Available online at www.sciencedirect.com
SCIENCE DIRECT
Physica A 362 (2006) 495–503
PHYSICA A
www.elsevier.com/locate/physa

Estimating the Number of HIV-infected gay sauna patrons in Taipei area

Ying-Hen Hsieh[a,*], Cathy W.S. Chen[b], Shen-Ming Lee[b], Yi-Ming A. Chen[c,d], Shiow-Ing Wu[d,e], Shu-Fen Lai[c,d], An-Lung Chang[f]

[a]Department of Applied Mathematics, National Chung Hsing University, Taichung 402, Taiwan
[b]Department of Statistics, Feng Chia University, Taiwan
[c]AIDS Prevention and Research Center, National Yang-Ming University, Taiwan
[d]Institute of Public Health, National Yang-Ming University, Taiwan
[e]Department of Health, Taipei City Government, Taiwan
[f]Taipei Municipal STD Control Center, Taiwan

Received 24 March 2005; received in revised form 28 August 2005
Available online 10 October 2005

Abstract

We make use of the voluntary HIV and syphilis test results conducted at five gay saunas in Taipei from August of 1999 to end of 2002 to estimate the number of HIV-positive gay saunas patrons in Taipei area by utilizing Hierarchical Bayes method in Generalized Removal Model for Open Populations (GERMO). Considering the effect of a nearby anonymous HIV quick test program on the gay sauna HIV serotesting data, we make use of the association between HIV and syphilis serotesting results from the gay sauna program to amend possible measurement error occurred at the time of data collection by utilizing the regression calibration method. The median estimates for the number of HIV-positive patrons of the five gay saunas increase from 120 (95% CI: 76.5–159.0) during the first half of 2000 to 224 (95% CI: 171.0–265.5) for the second half of 2002. The result, indicating two-fold increase within two and half years, confirms that the gay sauna patrons in Taipei area are at high risk for HIV infection.

Keywords: HIV/AIDS; Hierarchical Bayes methods; Measurement error; MSM; Syphilis; Taiwan

1. Introduction

Male–male sexual activity has existed across cultural barrier throughout the history of mankind. However, due to traditional cultural norms and social stigma in many countries, the men who have sex with men (MSM) population have been elusive and difficult to track. In many countries around the world, the incidence of Human Immunodeficiency virus (HIV) infections is most prevalent among MSM populations [1–6]. Consequently, regardless of the degree of openness of the society in question, the greatest difficulty in determinating the HIV prevalence in a certain MSM population is often the lack of knowledge concerning the

*Corresponding author. Tel./fax: +886 4 22853949.
E-mail address: hsieh@amath.nchu.edu.tw (Y.-H. Hsieh).

0378-4371/$ - see front matter
doi:10.1016/j.physa.2005.08.068

actual size of that population group [7–9]. Fining a solution to this endeavor is especially crucial given that the MSM population is often used as a sentinel population for HIV epidemiology [7,10].

Recently, an estimation method named "Generalized Removal Model for Open populations", or GERMO, was proposed to estimate the HIV-infected population size among specific population groups such as intravenous drug users (IVDU), sex workers, and sexually active individuals in a society [11–14]. Moreover, the results from the estimation procedure can be used to estimate the size of that hard-to-count, high-risk population group [15], which is crucial for the purpose of public health intervention measures. Recent developments in statistical computing have made Bayesian analysis accessible to researchers in epidemiology and other fields. The innovation, known as Markov chain Monte Carlo (MCMC), has facilitated the estimation of complex models that are difficult to estimate using alternative methods. In this work, we will implement Hierarchical Bayes (HB) method in GERMO to determine the HIV-infected population size among the gay sauna patrons in Taipei area. The advantage of an HB analysis is that we can specify priors via data information, which is the major improvement of present work over the procedure used in Ref. [12]. The predictive superiority of HB method lies in the freedom afforded by MCMC to specify more realistic models, and the ability to consider uncertainty in model parameters.

