

# **Time-dependent ROC Curve Analysis: Revision and Comparison**

國立中央大學統計研究所

碩士論文

指導教授：曾議寬 博士

研究生：侯昱宇

# 大綱

1

緒論

2

統計方法

3

模擬研究

4

資料分析

5

結論



# 緒論

# 緒論

## ROC 曲線：

- 受試者特徵曲線(Receiver Operating Characteristic curve，簡稱 ROC 曲線)，常用於醫學上，用於評估生物指標對疾病的區分能力。
- 傳統的定義方式，僅適用於二元分佈的資料，為了推廣至存活資料的應用，本研究採用 Heagerty. & Zheng. (2005) 提出的時間相依敏感度與特異度，進而得到時間相依ROC曲線。

# 符號定義

- $T_i$  : 第  $i$  位受試者存活時間
- $X_i$  : 第  $i$  位受試者共變量資料
- $M_i = X_i^T \beta$  : 第  $i$  位受試者模型分數
- $C_i$  : 第  $i$  位受試者設限時間，與  $T_i$  獨立
- $V_i = \min(T_i, C_i)$  : 第  $i$  位受試者觀測時間
- $\Delta_i = I(T_i \leq C_i)$  : 設限指標
- $R_i(t) = I(V_i \geq t)$  : 風險指標
- $N_i^*(t) = I(T_i \leq t)$  : 計數函數
- $dN_i^*(t) = N_i^*(t) - N_i^*(t-)$  : 增量函數

## 時間相依敏感度與特異度( $\mathbb{C} / \mathbb{D}$ )

累積敏感度 (Cumulative sensitivity) 與動態特異度 (Dynamic specificity) 定義：

- $Sensitivity^{\mathbb{C}}(c, t) : P(M_i > c | T_i \leq t) = P(M_i > c | N_i^*(t) = 1) = TP_t^{\mathbb{C}}(c)$
- $Specificity^{\mathbb{D}}(c, t) : P(M_i \leq c | T_i > t) = 1 - P(M_i > c | N_i^*(t) = 0) = 1 - FP_t^{\mathbb{D}}(c)$
- $ROC = \{(FP_t^{\mathbb{D}}(c), TP_t^{\mathbb{C}}(c)), c \in (-\infty, \infty)\}$
- 對特定時間點感興趣
- 常用於臨床相關應用
- 無法推導出一致性指標(Concordance)

## 時間相依敏感度與特異度( $\mathbb{C} / \mathbb{D}$ )

Heagerty 利用生存函數的 Kaplan-Meier 估計量，來估計時間相依敏感度與特異度：

- $Sensitivity^{\mathbb{C}}(c, t) : \widehat{Se}(c, t) = \frac{\{1 - \hat{S}(t|M_i > c)\}(1 - \hat{F}_M(c))}{1 - \hat{S}(t)}$
- $Specificity^{\mathbb{D}}(c, t) : \widehat{Sp}(c, t) = \frac{\hat{S}(t|M_i \leq c)\hat{F}_M(c)}{\hat{S}(t)}$

其中  $\hat{S}(t)$  為存活函數的估計， $\hat{S}(t|M_i > c)$  為條件存活函數的估計， $\hat{F}_M(c)$  為模型分數  $M$  的累積經驗分配函數

- 計算簡單
- 估計結果較不robust

## 時間相依敏感度與特異度( $\mathbb{C} / \mathbb{D}$ )

Heagerty 又提出另外一種方法，加權Kaplan-Meier 估計式，估計時間相依敏感度與特異度  
加權Kaplan-Meier 估計式如下：

$$\bullet \quad \hat{S}_{\lambda_n}(t|M_i) = \prod_{a \in T_n, a \leq t} \left( 1 - \frac{\sum_j K_{\lambda_n}(M_j, M_i) I(Z_j = a) \delta_j}{\sum_j K_{\lambda_n}(M_j, M_i) I(Z_j \geq a)} \right)$$

其中  $K_{\lambda_n}(M_j, M_i) = I(-\lambda_n < \hat{F}_M(M_i) - \hat{F}_M(M_j) < \lambda_n) : 0/1$  最鄰近點核函數(nearest neighbourhood kernel)， $\lambda_n$ : 平滑參數， $Z_j = \min(T_j, C_j)$ ， $\delta_j = I(Y_j = Z_j)$

- 可以在 C 與 M 有相關的情況下使用
- 估計結果較優



## 時間相依敏感度與特異度( $\mathbb{I}/\mathbb{D}$ )

事件敏感度 (incident sensitivity) 與動態特異度 (dynamic specificity) 定義：

- $Sensitivity^{\mathbb{I}}(c, t) : P(M_i > c | \mathbf{T}_i = \mathbf{t}) = P(M_i > c | dN_i^*(t) = 1) = TP_t^{\mathbb{I}}(c)$
- $Specificity^{\mathbb{D}}(c, t) : P(M_i \leq c | T_i > t) = 1 - P(M_i > c | N_i^*(t) = 0) = 1 - FP_t^{\mathbb{D}}(c)$
- $ROC = \{(FP_t^{\mathbb{D}}(c), TP_t^{\mathbb{I}}(c)), c \in (-\infty, \infty)\}$
- 用於事件發生時間點  $t$  為已知的資料
- 定義在風險集合下
- 可以推導出一致性指標 (Concordance)

則時間相依ROC曲線與時間相依AUC可以表示為：

- $ROC_t^{\mathbb{I}/\mathbb{D}}(p) = TP_t^{\mathbb{I}}\{[FP_t^{\mathbb{D}}]^{-1}(p)\} , p \in (0,1)$
- $AUC(t) = \int_0^1 ROC_t^{\mathbb{I}/\mathbb{D}}(p) dp = P(M_j > M_k | T_j = t, T_k > t)$

在此定義下，可以推導出**一致性指標**(concordance, Heagerty & Zheng -2005)，用來**衡量整體時間下的AUC**：

假設 $(M_j, T_j), (M_k, T_k)$ 彼此獨立

$$\begin{aligned}
 C &= P(M_j > M_k | T_j < T_k) = 2P(M_j > M_k, T_j < T_k) \\
 &= 2 \int_t P(\{M_j > M_k\} \cap \{T_j = t\} \cap \{t < T_k\}) dt \\
 &= 2 \int_t P(M_j > M_k | \{T_j = t\} \cap \{t < T_k\}) \times P(\{T_j = t\} \cap \{t < T_k\}) dt \\
 &= \int_t AUC(t) \times w(t) dt = E_T(AUC(T) \times 2 \times S(T))
 \end{aligned}$$

