Ascertaining HIV Underreporting in Low Prevalence Countries using the *Approximate Ratio of Underreporting*

Ying-Hen Hsieh[1], Hui-Ching Wang[1], Hector de Arazoza[2], Rachid Lounes[3], Shiing-Jer Twu[4], and Hsu-Mei Hsu[5]

[1]Department of Applied Mathematics, National Chung Hsing University, Taichung, Taiwan. [2]Dept. Ecuaciones Diferenciales, Facultad Matematica y Computacion, University of Havana, Havana , Cuba. [3]U.F.R. de Mathématiques et Informatique, Université René Descartes (Paris V), Paris, France. [4]National Health Research Institute, Taipei, Taiwan. [5]Department of Health, Taipei, Taiwan.

running head: Ascertaining HIV underreporting

Correspondence to: Ying-Hen Hsieh, Department of Applied Mathematics, National Chung Hsing University, Taichung, Taiwan 402. fax: 886-4-22853949, tel: 886-4-22853949, e-mail: hsieh@amath.nchu.edu.tw

Ascertaining HIV Underreporting in Low Prevalence Countries using the *Approximate Ratio of Underreporting*

Abstract

Underreporting of HIV/AIDS cases is a common problem in HIV epidemiology which often skews epidemiologic projections on which public health policy decisions are often based, especially in the cases of low HIV prevalence countries or in early phases of an emerging epidemic when the HIV incidence is still low, but might be growing rapidly. In this work, we propose a simple mathematical model with groups of known and undetected HIV-positives. Using this model with the annual HIV incidence data of new HIV cases and new AIDS cases detected at onset of symptoms, we are able to obtain an estimate for the number of undetected HIV-positives. Moreover, using Taiwan data of 1993-2000, we are able to predict the number of new cases in the next two years within 5% accuracy. We also give a *approximate ratio of underreporting* which approximates the magnitude of underreporting of HIV cases in low HIV prevalence settings. The procedure is illustrated with the HIV data of Taiwan and Cuba. The result shows that underreporting in Cuba is low, probably due to its intense contact tracing program. For Taiwan, the level of underreporting is higher, but has improved slightly since 1999. The method is useful as a simple tool to gauge the immediate impact of an emerging epidemic, as well as for the purpose of public health policy planning and short-term future projections.

Keywords: HIV/AIDS, epidemiology, mathematical model, underreporting, Taiwan, Cuba.

1 Introduction

Due to rapid advances in recent years, treatment of human immunodeficiency virus (HIV) infection is associated with high treatment and care costs and subject to large differences in prevalence between different areas of the world. Equitable distribution of resources requires precise information on HIV prevalence in specific regions. However, underreporting of the number of HIV or Acquired Immunodeficiency Syndrome (AIDS) cases is a common problem in HIV epidemiology which often skews epidemiologic projections on which public health policy decisions are to be made. The problem is especially severe in the cases of third world countries [4], or hard-to-reach populations (intravenous drug users, sex workers, etc.) where the population sizes are unknown [7]. The cause for underreporting could vary widely in different settings. It could be social-political as one must deal with specific high-risk groups in Europe and North America, or economical as the inadequacy of health service systems in the developing countries conceals the true scope of the epidemic. But one thing is certain: the old problem still persists in the new millennium, as the recent developments in South Africa and China can attest to.

HIV and AIDS surveillance systems are most often used to estimate the size of underreporting in HIV and AIDS cases [5]. Others make use of more detailed HIV information such as infection rate and behavior surveys to estimate HIV prevalence [13]. Moreover, in low-prevalence countries or during early phases of an emerging HIV epidemic, small samples of HIV incidence often produce unreliable estimates of total prevalent infections. However, early assessment of the magnitude of the problem is essential in facilitating long-term intervention policy of governments [1]. [2] and de Arazoza et al. [11] used a linear differential equations model to compute the number of undetected HIV-positives in Cuba using the data for HIV detection from Cuba's contact tracing program. However, only the number of known individuals living with HIV is used in the parameter estimation.

