Statistics and Its Interface Volume 0 (2014) 1–12

Estimating the sizes of populations at risk of HIV infection from multiple data sources using a Bayesian hierarchical model

Le Bao, Adrian E. Raftery*, and Amala Reddy

In most countries in the world outside of sub-Saharan Africa, HIV is largely concentrated in sub-populations whose behavior puts them at higher risk of contracting and transmitting HIV, such as people who inject drugs, sex workers and men who have sex with men. Estimating the size of these sub-populations is important for assessing overall HIV prevalence and designing effective interventions. We present a Bayesian hierarchical model for estimating the sizes of local and national HIV key affected populations. The model incorporates multiple commonly used data sources including mapping data, surveys, interventions, capture-recapture data, estimates or guesstimates from organizations, and expert opinion. The proposed model is used to estimate the numbers of people who inject drugs in Bangladesh.

Keywords and phrases: Capture-recapture, Expert opinion, Heterogeneity, HIV/AIDS epidemic, Injecting drug user, Key affected population, Mapping data, Markov chain Monte Carlo, Key affected population, Multiplier method.

1. INTRODUCTION

Since the 1950s, mortality had been declining and life expectancy had been increasing in both developed and developing countries until the global AIDS epidemic was reported. The AIDS epidemic caused a slowing down and in some cases even a reversal of these trends in the most severely affected countries due to increasing mortality. As a sexually transmitted disease, AIDS especially affects adolescents and young and middle-aged adults and has a damaging impact on labor supply, labor productivity, and families with AIDS patients. Reliable estimation and prediction of the HIV/AIDS epidemic can help policy makers and program planners efficiently allocate resources, as well as plan and manage interventions and treatment and care programs. Therefore, accurate estimation and projection of the epidemic is essential for HIV/AIDS-related programs.

In countries with low-level and concentrated epidemics, HIV has spread rapidly in the sub-populations that are most likely to acquire and transmit HIV, but is not well established in the general population. Unlike in countries where the epidemic has become generalized and data from pregnant women are used as a proxy for adult prevalence [17], there is no set of representative data that can be used to estimate adult prevalence in most of the countries with low-level and concentrated epidemics. Countries estimate the number of people living with HIV using models such as the Estimation and Projection Package (EPP) and the Asian Epidemic Model [45, 4, 18]. All these models require as inputs estimates of the sizes of key affected populations (KAPs) such as people who inject drugs (PWID), female sex workers and their clients, and men who have sex with men. However, few countries have estimates of the sizes of KAPs that are nationally accepted as reliable estimates, and existing estimates are often subject to large uncertainties.

Current national population size estimation approaches typically generate high and low estimates of both population size and HIV prevalence for KAPs, in addition to the best estimate, using various levels of inputs based on expert judgment. There is likely to be wide variation in how people decide on plausible bounds given the information they have [19]. These plausibility bounds are based on expert knowledge and so are to some extent subjective. As a result, they should not be interpreted as formal statistical confidence intervals [30].

There is often interest in estimates of the sizes of KAPs at different levels, such as the national level, and subnational levels corresponding to units such as provinces or districts. National estimates are important for policy purposes such as estimation and projection of the number of people infected with HIV, response planning and resource allocation. Within each country, subnational estimates are often used for better program planning and management, such as assessing and meeting the needs for commodities, human resources and other program elements, measuring population coverage, and monitoring and evaluating interventions. It is important to provide probabilistic estimates and projections for low-level and concentrated HIV epidemics at both national and subnational levels.

KAPs such as PWID and female sex workers are of great interest to researchers because their behavior affects the spread of HIV and other diseases [7]. Unfortunately, standard sampling and estimation techniques cannot be used for these populations because most of them are hard to

*Corresponding author.

reach, and often actively avoid being contacted for official purposes. Standard methods require the researcher to select sample members with a known probability of selection, and typically no sampling frames are available for these populations that would make this possible.

[28] reviewed methods for sampling hard-to-reach and hidden populations for HIV surveillance, including snowball sampling, targeted sampling, facility-based sampling, time-location sampling, respondent-driven sampling and conventional cluster sampling. [29] reviewed methodological obstacles to conducting surveillance with key affected populations, and proposed criteria for choosing a sampling strategy. The following methods are commonly used for estimating the size of populations at risk for HIV [47]:

- Census and enumeration methods are based on counting individuals in the key affected populations.
- The capture-recapture method has typically been used with detailed mapping that identifies "hotspots" where KAPs are found. Two independent samples are taken, the overlap is determined and the standard Petersen estimator is used for the population [33, 26].
- The multiplier method uses two independent data sources, typically with one providing a count of the KAP in a service and the other providing an estimate of the proportion of the KAP enrolled in the service. The resulting population size estimate is given by the same formula as the capture-recapture estimate.

Each method has its own advantages and disadvantages, and each data source provides information about the size of a KAP. However, those methods are often used without any uncertainty assessment. Sometimes, it is hard to explain the inconsistency between estimates from different methods, or to extrapolate the KAP size to the districts with no data.

Here we propose a Bayesian hierarchical model for estimating the size of a KAP at both district and national levels, as well as for assessing the uncertainty of the estimates. The model incorporates multiple commonly used data sources, including mapping data, surveys, interventions, capture-recapture data, estimates or guesstimates from organizations, and expert opinion. The district-level parameters are assumed to follow the same distribution, and hence the model allows sharing of information across districts. We apply the approach to data used to estimate the number of males who inject drugs in Bangladesh, and we compare the results with what was obtained using the methods agreed on by the Bangladesh Technical Group. This is a nationally constituted expert technical working group, which was chaired by the National AIDS/STD (sexually transmitted diseases) Programme, and included experts from the government, the International Centre for Diarrhoeal Disease Research, Bangladesh, the Centre for Health and Population Research, non-governmental organizations that carry out HIV interventions, and development partners.

The first size estimation process in Bangladesh began in 2003 in response to the need to provide UNAIDS with a national estimate of the number of HIV-infected people. The main KAPs in the country are people who inject drugs, female, male and transgender sex workers, clients of sex workers, men who have sex with men, and returnee external migrants. Family Health International provided technical assistance to the Bangladesh Technical Group. The goal was to reach consensus and produce data-informed estimates through a transparent collaborative process involving all the key stakeholders. The estimate was based on this collaborative process rather than on a unified statistical model. The final results were obtained by November 2004, and received government approval in December 2005. Our goal in this paper is to develop a formal statistical model and method for estimating the KAPs, combining the same data sources that were used by the expert technical working group.

In Section 2 we present our Bayesian hierarchical model and the Markov chain Monte Carlo algorithm used for estimating it. Section 3 describes the estimation of the number of males who inject drugs in Bangladesh. Section 4 shows results from simulated examples designed to assess the method and the potential impact of dependence between the probabilities of an individual being included in two different lists or capture occasions. In Section 5 we discuss outstanding issues and possible extensions.