其中  $w(t) = 2f(t)S(t), S(t) = P(T > t)$

當估計的範圍為一段時間  $(0, \tau)$  時，Concordance 的估計可對權重做修正改寫為：

$$C^\tau = \int_0^\tau AUC(t) \times w^\tau(t) dt = E_{\{T < \tau\}} \left( \frac{AUC(t) \times 2 \times S(T)}{W^\tau} \right)$$

$$\text{其中 } w^\tau(t) = \frac{2f(t)S(T)}{W^\tau}, \quad W^\tau = \int_0^\tau 2f(t)S(t)dt = 1 - S^2(\tau)$$



## 統計方法

# 比例風險模型

比例風險(Proportional Hazard, 簡稱 PH)模型, 又稱為 Cox (1972) 模型, 模型如下:

$$\lambda(t|M_i) = \lambda_0(t)\exp(M_i)$$

則時間相依敏感度可表示為:

$$TP_t^{\mathbb{I}}(c) = P(M > c | T = t) = \int_c^{\infty} f(m | T = t) dm$$

經推導, 可得給定事件發生時間 $t$ 下,  $M$ 的條件機率為:

$$f_M(m | T = t) = \frac{\lambda(t | m) f(m | T \geq t)}{\int \lambda(t | m) f(m | T \geq t) dm} \quad (1.1)$$

# 比例風險模型

給定存活時間 $T \geq t$ 下， $M$ 的條件機率可透過一致估計獲得：

$$\hat{P}(M > m \mid T \geq t) = \sum_{M_i > m} \frac{R_i(t)}{\sum_{i=1}^n R_i(t)}$$

則時間相依敏感度 $TP_t^{\mathbb{I}}(c)$ 的估計為：

$$\widehat{TP}_t^{\mathbb{I}}(c) = \hat{P}(M > c \mid T = t) = \sum_{M_i > c} \frac{\exp(M_i)R_i(t)}{\sum_{i=1}^n \exp(M_i)R_i(t)}$$

$FP_t^{\mathbb{D}}(c)$ 以經驗估計法估計：

$$\widehat{FP}_t^{\mathbb{D}}(c) = \hat{P}(M > c \mid T > t) = \sum_{M_i > c} \frac{R_i(t+)}{\sum_{i=1}^n R_i(t+)}$$

# 放寬比例風險模型

Heagerty & Zheng (2007) 於論文中提到，當比例風險不滿足，可使用放寬假設的時變係數模型(varying-coefficient)模型，模型如下：

$$\lambda(t|M_i) = \lambda_0(t)\exp(M_i\gamma(t))$$

$\gamma(t)$ 為時變係數函數，可由Grambsch & Therneau(1994)提出的平滑殘差估計法(smoothly residual-based methods)估計，則  $TP_t^{\mathbb{I}}(c)$  的估計可表示為：

$$\widehat{TP}_t^{\mathbb{I}}(c) = \hat{P}(M > c | T = t) = \sum_{M_i > c} \frac{\exp(M_i \hat{\gamma}(t)) R_i(t)}{\sum_{i=1}^n \exp(M_i \hat{\gamma}(t)) R_i(t)}$$

# 加速失效模型

加速失效(Accelerated failure time，簡稱 AFT )模型，風險模型如下：

$$\lambda(t|M_i) = \lambda_0(t \exp(M_i))\exp(M_i)$$

透過(1.1)式，僅需將風險函數以AFT模型代入，即可獲得 $TP_t^{\mathbb{I}}(c)$ 的估計：

$$\widehat{TP}_t^{\mathbb{I}}(c) = \hat{P}(M > c | T = t) = \sum_{M_i > c} \frac{\lambda_0(u_i) \exp(M_i) R_i(t)}{\sum_{i=1}^n \lambda_0(u_i) \exp(M_i) R_i(t)}$$

其中 $u_i = t \cdot \exp(M_i)$



# 加速失效模型

基線風險函數 $\lambda_0(u_i)$ ，使用Zeng & Lin (2007) 提出的核平滑(kernel smooth)估計式：

$$\hat{\lambda}_0(u) = \frac{(nu)^{-1} \sum_{i=1}^n \Delta_i K_{h_1}(L_i - \log u)}{n^{-1} \sum_{i=1}^n \int_{-\infty}^{L_i - \log u} K_{h_2}(s) ds}$$

- $L_i = \log V_i + M_i$
- $K_h(s) = \frac{1}{h} K(\frac{s}{h})$ ： $K$  為核函數，此處採用標準常態的機率密度函數
- 估計式中分子分母分別使用兩種最佳帶寬 (Jones,1990；Jones & Sheather,1991)，分別為  $h_1$  與  $h_2$

$h_1 = ((8\sqrt{2})/3)^{1/5} \sigma_1 n^{-1/5}$ ： $\sigma_1$  為未設限個體對應  $L_i$  的標準差

$h_2 = 4^{1/3} \sigma_2 n^{-1/3}$ ： $\sigma_2$  為所有個體對應  $L_i$  的標準差

# 比例勝算模型

比例勝算(Proportional Odds，簡稱 PO)模型假設如下：

$$\log\left(\frac{P(T \leq t | X)}{1 - P(T \leq t | X)}\right) = \log\left(\frac{P(T \leq t)}{1 - P(T \leq t)}\right) + X^T \beta$$

亦可改寫為：

$$\log\left(\frac{1 - S_X(t)}{S_X(t)}\right) = \log(G(t)) + X^T \beta$$

其中  $G(t) = \frac{1 - S(t|X=0)}{S(t|X=0)}$  為基線存活時間勝算，則風險模型可表示為：

$$\lambda(t | X) = \frac{g(t)}{\exp(-X^T \beta) + G(t)}$$

# 比例勝算模型

根據過去文獻， $G(t)$ 可以透過 Breslow 估計式  $\tilde{G}(t, \beta)$ ，在 $\beta$ 固定下估計：

$$\tilde{G}(t, \beta) = \int_0^t \frac{1}{S_0(s, \beta, \tilde{G})} dN^*(s)$$

其中  $S_0(t, \beta, G) = \sum_{i=1}^n R_i(t) \exp(X^T \beta) \lambda_0(\exp(X^T \beta) G(t-))$ ， $\lambda_0(t) = \frac{1}{1+t}$

$dN^*(t) = S_0(t, \beta, G)dG(t) + dM_i(t)$ ， $M_i(t)$ 為平賭計數過程(Counting process martingales)

# 比例勝算模型

$\beta$  的估計使用部分概似函數：

$$L(\beta) = \prod_{i=1}^n \prod_{t \geq 0} \left[ R_i(t) \exp(X_i^T \beta) dG(t) \lambda_0 \left( \exp(X_i^T \beta) G(t-) \right) \right]^{\Delta_i} \\ \times \exp \left\{ - \int_0^\infty R_i(t) \exp(X_i^T \beta) \lambda_0 \left( \exp(X_i^T \beta) G(t-) \right) dG(t) \right\}$$

