
P. Venkatesan

ABSTRACT

HIV incubation period is the random time between the HIV-infection and the onset of clinical AIDS. Distribution of this non-negative random variable is known as HIV incubation period distribution. The Back-calculation method reconstructs the past pattern of HIV infection and predicts the future number of AIDS cases with the present infection status. It depends on three important factors: incubation distribution, incidence curve and observed number of AIDS cases over time. This method is very popular and requires less information and assumptions. Lack of information about incubation distribution, the effect of intervention therapy on incubation period, and errors in reported AIDS incidence leads to uncertainties associated with this method. The incubation distribution is assumed to be exactly known in back-calculation methodology. Incubation period of HIV is very long and highly variable within and between cohorts. The current prevalence of HIV-infection and the corresponding pattern of incidence from the beginning of the epidemic to the present time are mainly estimated by means of back-calculation method. It calculates the most likely temporal distribution of infected individuals compatible with the number of observed AIDS cases starting from the suitable estimate of the incubation period, derived from the available data. Most of the projections formulated the problem of estimation of future AIDS cases as estimation of parameters in multinomial likelihood with unknown sample size by EM algorithm. In this paper the various approaches for modeling the incubation distribution are compared using real and simulated data under various infection density distributions. The projected minimum AIDS cases in India based on the reported data for the years 2003, 2004, 2005 and 2006 are around 27000, 44000, 70000 and 113000 respectively. The corresponding figures based on the adjusted data are around 67000, 100000, 150000 and 230000 respectively.

Department of Statistics, Tuberculosis Research Centre (ICMR), Chetpet, Chennai-600 031
venkatesanp@trccchennai.org,
INTRODUCTION
The Acquired Immune Deficiency Syndrome (AIDS) is a devastating disease caused by Human Immune deficiency Virus (HIV) that is transmitted through either sexual, other contacts in which body fluids are exchanged or using infected blood products. Following a few recognized cases among homosexual men in the United States in the year 1981, new cases of AIDS were subsequently reported in many countries throughout the world. It has now reached a pandemic proportion, as no country in the world is now free from HIV/AIDS. This epidemic ranks as one of the most destructive microbial scourges in human history and has posed a formidable challenge to the biomedical research and public health communities of the world.

The global pandemic of HIV infection comprises many different epidemics, each with its own dynamics e.g., time of introduction, population density, cultural and social issues. Spread of the epidemic has varied considerably between developed and developing countries, depending on the culture as well as other social and behavioral patterns. Incidence rates have been the highest in developing countries where heterosexual transmission is most common.

First case of AIDS was reported in India in the year 1986. Now India's entry into the third phase of the HIV epidemic, as envisaged from the increasing number of HIV infections detected even among housewives and children, signals major AIDS crisis in the offing. The first phase of HIV was recognized when rising trend of HIV prevalence was established among the Commercial Sex Workers (CSW) in 1988 and the professional blood donors in 1989. The second phase started after 1989 when several of the clients of CSW and blood recipients were found to be infected.

BIOLOGICAL ASPECTS OF HIV/AIDS AND ITS TRANSMISSION
AIDS is a condition in which the in-built immune mechanism of the human body breaks down completely. The process is gradual but ultimately suppresses the immune ability of the individuals. It is a medically accepted fact that the HIV is the causal agent of AIDS. Those who are affected by AIDS are susceptible to other opportunistic infections like Candida, Mucor Oomycosis and Aspergillus etc.

Within a few weeks of its entry into the blood stream of an individual, the HIV sets an insidious and progressive attack on the immune system of the individual by binding itself to one of the T4 helper cells. The viral attachment to the cell is initiated by means of antibodies to the virus. On infecting the T4 cell, the viral RNA is injected into the target cell along with viral transcriptase and viral integrase. The viral reverse transcriptase transcribes the viral RNA into viral DNA and the viral integrase bind the viral DNA with that of the host. The integrated viral DNA may remain latent or in an activated form. In the activated form, genomic RNA and messenger RNA are transcribed from the integrated viral DNA. The regulatory proteins such as tat and rev are translated from messenger...
RNA. These proteins together with the genomic RNA cause the production of new HIV viruses, which bud on the cell wall. These buds accumulate on the cell wall and after a random length of time from the time of infection, the infected cell disintegrates (i.e. the cell undergoes lyses) releasing a random number of HIV viruses and this process continues indefinitely. This mechanism by which the killing of the T4 cells takes place, the number of free virus in the blood destroys progressively the immune-competence of the host. The viral load (the amount of free HIV in the blood) increases following infection and peaks at the time of sero-conversion (i.e. the time at which detectable levels of antibodies develop in the blood) and falls to a level (called set-point) by about 2 years thereafter and then remains at that level throughout the asymptomatic period. A low level of set point corresponds to a low risk of AIDS and a high level of set point corresponds to a high risk of AIDS. The viral load measured at arbitrary base line is a good prognostic marker at all stages of the development of AIDS. On the other hand, the T4 cell counts in the blood decrease rapidly in the initial months following infection and thereafter at a slower rate. After a period of several years, when the level of T4 cells declines to roughly 400 cells/mm$^3$, the person exhibits symptoms and signs associated with HIV disease. These symptoms and signs do not meet the surveillance definition of AIDS and we say that the person is suffering from AIDS related complex (ARC). The level of T4 cells further continues to decrease to roughly below 200 cells/mm$^3$ and AIDS defining diseases such as Kaposi's sarcoma and Pneumocystis carinii pneumonia occur. At this stage, the patient is said to have developed AIDS.