2. METHODS

2.1 Bayesian hierarchical model

There are 64 districts in Bangladesh, and the availability of data for HIV key affected population size estimation varies between districts and between KAPs. For the ith district, let n_i be the size of the target population that we want to estimate, such as male PWID. The data to be used for estimating n_i consist of:

- N_i: the size of a reference population, e.g. adult males as a reference population for male PWID.
- $X_i = (X_{i01}, X_{i10}, X_{i11})$: capture-recapture data, or two listings with known overlap, from the ith district. X_{i01} is the number observed in the second list but not the first, X_{i10} is the number observed in the first list but not the second, and X_{i11} is the overlap, i.e. the number observed in both lists. X_{i00} is the number not observed in either list and is unobserved. We denote by $X_{i1} = X_{i11} + X_{i10}$ the number in the first list and by $X_{i2} = X_{i11} + X_{i01}$ the number in the second list. By construction, $X_{i11} + X_{i10} + X_{i01} + X_{i00} = n_i$.
- Y_i: an incomplete count of the target population such that $Y_i < n_i$, for example a mapping observation, a survey or an intervention.
- Z_i: an estimate or guesstimate of n_i from other sources, which could be greater or less than n_i.

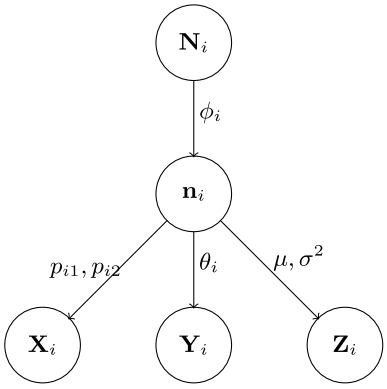

Figure 1. Data and Model Structure for Size Estimation in District i.

We use the following sampling models for the relationships between observed data and the target population size within the ith district:

(1)
$$\begin{cases} n_i|N_i,\phi_i & \sim \text{Binomial}(N_i,\phi_i), \\ X_{i1}|n_i,p_{i1} & \sim \text{Binomial}(n_i,p_{i1}), \\ X_{i2}|n_i,p_{i2} & \sim \text{Binomial}(n_i,p_{i2}), \\ X_{i00}|X_{i11}+X_{i10}+X_{i01},p_i & \sim \text{Negative Binomial}(X_{i11}+ \\ & \quad X_{i10}+X_{i01}, p_{i1}+p_{i1}-p_{i1}p_{i2}), \\ Y_i|n_i,\theta_i & \sim \text{Binomial}(n_i,\theta_i), \\ \log(Z_i)|n_i & \sim N(\mu+\log(n_i),\sigma^2), \end{cases}$$

where ϕ_i is the expected value of n_i/N_i, p_{ik} is the probability of inclusion in the kth overlapping list for $k = 1, 2$, and θ_i is the probability of being included in Y_i in the ith district. Since n_i is a count and Z_i is its guesstimate, we compare those two on the log scale. The parameter μ is the bias in $\log Z$ and σ^2 is its variance, both of which are constant across districts. Note that the normal distribution in the last equation of (1) is approximate, as Z_i and n_i are both integers. The conditional negative binomial distribution of X_{i00} follows from the assumption of independence between the two lists [13]. Figure 1 summarizes the data that may be available for size estimation in district i, and the parameters of our model.

To describe the heterogeneity of population proportions and inclusion probabilities across districts, we use the following between-district sampling models:

(2)
$$\begin{cases} \phi_i|a_0,b_0 & \sim \text{Beta}(a_0,b_0), \\ p_{i1}|a_1,b_1 & \sim \text{Beta}(a_1,b_1), \\ p_{i2}|a_2,b_2 & \sim \text{Beta}(a_2,b_2), \\ \theta_i|a_3,b_3 & \sim \text{Beta}(a_3,b_3), \end{cases}$$

where $\text{Beta}(a_m, b_m)$ denotes the beta distribution with mean $\pi_m = a_m/(a_m+b_m)$ and variance $\pi_m(1-\pi_m)/(a_m+b_m+1)$, for $m = 0, 1, 2, 3$. The hierarchical structure of our basic model represents the uncertainty in both the within-district sampling variability from equation (1) and the between-district sampling variability from equation (2).

We assign prior distributions to a_m, b_m, μ and σ^2, making it a Bayesian hierarchical model. We used the priors $p(a_m, b_m) \propto 1/(a_m + b_m)^2 I(a_m > 1, b_m > 1)$ for $m = 0, 1, 2, 3$, to represent vague prior information about ϕ, p_1, p_2 and θ [43, 13]. These priors are chosen so that the likelihood dominates the prior, in the sense that the prior is relatively flat over the part of parameter space in which the likelihood is substantial, and is not much greater outside this area. This also ensures that the results will be relatively insensitive to reasonable changes in these priors [8].

We use independent prior distributions for μ and σ^2, namely $\mu \sim N(\mu_0, \tau_0^2)$ and $\sigma^2 \sim \text{InverseGamma}(\nu_0/2, \nu_0\sigma_0^2/2)$. We set $\mu_0 = 0$ and $\tau_0 = \log(10)/2 = 1.15$, which implies that $\exp(\mu)$ is likely to be in the range $(0.1, 10)$, so that z_i is unlikely to be systematically biased by a factor of more than 10 in either direction. We chose $\nu_0 = 1$ and $\sigma_0 = \log(10)/2 = 1.15$ to represent weak prior information about σ^2.

In this hierarchical model, the district-level data (X_i, Y_i, Z_i, N_i) affect the district-level parameter estimates $n_i, p_i, \phi_i, \theta_i$, which in turn affect the national-level parameter estimates a_m and b_m. The national-level parameters will then influence the parameter estimates in other districts. Thus the model allows estimation for districts without data based on districts for which there are data.

2.2 Markov chain Monte Carlo algorithm

We estimate the joint posterior distribution of the parameters in our Bayesian hierarchical model by Markov chain Monte Carlo. Most of the parameters can be updated using Gibbs sampling. The algorithm is as follows:

1. Initialization:
   (a) Set the initial values $(a_0^{(0)}, b_0^{(0)}) = (2, 2000)$, $(a_m^{(0)}, b_m^{(0)}) = (2, 2)$ for $m = 1, 2, 3$.
   (b) For $i = 1, \ldots, d$, sample $\phi_i^{(0)} \sim \text{Beta}(a_0^{(0)}, b_0^{(0)})$, where d is the number of districts.
   (c) Sample $n_i^{(0)} \sim \text{Binomial}(N_i, \phi_i^{(0)})$. If $n_i^{(0)}$ is less than the minimum number that have been directly observed from the target population, namely $\hat{n}_{i\max}$, then replace it by a new $n_i^{(0)}$ satisfying $n_i^{(0)} - \hat{n}_{i\max} \sim \text{NegativeBinomial}(\hat{n}_{i\max}, 0.9)$.
   (d) Set $\mu^{(0)} = \text{mean}[\log(Z_i/n_i)]$, $\sigma^{2(0)} = \text{var}[\log(Z_i/n_i)]$.
   (e) Set the iteration number, k, to 1.
2. Update the within-district parameters, $i = 1, \ldots, d$:
   (a) Sample $p_{i1}^{(k-1)}$ from $\text{Beta}(X_{i1} + a_1^{(k-1)}, n_i^{(k-1)} - X_{i1} + b_1^{(k-1)})$.
   (b) Sample $p_{i2}^{(k-1)}$ from $\text{Beta}(X_{i2} + a_2^{(k-1)}, n_i^{(k-1)} - X_{i2} + b_2^{(k-1)})$.
   (c) Sample $\theta_i^{(k-1)}$ from $\text{Beta}(Y_i + a_3^{(k-1)}, n_i^{(k-1)} - Y_i + b_3^{(k-1)})$.