其中  $G(t)$  與  $dG(t)$  由  $\tilde{G}(t, \beta)$  與  $d\tilde{G}(t, \beta)$  代入，對  $\beta$  微分並整理可得：

$$\tilde{U}(\beta) = \sum_{i=1}^n \int_0^\infty \left\{ \frac{w'_i(t, \beta, \tilde{G})}{w_i(t, \beta, \tilde{G})} - \frac{S'_0(t, \beta, \tilde{G})}{S_0(t, \beta, \tilde{G})} \right\} dN_i(t) = 0$$

其中  $w_i(t, \beta, \tilde{G}) = \exp(X_i^T \beta) \lambda_0(\exp(X_i^T \beta) \tilde{G}(t-, \beta))$ ，

$$w'_i(t, \beta, \tilde{G}) = \frac{\partial}{\partial \beta} w_i(t, \beta, \tilde{G})，S'_0(t, \beta, \tilde{G}) = \frac{\partial}{\partial \beta} S_0(t, \beta, \tilde{G})$$

# 比例勝算模型

估計得到  $\hat{\beta}$  與  $\hat{G}(t)$  後，根據 (1.1) 式

$$f_M(m | T = t) = \frac{\lambda(t | m) f(m | T \geq t)}{\int \lambda(t | m) f(m | T \geq t) dm} \quad (1.1)$$

將風險函數以 PO 模型替換，則  $TP_t^{\mathbb{I}}(c)$  可被估計為：

$$\widehat{TP}_t^{\mathbb{I}}(c) = \hat{P}(M > c | T = t) = \sum_{M_i > c} \frac{(\exp(-M_i) + \hat{G}(t))^{-1} R_i(t)}{\sum_{i=1}^n (\exp(-M_i) + \hat{G}(t))^{-1} R_i(t)}$$

表 2.1：各種模型、方法下  $\widehat{TP}_t(c)$  形式

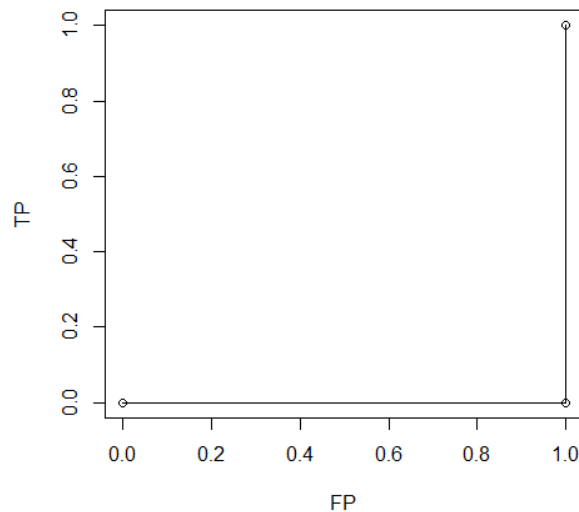
Model/Method	$\widehat{TP}_t(c)$
Cox	$\sum_{M_i > c} \frac{e^{M_i R_i(t)}}{\sum_{i=1}^n e^{M_i R_i(t)}}$
Residual smooth	$\sum_{M_i > c} \frac{e^{M_i \hat{\gamma}(t) R_i(t)}}{\sum_{i=1}^n e^{M_i \hat{\gamma}(t) R_i(t)}}$
AFT	$\sum_{M_i > c} \frac{\lambda_0(u_i) e^{M_i R_i(\psi_1^{-1}(u_i; \beta))}}{\sum_{i=1}^n \lambda_0(u_i) e^{M_i R_i(\psi_1^{-1}(u_i; \beta))}}$
PO	$\sum_{M_i > c} \frac{(e^{-M_i + G(t)})^{-1} R_i(t)}{\sum_{i=1}^n (e^{-M_i + G(t)})^{-1} R_i(t)}$
Kaplan-Meier (CD1)	$\frac{\{1 - \hat{S}(t   M_i > c)\} (1 - \hat{F}_M(c))}{1 - \hat{S}(t)}$
Nearest neighbour (CD2)	$\frac{(1 - \hat{F}_M(c)) - \hat{S}_{\lambda_n}(c, t)}{1 - \hat{S}_{\lambda_n}(t)}$



## 模擬研究

# 修正計算方式動機

Log time	AUC(t)	EST	SD
- 1.00	0.840	<b>0.845</b>	0.040
0.00	0.817	<b>0.820</b>	0.026
1.00	0.785	<b>0.786</b>	0.024
2.00	0.751	<b>0.750</b>	0.026
3.00	0.720	<b>0.720</b>	0.047
3.50	0.707	<b>0.710</b>	0.104
4.00	0.695	<b>0.572</b>	0.344
4.25	0.690	<b>0.409</b>	0.397
4.50	0.684	<b>0.324</b>	0.399



$$FP_t^{\mathbb{D}}(c) = P(M_i > c | T_i > t)$$

$$TP_t^{\mathbb{I}}(c) = P(M_i > c | T_i = t)$$



# 模擬步驟

1. 生成樣本數為  $n$  的隨機變數  $X \sim N(0,1)$ ,  $C \sim \text{Exp}(\lambda_c)$ ,  $U \sim \text{Unif}(0,1)$
2. 存活時間  $T$  根據風險模型假設以逆轉換法求得
3. 觀測時間  $V_i = \min(T_i, C_i)$  , 設限指標  $\Delta_i = I\{T_i \leq C_i\}$
4. 依據不同風險模型估計  $TP_t(c)$ 、 $FP_t(c)$ 、 $AUC$  及 Concordance
5. 計算標準差
6. 以上步驟重複 500 次

# Cox 模型模擬

本模擬比較新舊方法在不同設限率下，使用 Cox 模型估計時間相依 AUC 與 Concordance 的影響  
模擬設定如下：

- 存活時間來自 Lognormal 分配，樣本數為 200
- 參數設定為  $\mu = 0.5$ ， $\sigma = 2.5$ ， $\beta = 1.0$
- 風險模型為  $\lambda(t | X) = \frac{1}{\sigma t} \frac{\phi\left(\frac{\ln t - \mu}{\sigma}\right)}{1 - \Phi\left(\frac{\ln t - \mu}{\sigma}\right)} e^{\beta X}$
- 存活時間  $t = \exp\{\sigma \Phi^{-1}(1 - U^{\exp(-\beta X)}) + \mu\}$ ， $U \sim U(0, 1)$
- 設限時間參數  $\lambda_c$  在設限率 20%、40% 下分別為 0.01、0.18

