

計畫編號：DOH91-DC-1059

行政院衛生署疾病管制局九十一年度科技研究發展計畫

台灣地區愛滋病流行病學之數學及統計研究

研究報告

執行機構：國立中興大學

計畫主持人：謝英恆

研究人員：李燦銘、陳婉淑、涂醒哲、許須美、
陳宜民、吳秀英

執行期間：91年1月1日至91年12月31日

* 本研究報告僅供參考，不代表衛生署疾病管制局意見 *

摘要

在愛滋病流行病學中,愛滋病感染者與愛滋病患者人數被低估是一個常見的問題,造成政府訂定公共衛生政策的不定變數及困擾。這個問題在愛滋病流行率低的國家、愛滋病流行地區的流行初期以及難以估計的族群中(如靜脈注射或性工作者)尤其嚴重。因此,我們提出簡單的數學模型來計算隱藏的愛滋病感染者人數並用以了解一個社會中愛滋病疫情的實際嚴重性。我們用這個方法以及疾病管制局所收集 1984 年至今之台灣地區愛滋病疫情資料庫數據來推算台灣地區的愛滋病感染者人數低估的情形。結果顯示台灣地區的未病發愛滋病感染者實際人數(包括已知及尚未檢驗出感染者)約為已知未病發愛滋病感染者存活人數的兩倍。但愛滋病感染者人數持續增加中。如未加強現有防治措施,預計愛滋病感染者總人數將於 2007 年破萬。

另外台灣地區的愛滋病感染者中同性戀/雙性戀者佔很大比率。我們用統計學的 generalized removal model 從台北市立性病防治所的 1999 年底以來台北市同志三溫暖愛滋病篩檢數據來估計台北市同志三溫暖常客中未病發愛滋病感染者實際人數。

Abstract

Underreporting of HIV/AIDS cases is a common problem in HIV epidemiology which often skews epidemiologic projections on which public health policy decisions are to be made. The problem is especially severe in the cases of hard-to-reach populations (IDU's, sex workers, etc.), in low HIV prevalence countries, or in early phases of an HIV epidemic. In this work, we propose a simple mathematical model to compute the number of undetected HIV-positives and the approximate magnitude of underreporting of HIV cases in a society. The procedure is illustrated with the HIV data in Taiwan. The method is also useful for future projections and public health policy planning, given that the prevalence remains low.

We also make use of the generalized removal model for open populations to estimate the number of HIV-infected individuals who frequents the gay saunas in Taipei from the HIV serotesting data from the fall of 1999 to the first half of 2002. The result gives indication of the magnitude of HIV prevalence among gay group in Taipei who are customers of gay saunas.

第一節、前言：

在愛滋病的流行病學中，愛滋病感染者及愛滋病患者的人數被低估是一個很常見的問題，通常在政府的公共衛生政策決定中會被忽略。這樣的問題在第三世界國家[1]或者很難估計人數的族群(如靜脈注射或性工作者)特別嚴重。因此我們最常使用愛滋病感染者與愛滋病患者的監督系統來估計未檢驗出愛滋病感染者與愛滋病患者人數[2]。也有人用更詳細的愛滋病感染者訊息如感染率及愛滋病感染行為來估計愛滋病的流程度[3]。另外在流行率低的國家或是愛滋病流行初期的階段中，小樣本的愛滋病感染者數量對於整體的流行感染疫情常會產生不合理的估計。然而，評估這些重大的問題對於政府制定長期的公共衛生政策是有幫助且必要的。

因此，第二節我們提出一個簡單的數學模型。在第三節我們利用台灣地區的愛滋病感染者資料來做例子來展示如何使用愛滋病感染者的資料來計算愛滋病感染者的未知人數。第四節提出我們利用統計學的 generalized removal model 從台北市同志三溫暖愛滋病篩檢數據來估計台北市同志三溫暖常客中未病發愛滋病感染者實際人數的結果。第五節是本年度研究計劃之總結論及下年度研究計劃將繼續進行之研究。

第二節、數學模型：

我們考慮下列變數，在 t 時間函數如下：

$N(t)$ ---- 未感染者人數(number of susceptible individuals)
 $X(t)$ ---- 未檢驗出的未病發愛滋病感染者人數(number of living undetected HIV+ individuals)

$Y(t)$ ---- 已知未病發愛滋病感染者存活人數(number of living known detected HIV+ individuals)

$Z(t)$ ---- 愛滋病發才得知愛滋病患者(number of AIDS patients detected at AIDS symptoms)

$W(t)$ ---- 病發前已知愛滋病患者(number of previously known AIDS patients)

重要參數如下

μ ---- 未感染者死亡率(mortality rate of susceptibles)

$\bar{\mu}$ ---- 未檢驗出的未病發愛滋病感染者排除率(死亡率+病發率, removal rate of undetected HIV-positives)

$\bar{\mu}^*$ ---- 已檢驗出未病發愛滋病感染者排除率(removal rate of known HIV-positives)

$\bar{\mu}_0$ ---- 愛滋病患者死亡率(mortality rate of AIDS patients)

k ---- 愛滋病感染者檢驗率(detection rate of living undetected HIV-positives)

λ ---- 愛滋病感染者感染率(infection rate of HIV-positives)

δ ---- 未感染者移入率(recruitment rate of susceptibles)

