

# *Harvard University*

Harvard University Biostatistics Working Paper Series

---

*Year* 2018

*Paper* 213

---

## Cross-sectional HIV Incidence Estimation Accounting for Heterogeneity Across Communities

Yuejia Xu\*

Oliver B. Laeyendecker†

Rui Wang‡

\*University of Cambridge, yx299@cam.ac.uk

†Johns Hopkins University, olaeyen1@jhmi.edu

‡Harvard University, rwang@hsph.harvard.edu

This working paper is hosted by The Berkeley Electronic Press (bepress) and may not be commercially reproduced without the permission of the copyright holder.

<https://biostats.bepress.com/harvardbiostat/paper213>

Copyright ©2018 by the authors.

## Cross-sectional HIV Incidence Estimation Accounting for Heterogeneity Across Communities

Yuejia Xu<sup>1</sup>, Oliver Laeyendecker<sup>2,3</sup>, and Rui Wang<sup>4,5,\*</sup>

<sup>1</sup>MRC Biostatistics Unit, University of Cambridge, Cambridge CB2 0SR, U.K.

<sup>2</sup>National Institute of Allergy and Infectious Diseases, Baltimore, Maryland, 21205, U.S.A.

<sup>3</sup>Departments of Medicine and Epidemiology, Johns Hopkins University, Baltimore, Maryland, 21205, U.S.A.

<sup>4</sup>Department of Population Medicine, Harvard Medical School and Harvard Pilgrim Health Care Institute,  
Boston, Massachusetts 02215, U.S.A.

<sup>5</sup>Department of Biostatistics, Harvard T. H. Chan School of Public Health, Boston, Massachusetts 02115, U.S.A.

\**email*: rwang@hsph.harvard.edu

**SUMMARY:** Accurate estimation of HIV incidence rates is crucial for the monitoring of HIV epidemics, the evaluation of prevention programs, and the design of prevention studies. Traditional cohort approaches to measure HIV incidence require repeatedly testing large cohorts of HIV uninfected individuals with a HIV diagnostic test (e.g., enzyme-linked immunosorbent assay) for long periods of time to identify new infections, which can be prohibitively costly, time-consuming, and subject to loss to follow-up. Cross-sectional approaches based on the usual HIV diagnostic test and biomarkers of recent infection offer important advantages over standard cohort approaches, in terms of time, cost, and attrition. Cross-sectional sample usually consists of samples from different communities. However, small sample sizes limit the ability to estimate community-specific incidences and existing methods typically ignore heterogeneity in incidence across communities. We propose a permutation test for the null hypothesis of no heterogeneity in incidence rates across communities, develop a random effects model to account for this heterogeneity and to estimate community-specific incidences, and provide one way to estimate the coefficient of variation. We evaluate the performance of the proposed methods through simulation studies and apply the proposed methods to the data from the National Institute of Mental Health Project ACCEPT, a phase III randomized controlled HIV prevention trial in Sub-Saharan Africa, to estimate the overall and community-specific HIV incidence rates.

**KEY WORDS:** Biomarkers; Coefficient of variation; Permutation test; Random effects model.

## 1. Introduction

Accurate estimation of HIV incidence rates is crucial for the monitoring of local HIV epidemics, the evaluation of the impact of prevention programs, and the design of prevention studies. Traditional approaches to measure HIV incidence require repeatedly testing large cohorts of HIV uninfected individuals with a HIV diagnostic test [e.g., enzyme-linked immunosorbent assay (ELISA)] for long periods of time to identify new infections, which can be prohibitively costly and time-consuming. Such studies are also subject to differential loss to follow-up and behavioral modification. An alternative approach entails biomarkers of HIV disease progression that can distinguish recent from long-term infections (Brookmeyer and Quinn, 1995). In this strategy, subjects are administered the usual HIV diagnostic test (e.g., ELISA); then, those found to be positive are tested with biomarkers associated with recent infection (e.g., HIV-1 limiting antigen avidity enzyme immunoassay). The incidence estimate is obtained by carefully combining results from the two tests and external information about the average time individuals appear to be “recently infected”. This approach allows investigators to estimate incidence by testing blood samples at a single point in time, can provide quick and inexpensive estimates of HIV incidence rates and “may lead to a revolution in the way that worldwide HIV epidemics are routinely tracked” (Hallett, 2011).

Cross-sectional surveys usually consist of samples from multiple communities with varying HIV incidence rates. However, standard methods typically assume that the cross-sectional sample is a random sample of the population of interest. That is, observations on individuals in the sample are assumed to be independent. When the cross-sectional sample consists of individuals from multiple communities, this independence assumption is likely to be violated because observations on individuals in the same community are usually correlated. This within-cluster correlation results from variability in the underlying community-specific incidences. If there is heterogeneity in incidence across communities, then individuals in

the same community will tend to have responses that are more similar to each other than responses of individuals in different communities. Therefore, for clustered data, between-cluster variability and within-cluster correlation provide two different perspectives on the same underlying phenomenon. When analyzing clustered data, one must account for the between-cluster variability (or within-cluster correlation). Ignoring this correlation may lead to biased inference. To illustrate, we present a contour plot of the actual coverage of 95% confidence intervals for the overall incidence across 30 communities when heterogeneity is ignored, for varying magnitude of incidences and the heterogeneity across communities (Figure 1; Details in the Web Appendix A). As shown in Figure 1, the actual coverage of 95% confidence intervals obtained ignoring the heterogeneity can be substantially lower than the nominal level in many settings.

[Figure 1 about here.]

Furthermore, when incidence rates vary across communities, it would be useful to obtain community-specific incidences and to quantify the heterogeneity in incidence rates across communities. Small sample sizes and low incidence rates limit the ability to estimate community-specific incidences. For example, in the baseline survey of the Botswana Combination Prevention Project (BCPP), a cluster randomized trial involving 30 communities in Botswana designed to evaluate the effectiveness and cost-effectiveness of a combination prevention package (BCPP, 2013), among 30 communities, no individuals (among 70 to 150 individuals tested in each community) were identified as recent infections by the incidence assay in 8 communities. A direct application of the existing formula would lead to an incidence estimate of 0 in these communities. In this paper, we propose methods that can produce improved community-specific incidence estimates. We also propose a permutation test for the null hypothesis that there is no heterogeneity in incidence rates across communities. A test for heterogeneity is indicative of whether or not substantial heterogeneity across communities

exists. The proposed permutation test is easy to implement and its validity does not rely on parametric assumptions.

Another purpose of this paper is to propose an estimator for the coefficient of variation (CV) based on cross-sectional data. The CV captures the heterogeneity in outcomes across communities and provides equivalent information regarding variance inflation as the intraclass correlation coefficient (ICC) as described above. The statistical power of a cluster randomized trial can change substantially depending on the CV. As we noted in designing the BCPP, for a matched-pair cluster randomized trial with 15 pairs and a sample size of 300 within each community, the power to detect a 40% reduction in 3-year cumulative incidence from 2.5% to 1.5% decreases from 80% to 52% as CV increases from 0.20 to 0.45 (i.e., an increase in the ICC from 0.001 to 0.005). Obtaining accurate information on the CV is often a major stumbling block in cluster randomized trials (Rutterford et al., 2015). Methods to estimate the CV for HIV incidence based on cross-sectional data at baseline provide a practical way to obtain this essential information for the design of cluster randomized trials.

The rest of the paper is organized as follows. In Section 2, we propose a random effects model to account for heterogeneity of incidence rates across communities and to estimate community-specific incidences, develop a permutation test for the null hypothesis of no heterogeneity across community-specific incidences, and propose an estimator of the CV for HIV incidence. In Section 3, we present results from simulation studies. In Section 4, we apply the proposed methods to data from the NIMH Project ACCEPT to estimate the overall and community-specific incidence rates. We conclude with discussions in Section 5.

## 2. Methods

### 2.1 Notation and Background

Suppose that  $N$  subjects are randomly selected from an asymptomatic population, and each is tested with an ELISA and, if positive, tested with biomarkers of recent infection. We

consider the three-state longitudinal natural history model (S-Figure 1(a); Web Appendix B) of HIV seroconversion and subsequent reactivity to biomarkers of recent infection as in Wang and Lagakos (2009): State 1 represents the pre-seroconversion state (uninfected or infected but not seroconverted); State 2 represents the “recent infection” state, in which an infected individual is identified as a “recent infection” by the biomarker; and State 3 represents the “long-term infection” state in which an infected individual is classified as a “non-recent infection” by the biomarker. Let  $N_1$ ,  $N_2$ , and  $N_3$  denote the number of subjects who test negative on both tests (State 1), positive for both the diagnostic test and biomarkers of recent infection (State 2), and positive for the diagnostic test but negative for biomarkers of recent infection (State 3), respectively, so that  $N = N_1 + N_2 + N_3$ . The actual time spent in State 2 varies from person to person and is assumed to be independent of the time of seroconversion. We use  $\mu$ , commonly termed as the “mean window period”, to denote the mean population time in State 2. Let  $\lambda$  and  $1 - \phi$  denote the incidence rate and the prevalence of long-term infection at the time of the cross-sectional sample.

In the setting of one population of interest, Balasubramanian and Lagakos (2010) and Wang and Lagakos (2010) proposed a likelihood-based approach and derived the probability of an individual falling into one of the three states (uninfected, recent infection and long-term infection). The likelihoods considered in these earlier work are especially suited for settings where the incidence is low. Here we consider a modification of the likelihood that is more general and can also accommodate settings where the incidence is large. Let  $p_1$ ,  $p_2$ , and  $p_3$  denote the prevalence probabilities in State 1, 2, and 3, respectively, it follows that  $p_1 = \phi - \phi\mu\lambda$ ,  $p_2 = \phi\mu\lambda$ , and  $p_3 = 1 - \phi$ . Note that  $(N_1, N_2, N_3)$  follows a multinomial distribution with parameters  $(p_1, p_2, p_3)$ . Therefore, the likelihood function for  $(\phi, \lambda)$  based on  $(N_1, N_2, N_3)$  is given by

$$L(\phi, \lambda) \propto (\phi - \phi\mu\lambda)^{N_1} (\phi\mu\lambda)^{N_2} (1 - \phi)^{N_3}. \quad (1)$$

The maximum likelihood estimators of  $(\phi, \lambda)$  can be obtained by maximizing  $L(\phi, \lambda)$ , resulting in  $\hat{\phi} = (N_1 + N_2)/N$ , and  $\hat{\lambda} = N_2/\{(N_1 + N_2)\mu\}$ . To compute the estimators above,  $N_1$ ,  $N_2$ , and  $N_3$  are obtained from the cross-sectional sample, and  $\mu$  is typically assumed to be known. Estimators for the variances of  $(\hat{\phi}, \hat{\lambda})$  can be derived from the sample Fisher information:  $\widehat{\text{Var}}(\hat{\phi}) = N_3(N_1 + N_2)/N^3$ , and  $\widehat{\text{Var}}(\hat{\lambda}) = N_1N_2/\{(N_1 + N_2)^3\mu^2\}$ .

## 2.2 Random Effects Model

Suppose there are  $M$  communities. Let  $\lambda_i$  be the true community-specific incidence in community  $i$ , for  $i = 1, \dots, M$ . Let  $N_{1i}, N_{2i}, N_{3i}$ , and  $N_i$  denote the number of subjects in State 1, 2, 3, and the total number of subjects in community  $i$ , respectively.

**2.2.1 Fixed Effects Model vs. Random Effects Model.** Analogous to meta-analysis, here we discuss the concept of a fixed effects model and a random effects model in the setting of cross-sectional incidence estimation. Under the fixed effects model, we assume all communities have a common incidence rate  $\lambda^*$ , that is,  $\lambda_i = \lambda^*$ , for  $i = 1, \dots, M$ . The observed incidence rates  $\hat{\lambda}_i$  are distributed around  $\lambda^*$ , and each of  $\hat{\lambda}_i$  estimates the same underlying incidence rate  $\lambda^*$ . The difference in observed incidences can be attributed purely to random sampling error, which depends primarily on the size of the cross-sectional sample within each community. The overall incidence based on the fixed effects model can be estimated by  $\hat{\lambda}^* = \sum_{i=1}^M N_{2i}/\{(\sum_{i=1}^M N_{1i} + \sum_{i=1}^M N_{2i})\mu\}$ , and the corresponding variance can be estimated by  $\widehat{\text{Var}}(\hat{\lambda}^*) = \sum_{i=1}^M N_{1i} \sum_{i=1}^M N_{2i}/\{(\sum_{i=1}^M N_{1i} + \sum_{i=1}^M N_{2i})^3\mu^2\}$ .

When the cross-sectional sample consists of subjects from various communities, and many factors can lead to variations in incidence rates across different communities, the assumption underlying the fixed effects model that incidence rates are the same is likely to be violated. This motivated us to consider a more flexible random effects model to account for the heterogeneity of incidence rates across communities. Here we assume community-specific incidence rates differ by community and are a random sample drawn from a distribution. In

this case, we consider the overall incidence as the mean of the random effects distribution, that is,  $E(\lambda_i) = \lambda^*$ , and  $\text{Var}(\lambda_i) = \tau^2$ , where  $\tau^2$  denotes the between-community variability in true incidence rates. Under the random effects model, the variability of the observed incidence rates  $\widehat{\lambda}_i$  results from both the within-community sampling error and the variation of true underlying incidence rates  $\lambda_i$  across communities.

