

## A SIMULATION STUDY ON UNCERTAINTIES ASSOCIATED WITH BACKCALCULATION METHODOLOGY

RAVANAN.R <sup>1</sup>, VENKATESAN.P <sup>2</sup>

<sup>1</sup>Reader in Statistics, Presidency College, Chennai

<sup>2</sup>Deputy Director, TRC, ICMR, Chennai

### Abstract

The Backcalculation methodology depends on correct information about Incubation period distribution, Infection density and AIDS incidence. The uncertainties in any of the above three components results in the uncertainty of HIV/AIDS estimates obtained by Backcalculation methodology. The incubation period of HIV is very long and is highly variable within and between groups and also sensitive to the choice of incubation distribution. The infection density also adds to the uncertainty of the projected HIV/AIDS estimates. Most of the time, the non homogeneous Poisson process assumption for HIV infection curve is violated. The backcalculation also heavily depends on AIDS incidence data over a period of time. The reporting delay and under-reporting further adds to the uncertainties. This paper attempts using simulation techniques to examine the variations associated with backcalculation estimates for several key parameters of HIV/AIDS models. This study also attempts to quantify the extend of uncertainties due to variations in the incubation period and infection curve.

**Key words:** Backclaculation Methodology – Uncertainnity – Incubation Period and Infection Curve.

### I. INTRODUCTION

Monte-Carlo Simulation is one of the standard tools for comparing and validating statistical methodologies. In this paper an attempt has been made to use simulation techniques to examine the variation associated with the backcalculation estimates of several key parameters of HIV/AIDS. As already mentioned in the previous chapters the backcalculation methodology has certain uncertainties associated it. Through the simulation study, an attempt has been made to quantify the extent of uncertainties due to the variation in the incubation period distributions and infection curves. Several incubation period distributions and infection curves have been used in the simulation study, which were described in the previous chapters and the results are reported with variation associated with some key parameters of the backcalculation estimates of HIV/AIDS.

### II. GENERATION OF AIDS DATA

One of the main objectives of the present thesis is to give backcalculation estimates of some key disease measures of HIV/AIDS in India. The first step in this study has been designed to simulate the data structure of Indian Surveillance databases.

Following the approach of Brookmeyer and Gail (1988) and Ding (1996) we have assumed a multinomial distribution with unknown sample size for the reported AIDS cases. The reported AIDS cases were generated from this multinomial distribution. The likelihood of the data under multinomial distribution is given by

$$L(N, \theta) = L_1(N, P_{L+1}(\theta)) L_2(\theta) \quad [1]$$

Where

$$L_1(N, P_{L+1}(\theta)) = \frac{N!}{n!(N-n)!} [P_{L+1}(\theta)]^{N-n} [1 - P_{L+1}(\theta)]^n$$

and

$$L_2(\theta) = n! \prod_{j=1}^L \frac{[q_j(\theta)]^{x_j}}{x_j!}$$

with

$$q_j(\theta) = \frac{P_j(\theta)}{1 - P_{L+1}(\theta)} = \frac{P_j(\theta)}{\sum_{i=1}^L P_i(\theta)}$$

The likelihood function given in [1] is the conditional likelihood equation. Here  $X_j$  denote the number of AIDS cases reported in the interval  $[T_{j-1}, T_j]$ ,  $j = 1, 2, \dots, L$ . and  $X_{L+1}$  is the number of individuals infected before  $T_L$  who have not become AIDS by time  $T_L$ . One of the problems here is to estimate the total number of infection  $N$  before the time  $T_L$ . This number is the total number of AIDS cases that may develop from the number of infected individuals infected before  $T_L$ . We assume that all who have infected with HIV will ultimately become AIDS cases. For simulation of data under this model it is assumed that all the parameters of the model in equation [1] be known. According to NACO(2002) report by the end of 2002 around 4 million people are living with HIV infections. For the simulation study we have assumed  $N$  to be 4 millions and in the next step we simulate the AIDS cases  $X_j$ . The start of the infection in India is assumed to be 1981 and therefore  $T_0$  is taken to be January 1981. The reported AIDS cases are available from 1986 ( $T_1 = 1986$ ) to December 2002 ( $T_L = 2002$ ). There are 17 time points (yearly) at which yearly AIDS cases are available for India.