In this work we propose a mathematical model which makes complete use of the information usually available in the standard HIV incidence data, namely, the number of new asymptomatic HIV cases as well as the number of new AIDS cases, either by way of progression to AIDS-defined illness (ADI) of previously known asymptomatic HIV-infected individuals or of persons who,as previous asymptomatic HIV-positives, had not been known to health authorities prior to the onset of ADI. We make this important distinction because the latter group in particular contains valuable information on the true magnitude of underreporting of HIV cases and hence provides a method for more accurate estimation of the parameters in the model. Even in countries without adequate medical systems, the records of AIDS cases detected at onset can usually be found from hospital records. Therefore the procedure we propose is especially suitable for use by health policy makers in resource-poor countries wishing to ascertain the magnitude of an emerging epidemic, and to determine whether there is cause for concern.

The basic model is given in Section 2. In Section 3 the model for low HIV prevalence settings is described along with the procedure for parameter estimation. We will demonstrate how one can compute the number of unknown HIV-positives using HIV data by taking the Taiwan HIV data as an illustrative example. Some parameters are obtained through statistical methods, others are acquired by curve fitting with the HIV census data. In Section 4 we derive the *approximate ratio of underreporting*, or more precisely, the asymptotic ratio between the number of undetected HIV infections and the number of known HIV infections in low prevalence settings which sheds light on the magnitude of underreporting in the given population. The usefulness of the approximate ratio of underreporting will be illustrated with Cuban and Taiwanese HIV data. Concluding remarks will be given in Section 5.

2 The Mathematical Model

We consider the following variables, as function of time t in years:

$N(t)$ - number of susceptibles,

$X(t)$ - number of living undetected HIV-positives,

$Y(t)$ - number of living known HIV-positives,

$W(t)$ - number of AIDS patients who had been detected before onset of AIDS,

$Z(t)$ - number of AIDS patients detected at onset of AIDS symptoms.

The important parameters (or yearly rates) are given as follows:
μ - mortality rate of susceptibles,
$\bar{\mu}$ - removal (mortality and onset to AIDS) rate of undetected HIV-positives,
$\bar{\mu}^*$ - removal (mortality and onset to AIDS) rate of known HIV-positives,
$\bar{\mu_0}$ - mortality rate of AIDS patients,
λ - infection rate of HIV-positives,
k - detection rate of living undetected HIV-positives,
δ - recruitment rate of susceptibles.

A flow chart for the dynamics of the model is given in Fig. 1. Note that the removal rates in the definitions of $\bar{\mu}$ and $\bar{\mu}^*$ include mortality as well as progression to ADI of HIV-positive individuals, hence $\bar{\mu} \geq \mu$ and $\bar{\mu}^* \geq \mu$. Similarly, we have $\bar{\mu_0} \geq \mu$.

The model equations of the model are as follows:

$$\begin{aligned}
N^{'}(t) &= \delta N - \mu N - \lambda N(\frac{X}{N+X+Y}), \\
X^{'}(t) &= \lambda N(\frac{X}{N+X+Y}) - \bar{\mu}X - kX, \\
Y^{'}(t) &= kX - \bar{\mu}^* Y, \\
Z^{'}(t) &= (\bar{\mu} - \mu)X - \bar{\mu_0}Z, \\
W^{'}(t) &= (\bar{\mu}^* - \mu)Y - \bar{\mu_0}W.
\end{aligned} \tag{1}$$

This nonlinear model is different from the models proposed by de Arazoza et al. [2,3] in the use of a different removal rate $\bar{\mu}^*$ for known HIV-positives, which accounts for the time from infection to detection as well as the possible amelioration and prolonged progression to AIDS for detected HIV patients undergoing treatment. More recently, a nonlinear model for contact tracing with this consideration has been studied by [10]. Moreover, the assumption on proportionate recruitment of susceptibles in the present model allows us to consider population groups that are varying in size. We also assume the infections caused by detected

HIV-positives are negligible compared to those caused by the undetected HIV-positives (also see [2]). The model is most appropriate for modeling homosexual transmission of HIV. According CDC-Taiwan, the risk factor for 48.22% of all 5221 cases at the end of 2003 is homosexual/bisexual contact, while heterosexual contact accounts for another 39.84%. Hence the proposed HIV model is appropriately chosen. Finally we do not consider migration since in many low-prevalence countries such as Taiwan and Cuba, lack of massive migration is one of the reasons for the absence of sizable infections in the population.