*2.2.2 Random Effects Model Formulation.* Suppose that  $\lambda_i$  follow a Lognormal distribution and let  $\lambda_i = \lambda e^{v_i}$ , where  $v_i \stackrel{i.i.d}{\sim} N(0, \sigma^2)$ . We assume that data from  $M$  communities are independent and data from individuals in the same community are conditionally independent given  $v_i$ . Under this Lognormal random effects model, the overall incidence,  $\lambda^* = \lambda e^{\sigma^2/2}$ , and the between-community standard deviation of incidence,  $\tau = \lambda e^{\sigma^2/2} \sqrt{e^{\sigma^2} - 1}$ . Following (1), the conditional likelihood of community  $i$  is given by:

$$L_i(\phi, \lambda | v_i) \propto (\phi - \phi \mu \lambda e^{v_i})^{N_{1i}} (\phi \mu \lambda e^{v_i})^{N_{2i}} (1 - \phi)^{N_{3i}}, \quad (2)$$

and the marginal likelihood can be obtained by integrating over the random effects:

$$\begin{aligned} L(\phi, \lambda, \sigma^2) &= \prod_{i=1}^M \int_{v_i} L_i(\phi, \lambda | v_i) f(v_i | \sigma^2) dv_i \\ &\propto \prod_{i=1}^M \int_{v_i} \left\{ (\phi - \phi \mu \lambda e^{v_i})^{N_{1i}} (\phi \mu \lambda e^{v_i})^{N_{2i}} (1 - \phi)^{N_{3i}} \times \frac{1}{\sigma} \exp\left(-\frac{v_i^2}{2\sigma^2}\right) \right\} dv_i. \end{aligned} \quad (3)$$

There is no closed form solution for this integral. We will approximate this integral using numerical methods and then obtain the maximum likelihood estimators of  $(\phi, \lambda, \sigma^2)$  through the maximization of the approximated marginal likelihood.

Community-specific incidences  $\lambda_i$  can be estimated based on the empirical Bayes estimates of the random effects  $v_i$ . Unlike the standard community-specific incidence estimates, which can be calculated based on  $\widehat{\lambda}_i = N_{2i} / \{(N_{1i} + N_{2i})\mu\}$  using the observations from that community only, information is “borrowed” across communities under the proposed random effects model: the estimated community-specific incidences depend both on observations from that community and observations from other communities, such that the estimate



for each community is shrunk towards the overall incidence. This shrinkage effect is more pronounced for smaller communities and communities whose within-community variability is large relative to the between-community variability. The random effects model may lead to more efficient community-specific incidence estimates, and is especially useful in the HIV setting where the incidence is low and one single community usually has small sample size.

*2.2.3 Four-state Random Effects Model to Incorporate the False Recency Rate.* It has been noted that some infected individuals can repeatedly test positive for biomarkers of recent infection long after seroconversion (Novitsky et al., 2009; Claggett et al., 2012). In this case, the underlying assumption of the three-state model is violated, and the resulting incidence estimator is biased upwards by overestimating the number of subjects who are recently infected. Here we present an extension of the random effects model to accommodate this. Let  $1-p$ , commonly termed as the “false recency rate”, denote the proportion of subjects who would be tested as “recently infected” permanently in the HIV positive population. We highlight the importance of an accurate estimate of the false recency rate in order to estimate HIV incidence accurately. The estimation of  $p$  requires additional external information. This may be done via an augmented design where the cross-sectional sample is augmented with a longitudinal component (Wang and Lagakos, 2010), or by testing an independent sample of individuals with known long-term infection from the same population (Moyo et al., 2014).

We consider the four-state model as in Wang and Lagakos (2009) (S-Figure 1(b); Web Appendix B). State 1 and State 3 are defined the same way as in the three-state model, but we distinguish subjects being tested as recent infections into those who would eventually be classified as “long-term infected” (State 2) and those who would remain as “recently infected” indefinitely (State 4).  $N_1$  and  $N_3$  still represent the number of subjects in State 1 and State 3, but now  $N_2$  represents the total number of subjects in either State 2 or State 4. The three-state model is a special case of the four-state model ( $p = 1$ ).

The conditional likelihood of community  $i$  in (2) now generalizes to:

$$L_i(\phi, \lambda | v_i) \propto (\phi - \phi\mu\lambda e^{v_i})^{N_{1i}} \{\phi\mu\lambda e^{v_i} + (1-p)(1-\phi)\}^{N_{2i}} \{p(1-\phi)\}^{N_{3i}}, \quad (4)$$

and the marginal likelihood becomes

$$L(\phi, \lambda, \sigma^2) \propto \prod_{i=1}^M \int_{v_i} \left[ (\phi - \phi\mu\lambda e^{v_i})^{N_{1i}} \{\phi\mu\lambda e^{v_i} + (1-p)(1-\phi)\}^{N_{2i}} \{p(1-\phi)\}^{N_{3i}} \right. \\ \left. \times \frac{1}{\sigma} \exp\left(-\frac{v_i^2}{2\sigma^2}\right) \right] dv_i. \quad (5)$$

As before, the maximum likelihood estimators of  $(\phi, \lambda, \sigma^2)$  can be obtained by maximizing the approximated marginal likelihood. In the current formulation (4), we assume the same  $p$  across communities. This is often done in practice in part due to the lack of accurate information on community-specific  $p$ . However, if such information is available, these can be incorporated by modifying the conditional likelihood as  $L_i(\phi, \lambda | v_i) \propto (\phi - \phi\mu\lambda e^{v_i})^{N_{1i}} \{\phi\mu\lambda e^{v_i} + (1-p_i)(1-\phi)\}^{N_{2i}} \{p_i(1-\phi)\}^{N_{3i}}$ , and inference can proceed similarly.

**2.2.4 Implementation.** The proposed random effects model can be implemented in the SAS procedure NLMIXED (Kuss and McLerran, 2007; SAS Institute Inc., 2015). Adaptive Gaussian Quadrature method (Pinheiro and Bates, 1995) can be used to approximate the integral in (5). Sample codes for implementation are provided in the Web Appendix E.

Maximum likelihood estimators  $(\hat{\phi}, \hat{\lambda}, \hat{\sigma}^2)$  and their corresponding standard errors from the final Hessian matrix are displayed as a part of the standard output. The estimated incidence rate of the  $i$ th community is  $\hat{\lambda}e^{\hat{v}_i}$ , where  $\hat{v}_i$  is the empirical Bayes estimate of the random effects  $v_i$ . The overall incidence can be estimated as  $\hat{\lambda}^* = \sum_{i=1}^M \hat{\lambda}e^{\hat{v}_i}/M$ , and the variance of the overall incidence can be estimated as  $\widehat{\text{Var}}(\hat{\lambda}^*) = \left\{ \sum_{i=1}^M \widehat{\text{Var}}(\hat{\lambda}e^{\hat{v}_i}) + 2 \sum_{1 \leq i < j \leq M} \widehat{\text{Cov}}(\hat{\lambda}e^{\hat{v}_i}, \hat{\lambda}e^{\hat{v}_j}) \right\} / M^2$ , where  $\widehat{\text{Cov}}(\hat{\lambda}e^{\hat{v}_i}, \hat{\lambda}e^{\hat{v}_j}) = \widehat{\text{Var}}(\hat{\lambda})e^{\hat{\sigma}^2}$ . In addition, the between-community variability in true incidence rates can be estimated as  $\hat{\tau}^2 = \hat{\lambda}^2(e^{2\hat{\sigma}^2} - e^{\hat{\sigma}^2})$ .

**2.2.5 Extensions of the Random Effects Model.** The proposed random effects model can be extended to incorporate more complicated situations, including accounting for nested structures with more than two levels, incorporating random components in prevalence, and

estimating HIV incidence over time. Below we present the conditional likelihood corresponding to each case. Subsequent inferences are similar as described before.

When observations are obtained on subjects from communities, which are further nested within different sites, the conditional likelihood of community  $i$  can be written as

$$L_{i(j)}(\phi, \lambda | v_j, v_{i(j)}) \propto (\phi - \phi\mu\lambda e^{v_j+v_{i(j)}})^{N_{1i}} \{\phi\mu\lambda e^{v_j+v_{i(j)}} + (1-p)(1-\phi)\}^{N_{2i}} \{p(1-\phi)\}^{N_{3i}},$$

where  $v_j$  are “site effects”,  $j = 1, \dots, J$ , and  $v_{i(j)}$ ,  $i = 1, \dots, M_j$ , are the “community effects” that are nested within  $v_j$ . We assume  $v_j \sim N(0, \sigma_j^2)$ , and  $v_{i(j)} \sim N(0, \sigma_i^2)$ .

To incorporate varying prevalences across communities, the conditional likelihood becomes:

$$L_i(\phi, \lambda | v_i, w_i) \propto \{\phi e^{w_i}(1 - \mu\lambda e^{v_i})\}^{N_{1i}} \{\phi\mu\lambda e^{v_i+w_i} + (1-p)(1 - \phi e^{w_i})\}^{N_{2i}} \{p(1 - \phi e^{w_i})\}^{N_{3i}},$$

where  $\begin{pmatrix} v_i \\ w_i \end{pmatrix} \sim N\left(\begin{pmatrix} 0 \\ 0 \end{pmatrix}, \begin{pmatrix} \sigma_1^2 & \sigma_{12} \\ \sigma_{12} & \sigma_2^2 \end{pmatrix}\right)$ .

Finally, suppose that the dataset consists of a series of cross-sectional data over time, the conditional likelihood based on each round of data collection is:

$$L_i(\phi, \lambda | t_i) \propto (\phi - \phi\mu\lambda e^{t_i})^{N_{1i}} \{\phi\mu\lambda e^{t_i} + (1-p)(1-\phi)\}^{N_{2i}} \{p(1-\phi)\}^{N_{3i}},$$

where  $t_i \sim N(0, \sigma^2)$  is the round-specific random effects.

### 2.3 A Permutation Test

In this section, we propose a permutation test for the null hypothesis of no heterogeneity in incidence across communities, which is equivalent to testing the random effects variance component being 0,  $H_0 : \sigma^2 = 0$  versus  $H_1 : \sigma^2 > 0$ . Although it is natural to consider a likelihood ratio test (LRT) by comparing the likelihood maximized under  $H_0$  and that maximized without restrictions, in the current setting, the usual regularity condition that the null value is in the interior of the parameter space does not hold and the distribution of the LRT statistic under the null is no longer  $\chi^2$  and is hard to derive.

Let  $N_+ = \sum_{i=1}^M N_i$  denote the total number of observations and for each individual, let  $C_{ik}$  and  $S_{ik}$  denote the community membership index and the HIV infection state for  $i =$

$1, \dots, M$ , and  $k = 1, \dots, N_i$ . Let  $\mathbf{C} = (C_{11}, C_{12}, \dots, C_{M, N_M})^T$  and  $\mathbf{S} = (S_{11}, S_{12}, \dots, S_{M, N_M})^T$  denote the vector of community membership indices and the vector of HIV infection states, respectively. Let  $\mathbf{D}^{obs} = (\mathbf{C}^{obs}, \mathbf{S})$  denote the observed data matrix and  $\mathbf{D}^\ell = (\mathbf{C}^\ell, \mathbf{S})$  denote a permuted data matrix, where  $\mathbf{C}^\ell$  denotes a permutation of  $\mathbf{C}$ . When there is no heterogeneity in incidence and prevalence across communities, an individual's infection state confers no information about his/her community membership and he/she can be equally likely to come from any of the communities. We have  $C_{ik} \perp S_{ik}$ , where we use  $\perp$  to denote "is independent of". That is, the community indices can be viewed as random labels. This allows us to permute the vector of community indices, resulting in permuted datasets  $\mathbf{D}^\ell$  that are equally likely as the observed dataset. Under the null hypothesis  $H_0$ ,  $Pr\{\mathbf{D} = \mathbf{D}^\ell \mid \mathbf{C} \stackrel{p}{=} \mathbf{C}^{obs}, \mathbf{D} = (\mathbf{C}^{obs}, \mathbf{S})\} = 1/N_+!$ , where we use " $\stackrel{p}{=}$ " to denote "is a permutation of". That is, the observed data matrix,  $\mathbf{D}^{obs}$ , can be viewed as a randomly selected element from the set  $\prod$  consisting of  $N_+!$  matrices  $\mathbf{D}^\ell$ . It follows that if  $T = T(\mathbf{D})$  is a test statistic, for example, the usual LRT statistic  $T = -2(\ell_{\sigma^2=0}^{ML} - \ell_{\sigma^2 \geq 0}^{ML})$ , the observed value  $T(\mathbf{D}^{obs})$  can be viewed under  $H_0$  as a random sample of size 1 from the resulting permutation distribution of values  $\{T(\mathbf{D}^\ell) \mid \mathbf{D}^\ell \in \prod\}$ . This provides the basis for valid inferences about  $H_0$ .