將考慮的聯立微分方程模式如下：

$$\begin{aligned}
N'(t) &= \delta N - \mu N - \lambda N \left(\frac{X}{N+X+Y} \right) \\
X'(t) &= \lambda N \left(\frac{X}{N+X+Y} \right) - \bar{\mu} X - \kappa X \\
Y'(t) &= \kappa X - \bar{\mu}^* Y \\
Z'(t) &= (\bar{\mu} - \mu) X - \bar{\mu}_0 Z \\
W'(t) &= (\bar{\mu}^* - \mu) Y - \bar{\mu}_0 W
\end{aligned}$$

這個模型不同於[4]在於本模型利用從檢驗陽性到病發的時間來計算已檢驗出愛滋病感染者的病發率，以及考慮已檢驗出來為愛滋病感染者的病人接受治療(如雞尾酒療法 HAART)可能改善愛滋病患者繼續蔓延。我們也假設已檢驗出來的愛滋病感染者的感染率是可以忽略的。

第三節、愛滋病感染者流行率低的區域之模式分析

我們考慮愛滋病感染者流行率低的區域，也就是說

$$\frac{X+Y}{N+X+Y} \ll 1, \quad \frac{N}{N+X+Y} \approx 1$$

我們得到

$$\begin{aligned}
N'(t) &= (\delta - \mu) N - \lambda X \\
X'(t) &= \lambda X - (\bar{\mu} + \kappa) X \\
Y'(t) &= \kappa X - \bar{\mu}^* Y \\
Z'(t) &= (\bar{\mu} - \mu) X - \bar{\mu}_0 Z \\
W'(t) &= (\bar{\mu}^* - \mu) Y - \bar{\mu}_0 W
\end{aligned}$$

其中， $\bar{\mu} > \mu, \bar{\mu}_0 > \mu$ ，且 $\lambda > (\bar{\mu} + \kappa)$ 。

在這個動態的模型中，有兩個更有使用價值的參數，稍後將證明它的用途。

一個是 $P(t)$ ，在時間 t 檢驗出未病發的愛滋病感染者的累積數。另一個則

是 $M(t)$, 在時間 t 檢驗出已病發的愛滋病患者的累積數。這些變數的方程式如下:

$$P'(t) = kX$$

以及

$$M'(t) = (\bar{\mu} - \mu)X$$

其中, $P(t)+M(t)$ 為時間 t 時愛滋病感染者的檢驗累積數。上述方程式為線性, 因此很容易找出解。因此,

$$X(t) = X_0 e^{(\lambda - \bar{\mu} - k)t}$$

$$Y(t) = e^{-\bar{\mu}^* t} \left[Y_0 + \frac{kX_0}{\lambda - k + \bar{\mu}^* - \bar{\mu}} (e^{(\lambda - k + \bar{\mu}^* - \bar{\mu})t} - 1) \right]$$

$$Z(t) = e^{-\bar{\mu}_0 t} \left[Z_0 + \frac{(\bar{\mu} - \mu)X_0}{\lambda - k + \bar{\mu}_0 - \bar{\mu}} (e^{(\lambda - k + \bar{\mu}_0 - \bar{\mu})t} - 1) \right]$$

$$W(t) = e^{-\bar{\mu}_0 t} \left[W_0 (\bar{\mu}^* - \mu) \left(Y - \frac{kX_0}{\lambda - k + \bar{\mu}^* - \bar{\mu}} \right) \left(\frac{e^{(\bar{\mu}_0 - \bar{\mu}^*)t} - 1}{\bar{\mu}_0 - \bar{\mu}^*} \right) + \left(\frac{kX_0}{\lambda - k + \bar{\mu}^* - \bar{\mu}} \right) \left(\frac{e^{(\lambda - k + \bar{\mu}_0 - \bar{\mu})t} - 1}{\lambda - k + \bar{\mu}_0 - \bar{\mu}} \right) \right]$$

$$P(t) = P_0 + \frac{kX_0}{\lambda - k - \bar{\mu}} [e^{(\lambda - k - \bar{\mu})t} - 1]$$

$$M(t) = M_0 + \frac{(\bar{\mu} - \mu)X_0}{\lambda - k - \bar{\mu}} [e^{(\lambda - k - \bar{\mu})t} - 1]$$

其中, $X_0 = X(0), Y_0 = Y(0), Z_0 = Z(0), W_0 = W(0), P_0 = P(0)$, 以及 $M_0 = M(0)$ 為各個變數的起始值。我們主要目的是在既定的時間去估計某個中族群 $X(t)$ 的值。其他變數的資料則可從政府的愛滋病感染者資料中取得。表一列出台灣地區從 1984 年到 2000 年底檢驗出本國籍愛滋病感染者人數共 2969 人, 這是從台灣的疾病管制局的愛滋病感染者資料庫計算得出。 P_i 是第 i 年底已檢驗出為愛滋病感染者但未有其他症狀的累積數。模型中我們令 1993 年為 t 時間=0, 因為 1993 年以前的資料有較大的波動, 對我們的估計可能導致不必要的錯誤。1993 年底已檢驗出為愛滋病感染者但未有其他症狀的累