Since the first HIV/AIDS case in 1984, Taiwan has had a comparatively small HIV incidence. Through the end of 2002, there were a total of 4373 Taiwanese nationals known to be infected with HIV/AIDS, though studies have shown that there is some underreporting [16]. Among these 4373 cases, 4055 (92.7%) were males. Moreover, 2152 (49.2%) were reported to be homosexual/bisexual persons. Therefore, MSM account for almost half of the HIV-infected population in Taiwan. There are three main types of MSM populations in Taiwan distinguished by their meeting places: gay saunas, gay bars, and public parks, with very little interaction between the three groups [17]. By its nature, saunas provide a more convenient location for sex to take place. Consequently the gay sauna patrons conceivably are more sexually active and at higher-risk than the other groups. There are less than a dozen gay saunas in the Taipei area, with roughly the same number of gay saunas scattered in other major metropolitan areas in Taiwan. In August of 1999, a voluntary HIV and syphilis serotesting program was launched aiming at the customers of five gay saunas in Taipei, mainly in the Shi Meng Ding District. During the three years from 2000 to 2002, 81 HIV-positive cases were detected through this program. It is worthwhile to note that during this same time period, the total number of new HIV/AIDS cases is 1955, which means that roughly one out of 25 new HIV cases in Taiwan during these three years came from the customers of these five gay saunas in Taipei. Moreover, the prevalence rate of this gay sauna serotesting program over the same time period is 7.00% (81 out of 1191), much higher than that of any other group in Taiwan [18].

In a related study [19] of 589 gay sauna patrons in Taiwan which includes those taking part in the gay sauna serotesting program in Taipei, responses to questionnaire on sexual behavior from the 589 participants have also shown that the 21.7% of those participated in the program frequented the gay saunas at least once a week, 79.7% visited the saunas at once a month, and 91.7% of all participants have sex in the gay saunas with at least one or two persons during each visit. Moreover, only 43.5% of the participants always or often bring their own condoms when they visit the gay saunas and, despite of all these gay saunas having free condoms available on demand, almost half (46.6%) of all participants never asked for condoms even though 48.3% of them practice anal sex at least occasionally. Clearly, these are strong indications that safe sex is not being practiced in these establishments.

To gain insight into the true extension of HIV prevalence among the gay sauna patron population while the population size is unknown, we make use of the serotesting result to estimate the number of HIV-infected persons among the men having sex with men who frequent the five saunas in Taipei by utilizing the GERMO model. The idea is to consider the serotesting data as random samples from the gay sauna patrons in Taipei area. The result will give indications on the level of HIV underreporting in this concentrated, effusive population, as well as estimates for the size of gay sauna patron population.

2. Materials and methods

To provide anonymous HIV-1 antibody and syphilis tests, together with pre-test counseling and condoms, a team comprised of a researcher, a nurse, and peer educators paid regular visits to five Taipei gay saunas, four

Table 1
The HIV and syphilis serotesting result from the gay sauna testing program in Taipei area from August 1999 to the end of 2002

Time period	No. tested	No. of HIV+ (%)	No. of STS+ (%)
1999, Aug–Dec	58	3 (5.17)	9 (15.51)
2000, Jan–Jun	63	7 (11.11)	7 (11.11)
2000, Jul–Dec	273	24 (8.79)	36 (13.19)
2001, Jan–Jun	139	6 (4.32)	17 (12.23)
2001, Jul–Dec	178	12 (6.74)	14 (7.87)
2002, Jan–Jun	213	19 (8.92)	23 (10.80)
2002, Jul–Dec	265	13 (4.91)	28 (10.57)

of which were situated in Shi Meng Ding District, starting August of 1999. The visits were under prior consent from the gay sauna owners, and usually with posters at the entrances announcing the visits. Moreover, announcement was made over public announcement system of the establishments to encourage voluntary participation. In addition to taking blood samples, the participants were also encouraged, but not required, to fill in a questionnaire. HIV antibody test and serologic test for syphilis (STS) were carried out at the laboratory of Taipei Municipal STD Control Center (TSTDCC), also located in Shi Meng Ding. The subjects could contact TSTDCC by phone in two weeks to learn the test results, with those tested positive being provided the full opportunity for clinical healthcare at a clinic of his choice. The result of the serotests for HIV-1 and syphilis in six-month periods from August 1999 to the end of 2002 is given in Table 1.

