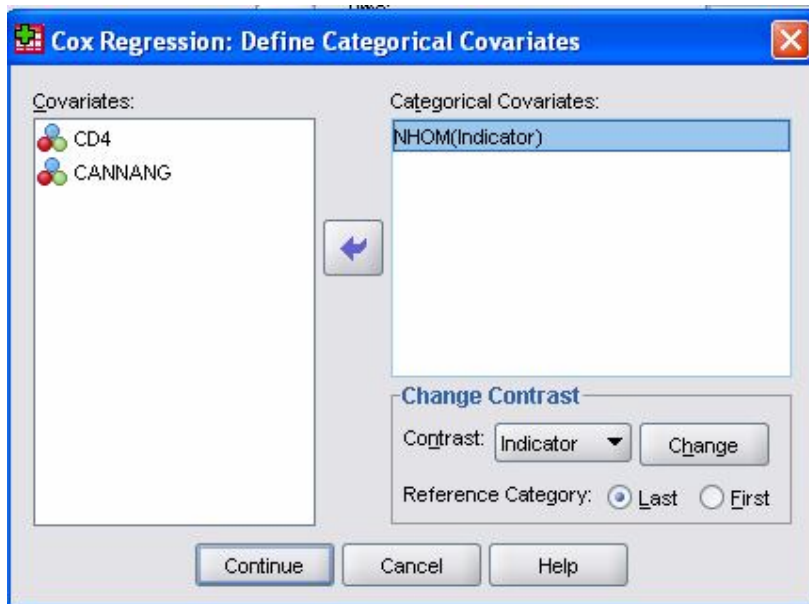
Cả 3 biến Nhóm, Cân nặng, CD4 vào ô Covariates



Nhấp vào nút Define Event để chọn kết cục bằng 1 (ta chỉ quan tâm đến tử vong)



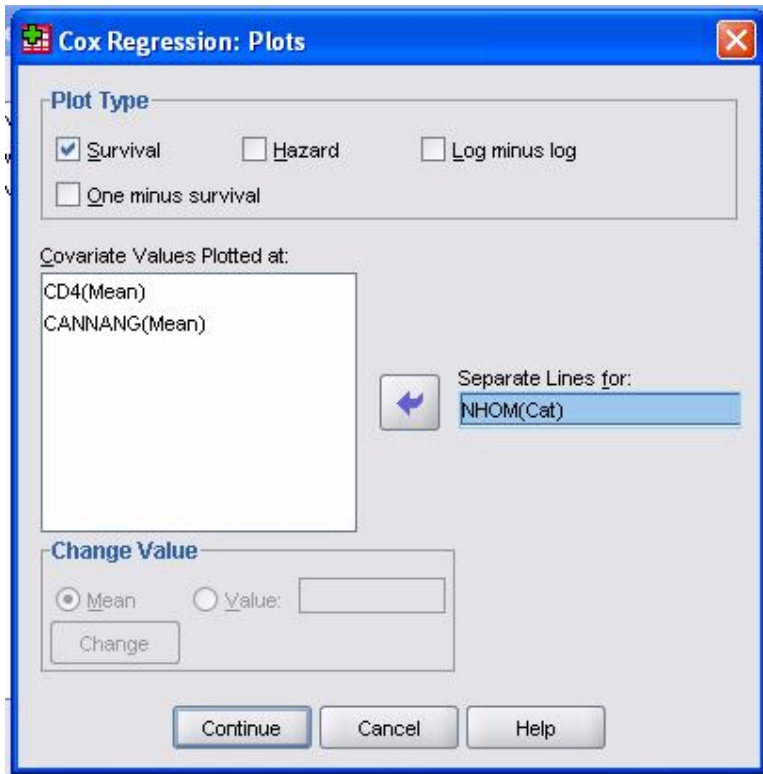
Nhấp nút Categorical ở góc trên phải màn hình, nhấp chọn biến phân loại, ta chỉ có 1 biến phân loại là nhóm (NHOM), chuyển Nhóm vào ô Categorical covariates như hình sau:



