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

Biometrics 000, 000–000 DOI: 000
000 0000

Cross-sectional HIV Incidence Estimation Accounting for Heterogeneity Across Communities

Yuejia Xu[1], Oliver Laeyendecker[2,3], and Rui Wang[4,5,*]

[1]MRC Biostatistics Unit, University of Cambridge, Cambridge CB2 0SR, U.K.

[2]Department of Medicine, Johns Hopkins University School of Medicine, Baltimore, Maryland 21205, U.S.A.

[3]Department of Epidemiology, Johns Hopkins School of Public Health, Baltimore, Maryland 21205, U.S.A.

[4]Department of Population Medicine, Harvard Medical School and Harvard Pilgrim Health Care Institute, Boston, Massachusetts 02215, U.S.A.

[5]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 the impact of prevention programs, and the design of prevention studies. Cross-sectional HIV incidence estimation based on a HIV diagnostic test [e.g., enzyme-linked immunosorbent assay (ELISA)] and biomarkers of recent infection (e.g., BED capture enzyme immunoassay, Sedia^{TM} HIV-1 limiting antigen avidity enzyme immunoassay) offers important advantages over the standard cohort study. Cross-sectional sample usually consists of samples from different communities. However, small sample sizes limit the ability to estimate community-specific incidence and existing standard 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 provide a 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 (NIMH) Project ACCEPT, a phase III randomized controlled trial of community mobilization, mobile testing, same-day results, and post-test support for HIV in Sub-Saharan Africa, to estimate the overall and community-specific HIV incidence rates.

Key words: Coefficient of variation; Incidence assay; Permutation test; Random effects model.

This paper has been submitted for consideration for publication in *Biometrics*

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 following large cohorts of HIV uninfected individuals for long periods of time and can be prohibitively costly and time-consuming. Such studies are also subject to differential loss to follow-up and behavioral modification. An alternative approach that can reduce cost and avoid lost to follow-up entails assays that can distinguish recent from long-term infections based on markers of HIV disease progression to estimate incidence (Brookmeyer and Quinn, 1995). In this strategy, subjects are administered a diagnostic test; then, those found to be positive are tested with biomarkers associated with recent infection. The incidence estimate is obtained by carefully combining results from the two tests and external information about the "window period", 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 from a cross-sectional sample, 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 for cross-sectional incidence estimation 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 considered 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. In the statistical sense, this correlation exists if and only if knowledge of one individual's outcome confers more information about the outcome

of another individual in the same community than it provides about the outcome of an individual in a separate community. In other words, 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.

[Figure 1 about here.]

Figure 1 presents 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 (horizontal axis) and the heterogeneity across communities, represented as the standard deviation of the 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 achieved and valid inferences on the incidence can be made. 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.

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 test the hypothesis that implementing an enhanced combination prevention package will impact the HIV/AIDS epidemic by significantly reducing HIV incidence

and will be cost-effective (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 valid community-specific incidence estimates and can accurately quantify the uncertainty around the incidence estimates. We also propose a permutation test for the null hypothesis that there is no heterogeneity in incidence rates across communities.

Another purpose of this paper is to propose an estimator for the coefficient of variation (CV) based on cross-sectional incidence data. The statistical power and required sample size for a cluster randomized trial can change substantially depending on the coefficient of variation. The coefficient of variation captures the heterogeneity in outcomes across communities and provides equivalent information regarding variance inflation as the intraclass correlation coefficient (ICC) as described above. 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 the coefficient of variation k increases from 0.20 to 0.45 (i.e., an increase in the ICC from 0.001 to 0.005). To achieve 80% power with a k of 0.45, assuming all else being fixed, the number of clusters required is almost doubled (15 pairs to 27 pairs). Obtaining accurate information on the coefficient of variation is often a major stumbling block in cluster randomized trials (Gail et al., 1992; Hayes and Bennett, 1999; Donner and Klar, 2000; Klar and Donner, 2001; Rutterford et al., 2015). Commonly-used approach is to examine sample size requirements for a range of values for the coefficient of variation, however, even the range can be arbitrary. Methods to estimate the coefficient of variation for HIV incidence based on cross-sectional data at baseline provide a practical way to obtain this essential information for the design of prevention studies.

The rest of the paper is organized as follows. In section 2, we propose a random effects model accounting for heterogeneity of incidence rates across communities, develop a permutation test for the null hypothesis of no heterogeneity across community-specific incidences, and propose an estimator of the coefficient of variation for HIV incidence under the random effects model framework. In section 3, we present results from simulation studies to evaluate the performance of the proposed methods. In section 4, we apply the proposed methods to data from the NIMH Project ACCEPT (HIV Prevention Trials Network [HPTN] 043), where the primary outcome was HIV incidence and was estimated with a cross-sectional multi-assay algorithm and antiretroviral drug screening assay (Coates et al., 2014), to estimate the overall and community-specific incidence rates. We discuss related issues and areas for further research 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 statistical model 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 by not seroconverted); State 2 represents the "recent infection" state, in which an infected individual is identified as a "recent infection" by the biomarkers; and State 3 represents the "long-term infection" state in which an infected individual is classified as a "non-recent infection" by the biomarkers of recent infection. The graphical representation of this three-state model is shown in the Supplementary Material S-Figure 1 (a). 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 recent infection (State 3), respectively, so that $N = N_1 + N_2 + N_3$.