積數為 457, 是 $P(0)$ 的真實值。另外 4 組資料, Y_i, M_i, Z_i , 以及 W_i 也都很類似。

資料 P_i 的定義本質是精確的, 就是歷年來檢驗出為愛滋病感染者的人數。 Y_i 則是在假設愛滋病感染者只有自然死亡以及病發為精確的。 資料顯示台灣地區有愛滋病感染者的人與其他年齡層的人比較有顯著的高死亡率(車禍或自殺等因素)。然而, 愛滋病感染因素而其差異性是不可能的。所以我們判斷 Y_i 資料中存在一些不精準。而 W_i 與 Y_i 有關, 因此也包含一些不精準。直覺上 M_i 和 Z_i 有最大的誤差, 因為他們是在初被診斷出有愛滋病感染時就已經有症狀出現。愛滋病患者在大部分的國家被低估是很常發生的, 所以我們判斷 M_i 和 Z_i 的資料比其他表中的資料更不精準。我們的目標是在我們的估計程序中盡可能的得出正確的資料。

模型中的一些必要參數可由其他來源得到。愛滋病感染者的 2969 人中有超過 80% 的人年齡約在 20-49 歲。在台灣地區這些年齡層之 5 年存活率的範圍是從 20-24 歲較高的 0.996 到 25-49 歲的 0.982。所以我們令自然死亡率為 0.002, 比 20-49 歲的所有人的 5 年存活率 0.990 還多。使用 Kaplan-Meier 方法, 我們得到從診斷到病發(所有愛滋病感染者人數扣掉還沒病發就已死亡的人共 2255 人)的平均存活時間(mean survival time)是 11.90 年。所以我們得到 $\bar{\mu}^* - \bar{\mu} = 1/11.90 = 0.086$ 。同樣的, 從病發到死亡(共 868 人)的平均存活時間是 5.45 年, 因此 $\bar{\mu}_0 = 1/5.45 = 0.184$ 。為估計愛滋病感染者的平均潛服期, 我們使用 1984 年到 2000 年經由輸血感染愛滋病的感染者從病發到死亡的人數共 44 人。理由是這群人經由被傳染的輸血, 通常在短短的幾個月被檢驗及診斷出為愛滋病感染者, 因此提供潛伏期的估計是一個好的方法。結果經由 Kaplan-Meier 方法得出是 12.20 年,

$$\bar{\mu} - \mu = 1/12.20 = 0.084.$$

為了得到正確的估計，我們利用 P_i 和 M_i 來估計理論參數 k ，進而估計 $X(t)$ 。從不同的方程式 $P(t)$ 和 $M(t)$ ，我們得到

$$\frac{\ln P'(t)}{\ln M'(t)} = \frac{k}{\bar{\mu} - \mu}$$

或 $k = (\bar{\mu} - \mu)P'(t)/M'(t)$ 。

表一中， P_i 和 M_i 的資料是時間 t 為 0 到 7 (即 1993 年到 2000 年)，我們利用 central difference 算出 ΔP_i 以及 ΔM_i ，則 $\Delta P_i = (P_{i+1} - P_{i-1})/2$ 以及 $\Delta M_i = (M_{i+1} - M_{i-1})/2$ 。 ΔP_i 與 ΔM_i 的值列於表二，當 $i=1, 6$ (1994 年到 1999 年)。此外，我們令 $\Delta P_i / \Delta M_i$ 近似於 $P'(t)/M'(t)$ 時間為 $t=1$ 到 6，使用 μ 和 $\bar{\mu}$ 的既定值得到 k 的平均值 $=0.283$ 。下面我們使用公式(2)及真實值 P_i ， $i=0$ 到 7 (1993 年到 2000 年) 以最小平方法來估計 kX_0 與 $\lambda - k - \bar{\mu}$ 。從這個程序我們得到 $kX_0 = 126.1$ 和 $\lambda - k - \bar{\mu} = 0.199$ 。接著我們得到 1993 年愛滋病感染者未知人數的起始值 $X_0 = 446$ ，感染率的估計值 $\lambda = 0.571$ 。

用我們估計出的參數，接著計算 $Y(t)$ ， $P(t)$ ， $W(t)$ ， $Z(t)$ 以及 $M(t)$ 的理論值， t 從 0 到 7，故我們可以跟實值作比較。將結果繪圖於圖 2-6。而 $P(t)$ 及 $M(t)$ 的理論估計與真實值則列於圖三與圖六，因此使用 P_i 和 M_i 來估計參數是適當的。其他參個估計值就較不完善。值得一提的是由於 $M(t)$ 與 $Y(t) + W(t)$ 有關，所以很直覺地， Y_i 低估時會相對應於 w_i 的高估，我們將結果列於圖二及圖四。

本章最主要的主題是估計潛在的愛滋病感染者人數 $X(t)$ 。對於這個主題我們從 $X(t)$ 的方程式中，以 X_0 ， λ ， k ， $\bar{\mu}$ 的估計值來計算 $X(t)$ 的理論值。結

果得出愛滋病感染者的累積數 $P(t)+M(t)$ ，時間 t 從 0 到 9(即 1993 年到 2002 年)，列於表三。結果顯示到今年底為愛滋病感染者但未檢驗出的人數達到 2674 人，台灣地區至今年底的愛滋病感染者累積人數達到 4661 人。特別注意的是由 1984 年到 2000 年的資料來估計 2001 年底的愛滋病感染者累積人數是 3764 人，而台灣地區 2000 年底的愛滋病感染者真實人數是 3595 人，兩者只有相差 4.7%。