The  $p$ -value is the proportion of test statistics calculated from permuted datasets that are more extreme than the observed test statistic under  $H_0$ . The total number of possible permutations is usually too large to enumerate all. In practice, we take a random sample of  $Q$  permutations, and approximate the  $p$ -value by  $1 + \sum_{q=1}^Q I\{T(\mathbf{D}^q) \geq T(\mathbf{D}^{obs})\} / (1 + Q)$ , where  $I(\cdot)$  is the indicator function. We add 1 to both the numerator and denominator to account for the fact that the observed dataset is also considered as one possible permutation.

#### 2.4 The Coefficient of Variation (CV)

The random effects model described in Section 2.2 provides a natural way to estimate the CV, defined as the between-cluster standard deviation  $\tau$  divided by the overall incidence

$\lambda^*$ . Under the Lognormal model, the CV for incidence can be expressed as  $CV = \tau/\lambda^* = \sqrt{\lambda^2(e^{2\sigma^2} - e^{\sigma^2})}/(\lambda e^{\sigma^2/2}) = \sqrt{e^{\sigma^2} - 1}$ . An estimate of the CV is given by  $\widehat{CV} = \sqrt{e^{\widehat{\sigma}^2} - 1}$ , and an estimate of the asymptotic variance of  $\widehat{CV}$  can be computed using the delta method,

$$\widehat{\text{Var}}(\widehat{CV}) = \frac{e^{2\widehat{\sigma}^2} \times \widehat{\text{Var}}(\widehat{\sigma}^2)}{4(e^{\widehat{\sigma}^2} - 1)}. \quad (6)$$

When the variability of incidences across communities is extremely small, an estimate of  $\text{Var}(\widehat{\sigma}^2)$  may not be available in the SAS output, so we are not able to estimate  $\widehat{\text{Var}}(\widehat{CV})$  using (6). Alternatively, we consider two commonly used bootstrap strategies for clustered data (Davison and Hinkley, 1997). One is the “cluster bootstrap”, where we randomly sample  $M$  communities with replacement, and for each selected community, all individuals within that community are included in the bootstrap sample (Field and Welsh, 2007). The other is the “subject bootstrap”, where for each community  $i$ ,  $i = 1, \dots, M$ , we draw a bootstrap sample of  $N_i$  subjects with replacement from the original sample (Roberts and Fan, 2004). When the number of communities is relatively small, but community sizes are relatively large, the “subject bootstrap” approach is usually preferred; for the settings involving a large number of communities with small sizes, “cluster bootstrap” may work better. In addition, “subject bootstrap” ensures that the total number of subjects remain the same as in the original sample; while “cluster bootstrap” may lead to varying total sample sizes and the difference in total sample sizes between a bootstrap sample and the original sample can be large depending on the heterogeneity in community sizes in the original sample.

### 3. Simulation Studies

#### 3.1 Fixed Effects Model vs. Random Effects Model

We conducted simulation studies to evaluate the performance of the proposed methods. Data was simulated from a multinomial distribution with state-specific prevalence  $p_{1i} = \phi - \phi\mu\lambda e^{v_i}$ ,  $p_{2i} = \phi\mu\lambda e^{v_i} + (1 - p)(1 - \phi)$ , and  $p_{3i} = p(1 - \phi)$ , where  $v_i \stackrel{i.i.d}{\sim} N(0, \sigma^2)$ , for  $i = 1, \dots, M$ . We assumed that the mean window period  $\mu$  is 0.5 years, and the prevalence of long-term

infection,  $1 - \phi = 0.25$  throughout. We considered combinations of  $\lambda = 0.01, 0.03, \text{ or } 0.05$ , and  $\sigma = 0.1, 0.3, \text{ or } 0.5$  when  $(M, N_i) = (30, 500)$ , or  $(20, 300)$ .

[Table 1 about here.]

Table 1 summarizes the simulation results when the false recency rate is 0 ( $p = 1$ ). Additional results for scenarios based on a false recency rate of 2%, varying community sizes, and sample size  $(M, N_i)$  settings similar to the Project ACCEPT are included in the Web Appendix C (S-Tables 1-3). Under all scenarios, the average of  $\hat{\lambda}^*$  are close to the truth for both models. When heterogeneity is small, both models achieve coverage rates that are close to the nominal level, with a slight increase in the width of the confidence interval from the random effects model. However, when heterogeneity is large, the actual coverage of the fixed effects model can be substantially lower than the nominal level due to severe underestimation of the standard errors. In contrast, confidence intervals based on the random effects model provide the empirical coverage at or close to the nominal level in all settings.

To examine the robustness to the mis-specification of the random effects distribution, we generated data under the same settings except that the random effects distribution was set to be Gamma or Weibull. The Lognormal model continues to provide accurate estimates of  $\lambda^*$  and leads to confidence intervals with coverage close to the nominal level (S-Table 4; Web Appendix C). We assessed the performance of the proposed random effects model in estimating community-specific incidences (Figure 2). Compared to the standard method, the results from the random effects model are associated with substantially smaller Mean Squared Errors (MSE) in all settings. The reduction in MSE is more substantial when there is a larger heterogeneity in community sizes (S-Figure 2; Web Appendix C).

[Figure 2 about here.]

### 3.2 Empirical Estimates of the Type I Error and Power for the Permutation Test

We evaluated the performance of the permutation test using similar data generation processes as before. We set  $\sigma^2 = 0$  to examine the type I error rate and varied  $\sigma^2$  from 0.2 to 0.8 to assess the power (S-Figure 3; Web Appendix C). In general, the empirical type I error estimates of the proposed permutation test are very close to the nominal level of 0.05. The power increases as  $\sigma$ ,  $M$ , or  $N_i$ , increases. For fixed  $\sigma$  and  $(M, N_i)$ , the power is higher for larger incidence.

### 3.3 The Coefficient of Variation

We next carried out simulation studies to assess the performance of the proposed CV estimator under the same settings as in Section 3.1 (Table 2).

[Table 2 about here.]

In line with the expectation that accurate estimation of CV requires a relatively large number of communities, the proposed estimator performs better in the settings where  $(M, N_i) = (30, 500)$  than those with  $(M, N_i) = (20, 300)$ : the average values of  $\widehat{CV}$  are in closer agreement with their true counterparts, and the standard errors of the estimates are smaller. In the settings we examined, the proposed estimator tends to underestimate the true CV. This underestimation becomes smaller as sample size increases, indicating that this might be a finite sample problem. When  $(M, N_i) = (20, 300)$  and the overall incidence  $\lambda^*$  is small ( $\lambda^* \approx 1\%$ ), we were not able to estimate standard errors of  $\widehat{CV}$  due to the instability of estimates for  $\sigma$ . For other settings, among the three methods of standard error estimation, the delta method generally performs well for large  $\tau$ . Both bootstrap approaches tend to underestimate the true variability of  $\widehat{CV}$  with the “subject bootstrap” method outperforming the “cluster bootstrap” approach. In contrast, when  $\tau$  is small, bootstrap methods provide more reliable standard error estimates than the delta method. When both  $\lambda^*$  and  $\tau$  are small,

the standard error estimates based on the “cluster bootstrap” approach are closer to the true sampling variability compared to those obtained from the “subject bootstrap” approach.

The confidence intervals in [Table 2](#) were obtained using the standard formulae  $\widehat{CV} \pm 1.96 \times \widehat{SE}(\widehat{CV})$ . We also evaluated other bootstrap confidence intervals: the bootstrap percentile confidence intervals, bootstrap pivotal confidence intervals, as well as one obtained by first constructing a confidence interval based on  $\log(\widehat{CV})$  and then exponentiating the endpoints. Results suggest that the standard symmetric confidence intervals perform the best with the coverage rates closest to the nominal level.

#### 4. Estimating Overall and Community-Specific Incidences in Project ACCEPT

We applied the proposed methods to the data from the NIMH Project ACCEPT (HPTN 043) (Coates et al., 2014). This is a phase III, community-randomized trial conducted in 34 communities at four sites in Africa (Soweto and Vulindlela, South Africa; Tanzania; and Zimbabwe). The primary endpoint, HIV incidence, was estimated from a cross-sectional sample via a multi-assay algorithm (MAA) to identify recent infections. The MAA applied in this study used four biomarkers: two serologic biomarkers (the BED-CEIA and an avidity assay) and two non-serologic biomarkers (CD4 cell count and HIV viral load), and the mean window period of this MAA was estimated to be 259 days ( $\mu = 0.71$  years) (Laeyendecker et al., 2013a). Each study sample was initially characterized based on the results of the two HIV rapid tests performed in-country, and those who had at least one reactive HIV rapid test results were further tested using MAA at the HPTN Network Laboratory. Each study participant’s HIV infection state can subsequently be classified as HIV-uninfected, recent infection, or long-term infection ([Table 3](#) (a)) (Laeyendecker et al., 2013b).

The estimated overall HIV incidence at each site and the associated uncertainty were calculated using the standard fixed effects model, the two-level random effects model (considered individuals as level-1 units, communities as level-2 clusters) stratified by the site, and



the three-level random effects model (considered individuals as level-1 units, communities as level-2 clusters, and sites as level-3 clusters) (Table 3 (b)). In addition, we performed the proposed permutation test and estimated the CV across multiple communities within each site based on the two-level random effects model. As in Laeyendecker et al. (2013b), we assumed a false recency rate of 0 ( $p = 1$ ) in the Project ACCEPT study population, since the MAA used in this study has been shown to reduce the misclassification rates substantially and to provide accurate incidence estimates (Laeyendecker et al., 2012, 2013a; Brookmeyer et al., 2013a,b; Eshleman et al., 2013). Additional sensitivity analysis results based on varying the false recency rate from 1% to 5% are presented in the Web Appendix D (S-Table 5).

[Table 3 about here.]

We observed a strong heterogeneity across community-specific incidences at site Soweto: The CV was substantial (estimated to be 0.742), and the  $p$ -value associated with the permutation test for  $\sigma^2 = 0$  indicates strong evidence against the null ( $p = 0.005$ ). The uncertainty of the overall incidence estimate was substantially underestimated using the standard fixed effects model. For the other three sites, both the CV and the permutation test  $p$ -value suggested modest heterogeneity. In these cases, the underestimation of uncertainty was minimal, and the standard model appeared to be adequate although using the random effects model only led to minimal efficiency loss. For all sites, the estimates of the overall incidence and its variability were almost identical regardless of the parametric assumptions about the random effects distribution (S-Table 6; Web Appendix D).

For the estimation of  $SE(\widehat{CV})$ , we presented estimates obtained using three different approaches, except for the site Tanzania, where  $SE(\widehat{CV})$  cannot be estimated using the delta method due to the extremely small variability in incidence rates across communities. Here we would prefer “subject bootstrap” over “cluster bootstrap” because the number of communities was relatively small, whereas the community sizes were large (28 out of 34

communities with  $N_i > 1000$ ). All methods led to the same conclusions, consistent with the permutation test results.

[Figure 3 about here.]

Community-specific incidence estimates are presented in [Figure 3](#). The community-specific incidence rates estimated from random effects models were associated with shorter confidence intervals compared to the standard method using information from that community only.

The point estimates for site-specific incidences from the two-level and the three-level models were virtually identical; but the three-level model led to larger standard errors for three of the four site-specific incidence estimates and larger or similar standard errors for community-specific incidence estimates in general. This may be due to the fact that the number of sites was relatively small and also that the additional assumption made about the distribution of site-specific incidences in the three-level model may be questionable given that the site Vulindlela appeared to be substantially different from the other three sites.

## 5. Discussion

In this paper, we develop a random effects model to estimate the overall and community-specific HIV incidence from a cross-sectional design. In the presence of heterogeneity in community-specific incidences, standard methods ignoring heterogeneity can lead to considerable biased inference because ignoring between-community heterogeneity leads to an underestimation of the uncertainty around the overall incidence estimate, while the proposed random effects model adequately accounts for this uncertainty. When the heterogeneity is negligible, the realized sample resembles an independent random sample, the standard method for estimating the overall incidence works well as expected; the random effects model yields similar results with minimal loss in efficiency because in such settings, the random effects model produces an estimate of between-cluster heterogeneity close to 0.

The random effects model leads to more efficient community-specific incidence estimates than applying the standard method to data from each community because it pools information from all communities. Assuming that the prevalence is constant across communities, we propose a permutation test for the null hypothesis of no heterogeneity in community-specific incidences. This test is easy to implement and its validity does not rely on parametric assumptions. The random effects model provides a natural way to estimate the CV, an essential parameter in the design and analysis of cluster randomized trials.

Our random effects model assumes a parametric distribution for the random effects. Both simulation studies and real data analysis suggest that the random effects model is robust to the mis-specification of the random effects distribution, when commonly-used distributions for positive-valued random variables such as Lognormal, Gamma and Weibull are specified because the shape of these distributions can be made similar by choice of parameter values.