The underlying assumption of this three-state model is that there are no individuals who remain positive for biomarkers of recent infection permanently, which implies that all infected individuals would transition into "long-term infected" eventually. The situation where some subjects remain positive for biomarkers of recent infection indefinitely will be discussed later in section 2.2.3. 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 μ, commonly termed as the "mean window period", to denote the mean population time in State 2. Let λ and $1-\phi$ denote the population 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 (i.e., no heterogeneity in incidence rates across communities), Balasubramanian and Lagakos (2010) and Wang and Lagakos (2010) proposed a likelihood-based framework 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 can be large. Let p_1, p_2, and p_3 denote the prevalence probabilities in State 1, 2, and 3, respectively, it follows that

$$\begin{cases} p_1 = \phi - \phi\mu\lambda \\ p_2 = \phi\mu\lambda \\ p_3 = 1 - \phi \end{cases}$$

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

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

The maximum likelihood estimators of (ϕ, λ) can be obtained by maximizing $L(\phi, \lambda)$, result-

ing in

$$\hat{\phi} = \frac{N_1 + N_2}{N}, \text{ and } \hat{\lambda} = \frac{N_2}{(N_1 + N_2)\mu}.$$

To compute the estimators above, N_1, N_2, and N_3 are obtained from the cross-sectional sample, and μ is typically assumed to be known. Estimators for the variances of $(\hat{\phi}, \hat{\lambda})$ can be derived from the sample Fisher information:

$$\widehat{\text{Var}}(\hat{\phi}) = \frac{(N_1 + N_2)N_3}{N^3}, \text{ and } \widehat{\text{Var}}(\hat{\lambda}) = \frac{N_1 N_2}{(N_1 + N_2)^3 \mu^2}.$$

2.2 *Random Effects Model*

Suppose there are M communities. Let λ_i denote the true community-specific incidence rate in community i, for $i = 1, 2, \ldots, M$. Let N_i denote the total number of subjects in community i, and N_{1i}, N_{2i}, N_{3i} denote the number of subjects in State 1, State 2, and State 3 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 λ^*, that is, $\lambda_i = \lambda^*$, for $i = 1, 2, \ldots, M$. The observed incidence rates $\hat{\lambda}_i$ are distributed around λ^*, and each of $\hat{\lambda}_i$ estimates the same underlying incidence rate λ^*. 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

$$\hat{\lambda}^* = \frac{\sum_{i=1}^{M} N_{2i}}{\left(\sum_{i=1}^{M} N_{1i} + \sum_{i=1}^{M} N_{2i}\right)\mu},$$

and the variance of the overall incidence based on the fixed effects model can be estimated by

$$\widehat{\text{Var}}(\hat{\lambda}^*) = \frac{\sum_{i=1}^{M} N_{1i} \sum_{i=1}^{M} N_{2i}}{\left(\sum_{i=1}^{M} N_{1i} + \sum_{i=1}^{M} N_{2i}\right)^3 \mu^2}.$$

Since the cross-sectional sample usually 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 in many situations. This motivated us to consider a more flexible random effects model, which takes into account the heterogeneity of incidence rates across communities. Contrary to the fixed effects model, we assume community-specific incidence rates are random samples drawn from a distribution and the true community-specific incidence rates λ_i differ by community. 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 τ^2 denotes the between-community variability in true incidence rates. Under the random effects model, the variability of the observed incidence rates $\hat{\lambda}_i$ results from both the within-community sampling error and the variation of true underlying incidence rates λ_i across communities.

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

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

and the marginal likelihood can be obtained by integrating over the random effects. This gives

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

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.

2.2.3 *Extension of the 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 (Hargrove et al., 2008; Wang and Lagakos, 2009; Novitsky et al., 2009; Claggett et al., 2012). Let $1-p$, commonly termed as the false recency rate, denote the proportion of these subject in the HIV positive population. In this case, the incidence estimator obtained based on the three-state model is biased upwards by overestimating the number of subjects who are recently infected. The random effects model proposed in section 2.2.2 can be easily extended to accommodate the false recency rate.

We consider the four-state model as in Wang and Lagakos (2009). A graphical demonstration of this four-state model is provided in the Supplementary Material S-Figure 1 (b). Under this model, a proportion (the false recency rate), $1-p$, of the infected population would be tested as "recently infected" permanently, whereas the remaining individuals would appear as "long-term infected" at some point after seroconversion. Under the four-state model, 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 corresponding to $p = 1$.

The conditional likelihood of community i in (2) can be extended to incorporate the false recency rate as follows:

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

and the marginal likelihood under the four-state model becomes

$$L(\phi,\lambda,\sigma^2) \propto \prod_{i=1}^{M} \int_{v_i} \Big[(\phi - \phi\mu\lambda e^{v_i})^{N_{1i}} \{\phi\mu\lambda e^{v_i} + (1-p)(1-\phi)\}^{N_{2i}} \{p(1-\phi)\}^{N_{3i}} \times \frac{1}{\sigma} \exp\Big(-\frac{{v_i}^2}{2\sigma^2} \Big) \Big] dv_i. \tag{4}$$

The likelihoods reduce to those presented in section 2.2.2 when $p = 1$. As before, the maximum likelihood estimators of $(\phi, \lambda, \sigma^2)$ can be obtained by maximizing the approximated marginal likelihood.

2.2.4 *Implementation.* The proposed random effects model can be implemented in the SAS procedure NLMIXED (SAS Institute Inc., 2014; Kuss and McLerran, 2007). Adaptive Gaussian Quadrature method as described in Pinheiro and Bates (1995) can be used to approximate the integral in (4). A sample code for implementation of the model is provided in the Appendix. The SAS program requires six macro variables to be input by the user, including the name of the input dataset (`dataset`) with two columns (subjects' community index and HIV infection state), the initial value for λ (`lambda_init`), ϕ (`phi_init`), and σ^2 (`sigma2_init`) for the optimization of (4), a user-specified μ (`mu`) based on the clade of HIV and the type of the assay used, and a user-specified p (`p`).

