Rで統計解析入門

(13) 生存時間解析〔後篇〕

本日のメニュー

1. 競合リスクに関する解析
2. 再発事象の解析

【復習】カプランマイヤー推定量

- 5人のがん患者さんに薬物療法を行い，「ガンの再発」をイベントとしてカプランマイヤー推定量によりイベント発生割合を計算する
- 何らかの理由でイベントを発生せずに観察を終了した場合は打ち切り

時間(日)	リスク集合	イベント	打ち切り	イベント無発生割合	イベント累積発生割合
0	5	-	-	100 %	0%
180	5	1	0	100×4/5 = 80%	0 + 100×1/5 = 20%
360	4	0	1	80%	20%
540	3	1	0	80×2/3 = 53.3%	20 + 80×1/3 = 46.7%
720	2	0	1	53.3%	46.7%
:	:	:	:	:	:

競合リスクとは

- 複数イベントにおいて，一方のイベントが観測されると他方のイベントは観測できないという場合がある
  - このようなイベントの関係を競合リスクイベントという
  - ガンの治療として薬物療法を行う場合，「ガンの再発」と「再発前の死亡」は競合リスクイベントとなる
    ⇒「再発した患者さん」においては「再発前の死亡」は起こらない
    ⇒「再発前に死亡した患者さん」においては「再発」は起こらない
- 競合リスクが存在する場合，カプランマイヤー推定量にはバイアスが入ることが知られている
  - 「ガンの再発」をイベントとした場合で「再発前の死亡」という競合リスクが存在する場合は、「再発前の死亡」を打ち切りとして扱ってしまうことでカプランマイヤー推定量にバイアスが入ってしまう

①の累積発生割合のカプランマイヤー推定

- 先ほどの 5 人のデータについて，イベントを「① ガンの再発」，打ち切りをもう一つのイベント「② 再発前の死亡」と定義し直して「① がんの再発」の累積発生割合を計算してみる
- その際，もう一つのイベント「② 再発前の死亡」が起きた場合は打ち切りと扱われることに注意する

時間 (日)	リスク集合	①が発生	②が発生	イベント無発生割合	①の累積発生割合
0	5	-	-	100 %	0%
180	5	1	0	?	?
360	4	0	1	?	?
540	3	1	0	?	?
720	2	0	1	?	?

① ガンの再発，② 再発前の死亡

①の累積発生割合のカプランマイヤー推定

- まず，180 日目にイベントが起こっているので，

  ①の無発生割合 ＝ 100×4/5 ＝ 80% ，①の累積発生割合 ＝ 20%

時間 (日)	リスク集合	①が発生	②が発生	イベント無発生割合	①の累積発生割合
0	5	-	-	100 %	0%
180	5	1	0	100×4/5 = 80 %	0 + 100×1/5 = 20 %
360	4	0	1		
540	3	1	0		
720	2	0	1		

① ガンの再発，② 再発前の死亡

①の累積発生割合のカプランマイヤー推定

- 次に，360 日目に打ち切り（イベント②）が起こっているので
  ①の無発生割合 ＝ 80%のまま，①の累積発生割合 ＝ 20%のまま
- しかし，実際は①も②も起こっていない人の割合は 60%（5人中3人）なので，「①の無発生割合」が 80%であるのは過大に推定しすぎ

時間 (日)	リスク集合	①が発生	②が発生	イベント無発生割合	①の累積発生割合
0	5	-	-	100 %	0%
180	5	1	0	100×4/5 = 80 %	0 + 100×1/5 = 20 %
360	4	0	1	80 %	20 %
540	3	1	0		
720	2	0	1		

① ガンの再発，② 再発前の死亡

①の累積発生割合のカプランマイヤー推定

- 次に，540 日目にイベントが起こっているので
  ①の無発生割合 ＝ 80%×2/3 ＝ 53.3%，①の累積発生割合 ＝ 46.7%
- 360 日目に「①の無発生割合」を 80 %と過大推定したため，この段階で「①の累積発生割合」が多めに足しこまれている

時間 (日)	リスク集合	①が発生	②が発生	イベント無発生割合	①の累積発生割合
0	5	-	-	100 %	0%
180	5	1	0	100×4/5 = 80 %	0 + 100×1/5 = 20 %
360	4	0	1	80 %	20 %
540	3	1	0	80×2/3 = 53.3 %	20 + 80×1/3 = 46.7 %
720	2	0	1		

① ガンの再発，② 再発前の死亡

①の累積発生割合のカプランマイヤー推定

- 720 日目も同様に計算し，「①の累積発生割合」は 46.7% となった

時間(日)	リスク集合	①が発生	②が発生	イベント無発生割合	①の累積発生割合
0	5	-	-	100 %	0%
180	5	1	0	100×4/5 = 80 %	0 + 100×1/5 = 20 %
360	4	0	1	80 %	20 %
540	3	1	0	80×2/3 = 53.3 %	20 + 80×1/3 = 46.7 %
720	2	0	1	53.3 %	46.7 %

① ガンの再発，② 再発前の死亡

②の累積発生割合のカプランマイヤー推定

- 先ほどの 5 人のデータについて，イベントを「② 再発前の死亡」，打ち切りをもう一つのイベント「① ガンの再発」と定義し直して「② 再発前の死亡」の累積発生割合を計算してみる
- その際，もう一つのイベント「① ガンの再発」が起きた場合は打ち切りと扱われることに注意する

時間 (日)	リスク集合	①が発生	②が発生	イベント無発生割合	②の累積発生割合
0	5	-	-	100 %	0%
180	5	1	0	?	?
360	4	0	1	?	?
540	3	1	0	?	?
720	2	0	1	?	?

① ガンの再発，② 再発前の死亡

②の累積発生割合のカプランマイヤー推定

- まず，180 日目に打ち切り（イベント①）が起こっているので
  ②の無発生割合 ＝ 100%のまま，②の累積発生割合 ＝ 0%のまま
- しかし，実際は①も②も起こっていない人の割合は 80%（5人中4人）なので，「②の無発生割合」が 100%であるのは過大に推定しすぎ

時間(日)	リスク集合	①が発生	②が発生	イベント無発生割合	②の累積発生割合
0	5	-	-	100 %	0%
180	5	1	0	100 %	0%
360	4	0	1		
540	3	1	0		
720	2	0	1		

① ガンの再発，② 再発前の死亡

②の累積発生割合のカプランマイヤー推定

- 360 日目にイベントが起こっているので,
  ②の無発生割合 ＝ 100×3/4 ＝ 75% , ②の累積発生割合 ＝ 25%
- 180 日目に「②の無発生割合」を 100 %と過大推定したため, この段階で「②の累積発生割合」が多めに足しこまれている

時間 (日)	リスク集合	①が発生	②が発生	イベント 無発生割合	②の累積発生割合
0	5	-	-	100 %	0%
180	5	1	0	100 %	0%
360	4	0	1	100×3/4 = 75%	0 + 100×1/4 = 25%
540	3	1	0		
720	2	0	1		

① ガンの再発, ② 再発前の死亡

②の累積発生割合のカプランマイヤー推定

- 次に，540 日目に打ち切り（イベント①）が起こっているので
  ②の無発生割合 ＝ 75%のまま，②の累積発生割合 ＝ 25%のまま
- しかし，実際は①も②も起こっていない人の割合は 40%（5人中2人）なので，「②の無発生割合」が 75%であるのは過大に推定しすぎ

時間 (日)	リスク集合	①が発生	②が発生	イベント無発生割合	②の累積発生割合
0	5	-	-	100 %	0%
180	5	1	0	100 %	0%
360	4	0	1	100×3/4 = 75%	0 + 100×1/4 = 25%
540	3	1	0	75 %	25 %
720	2	0	1		

