

§ 流行病學原理：資料分析 practice 5

流行病學資料分析 –
SAS
Practice 5
學號：b07401048
系級：醫學五

一. Risk ratio 的計算與統計檢定

(一)計算 Tx48 = 1 與 Tx48 = 0 兩組 30 天內死亡的 risk

1. Result :

(1) Tx48 = 1 組 30 天內死亡的 risk 為 53.49%。

(2) Tx48 = 0 組 30 天內死亡的 risk 為 28.07%。

2. Figure :

The FREQ Procedure

Frequency Row Pct	Table of Tx48 by Dead30			
	Tx48(Tx48)	Dead30(Dead30)		Total
		1	0	
1	1	23 53.49	20 46.51	43
0	0	32 28.07	82 71.93	114
Total		55	102	157

3. Description :

- (1) Tx48 = 1 組之病人為在尚未確定為 MRSA 感染時，即搶先於 48 小時內使用 glycopeptide 類的抗生素。該組病人有 43 人，在觀察時間（day2 到 day30）發生死亡者有 23 人，死亡的 risk 為 53.49%。
- (2) Tx48 = 0 組之病人並未搶先於 48 小時內使用 glycopeptide 類的抗生素。該組病人有 114 人，在觀察時間（day2 到 day30）發生死亡者有 32 人，死亡的 risk 為 28.07%。
- (3) 由於「是否搶先於 48 小時內使用 glycopeptide 類使用抗生素藥物」為不可逆事件，故適用 fixed cohort 的方法分析；又資料中的病人為住院病人，loss to follow-up 的人數為零，故是為 closed cohort，可以 risk 分析兩組死亡率。

(二)兩組死亡的 risk ratio 為何？是否達統計上的顯著？

1. Result :

- (1) Tx48 = 1 組之 risk 為 53.49%，Tx48 = 0 組之 risk 為 28.07%，兩者相除為 risk ratio = 1.9055。
- (2) 以雙尾 Z-test 檢定 1.9055 是否等於 1，檢定結果 p-value 為 0.0018，拒絕虛無假說，risk ratio 不等於 1，達統計上顯著。

2. Figure :

Relative Risk Test	
H0: P1 / P2 = 1 Wald Method	
Relative Risk	1.9055
Z	3.1202
One-sided Pr > Z	0.0009
Two-sided Pr > Z	0.0018
Column 1 (Dead30 = 1)	

3. Description :

- (1) 以雙尾 z-test 檢定 risk ratio 是否等於 1，虛無假設為 risk ratio 等於 1。檢定結果 p-value 為 0.0018，拒絕虛無假說，risk ratio 不等於 1，代表兩組之 risk 差異達統計上的顯著。
- (2) 此統計結果表示 Tx48 = 1 組在 30 天內死亡的 risk 顯著高於 Tx48 = 0 組，顯示在 48 小時內搶先使用 glycopeptide 類抗生素反而使 MRSA 病人 30 天內死亡的機率上升。
- (3) 根據結果，或許可推論「搶先於 48 小時內使用 glycopeptide 抗生素」為 MRSA 病人於 30 天內死亡的危險因子，原因可能是藥物副作用等因素。
- (4) 然而，此推論尚須小心求證。由於「是否於 48 小時內使用 glycopeptide 類抗生素」在現行醫療中並無一定的準則，醫師需根據病人整體狀況及經驗判斷是否於 48 小時內使用 glycopeptide 類藥物。雖然此 dataset 收案的皆為 MRSA 病人，但儘管同為 MRSA 病人，在前 48 小時病情嚴重程度也可能不同。或許本 dataset 收案的搶先於 48 小時使用 glycopeptide 類抗生素的病人（即 Tx48 = 1 組）皆是在 48 小時內病況非常嚴重的病人，當時醫師判斷需儘早使用 glycopeptide 類藥物進行搶救；而未於 48 小時內搶先使用 glycopeptide 類藥物的病人（即 Tx48 = 0 組）可能在 48 小時內的病況尚可，當時醫師判斷維生醫療可支撐病人至血液培養結果公布，不需於 48 小時內搶先投藥。若是如此，代表 Tx48 = 1 組「病況的 baseline」差於 Tx48 = 0 組，而此差異也將影響 30 天內死亡的 risk。以上僅為一個假設，但若要推論「搶先於 48 小時內使用 glycopeptide 抗生素」是 MRSA 病人於 30 天

