International Journal of Statistical Sciences *ISSN 1683–5603*
Vol. 8, 2008, pp 1-22

On Prior Distribution of HIV, Distribution of Incubation Period of AIDS and BLUE of HIV Population

Suddhendu Biswas[1], Sejong Bae and Karan P. Singh
Department of Biostatistics, School of Public Health
University of North Texas Health Science Center
3500 Camp Bowie Blvd.
Fort Worth, TX 76107, USA
E-mail: suddhendubiswas@yahoo.co.in

[Received January 12, 2008; Revised June 8, 2008; Accepted August 3, 2008]

Abstract

The objective of this paper is to develop a methodology for the Best Linear Unbiased Estimator (BLUE) of HIV population vies avis the projection of AIDS. This has been accomplished by developing a suitable informative prior distribution of HIV and an incubation period distribution of AIDS based on the reduction of CD-4 cells; the latter being a state through which an HIV positive individual passes while reaching the condition of full blown AIDS. Commenges and Etcheverry (1993) obtained Empirical Prior distribution for HIV using a Poisson process without taking into consideration of (i) deaths of HIV population in the state of HIV, (ii) Transition to AIDS during the process of growth of HIV population. The present approach comprises of replacing the Poisson Process by Kendall's Generalized birth and death process (1948), that does take into consideration of deaths of HIV population in the state of HIV. Also for consideration of transition to AIDS during the process of growth of HIV population, a more meaningful incubation period distribution based on biological consideration of decay of CD-4 cells has been developed and applied to the present situation rather than an inflexible Weibull distribution as used by many of the earlier investigators. Estimates for the BLUE of HIV population for 16 years are obtained first by obtaining a prior distribution of HIV population on time based on Kendall's birth and death process supplementing by using the data of the US official Statistics. Then, by setting a linear regression model of AIDS on HIV under Gauss Aitken set- up while the incidence data of AIDS from the prior distribution of HIV and the newly developed incubation period distribution is obtained.

Keywords and Phrases: HIV, CD4-Count, and AIDS; Sero-Positivity; Incubation Period; Prior Distribution; Best Linear Unbiased Estimator; Kendall's Birth and Death Process; US Official Statistics: Gauss-Aitken least squares.

AMS Classification: 62P10.

[1]At present in the Department of Mathematics & Statistics, School of Applied Sciences, Amity University, Noida (UP)

1 Introduction

The Center for Disease Control and Prevention (CDC) defined individuals with AIDS, who have CD-4 counts less than 200 cells per cubic millimeter or a CD-4 person less than 14% (the 1992 surveillance definition of CDC, Monroe, 1996; Dwyer, 1995). It is also revealed that after a long follow-up of HIV positive men, only 69% were diagnosed with AIDS after being HIV positive for 13 1/2 years (Buchbinder, et al, 1993). While 67% of patients infected with HIV will have developed AIDS within ten years of infection (Dwyer, 1995), others stay well longer. The misclassification with respect to the causes of deaths of HIV positive individuals mostly happens because the incubation period of AIDS is usually abnormally long even to the extent of 30 years or even more (Becker, 1991).

A good approach given by Becker (1989) is to use a non-parametric maximum likelihood technique incorporating modified E-M algorithm. Commenges and Etcheverry (1993) proposed a better technique exploring the pattern of growth in HIV population. The technique uses empirical Bayes approach for obtaining prior distribution of HIV based on the transmission of HIV infection from carrier to susceptible at varying Poisson rates over time. The authors considered only the data on AIDS vector and the transition matrix. By applying the Gauss Aitken generalized least squares method on the linear model, the BLUE of fresh incidences of HIV over time was re-estimated. However, the authors did not consider the double decrement pattern of HIV positive individuals viz the mortality of HIV positive individuals in the state of HIV and the transition of HIV affected individuals to the state of AIDS in the course of growth of HIV population.

The motivation of this paper is to extend the technique of Commences and Etcheverry (1993) by replacing the Weibull distribution. A meaningful alternative incubation period distribution that is used in this paper is based on a biological consideration on the reduction of CD-4 cells. Therefore, its validity is not for a short range. The Incubation period distribution is based on the assumption of a log-normal distribution of the number of CD-4 cells for HIV free individuals. A theoretical justification of this choice is given in the Appendix. Following Berman (1990), an exponential decay rate of CD-4 cells is chosen for HIV positive individuals, that leaves 5 to 10 percent of HIV patients not reaching the state of AIDS in 30 years in conformity with similar findings in Buchbinder, et.al (1993).