① ガンの再発，② 再発前の死亡

②の累積発生割合のカプランマイヤー推定

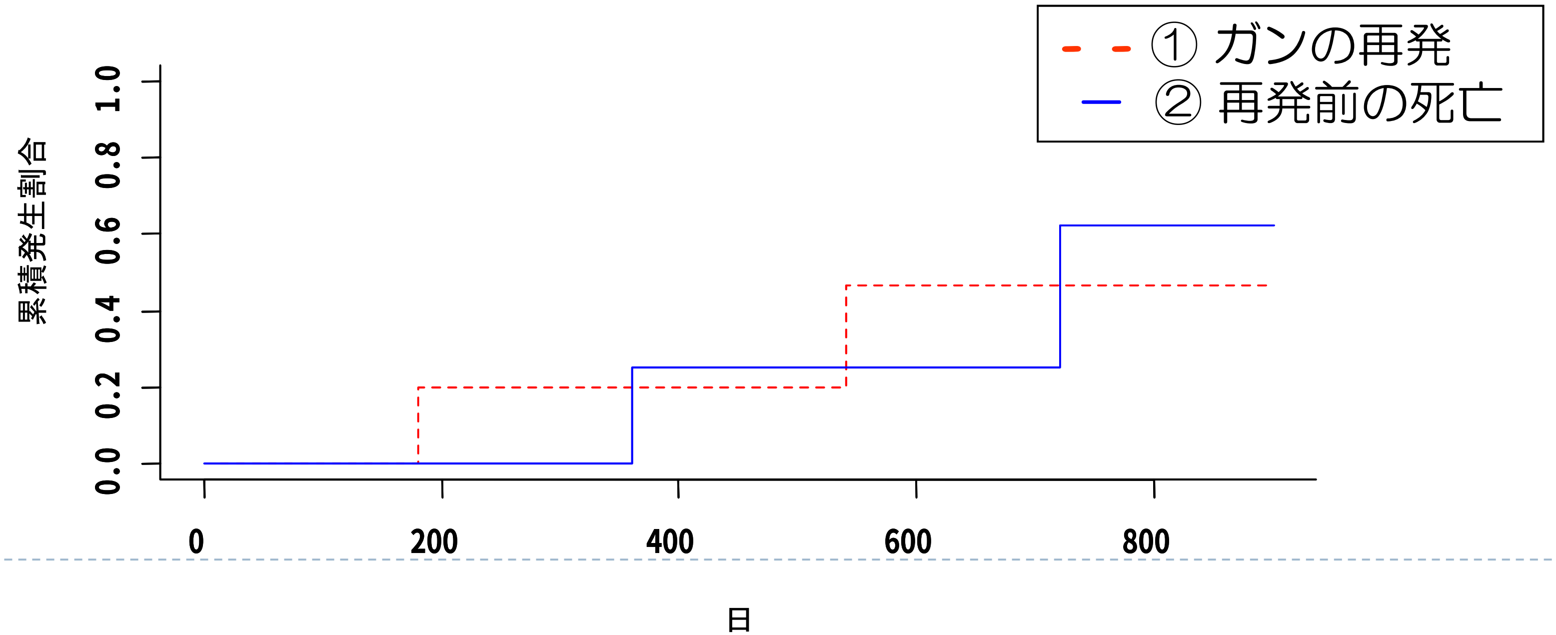
- 次に，720 日目にイベントが起こっているので
  ②の無発生割合 ＝ 75%×1/2 ＝ 37.5%，②の累積発生割合 ＝ 62.5%
- 540 日目に「②の無発生割合」を 75 %と過大推定したため，この段階で「②の累積発生割合」が多めに足しこまれている

時間 (日)	リスク集合	①が発生	②が発生	イベント無発生割合	②の累積発生割合
0	5	-	-	100 %	0%
180	5	1	0	100 %	0%
360	4	0	1	100×3/4 = 75%	0 + 100×1/4 = 25%
540	3	1	0	75 %	25 %
720	2	0	1	75×1/2 = 37.5 %	25 + 75×1/2 = 62.5 %

① ガンの再発，② 再発前の死亡

競合リスクがある時のカプランマイヤー推定の問題点

- 累積発生割合：①＝ 46.7%，②＝ 62.5% ⇒ 足すと 100% を超える…
- ①と②の2つの状態しか取り得ないはずなのに 100% を超えてしまうというおかしな現象は，興味のあるイベントの発生割合を求める際に，もう一方のイベント（競合リスク）が起きた時に打ち切りと扱い，無発生割合を過大推定することが原因となっている



前頁のグラフを作成するプログラム

```
> library(cmprsk)
> ALL <- data.frame(time  =c(180, 360, 540, 720, 900),
+                   censor=c(1, 2, 1, 2, 0), group=rep("A", 5) )
> EVENT        <- All                         # 1 をイベント,それ以外を打ち切り(0)
> EVENT$censor <- ifelse(ALL$censor==1, 1, 0)
> COMP         <- All                         # 2 をイベント,それ以外を打ち切り(0)
> COMP$censor  <- ifelse(ALL$censor==2, 1, 0)
> result  <- cuminc(EVENT$time, EVENT$censor, cencode=0)
> result2 <- timepoints(result, EVENT$time)
> result2$"est"                               # イベント①の累積発生割合
    180 360       540       720       900
1 1 0.2 0.2 0.4666667 0.4666667 0.4666667
> plot(result, col="red", lty=2, curvlab="", ann=F)
> par(new=T)
> result  <- cuminc(COMP$time, COMP$censor, cencode=0)
> result2 <- timepoints(result, COMP$time)
> result2$"est"                               # イベント②の累積発生割合
    180  360  540   720   900
1 1   0 0.25 0.25 0.625 0.625
> plot(result, col="blue", curvlab="", xlab="日", ylab="累積発生割合")
```

競合リスクとは

- 競合リスクがある場合に，興味のあるイベントの発生割合をカプラン・マイヤー推定で求めると，もう一方のイベント（競合リスク）が起きた時に打ち切りと扱い，無発生割合を過大推定してしまう
- これを解消する方法が Cumulative Incidence Function（イベントごとに累積発生割合にきちんと計算する）という方法
- 次頁以降で，この方法により「① がんの再発」と「② 再発前の死亡」の発生割合を計算してみる

発生割合を算出するルール（競合リスク版）

1. 0 日目の「イベントの無発生割合」を 1（100%），他を 0% とする
2. 打ち切りが起こった場合は「直前までの無発生割合」をそのまま引き継ぎ，生き残っている人の数（リスク集合）を減らす
3. イベントが発生した時点で，イベントごとに以下の計算を実行する
   ⇒ 例として，イベント①と②が同日に発生したとする
   「イベントの無発生割合」＝「直前までの無発生割合」
   ×「この瞬間に生き残っている人の割合」
   「イベント①の累積発生割合」＝「直前までの累積発生割合」
   ＋「直前までの無発生割合」×「イベント①を発生した人の割合」
   「イベント②の累積発生割合」＝「直前までの累積発生割合」
   ＋「直前までの無発生割合」×「イベント②を発生した人の割合」
4. 同じ日にイベントと打ち切りが起こった場合は，先にイベントが起こりその次の瞬間に打ち切りが起こったとする

Cumulative Incidence Function による推定 (i)

- Cumulative Incidence Function（イベントごとに累積発生割合にきちんと計算する）という方法により「① がんの再発」と「② 再発前の死亡」の発生割合を計算してみる

日	N	①	②	無発生割合	①の発生割合	②の発生割合
0	5	—	—	100 %	0 %	0 %
180	5	1	0	?	?	?
360	4	0	1	?	?	?
540	3	1	0	?	?	?
720	2	0	1	?	?	?
900+	1	0	0	?	?	?

※ N：リスク集合，① ガンの再発，② 再発前の死亡，+：打ち切り（900日目）

Cumulative Incidence Function による推定 (i)

- 180 日目にイベントが起こっているので，ルール 3 を適用し，
  - 無発生割合 ＝ 100×4/5 ＝ 80%
  - ①の累積発生割合 ＝ 0 + 100×1/5 ＝ 20%
  - ②の累積発生割合 ＝ 0% のまま

日	N	①	②	無発生割合	①の発生割合	②の発生割合
0	5	—	—	100 %	0 %	0 %
180	5	1	0	100×4/5 = 80 %	0 + 100×1/5 = 20 %	0 %
360	4	0	1			
540	3	1	0			
720	2	0	1			
900+	1	0	0			

※ N：リスク集合，① ガンの再発，② 再発前の死亡，+：打ち切り（900日目）

Cumulative Incidence Function による推定 (i)

- 360 日目にイベントが起こっているので，ルール 3 を適用し，
  - 無発生割合　　　　＝ 80×3/4 ＝ 60%
  - ①の累積発生割合 ＝ 20% のまま
  - ②の累積発生割合 ＝ 0 + 80×1/4 ＝ 20%

日	N	①	②	無発生割合	①の発生割合	②の発生割合
0	5	—	—	100 %	0 %	0 %
180	5	1	0	100×4/5 = 80 %	0 + 100×1/5 = 20 %	0 %
360	4	0	1	80×3/4 = 60 %	20 %	0 + 80×1/4 = 20 %
540	3	1	0			
720	2	0	1			
900+	1	0	0			

※ N：リスク集合，① ガンの再発，② 再発前の死亡，+：打ち切り（900日目）

Cumulative Incidence Function による推定 (i)

- 540 日目にイベントが起こっているので，ルール 3 を適用し，
  - 無発生割合　　　　　＝ 60×2/3 ＝ 40%
  - ①の累積発生割合 ＝ 20 + 60×1/3 ＝ 40%
  - ②の累積発生割合 ＝ 20% のまま

