Statistical Model and Method in Ensuring Validity of HIV Data

M. A. Shah

ABSTRACT

Human immunodeficiency Virus (HIV)/ Acquired Immune Deficiency Syndrome (AIDS) is spreading very fast over the globe. The urban infrastructure in Bangladesh is most developed in the three cities, Dhaka, Chittagong and Sylhet attracting large foreign investments and greater volumes of commerce and trade. These three cities are also experiencing an increasing influx of people from across the country. Keeping pace with the rapid urbanization, HIV is also going up. But conventional surveys, conducted so far, seems underestimated the number of HIV infected cases. Proper estimation of HIV infected people is, therefore, essential to effectively address the issue. That purpose Pure Birth process model has been formulated to estimate the HIV for the age group 15-49 years.

Key words: AIDS, HIV, Poisson process, pure birth process model.

INTRODUCTION

Acquired Immune Deficiency Syndrome (AIDS) has become one of the most serious epidemics or rather pandemic which has spread to every continent. AIDS is the full-blown disease of Human-Immune Deficiency Virus (HIV) infection. It has already claimed the lives of more than 23 million and to this 3 million of deaths are added every year. The number of HIV infected people is increasing with a galloping pace and the situation is quite alarming posing a challenge to the mankind. According to WHO report 2002, an estimated 42 million people throughout the globe are currently living with HIV.

It was estimated that at the end of 2001 there were 4.2 million people living with HIV/AIDS in South Asia - Afghanistan, Bangladesh, Bhutan, India, Maldives, Nepal, Pakistan and Sri Lanka. Much is now known about the disease in South Asia. Recently, a report on the AIDS cases came out in the daily newspaper according to that report, more than 13 thousand people are carrying HIV in Bangladesh and the experts have classified the AIDS situation as concentrated epidemic.1 Ironically, the government was downplaying the actual AIDS cases, just informing only 282 and taking credit from the international community. But in reality, the actual figure would be far higher than the reported one. Due to stigma and fear of discrimination, HIV infected people are so scared to come forward with their diseases and ask for treatment and health care. It can make people hide their HIV status amidst fear of rejection from their loved ones. Although the HIV prevalence rate is still low in Bangladesh, behavioral patterns suggest that the number of people infected with HIV could reach epidemic proportions unless concerted efforts are undertaken to prevent it.

The first HIV-positive person diagnosed in Bangladesh was in 1989. Since then it is increasing steadily to become approximately 7,500 in 1994 and to become 11,000 in 2005. According to International Center for Diarrhoeal Disease Research, Bangladesh (ICDDRB) the number of cases of HIV/AIDS has increased to 1495 at the end of 2008. However UNAIDS estimates that the number of people living with HIV in the country may be as high as 12,000, which is within the range of the low estimate by UNICEF's State of the World's Children Report 2009. Although the number of HIV/AIDS cases are increasing steadily over the years, the estimates are ambiguous. The overall prevalence

Authors' Information:
Professor Dr. M A Shah, Professor, Department of Statistics, University of Rajshahi, Rajshahi, Bangladesh.
Mobile: 01555-312554, E-mail: aaz61@yahoo.com
of HIV in Bangladesh is less than 1%. According to government sources, there were 363 reported HIV-positive cases as of December 2003. Of them, 12.3% are adolescents and youth (aged 15-24) and 3.7% are infants and children. However, it is believed that the actual number of cases is much higher. According to UNAIDS, by the end of 2001 the estimated number of adults in Bangladesh living with HIV, irrespective of whether or not they had developed symptoms of AIDS, was 13,000. More recent data of National AIDS/STD program demonstrate that new HIV, new AIDS and new AIDS deaths in 2010 were 343, 231 and 37 respectively. The underlying causes of the HIV/AIDS epidemic include poverty, gender inequality and high mobility of the population. All of which are commendably high in Bangladesh. Bangladesh is a densely populated country with its 30% population living in urban area. Continuous influx of people into the urban areas from across the country making the cities unsafe and unhealthy for future living. Emigration to other countries for employment is also very common, particularly amongst younger people, largely to the Middle East, followed by Singapore and Malaysia.