**APPROACHES FOR HIV/AIDS MODELING**

In general, HIV/AIDS epidemic modeling is categorized based on the following four broad approaches but not mutually exclusive ones:

**Deterministic Models:** In this type of modeling the parameters such as number of susceptible individuals, infected individuals and number of AIDS cases are assumed to be deterministic. These models are described by system of differential or integral equations. The progression of the epidemic is studied using these equations. Some of the deterministic models for AIDS epidemic were developed.

**Stochastic Models:** Stochastic models assume that some of the key parameters are random variables. It is assumed that the HIV epidemic is a continuous time stochastic process. Stochastic models are considered to be more realistic than deterministic models and with some special assumptions the results of deterministic models can be approximated through stochastic models. But there are studies, which show that stochastic models give a better interpretation of the epidemic than the deterministic models.

**Statistical models:** The statistical models are developed based on AIDS
epidemiological and survey data. These models make full use of the available data compared to deterministic and stochastic models. But in this type of modeling the mechanism and prior information system is usually not taken into consideration. The back-calculation approach for HIV/AIDS projection can be categorized into this type of modeling\textsuperscript{12-14}.

**State Space Models:** Wu and Tan (1995)\textsuperscript{15} have introduced the state space models for AIDS epidemic, which takes the advantage of stochastic and statistical models. The state space models were originally proposed\textsuperscript{16} for engineering control and communication. This model is also used\textsuperscript{10-11,17} in projection and detailed description of the state space models is given by Tan (2000)\textsuperscript{1}.

**PROJECTION OF HIV/AIDS**

Projections of HIV/AIDS using the statistical modeling approach are done based on the following three methods:(i). Fitting a model to the incidence of HIV/AIDS and extrapolating the curves into the future\textsuperscript{18}. The estimates obtained using this method depends on the mathematical function used and hence some function can produce anomalous results. This method is also less efficient as this does not include important information on the epidemic like incubation period, infection density and nature of the spread of the epidemic.(ii). The next approach is based on modeling the dynamics of the epidemic\textsuperscript{19,20}. This approach requires certain knowledge about mixing pattern of HIV individual with probabilities of infection per contact, size of high risk behavior group, probabilities of infection through blood product, needle sharing etc. In developing countries like India, knowledge about these key parameters is incomplete. Also stochastic modeling of the epidemic demands many parameters, which are generally difficult to estimate due to limitation of appropriate data especially in the Indian context. There is a lot of literature on deterministic and stochastic models for the spread of HIV epidemic\textsuperscript{21}. (iii). One of the most popular method used for projection of HIV/AIDS is the back calculation method\textsuperscript{21,22}. This method is used to reconstruct the past pattern of HIV infection and to predict the future number of AIDS cases, apart from knowing the present infection status. This method depends on three important factors namely, the incubation period distribution, incidence curves and the observed number of AIDS cases over a time period. There are also uncertainties associated with this approach because lack of certain information about incubation period distribution, the effect of intervention therapy on incubation period and errors in reported AIDS incidence. However back calculation method is very popular, as it requires few information and assumptions and thus easy to apply.

**Back-calculation Methodology**

The back calculation method for short-term projection of AIDS epidemic was introduced\textsuperscript{21,22}. The infection rate represents the number of new HIV infections per unit time at calendar time. Sero-conversion refers to infected individuals remain sero-negative until
they develop detectable HIV antibodies. The sero-conversion period is usually short and it has a median of 2 to 3 months. The incubation period is the time duration between sero-conversion to AIDS. The incubation periods are highly variable across groups. Several models for incubation period distribution are used to describe the incubation period.

This method uses a form of infection curve, either parametric or non-parametric, for the number of past HIV infections or equivalently a density function for HIV infections. The time between HIV infection and the diagnosis of AIDS is known as incubation time and it is modeled by a known distribution. Future AIDS incidence can be projected by the estimated infection curve and the known incubation distribution.

Most of the research on the projection of AIDS epidemic is based on the back calculation method. There are various sources of uncertainties associated with the estimates of AIDS based on back-calculation method. Major sources of uncertainties may be due to the inaccuracies of reported AIDS cases, the assumptions about the incidence curve and the incubation distribution. The inaccuracies in the reported AIDS cases may be due to the reporting delays, under diagnosis and underreporting. Reporting delays of AIDS incidence has also been modeled. It was observed that only a few studies address the problem of underreporting. Various forms of incidence curves and several parametric incubation distributions have been used in the literature. Uncertainties of AIDS incubation time and its effect on back calculation estimates are discussed.

The methodology to include information from the results of surveillance to improve the accuracy of projection of AIDS cases was given by Bellico and Marschner. Bayesian approaches for AIDS projection have also received a lot of attention. Swaminathan et al., (2000) presented the survival experience of a group of HIV patients. Venkatesan (2002) discussed the various mathematical and statistical approaches for projection of HIV/AIDS epidemic. Projections using back calculation method are explored. In this work we have concentrated mainly on understanding the current state of epidemic and forecasting the future path via back calculation method. The broad objective is to provide HIV/AIDS estimates under different model assumptions using surveillance data for India.

Models for Incubation Period

The HIV incubation period is the random time between the HIV infection and the onset of clinical AIDS symptoms. The probability distribution of this non-negative random variable is known as HIV incubation period distribution. Medley et al. (1987) showed that the incubation period of HIV is known to be very long and it is highly variable within and between cohorts.

The analysis of incubation period is very important in AIDS epidemic studies. Incubation period distribution is
assumed to be exactly known in back calculation methodology. Several studies have observed that back calculation estimates are very sensitive to the choice of incubation period distribution. It has also been shown that the HIV incubation distribution is significantly affected by age, treatment by antiviral drugs and other opportunistic infections. Hence for estimation of the HIV infection and projection of future HIV prevalence and AIDS, it is very important to study the HIV incubation distribution under different conditions.