The proposed method can account for the false recency rate (assumed to be known), multi-level clustering, or varying prevalences across communities. While the false recency rate does not affect the comparative conclusions between the random and the fixed effects model, reliable cross-sectional incidence estimation requires accurate knowledge of the false recency rate. The proposed model is also applicable to other settings with correlated data, e.g., a series of cross-sectional data are collected over time. Extensions to incorporate temporal dependence in incidence are possible through the modification of the random effects distribution: instead of assuming  $\lambda_i$ 's are independently distributed, we can consider the joint distribution of  $(\lambda_1, \dots, \lambda_M)$  with a more complex correlation structure (e.g., autoregressive or banded Toeplitz).

## 6. Supplementary Materials

Web Appendices referenced in Sections 1-4 and sample codes for implementing the proposed method are available with this article at the *Biometrics* website on Wiley Online Library.

## ACKNOWLEDGEMENTS

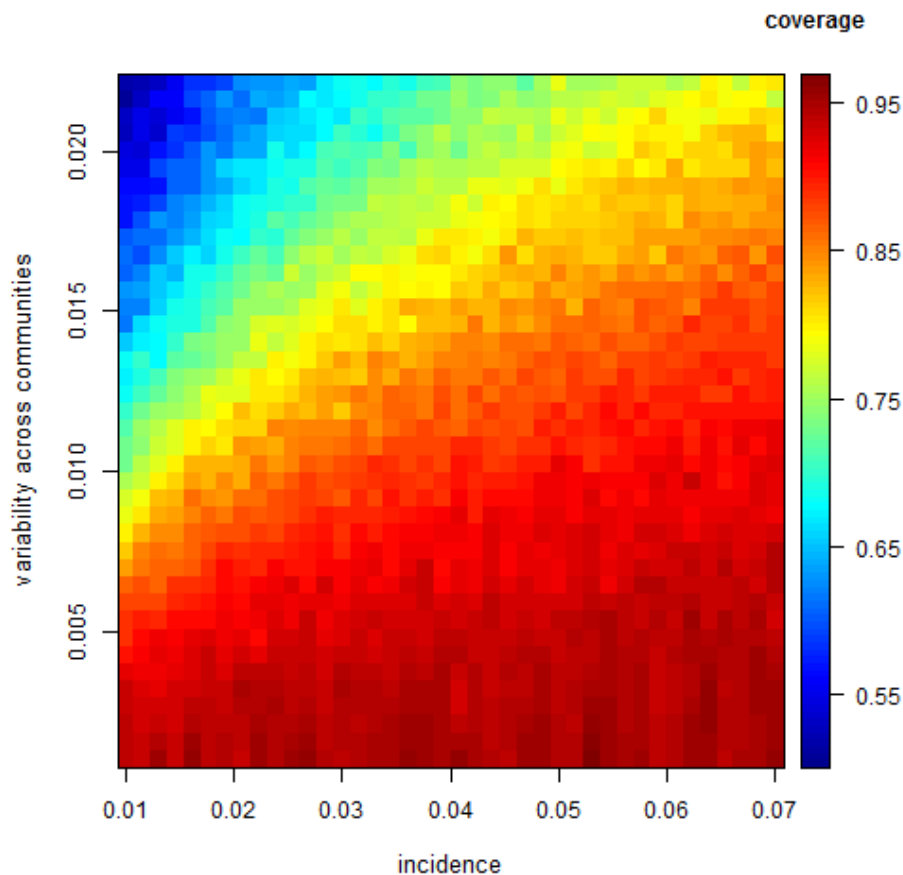
We thank the editor, associate editor and two referees for their insightful comments and suggestions, which led to an improved paper. This research was supported by grant R37 AI51164 from the U.S. National Institute of Allergy and Infectious Diseases (NIAID). In addition, this work was supported by HPTN Protocol 043 through contracts U01AI068613/UM1A068613 (HPTN Network Laboratory - Susan Eshleman, PI); U01AI068617/UM1AI068617 (SCHARP - Deborah Donnell, PI); U01AI068619/UM1AI068619 (HIV Prevention Trials Network - Sten Vermund/Wafaa El - Sadr, PIs); R01 AI095068 NIAID (Susan Eshleman, PI); and by the Office of AIDS Research, and the Division of Intramural Research, NIAID.

## REFERENCES

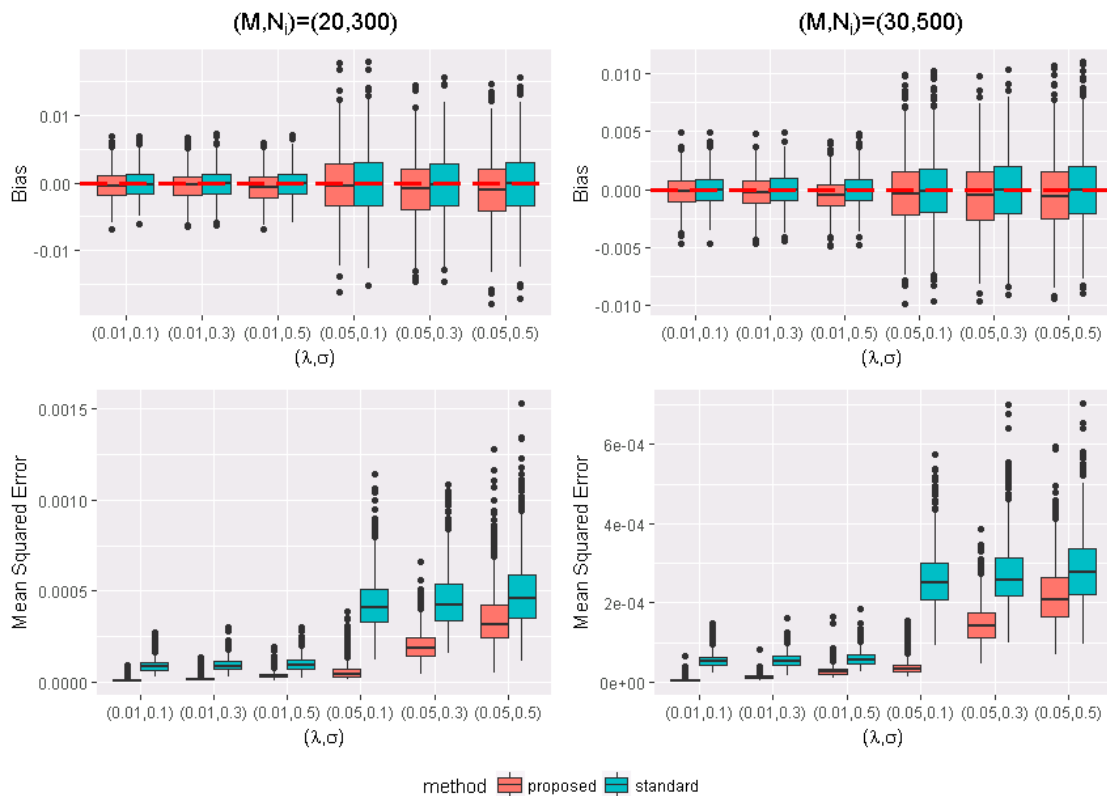
- Balasubramanian, R. and Lagakos, S. W. (2010). Estimating HIV incidence based on combined prevalence testing. *Biometrics* **66**, 1–10.
- BCPP (2013). Botswana Combination Prevention Project (BCPP) - NCT01965470. <http://clinicaltrials.gov/ct2/show/record/NCT01965470>. Accessed: 2015-08-28.
- Brookmeyer, R., Konikoff, J., Laeyendecker, O., and Eshleman, S. H. (2013a). Estimation of HIV incidence using multiple biomarkers. *American Journal of Epidemiology* **177**, 264–272.
- Brookmeyer, R., Laeyendecker, O., Donnell, D., and Eshleman, S. H. (2013b). Cross-sectional HIV incidence estimation in HIV prevention research. *Journal of Acquired Immune Deficiency Syndrome* **63**, S233–S239.
- Brookmeyer, R. and Quinn, T. C. (1995). Estimation of current human immunodeficiency virus incidence rates from a cross-sectional survey using early diagnostic tests. *American Journal of Epidemiology* **141**, 166–172.
- Claggett, B., Lagakos, S. W., and Wang, R. (2012). Augmented cross-sectional studies with

- abbreviated follow-up for estimating HIV incidence. *Biometrics* **68**, 62–74.
- Coates, T. J., Kulich, M., Celentano, D. D., Zelaya, C. E., Chariyalertsak, S., Chingono, A., et al. (2014). Effect of community-based voluntary counselling and testing on HIV incidence and social and behavioural outcomes (NIMH Project Accept; HPTN 043): a cluster-randomised trial. *The Lancet Global Health* **2**, e267–e277.
- Davison, A. C. and Hinkley, D. V. (1997). *Bootstrap Methods and their Application*. Cambridge: Cambridge University Press.
- Eshleman, S. H., Hughes, J. P., Laeyendecker, O., Wang, J., Brookmeyer, R., Johnson-Lewis, L., et al. (2013). Use of a multifaceted approach to analyze HIV incidence in a cohort study of women in the United States: HIV Prevention Trials Network 064 Study. *The Journal of Infectious Disease* **207**, 223–231.
- Field, C. A. and Welsh, A. H. (2007). Bootstrapping clustered data. *Journal of the Royal Statistical Society Series B* **69**, 369–390.
- Hallett, T. B. (2011). Estimating the HIV incidence rate—recent and future developments. *Current Opinion in HIV and AIDS* **6**, 102–107.
- Kuss, O. and McLerran, D. (2007). A note on the estimation of the multinomial logistic model with correlated responses in SAS. *Computer Methods and Programs in Biomedicine* **87**, 262–269.
- Laeyendecker, O., Brookmeyer, R., Cousins, M. M., Mullis, C. E., Konikoff, J., Donnell, D., et al. (2012). HIV incidence determination in the United States: A multi-assay approach. *The Journal of Infectious Diseases* **207**, 232–239.
- Laeyendecker, O., Kulich, M., Donnell, D., Komarek, A., Omelka, M., Mullis, C. E., et al. (2013a). Development of methods for cross-sectional HIV incidence estimation in a large, community randomized trial. *PLoS ONE* **8**, e78818.
- Laeyendecker, O., Piwowar-Manning, E., Fiamma, A., Kulich, M., Donnell, D., Bassuk, D.,

- et al. (2013b). Estimation of HIV incidence in a large, community-based, randomized clinical trial: NIMH Project Accept (HIV Prevention Trials Network 043). *PLoS ONE* **8**, e68349.
- Moyo, S., LeCuyer, T., Wang, R., Gaseitsiwe, S., Weng, J., Musonda, R., et al. (2014). Evaluation of the false recent classification rates of multiassay algorithms in estimating HIV type 1 subtype C incidence. *AIDS Research and Human Retroviruses* **30**, 29–36.
- Novitsky, V., Wang, R., Kebaabetswe, L., Greenwald, J., Rossenkhan, R., Moyo, S., et al. (2009). Better control of early viral replication is associated with slower rate of elicited antiviral antibodies in the detuned EIA during primary HIV-1C infection. *Journal of Acquired Immune Deficiency Syndromes* **52**, 265–272.
- Pinheiro, J. C. and Bates, D. M. (1995). Approximations to the log-likelihood function in the nonlinear mixed-effects model. *Journal of Computational and Graphical Statistics* **4**, 12–35.
- Roberts, J. and Fan, X. (2004). Bootstrapping within the multilevel/hierarchical linear modeling framework: A primer for use with SAS and SPLUS. *Multiple Linear Regression Viewpoints* **30**, 23–33.
- Rutterford, C., Copas, A., and Eldridge, S. (2015). Methods for sample size determination in cluster randomized trials. *International Journal of Epidemiology* **44**, 1051–1067.
- SAS Institute Inc. (2015). *SAS/STAT 14.1 User's Guide*. SAS Institute Inc, Cary, NC.
- Wang, R. and Lagakos, S. W. (2009). On the use of adjusted cross-sectional estimators of HIV incidence. *Journal of Acquired Immune Deficiency Syndromes* **52**, 538–547.
- Wang, R. and Lagakos, S. W. (2010). Augmented cross-sectional prevalence testing for estimating HIV incidence. *Biometrics* **66**, 864–874.