2 Development of the Model - A Diagrammatic Representation

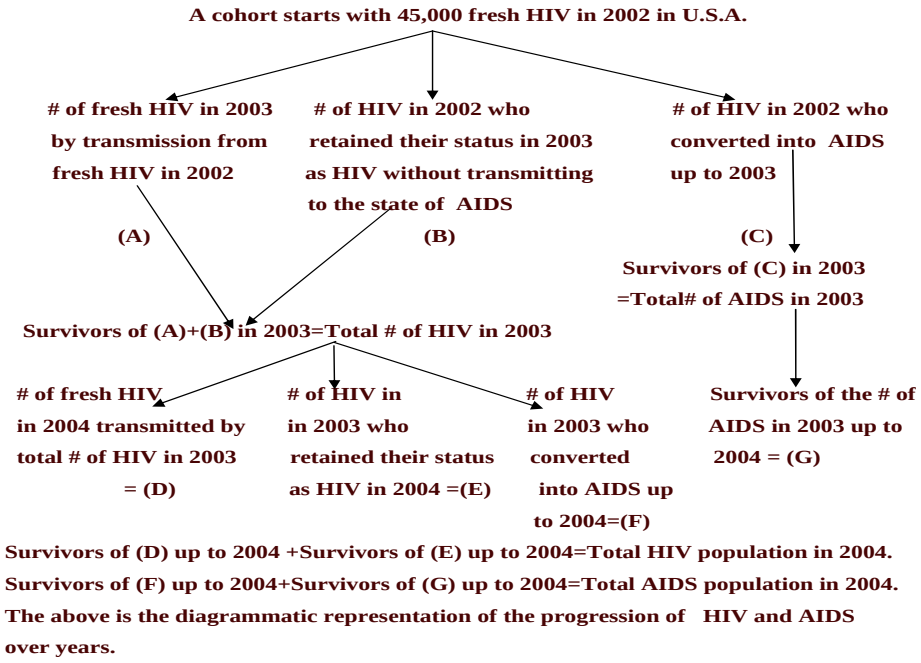

Figure 1: Diagrammatic representation of growth of HIV and AIDS over the years

3 Method of Obtaining Prior Distribution of HIV from Kendall's Birth and Death Process (1948)

Let $P_n(t)$ = Probability of n number of HIV in time t satisfying the Kolmogorv difference equation
$P_n(t+\delta t) = Pn(t)[1-n(\lambda(t)+\mu(t))\delta t+0(\delta t)]+P_{n-1}(t)[(n-1)\lambda(t)\delta t+o(\delta t)+P_{n+1}[(n+1)\lambda(t)\delta t + o(\delta t)]$
where $\lambda(t) = \lambda t^{\beta-1}$ and $\mu(t) = \mu t^{\alpha-1}$ are the Weibull birth and death intensities.
Then, we have the differential equation given by Kendall (1948).

$$\frac{\partial P_n(t)}{\partial t} = -n[\lambda(t) + \mu(t)]P_n(t) + \lambda(t)(n-1)P_{n-1}(t) + \mu(t)(n+1)P_{n+1}(t) \quad (1)$$

with the initial conditions $P_1(0) = 1, P_0(0) = 0, P_x(0) = 0$ for $n \neq 1$

$$P_0(t) = 0 \text{ for } t \neq 0,\ P_n(t) = 0 \text{ for } n < 0; \lambda = \lambda(t) \text{ and } \mu = \mu(t) \tag{2}$$

The solution if given by $$P_n(t) = \eta^{n-1}(1-\eta_t)(1-\xi_t) \tag{3}$$

where

$$\eta_t = 1 - \frac{1}{W(t)} \tag{4}$$

and

$$\xi_t = 1 - \frac{e^{-\rho(t)}}{W(t)} \tag{5}$$

$\rho(t)$ and $W(t)$ are given by

$$\rho(t) = \int_0^t [\mu(\tau) - \lambda(\tau)]d\tau \tag{6}$$

$$W(t) = 1 + e^{-\rho(t)} \int_0^t e^{\rho(\tau)}\lambda(\tau)d\tau \tag{7}$$

$$E(n_t) = \frac{(1-\xi_t)}{(1-\eta_t)} = e^{-\rho(t)} \tag{8}$$

$$Var(n_t) = \frac{(1-\xi_t)(\xi_t+\eta_t)}{(1-\eta_t)^2} \tag{9}$$

The Prior Distribution of HIV is given by (3) and the generation of HIV data have been obtained by estimating parameters $\mu(t)$, $\lambda(t)$, $\rho(t)$, $W(t)$, $\xi(t)$ and $\eta(t)$ by US Official Statistics data for 2002. The trend of the data for the future 16 year period being determined by taking random observations from two Uniform distributions (one from HIV infectivity and the other for Deaths) with stipulated maxima and minima based on the actual data subject to the condition that the survival function both from births and deaths are increasing over time.

3.1 Significance of the Parameters in Terms of HIV Population

$\lambda(t)$ and $\mu(t)$ are the intrinsic birth and death parameters respectively of an HIV population which starts initially with a single HIV. $\rho(t)$ is the negative of the growth parameter at any time t. $W(t)$ is the following function of $\rho(t)$ and $\lambda(t)$ (or $\lambda(t)$ or $\mu(t))W(t)$ is given by

$$W(t) = 1 + e^{-\rho(t)} \int_0^t e^{\rho(\tau)} \lambda(\tau) d\tau$$

Average size of the HIV population n_t at time t is given by $E(n_t) = \dfrac{1-\xi_t}{1-\eta_t}$ and $var(n_t) = \dfrac{(1-\xi_t)(\xi_t+\eta_t)}{(1-\eta_t)^2}$. Given ξ_t and η_t we can obtain the average size and the variance of HIV population at any time t where ξ_t and η_t are given by $\xi_t = 1 - \dfrac{e^{-\rho(t)}}{W(t)}$ and $\eta_t = \dfrac{1}{W(t)}$ and the distribution of n at any time t is given by $P_n(t) = \eta_t^{n-1} \times (1-\eta_t)(1-\xi_t)$.

4 Existing Data of HIV & AIDS in USA and in the World

Region	Epidemic Started	People living with HIV/AIDS	People newly infected with HIV 2002	% of HIV - positive adults who are women	Main modes of transmission (#) for adults living with HIV/AIDS
Western Europe	Late 70's early 80's	570,000	30,000	25%	MSM, IDU
North America	Late 70's early 80's	980,000	45,000	20%	MSM, IDU, Hetero
Australia & New Zealand	Late 70's early 80's	15,000	500	7%	MSM

Number of people living with HIV in 2002	
Women	19.2 Million
Men	19.4 Million
Children < 15 years	3.2 Million
Total	**42 Million**

During 2002, some 5 million people became infected with the human immunodeficiency virus (HIV) which causes AIDS.

Number of people infected with HIV in 2002	
Women	2 Million
Men	2.2 Million
Children < 15 years	800,000
Total	**5 Million**

(Source: UNAIDS /WHO: HIV and AIDS/ Global HIV and AIDS-Statistical Information and Tables 2002 - HTML Document.)

5 Estimation of the Stock of HIV and AIDS in 2002 in USA

The ratio of Number of People living with HIV and the number of people fresh infected is 42/5 in the world as a whole. Assuming the same ratio for USA, the estimated number of persons living with HIV is 42/5*45,000=378,000 in 2002.
Therefore, estimated number of persons living with AIDS in 2002 = 980,000 - 378,000 = 602,000.
Also, the estimated stock of HIV in 2002 = 378,000 + 45,000 = 423,000 in USA and the Estimated stock of AIDS in 2002 = 602,000 in USA.