3 Model for Low HIV Prevalence Settings

Now we will consider the model in low HIV prevalence settings. When the population in question is at low HIV prevalence,

$$\frac{X+Y}{N+X+Y} << 1, \frac{N}{N+X+Y} \approx 1.$$

We then have the following system of linear differential equations, with $'$ denoting derivative:

$$\begin{aligned} N^{'}(t) &= (\delta-\mu)N - \lambda X, \\ X^{'}(t) &= \lambda X - (\bar{\mu}+k)X, \\ Y^{'}(t) &= kX - \bar{\mu}^{*}Y, \\ Z^{'}(t) &= (\bar{\mu}-\mu)X - \bar{\mu_0}Z, \\ W^{'}(t) &= (\bar{\mu}^{*}-\mu)Y - \bar{\mu_0}W. \end{aligned} \tag{2}$$

with $\delta > \mu$, $\bar{\mu} \geq \mu$, $\bar{\mu}^{*} \geq \mu$, $\bar{\mu_0} \geq \mu$, and $\lambda > \bar{\mu} + k$. The first inequality intuitively assumes recruitment of susceptible population is greater than the mortality. In the last inequality, we assume that the unknown HIV-infected population is increasing at the low-prevalence stage of an epidemic. The assumption is reasonable either for the early (low-prevalence) stages of an emerging epidemic, or for a slowly growing epidemic like HIV in Taiwan and Cuba.

There are two more variables which do not appear in the dynamics of our model, but will prove useful later. One is $P(t)$ - the accumulated number of detected asymptomatic

HIV-positives at time t, the other is $M(t)$ - the accumulated number of AIDS cases detected at onset of ADI. The equations for these variables are:

$$P^{'}(t) = kX$$

and

$$M^{'}(t) = (\bar{\mu} - \mu)X.$$

Note that $P(t) + M(t)$ equals the accumulated number of detected HIV cases at time t. The analytic solutions of linear system (2) can be easily obtained and are given below:

$$\begin{aligned}
X(t) &= X_0 e^{(\lambda - \bar{\mu} - k)t}, \\
Y(t) &= e^{-\bar{\mu}^* t}[Y_0 + \frac{kX_0}{\lambda - k + \bar{\mu}^* - \bar{\mu}}(e^{(\lambda - k + \bar{\mu}^* - \bar{\mu})t} - 1)], \\
Z(t) &= e^{-\bar{\mu_0} t}[Z_0 + \frac{(\bar{\mu} - \mu)X_0}{\lambda - k + \bar{\mu_0} - \bar{\mu}}(e^{(\lambda - k + \bar{\mu_0} - \bar{\mu})t} - 1)], \\
W(t) &= e^{-\bar{\mu_0} t}\{W_0 + (\bar{\mu}^* - \mu)[(Y_0 + \frac{kX_0}{\lambda - k + \bar{\mu}^* - \bar{\mu}})\frac{e^{(\bar{\mu_0} - \bar{\mu}^*)t} - 1}{\bar{\mu_0} - \bar{\mu}^*} \\
&\quad + (Y_0 + \frac{kX_0}{\lambda - k + \bar{\mu}^* - \bar{\mu}})\frac{e^{(\lambda - k + \bar{\mu_0} - \bar{\mu})t} - 1}{\lambda - k + \bar{\mu_0} - \bar{\mu}}]\}, \\
P(t) &= P_0 + \frac{kX_0}{\lambda - k - \bar{\mu}}[e^{(\lambda - k - \bar{\mu})t} - 1], \\
M(t) &= M_0 + \frac{(\bar{\mu} - \mu)X_0}{\lambda - k - \bar{\mu}}.[e^{(\lambda - k - \bar{\mu})t} - 1]
\end{aligned}$$

where $X_0 = X(0)$, $Y_0 = Y(0)$, $Z_0 = Z(0)$, $W_0 = W(0)$, $P_0 = P(0)$, and $M_0 = M(0)$ are the initial numbers of the respective groups.