(d) Sample $\phi_i^{(k)}$ from Beta$(n_i^{(k-1)} + a_0^{(k-1)}, N_i - n_i^{(k-1)} + b_0^{(k-1)})$.

(e) Sample $n_i^{(k)}$ by using the Metropolis-Hastings (MH) algorithm with proposal distribution Poisson$(n_i^{(k-1)})$, as Gibbs sampling is not available.

3. Update the between-district parameters:

We sample $\alpha_m = \log(a_m/(a_m + b_m))$ and $\beta_m = \log(a_m + b_m)$ instead of a_m and b_m. The prior density of the transformed parameters α_m and β_m is then $p(\alpha_m, \beta_m) \propto \exp(\alpha_m) I(\alpha_m < 0, \beta_m > \log(2))$.

(a) For $m = 0, 1, 2, 3$, update $\alpha_m^{(k)}$ by using the MH algorithm with the following proposal distribution: $N(\alpha_m^{(k)}, 0.25)$ truncated below at $-\beta_m^{(k)}$ and above at $\log(1 - \exp(-\beta_m^{(k)}))$. This ensures that $a_m^* > 1$ and $b_m^* > 1$, as required by the prior distribution.

(b) For $m = 0, 1, 2, 3$, update $\beta_m^{(k)}$ by using the MH algorithm with the following proposal distribution: $N(\beta_m^{(k)}, 1)$ truncated below at the lower bound $\min(-\alpha_m^{(k)}, -\log(1-\exp(\alpha_m^{(k)})))$. This ensures that $a_m^* > 1$ and $b_m^* > 1$, as required by the prior distribution.

(c) Update $\mu^{(k)}$ from $N\left(\sum_i \log(Z_i/n_i)\tau^{2(k)}/\sigma^{2(k-1)}, \tau^{2(k)}\right)$, where $\tau^{2(k)} = 1/(1/\tau_0^2 + \ell/\sigma^{2(k-1)})$ and ℓ is the number of districts that have Z_i available.

(d) Update $\sigma^{2(k)}$ from InverseGamma $(\nu_0 + l)/2, (\nu_0\sigma_0^2 + \sum_i(\log(Z_i/n_i) - \mu^{(k)})^2)/2)$.

(e) Set $k \longleftarrow k + 1$ and return to 2.

Note that not all the sources of data $(Y_i, X_{i1}, X_{i2}, Z_i)$ are generally available at the district level. We used only the districts where at least one data source was available to estimate the model parameters. For the remaining districts, we imputed n_i from the hierarchical structure as follows: at each iteration, sample $\phi_i^{(k)} \sim$ Beta$(a^{(k)}, b^{(k)})$, and then sample $n_i^{(k)} \sim$ Binomial$(N_i, \phi_i^{(k)})$. To obtain the posterior distribution of the total size of the KAP, we summed n_i over all districts for each MCMC iteration. At each iteration, the prevalence of the KAP was estimated by dividing the size of the KAP by the population size.

The run-length diagnostic of [38] was used to assess the convergence of the Markov chain. It uses a relatively short pilot run of the Markov chain to determine the number of iterations and the degree of thinning needed to estimate the quantiles of interest to the desired level of accuracy. A longer Markov chain was then run with length determined by the Raftery-Lewis diagnostic. Convergence was also checked using trace plots and autocorrelation function estimates.

3. ESTIMATING THE NUMBER OF MALES WHO INJECT DRUGS IN BANGLADESH IN 2004

Bangladesh has transitioned from a low-level epidemic to a concentrated epidemic, with especially high rates among people who inject drugs (PWID) [2]. We applied the Bayesian hierarchical model to data on the number of PWID in Bangladesh in 2004 from several sources and sampling methods. PWID were defined as male drug users who had taken drugs primarily intravenously in the previous three to six months. Female PWID were excluded because the data indicated that there were few of them, and because many of them would have been already counted as female sex workers.

3.1 Results from the multiplier method

[40] described the 2004 size estimation procedure of PWID in Bangladesh using a multiplier method that led to an estimate of 20,000 to 40,000. The data used are shown in Table 1. We now summarize the previously used muliplier method.

The most nationally comprehensive data on PWID at the time were from the National Assessment of the Situation and Response to Opioid/Opiate Use in Bangladesh (NASROB) which surveyed 24 districts out of the 64 in the country. For NASROB, information was collected in 2001 from drug-using key informants at mapped public drug spots and secondary sources to derive a sampling frame of PWID for further survey. An estimate was made that there were 4,952 injectors in the 24 districts surveyed [32]. However, this was known to be an underestimate since the NASROB was not intended as a size estimation exercise, and was based on a comparison with the number of PWID reached by interventions.

CARE Bangladesh was the only non-governmental organization with PWID interventions at that time, and it provided service delivery data from a Needle Exchange Program (NEP), which contained the numbers of enrolled PWID in three cities from 2001 to 2003. The 2002 Behavioral Surveillance Surveys (BSS) in four cities included PWID who reported that they had enrolled in an NEP intervention in the preceding year. Hence, 2001 CARE NEP intervention enrollments and 2002 BSS formed two independent sources of data on PWID, where the intervention coverage data from BSS could be used to calculate a multiplier to inflate the NASROB estimates for Dhaka, Rajshahi and Chapai Nawabganj.

The NASROB counts for the remaining districts were also multiplied, using the Dhaka-derived multiplier 2.7 to inflate the NASROB counts in districts whose population densities were higher than 1,000 persons per km^2, and using the NASROB figure directly for districts with lower population densities. The resulting estimate was that there were

Table 1. Data for Estimating the Number of Males Who Inject Drugs. The last column shows the number of adult males in each district, from the 2001 Bangladesh Census

Index	District	NASROB Y	BSS 2002 X_1	NEP 2001 X_2	RSA Estimate Z	Population Size (Adult Males)
Dhaka Division, 8 districts						
1	Dhaka	2748	1759 PWID (46% in NEP)	3436	4287	2,364,000
2	Gazipur	258				524,500
3	Narayanganj	85				562,500
4	Manikganj	12				316,500
5	Narsingdi	0			76	480,000
6	Kishoreganj	0				635,000
7	Mymensingh	0			1000	1,123,500
8	Faridpur	2			50	433,500
Chittagong Division, 4 districts						
9	Chittagong	67				1,707,000
10	Chandpur	202	161		400 ∼ 500	538,000
11	Brahmanbaria	2				585,500
12	Cox's Bazaar	1			500	452,500
Rajshahi Division, 9 districts						
13	Rajshahi	770	710 PWID	376		579,000
14	Chapai Nawabganj	400	201 PWID (60% in NEP)	191		355,500
15	Pabna	99			140	551,000
16	Sirajganj	103			18	692,000
17	Bogra	20			18	759,500
18	Dinajpur	35			350	669,000
19	Rangpur	NA			350	645,000
20	Naogaon	NA			550	602,000
21	Joypurhat	NA			180	215,500
Khulna Division, 4 districts						
22	Khulna	14				605,500
23	Jhenaidah	9				397,500
24	Jessore	72				625,000
25	Satkhira	7				465,500
Barisal Division, 1 district						
26	Barisal	46			500	582,500
Sylhet Division, 2 districts						
27	Sylhet	0				652,000
28	Maulvi Bazar	NA			162	402,000

approximately 13,000 PWID in the 24 districts with a NASROB survey.