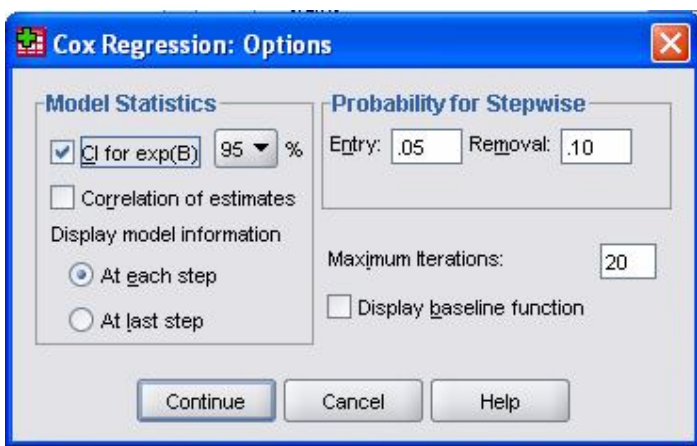
Nhấp continue. Vào tiếp hộp thoại Plots ngay dưới Categorical

Nhấp dấu nháy vào ô Survival trong Plot type

Chuyển NHOM vào ô Separate Lines for bên phải rồi nhấp Continue



Nhấp tiếp vào hộp thoại Options, mở màn hình sau, đánh dấu nhảy vào ô CI for exp(B) [có nghĩa exponential B với khoảng tin cậy 95%, dùng để tính Hazard risk tương tự như tính odds ratio trong phân tích hồi qui logistic], nhấp Continue



Cuối cùng nhấp OK sẽ xuất hiện kết quả sau (lưu ý ở đây chỉ giải thích các bảng chính)

➔ Cox Regression

[DataSet1] E:\BENHVIEN\BAIVIET\SURVIVAL ANALYSIS\AIDS.sav

Case Processing Summary

		N	Percent
Cases available in analysis	Event ^a	20	33.3%
	Censored	40	66.7%
	Total	60	100.0%
Cases dropped	Cases with missing values	0	.0%
	Cases with negative time	0	.0%
	Censored cases before the earliest event in a stratum	0	.0%
	Total	0	.0%
Total		60	100.0%

a. Dependent Variable: THOIGIAN

Như vậy số ca còn để phân tích là 60 trong đó có 20 ca là events (sự kiện đã xảy ra, ở đây là tử vong) và 40 ca censored (còn sống và bỏ cuộc).

Variables in the Equation

	B	SE	Wald	df	Sig.	Exp(B)	95.0% CI for Exp(B)	
							Lower	Upper
NHOM	-.103	.543	.036	1	.849	.902	.311	2.612
CANNANG	-.196	.053	13.867	1	.000	.822	.741	.911
CD4	-.028	.007	15.870	1	.000	.973	.959	.986

Các biến trong phương trình:

Cột 1 : có 3 biến NHOM, CAN NANG, CD4

Cột 2: hệ số β của biến cân nặng và biến CD4 đều âm, có nghĩa càng thấp càng tăng nguy cơ tử vong.