# Lognormal Cox 模型模擬結果

20 % censoring (n = 200)						40 % censoring (n = 200)					
Log	Target value	original		revision		Log	Target value	original		revision	
Time	AUC(t)	EST	SD	EST	SD	Time	AUC(t)	EST	SD	EST	SD
- 2.50	0.759	<b>0.759</b>	0.026	<b>0.759</b>	0.026	- 2.50	0.759	<b>0.759</b>	0.028	<b>0.759</b>	0.028
- 2.00	0.754	<b>0.755</b>	0.026	<b>0.755</b>	0.026	- 2.00	0.754	<b>0.755</b>	0.027	<b>0.755</b>	0.027
- 1.50	0.746	<b>0.747</b>	0.024	<b>0.747</b>	0.024	- 1.50	0.746	<b>0.747</b>	0.025	<b>0.747</b>	0.025
- 1.00	0.734	<b>0.735</b>	0.022	<b>0.735</b>	0.022	- 1.00	0.734	<b>0.735</b>	0.022	<b>0.735</b>	0.022
- 0.50	0.721	<b>0.721</b>	0.021	<b>0.721</b>	0.021	- 0.50	0.721	<b>0.721</b>	0.023	<b>0.721</b>	0.023
0.00	0.709	<b>0.709</b>	0.022	<b>0.709</b>	0.022	0.00	0.709	<b>0.708</b>	0.023	<b>0.709</b>	0.023
0.50	0.698	<b>0.697</b>	0.023	<b>0.697</b>	0.023	0.50	0.698	<b>0.697</b>	0.025	<b>0.697</b>	0.025
1.00	0.689	<b>0.688</b>	0.026	<b>0.688</b>	0.027	1.00	0.689	<b>0.687</b>	0.034	<b>0.688</b>	0.034
1.50	0.681	<b>0.682</b>	0.036	<b>0.684</b>	0.036	1.50	0.681	<b>0.681</b>	0.066	<b>0.688</b>	0.063
2.00	0.675	<b>0.671</b>	0.069	<b>0.679</b>	0.065	2.00	0.675	<b>0.627</b>	0.219	<b>0.686</b>	0.119
2.50	0.670	<b>0.640</b>	0.191	<b>0.674</b>	0.113	2.50	0.670	<b>0.549</b>	0.302	<b>0.667</b>	0.135
3.00	0.665	<b>0.531</b>	0.323	<b>0.669</b>	0.141	3.00	0.665	<b>0.531</b>	0.311	<b>0.661</b>	0.135
C	0.726	<b>0.726</b>	0.019	<b>0.726</b>	0.019	C	0.726	<b>0.726</b>	0.020	<b>0.727</b>	0.020

# Cox 模型模擬

本模擬比較新舊方法在不同設限率下，使用 Cox 模型估計時間相依 AUC 與 Concordance 的影響  
模擬設定如下：

- 存活時間來自 Weibull 分配，樣本數為 200
- 參數設定為  $\alpha = 1.5$ ， $\lambda = 0.25$ ， $\beta = 1.0$
- 風險模型為  $\lambda(t | X) = \alpha \lambda t^{\alpha-1} e^{\beta x}$
- 存活時間  $t = (-\frac{1}{\lambda e^{\beta x}} \ln U)^{\frac{1}{\alpha}}$ ， $U \sim U(0, 1)$
- 設限時間參數  $\lambda_c$  在設限率 20%、40% 下分別為 0.007、0.027

# Weibull Cox 模型模擬結果

20 % censoring (n = 200)						40 % censoring (n = 200)					
Log	Target value	original		revision		Log	Target value	original		revision	
Time	AUC(t)	EST	SD	EST	SD	Time	AUC(t)	EST	SD	EST	SD
- 2.50	0.759	<b>0.758</b>	0.027	<b>0.758</b>	0.027	- 2.50	0.759	<b>0.759</b>	0.029	<b>0.759</b>	0.029
- 2.00	0.757	<b>0.757</b>	0.026	<b>0.757</b>	0.026	- 2.00	0.757	<b>0.757</b>	0.028	<b>0.757</b>	0.028
- 1.50	0.754	<b>0.753</b>	0.024	<b>0.753</b>	0.024	- 1.50	0.754	<b>0.754</b>	0.027	<b>0.754</b>	0.027
- 1.00	0.748	<b>0.748</b>	0.023	<b>0.748</b>	0.023	- 1.00	0.748	<b>0.749</b>	0.026	<b>0.749</b>	0.026
- 0.50	0.740	<b>0.740</b>	0.022	<b>0.740</b>	0.022	- 0.50	0.740	<b>0.740</b>	0.024	<b>0.740</b>	0.024
0.00	0.728	<b>0.728</b>	0.021	<b>0.728</b>	0.021	0.00	0.728	<b>0.728</b>	0.023	<b>0.728</b>	0.023
0.50	0.714	<b>0.714</b>	0.020	<b>0.714</b>	0.020	0.50	0.714	<b>0.715</b>	0.024	<b>0.715</b>	0.024
1.00	0.699	<b>0.699</b>	0.022	<b>0.700</b>	0.022	1.00	0.699	<b>0.701</b>	0.028	<b>0.702</b>	0.028
1.50	0.684	<b>0.685</b>	0.032	<b>0.686</b>	0.032	1.50	0.684	<b>0.687</b>	0.054	<b>0.691</b>	0.053
2.00	0.670	<b>0.665</b>	0.083	<b>0.676</b>	0.070	2.00	0.670	<b>0.592</b>	0.284	<b>0.687</b>	0.142
2.50	0.658	<b>0.440</b>	0.358	<b>0.642</b>	0.152	2.50	0.658	<b>0.373</b>	0.380	<b>0.627</b>	0.159
C	0.726	<b>0.726</b>	0.018	<b>0.726</b>	0.018	C	0.726	<b>0.726</b>	0.020	<b>0.727</b>	0.020

# AFT 模型模擬

本模擬比較新舊方法在不同設限率下，使用 AFT 模型估計時間相依 AUC 與 Concordance 的影響  
模擬設定如下：

- 存活時間來自 Lognormal 分配，樣本數為 200
- 參數設定為  $\mu = 1.0$ ， $\sigma = 2.5$ ， $\beta = 1.0$
- 風險模型為  $\lambda(t | X) = \frac{1}{\sigma t} \frac{\phi\left(\frac{\ln t - \beta X - \mu}{\sigma}\right)}{1 - \Phi\left(\frac{\ln t - \beta X - \mu}{\sigma}\right)} e^{\beta X}$
- 存活時間  $t = \exp\{\sigma\Phi^{-1}(1 - U) + \beta X + \mu\}$ ， $U \sim U(0, 1)$
- 設限時間參數  $\lambda_c$  在設限率 20%、40% 下分別為 0.08、0.15