To make PWID size estimates for the remaining 40 districts of Bangladesh, the average number of PWID as a proportion of the adult male population (0.03%) was calculated from the 24 NASROB districts. Using this prevalence of injection drug use, an additional 5,000 PWID were assigned to the remaining districts. After combining the district estimates, the Bangladesh Technical Group settled on a national size range of 10,000 ∼ 20,000 PWID. The range was further multiplied by 2 using CARE Bangladesh Rapid Situation Assessment (RSA) data, raising the final national size range to 20,000 ∼ 40,000 PWID in Bangladesh in 2004.

3.2 Results from the Bayesian hierarchical model

We used the data on PWID from 28 districts in Bangladesh, shown in Table 1. For district i, we let n_i be the number of adult male PWID and N_i be the size of the adult male population. Let ϕ_i be the expected prevalence of intravenous drug use among adult males, and let $X_i = (X_{i1}, X_{i2})$ be the multiplier data consisting of X_{i1}, the number of PWID who participated in the BSS 2002 survey, and X_{i2}, the number of PWID enrolled in the NEP program. Also, we denote by Y_i the number of PWID included in the NASROB survey, and by Z_i the RSA estimate.

We ran the MCMC algorithm for 500,000 iterations,

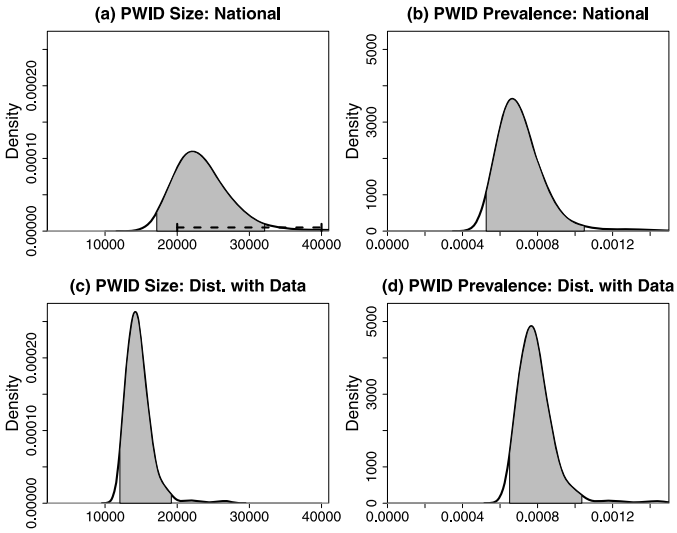

Figure 2. Posterior density of PWID population size (left) and PWID prevalence (right). The 95% credible intervals are shown by the shaded areas. (a) The posterior distribution of PWID size at the national level. The dashed horizontal line indicates the Bangladesh Technical Group's size estimate. (b) The posterior distribution of PWID prevalence at national level. (c) The posterior distribution of PWID size for districts with data. (d) The posterior distribution of PWID prevalence for districts with data.

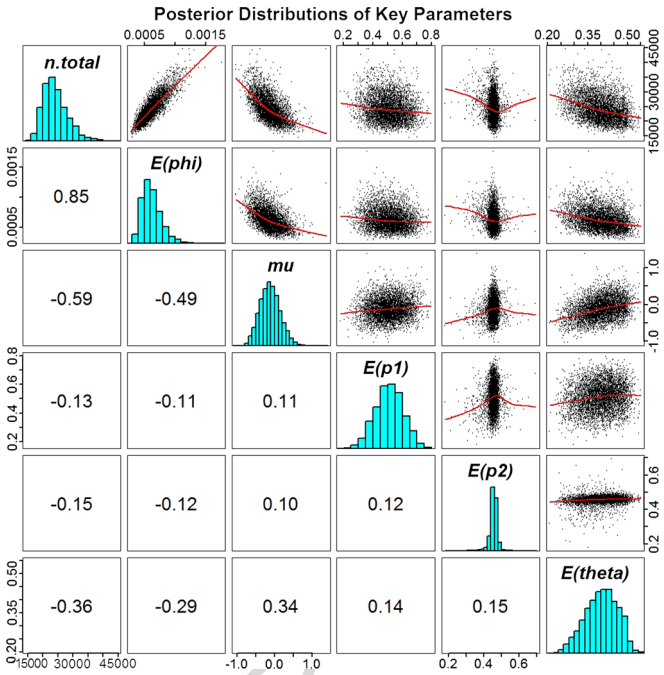

Figure 3. Marginal Posterior Distributions of Parameters in the Bayesian Hierarchical Model for Estimating the Number of PWID: the national PWID size, the expected PWID prevalence, the bias of expert estimate, the expected NASROB participation rate, the expected BSS participation rate, and the expected NEP participation rate. The upper panels are pairs plots, the lower panels are Pearson correlations, and the panels on the diagonal are histograms.

dropping the first 5,000 iterations as burn-in, and keeping every 100th scan. The diagnostic of [38], as implemented in the `raftery.diag` function in the `coda` R package [35], indicated that this was sufficient to reach the area of substantial posterior density and to explore it adequately, as well as to achieve approximate independence of the posterior samples. This took 55 minutes of CPU time to run.

Figure 2(a) shows the posterior distribution of the total number of male PWID in Bangladesh in 2004. The posterior median is 22,454 and the 95% Bayesian confidence interval is [17,207, 32,100]; the half-length of the interval is 7,446. The Bayesian interval is narrower than the Bangladesh Technical Group's estimate of 20,000 $\sim$ 40,000, but overlaps with it substantially. Figure 2(b) shows the posterior distribution of PWID prevalence at the national level, which has median 0.07%, and 95% confidence interval [0.05%, 0.10%].

The adult male population size in Bangladesh in 2004 is estimated to have been 31.3 million. Of these, 18.5 million lived in the 28 districts for which there is at least one data source, and 12.8 million lived in the 36 districts without any data. Our analysis treats PWID prevalence in the different districts as exchangeable *a priori*, in particular implying that districts without data are similar to districts with data in terms of PWID prevalence. This may not be the case, for example if data collection efforts have focused on the districts with the most PWID.

As an extreme sensitivity analysis, we computed the posterior distribution of the number and prevalence of PWID in the districts with data only, shown in Figures 2(c) and 2(d). This could be viewed as an extreme solution, corresponding to the assumption that there are no PWID in the districts without data. The posterior median is 14,700 with 95% credible interval [12,300, 19,200]. The interval is much narrower than for the PWID population for all districts. The lower bound is not much lower than for the whole PWID population, 12,300 compared to 17,207. The upper bound is much lower, however.