第四節、台北市同志三溫暖常客中未病發愛滋病感染者人數的估計

一、統計方法

近年來我們利用古巴與泰國之愛滋病感染者人數的數據來推算古巴及泰國二地實際的愛滋病感染者人數，已有不錯之結果(參見 Hsieh, Chen, and Lee 2000 Stat. Med. AIDS)，我們也一直希望能研究台灣地區實際的愛滋病感染人數。首先我們整理台北市立性病防治所現有的同志三溫暖的愛滋病感染者之篩檢結果資料，得出 5 組資料，將民國 89 年 1 月至 6 月的愛滋病感染者人數及篩檢人數分別加總，得出第一組資料，民國 89 年 7 月至 12 月的愛滋病感染者人數及篩檢人數分別加總，得出第二組資料，民國 90 年 1-6 月的資料為第三組，民國 90 年 7~12 月的資料得到第四組資料，民國 91 年 1~6 月的資料得到第五組資料，共 5 組。基本上是以半年為一抽樣時間。由於其是在特定地區所做之篩檢，所以我們假設篩檢人數與愛滋病感染者被篩檢到之機率成正比，設 e_j^* 為第 j 期篩檢之人數，定義 $e_1 = 1$ ， $e_j = \frac{e_j^*}{e_1^*}$ ，

$j=1, 2, \dots, s$, 及 P 為第一期愛滋病感染者被篩檢到的機率, pe_j 為第 j 期可被篩選到之愛滋病感染者的機率。定義 N_j 為到第 j 期以前的愛滋病感染者總數, u_j 為第 j 期新篩檢到的愛滋病感染者人數, M_j 表示到第 $j-1$ 期已觀測到愛滋病感染者之數目, 若不考慮愛滋病感染者死亡人數則 $M_j = u_1 + u_2 + \dots + u_{j-1}$, 在 $N_j - M_j$ 給定及 pe_j^* 下,

$$u_j \sim \text{Bin}(N_j - M_j, pe_j^*), \text{ 此時其概似函數為 } L \propto \prod_{j=1}^s (pe_j^*)^{u_j} (1 - pe_j^*)^{N_j - M_j - u_j} \binom{N_j - M_j}{u_j},$$

其中

$M_{j+1} = M_j + U_j$, 為至第 j 期已觀測到愛滋病感染者之人數。若有死亡人數則取 $M_j = \phi M_{j-1} + u_{j-1}$, 為至第 j 期觀測愛滋病感染者平均存活之人數, 其中 ϕ 為已知存活率。在此我們使用貝氏分析來估計 N_j 與 P 。我們考慮的事前分配為 $\pi(N, P) = \pi(N_1, N_2, \dots, N_s) \pi(p)$, 其中取 $\pi(p) = I_{[\alpha/4, 7\alpha/4]}(p)$,

u_j 有關 N_j 之決定是用 empirical 決定的, 詳見附錄之說明, 此時 P 的條件事後分配為

$$\pi(P | N_1, \dots, N_s, Data) \propto P^{u_1 + \dots + u_s} \prod_{j=1}^s (1 - pe_j)^{N_j - M_{j+1}}$$

, 則及 N_j 的條件事後分配為

$$\pi(N_j | N_{(-j)}, P, Data) = \frac{\binom{N_j - M_j}{u_j} (pe_j)^{u_j} (1 - pe_j)^{N_j - M_{j+1}}}{\sum_{N_j = \max\{N_{j-1}, M_{j+1}\}}^{N_{j+1}} \binom{N_j - M_j}{u_j} (pe_j)^{u_j} (1 - pe_j)^{N_j - M_{j+1}}}$$

其中, $N_{(-j)} = (N_1, N_2, \dots, N_{j-1}, N_{j+1}, \dots, N_s)$, 事實上, $N_j | N_{(-j)}, p, Data$ 為截尾之負二項分配, 給定起始值 $\{N_1^{(0)}, \dots, N_s^{(0)}\}$ 後, 利用 Gibbs 抽樣法得到 Gibbs sequence $\{P^{(h)}, N^{(k)}\}$, 在用此資料的平均數及中末數來推估 N_1, \dots, N_s 及 p 。我們採用 Gibbs 抽樣重複執行 9 萬次, 截去前面 3 萬個不平穩的觀察值, 而為達到較獨立的觀察值, 每間隔 5 個觀察

值取一個樣本，共收集了 1 萬 2 千個觀察值。求得平均數及標準差。估計參數 P 時，所使用的事前機率是來自 $U(\frac{\alpha}{4}, \frac{7\alpha}{4})$ ，其中 $\alpha =$

$$E(P) = \frac{E(u_1)}{N_1} \approx 1 - \frac{E(u_2)e_2}{E(u_1)e_1} \text{ (相關說明參見附錄) }，\text{ 因此可得 } \alpha \text{ 之估計值}$$

$\hat{\alpha}$ ，帶入求得，由於 P 的條件事後分配不具有封閉形式的分配，所有無法由分配中生成隨機變數，我們藉由 etropolis-Hastings 演算法求得。本文與 Chen, Lee, Hsieh & Ungchuask(1999) 及 Hsieh, Chen & Lee(stat in Med)(2000) 之不同處在第 j 期被觀測到之機率為 Pe_j^* ，其中 e_j^* 為已知，而 Chen et al (1999) 及 Hsieh al et.(2000) 則為 p_j ，而 p_j 之事前分配為 Beta(α, β)， α, β 之決定是用 empirical 方法，但其決定方法是由 e_j^* 及其他訊息來決定的，而本文則是