However, starting in 2001, an anonymous quick HIV-testing program was launched at TSTDCC in close proximity to where four of the five gay saunas are located. Person taking the quick test can come to TSTDCC voluntarily for blood sample and call back a week later to learn the testing result. Comparing the serotesting results of 2000 and 2001 in Table 1 shows that, from the half year prior to the initiation of the quick testing program (July–December, 2000) to the first half year of the quick test (January–June, 2001), there is a sharp decrease in the number of tests taken, as well in the number of HIV-positive tests and the prevalence rate. Since several of the gay saunas taking part in the gay sauna testing programs are within short walking distance from TSTDCC, those patrons who are more inclined to participate in the program are also more likely to want to take the voluntary quick test at TSTDCC. It is therefore reasonable to conclude that the quick test program at TSTDCC has affected the data collected from the gay sauna testing program by detecting some HIV-positive persons who otherwise might have been tested and detected through the gay sauna testing program. Unfortunately, due to the anonymous nature of both programs, we are unable to identify the overlapping population. However, the quick test program has detected many HIV-positive persons who are MSM. From the follow-up (also voluntary) questionnaire for the HIV-positive persons, the quick test program has identified at least 21 HIV-positive MSMs during January–November of 2001 — more than the number of HIV-positive persons detected through the gay sauna testing program during the same period. Although there is no data on how many of these 21 persons were in fact gay sauna patrons, it appears safe to assume that this program has had some effect on the gay sauna testing. We propose the following procedure to rectify this problem.

2.1. Correction for measurement error

We will consider the effect of quick test on the gay sauna serotesting data as a measurement error which occurred at the time of data collection. Moreover, it can be considered as an extreme case of missing data. Data measured with error are very common in epidemiological studies. The regression calibration method is a useful tool in the missing data and measurement error problem [20]. In medical studies, this method is often used for correction of measurement error in covariates.

The association between HIV infection and sexual-transmitted diseases (STD), in particular syphilis and gonorrhea, is well-documented [21–25]. Therefore we will make use of the syphilis serotesting data in our effort to correct measurement errors in the HIV serotesting data caused by the initiation of the quick test program after 2001, by assuming that the HIV quick test program has no affect on the syphilis test results. We

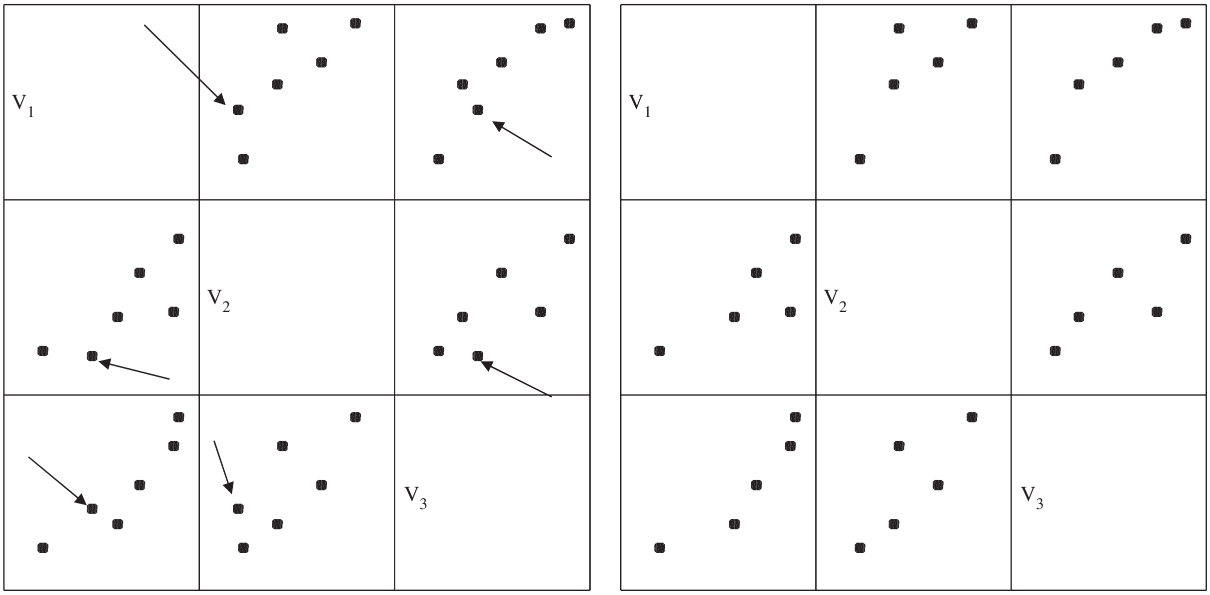

Fig. 1. (a) Pairs plot for the Taipei gay saunas serotesting data in Table 2. V_{ij}, $j = 1, 2, 3$, has been simplified to V_j for the sake of brevity. V_1, V_2, and V_3 denote the columns and rows for gay saunas serotests, HIV+, and STS+, respectively. The arrows point to the data points to be deleted. (b) Pairs plot for the Taipei gay saunas serotesting data in Table 2 with the data for time $i = 3$ deleted. V_{ij}, $j = 1, 2, 3$, has been simplified to V_j for the sake of brevity. V_1, V_2, and V_3 denote the columns and rows for gay saunas serotests, HIV+, and STS+, respectively.