The histograms and pairs plots in Figure 3 show the marginal posterior distributions of several parameters of interest. These include the total number of PWID, the prevalence of intravenous drug use in the adult male population, and $E(\phi_i)$, which has mean 0.62/1000 and standard deviation 0.16/1000. Also shown are results for μ, the bias parameter for the RSA estimates, which has mean -0.119 and standard deviation 0.272, the probability of participation in the BSS survey $E(p_{i1})$, which has mean 0.51 and standard deviation 0.09, the probability of enrollment in the NEP intervention $E(p_{i2})$, which has mean 0.46 and standard deviation 0.03, and the probability of participation in

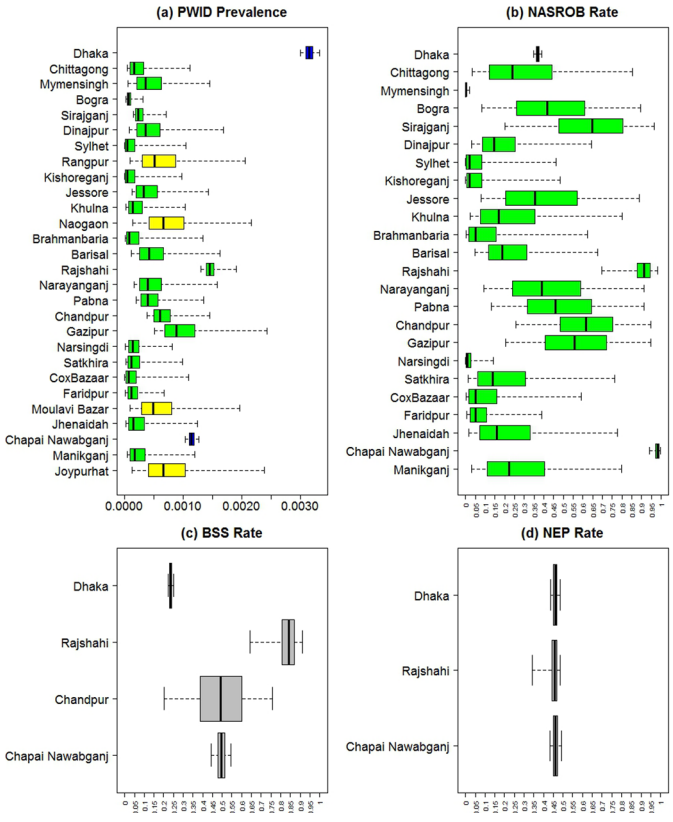

Figure 4. Posterior distributions: (a) PWID prevalence rates ϕ_i: blue for districts where multiplier data were available, green for districts with NASROB survey data but no capture-recapture data, yellow for districts with only an RSA estimate; (b) NASROB participation rates θ_i; (c) BSS participation rates p_{i1}; (d) NEP participation rates p_{i2}. For each boxplot, the box shows the posterior interquartile range and the dashed line goes from the .025 posterior quantile to the .975 posterior quantile.

the NASROB survey $E(\theta_i)$, which has mean 0.40 and standard deviation 0.06.

Figure 4 shows the posterior distribution of the PWID prevalence rates ϕ_i and the rates of participation in NASROB, θ_i, BSS, p_{i1}, and the NEP, p_{i2}, in each district. The districts are ordered by adult male population size, with the biggest population at the top and the smallest at the bottom. The PWID prevalence rate is highest in the capital, Dhaka, where it is well estimated. The participation rates in NASROB and BSS varied widely, while the participation rate in the NEP varied little between the three districts where the NEP was active.

Finally, we evaluated the contributions of the different data sources by removing each data source individually, one at a time, and recomputing the estimate with the remaining sources. Table 2 summarizes the posterior median and 95% credible interval of the total number of PWID in the absence of each data source individually. CRC stands for

Table 2. Posterior median and 95% credible interval of the total number of PWID in the absence of different types of data source

	All Data	NASROB Excluded	RSA Excluded	CRC Excluded
0.025 quantile	17,207	38,101	16,868	25,366
posterior median	22,454	62,598	21,878	65,626
0.975 quantile	32,100	99,974	33,596	195,532

the overlap between BSS and NEP, which form the capture-recapture data. We did not remove BSS or NEP data individually because they played similar roles to the NASROB data in Figure 1; removing any one of them would lead to the removal of the capture-recapture data as a whole. The overlap between BSS and NEP was available only in 3 districts, but its removal has the largest impact on the size estimates, particularly on the upper bound of the posterior interval. Without the presence of capture-recapture data, it is hard to distinguish between a low population size with a high probability of being counted or overestimates in RSA, and vice versa. This suggests that adding a few questions in the surveys to better understand the overlap of participants between data sources could improve the estimation method. The contribution of NASROB was also substantial because it was available in 24 districts.

The result was not greatly impacted by removing the RSA guesstimates: the 95% credible interval expanded slightly from [17,207, 32,100] to [16,868, 33,596]. This is because we did not have much information about the magnitude of its bias, and assumed a priori that the RSA guesstimates could be biased by a factor of more than 10 in either direction. If we had more information on the bias and variation the RSA guesstimates, this could be turned into more informative priors. However, even then the effect of the RSA guesstimates on the final result is modest.

To assess this, we calculated the effect of changing τ_0 and σ_0 from $\tau_0 = \sigma_0 = \log(10)/2 = 1.15$ to $\tau_0 = \sigma_0 = \log(2)/2 = 0.34$, implying that the guesstimate z_i is unlikely to be systematically biased by a factor of more than 2 in either direction, instead of the factor of 10 we have been using. Then the 95% credible interval shrinks only slightly from [17,207, 32,100] to [17,915, 31,078]. Even changing the factor to 1.1, so that $\tau_0 = \sigma_0 = \log(1.1)/2 = 0.48$, does not change the 95% credible interval much, to [18,426, 32,489], and the posterior median becomes 23,300 which is still close to our original estimate, 22,454.

In situations with less information from other sources, however, the RSA guesstimates could have a bigger impact on the final estimates if we had more information about their bias and measurement error magnitude. If we remove the capture-recapture data, comparing with the last column of Table 2, $\tau_0 = \sigma_0 = \log(1.1)/2 = 0.48$ provides a relatively

narrow 95% credible interval, [19,554, 65,507] with posterior median 28,882. Overall, the RSA guesstimates would increase our knowledge substantially only if we had enough prior information about their bias and variation.

4. SIMULATION STUDY AND SENSITIVITY ANALYSIS: DEPENDENCE AMONG CAPTURE PROBABILITIES

We did not know the actual number of HIV high-risk group members in any district, and so we were not able to assess the model directly by comparing estimates with true values. Instead we used simulation to assess the sensitivity of the results to the model assumptions. We have already seen that the results are not very sensitive to substantial changes in the precision parameter of the prior distribution of μ, the bias of the expert guesstimates. We also found that the results were relatively insensitive to reasonable changes in the prior distributions of the capture probabilities and participation rates, (results not shown).

As shown in the previous section, the capture-recapture data dominate the results. Following the model checking procedure suggested by [12], we tested the assumption of independence between capture and recapture. From each scanned posterior sample, we created a two by two contingency table given the observed data and the imputed hidden population size, and we then calculated the p-value for the Chi-square independence test. The p-values across posterior samples were approximately uniformly distributed between 0 and 1, indicating that the independence assumption is not inappropriate.

We assessed sensitivity of the results to the model assumption of individual homogeneity in capture probabilities in two different ways. First, we simulated datasets with heterogeneity in capture probabilities and assessed the resulting estimation bias in our method. Second, we modified our Bayesian hierarchical model and estimation method to incorporate specified levels of dependence between capture and recapture of the kind that could arise from heterogeneity, and applied it to our dataset for a range of values of the dependence parameter.

We now describe our simulation study of between-individual heterogeneity in capture probabilities. [31] and [36] discussed eight models, including all combinations of time effect, individual effect and behavioral response in capture probabilities. Our proposed model corresponds to what they called model M_t, which assumes independent capture probabilities between capture occasions, but equal capture probabilities on any particular occasion. Positive correlation between capture probabilities can lead this model to tend to underestimate population size [42]. An alternative model, M_{ht}, assumes that capture probabilities vary by occasion and by individual but are independent of the capture history. We call this variation individual heterogeneity; it can induce dependence of capture probabilities between capture occasions.