# Lognormal AFT 模型模擬結果

20 % censoring (n = 200)						40 % censoring (n = 200)					
Log	Target value	original		revision		Log	Target value	original		revision	
Time	AUC(t)	EST	SD	EST	SD	Time	AUC(t)	EST	SD	EST	SD
- 5.00	0.731	<b>0.738</b>	0.093	<b>0.738</b>	0.093	- 5.00	0.731	<b>0.738</b>	0.096	<b>0.738</b>	0.096
- 4.00	0.701	<b>0.704</b>	0.063	<b>0.704</b>	0.063	- 4.00	0.701	<b>0.704</b>	0.066	<b>0.704</b>	0.066
- 3.00	0.672	<b>0.674</b>	0.050	<b>0.674</b>	0.050	- 3.00	0.672	<b>0.673</b>	0.052	<b>0.673</b>	0.052
- 2.00	0.646	<b>0.646</b>	0.040	<b>0.646</b>	0.040	- 2.00	0.646	<b>0.647</b>	0.040	<b>0.647</b>	0.040
- 1.00	0.623	<b>0.624</b>	0.035	<b>0.624</b>	0.035	- 1.00	0.623	<b>0.623</b>	0.038	<b>0.623</b>	0.038
0.00	0.604	<b>0.604</b>	0.032	<b>0.604</b>	0.033	0.00	0.604	<b>0.604</b>	0.035	<b>0.604</b>	0.035
1.00	0.588	<b>0.588</b>	0.030	<b>0.588</b>	0.030	1.00	0.588	<b>0.585</b>	0.040	<b>0.585</b>	0.040
2.00	0.575	<b>0.576</b>	0.032	<b>0.576</b>	0.032	2.00	0.575	<b>0.572</b>	0.063	<b>0.573</b>	0.063
3.00	0.564	<b>0.564</b>	0.048	<b>0.565</b>	0.048	3.00	0.564	<b>0.554</b>	0.203	<b>0.563</b>	0.162
4.00	0.556	<b>0.562</b>	0.117	<b>0.563</b>	0.114	4.00	0.556	<b>0.490</b>	0.264	<b>0.556</b>	0.167
5.00	0.549	<b>0.467</b>	0.297	<b>0.555</b>	0.174	5.00	0.549	<b>0.490</b>	0.264	<b>0.556</b>	0.167
C	0.623	<b>0.623</b>	0.024	<b>0.623</b>	0.024	C	0.623	<b>0.626</b>	0.027	<b>0.626</b>	0.027

# AFT 模型模擬

本模擬比較新舊方法在不同設限率下，使用 AFT 模型估計時間相依 AUC 與 Concordance 的影響  
模擬設定如下：

- 存活時間來自 Weibull 分配，樣本數為 200
- 參數設定為  $\alpha = 1.5$ ， $\lambda = 0.05$ ， $\beta = 1.0$
- 風險模型為  $\lambda(t | X) = \alpha \lambda t^{\alpha-1} e^{\alpha \beta x}$
- 存活時間  $t = (-\frac{1}{\lambda e^{\alpha \beta x}} \ln U)^{\frac{1}{\alpha}}$ ， $U \sim U(0, 1)$
- 設限時間參數  $\lambda_c$  在設限率 20%、40% 下分別為 0.02、0.07



## Weibull AFT 模型模擬結果

20 % censoring (n = 200)						40 % censoring (n = 200)					
Log	Target value	original		revision		Log	Target value	original		revision	
Time	AUC(t)	EST	SD	EST	SD	Time	AUC(t)	EST	SD	EST	SD
- 1.00	0.840	<b>0.845</b>	0.040	<b>0.845</b>	0.040	- 1.00	0.840	<b>0.845</b>	0.041	<b>0.845</b>	0.041
0.00	0.817	<b>0.820</b>	0.026	<b>0.820</b>	0.026	0.00	0.817	<b>0.819</b>	0.028	<b>0.819</b>	0.028
1.00	0.785	<b>0.786</b>	0.024	<b>0.786</b>	0.024	1.00	0.785	<b>0.784</b>	0.027	<b>0.784</b>	0.027
2.00	0.751	<b>0.750</b>	0.026	<b>0.750</b>	0.027	2.00	0.751	<b>0.749</b>	0.033	<b>0.749</b>	0.033
3.00	0.720	<b>0.720</b>	0.047	<b>0.722</b>	0.047	3.00	0.720	<b>0.719</b>	0.102	<b>0.731</b>	0.090
3.50	0.707	<b>0.710</b>	0.104	<b>0.721</b>	0.089	3.50	0.707	<b>0.586</b>	0.353	<b>0.718</b>	0.171
4.00	0.695	<b>0.572</b>	0.344	<b>0.704</b>	0.166	4.00	0.695	<b>0.459</b>	0.401	<b>0.674</b>	0.179
4.25	0.690	<b>0.409</b>	0.397	<b>0.649</b>	0.180	4.25	0.690	<b>0.457</b>	0.401	<b>0.673</b>	0.179
4.50	0.684	<b>0.324</b>	0.399	<b>0.624</b>	0.176	4.50	0.684	<b>0.457</b>	0.401	<b>0.673</b>	0.179
C	0.790	<b>0.788</b>	0.016	<b>0.788</b>	0.016	C	0.790	<b>0.788</b>	0.018	<b>0.788</b>	0.018

# PO 模型模擬

本模擬比較新舊方法在不同設限率下，使用 PO 模型估計時間相依 AUC 與 Concordance 的影響  
模擬設定如下：

- 存活時間來自 Lognormal 分配，樣本數為 200
- 參數設定為  $\mu = 1.0$ ， $\sigma = 0.3$ ， $\beta = 1.0$
- 風險模型為 
$$\lambda(t | X) = \frac{\phi\left(\frac{\ln t - \mu}{\sigma}\right)e^{\beta X}}{\left(1 - \Phi\left(\frac{\ln t - \mu}{\sigma}\right)\right)^2 \cdot \left(\left(1 - \Phi\left(\frac{\ln t - \mu}{\sigma}\right)\right)^{-1} - 1\right)e^{\beta X} + 1}$$
- 存活時間  $t = \exp\left\{\sigma\Phi^{-1}\left(1 - \left((U^{-1} - 1)e^{-\beta X} + 1\right)^{-1}\right) + \mu\right\}$ ， $U \sim U(0, 1)$
- 設限時間參數  $\lambda_c$  在設限率 20%、40% 下分別為 0.06、0.16