Table 2
The estimates of Spearman's ρ correlation coefficients

		V_{i1}	V_{i2}	V_{i3}
$N = 6$ ($i = 1,\ldots,6$)	V_{i1}		0.886[a] (0.019)	0.943[a] (0.005)
	V_{i2}	0.886[a] (0.019)		0.771 (0.072)
	V_{i3}	0.943[a] (0.005)	0.771 (0.072)	
$N = 5$ ($i = 1,2,4,5,6$)	V_{i1}		0.900[a] (0.037)	1.000[a] (0.000)
	V_{i2}	0.900[a] (0.037)		0.900[a] (0.037)
	V_{i3}	1.000[a] (0.000)	0.900[a] (0.037)	

The values in parentheses are the two-sided p values for $H_0 : \rho = 0$.
[a]Denotes the p value being significant at 5%.

let V_{i1}, V_{i2}, and V_{i3} denote the respective numbers of serotests taken at gay saunas, those tested HIV-positive, and those tested STS-positive at time i. We use scatter plots and the Spearman's ρ correlation test to examine the linear relationship between those tested HIV-positive and those tested STS-positive. The plots in Fig. 1 show that there is a strong positively linear relationship between V_{i1} and V_{i2}, as well as between V_{i2} and V_{i3}. The estimates of Spearman's ρ correlation coefficients and p values of $H_0 : \rho = 0$ for any two variables are given in Table 2. The details of the regression calibration method and the scatter plots are given in Appendix 1.

The predicted value of V_{32} from the correction of measurement error is

$$\hat{V}_{i2} = 0.65943\,V_{i3}.$$

Consequently we use $\hat{V}_{32} = 11$ to correct for measurement error in replacement of $V_{32} = 6$ with $V_{33} = 17$. We use the notation u_j to replace V_{i2} after amending the measure error.

2.2. Generalized removal model for open populations (GERMO)

In our model, we make an assumption that the probability of HIV infections in jth sample is proportional to the jth sample size of serotests. We let N_j be the total number of subjects in HIV-infected population just before time t_j. As noted earlier, u_j is the number of HIV-infected individuals detected in the jth sample and p denotes the seroprevalence rate in the first sample. Therefore, $M_{j+1} = u_1 + \cdots u_j$ is the number of observed HIV-infected individuals in the first j samples.

We adopt a HB (Hierarchical Bayes) approach, which allows us to estimate $N_1, \ldots, N_s$, where s is the number of samples taken, and p simultaneously. Specifically, we can generate approximated samples from the posterior distribution of unknown parameters via MCMC methods. The estimation procedure of Bayesian analysis is outlined as follows:

Step 1: A Bayesian analysis requires a prior distribution for each parameter in the model. Suppose $\mathbf{\Theta} = (\mathbf{N}, p)$. We consider $\mathbf{N}$ and p a priori independent. The detailed description of the prior information is given in Appendix 2.

Step 2: Sample iteratively from $p(\mathbf{\Theta} \mid D)$, with the vector $D = (u_1, u_2, \ldots, u_s)$, to generate a posterior sample $\Theta^1, \ldots, \Theta^n$, where n is the total number of iterations. The sampling is done in two blocks, including $(N_1, N_2, \ldots, N_s)$ and p.

Step 3: Form $\hat{\mathbf{\Theta}}$, the point estimate of $\mathbf{\Theta}$, as the sample mean of the posterior sample:

$$\hat{\Theta} = \frac{1}{n-m} \sum_{k=m+1}^{n} \mathbf{\Theta}^k,$$

where m is the number of burn-in iterations to attain convergence. More details of the above procedures are presented in Appendix 2.

We assume that the natural mortality rate of the sexually active population during this time period is small compared to the AIDS-related death rate, since the majority (93 percent) of the subjects in question is of age 20–48 when the natural mortality is low. Moreover, small variation in natural mortality does not affect the result of our estimation [12].

