

競爭風險與 R 語法應用

張玉媚 東海大學統計系 副教授

存活資料中，死亡為最常見的感興趣事件(event)。但在醫學研究的實務應用，病人在追蹤過程中可能遭遇二種或兩種以上不同的事件，而感興趣的事件為其中一種。在此情形下，先出現某種事件便會影響其他事件發生的可能性，此即競爭風險。例如，白血病患者骨髓移植後，可能發生的事件有二種：白血病復發(relapse)以及移植失敗造成的死亡(death)。如果我們要探討的是白血病復發率，那麼移植失敗造成的死亡(death)即為復發(relapse)的競爭風險事件(competing risk event)。當病人因接受移植失敗而死亡，我們便無法觀察到病人復發的時間。又例如，65 歲以上的口腔癌病患，可能因為口腔癌而死亡，亦可能因跌倒意外、中風而死亡。此時口腔癌、跌倒意外及中風造成的死亡即為競爭風險事件，這些事件發生的機會就是競爭風險。

針對白血病患者骨髓移植研究，在進行統計分析時，當白血病復發為感興趣事件時，移植失敗造成的死亡該如何處理？以下利用 15 筆白血病患者骨髓移植研究的資料，介紹兩種方法估計復發的發生率。

資料名稱為 **ALL**，資料列於下表，第一列為變數名稱，其中 T1 是死亡時間，T2 是死亡或復發時間，d1 為死亡指標(1:死亡;0:設限)，d2 為復發指標(1:復發;0:設限)，d3 為死亡或復發指標(1:復發；0:設限)，rx 為治療方式(1: 處理組；0:對照組)。

T1	T2	d1	d2	d3	rx
269	110	1	1	1	0
350	332	1	0	1	0
530	530	0	0	0	1
996	996	0	0	0	1
226	226	0	0	0	1
418	418	1	0	1	0
276	276	1	0	1	1
156	104	1	1	1	1

781	609	1	1	1	0
371	230	1	1	1	0
526	526	1	0	1	1
110	74	1	1	1	0
243	122	1	1	1	1
466	466	1	0	1	1
262	192	1	1	1	0

方法 1. Kaplan-Meier (KM)法

如果只關心白血病復發這單一事件，則其他情況都視為設限(censor)。這時使用 Kaplan-Meier (Kaplan and Meier, 1958)法估計存活函數(survival function)來呈現不同時間點所對應的存活率。此法是假設病人不會因為非復發而死亡，反過來如果關心的是死亡風險，也是假設復發事件已排除，也就是假設事件時間跟設限時間獨立。我們可以利用 R 軟體中 library(survival)的 survfit 語法來得到 KM 估計。

R 語法(KM 估計)

```
library(survival)
data.all=read.table("d:/ALL.txt",header=TRUE)
attach(data.all)
km2=survfit(Surv(T2,d2) ~1,data=data.all)    # KM 估計
cbind(km2$time,km2$n.event,km2$n.risk,km2$surv,1-km2$surv) #列出想要的結果
```

我們將結果整理如下表，表中符號定義如下：

t_{re} = 觀察的復發時間(從病人接受移植後到復發事件發生所經歷的時間)

d_{re} = 復發人數

y_{re} = 涉險集合(at-risk set)，即復發時間大於或等於時間 t 的集合

KM_{re} = 免於復發的機率 (復發的存活率)

$1 - KM_{re}$ = 復發的累積發生率

使用 Kaplan-Meier (KM)方法估計復發的累積發生率

ID	t_{re}	d_{re}	y_{re}	KM_{re}	$1 - KM_{re}$
1	74	1	15	0.933	0.067
2	104	1	14	0.867	0.133
3	110	1	13	0.800	0.200
4	122	1	12	0.733	0.267
5	192	1	11	0.667	0.333
6	226	0	10	0.667	0.333
7	230	1	9	0.593	0.407
8	276	0	8	0.593	0.407
9	332	0	7	0.593	0.407
10	418	0	6	0.593	0.407
11	466	0	5	0.593	0.407
12	526	0	4	0.593	0.407
13	530	0	3	0.593	0.407
14	609	1	2	0.296	0.704
15	996	0	1	0.296	0.704