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 and saved in the dataset `mle`. The predicted incidence rate of the ith community is $\hat{\lambda}e^{\hat{v}_i}$, where $\hat{\lambda}$ is the maximum likelihood estimate of λ, and $\hat{v}_i$ is the empirical Bayes estimate of the random effect v_i. Estimates of incidence rates $\hat{\lambda}e^{\hat{v}_i}$ and their variances $\widehat{\mathrm{Var}}(\hat{\lambda}e^{\hat{v}_i})$ computed using the delta method are stored in the `pred_comm_spec` dataset. The overall incidence can thereby be estimated as

$$\hat{\lambda}^* = \frac{1}{M} \sum_{i=1}^{M} \hat{\lambda} e^{\hat{v}_i},$$

and the variance of the overall incidence can be estimated as

$$\widehat{\text{Var}}(\hat{\lambda}^*) = \frac{1}{M^2}\left\{\sum_{i=1}^{M}\widehat{\text{Var}}(\hat{\lambda}e^{\hat{v}_i}) + 2\sum_{1\leqslant i<j\leqslant M}\widehat{\text{Cov}}(\hat{\lambda}e^{\hat{v}_i}, \hat{\lambda}e^{\hat{v}_j})\right\},$$

where $\widehat{\text{Cov}}(\hat{\lambda}e^{\hat{v}_i}, \hat{\lambda}e^{\hat{v}_j}) = \widehat{\text{Var}}(\hat{\lambda})e^{\widehat{\sigma^2}}$.

In addition, the between-community variability in true incidence rates can be estimated as $\hat{\tau}^2 = \hat{\lambda}^2(e^{2\widehat{\sigma^2}} - e^{\widehat{\sigma^2}})$.

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 distribution of the LRT statistic under the null is no longer χ^2 and is hard to derive. The usual regularity condition that the null value is in the interior of the parameter space does not hold. In the linear mixed model with one variance component, it has been shown that the asymptotic distribution of LRT statistic under H_0 is $(1 - a_M)\chi_0^2 + a_M\chi_1^2$, where a_M is a function of M (Crainiceanu and Ruppert, 2004). However, the analytical expression for the distribution of the LRT statistic is hard to derive in more general settings such as unbalanced designs, or under the generalized linear model with non-identity link functions. To circumvent those limitations, Fitzmaurice et al. (2007) proposed a permutation test for variance components in multilevel generalized linear mixed models and demonstrated that the permutation test performs better than tests based on a mixture of χ^2 distributions. These results prompt us to consider a permutation test in the current setting.

Under the null hypothesis of no heterogeneity in incidence rates across communities, the community indices would be random labels and any permutation of the community indices is equally likely. The observed test statistic, for example, the usual LRT statistic

$T = -2(\ell^{ML}_{\sigma^2=0} - \ell^{ML}_{\sigma^2 \geqslant 0})$, can be viewed as a random sample of size 1 from the permutation distribution generated by permuting subjects' community labels.

The p-value is the proportion of test statistics calculated from the permutated dataset that are greater than or equal to the observed test statistic. 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

$$\frac{1 + \sum_{q=1}^{Q} I(T_q \geqslant T_{obs})}{1 + Q},$$

where T_{obs} is the test statistic calculated from the observed dataset, T_q is the test statistic calculated based on the qth permuted dataset, and $I(.)$ 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 (Phipson and Smyth, 2016).

2.4 *The Coefficient of Variation*

When the cross-sectional sample consists of individuals from multiple communities, it is no longer an independent sample of the population of interest 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 rates across communities, then persons in the same cluster will tend to have responses that are more similar to each other than responses of persons in different clusters. The between-cluster variability is usually measured by the coefficient of variation, defined as the between-cluster standard deviation τ divided by the overall incidence λ^*.

The random effects model described in section 2.2 provides a natural way to estimate the coefficient of variation. Under the current lognormal model, the coefficient of variation for incidence can be expressed as

$$CV = \frac{\tau}{\lambda^*} = \frac{\sqrt{\lambda^2(e^{2\sigma^2} - e^{\sigma^2})}}{\lambda e^{\frac{\sigma^2}{2}}} = \sqrt{e^{\sigma^2} - 1}.$$

An estimate of the coefficient of variation 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,

$$\widehat{\text{Var}}(\widehat{CV}) = \frac{e^{2\widehat{\sigma^2}} \times \widehat{\text{Var}}(\widehat{\sigma^2})}{4(e^{\widehat{\sigma^2}} - 1)}, \tag{5}$$

where $\widehat{\sigma^2}$ and $\widehat{\text{Var}}(\widehat{\sigma^2})$ can be obtained from the standard output by implementing the code provided in the Appendix.

We note that when the variability of incidences across communities is extremely small, PROC NLMIXED may fail to provide an estimate of $\text{Var}(\widehat{\sigma^2})$, so we are not able to estimate $\text{Var}(\widehat{CV})$ using (5). In such cases, the bootstrap method may be employed instead. To ensure that the generated bootstrap samples maintain the same dependence structure as the original data, the application of the bootstrap approach to clustered data requires special consideration (Davison and Hinkley, 1997; Ren et al., 2010). Here we describe two common strategies. 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 (Davison and Hinkley, 1997; Field and Welsh, 2007). The other is the "subject bootstrap", where for each community i, $i = 1, 2, \ldots, M$, we draw a bootstrap sample of N_i subjects with replacement from the original sample (Roberts and Fan, 2004). The choice between the two strategies depends on the specific settings. When the number of communities is relatively small, but community sizes are relatively large, the "subject bootstrap" approach is usually preferred; on the other hand, 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 (Sherman and le Cessie, 1997).

3. Simulation Studies

3.1 *Fixed Effects Model vs. Random Effects Model*

We conducted simulation studies under different settings to evaluate the performance of the proposed methods based on the use of the random effects model. Data was simulated from a multinomial distribution with state-specific prevalence

$$\begin{cases} p_{1i} = \phi - \phi\mu\lambda e^{v_i} \\ p_{2i} = \phi\mu\lambda e^{v_i} + (1-p)(1-\phi) \\ p_{3i} = p(1-\phi) \end{cases}$$

where $v_i \overset{i.i.d}{\sim} N(0, \sigma^2)$, for $i = 1, 2, \ldots, M$. We assumed that the mean window period μ is 0.5 years, the false recency rate is 0 ($p = 1$), and the prevalence of long-term infection, $1-\phi = 0.25$ throughout. We considered combinations of $\lambda = 0.01$, 0.03, or 0.05, and $\sigma = 0.1$, 0.3, or 0.5. We examined both situations when $(M,\ N_i) = (30, 500)$ and (20, 300).