3. Results

Using GERMO, we estimated number of HIV-infected persons among the gay sauna patrons in Taipei during the half-year periods from 2000 to 2002. The result of our estimation procedure is given in Table 3. The median estimates give the estimated number of HIV-positive persons among the patrons of these five gay saunas in Taipei during the half-year periods. As there are dozens of similar gay sauna scattered in every metropolitan area in Taiwan, it is clear that: (i) The gay sauna patrons are, without a doubt, a prominent high-risk group in Taiwan for the spread of HIV/AIDS; (ii) This serotesting program has been highly effective in detection of new cases, with 4.14% of all new HIV cases in Taiwan detected through this program during 2000–2002, and subsequently the prevention of additional cases. Furthermore, it provides useful data to carry

Table 3
Summary statistics of the total number of HIV-infected gay sauna patrons in Taipei for the six-month periods during 2000–2002

Time period	Median	Mean	SD	95% C.I.
2000, Jan–Jun	120	118.5	21.4	76.5–159.0
2000, Jul–Dec	150	149.5	18.0	111.5–180.5
2001, Jan–Jun	166	164.6	18.6	124.5–197.0
2001, Jul–Dec	186	184.4	19.4	142.5–217.0
2002, Jan–Jun	207	206.1	20.7	159.5–243.0
2002, Jul–Dec	224	222.7	22.9	171.0–265.5

For the median estimates, all fractions were rounded off to one.

Table 4
The estimated number of undetected HIV-positive gay sauna patrons in Taipei for the six-month periods during 2000-2002, with all fractions rounded off to one

Time period	Estimated no. of undetected HIV-positive persons	95% C.I.
2000, Jan–Jun	113	70–52
2000, Jul–Dec	126	88–157
2001, Jan–Jun	155	114–186
2001, Jul–Dec	174	131–205
2002, Jan–Jun	188	141–224
2002, Jul–Dec	211	158–253

Table 5
The estimated gay sauna patron population size in Taipei area for the six-month periods during 2000-2002

Time period	Mean estimate of no. of gay sauna patrons in Taipei area	95% C.I.
2000, Jan–Jun	1080	693–1431
2000, Jul–Dec	1706	1365–2059
2001, Jan–Jun	2096	1578–2487
2001, Jul–Dec	2760	2122–3220
2002, Jan–Jun	2321	1794–2724
2002, Jul–Dec	4562	3483–5418

out studies which enhance our understanding of this highly effusive and hard-to-count but at the same time important population group for surveillance and intervention purposes.

To gauge the information provided by this estimation procedure, we further compute the estimated number of undetected HIV-positive persons during each half-year, by subtracting the number of detected HIV cases in Table 1 (with corrected data $\hat{V}_{32} = 11$) from the median estimates of HIV-infected population in Table 3. The result is given in Table 4. The result shows a steady increase of the undetected HIV-positive persons among the gay sauna patrons, almost doubling in three years. There are alarmingly high percentages (almost half) of persons who do not use condoms and persons who engage in anal sex in the gay saunas [18]. Moreover, almost one out of three HIV-positive gay sauna patrons participated in a related study [18] describe them as bi-sexual while 35% of all participants in the same study are of age 29 or under, with 82.6% of these people still unmarried. Clearly, the gay sauna patron group is potentially a core group for the spread of HIV/AIDS in the community, as well as being a high-risk group. The present result highlights the need for public health officials to target the concentrated group of gay sauna customers for prevention measures, prominent among which are education programs on the danger of HIV transmission given their current sexual behavior and the importance of practicing safe sex.

4. Discussions

One can further make use of the estimated number of HIV-positive persons in a population group to make estimates of the actual population size by simply dividing the estimated number of HIV-infected person by the HIV seroprevalence rate [15]. To illustrate, we divide the median estimates for the HIV-infected population in Table 3 by the HIV prevalence rate for each half year in Table 1 (with corrected data $\hat{V}_{32} = 11$) to obtain the estimates of the gay sauna patron population in Taipei area in Table 5. The result indicates an increasing trend in the gay sauna patron population, although the last estimate might seem unrealistic in view of the previous estimates. Seasonality in data might also be a factor, as the estimates for the second half-years are always much higher than those of the first half-years.

Although one of the gay saunas is not situated near TSTSCC where the HIV quick test program took place, we are unable to make any inference or measurement correction by comparing the post-2001 data between this gay sauna and the other four — mainly due to the lack of knowledge of the mobility of gay sauna patrons in the Taipei area. Hence we know, to some certainty, that the willingness for testing of patrons of the four nearby gay saunas is affected by the quick test program. Whether the customers of the one gay sauna not situated in Shi Meng Ding were influenced by this program depends on how likely they are to come to Shi Meng Ding to visit the gay saunas, which cannot be ascertained without further studies aimed toward the social mixing of the MSM group in Taipei.