日	N	①	②	無発生割合	①の発生割合	②の発生割合
0	5	―	―	100 %	0 %	0 %
180	5	1	0	100×4/5 = 80 %	0 + 100×1/5 = 20 %	0 %
360	4	0	1	80×3/4 = 60 %	20 %	0 + 80×1/4 = 20 %
540	3	1	0	60×2/3 = 40 %	20 + 60×1/3 = 40 %	20 %
720	2	0	1			
900+	1	0	0			

※ N：リスク集合，① ガンの再発，② 再発前の死亡，+：打ち切り（900日目）

Cumulative Incidence Function による推定 (i)

- 720 日目にイベントが起こっているので，ルール 3 を適用し，
  - 無発生割合　　　　＝ 40×1/2 ＝ 20%
  - ①の累積発生割合 ＝ 40% のまま
  - ②の累積発生割合 ＝ 20 + 40×1/2 ＝ 40%

日	N	①	②	無発生割合	①の発生割合	②の発生割合
0	5	—	—	100 %	0 %	0 %
180	5	1	0	100×4/5 = 80 %	0 + 100×1/5 = 20 %	0 %
360	4	0	1	80×3/4 = 60 %	20 %	0 + 80×1/4 = 20 %
540	3	1	0	60×2/3 = 40 %	20 + 60×1/3 = 40 %	20 %
720	2	0	1	20×1/2 = 20 %	40 %	20 + 40×1/2 = 40 %
900+	1	0	0			

※ N：リスク集合，① ガンの再発，② 再発前の死亡，+：打ち切り（900日目）

Cumulative Incidence Function による推定 (i)

- 900 日目に打ち切りが起こっているので，ルール 2 を適用し，分母のみ 1 → 0 人に減らす
  - 無発生割合 ＝ 20%
  - ①の累積発生割合 ＝ 40%
  - ②の累積発生割合 ＝ 20%

日	N	①	②	無発生割合	①の発生割合	②の発生割合
0	5	—	—	100 %	0 %	0 %
180	5	1	0	100×4/5 = 80 %	0 + 100×1/5 = 20 %	0 %
360	4	0	1	80×3/4 = 60 %	20 %	0 + 80×1/4 = 20 %
540	3	1	0	60×2/3 = 40 %	20 + 60×1/3 = 40 %	20 %
720	2	0	1	20×1/2 = 20 %	40 %	20 + 40×1/2 = 40 %
900+	1	0	0	20 %	40 %	40 %

※ N：リスク集合，① ガンの再発，② 再発前の死亡，+：打ち切り（900日目）

Cumulative Incidence Function による推定 (ii)

- 別の例で Cumulative Incidence Function により「① がんの再発」と「② 再発前の死亡」の発生割合を計算してみる
- 180 日目にイベントが起こっているので，ルール 3 を適用し，
  - 無発生割合　　　＝ 100×4/5 ＝ 80%
  - ①の累積発生割合 ＝ 0 + 100×1/5 ＝ 20%
  - ②の累積発生割合 ＝ 0% のまま

日	N	①	②	無発生割合	①の発生割合	②の発生割合
0	5	―	―	100 %	0 %	0 %
180	5	1	0	100×4/5 = 80 %	0 + 100×1/5 = 20 %	0 %
360	4	1	2			
900+	1	0	0			

※ N：リスク集合，① ガンの再発，② 再発前の死亡，+：打ち切り（900日目）

Cumulative Incidence Function による推定 (ii)

- 360 日目にイベントが起こっているので，ルール 3 を適用し，
  - 無発生割合　　　　＝ 80×1/4 ＝ 20%
  - ①の累積発生割合 ＝ 20 + 80×1/4 ＝ 40%
  - ②の累積発生割合 ＝ 0 + 80×1/4 ＝ 40%
- 900 日目に打ち切りが起こっているので，ルール 2 を適用し，
  分母のみ 1 → 0 人に減らす（割合は 360日目のものと同じ）

日	N	①	②	無発生割合	①の発生割合	②の発生割合
0	5	—	—	100 %	0 %	0 %
180	5	1	0	100×4/5 = 80 %	0 + 100×1/5 = 20 %	0 %
360	4	1	2	80×1/4 = 20 %	20 + 80×1/4 = 40 %	0 + 80×2/4 = 40 %
900+	1	0	0	20 %	40 %	40 %

※ N：リスク集合，① ガンの再発，② 再発前の死亡，+：打ち切り（900日目）

Cumulative Incidence Function による推定 (iii)

- さらに別の例で「① がんの再発」と「② 再発前の死亡」の割合を計算
- 180 日目に打ち切り：ルール 2 を適用し，分母のみ減らす
- 360 日目にイベントが起こっているので，ルール 3 を適用し，
  - 無発生割合　　　　＝ 100×3/4 ＝ 75%
  - ①の累積発生割合 ＝ 0 + 100×1/4 ＝ 25%
  - ②の累積発生割合 ＝ 0% のまま

日	N	①	②	無発生割合	①の発生割合	②の発生割合
0	5	—	—	100 %	0 %	0 %
180+	5	0	0	100 %	0 %	0 %
360	4	1	0	100×3/4 = 75 %	0 + 100×1/4 = 25 %	0 %
540+	3	0	1			
900	1	1	0			

※ N：リスク集合，① ガンの再発，② 再発前の死亡，+：打ち切り（180,540日目）

Cumulative Incidence Function による推定 (iii)

- 540 日目にイベントと打ち切りが起こっているので，ルール 4 より先にイベントの処理 ⇒ まずルール 3 を適用（処理後の分母は 3 → 2 人）
  - 無発生割合 ＝ 75×2/3 ＝ 50%
  - ①の累積発生割合 ＝ 25% のまま
  - ②の累積発生割合 ＝ 0 + 75×1/3 ＝ 25%

次に，ルール 2 より分母を 2 → 1 人に減らす

日	N	①	②	無発生割合	①の発生割合	②の発生割合
0	5	―	―	100 %	0 %	0 %
180+	5	0	0	100 %	0 %	0 %
360	4	1	0	100×3/4 = 75 %	0 + 100×1/4 = 25 %	0 %
540+	3	0	1	75×2/3 = 50 %	25 %	0 + 75×1/3 = 25 %
900	1	1	0			

※ N：リスク集合，① ガンの再発，② 再発前の死亡，+：打ち切り（180,540日目）

Cumulative Incidence Function による推定 (iii)

- 900 日目にイベントが起こっているので，ルール 3 を適用し，
  - 無発生割合　　　＝ 50×0/1 ＝ 0%
  - ①の累積発生割合 ＝ 25 + 50×1/1 ＝ 75%
  - ②の累積発生割合 ＝ 25% のまま

日	N	①	②	無発生割合	①の発生割合	②の発生割合
0	5	—	—	100 %	0 %	0 %
180+	5	0	0	100 %	0 %	0 %
360	4	1	0	100×3/4 = 75 %	0 + 100×1/4 = 25 %	0 %
540+	3	0	1	75×2/3 = 50 %	25 %	0 + 75×1/3 = 25 %
900	1	1	0	50×0/1 = 0 %	25 + 50×1/1 = 75 %	25 %

※ N：リスク集合，① ガンの再発，② 再発前の死亡，+：打ち切り（180,540日目）

Cumulative Incidence Function による推定

```
> # 競合リスクに関する解析例(i)
> result  <- cuminc(ALL$time, ALL$censor, cencode=0)
> result2 <- timepoints(result, ALL$time)
> result2$"est"
    180  360  540  900
1 1   0 0.25 0.25 0.75
1 2   0 0.00 0.25 0.25
> plot(result, col=1:2, xlab="日", ylab="累積発生割合",
+      curvlab=c("①","②"))
> # 競合リスクに関する解析例(ii)
> ALL <- data.frame(time  =c(180, 360, 360, 360, 900),
+                   censor=c(1, 2, 1, 2, 0),
+                   group =rep("A", 5) )
> result  <- cuminc(ALL$time, ALL$censor, cencode=0)
> result2 <- timepoints(result, ALL$time)
> result2$"est"
    180 360 900
1 1 0.2 0.4 0.4
1 2 0.0 0.4 0.4
> plot(result, col=1:2, xlab="日", ylab="累積発生割合",
+      curvlab=c("①","②"))
>
```

Cumulative Incidence Function による推定

```
> # 競合リスクに関する解析例(iii)
> ALL <- data.frame(time  =c(180, 360, 540, 540, 900),
+                   censor=c(0, 1, 2, 0, 1),
+                   group =rep("A", 5) )
> result  <- cuminc(ALL$time, ALL$censor, cencode=0)
> result2 <- timepoints(result, ALL$time)
> result2$"est"
    180  360  540  900
1 1   0 0.25 0.25 0.75
1 2   0 0.00 0.25 0.25
> plot(result, col=1:2, xlab="日", ylab="累積発生割合",
+      curvlab=c("①","②"))
```