6 Estimation of Birth and Death Intensities of HIV

The official birth rate of HIV is 4% per annum in 2002. However, assuming the birth rate lying between 4% to 6% and the corresponding survival Rates lying between 94 to 96 percent; subject to the condition that the survival rate increases randomly during 15 years (by generating 15 random observations from a uniform distribution with minimum = 0.94 and maximum = 0.96 and arranging them in increasing order), the Weibull birth and death intensities given by $\lambda(t) = \lambda t^{\beta-1}$ and $\mu(t) = \mu t^{\alpha-1}$ are estimated. We assume both to have a decreasing pattern. i.e. $0 < \alpha < 1$ and $0 < \beta < 1$. Then,

$$\rho(t) = \left[\frac{\mu t^{\alpha}}{\alpha} - \frac{\lambda t^{\beta}}{\beta}\right] \tag{10}$$

$$W(t) = e^{\frac{\lambda t^{\beta}}{\beta} - \frac{\mu t^{\alpha}}{\alpha}}\left[1 + \mu \int_0^t \tau^{\alpha-1} e^{\frac{\mu\tau^{\alpha}}{\alpha} - \frac{\lambda\tau^{\beta}}{\beta}} d\tau\right] \tag{11}$$

By the mean value Theorem

$$\cong e^{\frac{\lambda t^{\beta}}{\beta} - \frac{\mu t^{\alpha}}{\alpha}}\left[1 + \left(e^{\frac{\mu}{\alpha}(\frac{t}{2})^{\alpha} - \frac{\lambda}{\beta}(\frac{t}{2})^{\beta}}\right)\left(\frac{\mu t^{\alpha}}{\alpha}\right)\right] \tag{12}$$

$$\xi(t) = 1 - \frac{e^{-\rho(t)}}{W(t)} \tag{13}$$

$$\eta(t) = 1 - \frac{1}{W(t)} \tag{14}$$

$$E[n(t)] = e^{-\rho(t)} = \frac{(1 - \xi(t))}{(1 - \eta(t))} \tag{15}$$

$$var(n(t)) = \frac{(1-\xi(t))(\xi(t)+\eta(t))}{(1-\eta(t))^2} \tag{16}$$

by rank regression method, we have, i.e. by regressing $\log(-\log R(t))$ on $\log t$; where $R(t) = \exp\left(\frac{-\lambda t^\beta}{\beta}\right)$ for birth survival function

$= \exp\left(\frac{-\mu t^\alpha}{\alpha}\right)$ for death survival function with corresponding Weibull Hazard rates

$\lambda(t) = \lambda t^{\beta-1}$ and $\mu(t) = \mu t^{\alpha-1}$ respectively we have estimated the parameters as $\widehat{\lambda} = .0052$, $\widehat{\beta} = .125$, $\widehat{\mu} = .0050$ and $\widehat{\alpha} = .0469$ based on Official Data of US (2002): (Sources : US Census (2000), (ii) HIV and AIDS / Global HIV and AIDS - Statistical Information and Tables 2002, (iii) Tai Webster - Emory Report, April 7, 2004.

In view of the fact, that both the birth and death rates of HIV are underestimated; the former because of lack of reporting and the latter due to the fact that majority of deaths due to HIV are ascribed to Opportunistic infections (like Hepatitis - 2, Cardio -respiratory failure, Tuberculosis, Cancer etc.), we have assumed (with minor modification of the figures reported in the above sources) that: (i) mortality rate (adjusted for opportunistic infection) is 8 to 10 percent with the further assumption that with in the ranges the rates decrease with time. Data for the same has been generated from uniform distribution with (min = 0.94, max = 0.96) for survival probabilities from HIV infection and deaths per annum respectively after sorting 15 observations from each in increasing order, as survival from HIV infection and survival from death due to HIV (and other opportunistic infections) are assumed to be increasing functions of time.

7 Estimation of Trend of Infectivity and Death Rates and of HIV in 16 Years Based on the above Data

Infectivity rate for HIV at time t $(t = 1, 2, \cdots, 15)$ is given by $\lambda(t) = \lambda t^{\beta-1}$. Substituting $\widehat{\lambda} = .0052$ and $\widehat{\beta} = 0.125$, we get the infectivity rates for 15 years which are graphically y shown in figure 2.

Remark: The above pattern in the trend in the infectivity is same as that of Commenges And Etcheverry (1993).

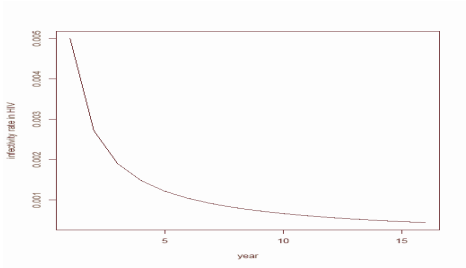

Figure 2: Trend in the infectivity rates in HIV

Similarly the trend in the HIV death rate is given by Weibull intensities $\mu(t) = \mu t^{\alpha-1}$ for $t = 1, 2, \cdots, 15$. On substitution of $\widehat{\mu} = .0050$ and $\widehat{\alpha} = .0469$, we get the trend in the HIV death rates for 15 years which are shown in figure 3.

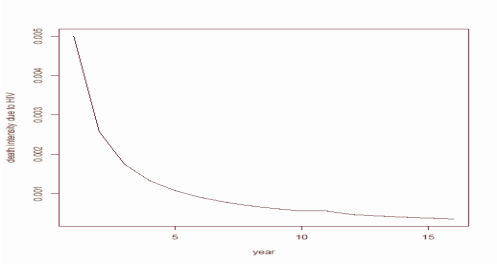

Figure 3: Trend in the HIV death intensity for 16 years

8 Estimation of the Pattern of HIV Incidences During 16 Years that Start with One HIV Positive Individual

On substitution of the estimated values of $\lambda(t)$ and $\mu(t)$ in $\rho(t)$, $W(t)$, $\xi(t)$ and $\eta(t)$ in (10), (12), (13) and (14) respectively and assuming infection process starts with 1 HIV positive individual, we have

$$E[n(t)] = \frac{1 - \xi(t)}{1 - \eta(t)} \tag{17}$$

from (17) which is the mean of the prior distribution of HIV population for $t = 0, 1, 2, \cdots, 15$ that started with one infected HIV individual. This is given in Table 1.

