

# Estimating the number of Cubans infected sexually by human immunodeficiency virus using contact tracing data

Ying-Hen Hsieh,<sup>a</sup> Hector de Arazoza,<sup>b</sup> Shen-Ming Lee<sup>c</sup> and Cathy WS Chen<sup>c</sup>

<b>Background</b>	To estimate the yearly number of people in Cuba who are living with human immunodeficiency virus (HIV) and were infected through sexual contact but who have not developed acquired immunodeficiency syndrome (AIDS). Estimation was made directly from the yearly HIV seroprevalence data of the Cuban Partner Notification Programme from 1991 to 2000.
<b>Methods</b>	The generalized removal model for open populations is utilized for the estimation. The total number of known HIV-infected Cubans at each sampling time is used in the prior to provide more reasonable approximations.
<b>Results</b>	We estimated a yearly survival rate of 93%. The median estimates for the number of all living asymptomatic HIV-positive Cubans, infected by sexual contact, tripled from 714 in 1991 to 2170 in 2000. The number of unknown HIV-positive Cubans infected sexually is estimated to range from 174 in 1991 to 401 in 2000.
<b>Conclusions</b>	A consistent increase in the number of sexually infected HIV-positive individuals in Cuba from 1991 to 2000 is evident from the estimates. From 1996 onwards more sexually active homosexual/bisexual contacts were traced and consequently more sexually-infected HIV-positives were detected. A consequence of increased detection is the levelling off and subsequent decrease in the number of unknown HIV-positives during this time period. The estimation procedure is useful in estimating prevalent population sizes of epidemiological and public health interest.
<b>Keywords</b>	Cuba, epidemiology, HIV/AIDS, Latin America, sexual contact, Bayes statistics, contact tracing
<b>Accepted</b>	January 11, 2002

In December 2000, the United Nations AIDS Programme (UNAIDS) and the World Health Organization (WHO) *Report on the Global HIV/AIDS Epidemic*<sup>1</sup> gave an estimate of 36.1 million people living with human immunodeficiency virus (HIV) worldwide, of which 390 000 live in the Caribbean region. Given the large number of HIV infections in the region (2.3% adult prevalence rate), it is remarkable that Cuba, with a population of 11 million, has had only 3230 cases of HIV-seropositivity by the end of 2000. Prominent among the many reasons for the low prevalence is the National Programme on HIV/AIDS established by the Cuban government in 1983.<sup>2,3</sup> Initially the

programme consisted mainly of testing blood donations and a hospital surveillance system to detect AIDS-related illness. The first HIV-seropositives were detected at the end of 1985. In 1986, when the first AIDS case was diagnosed, an HIV screening programme was initiated for travellers to countries with reported cases of HIV infection. The programme was systematically expanded to include various groups such as patients with sexually transmitted diseases (STD), pregnant women, and hospital patients. Roughly 23 million tests were performed between April 1986 and December 2000 and those with confirmed HIV-seropositivity were treated and placed in controversial sanatorium facilities (see, e.g. refs. 2,4). The sanatorium programme was designed to reduce the probability of transmission by HIV-positive individuals and has gone through several phases. Currently, after their initial admission to a sanatorium, most HIV-infected residents are permitted to return voluntarily to their communities following an evaluation and treatment period.

Another important aspect of the Cuban national AIDS programme is the 'Partner Notification Programme' (PNP) based on contact tracing and screening of the sexual partners of

<sup>a</sup> Department of Applied Mathematics, National Chung-Hsing University, Taichung, Taiwan.

<sup>b</sup> Departamento Ecuaciones Diferenciales, Facultad Matematica-Computacion, Universidad de la Habana, San Lazaro y L Habana 4, Cuba.

<sup>c</sup> Department of Statistics, Feng-Chia University, Taichung, Taiwan.