For simulating  $X_j$ , the probability of becoming AIDS in the  $j^{th}$  interval is given by

$$p_j(\theta) = \frac{1}{N} \int_{T_0}^{T_j} I(\theta, t) [F(T_j - t) - F(T_{j-1} - t)] dt \quad [2]$$

must be known. Calculation of these probabilities needs the infection curve and infection period distribution is known. Hence in our simulation study, we have assumed the incubation period distribution as the Weibull distribution with median 10 and 15 years based on various reports. The Weibull model given by Brookmeyer and Goedert (1989) is  $F(t) = 1 - \exp(-0.0021t^{2.516})$ . Based on this model median incubation period with 15 years is  $F(t) = 1 - \exp(-0.000762t^{2.516})$ . This Weibull distribution has been used as incubation period model because this model has been extensively used in the literature. The infection density  $I(\theta, t)$  is assumed to be logistic prevalence model given by

$$I(\theta, t) = \frac{\theta_1 \theta_3 e^{\theta_2 + \theta_3 t}}{k(1 + e^{\theta_2 + \theta_3 t})^2} \quad [3]$$

where,  $\theta_1$ ,  $\theta_2$  and  $\theta_3$  are the parameters of infection curve and  $k$  is a normalizing constant which makes the infection curve as infection density as given in section 3.5. The parameters of the model estimated to be  $\theta_1 = 1.77$ ,  $\theta_2 = -19.86$  and  $\theta_3 = 0.33$ . The choice of infection curve and its parameters were made based on previous experience on using backcalculation methodology for Indian AIDS data.

Once  $p_j(\theta)$  and  $N$  are known, several samples from multinomial distribution can be selected and the behaviour of the estimates for  $N$  and  $\theta_j$  can be studied. But the aim of the present simulation study is not to examine the behavior of the estimates of  $N$  and  $\theta_j$  instead we would like to analyse the uncertainties of the backcalculation estimates of  $N$  and the projected AIDS cases due to alternative choice of incubation period distributions and infection curves. Therefore only two sets of AIDS data with 17 time points was generated by taking  $\hat{X}_j = N p_j(\theta)$ ,  $j = 1, 2, \dots, 17$ . This is the expected AIDS case in the  $j^{th}$  interval corresponding to Weibull model with median 10 and 15 years. The two basic sets of AIDS data used in our simulation study are the following:

Set I:  $\hat{X}_j = N p_j(\theta)$ ,  $j = 1, 2, \dots, 17$ .  
 $\theta = (\theta_1, \theta_2, \theta_3) = (1.77, -19.86, 0.33)$

Weibull model with median incubation period 10 years is

$$F(t) = 1 - \exp(-0.002100t^{2.516}) \quad [4]$$

Set II:  $\hat{X}_j = N p_j(\theta)$ ,  $j = 1, 2, \dots, 17$ .

$$\theta = (\theta_1, \theta_2, \theta_3) = (1.77, -19.86, 0.33)$$

Weibull model with median incubation period 15 years is

$$F(t) = 1 - \exp(-0.000762t^{2.516}) \quad [5]$$

### III. SELECTION OF ALTERNATIVE INCUBATION PERIOD DISTRIBUTION

It is well known that the backcalculation estimates are very sensitive to changes in the incubation period distribution. Therefore one of our objectives is to quantify the uncertainties associated with these alternative models. In this paper, we have used ten models for incubation period distribution via. Weibull model, Gamma model, Log-logistic model, Log-normal model, Generalized exponential model, Generalized log-logistic model, Generalized gamma model, Mixed Weibull model, Change point model with Weibull hazard and Immune Invasion level model

To use the above models in our analysis the parameters of these models must be specified. The parameters can be estimated only when the data on incubation period is available. In the absence of data for incubation period distribution for Indian population, we have decided to use the incubation period data simulated from the assumed Weibull model with median 10 and 15 years as described in the previous section. All the nine models excluding the Weibull model were fitted to 5000 observations simulated from Weibull model. Such alternative models help us to understand the effects of incorrect specification of incubation period distributions on the backcalculation estimates.