Table 3. Percentiles of the Distribution of Relative Errors of the Posterior Median for the Simulated Examples With and Without Individual-level Heterogeneity of Capture Probabilities

No individual effect					
Percentiles	2.5%	25%	50%	75%	97.5%
total number of PWID	−0.172	−0.068	−0.035	0.003	0.049

Individual effect: $\mathbf{c} = (0.4, 0.8, 1.2, 1.6)$					
Percentiles	2.5%	25%	50%	75%	97.5%
total number of PWID	−0.323	−0.180	−0.151	−0.119	−0.082

Individual effect: $\mathbf{c} = (0.1, 1.0, 1.3, 1.6)$					
Percentiles	2.5%	25%	50%	75%	97.5%
total number of PWID	−0.307	−0.234	−0.204	−0.177	−0.122

In two-occasion capture-recapture, there are three observed numbers, X_{i11}, X_{i10}, X_{i01}, and three parameters, p_{i1}, p_{i2} and n_i to be estimated in M_t, so that there is not enough information to estimate individual heterogeneity. Therefore we carried out a simulation study, modeled on our PWID data, to investigate the potential bias caused by individual heterogeneity. We generated simulated PWID data by using the posterior mean of the parameters of our model, as follows. For districts $i = 1, 2, \ldots, 64$:

- sample ϕ_i, the expected prevalence, from Beta$(1.2, 2000)$;
- sample n_i, the PWID size, from Binomial(N_i, ϕ_i);
- if NASROB data are available, sample Y_i from Binomial(n_i, θ_i);
- if a CARE Bangladesh guesstimate is available, sample Z_i from LogNormal$(\log(n_i) - 0.1, 1)$.

In districts with capture-recapture data, individual heterogeneity was constructed similarly to the M_{ht} experiments in [31]. Individuals were randomly assigned to one of four categories with multipliers $\mathbf{c} = (c_1, c_2, c_3, c_4)$, and the capture probability on the tth occasion for an individual in the kth category was $p_{ikt} = c_k p_{it}$. We simulated datasets from two scenarios, each with a different value of $\mathbf{c}$. In the first scenario, $\mathbf{c} = (0.4, 0.8, 1.2, 1.6)$, corresponding to strong heterogeneity. In the second scenario, $\mathbf{c} = (0.1, 1.0, 1.3, 1.6)$, also corresponded to strong heterogeneity, but this time with particularly low capture probability in one category. For each scenario, we generated 100 datasets and then applied the Bayesian hierarchical model to each dataset. We considered the posterior median as the point estimate and calculated the relative error for the estimated total number of PWID for each simulation.

Table 3 shows the distribution of relative errors over the 100 simulations. The top panel confirms that in the absence of individual heterogeneity among capture probabilities, there is no substantial bias. The lower two panels show that when there is strong heterogeneity among individuals,

the proposed model does tend to underestimate the population size, so that the resulting estimates are conservative. The heterogeneity in both our simulated scenarios is fairly extreme, and in our experiments the population size was rarely underestimated by more than 30%, so this could be viewed as a practical bound on the amount of underestimation to be expected in situations like the one we are considering.

We now describe our second way of assessing sensitivity to heterogeneity in capture probabilities. This consists of modifying our Bayesian hierarchical model and MCMC estimation method to include a user-specified assumed level of positive dependence between capture probabilities. Such dependence could arise from heterogeneity. We then assess the difference between the estimate from the modified method, assuming that the dependence parameter is known, and the estimate from our original method that ignores heterogeneity. These differences, for a range of plausible values of the dependence parameter, give an idea of the possible bias resulting from ignoring heterogeneity.

To describe the dependence between two capture occasions, we define ρ as the ratio of the joint probability of capture on both occasions to the product of the two marginal capture probabilities:

$$\rho = \frac{P(\text{included in both capture occasions})}{P(\text{included in the first capture})} \times \frac{1}{P(\text{included in the second capture})}. \tag{3}$$

If there were a third capture occasion or individual records to match the capture-recapture data with other data sources, we could estimate ρ directly. However, such information was not available for the 2004 size estimation exercise.

To reflect the effect of dependence, the only modification in the MCMC sampling procedure is that the probability in the negative binomial distribution in the fourth equation of (1) becomes:

$$P(\text{captured on either the first or second occasion}) = p_1 + p_2 - \rho p_1 p_2.$$

To specify a range of plausible values of the dependence parameter ρ, we first note that when there is no dependence, $\rho = 1$. Further, in the presence of individual-level heterogeity, the capture occasions will be positively related, in which case $\rho > 1$. Finally, we note that in the simulation study, $\mathbf{c} = (0.4, 0.8, 1.2, 1.6)$ corresponds to $\rho = \frac{0.4^2+0.8^2+1.2^2+1.6^2}{4} = 1.2$; and $\mathbf{c} = (0.1, 1.0, 1.3, 1.6)$ corresponds to $\rho = \frac{0.1^2+1.0^2+1.3^2+1.6^2}{4} = 1.315$. These values correspond to fairly extreme levels of heterogeneity, and so their range would seem adequate to capture most likely levels of dependence. We therefore considered this range and expanded it slightly, considering values of ρ between 1 and 1.35.

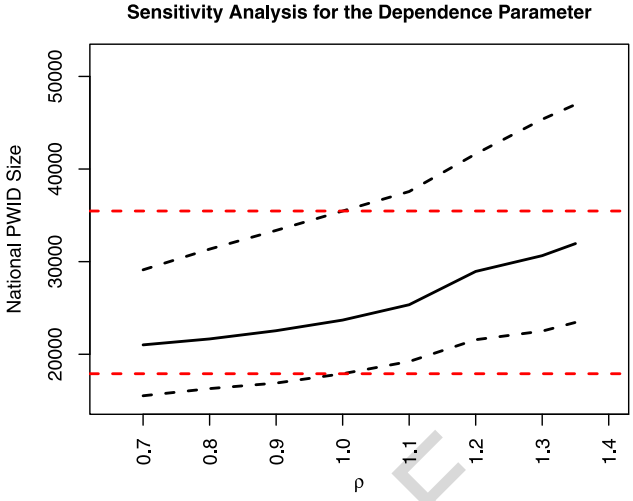

Figure 5. Results from Modified Model With Known Dependence Between Capture Probabilities for the Bangladesh PWID Data: The solid line is the posterior median for different values of the dependence parameter, ρ. The black dashed lines are 95% credible intervals. The red dashed lines are the 95% credible intervals when $\rho = 1$.

Figure 5 shows how the the posterior median and 95% credible interval vary with ρ. Both the point estimate and interval estimates of national PWID increase with ρ. For the most extreme value considered, $\rho = 1.35$, the PWID estimate is 31,949, compared with 22,454 ignoring heterogeneity. Thus the ratio of the estimate ignoring heterogeneity to the one that takes account of it is 0.703. This is in line with the conclusion from our simulation study that ignoring heterogeneity in these data is unlikely to lead to a downward bias of much more than about 30% in practical cases.

5. DISCUSSION

We have presented a Bayesian hierarchical model for estimating the size of populations at higher risk of HIV, which is easy to implement and to communicate to users. The hierarchical approach is attractive because it pools local and national information, provides estimates for all districts with their uncertainties, and incorporates multiple data sources. The basic model follows the assumptions made by the Bangladesh Technical Group, such as independent capture/inclusion probabilities and exchangeability of the district level parameters. It also takes account of two major sources of heterogeneity between districts: heterogeneity in the KAP as a proportion of the total, and heterogeneity in the probability of members of the KAP being included in the available data sources.