$P \sim U(\frac{\alpha}{4}, \frac{7\alpha}{4})$ 二者是不同的，本文 Pe_j^* 之結果能適當反應出當期篩檢人數之結果，其結果更具反應出實際情況。執行 Gibbs 抽樣的過程中，我們檢視疊代的過程中參數估計值的收斂程度是計算 potential scale 值來判斷疊代次數是否足夠。令 (potential scale)

$$\sqrt{\hat{R}} = \sqrt{\frac{\hat{\text{var}}(N_j)}{W}}$$

$$\hat{\text{var}}(N_j) = \frac{n-1}{n}W + \frac{1}{n}B$$

其中， b 表示各列馬可夫鏈的組間差異， w 表示一列中馬可夫鏈了組內差異， n 為取樣次數。若所有序列之 \hat{R} 不接近 1，則需繼續執行。假如 \hat{R} 較大，則我們有理由相信未來的模擬程序可能會改善我們對目標分配的推論。一但 \hat{R} 很接近 1 時，表示由 Gibbs 抽樣所模擬的資料會收斂到目標分配。

1.1 Gibbs 抽樣

此方法主要是一種自邊際分配抽樣而得到估計值隨機分布的技巧，不用透過複雜積分的一種方法，應用過程是先由條件分配中製造出一串馬可夫鏈，再藉由馬可夫鏈的分布來解釋我們的目標方佈，詳細內容在 Geman and Geman(1984)的文獻中被討論過，Gelfand and Smith(1990)將 Gibbs 抽樣運用在計算近似的邊際分配中，引起統計學家的廣泛重視。詳細方法介紹可參考 Casella and George(1992)及 Gilks, Richardson and Spiegelhalter(1996)的文獻。

1.2 Metropolis-Hastings 演算法

定義一函數 $g(\theta_1)$ ，令為 $f(\theta_1)$ 的轉換函數，則 M-H 演算法的過程如下；

(1) 首先給定參數 θ_1 的起始值，表示為 $\theta_1^{(0)}$

(2) 經由轉換函數 $g(\theta_1)$ 轉換生成 θ_1^*

(3) 在連續的均勻分配 $U(0, 1)$ 中隨機抽取一數值 P^* ，與

$$P = \min\left\{1, \frac{f(\theta_1^*)g(\theta_1^{(0)})}{f(\theta_1^{(0)})g(\theta_1^*)}\right\} \text{ 比較，所以 } p^* > p, \text{ 則 } \theta_1^{(1)} \text{ 的值以 } \theta_1^* \text{ 取代；}$$

否則 $\theta_1^{(1)}$ 仍為 $\theta_1^{(0)}$ 。

(4) 接著再以 $\theta_1^{(1)}$ 當作步驟 1 的起始值，重複步驟 (2) 及步驟 (3) 疊代出 $\theta_1^{(2)}$ ，依此類推重複這樣的動作，直到疊代出的 $\theta_1^{(j)}$ 值達到平穩為止。

二. 模擬研究與實例分析

2.1 模擬研究

本節我們做模擬研究可對結果與真實情況做一比較。對於每一個資料集採用 Gibbs 抽樣重複執行 9 萬次，截去前面 3 萬個不平穩的觀察值，而為達到較獨立的觀察值，每間隔 5 個觀察值取一個樣本，共收集了 1 萬 2 千個觀察值。我們總共模擬 500 組資料，再計算出此 1 萬 2 千個觀察值的平均數，中位數，標準差，95%的貝氏信賴區間。我們總共做了 9 組 n 與 p 的組合，列於表一至表九。從表一至表三中可看出推估 N_1, N_2, N_3, N_4 時，平均數較接近真正值，而估計 N_5 時，以中位數表現較好。推估 N_1, N_2, N_3, N_4, N_5 時的標準差有遞增的情形，信賴區間則有右長尾的情況。表四至表六所推估的 N 為平均數比中位數的表現較好。表七至表九則估計 N 以中位數較佳。整體而言我們所提出之估計平均數均較中位數接近真正之 N 值。

表一：模式模擬共 500 組資料， $(P_1, P_2, P_3, P_4, P_5) = (0.15, 0.225, 0.12, 0.225, 0.3)$

	MED	MEAN	SD	95%CI	
				lower bound	upper bound
N1=200	180.6	199.4	71.5	121.4	395.5
N2=250	233.9	256.9	79.0	178.3	477.9
N3=300	279.7	303.4	84.2	217.4	538.9
N4=350	329.3	356.0	89.2	270.5	607.6
N5=380	375.4	404.4	93.5	319.3	671.0

表二：模式模擬共 500 組資料， $(P_1, P_2, P_3, P_4, P_5) = (0.15, 0.225, 0.12, 0.225, 0.3)$

	MED	MEAN	SD	95%CI	
				lower bound	upper bound
N1=250	221.3	240.9	75.2	153.8	448.9
N2=300	277.0	300.2	80.0	215.7	525.8

N3=350	326.3	349.8	83.1	259.6	584.0
N4=400	375.1	400.9	85.8	313.2	645.1
N5=430	421.3	448.6	88.0	363.0	702.2

表三：模式模擬共 500 組資料， $(P_1, P_2, P_3, P_4, P_5) = (0.15, 0.225, 0.12, 0.225, 0.3)$

	MED	MEAN	SD	95%CI	
				lower bound	upper bound
N1=300	265.3	285.1	79.0	190.5	501.6
N2=350	325.4	348.5	82.0	258.5	579.6
N3=400	377.5	400.3	83.4	306.5	635.3
N4=450	427.8	452.7	84.3	364.4	693.3
N5=480	477.4	503.2	84.8	417.9	748.0