[Table 1 about here.]

Table 1 summarizes the simulation results for both the fixed and the random effects model under various configurations, based on 1000 simulation runs. The corresponding overall incidence λ^* and the between-community standard deviation τ are also presented. In each setting, we calculated the average ($\hat{E}(\hat{\lambda}^*)$) and standard error ($\widehat{SE}(\hat{\lambda}^*)$) of overall incidence estimates from 1000 runs. We also obtained the average of likelihood-based estimates of the standard errors ($\hat{E}(\hat{s})$) over 1000 simulated studies to evaluate the accuracy of the variance estimates. The coverage of the 95% confidence interval for λ^*, which refers to the proportion of simulations in which the true λ^* is contained in the nominal 95% confidence interval, together with the average width of the nominal 95% confidence interval are also presented.

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 for the random effects model.

However, when heterogeneity is large, the actual coverage level of the fixed effects model can be substantially lower than the nominal level. In general, the fixed effects model ignores the heterogeneity of incidence rates across communities. While ignoring the heterogeneity does not affect the point estimates for the incidence rates, it can lead to a severe underestimation of the uncertainty associated with the point estimates. As can be seen from the comparison of column $\widehat{SE}(\hat{\lambda}^*)$ and $\hat{E}(\hat{s})$, the true sampling variability of the overall incidence estimates is underestimated by the likelihood-based estimates of the standard error using the fixed effects model, especially when the between community heterogeneity is large. In contrast, confidence intervals based on the random effects model provide the empirical coverage level at or close to the nominal level in all settings. When the heterogeneity is small, the average width of confidence intervals based on the random effects model is only slightly larger than that from the fixed effects model.

To examine the robustness of the proposed model to the mis-specification of the random effect distribution, we performed simulation studies under the same settings except that the current underlying distributions for λ_i are no longer lognormal. We considered two commonly used distributions for time-to-event data, Weibull and Gamma. The proposed lognormal random effects model provides accurate estimates of λ^* and leads to confidence intervals with coverage close to the nominal level when the underlying distribution of the random effect is Weibull or Gamma (Supplementary Material S-Table 1).

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

We evaluated the performance of the proposed permutation test using the similar data generation process as before. We assumed that the mean window period μ is 0.5 years, the false recency rate is 0 ($p = 1$), and the prevalence of long-term infection, $1-\phi = 0.25$. We considered 4 simulation configurations, corresponding to combinations of $(M,\ N_i) = (30, 500)$, $(20, 300)$, and $\lambda = 0.01, 0.05$. We set $\sigma^2 = 0$ to examine the type I error rate and varied σ^2

from 0.2 to 0.8 to assess the power. Within each simulation configuration, we performed 1000 simulation replications and within each replication, we used 2000 permutations to examine type I error and 1000 permutations to assess the power.

[Figure 2 about here.]

Figure 2 presents the empirical type I error and power estimates of the proposed permutation test under various settings for a 0.05 level test of the null hypothesis: no heterogeneity of incidence rates across communities. In general, we note that the empirical type I error estimates of the permutation test are very close to the nominal level of 0.05. The power of the permutation test increases as σ, the number of communities, M, or the community size, N_i, increases. In addition, for fixed σ and $(M,\ N_i)$, the power is higher when the incidence is larger.

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. We compared the average values of $\widehat{CV}$ to the truth. In addition, we considered the estimates of the standard errors and the actual coverage of the 95% confidence intervals using three different approaches described in section 2.4 (the "cluster bootstrap", the "subject bootstrap", and the delta method). For the two bootstrap methods, we took $B = 500$ bootstrap replications. Results are presented in 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 λ^* is small ($\lambda^* \approx 0.01$), we were not able to estimate standard errors of $\widehat{CV}$ due to the instability of estimates for σ. For other settings, among the three methods of standard error estimation, the delta method generally performs well for larger τ (e.g., $\tau > 0.01$). 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 τ is small, bootstrap methods provide more reliable standard error estimates than the delta method. When both λ^* and τ 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 $\widehat{CV} \pm 1.96 \times \widehat{SE}(\widehat{CV})$. In addition to these standard symmetric confidence intervals, 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 Community-Specific Incidences in Project ACCEPT

We applied the proposed method 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 of this study, HIV incidence, was estimated from cross-sectional data 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 assumed to be 259 days. Each study sample was initially characterized based on the results of the two HIV rapid tests performed in-country, and then those who had at least one reactive HIV rapid test results were further assessed using MAA at the HPTN Network Laboratory. Each study participant's HIV status can subsequently be determined, and classified as positive recent infections, positive long-term infections, or negative (HIV-uninfected) (Laeyendecker et al., 2013). The total number of communities and the number of subjects in each state at each of the four site are provided in Table 3 (a).

The estimated overall HIV incidence at each site and the associated uncertainty were calculated using both the standard fixed effects model and the proposed random effects model. In addition, we performed the proposed permutation test with $Q = 2000$ permutations and estimated the coefficient of variation across multiple communities within each site. Results are presented in Table 3 (b).

[Table 3 about here.]

Results imply that there is a strong heterogeneity of community-specific incidences at site Soweto. The coefficient of variation is 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 ($p = 0.005$). While the point estimates of the overall incidence based on the standard fixed effects model and proposed random effects model are almost identical, its uncertainty is substantially underestimated using the standard fixed effects model. For the other three sites, both the coefficient of variation and the permutation test p-value suggest modest heterogeneity. In these cases, the underestimation of uncertainty is minimal, and the standard fixed effects model appears to be adequate although using the proposed random effects

model only leads to minimal efficiency loss (the standard error estimates are similar). For the estimation of $SE(\widehat{CV})$, we present 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 in Tanzania. Here we prefer "subject bootstrap" over "cluster bootstrap" to obtain the standard errors for the CV because the number of communities is relatively small ($M \leqslant 10$) at each site, whereas the total number of subjects in each community is large (28 out of 34 communities in Project ACCEPT have community size $N_i > 1000$). Although there are slight differences in $\widehat{SE}(\widehat{CV})$ using different methods, all methods lead to the same inference, consistent with the permutation test result.