# Lognormal PO 模型模擬結果

20 % censoring (n = 200)						40 % censoring (n = 200)					
Log	Target value	original		revision		Log	Target value	original		revision	
Time	AUC(t)	EST	SD	EST	SD	Time	AUC(t)	EST	SD	EST	SD
0.00	0.760	<b>0.757</b>	0.034	<b>0.757</b>	0.034	0.00	0.760	<b>0.753</b>	0.039	<b>0.753</b>	0.039
0.25	0.755	<b>0.752</b>	0.033	<b>0.752</b>	0.033	0.25	0.755	<b>0.748</b>	0.037	<b>0.748</b>	0.037
0.50	0.730	<b>0.728</b>	0.028	<b>0.728</b>	0.028	0.50	0.730	<b>0.725</b>	0.032	<b>0.725</b>	0.032
0.75	0.683	<b>0.682</b>	0.024	<b>0.682</b>	0.024	0.75	0.683	<b>0.681</b>	0.027	<b>0.681</b>	0.027
1.00	0.631	<b>0.631</b>	0.020	<b>0.631</b>	0.020	1.00	0.631	<b>0.630</b>	0.024	<b>0.630</b>	0.024
1.25	0.579	<b>0.580</b>	0.020	<b>0.580</b>	0.020	1.25	0.579	<b>0.578</b>	0.024	<b>0.578</b>	0.024
1.50	0.533	<b>0.535</b>	0.040	<b>0.540</b>	0.038	1.50	0.533	<b>0.537</b>	0.080	<b>0.553</b>	0.056
1.75	0.506	<b>0.345</b>	0.298	<b>0.533</b>	0.110	1.75	0.506	<b>0.212</b>	0.294	<b>0.547</b>	0.100
2.00	0.501	<b>0.027</b>	0.298	<b>0.507</b>	0.040	2.00	0.501	<b>0.047</b>	0.169	<b>0.512</b>	0.055
C	0.669	<b>0.668</b>	0.022	<b>0.668</b>	0.022	C	0.669	<b>0.666</b>	0.025	<b>0.666</b>	0.025

# PO 模型模擬

本模擬比較新舊方法在不同設限率下，使用 PO 模型估計時間相依 AUC 與 Concordance 的影響  
模擬設定如下：

- 存活時間來自 Loglogistic 分配，樣本數為 200
- 參數設定為  $\mu = 0.5$ ， $\sigma = 0.2$ ， $\beta = 1.0$
- 風險模型為  $\lambda(t | X) = \frac{1}{\sigma t} \frac{\exp\{\frac{\ln t - \mu}{\sigma}\}}{(1 + \exp\{\frac{\ln t - \mu}{\sigma}\})} e^{\beta x}$
- 存活時間  $t = \exp\{\sigma[\ln(U^{-1} - 1) - \beta x] + \mu\}$ ， $U \sim U(0, 1)$
- 設限時間參數  $\lambda_c$  在設限率 20%、40% 下分別為 0.12、0.30

## Loglogistic PO 模型模擬結果

20 % censoring (n = 200)						40 % censoring (n = 200)					
Log	Target value	original		revision		Log	Target value	original		revision	
Time	AUC(t)	EST	SD	EST	SD	Time	AUC(t)	EST	SD	EST	SD
- 1.00	0.760	<b>0.755</b>	0.035	<b>0.755</b>	0.035	- 1.00	0.760	<b>0.753</b>	0.039	<b>0.753</b>	0.039
- 0.50	0.754	<b>0.751</b>	0.033	<b>0.751</b>	0.033	- 0.50	0.754	<b>0.749</b>	0.037	<b>0.749</b>	0.037
0.00	0.718	<b>0.716</b>	0.027	<b>0.716</b>	0.027	0.00	0.718	<b>0.715</b>	0.029	<b>0.715</b>	0.029
0.25	0.679	<b>0.677</b>	0.023	<b>0.677</b>	0.023	0.25	0.679	<b>0.677</b>	0.026	<b>0.677</b>	0.026
0.50	0.631	<b>0.630</b>	0.020	<b>0.630</b>	0.020	0.50	0.631	<b>0.630</b>	0.024	<b>0.630</b>	0.024
0.75	0.583	<b>0.583</b>	0.019	<b>0.583</b>	0.019	0.75	0.583	<b>0.583</b>	0.024	<b>0.584</b>	0.024
1.00	0.544	<b>0.545</b>	0.027	<b>0.570</b>	0.027	1.00	0.544	<b>0.543</b>	0.044	<b>0.569</b>	0.041
1.25	0.519	<b>0.520</b>	0.102	<b>0.565</b>	0.075	1.25	0.519	<b>0.460</b>	0.233	<b>0.563</b>	0.103
1.50	0.507	<b>0.321</b>	0.291	<b>0.562</b>	0.106	1.50	0.507	<b>0.224</b>	0.304	<b>0.555</b>	0.104
1.75	0.502	<b>0.121</b>	0.249	<b>0.530</b>	0.081	1.75	0.502	<b>0.124</b>	0.253	<b>0.531</b>	0.084
2.00	0.501	<b>0.067</b>	0.193	<b>0.516</b>	0.061	2.00	0.501	<b>0.111</b>	0.241	<b>0.528</b>	0.079
C	0.670	<b>0.667</b>	0.022	<b>0.667</b>	0.022	C	0.670	<b>0.667</b>	0.024	<b>0.667</b>	0.024

# 模擬步驟

第二部分生成 $(M_i, \log T_i)$ 來自二元常態分佈，針對不同的統計方法、模型，比較各結果之間的差異。

1. 生成樣本數為  $n$  的隨機變數  $(M_i, \log T_i) \sim BVN(0, 0, 1, 1, -0.7), C \sim \text{lognormal}(\mu, 1)$
2. 觀測時間  $V_i = \min(T_i, C_i)$ ，設限指標  $\Delta_i = I\{T_i \leq C_i\}$
3. 依據不同風險模型估計  $TP_t(c)$ 、 $FP_t(c)$ 、 $AUC$  及 Concordance
4. 計算標準差
5. 以上步驟重複 1000 次

# Bivariate normal 模擬結果

20 % censoring (n = 200)													
Log	Target value	Cox model		residual smooth		PO model		AFT model		CD1		CD2	
Time	AUC(t)	EST	SD	EST	SD	EST	SD	EST	SD	EST	SD	EST	SD
- 2.00	0.884	0.754	0.028	0.880	0.040	0.848	0.024	0.885	0.020	0.917	0.060	0.854	0.062
- 1.50	0.833	0.746	0.026	0.840	0.032	0.821	0.021	0.836	0.016	0.884	0.044	0.818	0.048
- 1.00	0.782	0.735	0.024	0.777	0.031	0.786	0.019	0.783	0.014	0.857	0.034	0.774	0.039
- 0.50	0.734	0.726	0.021	0.727	0.034	0.743	0.019	0.736	0.015	0.836	0.030	0.731	0.033
0.00	0.693	0.717	0.021	0.688	0.037	0.698	0.020	0.694	0.016	0.831	0.031	0.704	0.029
0.50	0.660	0.711	0.024	0.663	0.041	0.655	0.023	0.662	0.020	0.837	0.032	0.692	0.029
1.00	0.634	0.703	0.039	0.645	0.060	0.615	0.032	0.636	0.032	0.858	0.041	0.697	0.035
1.50	0.614	0.693	0.092	0.653	0.100	0.607	0.076	0.639	0.071	0.880	0.062	0.714	0.070
2.00	0.598	0.647	0.163	0.642	0.159	0.598	0.126	0.638	0.166	0.876	0.105	0.628	0.253
C	0.741	0.728	0.020	0.744	0.018	0.745	0.018	0.749	0.011				