Correspondence: Ying-Hen Hsieh, Department of Applied Mathematics, National Chung-Hsing University, Taichung, Taiwan 402. E-mail: yhhshieh@dragon.nchu.edu.tw

known HIV-positives. This also began in 1986 and the idea is to search out asymptomatic carriers before they develop AIDS. Indeed, the result is impressive given that 55% of those detected with HIV in Cuba have not developed AIDS. General screening and subsequent admission to a sanatorium have been reduced in recent years due to economic constraints and an evolving sanatorium policy and so the PNP has gradually assumed added importance. Since approximately 90% of the reported AIDS cases in Cuba by the end of 1997 were acquired by sexual (hetero-, homo-, or bisexual) contact,<sup>5</sup> the number of HIV-positives detected via contact tracing should give a good indicator of the size of the HIV-infected population in general. Furthermore, recent growth in tourism has led to a re-emergence of prostitutes in the last few years.<sup>6</sup> Perhaps understandably, recent data have shown an increase in the number of HIV-positives, starting in 1996.<sup>7,8</sup> Hence it is especially worthwhile from the public health point of view to focus our attention on estimating the population size of HIV-positives infected through sexual activity.

Recent AIDS data have shown that approximately 14% of AIDS cases were unknown to the Health Authority before developing AIDS symptoms. To obtain an estimate of the size of unknown HIV-positive population, Arazoza<sup>7</sup> and Lounes<sup>8</sup> applied a mathematical model to compute the theoretical numbers of the known and unknown HIV-positives infected by sexual contact in Cuba. Their results indicate that roughly 20% to 30% of the HIV asymptomatic carriers have not been detected. Recently a 'generalized removal model for open populations' was proposed by Hsieh *et al.*<sup>9</sup> which uses an empirical Bayes approach to estimate the number HIV-infected people in a hidden, hard-to-count population without any knowledge of the population size. The method was employed in a preliminary study of recent trends in HIV infections in Cuba.<sup>10</sup> Here we use this method to estimate the known and unknown numbers of HIV-positives in Cuba infected via sexual contact. We will make use of HIV seroprevalence data from the PNP from 1991 to 2000. In contrast to Hsieh *et al.*,<sup>9</sup> we have additional information in the Cuban data set, i.e. the total number of known individuals in the HIV-infected population at each sampling time. We will employ this additional knowledge in our priors to improve our estimates.

The paper is organized as follows. We briefly describe the data and the statistical method used. We then give the results of our estimates, followed by concluding remarks and comments. Statistical details are given in the Appendix.

## Data and Methods

### Cuban partner notification HIV seroprevalence data

Ever since the emergence of the global HIV pandemic, contact tracing has been debated as a control measure for HIV/AIDS.<sup>11,12</sup> In Cuba, the PNP traces the sexual contacts of known HIV-seropositives.<sup>2,3,4,7,8</sup> The programme is carried out by Epidemiology Departments at all levels of the Cuban health system. Following partner notification, sexual contacts are interviewed, then tested for HIV every 3 months for a period of one year after the last sexual contact with an HIV-positive person. They are observed as long they remain in contact. We wish to estimate the number of HIV-positives, infected through sexual contact, by using the HIV seroprevalence data of the PNP. For our purpose the data were treated statistically as if they were a random sample of all sexually active individuals in Cuba, although clearly the method of their ascertainment inevitably introduced some selection bias.

One of the main focuses of this study is to obtain an estimate for the number of *unknown* HIV-positives in the sexually active population in recent years who have not developed AIDS. Table 1 gives the accumulated yearly number of known HIV-positives living in Cuba at the end of each year from 1991 to 2000. Table 2 is the seroprevalence data from 1991 to 2000 with number of contacts tested, number of HIV-positives detected, and the percentage of positive tests for each year listed.

A recent study gives an estimate of the incubation period for HIV-infected Cubans at roughly 9 years.<sup>8</sup> Consequently, most of those HIV-positives who were infected in or before 1990 have already developed AIDS by 2000. Therefore we choose to use only the prevalence data from 1991 to 2000. Note the number of known HIV-positives includes all modes of transmission. Also note the sharp increase starting 1996 in all three sets of numbers. The increase is caused by the large numbers of homosexual and bisexual HIV-positives detected since 1996. These people tend to be more sexually active and have a larger number of sexual contacts. Subsequently more of their contacts have been traced and tested positive since 1996.