The incubation period models are similar to survival models based on non-negative random variables and can be fitted using either parametric or semi-parametric approach. A review of some of the parametric models for incubation period is discussed below.

**Weibull and Gamma Models:** Weibull and gamma models are the most commonly used for many real data applications and in particular for back calculation approach. Between the two, Weibull model is a popular candidate for HIV incubation period because of its nice properties viz., it is proportional hazard as well as accelerated failure time model. The Weibull distribution function is given by

\[ F(t) = 1 - e^{-\left(\lambda t^\alpha\right)} \quad \lambda > 0, \alpha > 0 \text{ and } t > 0 \quad (1) \]

The earliest studies of Weibull incubation period have been attempted and fitted Weibull model to study the incubation period distribution for transfusion associated AID cases

\[ F(t) = 1 - e^{-0.0243 t^{2.286}} \quad (2) \]

These parameter values correspond to a median incubation period of 4.3 years. The incubation period of patients infected by blood transfusion was studied using Weibull distribution for Children (0-4 years) and estimated the parameters as follows; \( \alpha = 1.9390, \lambda^\alpha = 0.1566 \), for others (5-59 years) as \( \alpha = 2.3960, \lambda^\alpha = 0.0048 \). Boldson et al. used gamma, Weibull and log-normal models for incubation time of cohort study from San Francisco AIDS cases. The fitted Weibull model for their data is given by

\[ F(t) = 1 - \exp(-0.001296 t^{2.5}) \quad (3) \]

The Weibull HIV incubation model used by Anderson et al. (1986) is given by

\[ F(t) = 1 - \exp(-0.1190 t^{1.974}) \quad (4) \]

Brookmeyer and Goedert used the Weibull incubation period distributions based on the study of hemophiliacs over 20 years of age. The fitted Weibull model for their data is given by

\[ F(t) = 1 - e^{-0.0021 t^{2.516}} \quad (5) \]

This estimate corresponds to a median incubation of 10 years.

Based on 732 HIV-positive hemophiliacs enrolled in Italian registry, Chiarotti et al. estimated the incubation distributions assuming three different parametric models: uniform, uniform in three sub intervals and truncated Weibull under two approaches namely the median and median of three random values. There are altogether six different approaches to estimate the incubation time of individuals. They found that the
incubation times obtained using first and third are similar. Therefore they reported only four estimates. The estimates of the four models parameters are \( \alpha = 2.9, 2.4, 2.9, 2.6 \) and \( \lambda^* = 0.003668, 0.001039, 0.000446, 0.000845 \). The median incubation time are 13.5, 15, 12.6 and 13.3 respectively.

Munoz and Xu\(^{53}\) using Multicenter AIDS Cohort Study (MACS) estimated this Weibull model.

\[
F(t) = 1 - e^{-0.052087 \cdot 1.285347^t} \tag{6}
\]

The median incubation period using this model is 7.5 years. Other important studies also used the Weibull model for HIV incubation period\(^8,39,41,54\).

The gamma distribution is another important parametric distribution used to model incubation period of HIV/AIDS. The gamma density function is

\[
g(x) = \frac{1}{\Gamma(k) \lambda^k} x^{k-1} e^{-\lambda x} \quad x > 0, \lambda > 0 \tag{7}
\]

One of the earliest studies\(^8\) that used gamma model for incubation period of HIV estimated parameter for the gamma models are as follows. \( k = 2.669, 2.473 \) and \( \alpha = 0.911 \) and 11.001. The parameters estimated from gamma model based on the San Francisco AIDS data gave \( k = 3.130 \) and \( \alpha = 5.715 \) years. Freund and Book\(^{55}\) fitted gamma model with \( k = 3 \) (Erlang form) to the San Francisco AIDS data and the estimate obtained for the parameter \( \sigma = 2.660 \) years.

Log-logistic and Log-normal models:

Lawless and Sun\(^56\) used the log-logistic model for HIV incubation period, in addition Chiarotti et al.\(^52\) used log-logistic model and generalized exponential model for their data. The log-normal distribution has been used \(^{50,57}\) for HIV incubation period. Recent studies\(^{53,58}\) have shown that lognormal distribution fits well than Weibull model.

The log-logistic distribution function is

\[
H(t) = 1 - [1 + (\lambda t)^\mu]^{-1} \quad \lambda > 0, \mu > 0, t > 0 \tag{8}
\]

The distribution function of lognormal is

\[
F(t) = \Phi \left( \frac{\log t - \mu}{\sigma} \right) \tag{9}
\]

where

The distribution function generalized log-logistic is

\[
g(x) = \frac{1}{\Gamma(m_1, m_2) \lambda^{m_1}} \int_0^x x^{m_1-1} (1+x)^{-m_2-1} dx \quad x > 0, m_1 > 0, m_2 > 0 \tag{10}
\]

The generalized log-logistic distribution reduces to log-logistic distribution when This has been shown\(^{39,40}\) that the three parameters generalized log-logistic distribution with fits better than the log-logistic distribution for data on cancer survival analysis. Stacy\(^{61}\) introduced a generalization of gamma distribution with three parameters. This model is a generalization of many survival distributions. The convolution of exponential distribution has been used as incubation model for HIV.

Longini et al.\(^{62}\) used a staged Markov model to estimate the distribution and mean length of the incubation period from a cohort study of 603 HIV infected individuals who have been followed through various stages of infection. They used the generalized gamma model to describe the transition probabilities of the Markov model.