[Figure 3 about here.]

In addition, we estimated community-specific incidence rates using both the standard method and the proposed method based on the random effects model (see Figure 3). The community-specific incidence rates estimated from the proposed random effects model are associated with shorter confidence intervals because they incorporate information from other communities, as compared to the standard community-specific incidence rates, which were calculated based on $\hat{\lambda}_i = \frac{N_{2i}}{(N_{1i}+N_{2i})\mu}$ using the information from that community only.

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. The 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.

In addition to better quantifying the uncertainty around the overall incidence estimate, the random effects model also leads to more efficient estimates of community-specific incidences than applying the standard method to data from each community only because the random effects model pools information from all communities. This is especially useful in the setting of HIV because one single community usually has small sample size and small number of recent infections. We also propose a permutation test for the null hypothesis of no heterogeneity across community-specific incidences. This test is easy to implement and found to perform well in simulation studies. Moreover, the random effects model we consider provides a natural way to estimate the coefficient of variation, an essential parameter in the design and analysis of cluster randomized trials.

Our random effects model assumes a lognormal distribution for the random effects. Our simulation studies suggest that the lognormal model is robust to the mis-specification of several distributions that are commonly used for time-to-event data such as Weibull and Gamma because the shape of these distribution functions can be made similar by choice of parameter values. The proposed methods extend in a straightforward way to the four-state model considered in Wang and Lagakos (2009). When p is not known, we can estimate p through an augmented design where the cross-sectional sample is augmented with a longitudinal component (Wang and Lagakos, 2010), or by incorporating other sources of information, such as combining cohort and cross-sectional data. In addition, the current framework can be extended to incorporate varying prevalence across communities by incor-

porating random effects on ϕ. Our proposed model is also applicable to longitudinal settings under the assumption that individuals tested in different rounds are independent.

Acknowledgements

This research was supported by grant R37 AI51164 from the National Institutes of Health. In addition, this work was supported by HPTN Protocol 043 through contracts U01AI068613/UM 1A068613 (HPTN Network LaboratorySusan Eshleman, PI); U01AI068617/UM1AI068617 (SCHARP Deborah Donnell, PI); and U01AI068619/UM1AI068619 (HIV Prevention Trials Network Sten Vermund/Wafaa El-Sadr, PIs) of the Division of AIDS of the U.S. National Institute of Allergy and Infectious Diseases; and by the Office of AIDS Research of the U.S. National Institutes of Health. Additional support was provided by the Division of Intramural Research, National Institute of Allergy and Infectious Diseases, National Institutes of Health. Views expressed are those of the authors and not necessarily those of sponsoring agencies. The authors thank the NIMH Project ACCEPT Study Team for providing the data.

Supplementary Materials

Figures of the three-state model and the four-state model referenced in section 2, and simulation results under model mis-specification referenced in section 3 are available in the Supplementary Materials.

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. 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., Gray, G., Mbwambo, J. K., Morin, S. F., Richter, L., 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.

Crainiceanu, C. and Ruppert, D. (2004). Likelihood ratio tests in linear mixed models with one variance component. *Journal of the Royal Statistical Society Series B* **66,** 165–185.

Davison, A. C. and Hinkley, D. V. (1997). *Bootstrap Methods and their Application*. Cambridge: Cambridge University Press.

Donner, A. and Klar, N. (2000). *Design and analysis of cluster randomization trials in health research*. Wiley.

Field, C. A. and Welsh, A. H. (2007). Bootstrapping clustered data. *Journal of the Royal Statistical Society Series B* **69,** 369–390.

Fitzmaurice, G. M., Lipsitz, S. R., and Ibrahim, J. G. (2007). A note on permutation tests for variance components in multilevel generalized linear mixed models. *Biometrics* **63,** 942–946.

Gail, M. H., Byar, D. P., Pechacek, T. F., Corle, D. K., Group, C. S., et al. (1992). Aspects of statistical design for the Community Intervention Trial for Smoking Cessation (COMMIT). *Controlled Clinical Trials* **13,** 6–21.

Hallett, T. B. (2011). Estimating the HIV incidence rate–recent and future developments. *Current Opinion in HIV and AIDS* **6,** 102–107.

Hargrove, J. W., Humphrey, J. H., Mutasa, K., Parekh, B. S., McDougal, J. S., Ntozini, R., Chidawanyika, H., Moulton, L. H., Ward, B., Nathoo, K., et al. (2008). Improved HIV-1 incidence estimates using the BED capture enzyme immunoassay. *AIDS* **22,** 511–518.

Hayes, R. and Bennett, S. (1999). Simple sample size calculation for cluster-randomized trials. *International Journal of Epidemiology* **28,** 319–326.

Klar, N. and Donner, A. (2001). Current and future challenges in the design and analysis of cluster randomization trials. *Statistics in Medicine* **20,** 3729–3740.

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., Piwowar-Manning, E., Fiamma, A., Kulich, M., Donnell, D., et al. (2013). Estimation of HIV incidence in a large, community-based, randomized clinical trial: NIMH Project Accept (HIV Prevention Trials Network 043). *PLoS ONE* **8,** 1–9.

Novitsky, V., Wang, R., Kebaabetswe, L., Greenwald, J., Rossenkhan, R., Moyo, S., Musonda, R., Woldegabriel, E., Lagakos, S., and Essex, M. (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.

Phipson, B. and Smyth, G. K. (2016). Permutation p-values should never be zero: calculating exact p-values when permutations are randomly drawn. *ArXiv e-prints* .

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.

Ren, S., Lai, H., Tong, W., Aminzadeh, M., Hou, X., and Lai, S. (2010). Nonparametric bootstrapping for hierarchical data. *Journal of Applied Statistics* **37,** 1487–1498.