# Bivariate normal 模擬結果

40 % censoring (n = 200)													
Log Time	Target value	Cox model		residual smooth		PO model		AFT model		CD1		CD2	
	AUC(t)	EST	SD	EST	SD	EST	SD	EST	SD	EST	SD	EST	SD
- 2.00	0.884	<b>0.768</b>	0.031	<b>0.881</b>	0.042	<b>0.850</b>	0.024	<b>0.885</b>	0.031	<b>0.911</b>	0.061	<b>0.885</b>	0.062
- 1.50	0.833	<b>0.759</b>	0.028	<b>0.836</b>	0.034	<b>0.823</b>	0.022	<b>0.836</b>	0.027	<b>0.888</b>	0.043	<b>0.819</b>	0.048
- 1.00	0.782	<b>0.749</b>	0.026	<b>0.773</b>	0.035	<b>0.787</b>	0.020	<b>0.784</b>	0.026	<b>0.858</b>	0.036	<b>0.773</b>	0.040
- 0.50	0.734	<b>0.738</b>	0.023	<b>0.726</b>	0.039	<b>0.744</b>	0.020	<b>0.735</b>	0.027	<b>0.837</b>	0.033	<b>0.731</b>	0.035
0.00	0.693	<b>0.729</b>	0.024	<b>0.689</b>	0.044	<b>0.700</b>	0.022	<b>0.693</b>	0.030	<b>0.830</b>	0.036	<b>0.703</b>	0.031
0.50	0.660	<b>0.721</b>	0.032	<b>0.666</b>	0.053	<b>0.657</b>	0.029	<b>0.661</b>	0.049	<b>0.837</b>	0.040	<b>0.690</b>	0.031
1.00	0.634	<b>0.710</b>	0.068	<b>0.658</b>	0.084	<b>0.625</b>	0.056	<b>0.646</b>	0.099	<b>0.857</b>	0.063	<b>0.695</b>	0.048
1.50	0.614	<b>0.683</b>	0.149	<b>0.640</b>	0.150	<b>0.621</b>	0.125	<b>0.633</b>	0.169	<b>0.857</b>	0.103	<b>0.650</b>	0.196
2.00	0.598	<b>0.626</b>	0.161	<b>0.618</b>	0.158	<b>0.590</b>	0.127	<b>0.615</b>	0.167	<b>0.854</b>	0.114	<b>0.536</b>	0.285
C	0.741	<b>0.740</b>	0.021	<b>0.744</b>	0.020	<b>0.748</b>	0.019	<b>0.751</b>	0.020				



# Bivariate normal 模擬結果

20 % censoring (n = 500)													
Log Time	Target value	Cox model		residual smooth		PO model		AFT model		CD1		CD2	
	AUC(t)	EST	SD	EST	SD	EST	SD	EST	SD	EST	SD	EST	SD
- 2.00	0.884	<b>0.751</b>	0.018	<b>0.886</b>	0.025	<b>0.851</b>	0.015	<b>0.885</b>	0.020	<b>0.918</b>	0.034	<b>0.867</b>	0.041
- 1.50	0.833	<b>0.743</b>	0.017	<b>0.840</b>	0.020	<b>0.824</b>	0.013	<b>0.835</b>	0.016	<b>0.886</b>	0.026	<b>0.820</b>	0.035
- 1.00	0.782	<b>0.733</b>	0.015	<b>0.774</b>	0.021	<b>0.787</b>	0.012	<b>0.783</b>	0.014	<b>0.857</b>	0.022	<b>0.771</b>	0.029
- 0.50	0.734	<b>0.724</b>	0.014	<b>0.725</b>	0.023	<b>0.745</b>	0.012	<b>0.734</b>	0.015	<b>0.836</b>	0.019	<b>0.729</b>	0.025
0.00	0.693	<b>0.716</b>	0.014	<b>0.688</b>	0.025	<b>0.699</b>	0.012	<b>0.693</b>	0.016	<b>0.829</b>	0.019	<b>0.703</b>	0.022
0.50	0.660	<b>0.710</b>	0.015	<b>0.662</b>	0.028	<b>0.654</b>	0.014	<b>0.661</b>	0.020	<b>0.837</b>	0.021	<b>0.696</b>	0.023
1.00	0.634	<b>0.704</b>	0.023	<b>0.642</b>	0.037	<b>0.613</b>	0.018	<b>0.634</b>	0.032	<b>0.856</b>	0.027	<b>0.705</b>	0.027
1.50	0.614	<b>0.695</b>	0.047	<b>0.635</b>	0.059	<b>0.610</b>	0.035	<b>0.619</b>	0.071	<b>0.884</b>	0.040	<b>0.732</b>	0.044
2.00	0.598	<b>0.676</b>	0.137	<b>0.653</b>	0.137	<b>0.609</b>	0.109	<b>0.615</b>	0.166	<b>0.883</b>	0.093	<b>0.723</b>	0.176
C	0.741	<b>0.726</b>	0.013	<b>0.743</b>	0.012	<b>0.746</b>	0.011	<b>0.747</b>	0.011				