### Methods

The 'generalized removal model for open populations'<sup>9</sup> proposed recently allows only recruitment (of new HIV-infected individuals) and deaths (removal of HIV-infected individuals due to development of AIDS) to occur during the sampling. In this paper, where the number to be estimated is the yearly number

**Table 1** Number of known human immunodeficiency virus (HIV)-infected individuals living in Cuba at the end of the years 1991–2000

Year	1991	1992	1993	1994	1995	1996	1997	1998	1999	2000
No. known HIV+	600	699	717	736	741	871	1099	1306	1507	1965

**Table 2** The Partner Notification Programme seroprevalence data, 1991–2000

Year	1991	1992	1993	1994	1995	1996	1997	1998	1999	2000
No. tested	905	1104	1121	811	773	1204	2037	2172	2241	4089
No. tested HIV <sup>a</sup> +	81	76	44	44	45	80	113	117	113	122
% tested HIV+	8.95	6.88	3.93	5.43	5.82	6.64	5.55	5.39	5.04	2.98

<sup>a</sup> Human immunodeficiency virus.

of HIV-infected individuals within the sexually active population in Cuba, there is no recapture of those HIV-positive individuals detected in previous samplings since it is reasonable to assume that those tested positive will not be tested again. Hence the removal model is the appropriate choice of model to work with. In each sample, a number of subjects (in our case, those with recent contact with known HIV-positives) are selected for testing. Moreover, the sample-taking would exclude anyone who has already developed AIDS symptoms, hence the estimate we obtain is the number of HIV-infected individuals who have not developed AIDS. It does not hinder public health assessment of the AIDS scenario because the size of population with AIDS symptoms can be easily counted from clinical records.

Since it is not possible to obtain a valid estimate of the HIV-infected population using maximum likelihood estimation, we propose a Bayesian estimation procedure. Bayesian inference of a population size for various models has been proposed in the literature (e.g. refs 13–15). The detailed derivation of the model is given in Hseih *et al.*<sup>9</sup> Differing from Hseih *et al.*, there is extra information in this data set for Cuba. Namely, we know the total number,  $R_j$  of known subjects in the HIV population infected by sexual contact (i.e. 90% of the known number of HIV-positives) just before time  $t_j$ . Intuitively, we must have  $N_j \geq R_j$ ; i.e. the number of HIV-infected subjects known to the health authority cannot exceed the number of all HIV-infected subjects. We will make use of this additional knowledge in our priors. Using this extra information our posterior estimates could provide more reasonable approximations.

Statistical details of the method are given in the Appendix. An empirical Bayes analysis of the model is implemented using the Gibbs sampler, a Markov chain Monte Carlo (MCMC) method. Detailed discussions can be found in Casella and George.<sup>16</sup> The Bayes estimates are based on Monte Carlo samples from the Gibbs sampler run of 6000 iterations after 2000 burn-in, and selecting every 5th sampled value.

## Results

We are interested in estimating the number of the HIV-positives infected sexually in Cuba during the years 1991 to 2000 using the annual HIV seroprevalence data of the PNP during the same period (as given in Table 2). In the framework of the generalized removal model, we let the sexually active population in Cuba be the hidden and hard-to-count population in question. Since the fraction of sexually active males is more than three times that of the females,<sup>5</sup> this population group consists mostly of males aged 15–49, both heterosexual and homo-bisexual, which account for 90% of all HIV-positives in Cuba. Arazoza *et al.*<sup>7</sup> gave a mean estimate of the HIV incubation time, i.e. the time from HIV infection to symptomatic AIDS (when patients usually die within a year) of HIV-positives in Cuba of 9.2 years