Cột 3: Sai số chuẩn SE của hệ số β

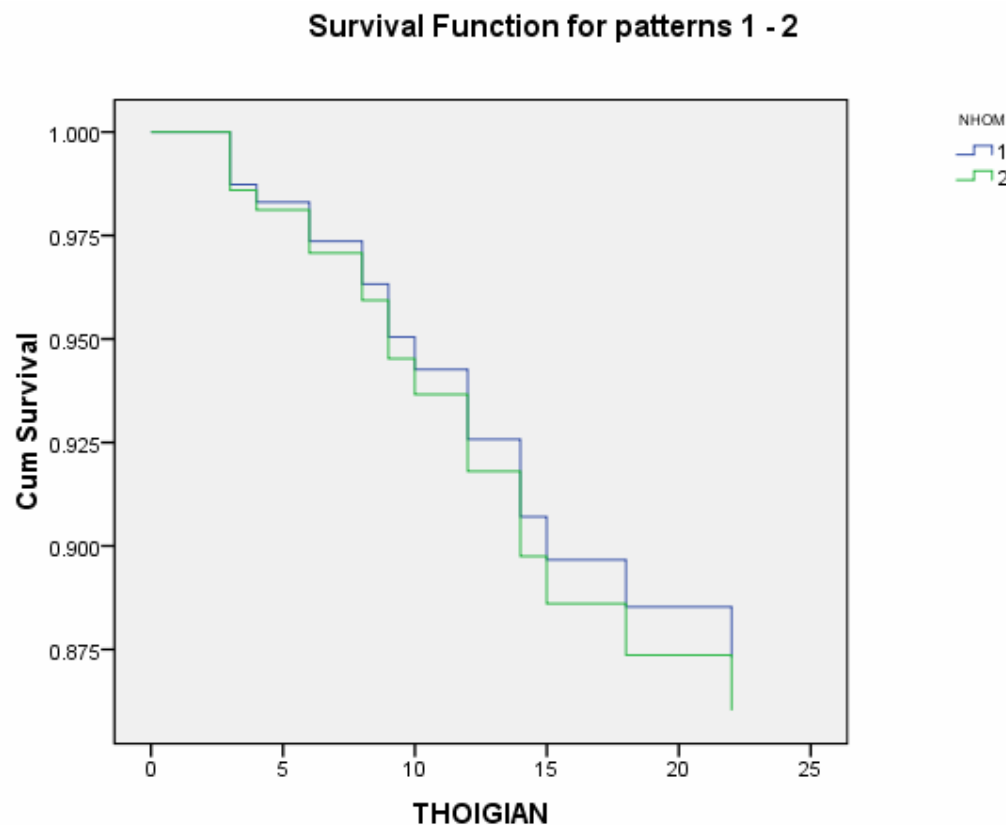
Cột 4 : Wald test tương tự chi bình phương (χ^2 test)

Cột 5: bậc tự do (df)

Cột 6: ý nghĩa thống kê hoặc trị số p. Kết quả không khác biệt giữa 2 nhóm điều trị ($p=0,849$), có khác biệt về cân nặng ($p=0,000$) và CD4 ($p=0,000$).

Cột 7: là Exponential (hàm mũ e) của hệ số β . ví dụ: $\text{Exp}(-0.103)=0.902$.

Cột này tương tự Odds ratio trong phân tích hồi qui logistic còn trong mô hình Cox được gọi là hazard risk (tỉ số nguy cơ)
Cột 8,9: Khoảng tin cậy 95% của tỉ số nguy cơ.



Và cuối cùng là biểu đồ hàm sống sót tích lũy giữa 2 nhóm (2 phác đồ điều trị) sau khi đã hiệu chỉnh cân nặng và lượng tế bào CD4 cho thấy 2 đường gần như nhau **không có sự khác biệt giữa 2 nhóm điều trị với $p=0,849$** . Chỉ có cân nặng thấp và lượng CD4 thấp làm tăng nguy cơ tử vong.

Như vậy không nên vội kết luận có sự khác biệt hiệu quả điều trị của 2 phác đồ trong phân tích đơn biến (kiểm định Log rank) cần tiến hành phân tích đa biến bằng mô hình Cox để hiệu chỉnh các yếu tố nhiễu hoặc các hiệp biến.

Tóm lại: Phân tích sống sót (Survival analysis) dựa vào sự khác biệt thời gian sống sót giữa 2 nhóm hoặc thời gian dẫn đến sự kiện xảy ra (chết, ung thư tái phát, hết sốt...). Phân tích này xử lý được khi đối tượng bỏ cuộc hoặc mất dấu

theo dõi. Dùng kiểm định Log rank trong phân tích đơn biến và mô hình Cox trong phân tích đa biến (hiệu chỉnh các yếu tố nhiễu hoặc các hiệp biến)

TS Nguyễn Ngọc Rạng, bsrang.blogspot.com

Tài liệu tham khảo:

1. Chan YH. Biostatistics 203. survival analysis, Singapore Med J, 2004 Vol 45(6):249.
2. Nguyễn Văn Tuấn. Phân tích sự kiện trong Phân tích số liệu và tạo biểu đồ bằng R, Nhà xuất bản Khoa học và Kỹ thuật, 2007, trang 238-259.