內死亡的危險因子，尚須小心求證，確定兩組病人是否有資料未記錄，但確實存在差異的變項，以免得出錯誤的結論。

(三)Code for Q1：

```
/* hw5 cohort study */
dm "odsresult" clear;
dm "log" clear;
proc import datafile = "C:\Users\Raymond\Desktop\sas data\Dataset 3.xlsx"
    dbms = xlsx
    out = work.dataset3
    replace;
run;
/* Q1 risk ratio of two groups*/
title "Q1";
/* create legal dataset */
data q1;
    set dataset3;
    if day30 <= 1 then delete;
run;

/* sort data */
proc sort data = q1;
    by descending tx48 descending dead30; /* descendingly sort, first sort tx48, then sort dead30 */
run;

/* calculate relrisk */
proc freq data = q1 order= data; /* to sort in order of data */
    tables tx48*dead30/ nopercnt nocol chisq relrisk (equal cl = wald);
run;
```

二. Rate ratio :

(一) 假設 induction period = 1，兩組死亡之 rate ratio 為何，是否達統計上的顯著？

1. Result :

- (1) 兩組的 rate ratio 為 2.24155。
- (2) Rate ratio 的 95%信賴區間為 1.28209 – 3.91902，不包含 1，代表達統計上的差異，兩組的 rate 不同。

2. Figures :

Contrast Estimate Results										
Label	Mean Estimate	Mean		L'Beta Estimate	Standard Error	Alpha	L'Beta		Chi-Square	Pr > ChiSq
		Confidence Limits					Confidence Limits			
incident rate for tx48= 1	0.0260	0.0168	0.0403	-3.6494	0.2236	0.05	-4.0876	-3.2111	266.36	<.0001
incident rate for tx48 = 0	0.0116	0.0082	0.0164	-4.4565	0.1768	0.05	-4.8030	-4.1100	635.54	<.0001

q2-1

Obs	rr	rr_lower	rr_upper
1	2.24155	1.28209	3.91902

3. Description :

- (1) 假設 induction period = 1，表示在 glycopeptide 投藥的一日內發生的死亡事件不可計入，即 Tx48 = 1 組於 day3 以後發生的事件才可以計入；Tx48 = 0 則無 induction period 的問題。
- (2) Tx48 = 1 組的 rate 是 0.0260，Tx48 = 0 的 rate 是 0.0116，兩組的 rate ratio 為 2.24155，95%CI 為 1.28209-3.91902，不包含 1，代表達統計上顯著，Tx48 = 1 組發生死亡的 incident rate 顯著大於 Tx48 = 0 組。

4. Code :

```
/* Q2 induction time , rate ratio */

/* induction time = 1 */
title "q2-1" ;
ods trace on;
ods output ParameterEstimates = para_induct1;
data induct1;
    set q1;
    if tx48 = 1 and day30 = 2 then delete;

    if tx48 = 1 then pt = day30 - 2;
    else if tx48 = 0 then pt = day30 - 1;

    logpt = log(pt);

    keep id tx48 day30 dead30 pt logpt;
run;

proc genmod data = induct1;
    class dead30(ref = "0") tx48 (ref = "0");
    model dead30 = tx48 / dist = poisson link = log offset = logpt;

    estimate "incident rate for tx48= 1" intercept 1 tx48 1 0;
    estimate "incident rate for tx48 = 0" intercept 1 tx48 0 1;

run;
ods trace off;

data rr_induct1;
    set para_induct1;

    if parameter = "Tx48" and level1 = "1";

    rr = exp(estimate);
    rr_lower = exp(lowerwaldcl);
    rr_upper = exp(upperwaldcl);

    keep rr rr_lower rr_upper;
run;

proc print
    data = rr_induct1;
run;
```