【参考】 Cumulative Incidence Function に関するモデル解析

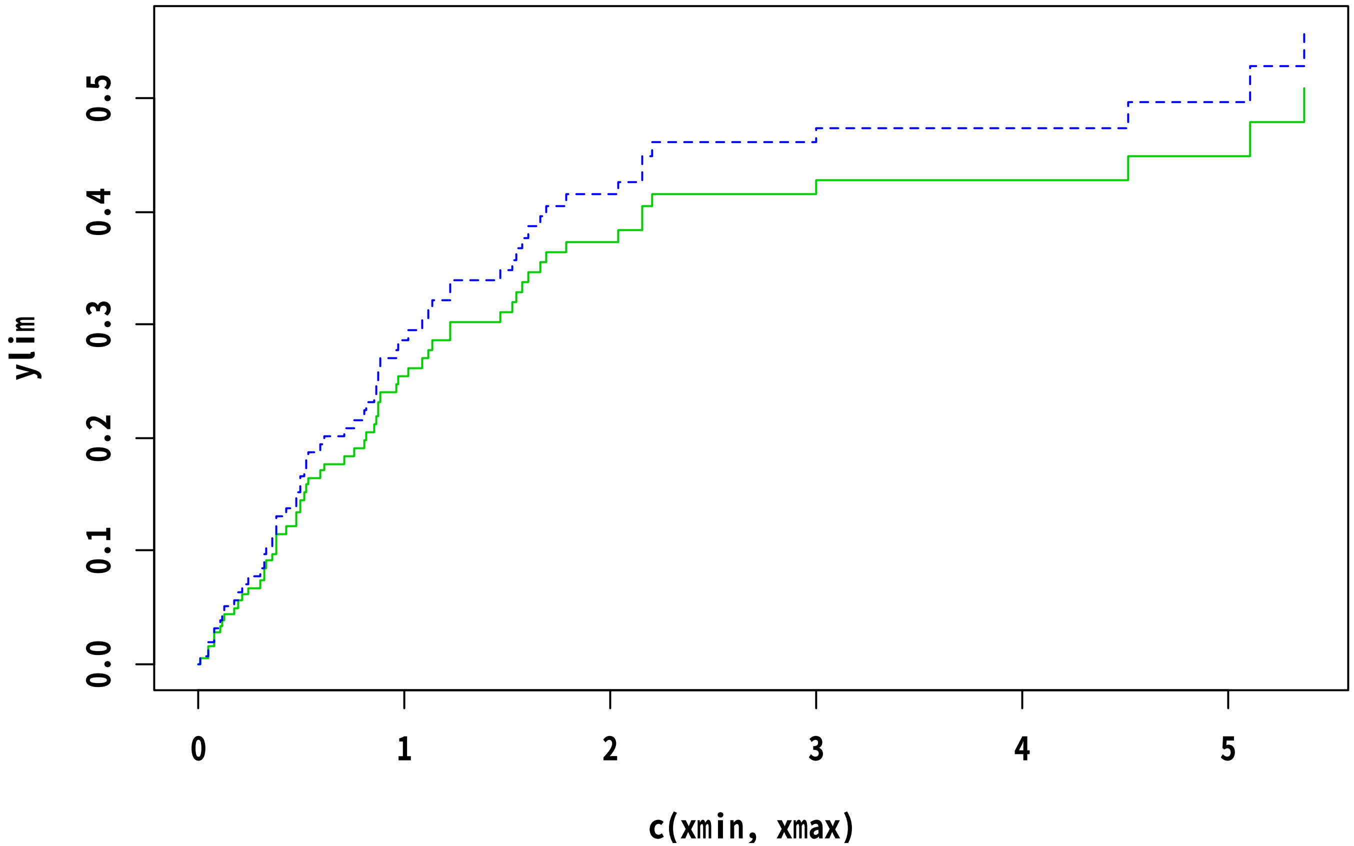
- Fine and Gray の方法によるモデル解析（共変量を入れた解析）
- 関数 crr の引数：failcode=1 ；興味のあるイベント
- 関数 crr の引数：cencode=0；打ち切り

```
> # CIFの比較
> set.seed(777);  ftime <- rexp(200)
> fstatus <- sample(0:2, 200, replace=TRUE)
> cov <- matrix(runif(600), nrow=200)
> dimnames(cov)[[2]] <- c('x1','x2','x3')
> ( result <- crr(ftime, fstatus, cov, failcode=2) )
convergence:  TRUE
coefficients:
      x1       x2       x3
-0.39880  0.09532 -0.23440
standard errors:
[1] 0.4379 0.4757 0.4035
two-sided p-values:
  x1   x2   x3
0.36 0.84 0.56
```

※ Fine JP and Gray RJ (1999) A proportional hazards model for the subdistribution of a competing risk. JASA 94:496-509.

【参考】 Cumulative Incidence Function に関するモデル解析

```
> result.p <- predict(result, rbind(c(.1,.5,.8),c(.1,.5,.2)))
> plot(result.p, lty=1:2, color=3:4)
```



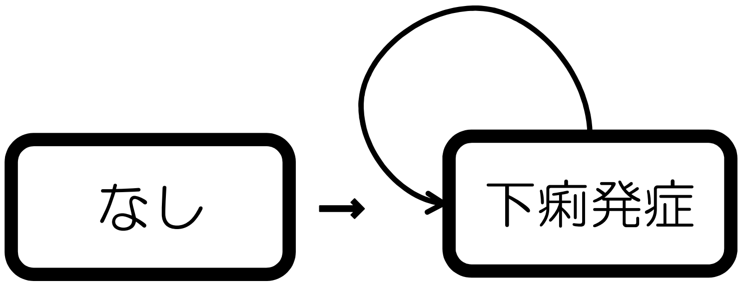
※ Fine JP and Gray RJ (1999) A proportional hazards model for the subdistribution of a competing risk. JASA 94:496-509.

本日のメニュー

1. 競合リスクに関する解析
2. 再発事象の解析

再発事象の解析

- これまでの内容は「各被験者のイベントは 0 回か 1 回」という仮定
- 以降では同一被験者で複数回イベントが起こり得ると仮定する

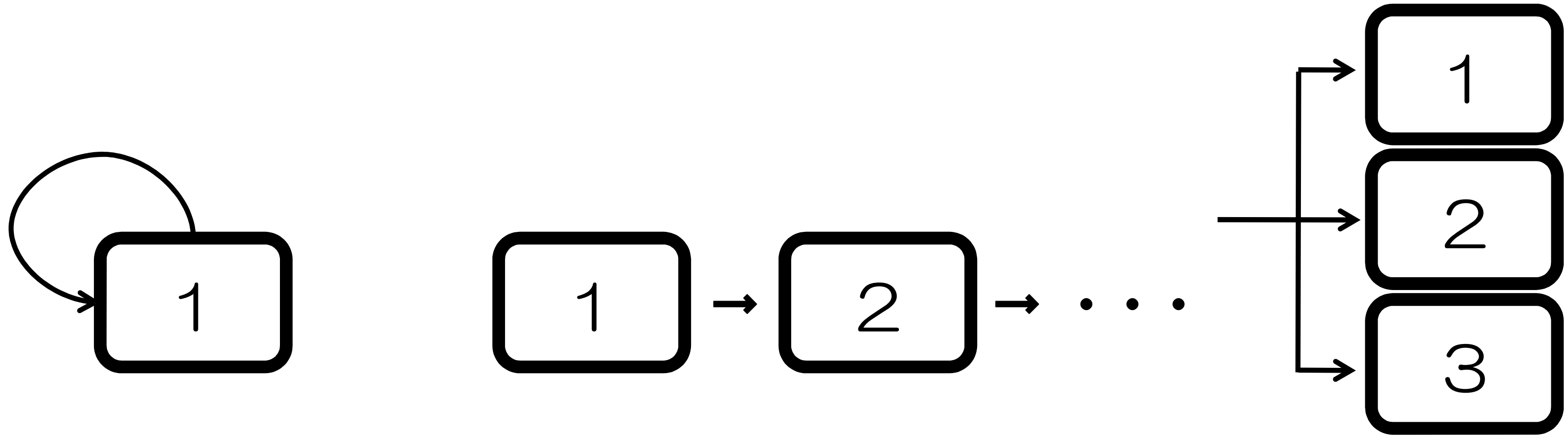


【下痢などの消化器症状】

【 ガン 】

再発事象の解析

- これまでの内容は「各被験者のイベントは 0 回か 1 回」という仮定
- 以降では同一被験者で複数回イベントが起こり得ると仮定する
- AG モデル　　：同じイベントが再発
- PWP モデル　：イベントが起こるたびに個体は新たな層へ
- WLW モデル：個体は最初から複数のイベントに対するリスク集合に