The data for the Weibull model has been generated using SAS IML software package and the parameters are estimated under different incubation model set up. The parameters of alternative models when the true underlying incubation period distribution is Weibull are presented in tables. Table 1 refers to a median incubation period of 10 years and table 2 refers to a median incubation period of 15 years.

**Table 1. Parameter estimates of incubation models (Simulated data using Weibull model with median 10 years  $\lambda = 0.0021$  and  $\alpha = 2.516$ )**

| Incubation Model            | Parameters   | Q <sub>1</sub> | Q <sub>2</sub> | Q <sub>3</sub> | F(7)   | F(10)  | F(13)  |
|-----------------------------|--|----------------|----------------|----------------|--------|--------|--------|
| Gamma                       | $\sigma = 2.209652$<br>$k = 4.650859$  | 6.80           | 9.56           | 12.98          | 0.2681 | 0.5386 | 0.7517 |
| Log-logistic                | $\lambda = 0.104152$<br>$\upsilon = 3.569923$  | 7.06           | 9.61           | 13.07          | 0.2445 | 0.5362 | 0.7468 |
| Log-normal                  | $\mu = 2.218545$<br>$\sigma = 0.510080$  | 6.52           | 9.20           | 12.97          | 0.2965 | 0.5654 | 0.7515 |
| Generalized Exponential     | $\lambda = 0.229635$<br>$\upsilon = 5.614001$  | 6.62           | 9.38           | 13.05          | 0.2849 | 0.5513 | 0.7475 |
| Generalized Log-logistic    | $m_1 = 0.940961$<br>$m_2 = 50.10286$<br>$\mu = 3.956737$<br>$\tau = 0.378575$                          | 7.03           | 9.99           | 13.20          | 0.2479 | 0.5011 | 0.7371 |
| Generalized Gamma           | $\lambda_1 = 0.197618$<br>$\lambda_2 = 0.194617$<br>$\lambda_3 = 3994.574$                             | 4.94           | 8.63           | 13.84          | 0.3951 | 0.5792 | 0.7188 |
| Mixed Weibull Model         | $\alpha_1 = 2.652815$<br>$\lambda_1 = 0.085546$<br>$\alpha_2 = 2.195428$<br>$\lambda_2 = 0.088324$     | 6.98           | 9.96           | 13.25          | 0.2517 | 0.5034 | 0.7338 |
| Change Point Model          | $\lambda_1 = 0.002148$<br>$\upsilon_1 = 2.504745$<br>$\lambda_2 = 0.001939$<br>$\upsilon_2 = 2.542406$ | 7.02           | 9.99           | 13.24          | 0.2489 | 0.5013 | 0.7344 |
| Immune Invasion level Model | $\lambda = 0.398128$<br>$\mu = 0.049518$   | 5.66           | 9.01           | 13.15          | 0.3506 | 0.5688 | 0.7504 |

**Table 2. Parameter estimates of incubation models (Simulated data using Weibull model with median 15 years  $\lambda = 0.000762$  and  $\alpha = 2.516$ )**