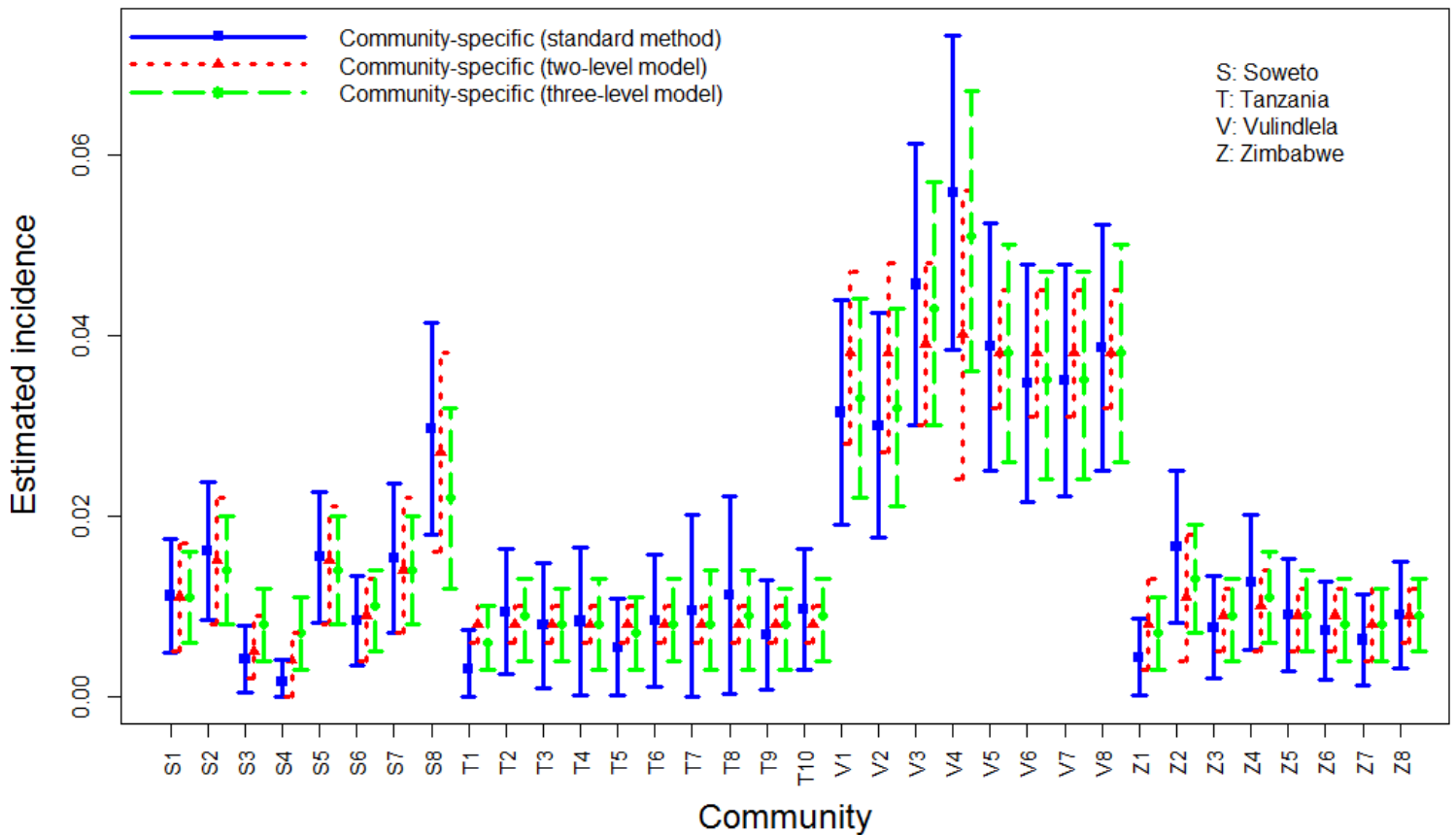


**Figure 1.** Contour plot of the actual coverage of 95% confidence intervals for the overall incidence across 30 communities (500 individuals in each community) ignoring heterogeneity, for varying incidences (horizontal axis) and heterogeneity across communities represented as the standard deviation across community-specific incidences (vertical axis). Different colors indicate whether or not the nominal coverage is attained with dark red representing the situations where the nominal coverage is attained. Results are based on 1000 experiments. Additional results on the generation of this figure are included in the Web Appendix A.

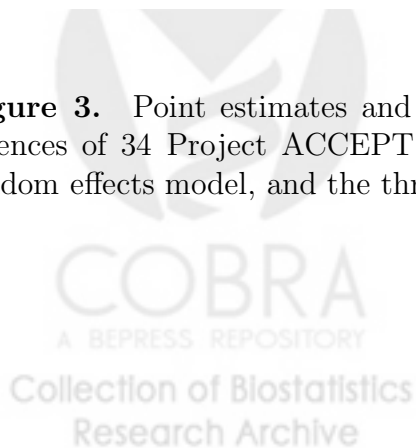


**Figure 2.** Boxplots for the bias  $(\sum_{i=1}^M (\hat{\lambda}_i - \lambda_i)/M)$  and the Mean Squared Errors (MSE)  $(\sum_{i=1}^M (\hat{\lambda}_i - \lambda_i)^2/M)$  when estimating the community-specific incidences  $\lambda_i$  using the standard method and the proposed method, based on 1000 simulations. Data generated under Lognormal random effects model with  $\mu = 0.5$  years,  $p = 1$ , and  $\phi = 0.75$ .





**Figure 3.** Point estimates and 95% confidence intervals of community-specific HIV incidences of 34 Project ACCEPT communities, using the standard method, the two-level random effects model, and the three-level random effects model.



**Table 1**

Simulation results for the standard fixed effects model vs. the proposed random effects model, with  $\mu = 0.5$  years,  $p = 1$ ,  $\phi = 0.75$ , and varying incidences. Estimation based on assuming the random effects follow the Lognormal distribution.  $\lambda^*$  and  $\tau$  denote the overall incidence and the between-community standard deviation.  $\hat{E}(\hat{\lambda}^*)$  and  $\widehat{SE}(\hat{\lambda}^*)$  denote average and standard error of estimates from 1000 simulations.  $\hat{E}(\hat{s})$  denotes average of likelihood-based estimates of the standard error from the 1000 experiments. Coverage denotes the proportion of simulations in which the true  $\lambda^*$  is contained in the nominal 95% confidence interval (CI), and width refers to the mean width of the nominal 95% CI.

$\lambda$ (%)	$\sigma$	$\lambda^*$ (%)	$\tau$ (%)	Standard Method (Fixed Effects Model)					Proposed Method (Random Effects Model)				
				$\hat{E}(\hat{\lambda}^*)$ (%)	$\widehat{SE}(\hat{\lambda}^*)$ (%)	$\hat{E}(\hat{s})$ (%)	Coverage (%)	Width (%)	$\hat{E}(\hat{\lambda}^*)$ (%)	$\widehat{SE}(\hat{\lambda}^*)$ (%)	$\hat{E}(\hat{s})$ (%)	Coverage (%)	Width (%)
$M = 30, N_i = 500$													
1	0.1	1.0	0.1	1.0	0.13	0.13	95.6	0.5	1.0	0.14	0.15	95.2	0.6
1	0.3	1.1	0.3	1.0	0.15	0.14	91.9	0.5	1.0	0.15	0.17	94.6	0.7
1	0.5	1.1	0.6	1.1	0.18	0.14	88.0	0.6	1.1	0.18	0.21	96.5	0.8
3	0.1	3.0	0.3	3.0	0.23	0.23	95.4	0.9	3.0	0.24	0.26	95.9	1.0
3	0.3	3.1	1.0	3.1	0.30	0.23	87.7	0.9	3.1	0.28	0.32	96.1	1.3
3	0.5	3.4	1.8	3.4	0.43	0.24	72.5	1.0	3.4	0.41	0.44	95.4	1.7
5	0.1	5.0	0.5	5.0	0.30	0.29	94.1	1.2	5.0	0.31	0.33	95.8	1.3
5	0.3	5.2	1.6	5.2	0.42	0.30	83.5	1.2	5.2	0.40	0.46	95.3	1.8
5	0.5	5.7	3.0	5.7	0.64	0.31	64.6	1.2	5.6	0.62	0.66	94.4	2.6
$M = 20, N_i = 300$													
1	0.1	1.0	0.1	1.0	0.21	0.21	92.9	0.8	1.0	0.21	0.25	96.4	1.0
1	0.3	1.1	0.3	1.1	0.22	0.21	92.7	0.8	1.0	0.22	0.27	95.6	1.1
1	0.5	1.1	0.6	1.1	0.27	0.22	87.8	0.9	1.1	0.27	0.31	93.7	1.2
3	0.1	3.0	0.3	3.0	0.37	0.36	93.7	1.4	3.0	0.38	0.41	95.0	1.6
3	0.3	3.1	1.0	3.2	0.44	0.37	92.6	1.5	3.1	0.43	0.47	95.5	1.9
3	0.5	3.4	1.8	3.4	0.56	0.38	83.6	1.5	3.3	0.56	0.62	94.3	2.4
5	0.1	5.0	0.5	5.0	0.46	0.47	95.0	1.8	5.0	0.46	0.53	96.9	2.1
5	0.3	5.2	1.6	5.2	0.61	0.47	87.4	1.9	5.2	0.61	0.65	94.5	2.6
5	0.5	5.7	3.0	5.6	0.84	0.49	74.4	1.9	5.6	0.85	0.89	93.7	3.5

**Table 2**

Simulation results for the coefficient of variation (CV) estimator based on the Lognormal random effects model, with  $\mu = 0.5$  years,  $p = 1$ ,  $\phi = 0.75$ , and varying incidences.  $\lambda^*$  and  $\tau$  denote the overall incidence and the between-community standard deviation.  $\widehat{E}(\widehat{CV})$  and  $\widehat{SE}(\widehat{CV})$  denote average and standard error of estimates from 1000 simulated studies.  $\widehat{E}(\widehat{s})$  refers to the mean of estimates of the standard error from the 1000 experiments, and coverage denotes the proportion of simulations in which the true CV is contained in the nominal 95% confidence interval. Non-NA represents the proportion of experiments in which the standard error of CV estimates can be calculated using the delta method. Estimates of the standard errors and the actual coverage of the 95% confidence intervals were obtained using three different approaches: the “cluster bootstrap”, the “subject bootstrap”, and the delta method. For the two bootstrap methods, we took 500 bootstrap replications.

$\lambda$ (%)	$\sigma$	$\lambda^*$ (%)	$\tau$ (%)	CV	$\widehat{E}(\widehat{CV})$	$\widehat{SE}(\widehat{CV})$	Cluster Bootstrap		Subject Bootstrap		Delta Method			
							$\widehat{E}(\widehat{s})$	Coverage (%)	$\widehat{E}(\widehat{s})$	Coverage (%)	$\widehat{E}(\widehat{s})$	Coverage (%)	Non-NA (%)	
$M = 30, N_i = 500$														
1	0.1	1.0	0.1	0.100	0.137	0.184	0.212	98.7	0.275	100.0	0.427	95.8	42.8	
1	0.3	1.1	0.3	0.307	0.248	0.237	0.224	96.4	0.287	100.0	0.355	100.0	63.9	
1	0.5	1.1	0.6	0.533	0.475	0.253	0.256	87.8	0.324	91.1	0.278	100.0	89.9	
3	0.1	3.0	0.3	0.100	0.091	0.114	0.103	93.7	0.117	96.5	0.195	94.0	46.9	
3	0.3	3.1	1.0	0.307	0.271	0.133	0.104	84.4	0.116	88.2	0.151	99.7	91.3	
3	0.5	3.4	1.8	0.533	0.495	0.136	0.119	88.6	0.125	91.2	0.136	94.0	99.6	
5	0.1	5.0	0.5	0.100	0.085	0.090	0.076	92.7	0.085	96.6	0.152	96.3	56.2	
5	0.3	5.2	1.6	0.307	0.280	0.101	0.075	85.3	0.082	89.6	0.095	99.6	97.0	
5	0.5	5.7	3.0	0.533	0.505	0.115	0.094	84.7	0.090	86.3	0.112	91.4	100.0	
$M = 20, N_i = 300$														
1	0.1	1.0	0.1	0.100	0.192	0.304	—	—	—	—	—	—	—	37.8
1	0.3	1.1	0.3	0.307	0.248	0.346	—	—	—	—	—	—	—	46.4
1	0.5	1.1	0.6	0.533	0.444	0.403	—	—	—	—	—	—	—	67.8
3	0.1	3.0	0.3	0.100	0.125	0.162	0.169	96.2	0.214	99.7	0.349	96.5	46.1	
3	0.3	3.1	1.0	0.307	0.241	0.198	0.182	81.9	0.219	99.8	0.267	100.0	70.9	
3	0.5	3.4	1.8	0.533	0.464	0.235	0.217	86.2	0.237	90.0	0.229	100.0	93.0	
5	0.1	5.0	0.5	0.100	0.096	0.123	0.120	94.4	0.149	98.7	0.253	97.0	46.1	
5	0.3	5.2	1.6	0.307	0.250	0.163	0.131	79.7	0.148	85.7	0.181	99.8	82.7	
5	0.5	5.7	3.0	0.533	0.488	0.173	0.152	87.6	0.159	90.9	0.171	96.1	98.8	

**Table 3**  
*Sample classification and analysis results of Project ACCEPT cross-sectional incidence data at each African site.*