Principal metropolitan cities of Bangladesh are Dhaka, Chittagong and Sylhet. Dhaka, the capital of Bangladesh, is one of the major cities of South Asia. Dhaka, along with its metropolitan area has a population of over 103,56500, making it the largest city in Bangladesh. It is the ninth largest city in the world. Chittagong has a population of over 39,20222, making it the second largest city in the country and is a major commercial and industrial center as much of Bangladesh's export and import passes through the port of Chittagong. According to a report released by International Institute for Environment and Development, Chittagong is among the ten fastest growing cities in the world. Sylhet is another major city in northeastern Bangladesh and approaching a population of 23,70000 people according to the Geo Names geographical database. The city, however, is currently known for its business boom - being one of the richest cities in Bangladesh, with new investments in hotels, shopping malls and luxury housing estates, brought mainly by expatriates living in Europe. The results of the 4th round of the National Serological Surveillance has shown an alarming increase in HIV rates among the injectable drug users (IDU) in Dhaka. In this group of people, prevalence has jumped from 1.7 per cent in 2002 to 4 per cent in 2003. This is just short of the 5 per cent mark required for a situation to be identified as a concentrated epidemic.

Laxmi et al. and Bashir et al. used statistical models for estimating the HIV rate for three different states of India. Lubboobi, Anderson and associates and Anderson and May have investigated the HIV in deterministic approach. Anderson and May have also considered vaccination strategies in age-structured populations. Most of the models so far investigated were within the framework of a deterministic environment. However, real environment is full of fluctuations. Further, the human behavior is stochastic rather than deterministic. Only very recently, it has been felt that stochasticity is the common feature of all the processes evolving time and space, be they related to natural science, social science, engineering and health science. The purpose of the present study is, therefore, to formulate a stochastic model for estimating the number of HIV infected persons in Dhaka, Chittagong and Sylhet divisions of Bangladesh.

**MATHEMATICAL FORMULATION**

In many situations the objective of an analysis consists of merely observing the number of units that enter the system. The model in which only the addition are counted and no subtractions take place are called Pure birth models. As such, the Pure birth models are not of much importance so far as their applicability to real life situation is concerned, but these are very important in the understanding of completely random arrival problems. Every new HIV infection can be treated as an addition to the HIV/AIDS family. Let be the rate of infection and be the initial population of the system. Thus the Pure birth process can be considered as statistical model for the analysis as follows.
We now wish to determine the probability on \( n \) additions in a time interval of length \( t \), denoted by \( P_n(t) \). Clearly, \( n \) will be an integer greater than or equal to zero. To do so, we shall first develop the differential equations governing the process in three different situations. For \( n > 0 \), there may be three mutually exclusive ways of having \( n \) units at time \( t + \delta t \).

(i) There are \( n \) units in the system at time \( t \) and no birth takes place during time interval \( \delta t \). Hence, there will be \( n \) units at time \( t + \delta t \).

(ii) Alternatively, there are \((n-1)\) units in the system at time \( t \), and one event takes place during \( \delta t \). Hence there will remain \( n \) units in the system at time \( t + \delta t \).

(iii) Further, there are \((n-2)\) units in the system at time \( t \), and two events take place during \( \delta t \). Hence there will remain \( n \) units in the system at time \( t + \delta t \).

Therefore, the probability of those two combined events explained in (i) will be

\[ P_n(t) = P_n(t) \left[ 1 - \lambda_n(t) \delta t + O(\delta t^2) \right] \]

the probability of those two combined events explained in (ii) will be

\[ P_{n-1}(t) \lambda_{n-1}(t) \delta t + O(\delta t^2) \]

and finally, the probability of those two combined events explained in (iii) will be

\[ P_{n-2}(t) \lambda_{n-2}(t) \delta t + O(\delta t^2) \]

Now, adding these three probabilities, we get the probability of events at time \( t + \delta t \),

\[ P_n(t + \delta t) = P_n(t) - \lambda_n(t) P_n(t) \delta t + O(\delta t^2) \]

The model is studied in considering three possible independent events \( n, n-1 \) and \( n-2 \). Since the probability of more than one or two events in \( \delta t \) is assumed to be negligible, other alternatives do not exist. Assuming that more than one person can not be infected simultaneously for the same source at the same time, and then \( O(\delta t^2) \to 0 \).