3.1 Parameter Estimation

Computing the values for $X(t)$ in a given population at a given time t is the focus of our investigation. Data for all other variables can be obtained from official government HIV data. The data of 2969 HIV-positive Taiwanese citizens detected from 1984 to the end of 2000 was obtained from the HIV databank of Center for Disease Control (CDC) of Taiwan (Table 1). For example, P_i is the accumulated number of detected asymptomatic HIV-positives at the end of year i. We will set $t = 0$ to be 1993 in our model, since the data for Z_i and W_i before 1993 are small and exhibit large fluctuations which might induce unnecessary errors in our parameter estimation. Therefore $P_0 = 457$, the real data value for accumulated number of detected asymptomatic HIV-positives at the end of year 1993, is the true value for $P(0)$. The other 4 sets of data, Y_i, M_i, Z_i, and W_i follow similarly.

The data for P_i is accurate by the nature of its definition, that is, the numbers of detection are exact. The data for Y_i would be accurate, as long as the removal out of this class is limited to natural mortality and progression to AIDS-defined illness (ADI). However, the data shows that the mortality of persons living with HIV in Taiwan has a conspicuously larger mortality rate (by car incidents, suicide, etc.) than others in their age group. Nevertheless, it is impossible to determine quantitatively if and how much their HIV-seropositive status contributes to this discrepancy. Therefore we conclude that some inaccuracy exists in the data Y_i. W_i is related to Y_i hence also contains some inaccuracy. Z_i and M_i are, intuitively, the data with most error since these groups of HIV-positive persons are detected initially only with diagnosis of ADI. Underreporting of AIDS is a common occurrence in most countries due to personal reasons so we conclude the data for Z_i and M_i are most inaccurate compared to others in the table. Our goal is to make use of as accurate data as possible in our estimation procedure.

Some parameters needed in the model can be obtained from other sources. Over 80% of the 2969 HIV-positive individuals in the dataset are of age 20-49. The 5-year survival rates of Taiwanese in these age groups ranges from a high of 0.996 for persons age 20-24 to a low of 0.982 for person age 45-49, so we let the natural mortality be $\mu = 0.002$ which yields a 5-year survival rate of 0.990 for all persons between 20-49. Using the well-known Kaplan-Meier method, we obtained the mean estimate for time from HIV diagnosis to ADI (from the cohort data of 2880 HIV-positive persons who had ADI before death) to be 7.89 years. So we have $\bar{\mu}^* = \mu + 1/8.77 = 0.116$. The time from ADI to death $\bar{\mu}_0$ is likewise estimated from the cohort data of 868 AIDS patients who have died and is 2.71 years, hence $\bar{\mu}_0 = 1/2.71 = 0.369$. For the HIV incubation time, we use the cohort data of 47 persons in

Taiwan who were infected by blood transfusion between 1984 to 2000 and had ADI before death. The reason is that this group of people are mostly diagnosed with HIV within a few months of infection when the information on contaminated blood became known, and hence provide a good estimate for the incubation time. The result, using the Kaplan-Meier method once again, is 11.28 years, which yields $\bar{\mu} = \mu + 1/11.28 = 0.091$.

In order to make as accurate an estimate as possible, we make use of P_i and M_i for our estimation of the theoretical parameter k which is needed in our theoretical estimation of $X(t)$. From the differential equations for $P(t)$ and $M(t)$, we have

$$\frac{P'(t)}{M'(t)} = \frac{k}{\bar{\mu} - \mu}$$

or $k = (\bar{\mu} - \mu)P'(t)/M'(t)$.

From the data for P_i and M_i for $t = 0$ to 7 (i.e. the years 1993-2000), we compute ΔP_i and ΔM_i by central difference at time i, i.e., $\Delta P_i = (P_{i+1} - P_{i-1})/2$ and $\Delta M_i = (M_{i+1} - M_{i-1})/2$. Furthermore, we let $\Delta P_i/\Delta M_i$ be the approximation to $P'(t)/M'(t)$ for $t = 1$ to 6. Using the known values of μ and $\bar{\mu}$ we obtain the average value of k_i ($k_i = (\bar{\mu} - \mu)\Delta P_i/\Delta M_i$) over the time $t = 1$ to 6, $k_{avg} = 0.283$, as our estimate for the detection rate k. Next we make use of the formula for $P(t)$ given earlier along with the real data P_i for $i = 0$ to 7 (year 1993 to 2000) to perform the least-square minimization for the estimated values of kX_0 and $\lambda - k - \bar{\mu}$. From this procedure we get $kX_0 = 126.1$ and $\lambda - k - \bar{\mu} = 0.199$. Subsequently we obtain an estimate for the initial number of unknown HIV-positives $X_0 = 446$ in year 1993, and the estimated infection rate is $\lambda = 0.571$.