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 Eldridge1, S. (2015). Methods for sample size determination in cluster randomized trials. *International Journal of Epidemiology* **44,** 1051–1067.

SAS Institute Inc. (2014). *SAS/STAT 13.2 Users Guide.* SAS Institute Inc, Cary, NC.

Sherman, M. and le Cessie, S. (1997). A comparison between bootstrap methods and generalized estimating equations for correlated outcomes in generalized linear models. *Communications in Statistics - Simulation and Computation* **26,** 901–925.

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.

Appendix

Example SAS Code for the Proposed Random Effects Model

```
PROC NLMIXED data = &dataset TECH = NEWRAP;
    parms lambda = &lambda_init phi = &phi_init sigma2 = &sigma2_init;
    bounds lambda > 0, 0 < phi <= 1, sigma2 >= 0;
    pred = lambda*exp(v);
    if state = 1 then prob = phi-phi*&mu*lambda*exp(v);
    if state = 2 then prob = phi*&mu*lambda*exp(v)+(1-&p)*(1-phi);
    if state = 3 then prob = &p*(1-phi);
    ll = log(prob);
    model state ~ general(ll);
```

```
    random v ~ normal(0,sigma2) subject = community;
    predict pred out = pred_comm_spec;
    ods output ParameterEstimates = mle (keep = Estimate StandardError);
RUN;
```

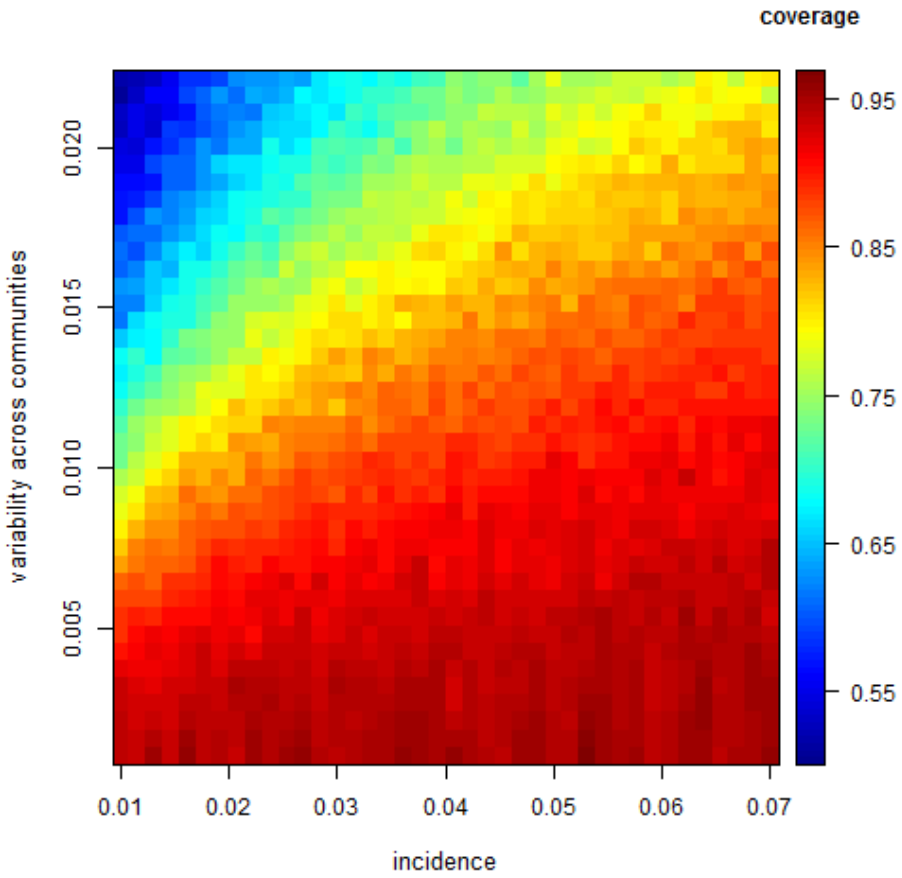


Figure 1. Contour plot of the actual coverage of 95% confidence intervals for the overall incidence across 30 communities ignoring heterogeneity, for varying incidences (horizontal axis) and heterogeneity across communities represented as the standard deviation across community-specific incidences (vertical axis). Dark red represents the situations where the nominal coverage is attained.