Table 1: HIV Population during 16 Years from an Epidemic that Started with 1 HIV Positive Individual.

0	-	1.0000000	8	-	1.0578230
1	-	0.9850563	9	-	1.0497230
2	-	1.0512223	10	-	1.0495658
3	-	1.0509672	11	-	1.0494189
4	-	1.0507101	12	-	1.0492789
5	-	1.0504770	13	-	1.0491469
6	-	1.0502645	14	-	1.0490210
7	-	1.0500690	15	-	1.0489008

The above figures have only been adjusted for the deaths of HIV in the state of HIV.

However, to estimate the HIV after elimination of those who pass to AIDS, we require the transition probability of passing to AIDS; for that we require the conversion rate to AIDS per year which is obtainable from the Incubation period distribution. This method is given in the next section.

9 Method of Obtaining $H = [q_{ji}]$

We denote, $E[X(t)] = \mu$, $var(X(t)) = \sigma^2$ and decay rate δ of CD-4 cell a sample of intravenous drug users in New York city. More specifically, let $X(t)$ be the # of T-4 cells in an HIV free individual. Obviously, $X(t)$ is a Stationary Stochastic Process. Denoting $Z(t) = \log(X(t))$ and assuming $Z(t)$ to be a Gaussian Process, if we consider the decay of T-4 cells of HIV individuals infected at time zero by an exponential rate δ so that the # of T-4 cells at time t for an HIV individual is $X(t)e^{-\delta t}$, then denoting the logarithm of the # of T-4 cells as $R(t)$, we have by denoting the time of detection as HIV positive as T, $[R(t)-R(t+t_1)]/t_1$ taken over all HIV positive individuals provides an estimate of δ, the decay cell parameter. If we consider the decay of T-4 cells of HIV individuals infected at time zero by an exponential rate δ so that the number of T-4 cells at any time t for an HIV positive individual is $X(t)e^{-\delta t} = e^{Z(t)-\delta t}$.

Let us denote the logarithm of the T-4 cells at time t by $R(t)$. Denoting, $R(t)$ as the number of T-4 cells in logarithmic scale at time t and $[R(t) - R(t + t_1)]/t_1$ for $t_1 > 0$ taken over all HIV individuals provides an estimate of δ. (the average change per unit time of $\log(T-4)$ between the first and second visits), the growth/decay rates of T-4 cells. Further, denting T as the duration of incubation period, we have,

$$\frac{R(T) - \mu}{\sigma} = \left[\left(\frac{Z(T) - \mu}{\sigma}\right) - \frac{\delta T}{\sigma}\right].$$

Therefore,

$$P\left[\frac{R(T) - \mu}{\sigma} = x \;\middle|\; \frac{\delta T}{\sigma} = t\right] \Rightarrow P\left[\frac{Z(T) - \mu}{\sigma} = x + t \;\middle|\; \frac{\delta T}{\sigma} = t\right] = \phi(x + t)$$

Since $\left[\dfrac{Z(T)-\mu}{\sigma}\right]$ is $N(0,1)$.

Therefore, the joint distribution of $\left[\dfrac{R(T)-\mu}{\sigma} = x,\ \dfrac{\delta T}{\sigma} = t\right]$ is $\phi(x+t)q(t)$, provided the random variables X and T are independent; where $\phi(z)$ is the ordinate of a standard normal variate at z and $q_T(t)$ is the distribution of T.

Following Berman (1993) and taking $q_T(t)$ to be exponential, the conditional density of $\dfrac{\delta T}{\sigma} = t$ given $\left[\dfrac{R(T)-\mu}{\sigma} = x\right]$ is given by $\phi(x+t)q(t)\left[\int\limits_0^\infty \phi(x+t)q(t)dt\right]^{-1}$.

Now,

$$\phi(x+t)q(t) = \frac{1}{\sqrt{2\pi}}\ \frac{1}{m}\exp\left(-\frac{1}{2}(x+t)^2\right)\exp\left(-\ \frac{1}{m}t\right)$$

Therefore,

$$\int\limits_0^\infty \frac{1}{\sqrt{2\pi}}\ \frac{1}{m}\ \left(\frac{-1}{2}(x+t+m^{-1})^2\right)e^{\frac{1}{2m^2}}e^{\frac{1}{2m}x}dt = \frac{1}{m}e^{\frac{1}{2m^2}}e^{\frac{1}{2m}x}\frac{1}{\sqrt{2\pi}}\int\limits_{x+m^{-1}}^\infty e^{\frac{-1}{2}\xi^2}d\xi$$

Therefore,

$$\frac{\phi(x+t)q(t)}{\int\limits_0^\infty \phi(x+t)q(t)dt} = \frac{e^{\frac{-1}{2}(x+t+m^{-1})^2}}{1-\frac{1}{\sqrt{2\pi}}\int\limits_\infty^{x+m^{-1}} e^{\frac{-1}{2}\xi^2}d\xi} \tag{18}$$

Thus,

$$P\left[\frac{\delta T}{\sigma} = t\middle|\frac{R(T)-\mu}{\sigma} = x\right] = \frac{e^{\frac{-1}{2}(x+t+m^{-1})^2}}{1-\frac{1}{\sqrt{2\pi}}\int\limits_\infty^{x+m^{-1}} e^{\frac{-1}{2}\xi^2}d\xi} = \frac{dnorm(x+t+m^{-1})}{1-pnorm(x+t+m^{-1})} \tag{19}$$

using $S-$ plus / R commands.

10 Estimation of the Parameters

Since X is log normally distributed with parameters μ and σ^2, therefore, $E(X) = \exp(\mu+\sigma^2/2)$ and $var(X) = \exp(2\mu+\sigma^2)(\exp(\sigma^2-1))$ $\log E(X) = \log(600 * 10\hat{}9 = 27.0202$ and assuming the $\sigma/\mu = 0.354/6.966 = 0.05281826$ (Berman) we get the estimates of μ and σ from the quadratic equation $\mu^2(.001241248) + \mu - 27.0202 = 0$ which gives $\widehat{\mu} = 26.17124$ and $\widehat{\sigma} = 1.423800$.