As with any estimation methods, there are limitations to our GERMO estimation procedure. Relevant discussions are given in detail in Ref. [12]. It suffices to say that, although widely useful in many applications of epidemiologic studies, one must exercise caution in determining how well the data corresponds to the assumptions made for the estimation procedure. For the present work, inconsistencies in the estimates as a result of the random sampling assumptions are very likely to occur, due to the use of data from the voluntary gay sauna serotesting program. Any improvement in this direction, however, is difficult since data of this nature (i.e. multiple sample data of effusive, hard-to-count populations) is difficult to obtain. Making use of the correlations between HIV+ and STS+ at the individual observation level, e.g., the propensity of HIV+ for a given subject who is STS+ which can be estimated via logistic regression, is a topic for future research.

Acknowledgements

The work is supported by Center for Disease Control of Taiwan grant DOH92-DC-1026 for which the authors are grateful. We thank the anonymous referees for valuable comments which significantly improved this article.

Appendix 1. Correction of measurement error

Let V_{i1}, V_{i2}, and V_{i3} denote the respective numbers of serotests taken at gay saunas, HIV+, and STS+ at time i. The pairs plot for the data is given in Fig. 1a with the pairs plot for the data for time $i = 3$ deleted in Fig. 1b.

The plots show that there is a strong positively linear relationship between V_{i1} and V_{i2}, as well as between V_{i2} and V_{i3}.

The estimates of Spearman's ρ correlation coefficients and p values of $H_0 : \rho = 0$ for any two variables are as given in Table 2. There is strongly significant linear relationship between any pair of two variables, (V_{i1}, V_{i2}) and (V_{i2}, V_{i3}). The existence of high correlations between the independent variables in a regression model is known as multicolinearity. There is highly positive correlation between V_{i1} and V_{i3} ($\hat{\rho} = 0.943$). The major concern is that the stability of the regression coefficients is affected by multicolinearity. To avoid serious multicolinearity, the variation of V_{i2} is explained by a simple regression line with an explanatory variable, V_{i3}. Consequently, we assume a simple linear regression model:

$$E(V_{i2}|V_{i1}, V_{i3}) = \beta_1 V_{i3}. \tag{1}$$

We have $V_{31} = 139$ and $V_{33} = 17$ which are the numbers of individuals tested and those with STS+, respectively at time $i = 3$. Instead of observing V_{32}, the data point V_{32} is treated as a missing value. We use the conditional expectation to estimate it. The information of $V_{32} \geqslant V_{32}^*$ is useful, where V_{32}^* is the observed value. We use max $\{\hat{\beta}_1 V_{i3}, V_{32}^*\}$ as a predicted value for V_{32} where $\hat{\beta}_1$ is the estimate which are obtained from (V_{i2}, V_{i3}), $i = 1,2,4,5,6$ using least square method.

Appendix 2. Implementation of HB method in GERMO

We consider a sequence of s samples taken from the seroprevalence data of gay saunas in Taipei. Let t_j, $j = 1,\ldots$ be the time when the jth sample is taken and let B_j be the number of HIV-infected individuals newly infected sexually between time t_j and time t_{j+1}. t_0 is some hypothetical initial time for the HIV-infected population and hence we assume that there are no infected individuals before t_0. We define N_j to be the total number of subjects in HIV-infected population just before time t_j, and $N_j = B_0 + \ldots + B_{j-1}$. Assume that the prevalence rate of HIV infections in jth sample is proportional to the jth sample size of serotests. Define

$e_j = e_j^*/e_1^*$, $j = 1, 2, \ldots, s$, where e_j^* denotes the number of individuals taken a serotest in the jth sample. Suppose that p denotes the seroprevalence rate in the first sample. Then pe_j is the seroprevalence rate in the jth sample.

The likelihood function can be obtained as follows:

$$L(\mathbf{B},\ p|\mathbf{D}) \propto \left\{ \prod_{j=1}^{s} \binom{N_j - M_j}{u_j} (pe_j)^{u_j} (1 - pe_j)^{N_j - M_{j+1}} \right\}, \tag{2}$$