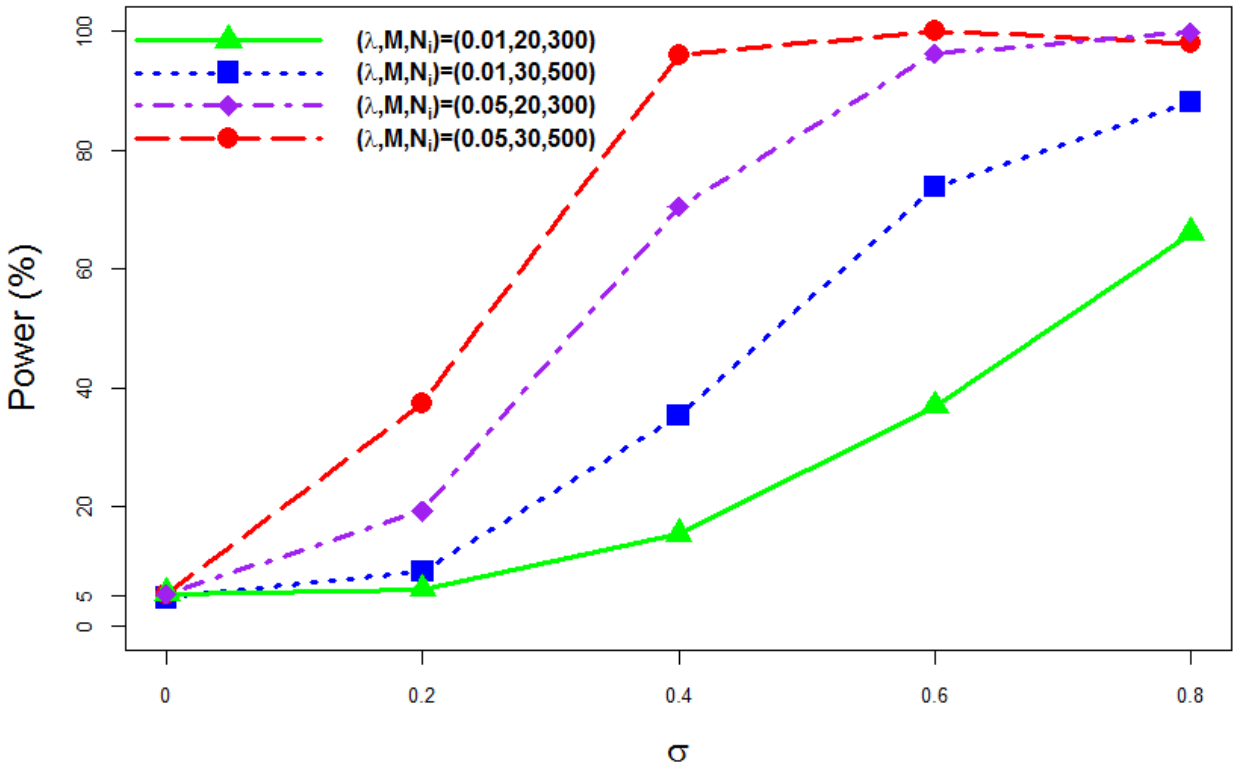


Figure 2. Simulation results for the empirical type I error and power estimates (expressed as percentages) of the proposed permutation test for a 0.05 level test of the null hypothesis: no heterogeneity of incidence rates across communities, based on 1000 replications. Data generated under the lognormal random effects model with $\mu = 0.5$, $p = 1$, and $\phi = 0.75$.

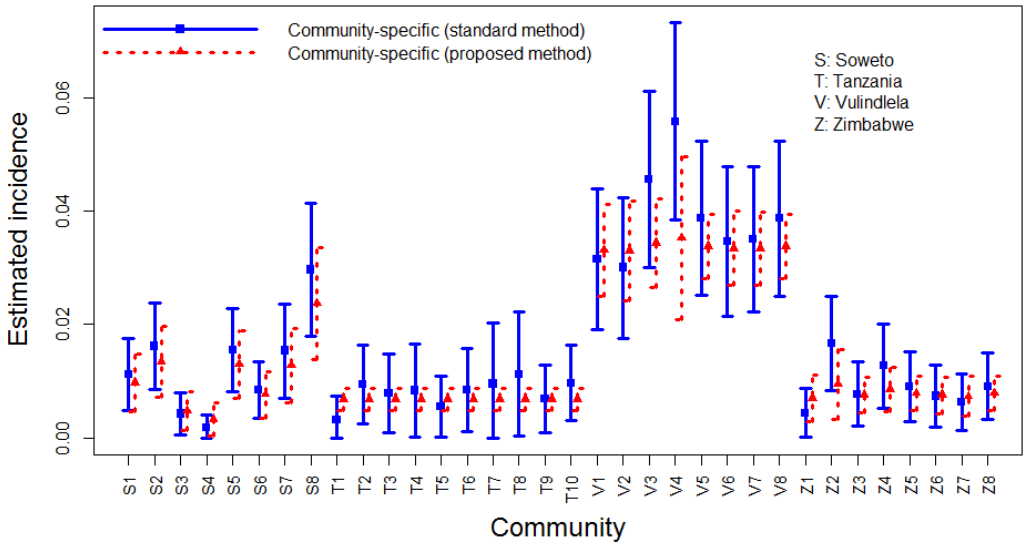


Figure 3. Point estimates and 95% confidence intervals of community-specific HIV incidences of 34 Project ACCEPT communities, using the standard method and the proposed method based on the random effects model.

Table 1

Simulation results for the standard fixed effects model vs. the proposed random effects model, with $\mu = 0.5$, $p = 1$, $\phi = 0.75$, and varying incidences. Estimation based on assuming the random effect follows lognormal distribution. $\hat{E}(\hat{\lambda}^*)$ 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 λ^* is contained in the nominal 95% confidence interval (CI), and width refers to the mean width of the nominal 95% CI.

				Standard Method (Fixed Effects Model)					Proposed Method (Random Effects Model)				
λ	σ	λ^*	τ	$\hat{E}(\hat{\lambda}^*)$	$\widehat{SE}(\hat{\lambda}^*)$	$\hat{E}(\hat{s})$	Coverage	Width	$\hat{E}(\hat{\lambda}^*)$	$\widehat{SE}(\hat{\lambda}^*)$	$\hat{E}(\hat{s})$	Coverage	Width
						$M = 30, N_i = 500$							
0.01	0.1	0.010	0.001	0.010	0.0013	0.0013	0.956	0.005	0.010	0.0014	0.0015	0.952	0.006
	0.3	0.011	0.003	0.010	0.0015	0.0014	0.919	0.005	0.010	0.0015	0.0017	0.946	0.007
	0.5	0.011	0.006	0.011	0.0018	0.0014	0.880	0.006	0.011	0.0018	0.0021	0.965	0.008
0.03	0.1	0.030	0.003	0.030	0.0023	0.0023	0.954	0.009	0.030	0.0024	0.0026	0.959	0.010
	0.3	0.031	0.010	0.031	0.0030	0.0023	0.877	0.009	0.031	0.0028	0.0032	0.961	0.013
	0.5	0.034	0.018	0.034	0.0043	0.0024	0.725	0.010	0.034	0.0041	0.0044	0.954	0.017
0.05	0.1	0.050	0.005	0.050	0.0030	0.0029	0.941	0.012	0.050	0.0031	0.0033	0.958	0.013
	0.3	0.052	0.016	0.052	0.0042	0.0030	0.835	0.012	0.052	0.0040	0.0046	0.953	0.018
	0.5	0.057	0.030	0.057	0.0064	0.0031	0.646	0.012	0.056	0.0062	0.0066	0.944	0.026
						$M = 20, N_i = 300$							
0.01	0.1	0.010	0.001	0.010	0.0021	0.0021	0.929	0.008	0.010	0.0021	0.0025	0.964	0.010
	0.3	0.011	0.003	0.011	0.0022	0.0021	0.927	0.008	0.010	0.0022	0.0027	0.956	0.011
	0.5	0.011	0.006	0.011	0.0027	0.0022	0.878	0.009	0.011	0.0027	0.0031	0.937	0.012
0.03	0.1	0.030	0.003	0.030	0.0037	0.0036	0.937	0.014	0.030	0.0038	0.0041	0.950	0.016
	0.3	0.031	0.010	0.032	0.0044	0.0037	0.926	0.015	0.031	0.0043	0.0047	0.955	0.019
	0.5	0.034	0.018	0.034	0.0056	0.0038	0.836	0.015	0.033	0.0056	0.0062	0.943	0.024
0.05	0.1	0.050	0.005	0.050	0.0046	0.0047	0.950	0.018	0.050	0.0046	0.0053	0.969	0.021
	0.3	0.052	0.016	0.052	0.0061	0.0047	0.874	0.019	0.052	0.0061	0.0065	0.945	0.026
	0.5	0.057	0.030	0.056	0.0084	0.0049	0.744	0.019	0.056	0.0085	0.0089	0.937	0.035