Thus,

\[ P_n(t + \delta t) = P_n(t) - \lambda_n(t) P_n(t) \delta t + O(\delta t^2) \]

Dividing both sides by \( \delta t \) and then taking limit \( \delta t \to 0 \), the first derivative results

\[ \frac{d}{dt} P_n(t) = \frac{d}{dt} P_n(t) - \lambda_n(t) P_n(t) + \lambda_{n+1}(t) P_{n+1}(t) - \lambda_{n-1}(t) P_{n-1}(t) + O(\delta t) \]

Since, \( E(N = n) = \lambda n \), Eq.(1) becomes

\[ P_n(t) = -\lambda(t) P_n(t) + (n+1) \lambda(t) P_{n+1}(t) + (n-1) \lambda(t) P_{n-1}(t) \]

for \( n > 2 \) (2)

**Solution Process:**

In order to solve the Eq.(2), we have used the probability generating function (p.g.f) method,

\[ P(u,t) = \sum_{n=0}^{\infty} P_n(t) u^n \]

Multiplying Eq.(2) by \( u^n \) and summing over \( n \), we have

\[ \sum_{n=0}^{\infty} P_n(t) u^n = \sum_{n=0}^{\infty} P_{n+1}(t) u^{n+1} + \lambda(t) \sum_{n=0}^{\infty} (n+1) P_{n+1}(t) u^n \]

so that

\[ \frac{\partial}{\partial t} P(u,t) = \sum_{n=0}^{\infty} P(t) u^n \text{ and } \frac{\partial}{\partial u} P(u,t) = \sum_{n=0}^{\infty} n P_n(t) u^{n-1} \]

Multiplying Eq.(5) by \( \lambda u, \lambda u^2 \text{ and } \lambda u^3 \) in succession we have

\[ \lambda u \frac{\partial P}{\partial u} = \lambda \sum_{n=0}^{\infty} n P_n(t) u^n \]

\[ \lambda u^2 \frac{\partial P}{\partial u} = \lambda \sum_{n=0}^{\infty} n^2 P_n(t) u^{n-1} \]

\[ \lambda u^3 \frac{\partial P}{\partial u} = \lambda \sum_{n=0}^{\infty} n^3 P_n(t) u^{n-2} \]

From the very beginning we have considered three consecutive possibilities, namely \( P_n(t) \), \( P_{n-1}(t) \) and \( P_{n-2}(t) \) we may assume in Eq.(5) and Eq.(5b), \( n \to (n-1) \). And in Eq,(5a) and Eq.(5c), \( n \to (n-2) \) we have

\[ \lambda u \frac{\partial P}{\partial u} = \lambda \sum_{n=0}^{\infty} (n-1) P_{n-1}(t) u^n \]

\[ \lambda u^2 \frac{\partial P}{\partial u} = \lambda \sum_{n=0}^{\infty} (n-2) P_{n-2}(t) u^n \]

Using Eq.(5), Eq.(5a), Eq.(5d) and Eq.(5e) in Eq.(4) we have

\[ \frac{d}{dt} P(u,t) = -\lambda u \frac{\partial P}{\partial u} + \lambda u^2 \frac{\partial P}{\partial u} + \lambda u^3 \frac{\partial P}{\partial u} - \lambda u \frac{\partial P}{\partial u} \]

or

\[ \frac{d}{dt} P(u,t) = -\lambda u \left[ u^3 \right] \frac{\partial P}{\partial u} - \lambda u \frac{\partial P}{\partial u} = 0 \]
With the help of Lagrange's linear equation we have,

\[ \frac{dt}{1} = \frac{dP}{0} \]  

and

\[ \frac{dt}{1} = \frac{du}{\lambda u(1-u-u^2)} \]  

Eq. (6a) gives,

\[ P = c_t \]  

where \( c \) is a constant of integration.

Further, Eq. (6b)

\[ 3\,dt = \int \frac{du}{u(1-u-u^2)} \]  

Therefore solving by partial fraction we have

\[ \lambda \,dt = -\frac{1}{u} \,du + \frac{2}{5\sqrt{5}} \int \frac{1}{u+\frac{1}{2}} \,du + \frac{2}{5\sqrt{5}} \int \frac{1}{u+\frac{1}{2}} \,du \]

\[ \lambda \,t = \log \left[ 1 \left( \frac{u+\frac{1}{2}}{u+\frac{1}{2}} \right)^\frac{2}{5\sqrt{5}} \left( \frac{u+\frac{1}{2}}{u+\frac{1}{2}} \right)^\frac{2}{5\sqrt{5}} \right] \]

\[ e^{\lambda t} = P \left( \frac{1}{u+\frac{1}{2}} \right)^{\frac{2}{5\sqrt{5}}} \left( \frac{u+\frac{1}{2}}{u+\frac{1}{2}} \right)^{\frac{2}{5\sqrt{5}}} \]

Using \( P = c (\text{constant}) \), we have

\[ P(u, t) = e^{\lambda t} \left( \frac{1}{u+\frac{1}{2}} \right)^{\frac{2}{5\sqrt{5}}} \left( \frac{u+\frac{1}{2}}{u+\frac{1}{2}} \right)^{\frac{2}{5\sqrt{5}}} \]  

(7)

It is mentionable that the rate of infection and initial population are \( \lambda \) and \( u \) respectively. Therefore,

\[ P_1(u, t) = P(u, t)/100 \]  

and

\[ P_2(u, t) = uP_1(u, t)/100 \]  

(8)

(9)

This form of Birth Process Model can be used to estimate the number of infected persons for the time interval \( t \) when the rate of infection \( \lambda \) and the initial population \( u \) are known.