The decay rate of CD-4 cells is estimated by assuming that in 30 years only 5% cells are not destroyed. This gives $\hat{\delta} = 0.099874$. Also the estimated exponential parameter, $m^{-1} = 0.421$ (Berman).
as an example, taking T, the stopping time (i.e., the incubation period as 5 years) $R(T)|_{T=5} = \log(600 * 10\hat{}9 * exp(-.9985774) * 5) = 28.62978)$, we get

$$\frac{R(5) - \mu}{\sigma} = \frac{28.62978 - 26.17124}{1.4238} = 1.725341 = x$$

$$P\left[\frac{\delta 5}{\sigma} = \frac{09985774 * 5}{1.4238} = t \middle| \frac{R(T) - \mu}{\sigma} = 1.725341 = x\right]_{T=5} = \frac{dnorm(x + t + m^{-1})}{1 - pnorm(x + t + m^{-1})} = 0.9575928$$

However, noting that (19) is not obtained from a proper probability distribution, the corresponding adjusted probability is given by

$$\frac{P\left[\frac{\delta 5}{\sigma} = \frac{09985774*5}{1.4238} = .3507304 = t \middle| \frac{R(5)-\mu}{\sigma} = 1.7531402 = x\right]}{\sum_{T=0}^{15} P\left[\frac{\delta T}{\sigma} = t \middle| \frac{R(T)-\mu}{\sigma} = x\right]} = \frac{0.9575928}{11.74305} = 0.0816 \tag{20}$$

Thus the probability distribution of incubation period of AIDS within a range of 15 years of incubation period is given in the following table 2 as follows:

Table 2: Probability Distribution of Incubation Period.

Year	Incubation Probability	Year	Incubation Probability
0	.0075	8	.0628
1	.1217	9	.0580
2	.1019	10	.0538
3	.0989	11	.0500
4	.0896	12	.0437
5	.0816	13	.0411
6	.0745	14	.0387
7	.0683	15	.0079

The distribution is graphically represented in figure 4.
Remarks: The graph in the figure 4 may be indicative of heterogeneity in the distribution of Incubation period of AIDS; possibly, because of genetic co-factors, a class of HIV infected individuals progress at significantly slow pace towards AIDS than the rest of the HIV population as revealed by many earlier investigators like Buchbinder et al (1993), Dywer. John. M (1995) etc. Buchbinder et. al (1993) maintain that 5% of the HIV affected individuals do not reach the state of AIDS even in 30 years.

However, since our study is limited to the extent of incubation period of only 15 years, therefore, we do not have enough data to support this hypothesis. However, this important aspect of Incubation period requires further probing. The traditional Weibull distribution considered by many of the earlier investigators is incapable of highlighting more facts on this issue.

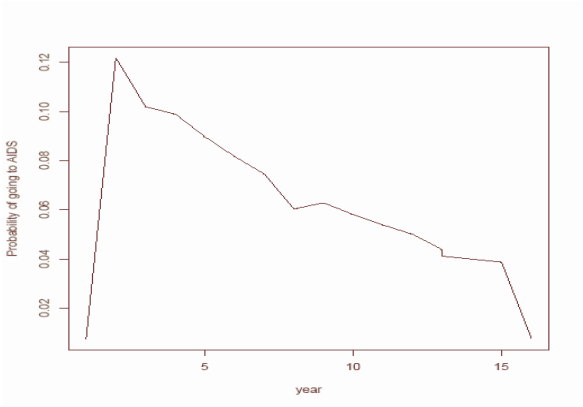

Figure 4: Probability Distribution of Incubation period for 15 year period. X = year, Y = Probability of Incubation period (in years)

11 Construction of Transition Matrix

Defining p_{ij} as the probability of passing to the state of AIDS in the j^{th} year given that the HIV positivity occurred in the i^{th} year ($i, j = 0, 1, 2, \cdots, 15$ and $j > i$), The Transition matrix $[H] = [p_{ji}]$ for $i, j = 0, 1, 2, \cdots, 15$ for $j > i$ has been constructed from table 4 by considering the homogeneity of the process i.e.

$$p_{ji} = p_{j-i}\theta \tag{21}$$

Given the transition Matrix the probability of transiting from HIV to AIDS in the adjusted HIV vector is obtained from unadjusted HIV vector by multiplying each entry in the column of the preceding table by $(1 - p_{00})$, since $p_{ii} = p_{00}$.

Table 3: Adjusted HIV Incidences (Adjusted for Conversion to AIDS) for an Epidemic that Started with One HIV Affected Individual.

Year	HIV population	Year	HIV population
0	0.992500	8	1.057823
1	1.05917	9	1.057823
2	1.058909	10	1.057349
3	1.058650	11	1.057208
4	1.058415	12	1.057075
5	1.058201	13	1.056948
6	1.058004	14	1.056827
7	1.057823	15	1.056710

Table 4: Adjusted AIDS Population during 16 Years.

Year	HIV population	Year	HIV population
0	0.00744375	8	0.74383136
1	0.12873099	9	0.80547326
2	0.23797804	10	0.86257142
3	0.34286254	11	0.91564700
4	0.43833960	12	0.92663900
5	0.52516591	13	0.97038726
6	0.60442455	14	1.01390833
7	0.67704589	15	1.05491063

12 Variance-Covariance Matrix of HIV Vector (V)

Defining p_{ji} = probability of going to the state of AIDS in j^{th} year given that HIV positivity occurs in the i^{th} year $j \geq i$; $i = 0, 1, 2, \cdots, 15$.