Mixture and Staging Model: One way to accommodate the variation between different groups of the population is by
using mixture models. Suppose certain proportions of infected individuals have an incubation period distribution and the remaining proportion have the incubation period distribution. Then the incubation period distribution for the entire population of infected individuals is a mixture of and is given by

\[ F(t) = \alpha F_1(t) + (1-\alpha) F_2(t) \quad 0 < \alpha < 1 \quad (11) \]

Auger et al.\textsuperscript{3} considered a mixture of two Weibull densities for the incubation period of Paedictric AIDS cases. The mixture of two Weibull distributions\textsuperscript{49} was also used in a study of incubation period distribution of sample individuals drawn from San Francisco AIDS data. The convolution equation for the incubation period comprising of two stages given by Brookmeyer and Liao (1990) is

\[ F(t) = \int_{0}^{t} f_1(u) F_2(t-u) du \quad (12) \]

where

\[ f_1(u) = h_1(u) \exp\left\{ -\int_{0}^{u} h_1(s) ds \right\} \quad (13) \]

\[ F_2(u) = 1 - \exp\left\{ -\int_{0}^{u} h_2(s) ds \right\} \quad (14) \]

Suitable changes should be made in the above formulations to account for calendar time of infection. Under staging models the incubation period is considered to be comprised of stages. The progression from time of infection to AIDS was assumed to occur in 3 stages\textsuperscript{44}. The stage 1 refers to HIV infection without immunological abnormalities, stage 2 is the development of pre-AIDS disease and stage 3 is the development of clinical AIDS. The incubation time of an individual by definition is the total time spent on stage 1 and stage 2. Therefore different models for these two stages can be assumed. Suppose the times spent on the two stages are not independent, then the time spent on the stage 2 can be conditioned on the time spent on stage 1. Under this case Mariotti and Cascio\textsuperscript{26} have given the survival functions for the second stage.

**Change Point and Contact Models:** A change point model and contact model for incubation period of HIV was proposed\textsuperscript{65}. Suppose the incubation time for an individual is \( t \). It is reasonable to assume that between 0 to \( t \), there is a time point at which the hazard of incubation changes. The point may be the time after infection when the individual realizes the threat of AIDS and seeks some kind of medication. Suppose the hazard before and after the change point is constant, then \( h(t) \) is given by

\[ h(t) = \begin{cases} \alpha & t \leq \tau \\ \beta & t > \tau \end{cases} \quad (15) \]

The survival function of the above change point model is given by

\[ S(t) = \exp\left\{ -\int_{0}^{t} h(x) dx \right\} = \begin{cases} e^{-\alpha t} & t \leq \tau \\ e^{-\alpha \tau} e^{-\beta(t-\tau)} & t > \tau \end{cases} \quad (16) \]

The other types of hazards considered were varying hazard and Weibull hazard.
A model for incubation time using the concept of contacts to immune system uses the random contact concept for formulation. Suppose at time $t = 0$, a member tested for HIV positive for the first time, experiences a random $N$ number of contacts to the immune system before he shows clinical symptoms to AIDS. The number of contacts $N$ experienced by the individual is assumed to follow a Poisson process with parameter $\lambda(>0)$. We assume that the system undergoes at least one invasion before the individual become AIDS in $(t, t+\Delta t)$. Using Laplace transformation and simple algebra we obtain the distribution function for an individual to become AIDS can be obtained as

$$F(t) = \frac{\lambda}{\mu} \int_0^t u(1-u)^{(\lambda/\mu)-1} e^{(\lambda/\mu)u} \, du \quad (17)$$

If $\lambda = \mu =$, then

$$F(t) = 1 - e^{-\lambda t} e^{1-e^{-\lambda t}} \quad (18)$$

$$h(t) = \lambda(1 - e^{-\lambda t}) \quad (19)$$

It can be noted that the hazard rate is increasing function of $t$. The median of the distribution can be obtained numerically by solving (17). The parameters of can be estimated by using the method of maximum likelihood.

**BACK-CALCULATION ESTIMATES OF HIV/AIDS IN INDIA**

The National AIDS Control Organization (NACO) publishes periodic estimates of HIV/AIDS based on its own projections and also the reports of UNAIDS and WHO. NACO also publishes the reported number of AIDS cases in India as well as in states and union territories in India. In 1986, the reported number of AIDS cases was 6 and in 2002, the number increased to 19324. The estimates published by various other agencies regarding the total number of HIV/AIDS cases in India seemed to be over estimated according to the many researchers. An attempt has been made in this work to give the estimates of the minimum number of people living with HIV/AIDS in India. We have taken 2002 as the base period and estimated the AIDS cases that will be reported to the system in the next four years. The estimated figures are compared with the observed ones for 2003, 2004 and 2005.

The basic data required for back calculation methodology is the number of AIDS cases over a period of time. NACO publishes monthly updates of the reported AIDS cases for the past few years. The monthly updates of the recent years also suffer reporting delay and therefore pooled yearly reported AIDS cases assumed to be more reliable. Moreover the data compiled by NACO is at unequal time intervals. Therefore in the present study only yearly reported AIDS cases were considered for projections.