| Incubation Model            | Parameters   | Q <sub>1</sub> | Q <sub>2</sub> | Q <sub>3</sub> | F(10)  | F(15)  | F(20)  |
|-----------------------------|--|----------------|----------------|----------------|--------|--------|--------|
| Gamma                       | $\sigma = 3.275354$<br>$k = 4.703653$  | 10.22          | 14.33          | 19.43          | 0.2373 | 0.5386 | 0.7712 |
| Log-logistic                | $\lambda = 0.069226$<br>$\upsilon = 3.609797$  | 10.66          | 14.45          | 19.59          | 0.2095 | 0.5339 | 0.7639 |
| Log-normal                  | $\mu = 2.624700$<br>$\sigma = 0.509970$  | 9.79           | 13.81          | 19.47          | 0.2638 | 0.5649 | 0.7665 |
| Generalized Exponential     | $\lambda = 0.153135$<br>$\upsilon = 5.627206$  | 9.94           | 14.08          | 19.59          | 0.2538 | 0.5508 | 0.7637 |
| Generalized Log-logistic    | $m_1 = 0.940954$<br>$m_2 = 50.10071$<br>$\mu = 4.359636$<br>$\tau = 0.378573$                          | 10.6           | 15.02          | 19.77          | 0.2203 | 0.4989 | 0.7605 |
| Generalized Gamma           | $\lambda_1 = 0.130078$<br>$\lambda_2 = 0.130073$<br>$\lambda_3 = 3978.496$                             | 7.39           | 12.89          | 20.67          | 0.3741 | 0.5814 | 0.7337 |
| Mixed Weibull               | $\alpha_1 = 2.652824$<br>$\lambda_1 = 0.057176$<br>$\alpha_2 = 2.195421$<br>$\lambda_2 = 0.059033$     | 10.61          | 15.04          | 19.76          | 0.2197 | 0.4982 | 0.7608 |
| Change Point Model          | $\lambda_1 = 0.000698$<br>$\upsilon_1 = 2.560998$<br>$\lambda_2 = 0.000691$<br>$\upsilon_2 = 2.544134$ | 10.53          | 15.02          | 19.80          | 0.2244 | 0.4989 | 0.7591 |
| Immune Invasion level Model | $\lambda = 0.401211$<br>$\mu = 0.020306$   | 8.65           | 13.65          | 19.68          | 0.3169 | 0.5640 | 0.7502 |

The quartiles Q<sub>1</sub>, Q<sub>2</sub> and Q<sub>3</sub> and the probability values at 7, 10 and 13 years viz., F(7), F(10) and F(13) are given in table 1. The values 7, 10 and 13 are first, second and third quartiles of the Weibull distribution. The quartiles and probability values of 7, 10 and 13 reflect that the log-logistic and generalized log-logistic distributions are similar to the Weibull distribution at these points. Generalized gamma model is totally different at these points. Similarly the fitted lognormal and immune Invasion level models are different from Weibull models at first and second quartiles.

The parameter estimates of the alternative models for data generated from Weibull with median 15 years are given in Table 2. The quartiles and probability values of Weibull distribution reveals that these parametric uncertainties in general reflect a incubation period distribution different from a Weibull model. The second and third quartiles of the fitted generalized log-logistic, mixed Weibull and change point models are very close to the second and third quartiles of the Weibull model. Similarly the first quartile of generalized exponential model is very close to that of the Weibull model. Hence the above ten alternatives models considered in the analysis explain wide range of incubation period distribution.

**IV. ALTERNATIVE INFECTION DISTRIBUTIONS**

Simulated AIDS data depends on the true incubation period distribution and infection density. The true incubation period distribution is taken to be Weibull and the infection density is taken to be normalized logistic prevalence curve. To understand the uncertainties of the backcalculation estimates due to alternative incubation period distribution, ten incubation models has been considered in the previous section. The uncertainties due to alternative infection densities are analysed based on the six infection densities via, Logistic Prevalence, Logistic incidence, Double Exponential incidence, Log-logistic incidence, Exponential incidence and Root Exponential incidence.

**V. UNCERTAINTIES OF BACKCALCULATION ESTIMATES**

In this section, backcalculation estimates of the parameters of the infection density, minimum size of epidemic and projected number of AIDS cases based on the simulated data are presented.

**5.1 Uncertainties due to Incubation Period Distribution**

For simulation of AIDS data, the incubation period distributions have been taken to be Weibull with median 10 years ( $\lambda = 0.0021$  and  $\alpha = 2.516$ ) and median 15 years ( $\lambda =$

0.000762 and  $\alpha = 2.516$ ). The infection density is assumed to be logistic prevalence. The following tables give the estimates of the minimum size of epidemic and projected AIDS cases.