此例中共有 15 個樣本，實驗一開始有 15 人被追蹤觀察，故涉險集合中包含 15 人。觀察 74 天時，觀察對象 1 發生復發(d_{re} 為 1 表示 1 人復發)，所以未發生復發的比率為 $(15-1)/15=0.933$ ，未發生復發的累積比率亦為 0.933，發生復發的累積發生率則是 $1-0.933=0.067$ ；觀察時間介於 74~104 天時，除了觀測對象 1 復發，其餘 14 人仍具復發風險，此時涉險集合中包含 14 人。觀測對象 2 於觀察 104 天時復發，所以未發生復發的比率為 $(14-1)/14=0.929$ ，未發生復發的累積比率為 $0.929*0.933=0.867$ ，因此發生復發累積發生率則是 $1-0.867=0.133$ ；觀察時間介於 276~332 天時，此時涉險集合中包含 8 人，觀測對象 8 於觀察 276 天後被設限(設限原因為死亡)，但因非復發，所以此期間未發生復發的比率為 1，未發生復發的累積比率為 $0.593*1=0.593$ ，發生復發的累積發生率仍為 $1-0.593=0.407$ ；觀察時間介於 609~996 天時，此時涉險集合中有 2 人，觀測對象 14 於觀察 609 天後發生復發，此期間未發生復發的比例則為 $1/2$ ，因此未發生復發的累積比例為 $0.593*1/2=0.296$ ，發生復發的累積發生率為 $1-0.296=0.704$ 。

方法 2. 累積發生函數(cumulative incidence function, CIC)

此法是假設病人可能經歷復發或死亡，估計發生感興趣事件的累積發生率。此法不須假設競爭事件彼此獨立，但需要假設競爭風險事件彼此互斥。我們可利用 R 軟體中 library(cmprsk) 的 cuminc 語法來得到 CIC 估計。

R 語法(CIC 估計)

```
library(cmprsk)
data.all$cd=d3          #新增變數 cd(1 表示死亡，2 表示復發，0 表示設限)
data.all$cd[d3==1 & d1==1]=1
data.all$cd[d3==1 & d2==1]=2
cf=cmprsk(T2,data.all$cd,cencode=0) # CIC for each competing risk
km=survfit(Surv(T2,d3)~1,data=data.all) # 免於復發或死亡的機率
CI=timepoints(cf) # 產生要估計的時間點
CI$est #列出 CIC 估計
```

我們將結果整理如下表，表中符號定義如下：

t_{re} = 觀察的復發時間

cd_{re} = 考慮競爭風險下，復發人數

cd_d = 考慮競爭風險下，死亡人數

cy = 考慮競爭風險下，復發或與死亡時間大於或等於 t 的集合(at-risk set)

KM = 免於復發或死亡的機率

ch_{re} = 考慮競爭風險下，發生復發事件的風險 = cd_{re}/cy

CI_{re} = 考慮競爭風險下，復發的累積發生率

CI_d = 考慮競爭風險下，死亡的累積發生率 = $1 - KM - CI_{re}$

使用 CIC 方法估計復發的累積發生率

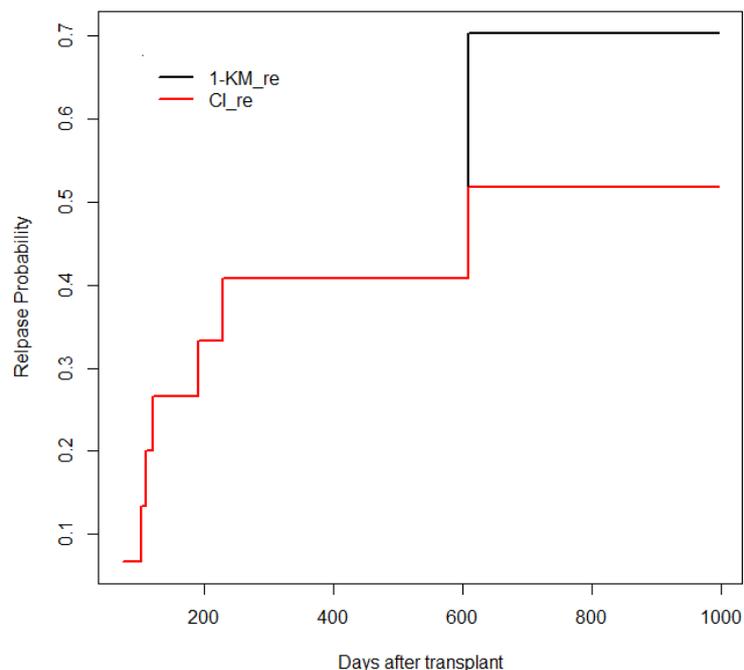
ID	t_{re}	cd	cd_{re}	cd_d	cy	KM	ch_{re}	CI_{re}	CI_d
1	74	2	1	0	15	0.933	0.067	0.067	0.000
2	104	2	1	0	14	0.867	0.071	0.134	0.000
3	110	2	1	0	13	0.800	0.077	0.200	0.000
4	122	2	1	0	12	0.733	0.083	0.267	0.000
5	192	2	1	0	11	0.667	0.091	0.333	0.000
6	226	0	0	0	10	0.667	0.000	0.333	0.000
7	230	2	1	0	9	0.593	0.111	0.407	0.000
8	276	1	0	1	8	0.519	0.000	0.407	0.074
9	332	1	0	1	7	0.444	0.000	0.407	0.148
10	418	1	0	1	6	0.370	0.000	0.407	0.222
11	466	1	0	1	5	0.296	0.000	0.407	0.296
12	526	1	0	1	4	0.222	0.000	0.407	0.370
13	530	0	0	1	3	0.222	0.000	0.407	0.370
14	609	2	1	0	2	0.111	0.500	0.518	0.370
15	996	0	0	0	1	0.111	0.000	0.518	0.370