3.2 The Theoretical Number of Unreported HIV-Positives

With our estimated parameters, we proceed to compute the theoretical values of $Y(t)$, $P(t)$, and $M(t)$ for $t = 0$ to 7 so we can compare with the real data values. The results are given in Figs. 2A-C. The theoretical numbers for $P(t)$ and $M(t)$ agree very well with the respective real data P_i and M_i, since the estimation of the parameters are geared toward optimization using P_i and M_i. The estimate for $Y(t)$ agrees less well with Y_i, probably due to underestimation caused by detected deaths of persons living with HIV attributable to suicide and traffic accidents which may or may not be related to their HIV seropositive status.

The main objective of this work is to estimate the number of undetected HIV-positives $X(t)$. For this purpose we can compute the theoretical number of $X(t)$ from the equation for $X(t)$ using the estimated values of X_0, λ, $\bar{\mu}$ and k. The result, along with the theoretical and actual numbers of accumulated known HIV cases $P(t) + M(t)$ for $t = 0$ to 9 (or 1993 to 2002), is given in Table 2. It asserts that the number of undetected persons living with HIV in Taiwan will reach 2674 by the end of 2002, and the accumulated number of HIV cases in Taiwan will reach 4661 by the end of 2002. To give some idea of the usefulness of our procedure for predicting HIV incidence, our result predicts that there will be 735 new HIV/AIDS cases with 173 new AIDS cases in 2002. The government data shows that in reality there were 715 new HIV/AIDS cases and 165 new AIDS cases in Taiwan during 2001, giving errors of 2.8% and 4.8%, respectively. Similarly, the predicted number of new HIV/AIDS cases in 2002 is 897, yielding an error of 2.4% from the true data of 878 cases. In other words, the method enables us to predict the number of new HIV cases for the next two years two years within 5% accuracy.

4 The Ratio of Underreporting

In this section, we will compute the *approximate ratio of underreporting*, or the approximate ratio of the number of living undetected HIV-positives to the number of known HIV-positives. The ratio of underreporting allows us to gain insights into the magnitude of the spread of disease in low-prevalence settings, either in the early stages of an emerging epidemic or in low prevalence communities.

The ratio of the number of living undetected HIV-positives to the number of known living HIV-positives at time t is as follows:

$$\frac{X(t)}{Y(t)} = [(\frac{1}{r_0} - \frac{1}{r^*})e^{-(\lambda - k + \bar{\mu}^* - \bar{\mu})t} + \frac{1}{r^*}]^{-1},$$

where $r_0 = X_0/Y_0$ is the initial ratio of the number of living undetected HIV-positives to the number of known HIV-positives at time $t = 0$ and the limiting ratio

$$r^* = (\lambda - k + \bar{\mu}^* - \bar{\mu})/k$$

is defined to be the *approximate ratio of underreporting.*

If $\lambda - k + \bar{\mu}^* - \bar{\mu}$ is nonpositive, all variables in the model except $N(t)$ will go to zero and there is no epidemic. On the other hand, if $\lambda - k + \bar{\mu}^* - \bar{\mu} > 0$, we have

$$r^* = \lim_{t\to\infty} \frac{X(t)}{Y(t)}.$$

Clearly, if the relevant parameters remain mostly unchanged over the time of the data set, then the ratio of the number of living undetected HIV-positives to the number of known living HIV-positives approaches r^* in time, hence the term "approximate ratio of underreporting".