**Table 3. Uncertainties due to alternative incubation distributions (True incubation period distribution Weibull with median 10 years  $\lambda = 0.0021$  and  $\alpha = 2.516$ )**

| Incubation Model            | Parameters of infection density   | $\hat{N}$ | $\chi^2$ | Projection of AIDS |        |        |        |
|-----------------------------|---|-----------|----------|--------------------|--------|--------|--------|
|                             |   |           |          | 2003               | 2004   | 2005   | 2006   |
| Weibull                     | $\theta_1 = 1.7749$<br>$\theta_2 = -19.8599$<br>$\theta_3 = 0.3343$         | 4055786   | 0.00     | 217533             | 303885 | 424511 | 593018 |
| Gamma                       | $\theta_1 = 1.9492$<br>$\theta_2 = -19.8871$<br>$\theta_3 = 0.3342$         | 3729064   | 29.45    | 220474             | 307962 | 430161 | 600844 |
| Log-logistic                | $\theta_1 = 1.7703$<br>$\theta_2 = -20.0398$<br>$\theta_3 = 0.3335$         | 4440024   | 97.23    | 220069             | 307194 | 428798 | 598539 |
| Log-normal                  | $\theta_1 = 1.6913$<br>$\theta_2 = -20.8357$<br>$\theta_3 = 0.3346$         | 4196915   | 63.36    | 220682             | 308376 | 430912 | 602132 |
| Gen. Expo.                  | $\theta_1 = 1.8108$<br>$\theta_2 = -19.8621$<br>$\theta_3 = 0.3346$         | 3909394   | 31.13    | 220710             | 308423 | 430987 | 602251 |
| Log-logistic                | Gen:<br>$\theta_1 = 1.7037$<br>$\theta_2 = -19.8599$<br>$\theta_3 = 0.3216$ | 2770510   | 113.67   | 216421             | 299288 | 413625 | 571368 |
| Gen. Gamma                  | $\theta_1 = 1.7737$<br>$\theta_2 = -19.8586$<br>$\theta_3 = 0.3392$         | 2346462   | 255.87   | 223477             | 313748 | 440476 | 618386 |
| Mixed Weibull               | $\theta_1 = 1.8170$<br>$\theta_2 = -19.8708$<br>$\theta_3 = 0.3344$         | 3917101   | 0.23     | 220654             | 308289 | 430726 | 601787 |
| Change Point Model          | $\theta_1 = 1.8540$<br>$\theta_2 = -19.8527$<br>$\theta_3 = 0.3344$         | 3827602   | 0.44     | 220635             | 308417 | 430560 | 601985 |
| Immune Invasion level Model | $\theta_1 = 1.8137$<br>$\theta_2 = -19.7514$<br>$\theta_3 = 0.3385$         | 4343189   | 231.85   | 223135             | 313044 | 439167 | 616089 |
| C.V                         | -   | 18.42     | -        | 1.70               | 1.93   | 2.22   | 2.55   |

**Table 4. Uncertainties due to alternative incubation distributions (True incubation period distribution Weibull with median 15 years  $\lambda = 0.000762$  and  $\alpha = 2.516$ )**