【 AG モデル 】　【 PWP モデル 】　【 WLW モデル 】

再発事象の解析

- 同一患者さんから 2 回以上のイベントが観測されうる場合に対応する解析手法を紹介する
- 例として，うつ病を患っている 5 人の患者さんのデータを用いる

ID	GROUP	EVENT	START	END	DIFF	STRATUM
1	A	1	0	1	1	1
1	A	1	1	2	1	2
1	A	0	2	4	2	3
2	A	1	0	6	6	1
2	A	1	6	8	2	2
3	B	0	0	3	3	1
4	B	1	0	5	5	1
4	B	0	5	7	2	2
5	B	1	0	9	9	1

再発事象の解析

- ID：患者さん，GROUP：薬剤の種類
- EVENT：1 → あり（イベント），0 → なし（打ち切り）
- START：観察開始時（年），END：観察終了時（年）
- DIFF：観察期間（年），STRATUM：各患者さんの何行目のデータか

ID	GROUP	EVENT	START	END	DIFF	STRATUM
1	A	1	0	1	1	1
1	A	1	1	2	1	2
1	A	0	2	4	2	3
2	A	1	0	6	6	1
2	A	1	6	8	2	2
3	B	0	0	3	3	1
4	B	1	0	5	5	1
4	B	0	5	7	2	2
5	B	1	0	9	9	1

再発事象の解析

- ID：患者さん，GROUP：薬剤の種類
- EVENT：1 → あり（イベント），0 → なし（打ち切り）
- START：観察開始時（年），END：観察終了時（年）
- DIFF：観察期間（年），STRATUM：各患者さんの何行目のデータか

```
> MULTI <- data.frame(ID     =c(1,1,1,2,2,3,4,4,5),
+                     GROUP  =c(rep("A",5), rep("B",4)),
+                     EVENT  =c(1,1,0,1,1,0,1,0,1),
+                     START  =c(0,1,2,0,6,0,0,5,0),
+                     END    =c(1,2,4,6,8,3,5,7,9),
+                     DIFF   =c(1,1,2,6,2,3,5,2,9),
+                     STRATUM=c(1,2,3,1,2,1,1,2,1) )

> MULTI$GROUP <- relevel(MULTI$GROUP, ref="B")      # カテゴリのベースを「B」に変更
```

再発事象の解析

- ID=1 の患者さんの場合（薬剤 A ）

1 行目：1回目のイベントを観察するための期間（0～1）→1年後にイベント発生

2 行目：2回目のイベントを観察するための期間（1～2）→2年後にイベント発生

3 行目：3回目のイベントを観察するための期間（2～4）→4年後に打ち切り

ID	GROUP	EVENT	START	END	DIFF	STRATUM
1	A	1	0	1	1	1
1	A	1	1	2	1	2
1	A	0	2	4	2	3
2	A	1	0	6	6	1
2	A	1	6	8	2	2
3	B	0	0	3	3	1
4	B	1	0	5	5	1
4	B	0	5	7	2	2
5	B	1	0	9	9	1

再発事象の解析

- ID=1 の患者さんの 2 行目の情報について
  - 2 回目のイベントである（STRATUM=2）であることを考慮するか，考慮しない（何回目であってもイベント発現率は同じと仮定する）か
  - 一番最初から何年か（END），2 回目の開始時（ 1 回目の終了時）から何年か（DIFF）

ID	GROUP	EVENT	START	END	DIFF	STRATUM
1	A	1	0	1	1	1
1	A	1	1	2	1	2
1	A	0	2	4	2	3
2	A	1	0	6	6	1
2	A	1	6	8	2	2
3	B	0	0	3	3	1
4	B	1	0	5	5	1
4	B	0	5	7	2	2
5	B	1	0	9	9	1

再発事象の解析

- Cox 回帰モデル（関数 coxph() ）を用いることで再発事象に関する解析を行うことが出来る
- 解析できるモデルは以下の通り
  - AG モデル：Andersen and Gill（1982）で提案されたモデル
  - PWP モデル：Prentice, Williams and Peterson（1981）で提案されたモデル
    - Conditional Probability（PWP-CP モデルと呼ぶ）
    - Gap Time（PWP-GT モデルと呼ぶ）
  - TT-R モデル：Kelly and Lim（2000）で提案されたモデル
    →「TT-R」は「Total Time-Restricted」の略
  - WLW モデル：Wei, Lin and Weissfeld（1989）で提案されたモデル

再発事象の解析

- Cox 回帰モデル（関数 coxph() ）を用いることで再発事象に関する解析を行うことが出来る
- 解析できるモデルは以下の通り
  - AG モデル　　　：イベントの独立性が非現実的
  - PWP-CP モデル ：試験期間全体を通じたモデル化をするときにお勧め
  - PWP-GT モデル：最終イベントからの時間をモデル化するときにお勧め
  - TT-R モデル　　： PWP-CP モデルとほぼ同様
  - WLW モデル　　：観測していないデータを補完するのが非現実的
    （たまに適する状況もある）

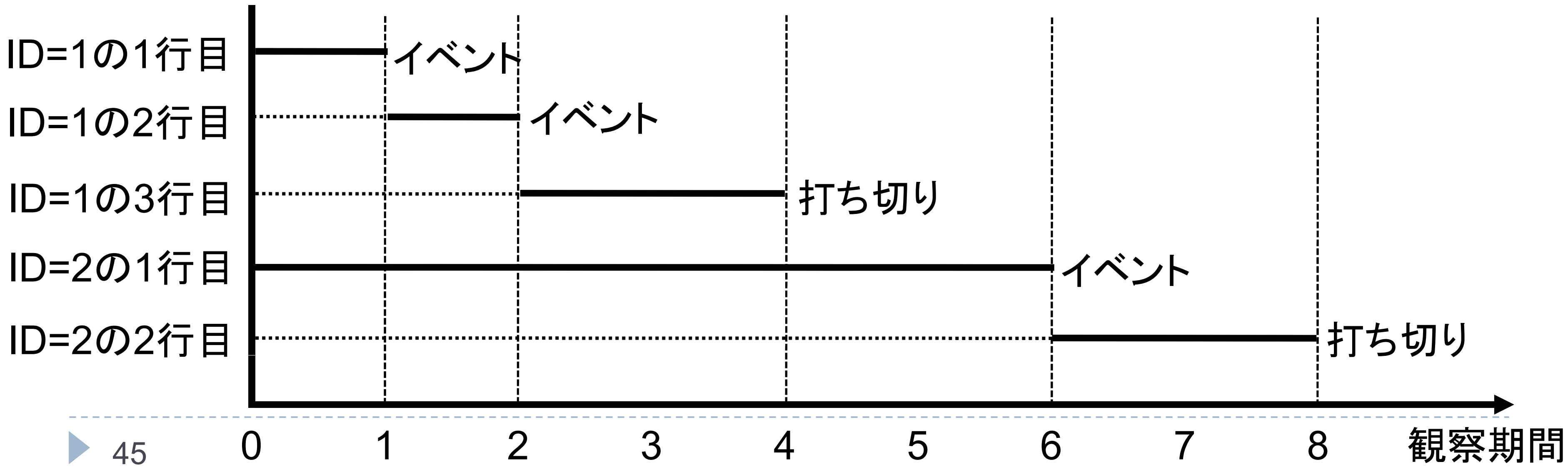
AG モデルを用いた解析（薬剤 A のみ抽出）

- 何回目のイベントであろうともイベントの起こりやすさ（正確にはベースラインハザード関数）は同じと仮定する
- ID=1 の患者さんの 2 行目（2回目のイベント）の観察開始時は，1 行目（ 1 回目のイベント）の観察終了時となるので，2 行目（2回目のイベント）の観察期間は 2-1 ＝ 1 年となる
- 使用する変数：ID，GROUP，EVENT，START，END

ID	GROUP	EVENT	START	END	DIFF	STRATUM
1	A	1	0	1	1	1
1	A	1	1	2	1	2
1	A	0	2	4	2	3
2	A	1	0	6	6	1
2	A	1	6	8	2	2

AG モデルを用いた解析（薬剤 A のみ抽出）

- 何回目のイベントであろうともイベントの起こりやすさ（正確にはベースラインハザード関数）は同じと仮定する
- ID=1 の患者さんの 2 行目（2回目のイベント）の観察開始時は，1 行目（ 1 回目のイベント）の観察終了時となるので，2 行目（2回目のイベント）の観察期間は 2-1 ＝ 1 年となる