Since, $H = [p_{ji}] = [p_{j-i,0}] = 1$ assuming the homogeneity of the process, therefore

$$\sum_{j=0}^{15} p_{ji} = \sum_{j=0}^{15} p_{j-i,0} = 1 \qquad \forall \qquad i = 0, 1, 2, \cdots, 15$$

and

$$\sum_{i} (y_j | x_i) = y_j; \qquad i, j = 0, 1, 2, \cdots, 15 \tag{22}$$

Therefore, the distribution of $Y = (y_0, y_1, y_2, \cdots, y_{15})$ is multinomial. The variance covariance matrix of Y can therefore be given as follows:

$$Var(X_k) = \sum_{i=0}^{k} x_i p_{0\ k-i}(1 - p_{0\ k-i}) \tag{23}$$

$$Covar(X_j, X_k) = \sum_{i=0}^{\min(j,k)} -[x_i, p_{0\ j-i} p_{0\ k-i}] \tag{24}$$

Since $p_{ji} = p_{j-i\ 0}$ because of the homogeneity of the process.

13 Methodology to Obtain the Best Linear Unbiased Estimator of the HIV Vector

Let, $X = (x_1, x_2, \cdots, x_{15})^t$ be the HIV vector of 15 observations. Since, no reliable data are available, we have built up an empirical prior distribution of the same as discussed in the preceding section. Similarly, let, $Y = (y_1, y_2, \cdots, y_{15})^t$ be the corresponding AIDS vector; and $E(Y|X) = HX$ with the dispersion matrix

$$D = D(Y|X) = B(X) \cdots (25) = B, \qquad \text{say} \qquad \Rightarrow E(Y) = H\overline{X} \cdots \tag{25}$$

$$var(Y) = E(Var(Y|X) + var(E(Y|X)) \tag{26}$$

Therefore,

$$V = \overline{B} + HS\ H^t \qquad \text{where} \qquad S = var(X) \tag{27}$$

Then,

$$\widehat{X} = \text{ BLUE of } X = (H^t V^{-1} H)^{-1} H^t V^{-1} Y = S(H^t V^{-1} Y) \text{ where } S = (H^t V^{-1} H)^{-1} \tag{28}$$

$\overline{X} = SH^t V^{-1}(Y - H\overline{X}) \cdots\cdots$

$$\Rightarrow \widehat{X} - \overline{X} = SH^t V^{-1}(Y - H\overline{X}) \tag{29}$$

is the regression of X and Y.

BLUE of HIV=ginv(t(H)%%ginv(V)%*% (H))%*%H%*%ginv(V)%*%YX.*

Table 5: HIV Population during 16 Years that Start with a Single Infected HIV Individual.

Year	HIV population	Year	HIV population
0	1.05070854	8	1.0400300
1	1.09884727	9	1.0609486
2	1.10092763	10	1.1607088
3	1.14450123	11	1.2376678
4	1.21711091	12	1.0111850
5	1.11740764	13	0.9823875
6	1.00262278	14	0.9449600
7	1.02561700	15	0.8988975

14 Estimates of HIV (2002-16) in USA (Adjusted for Transition to AIDS)

The existing total HIV Population is 423,000. Using the same we have given in tables 6 and 7 the estimated HIV population in USA from 2002-2016, on the basis of generated HIV population as well as BLUE of the HIV that started with one HIV affected individual (subject to adjustment for AIDS).

Table 6: Estimated HIV Population in USA from 2002-2015.

0	419827.5=2002*	8	447388.0=2009
1	448027.1=2003	9	447321.3=2010
2	447918.5=2004	10	447258.5=2011
3	447809.1=2005	11	447199.1=2012
4	447709.4=2005	12	447142.8=2013
5	447618.8=2006	13	447089.0=2014
6	447535.9=2007	14	447037.7=2015
7	447459.3=2008	15	446988.5=2016

* (adjusted for 1 year conversion to AIDS)

Table 7: Estimated HIV Population in USA during 2002-2015 Based on BLUE of HIV Population.

Formula: 423000 BlueX= BLUE of HIV from 2002-15 in USA.*

2002	444449.7	2009	433836.0
2003	464812.4	2010	439932.7
2004	465692.4	2011	448781.2
2005	484124.0	2012	523533.5
2006	514837.9	2013	427731.3
2007	472663.4	2014	415549.9
2008	424109.4	2015	399718.1

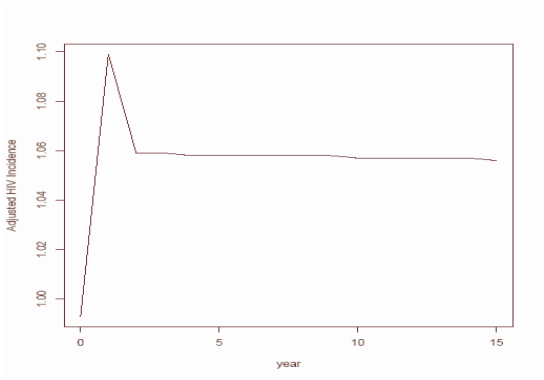

Figure 5: HIV population (adjusted for elimination into AIDS) for 16 years that Started with a single HIV affected individual

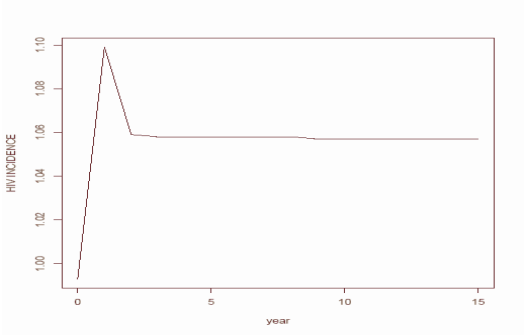

Figure 6: Best linear estimates (BLUE) of HIV population during 16 years that started with a single HIV affected individual

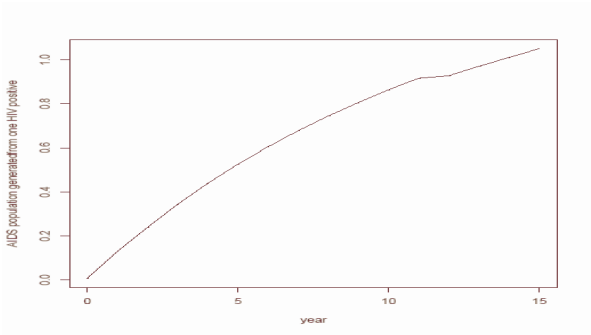

Figure 7: AIDS population during 16 years that was generated from a single affected Individual