| Incubation Model            | Parameters of infection density                                     | $\hat{N}$ | $\chi^2$ | Projection of AIDS |        |        |        |
|-----------------------------|---|-----------|----------|--------------------|--------|--------|--------|
|                             |   |           |          | 2003               | 2004   | 2005   | 2006   |
| Weibull                     | $\theta_1 = 1.7941$<br>$\theta_2 = -19.8699$<br>$\theta_3 = 0.3343$ | 4012478   | 0.00     | 97176              | 135802 | 189752 | 265110 |
| Gamma                       | $\theta_1 = 2.1667$<br>$\theta_2 = -19.1384$<br>$\theta_3 = 0.3285$ | 4119278   | 158.78   | 111747             | 155264 | 215695 | 299619 |
| Log-logistic                | $\theta_1 = 2.0608$<br>$\theta_2 = -19.0373$<br>$\theta_3 = 0.3258$ | 4881722   | 307.78   | 111045             | 153876 | 213196 | 295355 |
| Log-normal                  | $\theta_1 = 1.6839$<br>$\theta_2 = -20.5469$<br>$\theta_3 = 0.3282$ | 5225986   | 406.34   | 111587             | 154994 | 215259 | 298934 |
| Gen. Expo.                  | $\theta_1 = 1.7375$<br>$\theta_2 = -19.8925$<br>$\theta_3 = 0.3291$ | 4959690   | 188.07   | 111860             | 155510 | 216166 | 300456 |
| Gen. Log-logistic           | $\theta_1 = 1.9191$<br>$\theta_2 = -19.0362$<br>$\theta_3 = 0.3209$ | 5228504   | 60.70    | 111120             | 153687 | 212394 | 293348 |
| Gen. Gamma                  | $\theta_1 = 1.1984$<br>$\theta_2 = -23.9688$<br>$\theta_3 = 0.3399$ | 3083489   | 111.26   | 114771             | 161249 | 226540 | 318261 |
| Mixed Weibull               | $\theta_1 = 1.6088$<br>$\theta_2 = -21.9522$<br>$\theta_3 = 0.3313$ | 5106661   | 6.08     | 112584             | 156865 | 218527 | 304398 |
| Change Point Model          | $\theta_1 = 1.7907$<br>$\theta_2 = -19.8727$<br>$\theta_3 = 0.3311$ | 4616337   | 8.37     | 112517             | 156829 | 218199 | 304228 |
| Immune Invasion level Model | $\theta_1 = 1.4355$<br>$\theta_2 = -22.9673$<br>$\theta_3 = 0.3374$ | 3442759   | 47.19    | 114112             | 159939 | 224157 | 314146 |
| C.V                         | -   | 22.55     | -        | 15.69              | 15.32  | 14.96  | 14.66  |

In the above tables, the estimates under the Weibull model have been considered as the correct figures because the simulation was done assuming Weibull model. The estimates under alternative models represent the uncertainties due to deviation from the assumed incubation period distribution. The  $\hat{N}$  values are very much

affected due to the changes in the incubation distribution. The values of the coefficient of variation are computed for  $\hat{N}$  and projected AIDS cases based on the true incubation Weibull model. Hence the value of the Weibull in respect of  $\hat{N}$  and projected AIDS cases are taken as true mean value.

The deviations from these values are taken for computing standard deviation. The co-efficient of variation is very high in the case of  $\hat{N}$ . The estimates are also very sensitive to alternative specification of the median incubation period. The co-efficient of variation for  $\hat{N}$  changes from 18% to 23% where as for Projected AIDS cases the co-efficient of variation changes from 2% to 15%. Therefore the estimates are very sensitive to the changes in the parameters of the incubation period distribution. Short-term prediction with long incubation period gives reduced estimates compared to short incubation period models. But the  $\hat{N}$  is very much affected with long incubation period distribution. It can be noted that the projected AIDS cases obtained under all other alternative models are very close to each other although they are different from Weibull model. This may be due to the fact that the parameter estimates of all the alternative models were obtained using the Weibull generated data of incubation period distribution. This suggests us that wrong model selection results in uncertainty of the projected AIDS cases but these uncertainties are not volatile across alternative wrong specifications. But this is not the case for the parameter  $\hat{N}$ . The  $\hat{N}$  values are very much affected by various choices for the incubation period. Hence the ultimate number of AIDS cases that may develop in a region cannot be specifically measured under backcalculation methodology although short-term prediction is possible.

**5.2 Uncertainties due to alternative infection densities**

The data of AIDS cases were generated using logistic prevalence infection density. In the last section uncertainties of backcalculation estimates for alternative incubation period distribution were studied. Another potential source of uncertainties is assumed infection curve or infection density. The following tables give variation in the backcalculation estimates due to alternative choices of infection densities corresponds to median incubation period of 10 and 15 years respectively.