(SD 0.273, median 8.42 years, 95% CI : 7.5–8.83). For our work, we defined the yearly survival rate  $\phi$  for HIV-infected individuals (i.e. the mean probability that an infected individual will not develop AIDS during the 12 months between samples) as 93%. A yearly survival rate of 93% implies that the survival rate after 9.5 years is approximately 50% ( $0.93^{9.55} = 0.5$ ), resulting in a median survival time of approximately 9.5 years. For a detailed discussion on the choice of constant survival rate, see Hseih *et al.*<sup>9</sup> For the yearly number of known HIV-positives by sexual contact  $R_j$ , we use the yearly number of known HIV-positives in Table 1, but multiplied by 90% since data have shown that approximately 90% of all HIV-positives in Cuba are infected sexually.<sup>5</sup> Table 3 lists median, mean, standard error, and 95% CI for  $N_j$  obtained from 2.5% and 97.5% quantiles.

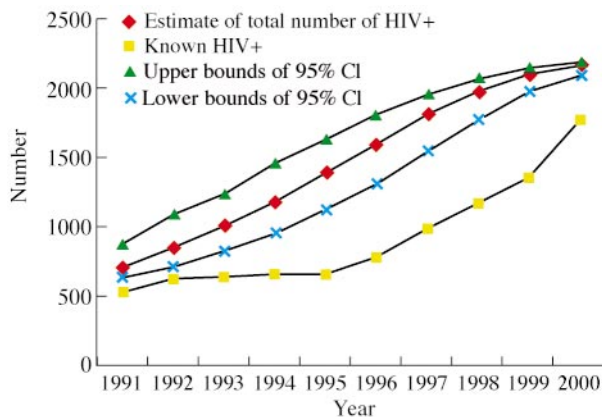
For the purpose of comparing our result with other studies, we consider the following. In Figure 1, we plot our estimate of the total number of HIV-positives infected by sexual contact using a 93% survival rate with the 95% CI against 90% of the number of known HIV-positives in Table 1. The estimates are much higher, mainly due to the discrepancy caused by the unknown HIV-positives unaccounted for in the government data. Our result also shows there is consistent increase in the number of sexually infected HIV-positives in Cuba throughout the decade, despite the relatively unchanged number in the known data from 1992 to 1995, and the sharp increase from 1996 to 2000. To further understand our result in the context of public health concerns, we consider the estimate in terms of undetected HIV-positives.

Arazoza *et al.*<sup>7</sup> used a mathematical model to compute the numbers of known and unknown HIV-positives by sexual contact in Cuba. Here we compute the estimated number of unknown HIV-positives in Cuba by subtracting 90% of the known HIV-positives (which approximates the number of known HIV-positives by sexual contact) from our estimated total number in Table 3. Estimate I in Table 4 is our estimate of unknown HIV-positives. Estimate II is the estimate by Arazoza *et al.*<sup>7</sup> Our results show the increase in the number of unknown HIV-positives starts to level off in 1996, resulting in a decrease in number of unknown HIV-positives since 1998. This is consistent with the sharp increase in the number of HIV-positive contacts detected since 1996.

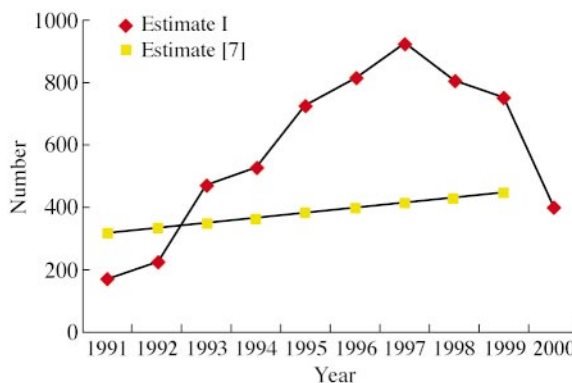
We give the plots of our estimate of unknown HIV-positives and the theoretical estimate from Arazoza *et al.*<sup>7</sup> in Figure 2. Contrary to the theoretical estimate of Arazoza *et al.*,<sup>7</sup> our estimate clearly shows that a sharp increase in the number of unknown HIV-positives starting in 1992 was brought under control by the large number of detections from 1996 onwards. Note that this change of trend is not captured by the estimate in Arazoza *et al.*,<sup>7</sup> mainly because the model in Arazoza *et al.* is a deterministic differential equations model which yields an estimate that is essentially a smooth curve over the time period considered.