The starting point of the infection $T_0$ is taken to be 1980 for India. Ten incubation period distributions are used in the present projection of AIDS cases in India. The estimates of minimum size of the epidemic and future number of AIDS cases are obtained assuming different median incubation periods. For the
Table 1. Estimates of HIV incidence and projection of AIDS under logistic prevalence infection density

<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>When median incubation= 8 years</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Weibull</td>
<td>4243202</td>
<td>280156</td>
<td>26795</td>
<td>43327</td>
<td>70060</td>
<td>113286</td>
<td></td>
</tr>
<tr>
<td>Gamma</td>
<td>3023853</td>
<td>160377</td>
<td>26841</td>
<td>43433</td>
<td>70282</td>
<td>113728</td>
<td></td>
</tr>
<tr>
<td>Log-logistic</td>
<td>4141256</td>
<td>301222</td>
<td>26797</td>
<td>43332</td>
<td>70069</td>
<td>113302</td>
<td></td>
</tr>
<tr>
<td>Log-normal</td>
<td>4378305</td>
<td>249908</td>
<td>26816</td>
<td>43375</td>
<td>70157</td>
<td>113482</td>
<td></td>
</tr>
<tr>
<td>Generalized Exponential</td>
<td>4446099</td>
<td>146690</td>
<td>26846</td>
<td>43444</td>
<td>70304</td>
<td>113771</td>
<td></td>
</tr>
<tr>
<td>Generalized Log-logistic</td>
<td>4052072</td>
<td>306432</td>
<td>26680</td>
<td>43035</td>
<td>69410</td>
<td>111944</td>
<td></td>
</tr>
<tr>
<td>Generalized Gamma</td>
<td>3624674</td>
<td>146772</td>
<td>26840</td>
<td>43432</td>
<td>70069</td>
<td>113302</td>
<td></td>
</tr>
<tr>
<td>Mixed Weibull</td>
<td>4219371</td>
<td>280259</td>
<td>26789</td>
<td>43321</td>
<td>70050</td>
<td>113280</td>
<td></td>
</tr>
<tr>
<td>Change Point</td>
<td>4287475</td>
<td>279968</td>
<td>26780</td>
<td>43325</td>
<td>70065</td>
<td>113289</td>
<td></td>
</tr>
<tr>
<td>Immune Invasion</td>
<td>2917156</td>
<td>175218</td>
<td>26829</td>
<td>43406</td>
<td>70226</td>
<td>113616</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>When median incubation= 10 years</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Weibull</td>
<td>4681218</td>
<td>444462</td>
<td>26475</td>
<td>43265</td>
<td>69925</td>
<td>113013</td>
<td></td>
</tr>
<tr>
<td>Gamma</td>
<td>3814500</td>
<td>255189</td>
<td>26824</td>
<td>43394</td>
<td>70199</td>
<td>113561</td>
<td></td>
</tr>
<tr>
<td>Log-logistic</td>
<td>4104232</td>
<td>488436</td>
<td>26766</td>
<td>43258</td>
<td>69911</td>
<td>112987</td>
<td></td>
</tr>
<tr>
<td>Log-normal</td>
<td>4250104</td>
<td>385640</td>
<td>26753</td>
<td>43326</td>
<td>70055</td>
<td>113273</td>
<td></td>
</tr>
<tr>
<td>Generalized Exponential</td>
<td>4613632</td>
<td>251003</td>
<td>26826</td>
<td>43398</td>
<td>70206</td>
<td>113576</td>
<td></td>
</tr>
<tr>
<td>Generalized Log-logistic</td>
<td>4706203</td>
<td>550776</td>
<td>26676</td>
<td>43023</td>
<td>69383</td>
<td>111885</td>
<td></td>
</tr>
<tr>
<td>Generalized Gamma</td>
<td>4074022</td>
<td>199972</td>
<td>26832</td>
<td>43412</td>
<td>70237</td>
<td>113638</td>
<td></td>
</tr>
<tr>
<td>Mixed Weibull</td>
<td>4730853</td>
<td>443742</td>
<td>26731</td>
<td>43242</td>
<td>69927</td>
<td>113007</td>
<td></td>
</tr>
<tr>
<td>Change Point</td>
<td>4732736</td>
<td>442925</td>
<td>26725</td>
<td>43235</td>
<td>69919</td>
<td>113020</td>
<td></td>
</tr>
<tr>
<td>Immune Invasion</td>
<td>3912703</td>
<td>255182</td>
<td>26816</td>
<td>43374</td>
<td>70158</td>
<td>113479</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>When median incubation= 12 years</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Weibull</td>
<td>5069775</td>
<td>657874</td>
<td>26750</td>
<td>43216</td>
<td>69820</td>
<td>112799</td>
<td></td>
</tr>
<tr>
<td>Gamma</td>
<td>4219708</td>
<td>405246</td>
<td>26804</td>
<td>43345</td>
<td>70095</td>
<td>113352</td>
<td></td>
</tr>
<tr>
<td>Log-logistic</td>
<td>4630203</td>
<td>750856</td>
<td>26736</td>
<td>43186</td>
<td>69757</td>
<td>112675</td>
<td></td>
</tr>
<tr>
<td>Log-normal</td>
<td>4964605</td>
<td>570339</td>
<td>26774</td>
<td>43273</td>
<td>69945</td>
<td>113052</td>
<td></td>
</tr>
<tr>
<td>Generalized Exponential</td>
<td>5412130</td>
<td>451076</td>
<td>26795</td>
<td>43325</td>
<td>70052</td>
<td>113266</td>
<td></td>
</tr>
<tr>
<td>Generalized Log-logistic</td>
<td>5144634</td>
<td>883083</td>
<td>26674</td>
<td>43018</td>
<td>69369</td>
<td>111856</td>
<td></td>
</tr>
<tr>
<td>Generalized Gamma</td>
<td>4708975</td>
<td>259962</td>
<td>26825</td>
<td>43396</td>