where $\mathbf{D} = (u_1,\ldots,u_s)$, $\mathbf{B} = (B_0,\ldots,B_{s-1})$, and u_j the number of distinct HIV-infected individuals found in the jth sample. Therefore, $M_{j+1} = u_1 + \cdots u_j$ is the number of observed HIV-infected individuals in the first j samples. We assume that $u_0 = 0$, i.e., there is no observed infected individuals before the first sample. Suppose that the prior distribution of $(\mathbf{N}, p)$ where $\mathbf{N} = (N_1,\ldots,N_s)$ is given by $\pi(\mathbf{N}, p) = \pi(N_1,\ldots,N_s)\pi(p)$. This asserts that $\mathbf{N}$ and p are a priori independent. It is very difficult to draw $(N_1, N_2, \ldots, N_s)$ from the posterior distribution of $\pi(N_j|\mathbf{N}_{(-j)}, p, \mathbf{D})$ and simultaneously satisfied the restriction $N_1 \leqslant N_2 \leqslant \ldots \leqslant N_s$ when p is extremely small or large. Therefore, the prior of p is restricted to be within the interval $\left(\frac{\alpha}{4}, \frac{7\alpha}{4}\right)$, where $\alpha = E(pe_1)$. Subsequently, we assume that $\pi(p) = I\left(\frac{\alpha}{4} < p < \frac{7\alpha}{4}, pe_j \leqslant 1\right)$, $I(\cdot)$ is an indicator function that $I(A) = 1$ if the event A is true. The value of α is determined by solving the equation $\alpha = E(pe_1) \approx 1 - \left\{E(u_2)e_1^*\right\}/\left\{E(u_1)e_2^*\right\}$. It is reasonable to assume a prior of N to be $\pi\ (N_1,\ldots,N_s) = I\ (N_1 \leqslant \ldots \leqslant N_s)$. We call this HB method in GERMO.

Such priors lead to conditional posteriors of the forms:

$$\pi(p|\mathbf{N}, \mathbf{D}) \propto p^{u_1+\ldots+u_s} \prod_{j=1}^{s} (1 - pe_j)^{N_j - M_{j+1}} \times I\left(\frac{\alpha}{4} < p < \frac{7\alpha}{4}, pe_j \leqslant 1\right) \tag{3}$$

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

where $\mathbf{N}_{(-j)}$ denotes the vector $\mathbf{N}$ with the N_j deleted. $(N_j - M_{j+1})$ follows a truncated negative binomial distribution. Subsequently one can easily implement the Gibbs sampler to generate $(N_j - M_{j+1})$ from the truncated negative binomial in Eq. (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 jth sample to be ϕ. 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

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

We generate an approximated sample $(N_j^{(m+1)} - M_j^{(m+1)}), \ldots, (N_j^{(n)} - M_j^{(n)})$ from a truncated negative binomial posterior distribution by using MCMC methods, where m is the number of "burn-in" iterations for convergence and n is the total number of iterations. In our study, $m = 3000$ and $n = 10{,}000$. The convergence of the Gibbs samplers are monitored by examining several plots. We use $\hat{N}_{j+1} = \phi N_j + B_j$, $\hat{M}_{j+1} = \phi M_j + u_j$ from Eq. (5) in order to make the adjustment.

References

[1] D. Osmand, K. Page, J. Wiley, K. Garrett, H.W. Sheppard, A.R. Moss, L. Schrager, W. Winkelstein, Human immunodeficiency virus infection in homosexual/bisexual men, age 18–29: the San Francisco Young Men's Health Study, Am. J. Public Health 77 (1994) 685–689.