表四：模式模擬共 500 組資料， $(P_1, P_2, P_3, P_4, P_5) = (0.13, 0.25, 0.1, 0.2, 0.15)$

	MED	MEAN	SD	95%CI	
				lower bound	upper bound
N1=200	165.9	181.9	64.0	108.0	356.5
N2=250	217.9	238.4	70.5	166.3	435.8
N3=300	265.2	286.4	72.2	213.5	490.4
N4=350	316.3	338.9	76.9	261.6	555.7
N5=380	365.7	390.9	87.0	302.0	636.7

表五：模式模擬共 500 組資料， $(P_1, P_2, P_3, P_4, P_5) = (0.13, 0.25, 0.1, 0.2, 0.15)$

	MED	MEAN	SD	95%CI	
				lower bound	upper bound
N1=250	196.8	215.5	71.9	133.0	414.1
N2=300	250.5	273.5	77.6	194.1	492.5
N3=350	304.0	327.1	77.3	248.7	547.3
N4=400	355.3	379.4	80.2	298.8	608.5

N5=430	404.7	430.6	88.3	340.0	682.9
--------	-------	-------	------	-------	-------

表六： 模式模擬共 500 組資料， $(P_1, P_2, P_3, P_4, P_5) = (0.13, 0.25, 0.1, 0.2, 0.15)$

	MED	MEAN	SD	95%CI	
				lower bound	upper bound
N1=300	236.0	254.5	75.9	165.6	455.8
N2=350	294.2	316.7	79.8	234.0	533.0
N3=400	351.0	373.0	77.9	293.0	585.4
N4=450	404.2	426.7	79.6	345.7	643.6
N5=480	456.1	480.1	86.7	389.8	716.7

表七： 模式模擬共 500 組資料， $(P_1, P_2, P_3, P_4, P_5) = (0.2, 0.2, 0.2, 0.2, 0.2)$

	MED	MEAN	SD	95%CI	
				lower bound	upper bound
N1=200	244.1	268.9	90.6	170.5	518.5
N2=250	303.3	331.8	98.3	230.1	605.7
N3=300	353.5	384.5	104.6	278.9	677.8
N4=350	409.2	442.6	111.6	331.1	756.9
N5=380	463.6	500.2	122.8	377.1	846.7

表八： 模式模擬共 500 組資料， $(P_1, P_2, P_3, P_4, P_5) = (0.2, 0.2, 0.1, 0.15, 0.25)$

	MED	MEAN	SD	95%CI	
				lower bound	upper bound
N1=200	237.8	264.8	94.1	164.6	525.0
N2=250	294.5	325.6	103.6	220.1	614.6
N3=300	346.2	379.5	113.8	260.6	696.1
N4=350	402.9	440.5	124.1	314.4	788.0
N5=380	456.6	499.1	135.5	366.3	881.3

表九： 模式模擬共 500 組資料， $(P_1, P_2, P_3, P_4, P_5) = (0.18, 0.1, 0.15, 0.25, 0.15)$

	MED	MEAN	SD	95%CI	
				lower bound	upper bound
N1=200	200.6	220.8	74.6	138.3	428.0
N2=250	254.7	277.5	84.0	184.6	512.0
N3=300	305.2	331.6	91.7	234.2	589.6
N4=350	359.1	389.2	98.0	291.8	668.8

N5=380	417.7	449.3	107.9	338.1	755.2
--------	-------	-------	-------	-------	-------

2.2 實例分析

我們整理台北市立性病防治所現有的同志三溫暖的愛滋病感染者之篩檢結果資料，從民國 89 年 1 月至 91 年 6 月，以半年為一抽樣時間得出 5 組資料。設 e_j^* 為第 j 期篩檢之人數，定義 $e_1 = 1$ ， $e_j = \frac{e_j^*}{e_1^*}$ ， $j=1, 2, \dots, s$ ， P_1 為第一期愛滋病感染者被篩檢到的機率， $P_j = pe_j$ 為第 j 期可被篩選到之愛滋病感染者的機率。定義 N_j 為到第 j 期以前的愛滋病感染者總數， u_j 為第 j 期新篩檢到的愛滋病感染者人數。表十為台北市立性病防治所所收集到之每半年篩檢人數及所篩檢到愛滋病感染者之數目。

表十

	1	2	3	4	5
u_j	7	24	6	12	19
e_j^*	63	273	139	178	213

我們採用 Gibbs 抽樣重複 9 萬次，每隔 5 個取一個，以使此分配之樣本較接近獨立，我們將此 1 萬 8 千個觀測值畫時間序列圖，可看出去掉 6000 筆後就趨近平穩之變化，可知如截掉原 Gibbs 序列之前 3 萬筆後，就趨近平穩之變化，所以下面之估計是截去前面 3 萬個不平穩的觀察值，而為達到較獨立的觀察值，每間隔 5 個觀察值取一個樣本，共收集了 1 萬 2 千個觀察值。估計結果如表十一：

表十一

	Median	Mean	SD	95%CI lower bound	95%CI upper bound
N1	108.1	107.6	25.7	56.0	156.6

N2	138.1	135.9	26.9	77.4	182.9
N3	153.9	151.3	29.4	87.4	202.4
N4	177.2	174.2	33.4	101.7	233.0
N5	215.3	212.5	41.4	125.0	288.0