<td>70203</td>
<td>113568</td>
<td></td>
</tr>
<tr>
<td>Mixed Weibull</td>
<td>5070029</td>
<td>653421</td>
<td>26732</td>
<td>43208</td>
<td>69815</td>
<td>112810</td>
<td></td>
</tr>
<tr>
<td>Change Point</td>
<td>4908890</td>
<td>660570</td>
<td>26724</td>
<td>43251</td>
<td>69803</td>
<td>112893</td>
<td></td>
</tr>
<tr>
<td>Immune Invasion</td>
<td>4253183</td>
<td>353892</td>
<td>26805</td>
<td>43348</td>
<td>70099</td>
<td>113362</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>When median incubation= 15 years</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Weibull</td>
<td>5293098</td>
<td>1093251</td>
<td>26729</td>
<td>43165</td>
<td>69704</td>
<td>112560</td>
<td></td>
</tr>
<tr>
<td>Gamma</td>
<td>4362512</td>
<td>812829</td>
<td>26763</td>
<td>43249</td>
<td>69888</td>
<td>112935</td>
<td></td>
</tr>
<tr>
<td>Log-logistic</td>
<td>4447016</td>
<td>1320676</td>
<td>26698</td>
<td>43091</td>
<td>69550</td>
<td>112254</td>
<td></td>
</tr>
<tr>
<td>Log-normal</td>
<td>5210566</td>
<td>965345</td>
<td>26741</td>
<td>43197</td>
<td>69777</td>
<td>112713</td>
<td></td>
</tr>
<tr>
<td>Generalized Exponential</td>
<td>6078422</td>
<td>1185040</td>
<td>26714</td>
<td>43130</td>
<td>69634</td>
<td>112423</td>
<td></td>
</tr>
<tr>
<td>Generalized Log-logistic</td>
<td>6585507</td>
<td>1585456</td>
<td>26672</td>
<td>43013</td>
<td>69359</td>
<td>111834</td>
<td></td>
</tr>
<tr>
<td>Generalized Gamma</td>
<td>5215866</td>
<td>365798</td>
<td>26817</td>
<td>43376</td>
<td>70160</td>
<td>113483</td>
<td></td>
</tr>
<tr>
<td>Mixed Weibull</td>
<td>5246183</td>
<td>1081218</td>
<td>26732</td>
<td>43171</td>
<td>69716</td>
<td>112585</td>
<td></td>
</tr>
<tr>
<td>Change Point</td>
<td>5323427</td>
<td>1109845</td>
<td>26748</td>
<td>43210</td>
<td>69689</td>
<td>112645</td>
<td></td>
</tr>
<tr>
<td>Immune Invasion</td>
<td>4661770</td>
<td>534254</td>
<td>26793</td>
<td>43319</td>
<td>70037</td>
<td>113235</td>
<td></td>
</tr>
<tr>
<td>Incubation Model</td>
<td>( \hat{N} )</td>
<td>Incidence in 2002</td>
<td>Projection of AIDS</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>--------------------------</td>
<td>-----------------</td>
<td>-------------------</td>
<td>--------------------</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>2003</td>
<td>2004</td>
<td>2005</td>
<td>2006</td>
<td></td>
</tr>
<tr>
<td><strong>When median incubation = 8 years</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Weibull</td>
<td>4580854</td>
<td>561695</td>
<td>64695</td>
<td>100801</td>
<td>152786</td>
<td>231785</td>
<td></td>
</tr>
<tr>
<td>Gamma</td>
<td>4042568</td>
<td>356453</td>
<td>66735</td>
<td>101312</td>
<td>153795</td>
<td>233433</td>
<td></td>
</tr>
<tr>
<td>Log-logistic</td>
<td>4752540</td>
<td>598002</td>
<td>66504</td>
<td>100819</td>
<td>152832</td>
<td>231671</td>
<td></td>
</tr>
<tr>
<td>Log-normal</td>
<td>4505245</td>
<td>512988</td>
<td>66601</td>
<td>101031</td>
<td>153241</td>
<td>232415</td>
<td></td>
</tr>
<tr>
<td>Generalized Exponential</td>
<td>3952652</td>
<td>330758</td>
<td>66756</td>
<td>101368</td>
<td>153902</td>
<td>233646</td>
<td></td>
</tr>
<tr>
<td>Generalized Log-logistic</td>
<td>4589124</td>
<td>545150</td>
<td>65874</td>
<td>99305</td>
<td>149699</td>
<td>225625</td>
<td></td>
</tr>
<tr>
<td>Generalized Gamma</td>
<td>3952681</td>
<td>323935</td>
<td>66732</td>
<td>101302</td>
<td>153772</td>
<td>233428</td>
<td></td>
</tr>
<tr>
<td>Mixed Weibull</td>
<td>4253141</td>
<td>562326</td>
<td>66510</td>
<td>100789</td>
<td>152764</td>
<td>231574</td>
<td></td>
</tr>
<tr>
<td>Change Point</td>
<td>4650105</td>
<td>561549</td>
<td>66502</td>
<td>100795</td>
<td>152799</td>
<td>231602</td>
<td></td>
</tr>
<tr>
<td>Immune Invasion</td>
<td>4101854</td>
<td>375978</td>
<td>66677</td>
<td>101712</td>
<td>153521</td>
<td>232975</td>
<td></td>
</tr>
<tr>
<td><strong>When median incubation = 10 years</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Weibull</td>
<td>4852649</td>
<td>867059</td>
<td>63356</td>
<td>100475</td>
<td>152148</td>
<td>230404</td>
<td></td>
</tr>
<tr>
<td>Gamma</td>
<td>4953246</td>
<td>546248</td>
<td>66644</td>
<td>101102</td>
<td>153374</td>
<td>232676</td>
<td></td>
</tr>
<tr>
<td>Log-logistic</td>
<td>4858625</td>
<td>935054</td>
<td>66341</td>
<td>100446</td>
<td>152125</td>
<td>230375</td>
<td></td>
</tr>
<tr>
<td>Log-normal</td>
<td>4751689</td>
<td>765549</td>
<td>66495</td>
<td>100786</td>
<td>152571</td>