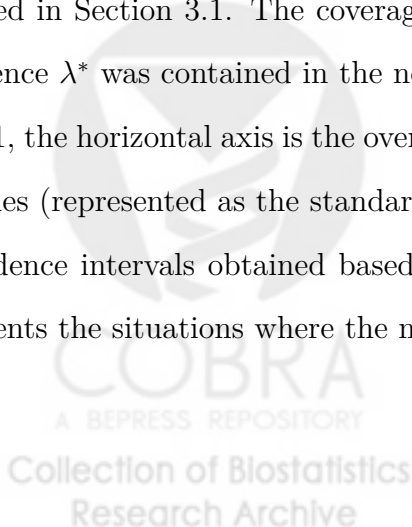
	Soweto	Tanzania	Vulindlela	Zimbabwe
<b>(a). Sample classification.</b>				
$M$ : Number of communities	8	10	8	8
$N_1$ : Uninfected	11962	8505	8197	10348
$N_2$ : Recent infections	101	47	230	67
$N_3$ : Long-term infections	1547	479	3400	1461
<b>(b). Analysis results.</b>				
<b>Standard fixed effects model</b>				
$\hat{\lambda}^*$ : Overall incidence by the standard model (%)	1.2	0.8	3.8	0.9
$\widehat{SE}(\hat{\lambda}^*)$ : Variability of the standard incidence estimate (%)	0.12	0.11	0.25	0.11
<b>Two-level random effects model</b>				
$\hat{\lambda}^*$ : Overall incidence by the two-level model (%)	1.3	0.8	3.8	0.9
$\widehat{SE}(\hat{\lambda}^*)$ : Variability of the two-level incidence estimate (%)	0.34	0.11	0.30	0.14
Permutation test $p$ -value	0.005	>0.99	0.27	0.17
$\widehat{CV}$ : Coefficient of variation	0.7420	0.0001	0.0631	0.1872
$\widehat{SE}_{\text{delta}}(\widehat{CV})$ : Variability of $\widehat{CV}$ by delta method	0.310	—	0.157	0.207
$\widehat{SE}_{\text{cluster}}(\widehat{CV})$ : Variability of $\widehat{CV}$ by cluster bootstrap	0.282	0.076	0.072	0.142
$\widehat{SE}_{\text{subject}}(\widehat{CV})$ : Variability of $\widehat{CV}$ by subject bootstrap	0.237	0.208	0.098	0.180
<b>Three-level random effects model</b>				
$\hat{\lambda}^*$ : Overall incidence by the three-level model (%)	1.2	0.8	3.8	0.9
$\widehat{SE}(\hat{\lambda}^*)$ : Variability of the three-level incidence estimate (%)	0.20	0.15	0.53	0.17

# Web-based Supplementary Materials for “Cross-sectional HIV Incidence Estimation Accounting for Heterogeneity Across Communities”

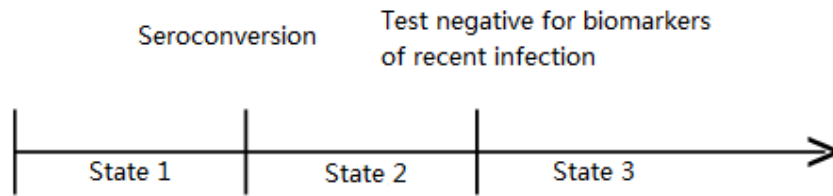
Yuejia Xu, Oliver Laeyendecker, and Rui Wang\*  
\*rwang@hsph.harvard.edu

## Web Appendix A: Details on the Generation of Figure 1

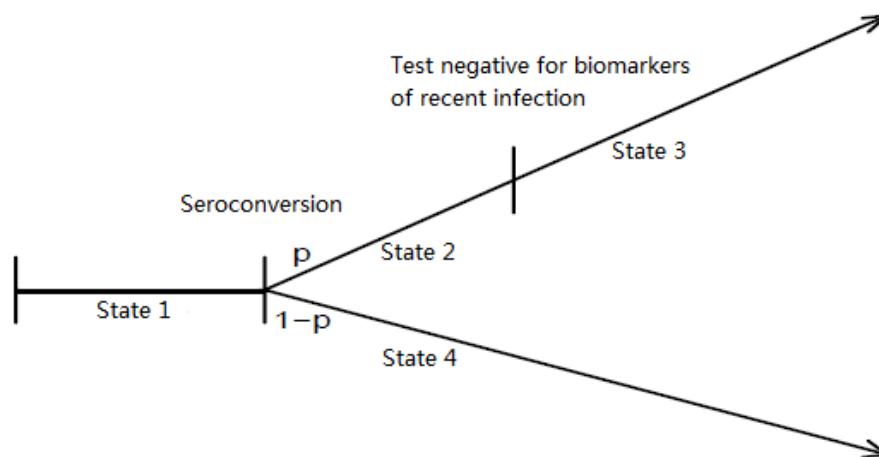
In this section, we clarify details on how Figure 1 (contour plot) in the main paper was produced. We assumed there were 30 communities ( $M = 30$ ), and each community had 500 individuals ( $N_i = 500$ ). The false recency rate was set to 0 ( $p = 1$ ), and the mean window period  $\mu$  was 0.5 years. We calculated the coverage of 95% confidence intervals for the overall incidence across 30 communities under the fixed effects model (ignoring heterogeneity), for varying incidences and heterogeneity across communities. Specifically, we considered a grid of incidence values ( $\lambda^*$ ) from 1% to 7% at an increment of  $\Delta\lambda^* = 0.1\%$ , and the variability of overall incidence (represented as the variance:  $\tau^2$ ) from 0.0001% to 0.05% at an increment of  $\Delta\tau^2 = 0.001\%$ . For each combination of incidence and its variability, we calculated  $\lambda$  and  $\sigma$  based on  $\lambda^* = \lambda e^{\sigma^2/2}$  and  $\tau^2 = \lambda^2(e^{2\sigma^2} - e^{\sigma^2})$ , and generated data under the corresponding  $\lambda$  and  $\sigma$  in the same way as described in Section 3.1. The coverage was calculated as the proportion of simulations in which the true incidence  $\lambda^*$  was contained in the nominal 95% confidence interval, based on 1000 simulation runs. In Figure 1, the horizontal axis is the overall incidence  $\lambda^*$ , and the vertical axis is the heterogeneity across communities (represented as the standard deviation:  $\tau$ ). Different colors indicate the actual coverage of 95% confidence intervals obtained based on the fixed effects model ignoring heterogeneity, where dark red represents the situations where the nominal coverage is achieved.



## Web Appendix B: Graphical Representation of Models



(a) Three-state longitudinal model of HIV seroconversion and subsequent reactivity to biomarkers of recent infection.



(b) Four-state longitudinal model in which a proportion,  $1 - p$ , of infected persons would be classified as “recently infected” indefinitely.

**S-Figure 1.** Graphical representation of models.

## Web Appendix C: Additional Simulation Results

We present additional simulation results in this section. In S-Table 1, we show the performance of the proposed random effects model when  $p = 0.98$  (false recency rate = 2%). Simulation settings were the same as those described in Section 3.1 of the main paper, except that the false recency rate was no longer assumed to be 0.

S-Table 2 corresponds to the setting with varying community sizes. We considered the case where  $\lambda = 0.03$ ,  $\sigma = 0.5$  (CV = 0.533),  $\mu = 0.5$  years,  $p = 1$ , and  $\phi = 0.75$ . Instead of assuming equal community

**S-Table 1**

Simulation results for the standard fixed effects model vs. the proposed random effects model, with  $\mu = 0.5$  years,  $p = 0.98$ ,  $\phi = 0.75$ , and varying incidences. Estimation based on assuming the random effects follow the Lognormal distribution.  $\lambda^*$  and  $\tau$  denote the overall incidence and the between-community standard deviation.  $\widehat{E}(\widehat{\lambda}^*)$  and  $\widehat{SE}(\widehat{\lambda}^*)$  denote average and standard error of estimates from 1000 simulations.  $\widehat{E}(\widehat{s})$  denotes average of likelihood-based estimates of the standard error from the 1000 experiments. Coverage denotes the proportion of simulations in which the true  $\lambda^*$  is contained in the nominal 95% confidence interval (CI), and width refers to the mean width of the nominal 95% CI.

$\lambda$ (%)	$\sigma$	$\lambda^*$ (%)	$\tau$ (%)	Standard Method (Fixed Effects Model)			Proposed Method (Random Effects Model)						
				$\widehat{E}(\widehat{\lambda}^*)$ (%)	$\widehat{SE}(\widehat{\lambda}^*)$ (%)	Coverage (%)	Width (%)	$\widehat{E}(\widehat{\lambda}^*)$ (%)	$\widehat{SE}(\widehat{\lambda}^*)$ (%)	Coverage (%)	Width (%)		
$M = 30, N_i = 500$													
1	0.1	1.0	0.1	1.0	0.19	0.20	96.6	0.8	1.0	0.20	0.26	98.9	1.0
1	0.3	1.0	0.3	1.1	0.20	0.21	96.4	0.8	1.0	0.22	0.27	98.6	1.1
1	0.5	1.1	0.6	1.1	0.23	0.21	92.9	0.8	1.1	0.24	0.30	99.0	1.2
3	0.1	3.0	0.3	3.0	0.28	0.28	94.2	1.1	3.0	0.28	0.31	96.5	1.2
3	0.3	3.1	1.0	3.1	0.33	0.28	90.8	1.1	3.1	0.33	0.37	96.6	1.4
3	0.5	3.4	1.8	3.4	0.45	0.29	79.1	1.1	3.4	0.45	0.49	95.2	1.9
5	0.1	5.0	0.5	5.0	0.34	0.33	93.6	1.3	5.0	0.34	0.37	96.2	1.5
5	0.3	5.2	1.6	5.2	0.46	0.34	85.7	1.3	5.2	0.45	0.49	95.2	1.9
5	0.5	5.7	3.0	5.7	0.65	0.35	69.3	1.4	5.7	0.66	0.70	95.4	2.7
$M = 20, N_i = 300$													
1	0.1	1.0	0.1	1.0	0.29	0.32	98.5	1.3	1.0	0.30	0.41	99.1	1.6
1	0.3	1.0	0.3	1.1	0.30	0.33	97.3	1.3	1.0	0.31	0.42	97.9	1.6
1	0.5	1.1	0.6	1.2	0.33	0.33	95.8	1.3	1.1	0.34	0.46	97.9	1.8
3	0.1	3.0	0.3	3.0	0.44	0.44	94.7	1.7	3.0	0.44	0.50	96.8	1.9
3	0.3	3.1	1.0	3.1	0.48	0.44	93.5	1.7	3.1	0.48	0.55	96.4	2.2
3	0.5	3.4	1.8	3.4	0.62	0.45	85.3	1.8	3.3	0.62	0.68	94.3	2.7
5	0.1	5.0	0.5	5.0	0.54	0.52	94.1	2.1	5.0	0.54	0.59	95.4	2.3
5	0.3	5.2	1.6	5.2	0.63	0.53	89.5	2.1	5.1	0.63	0.70	95.4	2.8
5	0.5	5.7	3.0	5.6	0.87	0.55	77.9	2.1	5.6	0.88	0.94	93.5	3.7

**S-Table 2**

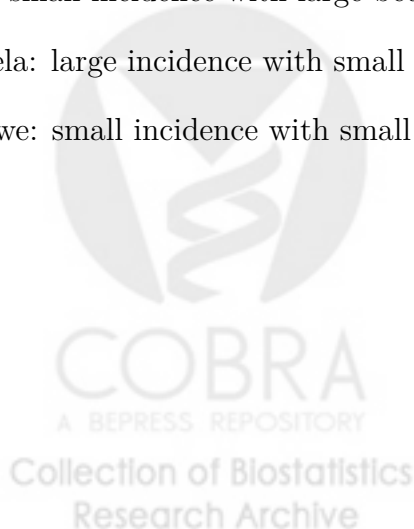
Simulation results for the standard fixed effects model vs. the proposed random effects model under settings with unequal cluster sizes generated from the discrete uniform distribution  $\mathcal{U}\{a, b\}$ , with  $\lambda = 0.03$ ,  $\sigma = 0.5$ ,  $\mu = 0.5$  years,  $p = 1$ , and  $\phi = 0.75$ . Estimation based on assuming the random effects follow the Lognormal distribution.  $\widehat{E}(\widehat{\lambda}^*)$  and  $\widehat{SE}(\widehat{\lambda}^*)$  denote average and standard error of estimates from 1000 simulations.  $\widehat{E}(\widehat{s})$  denotes average of likelihood-based estimates of the standard error from the 1000 experiments. Coverage denotes the proportion of simulations in which the true  $\lambda^*$  is contained in the nominal 95% confidence interval (CI), and width refers to the mean width of the nominal 95% CI.

$M$	$N_i$	Fixed Effects Model					Random Effects Model				
		$\widehat{E}(\widehat{\lambda}^*)$ (%)	$\widehat{SE}(\widehat{\lambda}^*)$ (%)	$\widehat{E}(\widehat{s})$ (%)	Coverage (%)	Width (%)	$\widehat{E}(\widehat{\lambda}^*)$ (%)	$\widehat{SE}(\widehat{\lambda}^*)$ (%)	$\widehat{E}(\widehat{s})$ (%)	Coverage (%)	Width (%)
$\lambda^* = 3.4\%$ , $\tau = 1.8\%$											
10	$\mathcal{U}\{500, 2000\}$	3.4	0.68	0.27	57.2	1.1	3.4	0.65	0.62	91.4	2.4
20	$\mathcal{U}\{200, 400\}$	3.4	0.57	0.38	80.4	1.5	3.3	0.56	0.61	93.9	2.4
20	$\mathcal{U}\{50, 550\}$	3.4	0.57	0.37	79.7	1.5	3.3	0.54	0.61	94.4	2.4
30	$\mathcal{U}\{100, 900\}$	3.4	0.44	0.25	72.2	1.0	3.3	0.43	0.47	94.3	1.8

sizes,  $N_i$  were generated from the discrete uniform distribution  $\mathcal{U}\{a, b\}$ . We examined the following scenarios of  $(M, N_i)$ :  $(10, \mathcal{U}\{500, 2000\})$ ,  $(20, \mathcal{U}\{200, 400\})$ ,  $(20, \mathcal{U}\{50, 550\})$ , and  $(30, \mathcal{U}\{100, 900\})$ .