從表十一可得平均數與 Median 均相當接近, 差距均小於 4, 標準差遞增的現象此部份與模擬結果相同。

第五節. 總結

1. 數學模式結果顯示台灣地區的未病發愛滋病感染者實際人數 (包括已知及尚未檢驗出感染者) 約為已知未病發愛滋病感染者存活人數的兩倍。但愛滋病感染者人數持續增加中。如未加強現有防治措施, 預計愛滋病感染者總人數將於 2007 年破萬。
2. 數學模式中參數估計結果 (1993-2000 平均值): 傳染基數 (Basic Reproduction Number) $R_0=1.239$, 平均 HIV 傳染率 (mean infection rate 或每次接觸平均 HIV 傳染機率 \times 每年未檢驗出愛滋病感染者平均接觸的未感染人數) 是 0.571, 平均感染至篩檢時間 (mean detection time) 是 3.55 年, 58% 愛滋病感染者是在愛滋病併發症症狀出現前兩星期內才檢驗出感染愛滋病毒。
3. 台灣地區的愛滋病感染者中同性戀/雙性戀者佔很大比率。我們應用統計上生態模式的廣義移除模式 (generalized removal model), 從台北市立性病防治所的 1999 年底以來台北市同志三溫暖愛滋病篩檢數據來估計台北市同志三溫暖常客中未病發愛滋病感染者實際人數。從 89 年 1 月至 91 年 6 月的 30 個月中, 累積已檢驗出愛滋病感染者人數為 68

人,而我們的結果顯示台北市同志三溫暖常客中,在 90 年 12 月至 91 年 6 月間尚未檢驗出的未病發愛滋病感染者實際人數平均為 197 人,可知為已檢驗出愛滋病感染者人數之三倍。另外在這 30 個月中,台北市同志三溫暖常客中未病發愛滋病感染者估計人數增加一倍。趨勢值得注意。

4. 在本年度研究計劃的第二年中,我們將進行下列工作:
 - a. 加入 2001 年疾病管制局愛滋病疫情資料庫數據。
 - b. 修改數學模式,加入已檢驗出愛滋病感染者的傳染率及在雞尾酒療法前後不同愛滋病併發症病發率、存活率。以探討已檢驗出愛滋病感染者對台灣地區愛滋病傳染之重要性、全面雞尾酒療法政策對台灣地區愛滋病傳染之影響等重要議題。
 - c. 繼續台北市同志三溫暖常客中未病發愛滋病感染者之統計分析並建立數學模式,以追蹤及掌握此高感染率高危險群之未來愛滋病傳染趨勢。

參考文獻：

1. Balint, G.A. (1998) Situation analysis of 愛滋病感染者/愛滋病患者 epidemic in sub-Saharan Africa, *East. Afr. Med. J.* **75**(12):684-686.
2. Barchielli, A., Buiatti, E., Balanti, C., Giovannetti, L., Acciai, S., and Lazzeri, V. (1995) Completeness of AIDS reporting and quality of AIDS death certification in Tuscany (Italy): a linkage study between. *Eur. J. Epidemiol.* **11**(5): 513-7.
3. Wijngaarden, J.K. and Jager, H.J. (1998) Methods for estimating HIV prevalence: A comparison of extrapolation from surveys on infection rate and risk behavior with back-calculation for the Netherlands. *Eur. J. Epidemiol.* **14**(7): 645-52.
4. Anderson, R.M., May, R.M. and McLean, A.R. (1988) Possible demographic consequences of AIDS in developing countries. *Nature* **332**, 228-234.
5. de Arazoza, H., Lounes, R., Hoang, T., and Interlan, Y. (2000) Modeling

- HIV epidemic under contact tracing - The Cuban case. *J. of Theoretical Medicine*. 2: 267-274.
6. Lounes, R, and de Arazoza, H. (1999) A two-type model for the Cuban national programme on HIV/AIDS, *IMA J. Math. Appl. Med. Biol.* 16:143-154.
 7. Hsieh, Y.-H. (2002) The changing faces of commercial sex in Thailand: its implications for HIV transmission. *JAIDS*. 30(5): 537-540.
 8. Hsieh, Y.-H., Lee, S.-M., Chen, C.W.S., and Arazoza, H. On the recent sharp increase of HIV infections in Cuba. (2001) *AIDS*. 15(3):426-428.
 9. Chen, C.W.S., Lee, S.-M., Hsieh, Y.-H., and Ungchusak, K. (1999). A unified approach to estimating population size for births only models, *Compu. Stat. Data Analysis*, 32:29-46.
 10. Hsieh, Y.-H., Chen, C.W.S., and Lee, S-M. (2000) Empirical Bayes approach to estimating the number of HIV-infected individuals in hidden and elusive populations. *Stat. Med.* 19: 3095-3108.
 11. Y.-H. Hsieh, H. de Arazoza, S.M. Lee, C.W.S. Chen. (2002) Estimating the number of Cubans infected sexually by HIV using contact tracing data. *Inter. J. Epidemiol.* 31: 679-683.
 12. Y.-H. Hsieh, H.C. Wang, S.J. Twu, and S-M. Hsu. Ascertaining underreporting in HIV cases in Taiwan. Preprint. (2002)
 13. Hsieh Y.H., Chen C.W.S., Lee, S.M., and de Arazoza H. (2001) On

the recent sharp increase in HIV detections in Cuba. *AIDS* **15**(3), 426-428.