Illustration:

**TABLE I(a): Probability for the possible values \( n=0,1 \) for the Classical Process**

<table>
<thead>
<tr>
<th>Time</th>
<th>n values</th>
<th>Total Probability</th>
</tr>
</thead>
<tbody>
<tr>
<td>0</td>
<td>0.90955301</td>
<td>0.08622744</td>
</tr>
<tr>
<td>2</td>
<td>0.8272867</td>
<td>0.15668569</td>
</tr>
<tr>
<td>3</td>
<td>0.7524611</td>
<td>0.21400444</td>
</tr>
<tr>
<td>4</td>
<td>0.68440323</td>
<td>0.25953124</td>
</tr>
<tr>
<td>5</td>
<td>0.62250108</td>
<td>0.2950717</td>
</tr>
<tr>
<td>6</td>
<td>0.5661977</td>
<td>0.32206004</td>
</tr>
<tr>
<td>7</td>
<td>0.5149868</td>
<td>0.3417525</td>
</tr>
</tbody>
</table>

**TABLE I(b): Probability for the possible values \( n=0,1,2 \) for the Classical Process**

<table>
<thead>
<tr>
<th>Time</th>
<th>n values</th>
<th>Total Probability</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>0.90955308</td>
<td>0.08622744</td>
</tr>
<tr>
<td>2</td>
<td>0.8272867</td>
<td>0.15668569</td>
</tr>
<tr>
<td>3</td>
<td>0.7524611</td>
<td>0.21400444</td>
</tr>
<tr>
<td>4</td>
<td>0.68440323</td>
<td>0.25953124</td>
</tr>
<tr>
<td>5</td>
<td>0.62250108</td>
<td>0.2950717</td>
</tr>
<tr>
<td>6</td>
<td>0.5661977</td>
<td>0.32206004</td>
</tr>
<tr>
<td>7</td>
<td>0.5149868</td>
<td>0.3417525</td>
</tr>
</tbody>
</table>

**TABLE I(c): Probability for the possible values \( n=0,1,2,3 \) for the Classical Process**

<table>
<thead>
<tr>
<th>Time</th>
<th>n values</th>
<th>Total Probability</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>0.90955308</td>
<td>0.08622744</td>
</tr>
<tr>
<td>2</td>
<td>0.8272867</td>
<td>0.15668569</td>
</tr>
<tr>
<td>3</td>
<td>0.7524611</td>
<td>0.21400444</td>
</tr>
<tr>
<td>4</td>
<td>0.68440323</td>
<td>0.25953124</td>
</tr>
<tr>
<td>5</td>
<td>0.62250108</td>
<td>0.2950717</td>
</tr>
<tr>
<td>6</td>
<td>0.5661977</td>
<td>0.32206004</td>
</tr>
<tr>
<td>7</td>
<td>0.5149868</td>
<td>0.3417525</td>
</tr>
</tbody>
</table>

**Caption of Tables and Figures**

When rate = \( \lambda = 0.079361 \), the probability that none and single AIDS cases are presented in Table 1(a). For \( n=0,1,2 \) and \( n=0,1,2,3 \), the corresponding probabilities are shown in Tables 1(b) & 1(c) for different times. It has been observed that the number of independent cases are increases during, \( \delta t \) the probability values are decreases and holds the Poisson property. So that we can not ignore the importance of two/three or more AIDS cases by epoch \( t \). In our theoretical development we have considered \( (n-2) \) events by epoch \( t \) and two events between \( t \) and \( (t+\delta t) \).

The above tabular results are also represents that the probabilities are gradually decreasing provided the total probabilities are close to one. The Figures 1(a)-(c), 2(a)-(c) and 3(a)-(c) represent the conception as narrated in the above tables. It is to be noted that we have examined in details for the Chittagong division only when \( t=1,2,3 \).
FIGURE 1(a)-(c): when t=1

FIGURE 2(a)-(c): when t=2
FIGURE 3(a)-(c) : when t=3
Using R-language we have determined the dependence of two or more independent cases for Dhaka and Sylhet divisions of Bangladesh and compared with the existing classical process and are presented in the Figures 4(a)-(c) and 5(a)-(c).