<td>231516</td>
<td></td>
</tr>
<tr>
<td>Generalized Exponential</td>
<td>4152645</td>
<td>538489</td>
<td>66223</td>
<td>101115</td>
<td>153416</td>
<td>232748</td>
<td></td>
</tr>
<tr>
<td>Generalized Log-logistic</td>
<td>4925448</td>
<td>975477</td>
<td>65818</td>
<td>9924</td>
<td>149575</td>
<td>225354</td>
<td></td>
</tr>
<tr>
<td>Generalized Gamma</td>
<td>4251006</td>
<td>434145</td>
<td>66885</td>
<td>101203</td>
<td>153566</td>
<td>233025</td>
<td></td>
</tr>
<tr>
<td>Mixed Weibull</td>
<td>4556245</td>
<td>867121</td>
<td>66355</td>
<td>100486</td>
<td>152164</td>
<td>230417</td>
<td></td>
</tr>
<tr>
<td>Change Point</td>
<td>4854662</td>
<td>864987</td>
<td>66341</td>
<td>100457</td>
<td>152149</td>
<td>230429</td>
<td></td>
</tr>
<tr>
<td>Immune Invasion</td>
<td>4526488</td>
<td>536745</td>
<td>66601</td>
<td>101004</td>
<td>153198</td>
<td>232317</td>
<td></td>
</tr>
<tr>
<td><strong>When median incubation = 12 years</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Weibull</td>
<td>5145678</td>
<td>1258796</td>
<td>66224</td>
<td>100228</td>
<td>151654</td>
<td>229441</td>
<td></td>
</tr>
<tr>
<td>Gamma</td>
<td>5215674</td>
<td>83486</td>
<td>66526</td>
<td>100855</td>
<td>152898</td>
<td>231755</td>
<td></td>
</tr>
<tr>
<td>Log-logistic</td>
<td>5043259</td>
<td>1395541</td>
<td>66195</td>
<td>100129</td>
<td>151442</td>
<td>229072</td>
<td></td>
</tr>
<tr>
<td>Log-normal</td>
<td>4921486</td>
<td>1099113</td>
<td>66374</td>
<td>100327</td>
<td>152255</td>
<td>230605</td>
<td></td>
</tr>
<tr>
<td>Generalized Exponential</td>
<td>4658415</td>
<td>912247</td>
<td>64847</td>
<td>100741</td>
<td>152702</td>
<td>231411</td>
<td></td>
</tr>
<tr>
<td>Generalized Log-logistic</td>
<td>5428503</td>
<td>1560546</td>
<td>65822</td>
<td>99218</td>
<td>149514</td>
<td>225243</td>
<td></td>
</tr>
<tr>
<td>Generalized Gamma</td>
<td>4628457</td>
<td>558025</td>
<td>66641</td>
<td>101105</td>
<td>153399</td>
<td>232712</td>
<td></td>
</tr>
<tr>
<td>Mixed Weibull</td>
<td>4856477</td>
<td>1253741</td>
<td>66258</td>
<td>100242</td>
<td>151674</td>
<td>229486</td>
<td></td>
</tr>
<tr>
<td>Change Point</td>
<td>4958115</td>
<td>1261466</td>
<td>66205</td>
<td>100266</td>
<td>151408</td>
<td>229454</td>
<td></td>
</tr>
<tr>
<td>Immune Invasion</td>
<td>4726489</td>
<td>733543</td>
<td>66533</td>
<td>100872</td>
<td>152914</td>
<td>231798</td>
<td></td>
</tr>
<tr>
<td><strong>When median incubation = 15 years</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Weibull</td>
<td>5891562</td>
<td>2051215</td>
<td>66124</td>
<td>99965</td>
<td>151089</td>
<td>228348</td>
<td></td>
</tr>
<tr>
<td>Gamma</td>
<td>5945188</td>
<td>1576427</td>
<td>66318</td>
<td>100389</td>
<td>151933</td>
<td>229987</td>
<td></td>
</tr>
<tr>
<td>Log-logistic</td>
<td>5518246</td>
<td>237985</td>
<td>65987</td>
<td>99662</td>
<td>150521</td>
<td>227314</td>
<td></td>
</tr>
<tr>
<td>Log-normal</td>
<td>5259995</td>
<td>1788556</td>
<td>66218</td>
<td>100759</td>
<td>151524</td>
<td>229216</td>
<td></td>
</tr>
<tr>
<td>Generalized Exponential</td>
<td>5348926</td>
<td>2146419</td>
<td>66074</td>
<td>99847</td>
<td>150897</td>
<td>228005</td>
<td></td>
</tr>
<tr>
<td>Generalized Log-logistic</td>
<td>6654897</td>
<td>279684</td>
<td>65818</td>
<td>99201</td>
<td>149451</td>
<td>225142</td>
<td></td>
</tr>
<tr>
<td>Generalized Gamma</td>
<td>4854981</td>
<td>775325</td>
<td>66602</td>
<td>101014</td>
<td>153187</td>
<td>232311</td>
<td></td>
</tr>
<tr>
<td>Mixed Weibull</td>
<td>5624189</td>
<td>2037019</td>
<td>66134</td>
<td>99987</td>
<td>151132</td>
<td>228438</td>
<td></td>
</tr>
<tr>
<td>Change Point</td>
<td>5548050</td>
<td>2065756</td>
<td>66111</td>
<td>99956</td>
<td>151021</td>
<td>228315</td>
<td></td>
</tr>
<tr>
<td>Immune Invasion</td>
<td>5218227</td>
<td>1091589</td>
<td>66465</td>
<td>100714</td>
<td>152605</td>
<td>231204</td>
<td></td>
</tr>
</tbody>
</table>
incubation period models Weibull, Gamma, log-logistic, log-normal and generalized exponential distribution prior estimates of their parameters were available. All these models have only two parameters and therefore one parameter was fixed based on the estimates available in the literature. The parameters of the generalized log-logistic, generalized gamma, mixed Weibull, change point and immune invasion level models were not available in the literature. Therefore parameters of these models were decided based on a simulation study.