15 Some Evidences about the Preference of Log-normal to Weibull as HIV Distribution for Incubation Period of AIDS

For the choice of models to describe the incubation for AIDS (especially between Weibull and log-normal), Munoz and Xu (1996) report that there are empirical evidences that, although the hazards of AIDS for an HIV seropositive increase. During the first six to eight years from sero-conversion, the curvature changes and the hazard function flattens out past 10 years from sero conversion. The attenuation of the hazard is most consonant with log normal model of HIV being an appropriate distribution of incubation period of AIDS rather than traditional Weibull. To exhibit the same the authors consider the hazard functions of three parameter. Logistic model, Weibull and log-normal models which are given as follows:

$$\text{log normal: } h_L(t) = \frac{1}{1-\phi\left(\frac{\log t-\beta}{\alpha}\right)} \cdot \frac{1}{\sqrt{2\pi}}\ \frac{1}{\alpha t} \exp\left[\frac{-1}{2}\left(\frac{\log t-\beta}{\alpha}\right)^2\right]$$

$$\text{Weibull : } h_W(t) = \frac{1}{\alpha t}(\exp(-\beta t)^{1-\alpha})$$

$$\text{Three parameter logistic : } h(t) = \frac{1}{1+\exp(-\beta_0+\beta_1(\log t)+\beta_2(\log t)^2)}$$

They further point out that the three parameter logistic model under certain conditions of β_0, β_1 and β_2 can take the shape of log-normal, Weibull and none of the either two also. Therefore, in consideration of Weibull and log normal being particular cases of three parameter logistic (which has been considered as a standard distribution for

the comparison of Weibull and log normal), they considered the proximities of logistic and Weibull to three parameter logistic respectively by their respective deviances from the three parameter logistic by the method of likelihood ratio test. The full data collected by the authors were utilized and the analysis has been carried on 10 times using a different set of random draws from the imputed times from sero-conversion to entry for the sero-prevalent cohort. For details of collection of data one may refer Munoz and Xu (1996). The results reproduced from Table II, page 2464 (Munoz and Xu (1996) are shown in the following table 8.

Table 8: Values of - 2 log λ Likelihood to Compare Three Parametric Models Using the Full Data on Ten Imputations.

	Models		
Imputation	-2log λ likelihood of three parameter logistic	Deviances in respect of three Parameter logistic model for lognormal	Deviances in resect of three Parameter Logistic Model for weibull
1	7822.62	4.90	35.03
2	7797.54	13.15	22.89
3	7823.60	6.03	28.60
4	7799.96	2.46	35.24
5	7812.92	4.68	32.00
6	7850.61	3.23	34.90
7	7818.66	2.16	40.76
8	7827.41	7.74	26.21
9	7874.45	11.65	16.32
10	7801.16	5.61	33.46

The table shows in all cases the deviance of log-normal model from the three parameter Logistic is lower than the corresponding deviance of the Weibull model and the latter significantly differs from the three parameter logistic model in all the ten cases. This Indicates that for the HIV distribution for the incubation of AIDS, the log-normal provides a much better fit in the context of three parameter logistic family. Finally, to corroborate our conjecture that log-normal gives a better representation of HIV distribution than Weibull, we again reproduce a second findings of Munoz and Xu (1996) relating to the HIV distribution by years (i.e. estimates of percentage free of AIDS after sero conversion in years) from table III, page 2465 of the same paper. That shows the relative performance of log-normal and Weibull in representing the HIV distribution while comparing them with standard distributions like three parameter. Logistic and Kaplan Meier (which is also considered in the literatures a fairly good Non-parametric model for describing HIV distribution especially for censored data).

Table 9: Estimates of Percentage Free of AIDS after HIV Transmission in Years.

Years from Sero conversion	Kaplan Meier (%)	*Three parameter logistic (%)	+ Lognormal (%)	+Weibull (%)
2	97.7	97.9	98.7	88.1
4	87.7	87.7	88.6	73.4
6	72.8	72.7	72.9	59.4
8	57.3	57.6	57.5	47.0
10	44.3	44.6	44.5	36.6
12	33.9	34.1	34.3	28.1
14	25.8	26.0	26.4	21.3
16		19.9	20.4	15.9
18		15.3	15.9	11.8
20		11.8	12.4	8.6

* Based on average of 10 imputations,
+ based on average observations over 5 years

A comparison of the observations of Kaplan Meier (based on full observations), Log normal and Weibull (left truncated at 5 years) clearly shows that the simple two parameter log normal model provides good fit as three parameter logistic model. In contrast, the Weibull model provides a very poor fit considerably underestimating the survival functions at all times.

16 Conclusion

We have attempted to make an innovative application of D. G. Kendall's generalized Birth and death process by obtaining a Prior distribution of HIV from the same. This distribution is supposed to be real as it can be utilized to take into account of the mortality of HIV patients in the state of HIV; as well as because of the fact that the parameters of the distribution have been estimated on the basis of available Statistical data of USA. The trend of HIV infectivity curve obtained by us compares fairly well with that of Commenges and Etcheverry. Whereas the trend in the HIV mortality curve cannot be compared with any other earlier investigators as HIV mortality, which is quite significant has not been considered by any earlier worker to the best of the knowledge of the Authors.