實驗一開始有 15 人被追蹤觀察，故涉險集合中包含 15 人。74 天前，沒有人復發或死亡，因此免於疾病的存活率是 1。觀察 74 天時，觀察對象 1 發生復發(cd 為 2 表示復發)，因此發生復發的比率是 $1/15=0.067$ ，所以發生復發的發生率為前期的存活率(KM)*當期的風險(ch_{re})= $1*0.067=0.067$ ，發生復發的累積發生率則是 $1-0.933=0.067$ ；觀察時間介於 74~104 天時，除了觀測對象 1 復發，其餘 14 人仍具復發或死亡風險，此時涉險集合中包含 14 人，免於疾病的存活率是 0.933。觀測對象 2 於觀察 104 天時復發，所以發生復發的比例為 $1/14=0.071$ ，所以復發的發生率為 $0.933*0.071=0.067$ ，復發的累積發生率則為 $0.067+0.067=0.134$ ；觀察時間介於 276~332 天時，此時涉險集合中包含 8 人，觀測對象 8 於觀察 276 天後死亡，但因非復發，所以此期間發生復發的比率為 0，復發的發生率為 $0.593*0=0$ ，因此復發的累積發生率仍為 0.407；觀察時間介於 609~996 天時，此時涉險集合中有 2 人，觀測對象 14 於觀察 609 天後發生復發，所以此期間發生復發的比率為 $1/2$ ，復發的發生率為 $0.222*1/2=0.111$ ，發生復發的累積發生率則為 $0.407+0.111=0.518$ 。

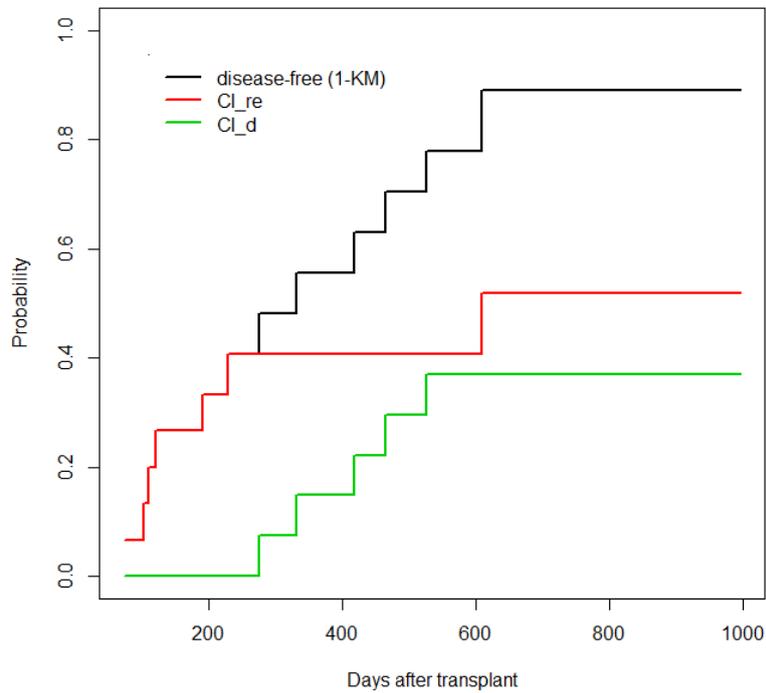
我們將表 1 的 $1-KM_{re}$ 跟表 2 的 CI_{re} 繪於圖一。由圖可以看出時間在 609 及以後的復發發生率不同。主要是因為 CIC 方法允許病人可能經歷復發或死亡，因此免於疾病的存活率會較低，估計復發的風險時不會高估。此外表 2 中的最後一欄是死亡的累積發生率，也就是 $1-KM-CI_{re}$ 。我們將結果會於圖二以呈現復發與死亡發生機率的相互影響。可以看出時間 609 天時，復發的累積發生率為 0.518，死亡的累積發生率為 0.370，免於疾病的存活率為 $1-0.518-0.378=0.111$ 。繪圖參考 R 語法如下：

```
plot(km2$time,1-km2$surv,type="s",lty=1,col=1,lwd=2,ylab="Relapse
Probability",xlab="Days after transplant")
points(km2$time,CI$est[2,],type="s",col=2,lwd=2)
legend(locator(2),c("1-KM_re","CI_re"),lwd=2,col=c(1,2))

plot(km$time,1-km$surv,type="s",lty=1,lwd=2,ylim=c(0,1),ylab="Probability",xlab="Days
after transplant")
points(km$time,CI$est[1,],type="s",col=3,lwd=2)
points(km$time,CI$est[2,],type="s",col=2,lwd=2)
legend(locator(2),c("disease-free (1-KM)","CI_re","CI_d"),lwd=2,col=c(1,2,3))
```