We also performed simulation studies under sample size settings similar to the Project ACCEPT (S-Table 3). Specifically, we considered  $(M, N_i) = (10, 1500)$ , corresponding to settings with fewer communities and larger number of individuals per community. We simulated data under three different scenarios, mimicking situations of Soweto, Vulindlela, and Zimbabwe, respectively (in terms of the overall incidence,  $\lambda^*$ , and heterogeneity of incidence,  $\tau$ ).

1. Soweto: small incidence with large between-community heterogeneity.
2. Vulindlela: large incidence with small between-community heterogeneity.
3. Zimbabwe: small incidence with small between-community heterogeneity.





**S-Table 3**

Simulation results for the standard fixed effects model vs. the proposed random effects model, with  $(M, N_i) = (10, 1500)$ ,  $\mu = 0.71$  years,  $p = 1$ , and varying incidences. Estimation based on assuming the random effects follow the Lognormal distribution.  $\lambda^*$  and  $\tau$  denote the overall incidence and the between-community standard deviation.  $\widehat{E}(\widehat{\lambda}^*)$  and  $\widehat{SE}(\widehat{\lambda}^*)$  denote average and standard error of estimates from 1000 simulations.  $\widehat{E}(\widehat{s})$  denotes average of likelihood-based estimates of the standard error from the 1000 experiments. Coverage denotes the proportion of simulations in which the true  $\lambda^*$  is contained in the nominal 95% confidence interval (CI), and width refers to the mean width of the nominal 95% CI.

		Fixed Effects Model					Random Effects Model				
$\lambda^*$	$\tau$	$\widehat{E}(\widehat{\lambda}^*)$	$\widehat{SE}(\widehat{\lambda}^*)$	$\widehat{E}(\widehat{s})$	Coverage	Width	$\widehat{E}(\widehat{\lambda}^*)$	$\widehat{SE}(\widehat{\lambda}^*)$	$\widehat{E}(\widehat{s})$	Coverage	Width
(%)	(%)	(%)	(%)	(%)	(%)	(%)	(%)	(%)	(%)	(%)	(%)
$M = 10, N_i = 1500$											
Setting: Soweto											
1.2	0.92	1.3	0.31	0.11	55.6	0.4	1.2	0.31	0.29	90.4	1.1
Setting: Vulindlela											
4.0	0.24	4.0	0.24	0.23	93.2	0.9	4.0	0.24	0.26	95.1	1.0
Setting: Zimbabwe											
0.9	0.18	0.9	0.11	0.10	91.5	0.4	0.9	0.11	0.12	94.6	0.5

We investigated the robustness of the proposed model to the mis-specification of the random effects distribution. We performed simulation studies under the same settings as described in Section 3.1 of the main paper, except that the random effects distribution was assumed to be Gamma or Weibull. Corresponding results are summarized in S-Table 4.

In addition, we assessed the performance of the proposed method in estimating community-specific incidences and compared with the standard method under the settings where community sizes are unequal. Community sizes  $N_i$  were generated from the discrete uniform distribution  $\mathcal{U}\{a, b\}$ . We examined  $(M, N_i) = (20, \mathcal{U}\{50, 550\})$  and  $(M, N_i) = (30, \mathcal{U}\{100, 900\})$ . Boxplots summarizing bias and Mean Squared Errors (MSE) over 1000 simulations are displayed in S-Figure 2.

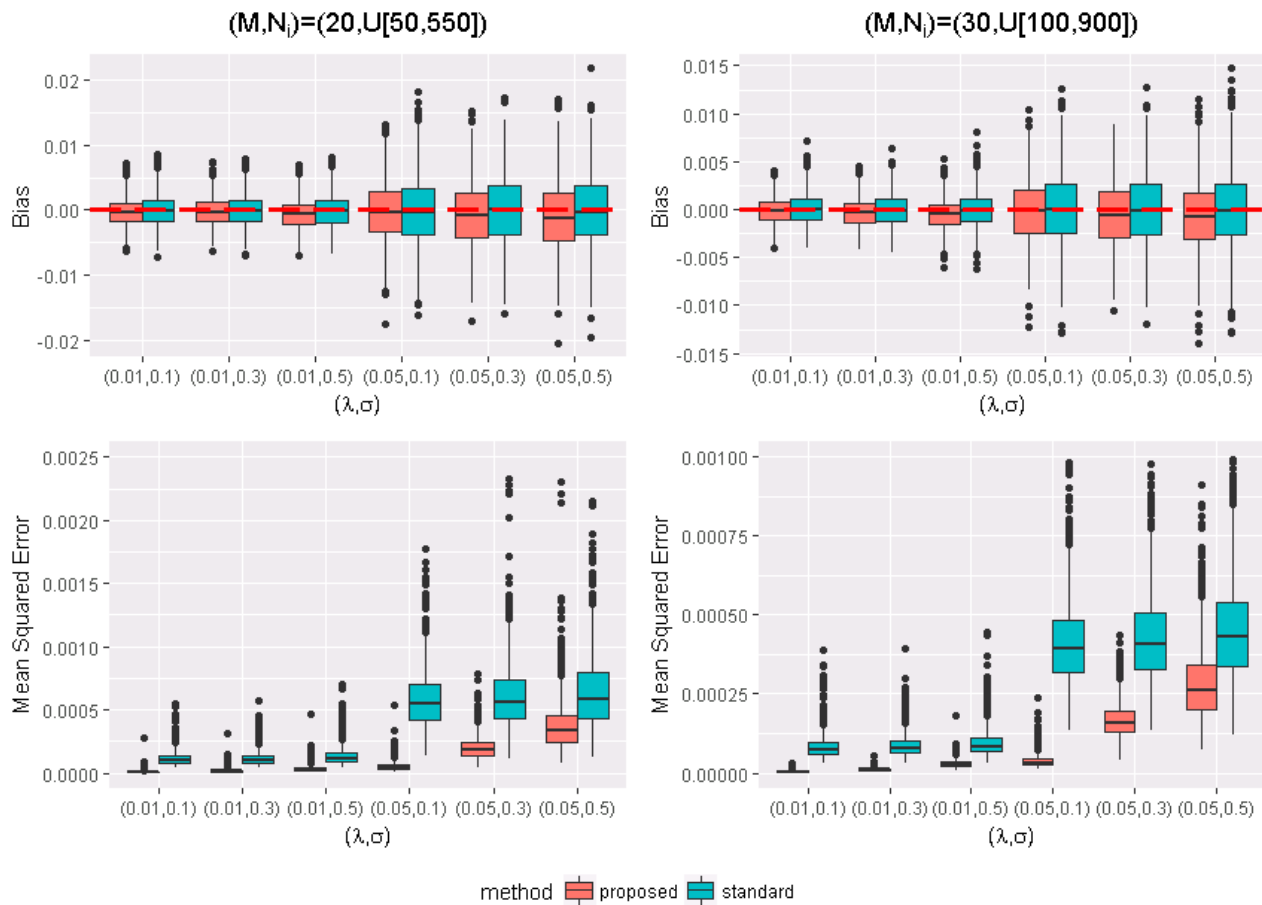
Simulation results for the empirical Type I error and power estimates of the proposed permutation test for a 0.05 level test of the null hypothesis that there is no heterogeneity of incidence rates across communities are shown in S-Figure 3.



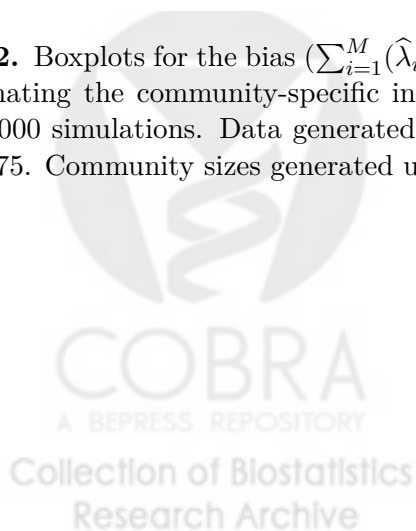
**S-Table 4**

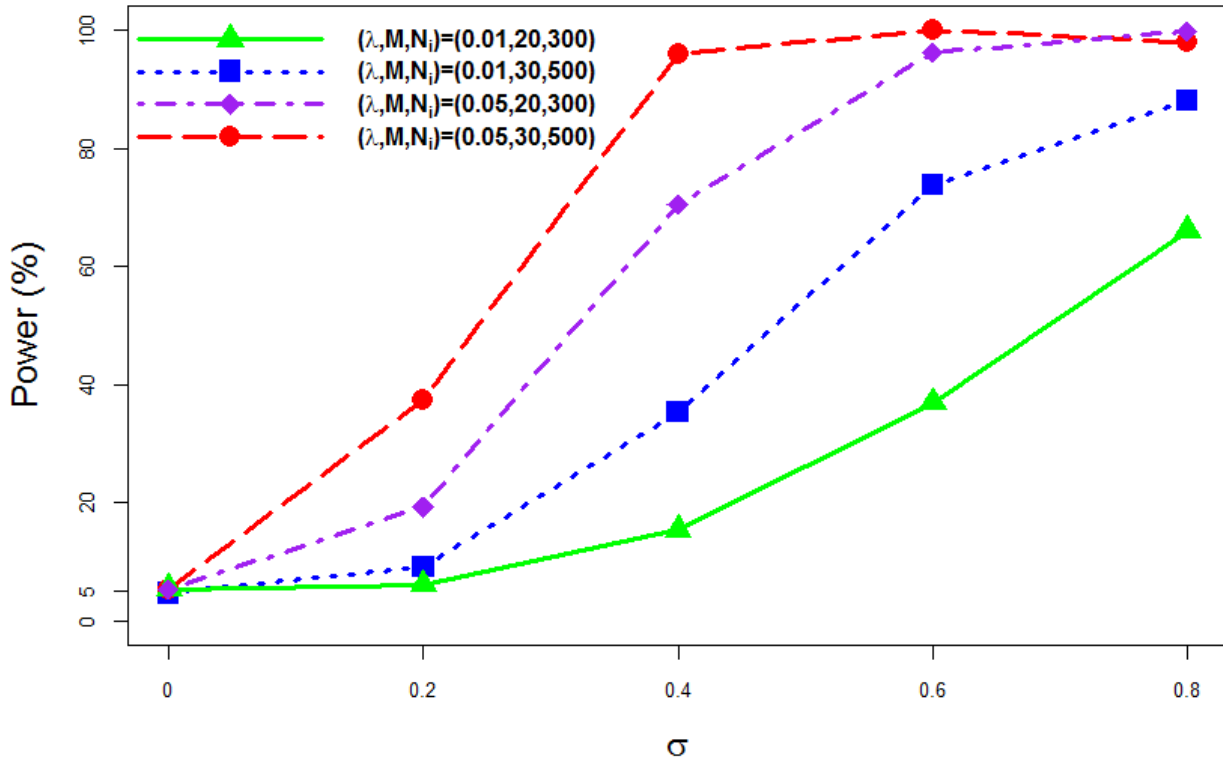
Simulation results for the mis-specification of the random effects distribution, with  $\mu = 0.5$  years,  $p = 1$ ,  $\phi = 0.75$ , and varying incidences. Data generated under Weibull or Gamma, while estimation based on assuming the Lognormal random effects.  $\lambda^*$  and  $\tau$  denote the overall incidence and the between-community standard deviation.  $\widehat{E}(\lambda^*)$  and  $\widehat{SE}(\lambda^*)$  denote average and standard error of estimates from 1000 simulations.  $\widehat{E}(\hat{s})$  denotes average of likelihood-based estimates of the standard error from the 1000 experiments. Coverage denotes the proportion of simulations in which the true  $\lambda^*$  is contained in the nominal 95% confidence interval (CI), and width refers to the mean width of the nominal 95% CI.