**Table 3** Estimated number of human immunodeficiency virus (HIV)-infected individuals by sexual contact in Cuba, 1991–2000

Year	1991	1992	1993	1994	1995	1996	1997	1998	1999	2000
Median estimate	714	857	1016	1188	1393	1602	1816	1981	2110	2170
Mean	728	870	1023	1198	1389	1594	1796	1966	2097	2162
SE	60	95	111	129	130	126	104	75	46	25
95% CI	651–886	721–1100	831–1245	962–1462	1128–1635	1318–1811	1555–1953	1776–2065	1980–2143	2099–2185



**Figure 1** The known number of human immunodeficiency virus (HIV)-positives by sexual contact in Cuba versus the estimated total (known + unknown) number of HIV-positives by sexual contact from 1991 to 2000 by generalized removal model



**Figure 2** The number of unknown HIV-positive people by sexual contact in Cuba from 1991 to 2000. The solid line is the estimate by generalized removal model and the broken line is the theoretical number computed by ref. 7

**Table 4** Estimated number of unknown HIV-infected individuals by sexual contact living in Cuba during the years 1991–2000

Year	1991	1992	1993	1994	1995	1996	1997	1998	1999	2000
Estimate I	174	227	470	525	726	818	926	805	753	401
Estimate II <sup>7</sup>	316	333	350	366	382	398	414	430	446	

### Discussion and Concluding Remarks

Our results show that, in agreement with the seroprevalence data, there is an increase in the number of HIV-positives in Cuba from 1996 to the present. The increase is consistent with the increase in the first half of the 1990s. Among the many possible underlying causes of the sharp increase in the government data is the fact that many more homosexual/bisexual contacts were traced and subsequently more HIV-positives detected from 1996 onwards. As a consequence, the number of unknown HIV-positives levelled off and has actually decreased since 1998.

We wanted to estimate the number of HIV-positives in the sexually (both heterosexual and homo-bisexual) active population. Hence we consider the sexual contacts of known HIV-positives who have been traced as a random sample of the sexually active population since they evidently are sexually active. However, as those traced and tested are known to have had contact with at least one HIV-positive in the past, the capture probability might be higher than it would be in a truly random sample. Consequently, this may result in some overestimation of the true numbers. On the other hand, it is generally unknown whether the contacts are made when the HIV-positive person is already infective, due to variance in infection time and progression of disease. Therefore the actual effect of this uncertainty on the estimate is not clear. In this respect, a possible future research direction is to improve the method by considering the detailed individual contact tracing data of the HIV-positives in Cuba. That would require a much more complicated and difficult model which is beyond our scope.

A full discussion of the advantages as well as the drawbacks and limitations of the generalized removal model is given in Hsieh *et al.*<sup>9</sup> It suffices to point out the difficulty in obtaining

information regarding hidden and elusive populations such as the sexually active population in a society. In practice, the dilemma has proved to be even more challenging in the context of HIV epidemic. This work and Hsieh *et al.*'s previous paper,<sup>9</sup> in which we estimated the number of HIV-infected people in elusive, hard-to-count population groups, demonstrate the usefulness of the generalized removal model, not just in estimating the HIV-infected population sizes, but any prevalent population size of epidemiological and public health interest. As long as two or more (non-overlapping) random samplings of the prevalence data are obtained, one can use it to make inference of the prevalent population size.

### Acknowledgements

Ying-Hen Hsieh, Shen-Ming Lee and Cathy WS Chen are supported by grants from National Science Council of Taiwan. The authors wish to thank the Sanatorium for Persons Living with HIV in Santiago de Las Vegas, Havana, Cuba for the data used in this work.