圖一、復發機率的比較



圖二、復發與死亡機率的相互影響

在醫學實務上，經常感興趣的是接受不同治療方式的病人其存活時間是否相同。Logrank test 是最常被利用於比較兩組或多組存活函數是否有差異的檢定方法。針對競爭風險，Gray(1988)提出一類似 logrank test 的檢定，分別就各競爭風險事件，進行兩組累積發生率的比較。利用 R 分析時，只要在 R 語法 `cuminc` 中加入 `group=rx` 的設定即可。語法與結果如下：

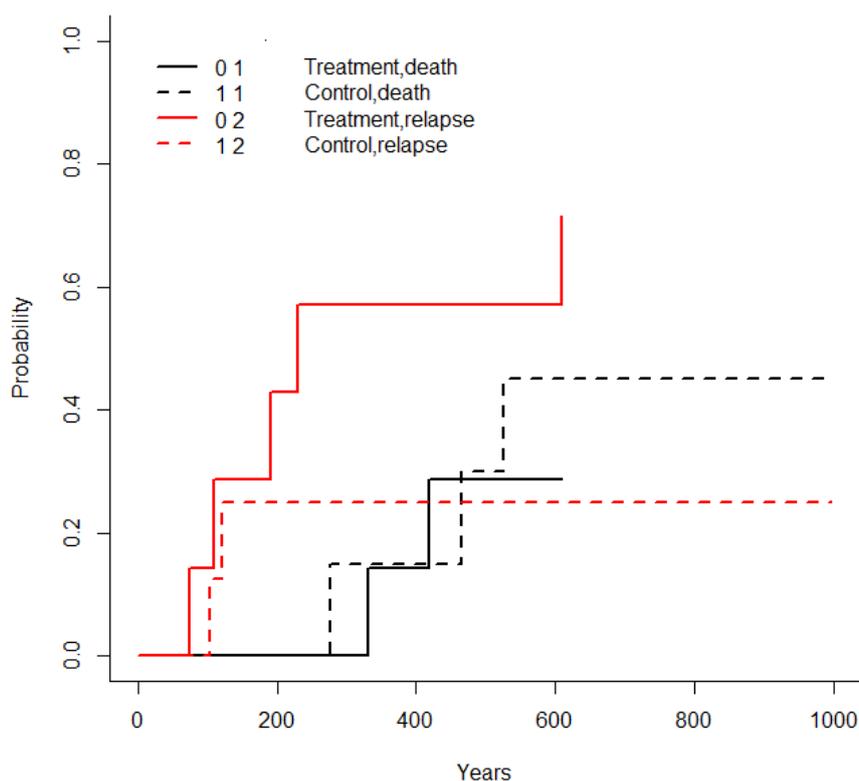
```
cfit1=cuminc(T2,data.all$scn, group=rx, cencode=0)
cfit1          #列出估計結果
cfit1$Tests   # 只列檢定結果
plot(cfit1,lwd=2,col=c(1,1,2,2),lty=c(1,2,1,2)) #圖三
```

不同競爭風險事件下，兩組 CIC 差異之檢定

Tests:

	stat	pv	df	
1	0.227	0.634	1	→ 檢定兩組的死亡累積發生率是否相同
2	1.788	0.181	1	→ 檢定兩組的復發累積發生率是否相同

若檢定處理組與對照組的累積發生率是否相同，由上表可以看出卡方檢定統計量為 1.788，p 值(pv)=0.181 > 0.05，所以不拒絕虛無假設，表示沒有足夠證據說明處理組與對照組的累積發生率有差異。由圖三可以亦可看出不管是死亡事件或復發事件，兩組的 CIC 沒有顯著差異，表示治療結果沒有不同。



圖三、處理組與對照組的死亡累積發生率及復發的累積發生率

參考文獻:

Gray RJ (1988). A class of K-sample tests for comparing the cumulative incidence of a competing risk, ANNALS OF STATISTICS, 16:1141-1154.

Kaplan EL, Meier P (1958). Nonparametric estimation from incomplete observations, Journal of American Statistical Association, 53:457-481.