		Weibull				Gamma					
$\lambda^*$ (%)	$\tau$ (%)	$\widehat{E}(\lambda^*)$ (%)	$\widehat{SE}(\lambda^*)$ (%)	$\widehat{E}(\hat{s})$ (%)	Coverage (%)	Width (%)	$\widehat{E}(\lambda^*)$ (%)	$\widehat{SE}(\lambda^*)$ (%)	$\widehat{E}(\hat{s})$ (%)	Coverage (%)	Width (%)
$M = 30, N_i = 500$											
1.0	0.1	1.0	0.14	0.15	96.8	0.6	1.0	0.14	0.16	96.2	0.6
1.1	0.3	1.0	0.15	0.17	95.5	0.7	1.0	0.15	0.17	95.2	0.7
1.1	0.6	1.1	0.18	0.22	94.6	0.8	1.1	0.18	0.21	94.9	0.8
3.0	0.3	3.0	0.24	0.26	94.8	1.0	3.0	0.23	0.26	96.8	1.0
3.1	1.0	3.1	0.29	0.33	95.7	1.3	3.1	0.29	0.33	95.8	1.3
3.4	1.8	3.3	0.40	0.47	95.8	1.8	3.4	0.40	0.46	96.0	1.8
5.0	0.5	5.0	0.30	0.33	96.0	1.3	5.0	0.32	0.33	95.5	1.3
5.2	1.6	5.2	0.43	0.47	95.6	1.9	5.1	0.43	0.46	94.5	1.8
5.7	3.0	5.6	0.63	0.73	96.8	2.9	5.6	0.64	0.70	95.1	2.7
$M = 20, N_i = 300$											
1.0	0.1	1.0	0.21	0.25	95.4	1.0	1.0	0.22	0.25	96.0	1.0
1.1	0.3	1.0	0.23	0.27	95.3	1.1	1.0	0.23	0.27	95.0	1.1
1.1	0.6	1.1	0.25	0.32	95.8	1.3	1.1	0.26	0.32	96.8	1.3
3.0	0.3	3.0	0.37	0.41	96.8	1.6	3.0	0.36	0.42	96.1	1.6
3.1	1.0	3.1	0.44	0.48	94.8	1.9	3.1	0.43	0.48	94.4	1.9
3.4	1.8	3.3	0.55	0.65	94.6	2.5	3.3	0.55	0.63	94.9	2.5
5.0	0.5	5.0	0.47	0.53	96.7	2.1	5.0	0.49	0.53	95.8	2.1
5.2	1.6	5.2	0.61	0.66	95.0	2.6	5.2	0.61	0.65	94.4	2.6
5.7	3.0	5.5	0.83	0.94	96.0	3.7	5.6	0.80	0.92	95.7	3.6



**S-Figure 2.** Boxplots for the bias ( $\sum_{i=1}^M (\hat{\lambda}_i - \lambda_i) / M$ ) and the Mean Squared Errors (MSE) ( $\sum_{i=1}^M (\hat{\lambda}_i - \lambda_i)^2 / M$ ) when estimating the community-specific incidences  $\lambda_i$  using the standard method and the proposed method, based on 1000 simulations. Data generated under Lognormal random effects model with  $\mu = 0.5$  years,  $p = 1$ , and  $\phi = 0.75$ . Community sizes generated under discrete uniform distributions.





**S-Figure 3.** Simulation results for the empirical type I error and power estimates (expressed as percentages) of the proposed permutation test for a 0.05 level test of the null hypothesis: no heterogeneity of incidence rates across communities. Data generated under the Lognormal random effects model with  $\mu = 0.5$  years,  $p = 1$ , and  $\phi = 0.75$ . Results based on 1000 simulation runs and for each experiment, 2000 permutations were used for type I error estimates and 1000 permutations were used for power estimates.

# Web Appendix D: Robustness of Project ACCEPT Analysis

Sensitivity analysis results based on varying the false recency rate from 1% to 5% in the Project ACCEPT are presented in S-Table 5. We re-estimated the overall incidence and its variability assuming the false recency rate to be 1%, 2%, 3%, 4%, and 5% for each site, except for site Zimbabwe, where the false recency rate of 4% and 5% are inadmissible given the small number of recent infections identified.

**S-Table 5**

*Sensitivity analysis results based on varying  $p$  in the Project ACCEPT data analysis.  $\hat{\lambda}_{fixed}^*$  and  $\widehat{SE}(\hat{\lambda}_{fixed}^*)$  represent the overall incidence and its standard error by the standard model.  $\hat{\lambda}_{random}^*$  and  $\widehat{SE}(\hat{\lambda}_{random}^*)$  denote the overall incidence and its standard error by the two-level random effects model. All results expressed in %.*

1 – $p$ : False Recency Rate (%)	0	1	2	3	4	5
<b>Site: Soweto</b>						
$\hat{\lambda}_{fixed}^*$	1.2	1.0	0.8	0.6	0.4	0.2
$\widehat{SE}(\hat{\lambda}_{fixed}^*)$	0.12	0.12	0.12	0.12	0.12	0.12
$\hat{\lambda}_{random}^*$	1.3	1.1	0.9	0.8	0.6	0.4
$\widehat{SE}(\hat{\lambda}_{random}^*)$	0.34	0.36	0.37	0.37	0.36	0.31
<b>Site: Tanzania</b>						
$\hat{\lambda}_{fixed}^*$	0.8	0.7	0.6	0.5	0.4	0.4
$\widehat{SE}(\hat{\lambda}_{fixed}^*)$	0.11	0.11	0.11	0.11	0.11	0.11
$\hat{\lambda}_{random}^*$	0.8	0.7	0.6	0.5	0.4	0.4
$\widehat{SE}(\hat{\lambda}_{random}^*)$	0.11	0.11	0.11	0.11	0.11	0.11
<b>Site: Vulindlela</b>						
$\hat{\lambda}_{fixed}^*$	3.8	3.3	2.7	2.1	1.5	0.9
$\widehat{SE}(\hat{\lambda}_{fixed}^*)$	0.25	0.25	0.25	0.25	0.25	0.26
$\hat{\lambda}_{random}^*$	3.8	3.3	2.7	2.1	1.5	0.8
$\widehat{SE}(\hat{\lambda}_{random}^*)$	0.30	0.31	0.31	0.33	0.35	0.39
<b>Site: Zimbabwe</b>						
$\hat{\lambda}_{fixed}^*$	0.9	0.7	0.5	0.3	—	—
$\widehat{SE}(\hat{\lambda}_{fixed}^*)$	0.11	0.11	0.11	0.11	—	—
$\hat{\lambda}_{random}^*$	0.9	0.7	0.5	0.3	—	—
$\widehat{SE}(\hat{\lambda}_{random}^*)$	0.14	0.15	0.16	0.18	—	—

We also explored the robustness of the Project ACCEPT analysis results to different parametric assumptions about the random effects distribution. For each site, we fit the two-level random effects model assuming a Lognormal, Gamma or Weibull random effects distribution (S-Table 6). For all sites, the overall incidence estimates and their standard errors are almost identical regardless of the random effects distribution assumed.

**S-Table 6**

*Robustness of data analysis to the random effects distributional assumption.  $\hat{\lambda}^*$  and  $\widehat{SE}(\hat{\lambda}^*)$  denote the overall incidence and its standard error by the two-level random effects model. All results expressed in %.*

	Lognormal	Gamma	Weibull
<b>Site: Soweto</b>			
$\hat{\lambda}^*$	1.3	1.2	1.2
$\widehat{SE}(\hat{\lambda}^*)$	0.34	0.31	0.30
<b>Site: Tanzania</b>			
$\hat{\lambda}^*$	0.8	0.8	0.8
$\widehat{SE}(\hat{\lambda}^*)$	0.11	0.11	0.11
<b>Site: Vulindlela</b>			
$\hat{\lambda}^*$	3.8	3.8	3.9
$\widehat{SE}(\hat{\lambda}^*)$	0.30	0.30	0.32
<b>Site: Zimbabwe</b>			
$\hat{\lambda}^*$	0.9	0.9	0.9
$\widehat{SE}(\hat{\lambda}^*)$	0.14	0.12	0.13

## Web Appendix E: Sample SAS Codes

The proposed models can be implemented in SAS PROC NLMIXED (SAS Institute Inc., 2015). In E.1, we provide sample codes for the Lognormal random effects model. The current SAS PROC NLMIXED procedure only allows a normal distribution for the random effects  $v_i$  (i.e., a Lognormal distribution for the community-specific incidence  $\lambda_i = \lambda e^{v_i}$ ). To implement the method for other random effects distributions, we apply the probability integral transformation (Hoel et al., 1971; Nelson et al., 2006) and

include sample codes in E.2 (Gamma) and E.3 (Weibull). We provide sample codes for implementing the three-level random effects model in E.4 and describe the required program inputs in E.5.

## E.1 Code for the Two-level Lognormal Random Effects Model

```
PROC NL MIXED data = &dataset TECH = NEWRAP;
  parms lambda = &lambda_init phi = &phi_init sigma2 = &sigma2_init;
  bounds lambda > 0, 0 < phi < 1, sigma2 >= 0;
  lambda_i = lambda*exp(v_i);
  if state = 1 then prob = phi-phi*&mu*lambda*exp(v_i);
  if state = 2 then prob = phi*&mu*lambda*exp(v_i)+(1-&p)*(1-phi);
  if state = 3 then prob = &p*(1-phi);
  ll = log(prob);
  model state ~ general(ll);
  random v_i ~ normal(0,sigma2) subject = community;
  predict lambda_i out = pred_comm_spec;
RUN;
```

## E.2 Code for the Two-level Gamma Random Effects Model

```
PROC NL MIXED data = &dataset TECH = NEWRAP;
  parms lambda = &lambda_init phi = &phi_init theta = &theta_init;
  bounds lambda > 0, 0 < phi < 1, theta > 0;
  lambda_i = lambda*w_i;
  p_i = CDF('NORMAL',a_i);
  if (p_i > 0.999999) then p_i = 0.999999;
  w_i = quantile('GAMMA',p_i,1/theta,theta);
  if state = 1 then prob = phi-phi*&mu*lambda*w_i;
  if state = 2 then prob = phi*&mu*lambda*w_i+(1-&p)*(1-phi);
  if state = 3 then prob = &p*(1-phi);
  ll = log(prob);
  model state ~ general(ll);
  random a_i ~ normal(0,1) subject = community;
  predict lambda_i out = pred_comm_spec;
RUN;
```

## E.3 Code for the Two-level Weibull Random Effects Model

```
PROC NL MIXED data = &dataset TECH = NEWRAP;
  parms lambda = &lambda_init phi = &phi_init theta = &theta_init;
  bounds lambda > 0, 0 < phi < 1, theta > 0;
  lambda_i = lambda*w_i;
  p_i = CDF('NORMAL',a_i);
  if (p_i > 0.999999) then p_i = 0.999999;
  w_i = quantile('WEIBULL',p_i,theta,1);
```

```

if state = 1 then prob = phi-phi*&mu*lambda*w_i;
if state = 2 then prob = phi*&mu*lambda*w_i+(1-&p)*(1-phi);
if state = 3 then prob = &p*(1-phi);
ll = log(prob);
model state ~ general(ll);
random a_i ~ normal(0,1) subject = community;
predict lambda_i out = pred_comm_spec;
RUN;

```

## E.4 Code for Three-level Random Effects Model

```

PROC NLMIXED data = &dataset TECH = NEWRAP;
  parms lambda = &lambda_init phi = &phi_init s1 = &s1_init s2 = &s2_init;
  bounds lambda > 0, 0 < phi < 1, s1 >= 0 s2 >= 0;
  lambda_1i = lambda*exp(v_1i);
  lambda_2i = lambda*exp(v_1i+v_2i);
  if state = 1 then prob = phi-phi*&mu*lambda*exp(v_1i+v_2i);
  if state = 2 then prob = phi*&mu*lambda*exp(v_1i+v_2i)+(1-&p)*(1-phi);
  if state = 3 then prob = &p*(1-phi);
  ll = log(prob);
  model state ~ general(ll);
  random v_1i ~ normal(0,s1) subject = site;
  random v_2i ~ normal(0,s2) subject = community(site);
  predict lambda_1i out = pred_site_spec;
  predict lambda_2i out = pred_comm_spec;
RUN;

```

## E.5 Description of the Required Inputs in E.1-E.4

SAS programs in E.1-E.4 require the following macro variables from the user:

- **dataset**: the name of the input dataset.
  - In E.1-E.3: a dataset with two columns (subjects' community indices and HIV infection states).
  - In E.4: a dataset with three columns (subjects' site indices , subjects' community indices, and HIV infection states) where individuals are nested within communities and communities are nested within sites.
- **mu**: a user-specified  $\mu$  (estimated externally).
- **p**: a user-specified  $p$  (estimated externally).



- `lambda_init`: the initial value for  $\lambda$ .
- `phi_init`: the initial value for  $\phi$ .
- – `sigma2_init` in E.1: the initial value for  $\sigma^2$ .
- `theta_init` in E.2 and E.3: the initial value for  $\theta$  (the scale parameter in the Gamma distribution/the shape parameter in the Weibull distribution).
- `s1_init` and `s2_init` in E.4: initial values for `s1` and `s2` (`s1` is the variance component of the Normal random effects at the site level, and `s2` is the variance component of the Normal random effects at the community level).

## References

- Hoel, P., Port, S., and Stone, C. (1971). *Introduction to Probability Theory*. Houghton Mifflin series in statistics. Houghton Mifflin.
- Nelson, K. P., Lipsitz, S. R., Fitzmaurice, G. M., Ibrahim, J., Parzen, M., and Strawderman, R. (2006). Use of the probability integral transformation to fit nonlinear mixed-effects models with nonnormal random effects. *Journal of Computational and Graphical Statistics* **15**, 39–57.
- SAS Institute Inc. (2015). *SAS/STAT 14.1 User's Guide*. SAS Institute Inc, Cary, NC.