**Table 5. Uncertainties due to alternative infection densities using Weibull incubation distribution with median 10 years (True infection density is logistic prevalence)**

| Incidence Curve     | Parameters  | N       | $\chi^2$ | Projection of AIDS |        |        |        |
|---------------------|---|---------|----------|--------------------|--------|--------|--------|
|                     |   |         |          | 2003               | 2004   | 2005   | 2006   |
| Logistic Prevalence | $\theta_1 = 1.7749$<br>$\theta_2 = -19.7799$<br>$\theta_3 = 0.3344$ | 4055786 | 0.00     | 217533             | 303885 | 424511 | 593018 |
| Logistic Incidence  | $\theta_1 = 1.5032$<br>$\theta_2 = -19.8607$<br>$\theta_3 = 0.3342$ | 4788710 | 0.00     | 220568             | 308126 | 430437 | 601296 |
| Double Exponential  | $\theta_1 = 1.1770$<br>$\theta_2 = 6.4198$<br>$\theta_3 = -0.0005$  | 7170988 | 0.39     | 220227             | 307329 | 428800 | 598171 |
| Log-logistic        | $\theta_1 = 0.0004$<br>$\theta_2 = 4.5026$                          | 4462734 | 16719.29 | 179363             | 224248 | 277159 | 338973 |
| Exponential         | $\theta_1 = .7155$<br>$\theta_2 = 0.3342$                           | 7198742 | 0.00     | 220467             | 308028 | 430349 | 601178 |
| Root Exponential    | $\theta_1 = 1.1453$<br>$\theta_2 = 7.8098$                          | 4289022 | 9060.42  | 187019             | 238715 | 301757 | 378059 |
| C.V                 | -   | 49.73   | -        | 10.10              | 15.18  | 20.23  | 25.12  |

**Table 6. Uncertainties due to alternative infection densities using Weibull incubation distribution with median 15 years (True infection density is logistic prevalence)**

| Incidence Curve     | Parameters   | $\hat{N}$ | $\chi^2$ | Projection of AIDS |        |        |        |
|---------------------|--|-----------|----------|--------------------|--------|--------|--------|
|                     |  |           |          | 2003               | 2004   | 2005   | 2006   |
| Logistic Prevalence | $\theta_1 = 1.79410$<br>$\theta_2 = -19.86990$<br>$\theta_3 = 0.33430$ | 4012478   | 0.00     | 97176              | 135802 | 189752 | 265110 |
| Logistic Incidence  | $\theta_1 = 1.80540$<br>$\theta_2 = -23.52610$<br>$\theta_3 = 0.33120$ | 4562066   | 7.27     | 112554             | 156810 | 218431 | 304237 |
| Double Exponential  | $\theta_1 = 1.17700$<br>$\theta_2 = 6.41987$<br>$\theta_3 = -0.00054$  | 8201781   | 9.34     | 112402             | 156451 | 217689 | 302812 |
| Log-logistic        | $\theta_1 = 0.00013$<br>$\theta_2 = 4.27507$                           | 4784651   | 6532.80  | 93596              | 113411 | 139465 | 181194 |
| Exponential         | $\theta_1 = 4.14553$<br>$\theta_2 = 0.33119$                           | 8236217   | 7.12     | 112432             | 156746 | 218358 | 304142 |
| Root Exponential    | $\theta_1 = 1.43810$<br>$\theta_2 = 7.41743$                           | 3707713   | 3733.29  | 97061              | 124501 | 158150 | 199093 |
| C.V                 | -  | 67.23     | -        | 12.30              | 14.49  | 18.18  | 21.25  |

The log-logistic and root exponential models tend to give expected AIDS cases, which are significantly different from the observed AIDS cases, and hence the chi-square values are very high. Also the projected AIDS under these infection densities are very low compared to other curves. The minimum size of the epidemic  $\hat{N}$  estimated using exponential and double exponential are very high compared other curves. But the projected values under these two models are comparable to other models. Due to very high  $\hat{N}$  values under these two models the coefficient of variation for  $\hat{N}$  is very high. Similarly the coefficient of variation for projected AIDS cases is inflated due to under estimates for two infection densities. For projection of AIDS cases it seems one can use logistic prevalence, logistic incidence, double exponential or exponential curves.