The backcalculation estimates were obtained using the conditional likelihood approach for the multinomial likelihood with unknown sample size. Although EM algorithm\textsuperscript{22} can also be used, it has been observed\textsuperscript{23} that both of the approaches yield almost similar estimates. Also the conditional likelihood approach converges at much faster rate and is easy to implement in a computer algorithm. SAS IML programming was done for the actual computation of the estimates. Non-linear optimization routines were extensively used for maximization of the conditional likelihood. The calculations were repeated under logistic incidence, double exponential and root exponential for the different median incubation period distributions. The estimates of the minimum number of HIV cases (\(N\)), HIV incidence for 2002 and short-term projection for next 4 years for four median times under logistic prevalence are presented in the Table 1.

**Adjustments for Reporting Delays and Under-Reporting:**

According to WHO Global HIV/AIDS Survey\textsuperscript{67}, the level of under reporting in the South-Eastern Asian Countries including India is about 80%. The various small studies conducted in different parts of India reported the level of under-reporting between 50% - 95 %. Vandal and Remis\textsuperscript{68} used 75% upward adjustment for projections for Canada surveillance data. In this work we assume 90% under-reporting during 1986-87 and it decreases to 60% in 2001-02 in an exponential decay form. The upward adjustments for different years are carried out. The reporting delays were modeled by many authors. In this work we have used the approach proposed\textsuperscript{69}. The over-dispersed Poisson regression model with modifications as adopted\textsuperscript{68} for the reporting delays was used and the adjusted figures are used for recalculating the estimates under the above model assumptions, median incubation distributions and infection curves. If the adjusted AIDS counts are taken to the upper bound of the under reporting and delay in reporting, the figure in the estimates give the upper limit of the number of cases expected. The adjusted estimates of the minimum number of HIV cases (\(N\)), HIV incidence for 2002 and short-term projection for next 4 years for four median times under logistic prevalence are presented in the Table 2.

**DISCUSSION**

It is generally observed that the short term projected AIDS cases do not vary much across various infection densities and
incubation period distributions. But the minimum size of the epidemic and HIV incidences are highly variable across the infection densities and incubation period distributions. The projected AIDS cases within a infection density across various incubation distributions are found to be very stable. But across the infection densities the variation is observed to be high. The projected AIDS cases using the log-logistic and root exponential infection densities are less compared to the estimates obtained using the other four infection densities. The projected AIDS estimates obtained using the four infection densities logistic prevalence, logistic incidence, exponential and double exponential seems to be more plausible compared to the estimates obtained using the other two models since the observed AIDS cases in the year 2000 were nearly 20000. Hence based on the four infection densities and 10 incubation period distributions with four possible median incubation periods (several possible combinations for the parameters of these distributions), we infer that the projected AIDS cases in India for the years 2003, 2004, 2005 and 2006 will be approximately around 27000, 43500, 70000 and 113500 respectively. It is to be noted that these estimates are based on the unadjusted AIDS incidence data. These estimates may not be the correct number of AIDS cases that may develop in India during these periods. The exact number of AIDS cases that may develop in India will be certainly higher than these figures and hence these figures can be taken to be a lower bound for possible number of AIDS cases. The HIV incidence reported for the year 2002 is calculated based on the relationship between the infection density and the infection or incidence curves. These figures are found to be highly variable and are not smoothed estimates. Hence these figures cannot be taken as exact number of HIV incidence in the year 2002. The minimum size of the epidemic denotes the total number of people who may ultimately become AIDS cases even if the new infections are currently stopped. Hence this figure can be taken to be the number of people living with HIV/AIDS in India and the adjusted AIDS counts can be taken to the upper bound. For all combinations of eight infection densities and ten incubation period distributions, the increase as median incubation period increases.

One limitation of this study is that the various staging models incorporating effect of therapy, different risk groups and other methods for reporting delays and underreporting were not considered due to non-availability and non-accessibility of the necessary data. AIDS incidence data among all cases reported up to Dec. 2002 and diagnosed before Dec. 2002 were adjusted for reporting delays and underreporting and the results were compared. Back-calculation was performed on the adjusted data using flexible models and conditional maximum likelihood estimation method. The models used were time variant, which allowed for testing availability and therapy effects on disease progression. The HIV incidence estimates on the
cumulative incidences ascertained through simulation account for variability associated with adjustments and random components of the models. The estimates obtained are lower than the estimates given by other international bodies even at extreme model assumption. Data used consists of predominantly adult population for back-calculation. There is an increase in trend among the children. The non-parametric backcalculation approach is an alternative which uses mild assumptions for projections.

REFERENCES


52. Chiarotti F, Palombi M, Schinaia N, Ghirardini A and Bellocco R. Median