AG モデルを用いた解析

- coef：ハザード比が exp(1.84) = 6.33 倍
- se(coef)：同一 ID のデータの相関を考慮していない標準誤差（見てはダメ）
- robust se：同一 ID のデータの相関を考慮したロバストな標準誤差は 0.909

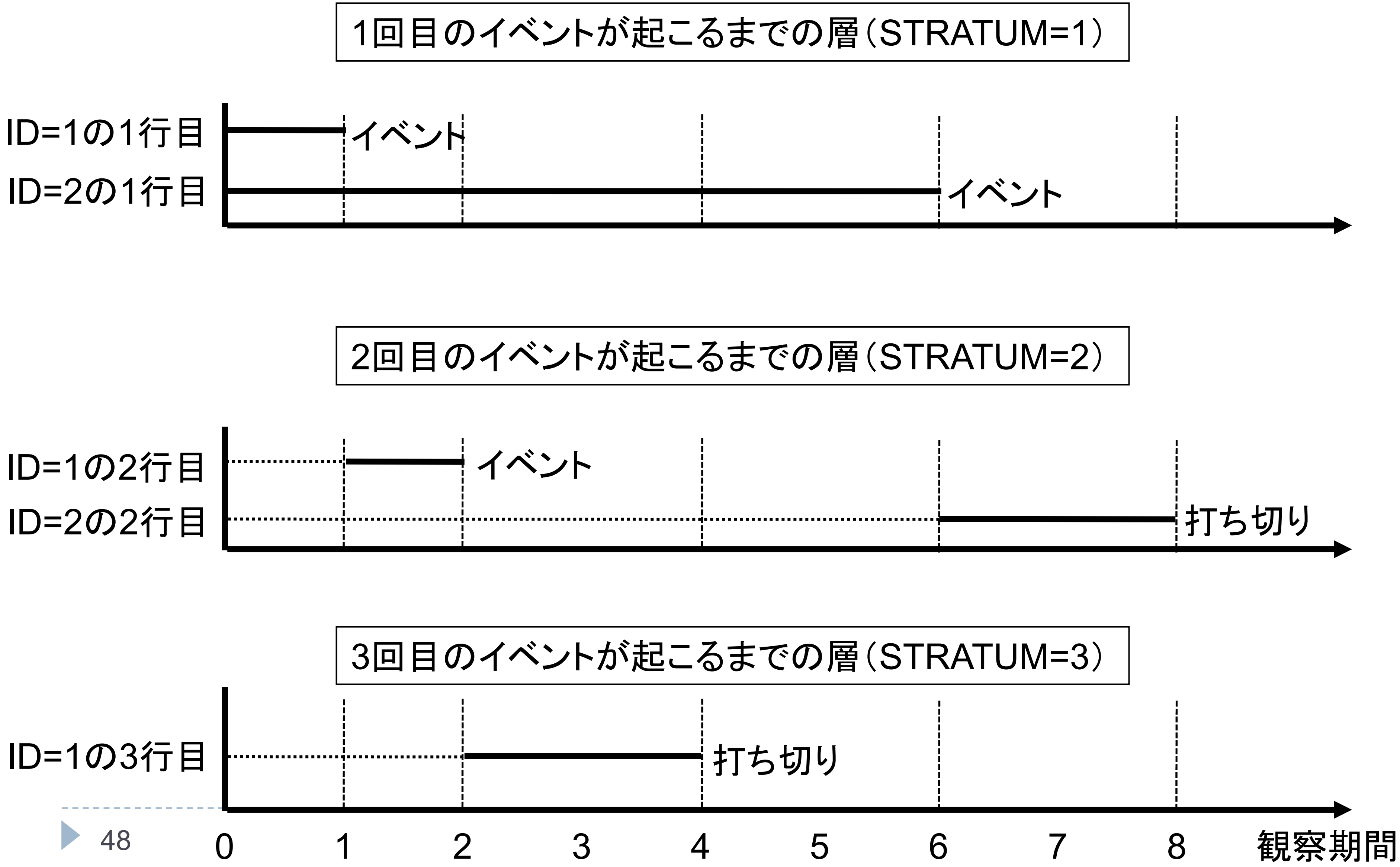
```
> library(survival)
> coxph(Surv(START,END,EVENT) ~ GROUP + cluster(ID), data=MULTI)
Call:
coxph(formula = Surv(START, END, EVENT) ~ GROUP + cluster(ID),
    data = MULTI)


       coef exp(coef) se(coef) robust se    z     p
GROUPA 1.84      6.33     1.12     0.909 2.03 0.042

Likelihood ratio test=3.51  on 1 df, p=0.0609  n= 9, number of events= 6
```

PWP-CP モデルを用いた解析（薬剤 A のみ抽出）

- 1 回目のイベントと 2 回目のイベントの起こりやすさ
  （正確にはベースラインハザード関数）は異なると仮定する
- よってリスク集合は「1 回目のイベントが起こるまでの層」
  「2 回目のイベントが起こるまでの層」
  「3 回目のイベントが起こるまでの層」に分けて定義される
- 使用する変数：ID，GROUP，EVENT，START，END，STRATUM

ID	GROUP	EVENT	START	END	DIFF	STRATUM
1	A	1	0	1	1	1
1	A	1	1	2	1	2
1	A	0	2	4	2	3
2	A	1	0	6	6	1
2	A	1	6	8	2	2

PWP-CP モデルを用いた解析（薬剤 A のみ抽出）



PWP-CP モデルを用いた解析

- coef：ハザード比が exp(1.07) = 2.92 倍
- se(coef)：同一 ID のデータの相関を考慮していない標準誤差（見てはダメ）
- robust se：同一 ID のデータの相関を考慮したロバストな標準誤差は 1.02

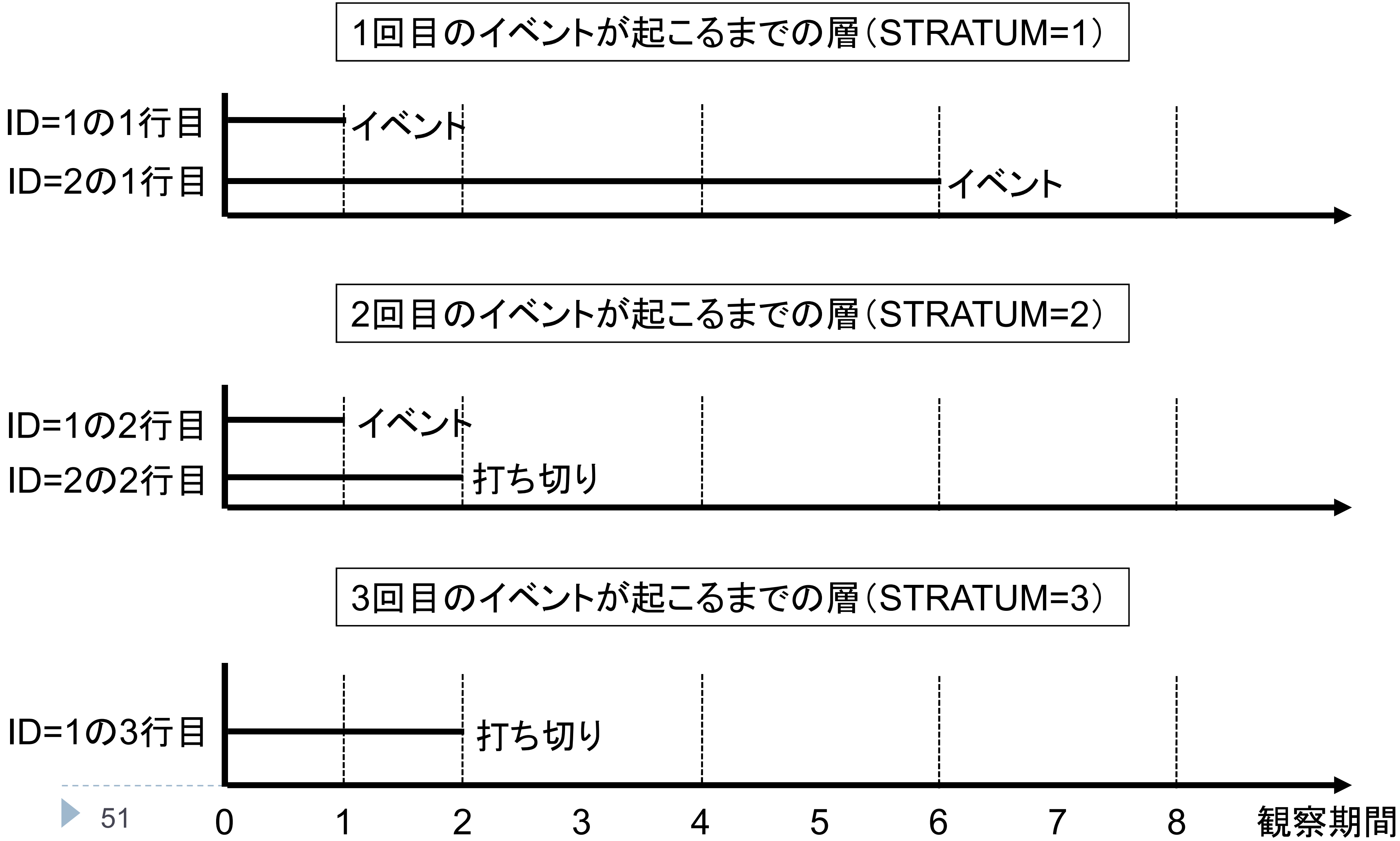
```
> coxph(Surv(START,END,EVENT) ~ GROUP + cluster(ID) + strata(STRATUM), data=MULTI)
Call:
coxph(formula = Surv(START, END, EVENT) ~ GROUP + cluster(ID) +
    strata(STRATUM), data = MULTI)


       coef exp(coef) se(coef) robust se    z    p
GROUPA 1.07      2.92     1.24      1.02 1.05 0.29

Likelihood ratio test=0.81  on 1 df, p=0.368  n= 9, number of events= 6
```