14. Chen, Y.-M. A., Chung, C.-L., Wu, S.-I., Hsu, H., and Jen, I. (2001) A Survey on the HIV-1, Syphilis and Sexual Behavior Among Patrons of Gay Saunas in Taiwan. 2001 Hong Kong AIDS Conference. Conference Proceedings, pp. 128-150.
15. Casella, G. and E.I. George, Explaining the Gibbs sampler, *The American Statistician*, **46** (1992) 167-174.
16. Gelfand, A.E. and A.F.M. Smith, Sampling-based approach to calculating marginal densities, *J. Amer. Statist. Assoc.*, **85** (1990) 398-409.
17. Geman, S. and D. Geman, Stochastic relaxation, Gibbs distributions and the Bayesian restoration of images, *IEEE Trans. on Pat. Anal. And Mach. Intel.*, **6** (1994) 721-741.
18. Gilks, W. R., Richardson, S., and Spiegelhalter, D. J. (1996) *Markov chain Monte Carlo in practice*, London: Chapman & Hall.

附錄

設 pe_j 為第 j 期可被篩選到之愛滋病感染者的機率，

其中， e_j 為相對第 j 期之補取力(catch effort)

若 e_j^* 為第 j 期篩檢之人數，則可令 $e_1 = \frac{e_1^*}{e_1^*} = 1$ ， $e_j = \frac{e_j^*}{e_1^*}$ ， $j=1,2,\dots,s$

令 N_j 為到第 j 期以前的愛滋病感染者總數

u_j 為第 j 期新觀測到的愛滋病感染者人數，其概似函數為

$$L \propto \prod_{j=1}^s \binom{N_j - M_j}{u_j} (pe_j)^{u_j} (1 - pe_j)^{N_j - M_j - u_j}$$

則 $u_j | N_j - M_j \sim Bin(N_j - M_j, pe_j)$

其中， M_j 表示到第 $j-1$ 期已觀測到愛滋病感染者之數目，若不考慮愛滋

病感染者死亡人數則 $M_j = u_1 + u_2 + \dots + u_{j-1}$, 若有死亡人數則

$$M_j = \phi M_{j-1} + u_{j-1},$$

ϕ 為存活率 , 並且得出

$$Eu_1 = N_1 pe_1$$

$$Eu_2 = (N_1(1 - pe_1)\phi + B_1)pe_2$$

其中 , $\phi = 1, B_1 = 0$, 因此

$$\alpha = EP = \frac{E(u_1)}{N_1} \sim 1 - \frac{E(u_2 | N_1)e_2}{E(u_1 | N_1)e_1}$$

$$\hat{\alpha} \approx 1 - \frac{u_2 e_2}{u_1 e_1}$$

在使用貝氏分析時 , P 的事前機率 $U(0, \alpha)$, 則其 N_j 及 P 的條件事後分配為

$$\begin{aligned} \prod (P | N_1, \dots, N_s, Data) &\propto P^{u_1 + \dots + u_s} \prod_{j=1}^s (1 - pe_j)^{N_j - M_{j+1}} \\ \prod (N_j | N_{j-1}, P, Data) &= \frac{\binom{N_j - M_j}{u_j} (pe_j)^{u_j} (1 - pe_j)^{N_j - M_{j+1}}}{\sum_{N_j = \max\{N_{j-1}, M_{j+1}\}}^{N_{j+1}} \binom{N_j - M_j}{u_j} (pe_j)^{u_j} (1 - pe_j)^{N_j - M_{j+1}}} \end{aligned}$$

事實上 , $N_j - M_j$ 為截尾之負二項分配 (truncate NB)

$$\max\{0, N_{j-1} - M_j\} \leq N_j - M_j \leq N_{j+1} - M_j$$

Table 1. Yearly numbers of the accumulated detected asymptomatic HIV- infectives (P_i), known persons living with HIV (Y_i), the accumulated AIDS cases detected at onset of ADI (M_i), and AIDS patients detected at onset of AIDS symptoms (Z_i), AIDS patients known before onset of AIDS (W_i).

Year (i)	P_i	Y_i	M_i	Z_i	W_i
1984	9	9	0	0	0
1985	24	24	0	0	0
1986	34	34	1	0	0

1987	46	45	1	0	1
1988	74	68	2	0	2
1989	111	102	8	3	2
1990	144	128	11	2	3
1991	224	196	23	6	6
1992	345	306	38	8	7
1993	457	402	64	17	10
1994	600	499	99	32	23
1995	776	626	154	49	39
1996	980	736	228	77	75
1997	1253	949	306	105	101
1998	1555	1183	410	155	114
1999	1917	1483	521	227	144
2000	2343	1840	626	281	173

Table 2. The rate of change of P_i and M_i , P_i and M_i , for $t=1$ to 6 (1994-1999).

Year (i)	P_i	M_i
1994	159.5	45
1995	190	64.5
1996	238.5	76
1997	287.5	91
1998	332	107.5
1999	394	108

Table 3. The theoretical numbers of $X(t)$ and $P(t)+M(t)$ for $t=0$ to 9 (1993-2002).

Year	$X(t)$	$P(t)+M(t)$
1993	446	521
1994	544	704
1995	664	926
1996	810	1198
1997	989	15229
1998	1206	1933

1999	1472	2428
2000	1796	3029
2001	2191	3764
2002	2674	4661