In the case of data generated assuming a median of 10 years, the changes in the incubation period very much affect only the minimum size of the epidemic.  $\hat{N}$  The coefficient of variation, which is in the order of 18 % under the true model accounting for variations in the incubation period, rises up to 65 % if the infection densities are also varied. The backcalculation estimate of the total epidemic size is very much affected by the changes in the incubation period. But the projected values are somewhat stable even if the incubation periods are changed. The present simulation study reveals an important aspect of backcalculation. The infection density has been to chosen judiciously. Based on the six parametric infection densities the simulation shows that the estimates are very much affected for wrong choice of the infection densities. The same pattern of uncertainty is reflected even in the case of median incubation period with 15 years. The coefficient of variation is also reflects the same in both cases.

**VI. CONCLUSION**

The simulation study has been attempted to quantify the uncertainties due to incubation period distributions and infection densities. Under the simple backcalculation methodology it has been observed that one can obtain reliable estimate of the projected AIDS cases under various choices of incubation period distributions.

But parametric infection density curve adds a larger source of uncertainty. Among the six models considered as infection density, except log-logistic and root exponential densities all other models give consistent values for  $\hat{N}$  and projected AIDS cases. In the absence of prior knowledge about infection density it is better to go for various choices of infection densities to obtain  $\hat{N}$  and projection of AIDS epidemic.

### ACKNOWLEDGEMENT

The authors whole heartedly thank Dr.T. Anbupalam for his kind comments and positive criticism.

### REFERENCES

1. Anbupalam, T., Ramanan, R. and Venkatesan, P. (2002). Backcalculation of HIV/AIDS in Tamilnadu; In *Bio statistical Aspects of Health and Epidemiology* (Eds; Pandey, C.M., Pradeep Mishra and Uttam Singh), Department of Biostatistics, Sanjay Gandhi Postgraduate Institute of Medical Research, Lucknow, India. 232-243.
2. Brookmeyer, R. (1991). Reconstruction and future trends of the AIDS epidemic in the United States. *Science*, **253**: 37-42.
3. Brookmeyer, R. (1996). AIDS, Epidemics and Statistics. *Biometrics*, **52**: 781-796.
4. Brookmeyer, R. and Gail, M.H (1986). Minimum size of the acquired immunodeficiency syndrome (AIDS) epidemic in the United States. *Lancet*, **2**: 1320-1322.
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. *Journal of the American Statistical Association*, **83**: 301-308.
6. Brookmeyer, R. and Gail, M.H. (1994). *AIDS epidemiology: A Quantitative Approach*. Oxford University Press, Oxford.
7. Dempster, A.P., Laird, N.M. and Rubin, D.B. (1977). Maximum likelihood from incomplete data via the EM algorithm. *Journal of the Royal Statistical Society, Series B*, **39**: 1-22.
8. Mariotti, S. and Cascilio, R. (1996). Sources of uncertainty in estimating HIV infection rates by backcalculation: Application to Italian data. *Statistics in Medicine*, **15**: 2669-2687.
9. Sanathanan, L. (1972). Estimating the size of a multinomial population. *Annals of Mathematical Statistics*, **43**: 142-152.
10. Venkatesan, P. (2002). Methods of projection of HIV/AIDS epidemic; In *Epidemiology, Health and Population*, (Eds; Anil Kumar), 143-155



**Dr. R. RAVANAN,**

He has 19 years of research, consultancy and teaching experience in the application of Statistics to Bio-statistics especially in the estimation of HIV/AIDS.. He has been teaching and training Advanced Operation Research, Stochastic Processes, Statistical Inference and Multivariate methods to Post Graduate students and researchers and has provided consultancy support to scientists in the field of biomedical sciences and was part of multi-disciplinary research teams in charge of providing statistical guidance. He has considerable experience in using statistical software's SPSS. He is life member of Indian Society for Medical Research.