PWP-GT モデルを用いた解析（薬剤 A のみ抽出）

- 1 回目のイベントと 2 回目のイベントの起こりやすさ
  （正確にはベースラインハザード関数）は異なると仮定する
- PWP-CP モデルと考え方は同じだが，観察開始時の定義が異なる
  ⇒ 観察期間は観察開始時から観察終了時までの期間とするが，
  その後，全ての行について観察開始時を 0 年に揃える
- 使用する変数：ID，GROUP，EVENT，DIFF，STRATUM

ID	GROUP	EVENT	START	END	DIFF	STRATUM
1	A	1	0	1	1	1
1	A	1	1	2	1	2
1	A	0	2	4	2	3
2	A	1	0	6	6	1
2	A	1	6	8	2	2

PWP-GT モデルを用いた解析（薬剤 A のみ抽出）



PWP-GT モデルを用いた解析

- coef：ハザード比が exp(1.53) = 4.62 倍
- se(coef)：同一 ID のデータの相関を考慮していない標準誤差（見てはダメ）
- robust se：同一 ID のデータの相関を考慮したロバストな標準誤差は 0.721

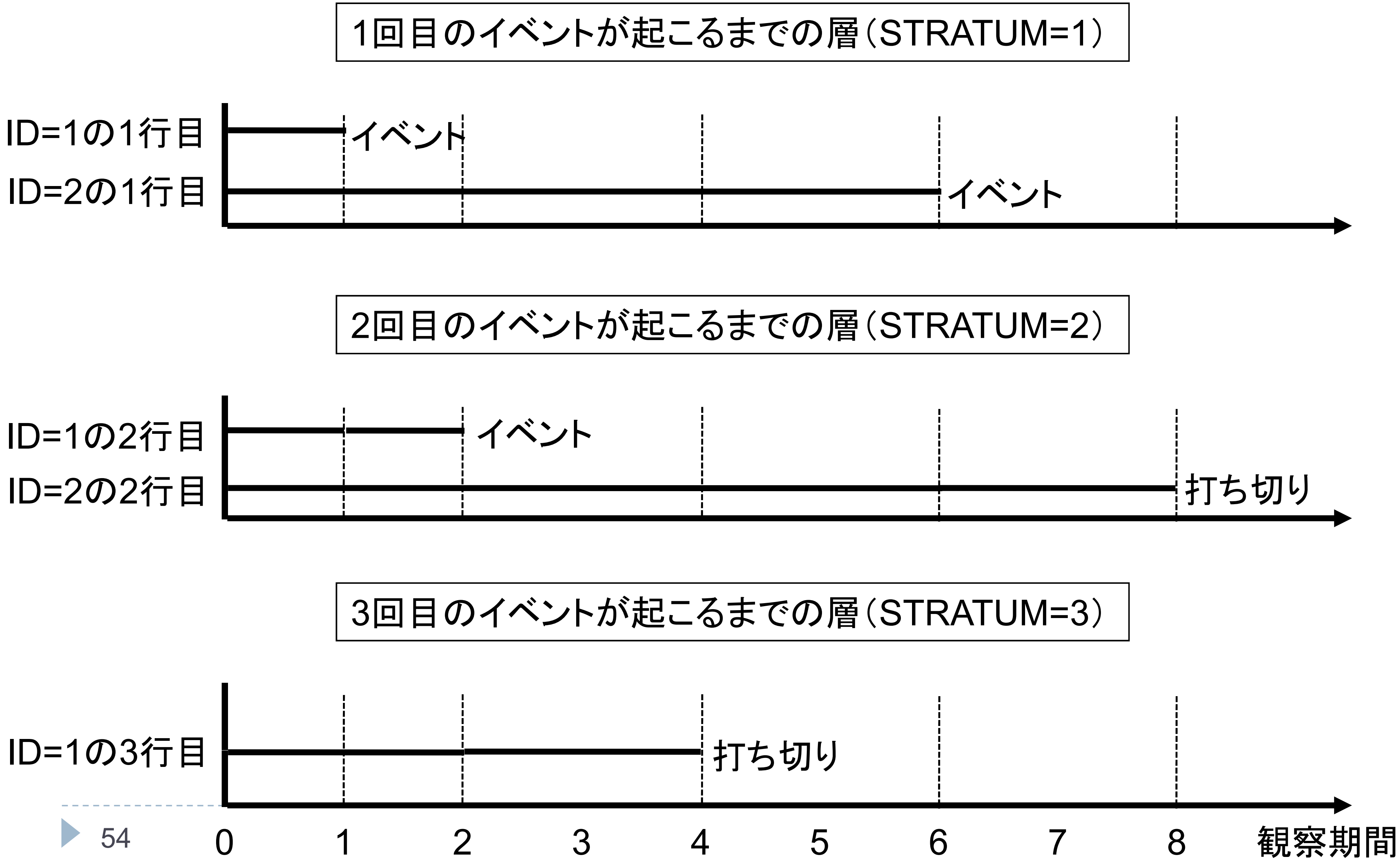
```
> coxph(Surv(DIFF,EVENT) ~ GROUP + cluster(ID) + strata(STRATUM), data=MULTI)
Call:
coxph(formula = Surv(DIFF, EVENT) ~ GROUP + cluster(ID) + strata(STRATUM),
    data = MULTI)


       coef exp(coef) se(coef) robust se    z     p
GROUPA 1.53      4.62     1.14     0.721 2.12 0.034

Likelihood ratio test=2.28  on 1 df, p=0.131  n= 9, number of events= 6
```

TT-R モデルを用いた解析（薬剤 A のみ抽出）

- 1 回目のイベントと 2 回目のイベントの起こりやすさ
  （正確にはベースラインハザード関数）は異なると仮定する
- PWP-GT モデルと考え方は同じだが，観察開始時の定義が異なる
  ⇒ 何回目のイベントでも観察開始時は 0 年とし，観察期間は 0 年から
  　観察終了時までの期間とする
- 使用する変数：ID，GROUP，EVENT，END，STRATUM

ID	GROUP	EVENT	START	END	DIFF	STRATUM
1	A	1	0	1	1	1
1	A	1	1	2	1	2
1	A	0	2	4	2	3
2	A	1	0	6	6	1
2	A	1	6	8	2	2

TT-R モデルを用いた解析（薬剤 A のみ抽出）



TT-R モデルを用いた解析

- coef：ハザード比が 1.27 = 3.55 倍
- se(coef)：同一 ID のデータの相関を考慮していない標準誤差（見てはダメ）
- robust se：同一 ID のデータの相関を考慮したロバストな標準誤差は 0.919