We have applied the method to estimating the number of males who inject drugs in Bangladesh using data from multiple listings and districts, namely mapping data, behavioral surveillance survey, service delivery data and capture-

recapture data. The model leads to narrower credible intervals than the intervals produced by the Bangladesh Technical Group, but overlaps with them substantially.

The Bangladesh Technical Group pointed out that there were fewer PWID outside the urban centers, and much of the data used for estimation was collected in the large cities rather than in the whole district. Therefore the participation rates in the data actually reflect both the participation rate in the urban area and the proportion of the district's population that is urban. If the proportion urban in each district was available, the accuracy of our estimates could be further improved by including it in our model.

Due to paucity of information, our basic model assumes that the district-level parameters are exchangeable and the district-level data are missing at random. More district level covariates have became available in recent years, and they could explain some of the district variation (e.g., urbanization measures). A possible improvement of the model would be to incorporate district-level covariates such as urbanization in a regression-type framework [14, 16, 15, 39, 48]. As an extreme sensitivity analysis, we computed the posterior distribution assuming that districts without data had no PWID, and found that the lower bound on population size was not too much smaller than in our main analysis.

Our model does not take account of differences in inclusion probabilities between individuals. It is possible that individual-level heterogeneity may introduce downward biases [42], and we have done some simulations of fairly extreme levels of heterogeneity to explore this. Ideally, we would incorporate individual-level heterogeneity into our model and estimate it as part of our method, but data to do this are not generally available. We recommend that data allowing the estimation of individual-level heterogeneity be collected in the future, such as three overlapping lists, rather than two [11].

There has been considerable research on the use of capture-recapture data for population size estimation; see the reviews by [21] [41], [37], [6], and [1]. Bayesian inference for capture-recapture data has been developed by [5], [43], [13], [27] and [46]. [3] studied individual heterogeneity and dependence. [23] developed a Bayesian approach to model capture-recapture data with covariates; this method could reduce the bias due to individual heterogeneity. For evaluating the U.S. Census, [9] and [10] poststratified the 2×2 table into poststrata with similar capture-recapture profiles and used Bayes factors for model comparison. [24] used Bayesian model averaging to incorporate model uncertainty into population size estimation. Here we have developed methods for multiple data sources including capture-recapture data.

Besides the commonly used data sources we have discussed, there are two more recent network-based methods that can provide data for population size estimation: respondent-driven sampling (RDS), and the network scale-up method. RDS is a chain-referral sampling method introduced by [20]; see [44] for inference from RDS data. [25] described the National HIV Behavioral Surveillance System (NHBS), in which the U.S. Center for Disease Control is using RDS for behavioral surveillance of high-risk HIV-related behaviors in PWID. RDS has the potential to provide information about population size as well as prevalence, although this has not yet been fully explored. It could be worth expanding our hierarchical model to incorporate information from RDS studies, although how to do this is not yet fully clear.

The network scale-up method is a social network estimator of the size of hidden or hard-to-count populations [22, 49]. The basic idea is that people's social networks are on average representative of the general population, and hence the average occurrence of any particular sub-population in personal networks reflects their prevalence in the general population. The method's advantage is that the estimation of the sizes of a KAP does not require reaching members of the at-risk population, but can be done by surveying respondents in the general population. However, there are still various factors that affect the accuracy of the final estimate that need to be resolved to make this method widely applicable. Once these issues are resolved, data from this method could be incorporated into our hierarchical approach.

Estimates of the sizes of KAPs can be politically sensitive due to the stigmatized nature of these populations in many countries. Hence non-data related issues can affect the final estimates [34, 40]. The Bayesian model may be useful for technical working groups in countries as it provides a tool that can be applied to multiple, biased size-related data sets, yielding a principled statistical basis for population size estimates, and confidence intervals.

ACKNOWLEDGEMENTS

This work was supported by NICHD grants R01 HD054511 and R01 HD070936, and Raftery's work was partly supported by a Science Foundation Ireland ETS Walton visitor award, grant number 11/W.1/I2079. Its contents are solely the responsibility of the authors and do not necessarily represent the official views of the National Institute of Child Health and Human Development or those of the United Nations or the Ministry of Health and Family Welfare, Government of Bangladesh. Its contents have not been formally edited and cleared by the United Nations. The authors are grateful to Peter Ghys, the Editor, the Associate Editor and two anonymous reviewers for helpful comments.

Received 24 October 2013

REFERENCES

[1] Amstrup, S., McDonald, T., and Manly, B. (2005). *Handbook of Capture-Recapture Analysis*. Princeton University Press.

[2] Azim, T., Rahman, M., Alam, M., Chowdhury, I., Khan, R., Reza, M., Rahman, M., Chowdhury, E., Hanifuddin, M., and Rahman, A. (2008). Bangladesh moves from being a low-prevalence nation for HIV to one with a concentrated epidemic in injecting drug users. *International Journal of STD & AIDS* **19**, 327–331.

[3] Basu, S. and Ebrahimi, N. (2001). Bayesian capture-recapture methods for error detection and estimation of population size: Heterogeneity and dependence. *Biometrika* **88**, 269–279. MR1841274

[4] Brown, T. and Peerapatanapokin, W. (2004). The Asian Epidemic Model: A process model for exploring HIV policy and programme alternatives in Asia. *Sexually Transmitted Infections* **80**, i19–i24.

[5] Castledine, B. (1981). A Bayesian analysis of multiple-recapture sampling for a closed population. *Biometrika* **67**, 197–210. MR0614956

[6] Chao, A. (2001). An overview of closed capture-recapture models. *Journal of Agricultural, Biological, and Environmental Statistics* **6**, 158–175.

[7] Commission on AIDS in Asia (2008). *Redefining AIDS in Asia: Crafting an effective response*. Oxford University Press, India, New Delhi.

[8] Edwards, W., Lindeman, H., and Savage, L. J. (1963). Bayesian statistical inference for psychological research. *Psychological Review* **70**, 193–242.

[9] Elliott, M. R. and Little, R. J. A. (2000). A Bayesian approach to combining information from a census, a coverage measurement survey and demographic analysis. *Journal of the American Statistical Association* **95**, 351–362.

[10] Elliott, M. R. and Little, R. J. A. (2005). A Bayesian approach to 2000 census evaluation using A.C.E. survey data and demographic analysis. *Journal of the American Statistical Association* **100**, 380–388. MR2170461

[11] Fienberg, S. E., Johnson, M., and Junker, B. W. (1999). Classical multilevel and Bayesian approaches to population size estimation using multiple lists. *Journal of the Royal Statistical Society, Series A* **162**, 383–405.