# Bivariate normal 模擬結果

40 % censoring (n = 500)													
Log Time	Target value	Cox model		residual smooth		PO model		AFT model		CD1		CD2	
	AUC(t)	EST	SD	EST	SD	EST	SD	EST	SD	EST	SD	EST	SD
- 2.00	0.884	<b>0.765</b>	0.020	<b>0.886</b>	0.025	<b>0.851</b>	0.015	<b>0.885</b>	0.019	<b>0.918</b>	0.035	<b>0.867</b>	0.041
- 1.50	0.833	<b>0.756</b>	0.018	<b>0.833</b>	0.021	<b>0.824</b>	0.014	<b>0.835</b>	0.016	<b>0.885</b>	0.027	<b>0.820</b>	0.035
- 1.00	0.782	<b>0.746</b>	0.016	<b>0.771</b>	0.024	<b>0.787</b>	0.013	<b>0.783</b>	0.015	<b>0.856</b>	0.023	<b>0.771</b>	0.029
- 0.50	0.734	<b>0.736</b>	0.015	<b>0.725</b>	0.025	<b>0.745</b>	0.013	<b>0.734</b>	0.016	<b>0.836</b>	0.021	<b>0.729</b>	0.025
0.00	0.693	<b>0.728</b>	0.015	<b>0.689</b>	0.029	<b>0.700</b>	0.014	<b>0.694</b>	0.019	<b>0.829</b>	0.021	<b>0.703</b>	0.023
0.50	0.660	<b>0.721</b>	0.019	<b>0.666</b>	0.035	<b>0.655</b>	0.017	<b>0.660</b>	0.028	<b>0.836</b>	0.026	<b>0.695</b>	0.024
1.00	0.634	<b>0.711</b>	0.037	<b>0.648</b>	0.053	<b>0.615</b>	0.028	<b>0.635</b>	0.057	<b>0.855</b>	0.038	<b>0.704</b>	0.033
1.50	0.614	<b>0.608</b>	0.103	<b>0.665</b>	0.108	<b>0.610</b>	0.088	<b>0.611</b>	0.140	<b>0.873</b>	0.071	<b>0.727</b>	0.093
2.00	0.598	<b>0.606</b>	0.164	<b>0.634</b>	0.160	<b>0.594</b>	0.123	<b>0.601</b>	0.177	<b>0.852</b>	0.122	<b>0.600</b>	0.263
C	0.741	<b>0.738</b>	0.014	<b>0.743</b>	0.013	<b>0.747</b>	0.012	<b>0.748</b>	0.012				

# 模擬結論

- 實驗後期，AUC 的估計容易受估計方法影響，修正後的表現更好。
- 一致性指標因為權重關係，較不受後期 AUC 改變所影響。
- 在同一筆資料上的模型比較，PO 模型的估計結果較好。



## 資料分析

# 愛滋病患者資料介紹

- 資料為 1990 年至 2003 年的台灣愛滋病患者資訊
- 使用當中 137 位愛滋病患者
- 資料記錄每位病人回診時的 CD4 數量及是否接受雞尾酒療法(HAART)
- 37 位患者接受雞尾酒療法，100 位接受標準療法
- 最終共有 68 位病患因愛滋病病發而亡

# 時間獨立共變數存活模型

將資料當作時間獨立共變數資料，取患者 CD4 指數第一筆測量值作共變數，分別考慮三種情形

不同風險模型下 Concordance 比較

- Model I : CD4

- Model II : 治療方法

- Model III : CD4 + 治療方法

Model	Cox			AFT			PO		
	EST	ASE	SD	EST	ASE	SD	EST	ASE	SD
I	0.611	0.027	0.032	0.567	0.027	0.040	0.737	0.023	0.020
II	0.594	0.009	0.023	0.511	0.008	0.022	0.598	0.008	0.010
III	0.675	0.023	0.026	0.646	0.021	0.033	0.797	0.019	0.021

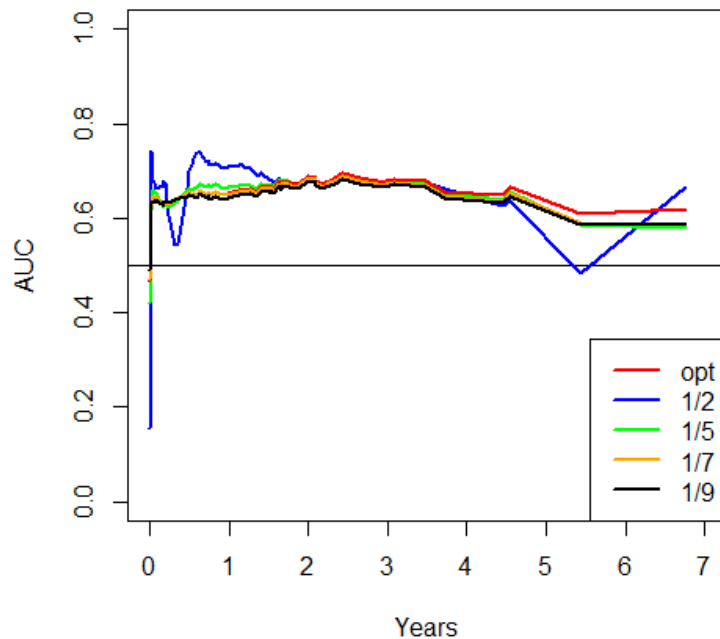
EST：一致性指標估計值，ASE：漸進理論估計標準差，SD：Bootstrap 估計標準差

# 時間獨立共變數存活模型

接著在 **Model III** 下，考慮AFT模型不同帶寬，對估計結果的影響

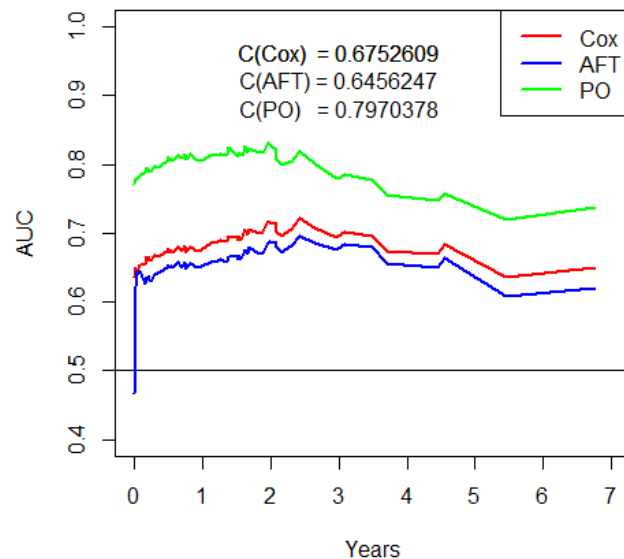
不同帶寬下 Concordance 比較

帶寬	C
1/2	0.6556438
1/5	0.6456529
1/7	0.6424481
1/9	0.6393196
opt	0.6456247



# 時間獨立共變數存活模型

- 根據 Concordance 結果，所有風險模型皆於 **Model III** 下預測效果最佳。
- **Model III** 下，AFT模型選擇最佳帶寬。
- AFT模型因基線風險函數使用核平滑估計法，估計第一點時會有偏差。
- 資料屬於時間獨立共變數時，PO模型有較好的預測能力。



Model III 下不同風險模型 AUC 變化





## 結論

- 實驗後期，各模型下 AUC 的估計容易受估計方法影響，修正後結果較佳。
- 各個模型結果比較，以 PO 模型的結果最佳。
- 實際資料分析，以 PO 模型最適合此筆資料。
- 未來可以考慮將修正過後的計算方式，應用到時間相依共變數的資料中。

A circular graphic with a gradient from orange to blue, surrounded by white lines, containing the text "Thank You!!".

Thank You!!