(二) 假設 induction period = 2，兩組死亡之 rate ratio 為何，是否達統計上的顯著？

1. Result：

(1) 兩組的 rate ratio 為 2.24631。

(2) Rate ratio 的 95%信賴區間為 1.27333 – 3.96276，不包含 1，代表達統計上的差異，兩組的 rate 不同。

2. Figures：

Contrast Estimate Results										
Label	Mean Estimate	Mean		L'Beta Estimate	Standard Error	Alpha	L'Beta		Chi-Square	Pr > ChiSq
		Confidence Limits					Confidence Limits			
incident rate for tx48= 1	0.0261	0.0166	0.0409	-3.6472	0.2294	0.05	-4.0969	-3.1976	252.74	<.0001
incident rate for tx48 = 0	0.0116	0.0082	0.0164	-4.4565	0.1768	0.05	-4.8030	-4.1100	635.54	<.0001

q2-2

Obs	rr	rr_lower	rr_upper
1	2.24631	1.27333	3.96276

3. Description：

(1) 假設 induction period = 2，表示在 glycopeptide 投藥的二日內發生的死亡事件不可計入，即 Tx48 = 1 組於 day4 以後發生的事件才可以計入；Tx48 = 0 則無 induction period 的問題。

(2) Tx48 = 1 組的 rate 是 0.0261，Tx48 = 0 的 rate 是 0.0116，兩組的 rate ratio 為 2.24631，95%CI 為 1.27333-3.96276，不包含 1，代表達統計上顯著，Tx48 = 1 組發生死亡的 incident rate 顯著大於 Tx48 = 0 組。

4. Code :

```
/* induction time = 2 */
title "q2-2" ;
ods trace on;
ods output ParameterEstimates = para_induct2;
data induct2;
    set q1;
    if tx48 = 1 and day30 <= 3 then delete;

    if tx48 = 1 then pt = day30 - 3;
    else if tx48 = 0 then pt = day30 - 1;

    logpt = log(pt);

    keep id tx48 day30 dead30 pt logpt;
run;

proc genmod data = induct2;
    class dead30(ref = "0") tx48 (ref = "0");
    model dead30 = tx48 / dist = poisson link = log offset = logpt;

    estimate "incident rate for tx48= 1" intercept 1 tx48 1 0;
    estimate "incident rate for tx48 = 0" intercept 1 tx48 0 1;

run;
ods trace off;

data rr_induct2;
    set para_induct2;

    if parameter = "Tx48" and level1 = "1";

    rr = exp(estimate);
    rr_lower = exp(lowerwaldcl);
    rr_upper = exp(upperwaldcl);

    keep rr rr_lower rr_upper;
run;

proc print
    data = rr_induct2;
run;
```

(三) 假設 induction period = 3，兩組死亡之 rate ratio 為何，是否達統計上的顯著？

1. Result：

(1) 兩組的 rate ratio 為 2.12346。

(2) Rate ratio 的 95%信賴區間為 1.17918 – 3.82392，不包含 1，代表達統計上的差異，兩組的 rate 不同。

2. Figures：

Contrast Estimate Results										
Label	Mean Estimate	Mean		L'Beta Estimate	Standard Error	Alpha	L'Beta		Chi-Square	Pr > ChiSq
		Confidence Limits					Confidence Limits			
incident rate for tx48= 1	0.0246	0.0153	0.0396	-3.7035	0.2425	0.05	-4.1788	-3.2281	233.17	<.0001
incident rate for tx48 = 0	0.0116	0.0082	0.0164	-4.4565	0.1768	0.05	-4.8030	-4.1100	635.54	<.0001

q2-3

Obs	rr	rr_lower	rr_upper
1	2.12346	1.17918	3.82392

3. Descriptions：

(1) 假設 induction period = 3，表示在 glycopeptide 投藥後的三日內發生的死亡事件不可計入，即 Tx48 = 1 組於 day5 以後發生的事件才可以計入；Tx48 = 0 則無 induction period 的問題。

(2) Tx48 = 1 組的 rate 是 0.0246，Tx48 = 0 的 rate 是 0.0116，兩組的 rate ratio 為 2.12346，95%CI 為 1.17918-3.82392，不包含 1，代表達統計上顯著，Tx48 = 1 組發生死亡的 incident rate 顯著大於 Tx48 = 0 組。