In view of this, the informative prior obtained by us is considered better than any other empirical prior that is developed on the basis of only growth behavior of HIV without considering the mortality pattern of HIV; which is sometimes quite high because of several opportunistic infections. While taking into account of the same, we have assumed little flexibility on the official Statistics figures in our analysis. Secondly,

departing from the traditional line of back projection of HIV from AIDS data, we have evolved an incubation period distribution with biological orientation based on the premises of progressive reduction of CD-4 cells; as and when an HIV positive individual transits towards full blown AIDS state. This kind of Incubation period distribution is considered to produce better correspondence with real figures than an inflexible Weibull distribution established by Munoz and Xu (1996) which may not be quite suitable for long term projection of AIDS; as the estimates of the parameters of the distribution undergo change during long time range. One important finding in our exercise is that the trend in the generated data of HIV from informative prior based on Kendall's birth and death process shows the same pattern as that of their BLUE obtained by Gauss-Aitken least squares method which produce maximum correlation between two consecutive observations of HIV incidences. Further work in this line can fruitfully be generalized with the break up of HIV data for several identifiable social classes who are more vulnerable to HIV and AIDS; that is very much applicable in the preventive and Action research program of HIV and AIDS.

Acknowledgements

The authors are grateful to anonymous referee for his valuable suggestions to improve the paper. They are also thankful to Dr. Pankaj Choudhary, Department of Mathematical Sciences, University of Texas at Dallas for his help in the preparation of the paper.

References

[1] Becker Niels G. (1989): A model of the spread of HIV infection and the Demographic impact of AIDS. *Statistics in Medicine, 8, 103-120.*

[2] Becker Niels G., Watson Lyndsey, F. and Karl, John B. (1991): A Method of Non-parametric Back projection and its application to AIDS data. *Statistics in Medicine, 10, 1527-1542.*

[3] Berman, Simon, M. (1990): A Stochastic Model for the distribution of HIV latency time based on T-4 counts. *Biometrika, 77(4), 733-741.*

[4] Boullion Thomas, L. and Odell Patrick, L. (1971): Generalized Inverse Matrices, *Wiley Inter Science.*

[5] Brookmeyer, R. and Gail, M. H. (1988): A method for obtaining short-term projections and lower bounds on the size of the AIDS epidemic. *J.Am.Stat.Assn, 83, 301-30.*

[6] Biswas, Suddhendu (1995): Applied Stochastic Processes. *John Wiley & Sons, New York.*

[7] Biswas, Suddhendu and Thomas, Mariamma (1995): On a continuous Branching Process and Birth, Death and Immigration Models for testing the efficacy of some Control Programmes for AIDS. *Biometrical Journal, 37, 235-249.*

[8] Biswas, Suddhendu and Gurung Amar, B. (2002): On an E-M Algorithm technique in making back projection from AIDS data. *Progress in Mathematics, 36 (1 & 2), 2, 221-2.*

[9] Biswas, Suddhendu, Grover, Gurprit and Gurung Amar Bahadur (2002): On the estimation of HIV mortality rate and evolution of the distribution of Incubation Period of AIDS. Biostatistical Aspects of Health and Population, Arvind Pandey (ed.), *Hindustan Publishing Corporation, Delhi, 131-141.*

[10] Buchbinder, Susan, Mann D., Lovie, L., et al (1993): Healthy Long term Positives; Genetic cofactors for delayed HIV disease progress, *IX International Conference on AIDS, Berlin, Germany, WS-BO3-2.*

[11] Commenges D. and Etcheverry B. (1993): An Empirical Bayes Approach to estimation of the incidence curve of HIV infection. *Statistics in Medicine, 12, 1317-1324.*

[12] Ding, Ye (1995): Computing back calculation estimates of AIDS epidemic. *Statistics in Medicine, 14, 1505-1512.*

[13] Dwyer John M. (1995): Biomedical Aspects of HIV and AIDS. *Current Science, 59(10), 823-827.*

[14] Kendall D. G. (1948): A generalised birth and death Process, *Annals of Mathematical Statistics, Vol. 19, 1-15.*

[15] Munoz Alvaro and XU Jane (1996): Models for Incubation period of AIDS and variations according to Age and Period, *Statistics in Medicine, Vol. 15, 2459-73.*

[16] Rees, M. (1989): The somber view of AIDS, *Nature, 326, 343-345.*

[17] Smith James Munroe (1996): AIDS and Society, *Prentice Hall, Inc. Simon and Schuster/AViacom Company, Upper Saddle River, New Jersey 07458.*

[18] Thomas, M., Biswas, Suddhendu, Sultana Rehena and Pandey Arvind (2003): Stochastic Models to study the Dynamics for HIV/AIDS. An Analytical Review, Biostatistical Aspects of Health and Population, Arvind Pandey (ed.), *Hindustan Publishing Corporation, Delhi, 106-131.*

[19] UNAIDS/WHO (2002): Global Statistical Information and Tables Epidemic-Annex, Country Profiles (http://www.journals.org/statistics.php♯globalstats).

[20] UNAIDS (2006): Report on Global Epidemic Annex2, Country Profiles (http://www.journals.org/statistics.php♯globalstats).

[21] Webster Tia (2004): US HIV transmission rate at roughly 4%, *Emory Report, April 7, 2004.*

[22] www.wrongdiagnosis.com/diagnosis/symptom : Statistics about HIV and AIDS.

Appendix

Let $X_{\nu+1}$ be the ♯ of CD-4 cells at an instant ν, we assume $X_{\nu+1} - X_\nu = \xi_{\nu+1} \cdot g(X_\nu)$ where $\xi_{\nu+1}$ is an impulse working at $\nu + 1$ and $g(X_\nu)$ is a function of X_ν.

$$\Rightarrow \frac{X_{\nu+1} - X_\nu}{g(X_\nu)} = \xi_{\nu+1} \Rightarrow \sum_{\nu=0}^{n-1} \frac{X_{\nu+1} - X_\nu}{g(X_\nu)} = \sum_{\nu=0}^{n-1} \xi_{\nu+1} \Rightarrow \sum_{\nu=0}^{n-1} \xi_{\nu+1} \sim \int_{X_0}^{X_n} \frac{dt}{g(t)}$$

Note that by Central limit theorem, since $\sum_{\nu=0}^{n-1} \xi_{\nu+1}$ is the sum of several independent randon variables, therefore, $\int_{X_0}^{X_n} \frac{dt}{g(t)}$ is normally distributed.

We have assumed $g(t) = t \Rightarrow \log \frac{X_n}{X_0}$; which proves that $\log X_n$ is normally distributed. This establishes the justification of the distribution of CD-4 cells as lognormal.