FIGURE 4(a)-(c) : when t = 1, 2, 3
**Classical and Proposed Methods:**

Pr\{No AIDS case observed\}=Pr\{N(t)=0\}, i.e., n=0. We have calculated in the Tables-2 for the increasing values of time parameter, t=1, 2, 3, 4, 5, 6, 7. We would like to point out here that our proposed model gives a far better description as compared to the classical model. The model would not be mere mathematical artifacts, but rather it will cover a much larger spectrum of the birth process.

**TABLE 2:**

<table>
<thead>
<tr>
<th>Time t</th>
<th>Classical</th>
<th>Proposed</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>0.9237064</td>
<td>0.9236605</td>
</tr>
<tr>
<td>2</td>
<td>0.8532335</td>
<td>0.8531911</td>
</tr>
<tr>
<td>3</td>
<td>0.7881373</td>
<td>0.7880981</td>
</tr>
<tr>
<td>4</td>
<td>0.7280074</td>
<td>0.7279713</td>
</tr>
<tr>
<td>5</td>
<td>0.6724651</td>
<td>0.6724317</td>
</tr>
<tr>
<td>6</td>
<td>0.6211604</td>
<td>0.6211295</td>
</tr>
<tr>
<td>7</td>
<td>0.5737698</td>
<td>0.5737413</td>
</tr>
</tbody>
</table>

**RESULTS AND DISCUSSIONS**

As mentioned in the introduction the rate of HIV infection represented by \( \lambda \), the time interval \( t \) stands for one year and population infected by HIV \( P_{2t} \). Thus for the age group (15–49) years of the population of Dhaka division, \( \lambda = 0.094802 \) and \( t=1 \). So that Equation (7) results, \( P(z,t) = 1.09944 \).
and the corresponding value of $P_1(z,t) = 0.010994$ and $P_2(z,t) = 62.6249$, is the number of infected persons from the normal population. According to the nature of the problem and size of the population, using Pure Birth process model, the rate of HIV infected persons $P_2(z,t)$, predicted for next ten years for inhabitants of Dhaka city are shown in Table-3. Therefore, $62.6249 \approx 63$ persons will be infected in a year $(t = 1)$ for the population of Dhaka for the age group (15-49 years) and the results are presented in Table-4.

| Table 4: Predicted number of HIV infection $P_1(t)$ for the age-group 15-49 with respect to time $(t)$ in Dhaka. |
|---|---|---|---|---|
| Time | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | 10 |
| HIV | 63 | 69 | 76 | 83 | 92 | 101 | 111 | 122 | 134 | 147 |

Further, $\sum P_2(t) = 998$. It implies that after 10 years there will be about 1000 more cases of HIV in the Dhaka division only and is presented in Figure-6.

![Figure 6: Predicted Number of HIV Infection in Dhaka within a Decade](image)

Finally, $\sum P_2(t) = 599$ and $\sum P_3(t) = 989$, assert that after 10 years there will be around 599 and 989 more cases of HIV cases in Chittagong and Sylhet divisions of Bangladesh and are presented in Figure-7-8.

![Figure 7: Predicted Number of HIV Infection in Chittagong within a Decade](image)
CONCLUSIONS

It has been observed that the expected number of HIV infection in Dhaka, Chittagong and Sylhet are 998, 559 and 989 respectively for the age group 15-19. However, the total number of infections in Chittagong is lower than those in Dhaka and Sylhet. Categorically the estimated HIV rate is much higher in Dhaka, the capital city of Bangladesh as compared to the second largest commercial city Chittagong. Bangladesh is in the unique position to succeed where several other developing countries have not succeeded to keep the HIV epidemic from expanding beyond the current level by initiating comprehensive and strategically viable preventative measures, avoiding a gradual spread of HIV infection from high-risk groups to the general population. Due to the limited access to voluntary counseling and testing services, very few Bangladeshis are aware of their HIV status. Although still regarded as a low prevalence country, Bangladesh remains extremely vulnerable to HIV epidemic. Providing real data on HIV dealing with district-wise independently following infrastructure concepts, parameters can be easily estimated and can be used in perfect decision making.

REFERENCES

1. Hossain MS. AIDS/HIV situation in Bangladesh: A Looming Threat, Scholars Journal 2005; 15, Personal communication to journal@scholarsbangladesh.com