### References

- 1 Joint United Nations Programme on HIV/AIDS and World Health Organization, *AIDS Epidemic Update: December 2000*. Geneva: UNAIDS/WHO, December 2000.
- 2 Granich R, Jacobs B, Mermin J, Pont A. Cuba's National Aids Program —The First Decade. *West J Med* 1995;**163**:139–44.
- 3 Pérez Avila J, Peña Torres R, Joanes Fiol J, Lantero Abreu M, Araza Rodríguez H. HIV control in Cuba. *Biomed Pharmacother* 1996;**50**: 216–19.



<sup>4</sup> Swanson JM, Gill AE, Wald K, Swanson KA. Comprehensive care and the sanatorium: Cuba's response to HIV/AIDS. *J Assoc Nurses AIDS Care* 1995;**6**:33–41.

<sup>5</sup> Joint United Nations Programme on HIV/AIDS and World Health Organization, *Epidemiological Fact Sheet on HIV/AIDS and Sexually Transmitted Diseases: Cuba*. Geneva: UNAIDS/WHO, June 1998.

<sup>6</sup> Burr C. Assessing Cuba's approach to contain AIDS and HIV. *Lancet* 1997;**350**:647.

<sup>7</sup> de Arazoza H, Lounes R, Hoang T, Interlan Y. Modeling HIV epidemic under contact tracing—The Cuban case. *J Theoretical Medicine* 2000;**2**:267–74.

<sup>8</sup> Lounes R, de Arazoza H. A two-type model for the Cuban national programme on HIV/AIDS. *IMA J Math Appl Med Biol* 1999;**16**:143–54.

<sup>9</sup> Hsieh YH, Chen CWS, Lee SM. Empirical Bayes approach to estimating the number of HIV-infected individuals in hidden and elusive populations. *Stat Med* 2000;**19**:3095–108.

<sup>10</sup> Hsieh YH, Chen CWS, Lee SM, de Arazoza H. On the recent sharp increase in HIV infections in Cuba. *AIDS* 2001;**15**:426–28.

<sup>11</sup> Burr C. The AIDS exception: Privacy vs public. *Atlant Month* June 1997;**279**:57–67.

<sup>12</sup> Rutherford G, Woo J. Contact tracing and the control of human immunodeficiency virus. *JAMA* 1988;**259**:3609–10.

<sup>13</sup> Lee SM, Chen CWS. Bayesian inference of population size for behavioral response models. *Statistica Sinica* 1998;**8**:1233–47.

<sup>14</sup> Chen CWS, Lee SM, Hsieh YH, Ungchusak K. A unified approach to estimating population size of births only model. *Computational Statistics and Data Analysis* 1999;**32**:29–46.

<sup>15</sup> George EI, Robert CP. Capture-recapture estimation via Gibbs sampling. *Biometrika* 1992;**79**:677–83.

<sup>16</sup> Casella G, George EI. Explaining the Gibbs sampler. *American Statistician* 1992;**46**:167–74.

## Appendix

### Statistical details

We consider a sequence of  $s$  samples taken from the seroprevalence data of the Partner Notification Programme. Let  $t_j$  be the time when the  $j$ th sample is taken and let  $B_j$  be the number of HIV-positive individuals newly infected sexually between time  $t_j$  and time  $t_{j+1}$ . Assume also that all subjects in the HIV-infected population just before time  $t_j$  who have not been detected in the first  $j - 1$  samples have the same capture probabilities  $P_j$  in the  $j$ th sample. We define  $N_j$  to be the total number of subjects in HIV-infected population just before time  $t_j$ , and  $N_j = B_0 + \dots + B_{j-1}$ .