[2] P.S. Sullivan, S.Y. Chu, P.L. Fleming, J.W. Ward, Change in AIDS incidence for men who have sex with men, United States 1990–1995, AIDS 11 (1997) 1641–1646.
[3] J.A. Cantania, D. Osmond, R.D. Stall, L. Pollack, J.P. Paul, S. Blower, D. Binson, J.A. Canchola, T.C. Mills, L. Fisher, K.H. Choi, T. Porco, C. Turner, J. Blair, J. Henne, L.L. Bye, T.J. Coates, The continuing epidemic among men who have sex with men, Am. J. Public Health 91 (6) (2001) 907–915.
[4] UNAIDS, AIDS epidemic update, Geneva, UNAIDS, December 2001.
[5] C.F. Caceres, HIV among gay and other men who have sex with men in Latin America and the Caribbean: a hidden epidemic?, AIDS 16 (suppl 3) (2002) S23–S33.
[6] J.M. Calleja, N. Walker, P. Cuchi, S. Lazzari, P.D. Ghys, F. Zacarias, Status of the HIV/AIDS epidemic and methods to monitor it in the Latin America and Caribbean region, AIDS 16 (Suppl 3) (2002) S3–S12.
[7] W. McFarland, C.F. Caceres, HIV surveillance among men who have sex with men, AIDS 15 (suppl 3) (2001) S23–S32.
[8] J.S. Ramon, M. Alvarenga, N. Walker, J.M. Garcia-Calleja, F. Zacarias, Estimating HIV/AIDS prevalence in countries with low-level and concentrated epidemics: the example of Honduras, AIDS 16 (Suppl 3) (2002) S18–S22.
[9] Y.S.A. Amirkhanian, J.A. Kelly, A.A. Kukharsky, O.I. Borodkina, J.V. Granskaya, R.V. Dyatlov, T.L. McAuliffe, A.P. Kozlov, Predictors of HIV risk behavior among Russian men who have sex with men: an emerging epidemic, AIDS 15 (3) (2001) 407–412.
[10] E. McCray, I.M. Onorato, Sentinel surveillance of human immunodeficiency virus type 1 and sexual behaviors in bisexual African–American and Hispanic men visiting a sexually transmitted disease clinic in New York City, Am. J. Epidemiol. 19 (1992) 235–241.
[11] C.W.S. Chen, S.M. Lee, Y.H. Hsieh, K. Ungchusak, A unified approach to estimating population size for births only models, Comput. Statist. Data Anal. 32 (1999) 29–46.
[12] Y.H. Hsieh, C.W.S. Chen, S.M. Lee, Empirical Bayes approach to estimating the number of HIV-infected individuals in hidden and elusive populations, Statist. Med. 19 (2000) 3095–3108.
[13] Y.H. Hsieh, S.M. Lee, C.W.S. Chen, H. Arazoza, On the recent sharp increase of HIV infections in Cuba, AIDS 15 (3) (2001) 426–428.
[14] Y.H. Hsieh, H. de Arazoza, S.M. Lee, C.W.S. Chen, Estimating the number of Cubans infected sexually by HIV using contact tracing data, Int. J. Epidemiol. 31 (2002) 679–683.
[15] Y.H. Hsieh, The changing faces of commercial sex in Thailand: its implications for HIV transmission, JAIDS 30 (5) (2002) 537–540.
[16] Y.H. Hsieh, H.-C. Wang, H. de Arazoza, R. Lounes, S-.J. Twu, H.-S. Hsu, Ascertaining HIV underreporting in low HIV prevalence countries, preprint.
[17] N.Y. Ke, H.H. Chung, S.J. Chang, Y.C. Ko, K.R. Wen, Self-awareness of the threat of HIV and its related sexual behavior— analysis of 108 MSM in southern Taiwan, Res. Nursing 4 (3) (1996) 285–297 (in Chinese).
[18] Center for Disease Control of Taiwan, The monthly statistics of HIV/AIDS in Taiwan CDC (Taiwan) Web site, available at http://203.65.72.83/ch/das/ShowPublication.ASP?RecNo=928. (in Chinese).
[19] S.F. Lai, Molecular epidemiology of HIV-1 infection in men who have sex with men from gay saunas in Taiwan, National Yang-Ming University, Master Thesis, 2003.
[20] R.J. Carroll, D. Ruppert, L.A. Stefanski, Nonlinear Measurement Error Models, Chapman & Hall, London, 1995.
[21] T.C. Quinn, R.O. Cannon, D. Glasser, S.L. Groseclose, W.S. Brathwaite, A.S. Fauci, E.W. Hook, The association of syphilis with risk of human immunodeficiency virus infection in patients attending sexually transmitted disease clinics, Arch. Int. Med. 150 (6) (1990) 1297–1302.
[22] L.H. Pereira, J.A. Embil, D.A. Haase, K.M. Manley, Prevalence of uman immunodeficiency virus in patient population of a sexually transmitted disease clinic, Association with syphilis and gonorrhea, Sex Transmission Dis. 19 (2) (1992) 115–120.
[23] D.A. Ansell, T.C. Hu, M. Straus, M. Cohen, R. Sherer, HIV and syphilis seroprevalence among clients with sexually transmitted diseases attending a walk-in clinic at Cook County Hospital, Sex Transmission Dis. 21 (2) (1994) 93–96.
[24] L.V. Torian, H.A. Makki, I.B. Menzies, High HIV seroprevalence in men diagnosed with gonorrhea, New York City Department of Health sexually transmitted diseases clinic, 1990–1997, AIDS 14 (2000) 187–195.
[25] L.V. Torian, H.A. Makki, I.B. Menzies, C.S. Murrill, I.B. Weisfuse, HIV infection in men who have sex with men, New York City Department of Health sexually transmitted diseases clinic, 1990–1999, Sex Transmission Dis. 29 (2002) 73–78.