4. Code :

```
/* induction time = 3 */
title "q2-3" ;
ods trace on;
ods output ParameterEstimates = para_induct3;
data induct3;
    set q1;
    if tx48 = 1 and day30 <= 4 then delete;

    if tx48 = 1 then pt = day30 - 4;
    else if tx48 = 0 then pt = day30 - 1;

    logpt = log(pt);

    keep id tx48 day30 dead30 pt logpt;
run;

proc genmod data = induct3;
    class dead30(ref = "0") tx48 (ref = "0");
    model dead30 = tx48 / dist = poisson link = log offset = logpt;

    estimate "incident rate for tx48= 1" intercept 1 tx48 1 0;
    estimate "incident rate for tx48 = 0" intercept 1 tx48 0 1;

run;
ods trace off;

data rr_induct3;
    set para_induct3;

    if parameter = "Tx48" and level1 = "1";

    rr = exp(estimate);
    rr_lower = exp(lowerwaldcl);
    rr_upper = exp(upperwaldcl);

    keep rr rr_lower rr_upper;
run;

proc print
    data = rr_induct3;
run;
```

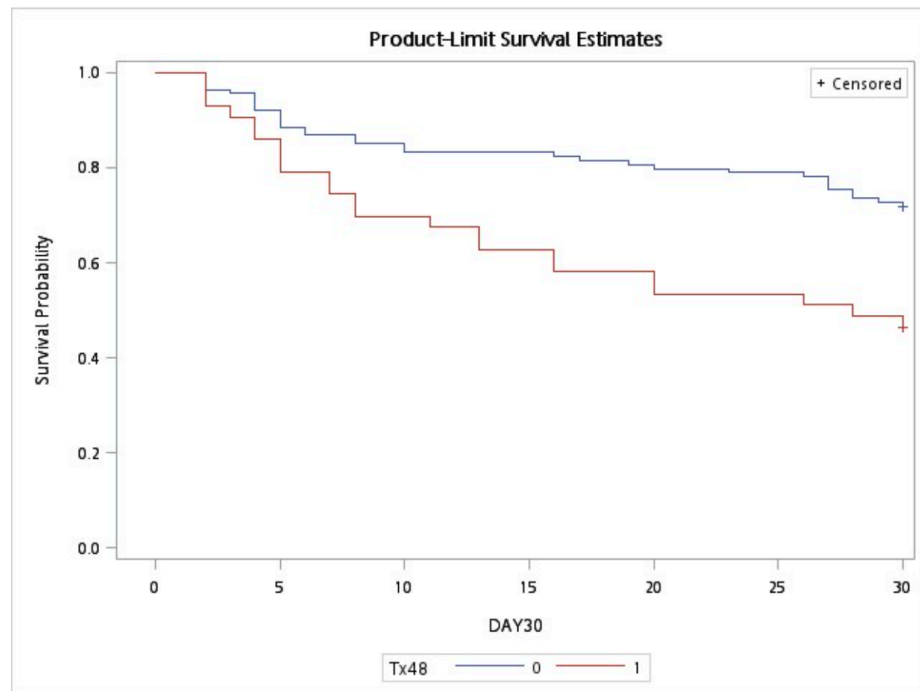
(四)不同的 induction period 假設是否會導致不同的結論？

1. 前三題假設的 induction period 分別為 1, 2 及 3 日，三者計算出的 rate ratio 分別為 2.24155, 2.24631 及 2.12346，三者僅有微幅差距，亦不影響結論，不論 induction period 為 1, 2 或 3，Tx48 = 1 組發生死亡的 incident rate 皆顯著大於 Tx48 = 0 組。以本題而言，induction period 的假設不會導致不同的結論。
2. 然而，induction period 在特定狀況確實會影響結論。比如 induction period 很長（數個月）或是收案數不多的情況下，確實可能影響結論。
3. 我認為 Induction period 的假設需考量醫學及生物方面的知識，訂立一個合理的 induction period，而非為了達到統計上的顯著性，而訂立一個背離醫學知識的 induction period。

三. Survival curve :

(一)作兩組 30 天內的 survival curve 。

1. Figures :



(二)是否達統計上的顯著？

1. Result :

(1) 以 log-rank test 檢定，p-value 為 0.0017，達統計上的顯著。

2. Figures :

Test of Equality over Strata			
Test	Chi-Square	DF	Pr > Chi-Square
Log-Rank	9.8983	1	0.0017
Wilcoxon	9.5703	1	0.0020
-2Log(LR)	9.6351	1	0.0019

(三)Code for Q3 :

```
/* q3 survival curve */  
title "q3";  
proc lifetest data = q1 plots = (s) method = km;  
  strata tx48;  
  time day30*dead30(0);  
run;
```