The likelihood function can be obtained as follows:

$$L(\mathbf{B}, \mathbf{P} \mid D) \propto \left\{ \prod_{j=1}^s \binom{N_j - M_j}{u_j} P_j^{u_j} (1 - P_j)^{N_j - M_{j+1}} \right\} \tag{1}$$

where  $D = \{u_1, \dots, u_s\}$ ,  $\mathbf{B} = (B_0, \dots, B_{s-1})$ ,  $\mathbf{P} = (P_1, \dots, P_s)$ , and  $u_j$  is the number of HIV-infected individuals by sexual contact captured in the  $j$ th sample. Therefore,  $M_{j+1} = u_1 + \dots + u_j$  is the number of observed HIV-infected individuals in the first  $j$  samples. We call this model a generalized removal model for open populations<sup>9</sup> due to the removal of the observed HIV-infected individuals from the (open) sexually active population,  $M_j$ ,  $j = 1, 2, \dots$ , in each sampling occasion.

Suppose that the prior distribution of  $(\mathbf{N}, \mathbf{P})$  where  $\mathbf{N} = (N_1, \dots, N_s)$  is given by  $\pi(\mathbf{N}, \mathbf{P}) = \pi(N_1, \dots, N_s) \pi(\mathbf{P})$ . This asserts that  $\mathbf{N}$  and  $\mathbf{P}$  are *a priori* independent. We assume that the priors of  $P_j$ 's are *a priori* independent and follow a Beta distribution  $Be(\gamma_1, \gamma_2)$ . In addition, let

$$\pi(N_j) \propto I_{S_j}(N_j) \tag{2}$$

where  $I_{S_j}(\cdot)$  is an indicator function of  $N_j$  and  $S_j$  is the set of  $N_j$  with  $N_j \geq R_j$ .

The assumption on the prior of  $\mathbf{N}$  in (2) is appropriate since intuitively  $N_j$  must be larger than  $R_j$ .

Such priors lead to conditional posteriors of the forms:

$$\pi(\mathbf{P} \mid \mathbf{N}, D) = \prod_{j=1}^s Be(u_j + \gamma_1, N_j - M_{j+1} + \gamma_2) \tag{3}$$

$$\pi(N_j \mid \mathbf{N}_{(-j)}, \mathbf{P}, D) = \frac{\binom{(N_j - M_{j+1}) + (u_j + 1) - 1}{u_j} P_j^{u_j + 1} (1 - P_j)^{N_j - M_{j+1}}}{\sum_{N_j = \max\{N_j, R_j, M_{j+1}\}}^{N_{j+1}} \binom{(N_j - M_{j+1}) + (u_j + 1) - 1}{u_j} P_j^{u_j + 1} (1 - P_j)^{N_j - M_{j+1}}} \tag{4}$$

where  $\mathbf{N}_{(-j)}$  denotes the vector  $\mathbf{N}$  with the  $N_j$  deleted.  $(N_j - M_{j+1})$  follows a truncated negative binomial with parameters  $u_j + 1$  and  $P_j$ , and  $\max\{N_{j-1}, R_j, M_{j+1}\} \leq N_j \leq N_{j+1}$ . Moreover, we know that  $N_1 \geq R_1$ . Subsequently one can easily implement the Gibbs sampler to generate  $(N_j - M_{j+1})$  from the truncated negative binomial in Equation (4), and therefore the estimates of  $N_j$  can be obtained. Since there are AIDS-related deaths during the process, we define the yearly survival rate specific to an HIV-infected individual between the  $(j - 1)$ th and  $j$ th sample to be  $\phi$ . The conditional expectations of  $M_{j+1}$  and  $N_{j+1}$  for  $(j + 1)$ th sample given  $M_j$  (the number of distinct HIV-infected individuals captured in the first  $j - 1$  samples) and  $N_j$  (the total number of subjects in HIV-infected population just before time  $t_j$ ), respectively, are

$$E(M_{j+1} \mid M_j) = \phi M_j + u_j \text{ and } E(N_{j+1} \mid N_j) = \phi N_j + B_j \tag{5}$$

We assume that the natural death-rate of the sexually active population during this time period is negligible compared to the AIDS-related death rate since the majority of the subjects in question are aged 15–49 when the natural mortality is low.