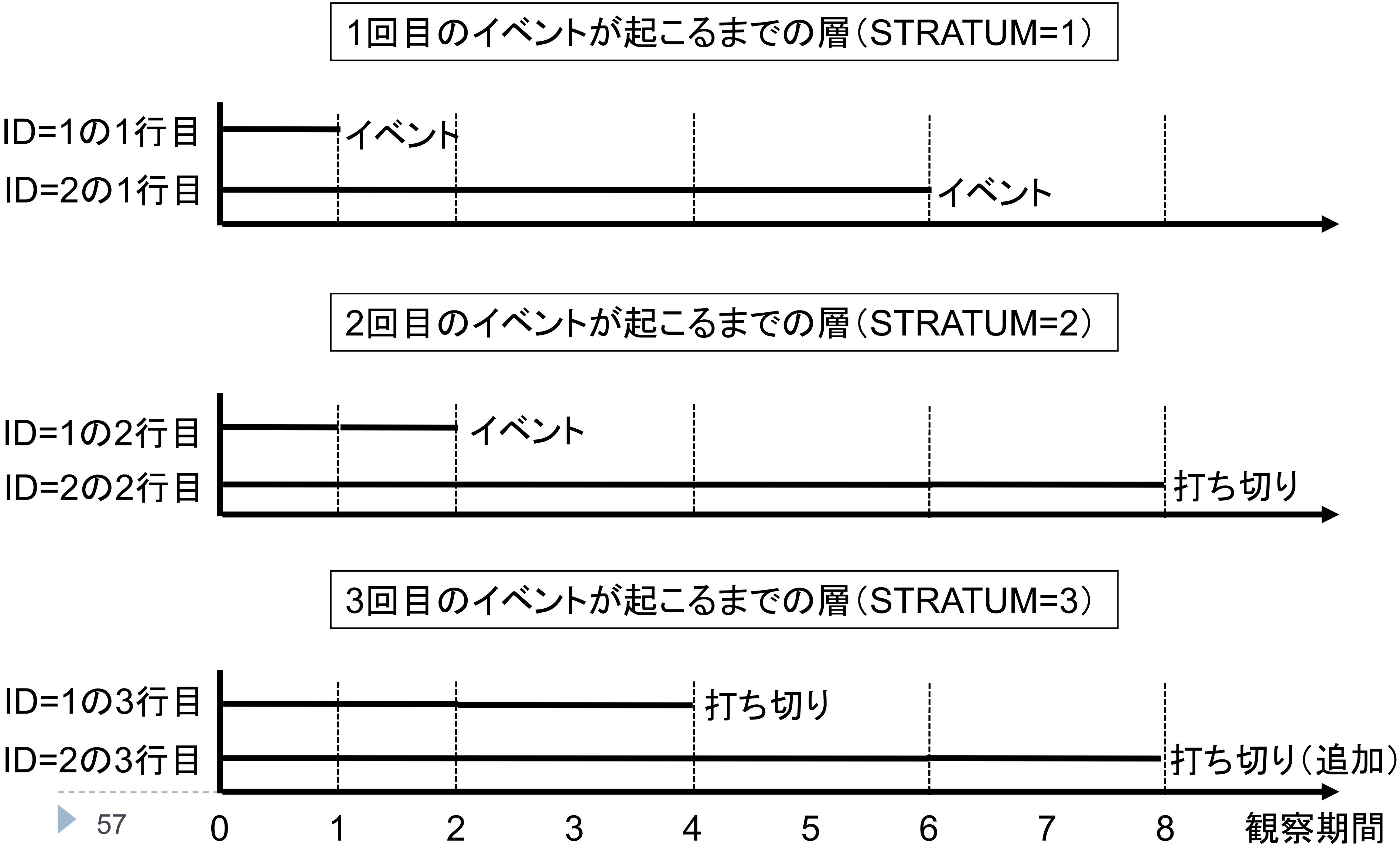
```
> coxph(Surv(END,EVENT) ~ GROUP + cluster(ID) + strata(STRATUM), data=MULTI)
Call:
coxph(formula = Surv(END, EVENT) ~ GROUP + cluster(ID) + strata(STRATUM),
    data = MULTI)


       coef exp(coef) se(coef) robust se    z    p
GROUPA 1.27      3.55     1.18     0.919 1.38 0.17

Likelihood ratio test=1.33  on 1 df, p=0.248  n= 9, number of events= 6
```

WLW モデルを用いた解析（薬剤 A のみ抽出）

- まず，全ての患者さんの中で最大の行数を求める（ここでは 3 となる）
- 次に，打ち切りの行を補完することで，全患者さんについて 3 行分のデータを起こす（以下の例では 6 行目のデータを追加することになる）
  ⇒ 要は，どの層にも全ての患者さんがリスク集合に含まれる様にする
- 使用する変数：ID，GROUP，EVENT，END，STRATUM

ID	GROUP	EVENT	START	END	DIFF	STRATUM
1	A	1	0	1	1	1
1	A	1	1	2	1	2
1	A	0	2	4	2	3
2	A	1	0	6	6	1
2	A	1	6	8	2	2
2	A	2		8		3

WLW モデルを用いた解析（薬剤 A のみ抽出）



WLW モデルを用いた解析

- coef：ハザード比が 5.52 倍
- se(coef)：同一 ID のデータの相関を考慮していない標準誤差（見てはダメ）
- robust se：同一 ID のデータの相関を考慮したロバストな標準誤差は 1.04

```
> MULTI2 <- data.frame(ID     =c(1,1,1,2,2,2,3,3,3,4,4,4,5,5,5),
+                      GROUP  =c(rep("A",6), rep("B",9)),
+                      EVENT  =c(1,1,0,1,1,0,0,0,0,1,0,0,1,0,0),
+                      END    =c(1,2,4,6,8,8,3,3,3,5,7,7,9,9,9),
+                      STRATUM=c(1,2,3,1,2,3,1,2,3,1,2,3,1,2,3) )
> MULTI2$GROUP <- relevel(MULTI2$GROUP, ref="B")   # カテゴリのベースを「B」に変更
> coxph(Surv(END,EVENT) ~ GROUP + cluster(ID) + strata(STRATUM), data=MULTI2)
Call:
coxph(formula=Surv(END, EVENT) ~ GROUP + cluster(ID) + strata(STRATUM), data=MULTI2)

       coef exp(coef) se(coef) robust se    z   p
GROUPA 1.71      5.52     1.12      1.04 1.64 0.1

Likelihood ratio test=2.97  on 1 df, p=0.0847  n= 15, number of events= 6
```

【参考】ポアソン回帰

- まず，データを「薬剤の種類」「イベント数」「総観察期間」のみに縮約する

ID	GROUP	EVENT	START	END	DIFF	STRATUM
1	A	1	0	1	1	1
1	A	1	1	2	1	2
1	A	0	2	4	2	3
2	A	1	0	6	6	1
2	A	1	6	8	2	2
3	B	0	0	3	3	1
4	B	1	0	5	5	1
4	B	0	5	7	2	2
5	B	1	0	9	9	1

ID	GROUP	COUNT	TIME
1	A	2	4
2	A	2	8
3	B	0	3
4	B	1	7
5	B	1	9

【参考】ポアソン回帰

- ポアソン回帰は以下のようなモデルについて分析を行う
  log(イベント数) ＝ log(観察期間) ＋β_0 ＋ β_1×薬剤　（β_0：切片）
- ポアソン回帰による薬剤 B に対する薬剤 A のイベント発生率の比は
  exp(1.1527) ＝ 3.167 となる

```
> POISSON <- data.frame(ID    =c(1,2,3,4,5),
+                       GROUP =c(rep("A",2), rep("B",3)),
+                       COUNT =c(2,2,0,1,1),
+                       TIME  =c(4,8,3,7,9) )
> POISSON$GROUP <- relevel(POISSON$GROUP, ref="B")   # カテゴリのベースを「B」に変更
> result <- glm(COUNT ~ GROUP, offset=log(TIME), family=poisson(log), data=POISSON)
> summary(result)
Coefficients:
            Estimate Std. Error z value Pr(>|z|)
(Intercept)  -2.2513     0.7071  -3.184  0.00145 **
GROUPA        1.1527     0.8660   1.331  0.18319
```

【参考】人年法によるイベント発現率の算出

- ちなみに，イベント発生率を「人年法」を使って計算すると
  - 薬剤 A のイベント発生率 ＝（2＋2） ÷（4＋8） ＝ 0.333
  - 薬剤 B のイベント発生率 ＝（0＋1＋1）÷（3＋7＋9）＝ 0.105
  - 薬剤 B に対する薬剤 A のイベント発生率の比 ＝ 0.333÷0.105 ＝ 3.171

となり，ポアソン回帰による薬剤 B に対する薬剤 A のイベント発生率の比と近い値となる

本日のメニュー

1. 競合リスクに関する解析
2. 再発事象の解析

参考文献

- 統計学（白旗 慎吾 著，ミネルヴァ書房）
- ロスマンの疫学（Kenneth J. Rothman 著，矢野 栄二 他翻訳，篠原出版新社）
- Competing Risks: A Practical Perspective（Melania Pintilie, Wiley）
- Applied Survival Analysis （Hosmer & Lemeshow, Wiley）
- The R Tips 第 2 版（オーム社）
- R 流！イメージで理解する統計処理入門（カットシステム）

Rで統計解析入門

終