Note that when $(\lambda - k + \bar{\mu}^* - \bar{\mu})t$ is sufficiently large for a specific time t, $\dfrac{X(t)}{Y(t)} \approx r^*$. That is, r^* gives an approximation to the magnitude of underreporting within that population at that time t. The expression for r^* is very intuitive. First, if r^* is negative than the removal from the unknown HIV-infective class (either by detection k or death and progression to AIDS $\bar{\mu}$) is so fast that there is no underreporting. Moreover, $\bar{\mu}^* - \bar{\mu}$ gives the difference in removal rates of the known HIV-positives and undetected HIV-positives due to progression of the disease while $\lambda - k$ measures the rate of change of the undetected HIV-positives disregarding such removal. If either (i) $\bar{\mu}^*$ is decreased (i.e. longer time from diagnosis to AIDS) due to early detection or treatment of known HIV patients, or (ii) the rate of infection λ is decreased, or (iii) detection k is increased, the resulting approximate ratio of underreporting will be smaller (i.e. less underreporting of HIV cases). Note that an increase in k produces the most dramatic effect, as it also appears in the denominator of r^*. Hence intensive detection strategy is the most effective way to combat underreporting when the HIV prevalence is low or at the early stages of the epidemic.

To illustrate the usefulness of the result we will use the examples of Cuba and Taiwan, both countries having very low HIV prevalence (0.35% for Cuba and 0.17% for Taiwan) even though the first HIV case was diagnosed in 1986 for Cuba and in 1984 for Taiwan.

In [2], the following relevant parameters for Cuba are given: $\lambda = 0.5594, k = 0.4554, \bar{\mu} = 0.1087 + 0.0053 = 0.1140$. From the data of 3627 HIV-positive persons in Cuba who were diagnosed with HIV before or at the onset of AIDS symptoms between 1986 and 2001, we used the Kaplan-Meier method to compute the mean time from diagnosis of HIV to AIDS to be 7.22 years. Subsequently the mean yearly progression rate from detection of HIV to AIDS is 0.1385 and $\bar{\mu}^* = 0.1385 + 0.0053 = 0.1438$. Therefore we have $r^* = 0.2938$ as the approximate ratio of underreporting in Cuba. The theoretical numbers of undetected HIV-positives $X(t)$ in Cuba were computed for 1987 to 1999 using the model in [2]. In Fig. 3A we give the plot of the approximate ratio of underreporting along with $X(t)/Y_i$ computed from 1991 to 1999, i.e. $t = 4$ to 12. The result shows good agreement with r^* for $t > 10$. Note that due to its extensive contact tracing program, the HIV detection rate in Cuba is very high ($k = 0.4554$) and subsequently the HIV prevalence has remained low [8,9]. Hence a large t is needed for $(\lambda - k + \bar{\mu}^* - \bar{\mu})t$ to be sufficiently large. Consider a hypothetical society at early stages of a growing HIV epidemic. Suppose its HIV detection is only half as effective as that of Cuba. With everything else $(\lambda, \bar{\mu}^*, \bar{\mu})$ being the same, it would require only 4 or 5 years of HIV data to obtain a reasonable approximate ratio of underreporting.

To further illustrate, we return to the example of Taiwan. From the estimates of the parameters in Section 3, we have $r^* = 1.018$. We also compute $X(t)/Y_i$ by using the theoretical numbers $X(t)$ and the real data Y_i. The results for t=0 to 9, or 1993 to 2000 are plotted in Fig. 3B, and agree well with the approximate ratio of underreporting. One intuitively plausible reason is that the low detection rate in Taiwan yields quick convergence to the approximate ratio of underreporting.

5 Concluding Remarks

1. In this work we estimated the number of unknown HIV-infected individuals using the information on the detected HIV cases before onset of ADI and those detected at the onset of ADI. From the results one can predict the number of HIV cases in

the future. Using HIV data from 1984-2000, the errors resulting from our prediction of the accumulated HIV cases numbers for 2001 and 2002 are only 4.7% and 6.6%, respectively. This margin of error is sufficiently small for the purpose of future policy and budget-planning.

2. The underlying assumption of the model is that only the asymptomatic HIV-positives are unknown to the health authority. The accuracy of the method then is contingent on the accurate reporting of all ADI cases, which may vary from country to country.