[12] Gelman, A., Mechelen, I. V., Verbeke, G., Heitjan, D., and Meulders, M. (2005). Multiple imputation for model checking: Completed-data plots with missing and latent data. *Biometrics* **61**, 74–85. MR2135847

[13] George, E. I. and Robert, C. P. (1992). Capture-recapture estimation via Gibbs sampling. *Biometrika* **79**, 677–683. MR1209469

[14] Ghosh, M. and Meeden, G. (1986). Empirical Bayes estimation in finite population sampling. *Journal of the American Statistical Association* **81**, 1058–1062. MR0867632

[15] Ghosh, M., Natarajan, K., Stroud, T. W. F., and Carlin, B. (1998). Generalized linear models for small area estimation. *Journal of the American Statistical Association* **93**, 273–282. MR1614644

[16] Ghosh, M. and Rao, J. N. K. (1994). Small area estimation: An appraisal. *Statistical Science* **9**, 55–76. MR1278679

[17] Ghys, P. D., Brown, T., Grassly, N. C., Garnett, G., Stanecki, K. A., Stover, J., and Walker, N. (2004). The UNAIDS estimation and projection package: A software package to estimate and project national HIV epidemics. *Sexual Transmitted Infections* **80**, i5–i9.

[18] Ghys, P. D., Walker, N., and Garnett, G. P. (2006). Improving analysis of the size and dynamics of AIDS epidemics. *Sexually Transmitted Infections* **82**, iii1–iii2.

[19] Grassly, N. C., Morgan, M., Walker, N., Garnett, G., Stanecki, K. A., Stover, J., Brown, T., and Ghys, P. D. (2004). Uncertainty in estimates of HIV/AIDS: The estimation and application of plausibility bounds. *Sexually Transmitted Infections* **80**, i31–i38.

[20] Heckathorn, D. (1997). Respondent-driven sampling: A new approach to the study of hidden populations. *Social Problems* **44**, 174–199.

[21] Hook, E. B. and Regal, R. R. (1995). Capture recapture methods in epidemiology: Methods and limitations. *Epidemiol Reviews* **17**, 243–264.

[22] Killworth, P. D., Johnsen, E. C., McCarty, C., Shelley, G. A., and Bernard, H. R. (1998). A social network approach to estimating seroprevalence in the United States. *Social Networks* **20**, 23–50.

[23] King, R. and Brooks, S. P. (2001). On the Bayesian analysis of population size. *Biometrika* **88**, 317–316. MR1844834

[24] King, R. and Brooks, S. P. (2008). On the Bayesian estimation of a closed population size in the presence of heterogeneity and model uncertainty. *Biometrics* **64**, 816–824. MR2526632

[25] Lansky, A., Sullivan, P. S., Gallagher, K. M., and Fleming, P. L. (2007). HIV behavioral surveillance in the U.S.: A conceptual framework. *Public Health Reports* **122**, 16–23.

[26] Lincoln, F. C. (1930). Calculating waterfowl abundance on the basis of banding returns. *United States Department of Agriculture Circular* **118**.

[27] Madigan, D. and York, J. C. (1997). Bayesian methods for estimation of the size of a closed population. *Biometrika* **84**, 19–31. MR1450189

[28] Magnani, R., Sabin, K., Saidel, T., and Heckathorn, D. (2005). Review of sampling hard-to-reach and hidden populations for HIV surveillance. *AIDS* **19**, s67–s72.

[29] Mills, S., Saidel, T., Magnani, R., and Brown, T. (2004). Surveillance and modelling of HIV, STI, and risk behaviours in concentrated HIV epidemics. *Sexually Transmitted Infections* **80**, ii57–ii62.

[30] Morgan, M., Walker, N., Gouws, E., Stanecki, K. A., and Stover, J. (2006). Improved plausibility bounds about the 2005 HIV and AIDS estimates. *Sexually Transmitted Infections* **82**, iii71–iii77.

[31] Otis, D. L., Burnham, K. P., White, G. C., and Anderson, D. R. (1978). Statistical inference from capture data on closed animal populations. *Wildlife Monographs* **62**, 3–135.

[32] Panda, S., Mallick, P., Karim, M., Sharifuzzaman, M., Ahmed, A. H., and Baatsen, P. (2002). ... what will happen to us...? national assessment on situation and responses on opioid/opiate use in Bangladesh (NASROB). *Dhaka: FHI, CARE, and HASAB*.

[33] Petersen, C. G. J. (1896). The yearly immigration of young plaice into the Limfjiord from the German Sea. *Report of the Danish Biological Station* **6**, 1–48.

[34] Pisani, E. (2006). Estimating the number of drug injectors in Indonesia. *International Journal of Drug Policy* **17**, 35–40.

[35] Plummer, M., Best, N., Cowles, K., and Vines, K. (2006). Coda: Convergence diagnosis and output analysis for mcmc. *R News* **6**(1), 7–11.

[36] Pollock, K. H. (1991). Modelling capture recapture, and removal statistics for estimation of demographic parameters for fish and wildlife populations: Past, present, and future. *Journal of the American Statistical Association* **86**, 225–238.

[37] Pollock, K. H. (2000). Capture-recapture models. *Journal of the American Statistical Association* **95**, 293–296.

[38] Raftery, A. E. and Lewis, S. M. (1992). How many iterations in the Gibbs sampler? in: *Bayesian Statistics 4*, Bernardo, J. M., Berger, J. O., Dawid, A. P., and Smith, A. F. M. eds., Oxford University Press, Oxford, U.K., pp. 763–773.

[39] Rao, J. N. K. (2003). *Small area estimation*. Hoboken, N. J., John Wiley. MR1953089

[40] Reddy, A., Hoqueb, M., and Kelly, R. (2008). HIV transmission in Bangladesh: An analysis of IDU programme coverage. *International Journal of Drug Policy* **19**, 37–46.

[41] Schwarz, C. and Seber, G. (1999). Estimating animal abundance: Review III. *Statistical Science* **14**, 427–456.

[42] Sekar, C. C. and Deming, W. E. (1949). On a method of estimating birth and death rates and the extent of registration. *Journal of the American Statistical Association* **44**, 101–115.

[43] Smith, P. J. (1991). Bayesian analyses for a multiple capture-recapture model. *Biometrika* **78**, 399–408. MR1131174

[44] Volz, E. and Heckathorn, D. (2008). Probability based estimation theory for respondent-driven sampling. *Journal of Offical Statistics* **24**, 79–97.

[45] Walker, N., Stover, J., Stanecki, K., Zaniewski, A. E., Grassly, N. C., Garcia-Calleja, J. M., and Ghys, P. D. (2004). The workbook approach to making estimates and projecting future scenarios of HIV/AIDS in countries with low level

and concentrated epidemics. *Sexually Transmitted Infections* **80**, i10–i13.

[46] Wang, X., He, C. Z., and Sun, D. (2007). Bayesian population estimation for small sample capture-recapture data using noninformative priors. *Journal of Statistical Planning and Inference* **137**, 1099–1118. MR2301466

[47] World Health Organization (2010). *Guidelines on Estimating the Size of Populations Most at Risk to HIV*. Geneva, World Health Organization.

[48] Zeger, S. L., See, L.-C., and Diggle, P. J. (1989). Statistical methods for monitoring the AIDS epidemic. *Statistics in Medicine* **8**, 3–21.

[49] Zheng, T., Salganik, M. J., and Gelman, A. (2006). How many people do you know in prison? *Journal of the American Statistical Association* **101**, 409–423. MR2256163

Le Bao
Department of Statistics
The Pennsylvania State Univeristy
USA
E-mail address: lebao@psu.edu

Adrian E. Raftery
Departments of Statistics and Sociology
University of Washington
USA
E-mail address: raftery@u.washington.edu

Amala Reddy
UNAIDS Regional Support Team for Asia and the Pacific
Thailand
E-mail address: reddya@unaids.org