3. We assume no heterogeneity for the different modes of HIV transmission rates within the infected population in Taiwan and Cuba, since sexual contact is the main mode of HIV transmission in both countries. Moreover, sexual contact is also the major contributing mode of infection in Cuba. While this might not be true for some low HIV-prevalence countries, e.g. countries with large groups of infection by IV drug use as well as sexual contacts, for which more complicated model might be called for, the present model which assumes homogeneity of disease transmission is appropriate for estimates of countries like Taiwan and Cuba as demonstrated here.

4. The accuracy of the estimate also depends on the accuracy of the estimated incubation time and time from ADI to death. In this work we use the Taiwan data from 1984 to 2000 with the Kaplan-Meier method to make the estimates. With the inclusion of HAART in the national health service in Taiwan starting in 1997, there is likely a significant change in both parameter values [12]. A model which considers the effect of HAART on the accuracy of the estimation is a worthwhile subject for future study.

5. The convergence of the ratio to the approximate ratio of underreporting takes eight or more years in the case of Cuba, mainly due to the low underreporting of HIV cases. This is due, at least in part, to the intense Partner Notification Program which vigorously tracks and screens the contacts of detected HIV-positive individuals.

6. In the case of Taiwan, the higher underreporting (i.e. lower detection rate) leads to

a quick convergence to the approximate ratio of underreporting. Note also that the low underreporting ratio in 1999-2000 probably reflects to some degree the effect of the recently implemented HIV serotesting program for gay saunas patrons in major cities in Taiwan. This program, which took place in the metropolitan areas of Taipei, Hsinchu, Taichung, and Kaohsiung, was initiated in summer of 1999 [6] and had been successful in detecting HIV cases among men who have sex with other men (MSM). For the year 2000 and in five gay saunas in Taipei alone, this program yielded 33 HIV-positives with a high seroprevalence rate of 9.5% (33/348). This program has obviously been significant for the HIV intervention measures in Taiwan since there were only 531 new HIV cases in all of Taiwan in 2000. As homosexuals and bisexuals constitute almost half of all HIV/AIDS cases in Taiwan (> 48% according to the most recent Taiwan CDC HIV data), this serotesting program which targets the most important high-risk group in Taiwan might be linked to the observed decrease in HIV underreporting in Taiwan. It also demonstrated the transient nature of any estimate of HIV-infected population size, that a change in intervention measures could easily alter the rate of change of the HIV-infected population groups.

7. We have used a mathematical model and standard government HIV data to develop a simple procedure by which one can obtain valuable information on the magnitude of HIV underreporting in a low prevalence setting. The procedure could be useful in many low HIV prevalence scenarios in the world, as well as for other epidemics of interest.

Acknowledgment. YHH is supported by grant DOH91-DC-1059 from CDC-Taiwan. The authors wish to thank the anonymous reviewer for constructive comments.

References

1. Anderson R. M., May R. M., and McLean A.R. Possible demographic consequences of AIDS in developing countries. *Nature* **332**(1988) pp.228-234.

2. de Arazoza H., Lounes R., Hoang T. and Interlan Y. Modeling HIV epidemic under contact tracing - The Cuban case. *J. of Theoretical Medicine.* **2**(2000) pp.267-274.

3. de Arazoza H. and Lounes R. A non-linear model for a sexually transmitted disease with contact tracing. *IMA J Math Appl Med Biol.* **19**(3)(2002.) pp.221-234.

4. Balint G. A. Situation analysis of HIV/AIDS epidemic in sub-Saharan Africa, *East. Afr. Med. J.* **75**(12)(1998) pp.684-686.

5. Barchielli A., Buiatti E., Balanti C., Giovannett, L. Acciai, S. and Lazzeri V. Completeness of AIDS reporting and quality of AIDS death certification in Tuscany (Italy): a linkage study between. *Eur. J. Epidemiol.* **11**(5)(1995) pp.513-517.

6. Chen Y.M.A., Chung C.L., Wu S.I., Hsu H. and Jen I. A Survey on the HIV-1, Syphilis and Sexual Behavior Among Patrons of Gay Saunas in Taiwan. 2001 Hong Kong AIDS Conference Proceedings. (2001) pp. 128-150.

7. Hsieh Y.-H., Chen C.W.S. and Lee S.M. Empirical Bayes approach to estimate the number of HIV-infected individuals in hidden and elusive populations. *Stat. Med.* **19**(2000) pp.3095-3108.

8. Hsieh Y.-H., Chen C.W.S., Lee S.M. and de Arazoza H. On the recent sharp increase in HIV detections in Cuba. *AIDS*, **15**(3)(2001) pp.426-428.

9. Hsieh Y.-H., de Arazoza H., Lee S.M. and Chen C.W.S. Estimating the number of Cubans infected sexually by HIV using contact tracing data. *Inter. J. Epidemiol.* **31**(2002) pp.679-683.

10. Hsieh Y.-H., de Arazoza Rodriguez H., Lounes R., and Joanes J. A Class of Models for HIV Contact Tracing in Cuba: Implications for Intervention and Treatment. To appear in *Deterministic and Stochastic Models for AIDS Epidemics and HIV Infection with Interventions*, ed.by Tan W.Y. (World Scientific, Singapore, 2005)

11. Lounes R. and de Arazoza, H. A two-type model for the Cuban national programme on HIV/AIDS, *IMA J. Math. Appl. Med. Biol.* **16**(1999) pp.143-154.

12. Tassie J.M., Graba, S., Lancar R., Deloumeaux J., Bentata M. and Costagliola D. Time to AIDS from 1992 to 1999 in HIV-1-infected subjects with known date of infection . *JAIDS* **30**(2002) pp.81-87.

13. Wijngaarden J.K. and Jager H.J. Methods for estimating HIV prevalence: A comparison of extrapolation from surveys on infection rate and risk behavior with back-calculation for the Netherlands. *Eur. J. Epidemiol.* **14**(7)(1998) pp.645-652.

Table 1. Yearly numbers of the accumulated detected asymptomatic HIV-infectives (Pi), known persons living with HIV (Yi), the accumulated AIDS cases detected at onset of ADI (Mi), and AIDS patients detected at onset of AIDS symptoms (Zi) , AIDS patients known before onset of AIDS (Wi).

Year(i)	P(i)	Y(i)	Mi	Zi	Wi
1984	9	9	0	0	0
1985	24	24	0	0	0
1986	34	34	1	0	0
1987	46	45	1	0	1
1988	74	68	2	0	2
1989	111	102	8	3	2
1990	144	128	11	2	3
1991	224	196	23	6	6
1992	345	306	38	8	7
1993	457	402	64	17	10
1994	600	499	99	32	23
1995	776	626	154	49	39
1996	980	736	228	77	75
1997	1253	949	306	105	101
1998	1555	1183	410	155	114
1999	1917	1483	521	227	144
2000	2343	1840	626	281	173

Table 2. The theoretical numbers of unknown HIV-positives X(t) and accumulated HIV/AIDS cases P(t)+M(t) compared to real data Pi+Mi in Taiwan for t=0 to 9 (1993-2002).

Year	X(t)	P(t)+M(t)	Pi+Mi
1993	446	521	521
1994	544	704	699
1995	664	926	930
1996	810	1198	1208
1997	989	1529	1559
1998	1206	1933	1965
1999	1472	2428	2438
2000	1796	3029	2969
2001	2191	3764	3595
2002	2674	4661	4373

Figure Captions

Fig. 1. The model flow diagram.

Fig. 2A. The theoretical number $Y(t)$ of known persons living with HIV in Taiwan from 1993 to 2000 computed from model plotted against the true number Y_i.

Fig. 2B. The theoretical number $P(t)$ of detected asymptomatic HIV-positives in Taiwan from 1993 to 2000 computed from model plotted against the true numberP_i.

Fig. 2C. The theoretical number $M(t)$ of detected asymptomatic HIV-positives in Taiwan from 1993 to 2000 computed from model plotted against true number M_i.

Fig. 3A. The approximate ratio of underreporting in Cuba computed from [5] plotted against the numbers $X(t)/Y_i$ from 1991 to 1999.

Fig. 3B. The approximate ratio of underreporting in Taiwan plotted against the numbers $X(t)/Y_i$ from 1993 to 2000.