Table 2

Simulation results for the coefficient of variation (CV) estimator based on the lognormal random effects model, with $\mu = 0.5$, $p = 1$, $\phi = 0.75$, and varying incidences. $\hat{E}(\widehat{CV})$ and $\widehat{SE}(\widehat{CV})$ denote average and standard error of estimates from 1000 simulated studies. $\hat{E}(\hat{s})$ 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.

							Cluster Bootstrap		Subject Bootstrap		Delta Method		
λ	σ	λ^*	τ	CV	$\hat{E}(\widehat{CV})$	$\widehat{SE}(\widehat{CV})$	$\hat{E}(\hat{s})$	Coverage	$\hat{E}(\hat{s})$	Coverage	$\hat{E}(\hat{s})$	Coverage	Non-NA
						$M = 30, N_i = 500$							
0.01	0.1	0.010	0.001	0.100	0.137	0.184	0.212	0.987	0.275	1.000	0.427	0.958	0.428
0.01	0.3	0.011	0.003	0.307	0.248	0.237	0.224	0.964	0.287	1.000	0.355	1.000	0.639
0.01	0.5	0.011	0.006	0.533	0.475	0.253	0.256	0.878	0.324	0.911	0.278	1.000	0.899
0.03	0.1	0.030	0.003	0.100	0.091	0.114	0.103	0.937	0.117	0.965	0.195	0.940	0.469
0.03	0.3	0.031	0.010	0.307	0.271	0.133	0.104	0.844	0.116	0.882	0.151	0.997	0.913
0.03	0.5	0.034	0.018	0.533	0.495	0.136	0.119	0.886	0.125	0.912	0.136	0.940	0.996
0.05	0.1	0.050	0.005	0.100	0.085	0.090	0.076	0.927	0.085	0.966	0.152	0.963	0.562
0.05	0.3	0.052	0.016	0.307	0.280	0.101	0.075	0.853	0.082	0.896	0.095	0.996	0.970
0.05	0.5	0.057	0.030	0.533	0.505	0.115	0.094	0.847	0.090	0.863	0.112	0.914	1.000
						$M = 20, N_i = 300$							
0.01	0.1	0.010	0.001	0.100	0.192	0.304	——	——	——	——	——	——	0.378
0.01	0.3	0.011	0.003	0.307	0.248	0.346	——	——	——	——	——	——	0.464
0.01	0.5	0.011	0.006	0.533	0.444	0.403	——	——	——	——	——	——	0.678
0.03	0.1	0.030	0.003	0.100	0.125	0.162	0.169	0.962	0.214	0.997	0.349	0.965	0.461
0.03	0.3	0.031	0.010	0.307	0.241	0.198	0.182	0.819	0.219	0.998	0.267	1.000	0.709
0.03	0.5	0.034	0.018	0.533	0.464	0.235	0.217	0.862	0.237	0.900	0.229	1.000	0.930
0.05	0.1	0.050	0.005	0.100	0.096	0.123	0.120	0.944	0.149	0.987	0.253	0.970	0.461
0.05	0.3	0.052	0.016	0.307	0.250	0.163	0.131	0.797	0.148	0.857	0.181	0.998	0.827
0.05	0.5	0.057	0.030	0.533	0.488	0.173	0.152	0.876	0.159	0.909	0.171	0.961	0.988

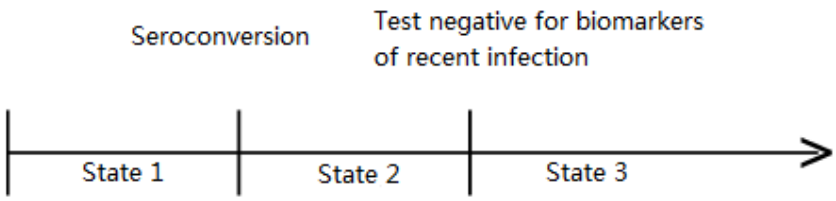
Table 3
Sample classification and analysis results of Project ACCEPT cross-sectional incidence data at each African site.

	Soweto	Tanzania	Vulindlela	Zimbabwe
(a). Sample classification.				
Number of communities (M)	8	10	8	8
Uninfected (N_1)	11962	8505	8197	10348
Recent infections (N_2)	101	47	230	67
Long-term infections (N_3)	1547	479	3400	1461
(b). Analysis results.				
Prevalence of long-term infection ($1-\hat{\phi}$)	0.114	0.053	0.288	0.123
Overall incidence by the standard model ($\hat{\lambda}^*_{fixed}$)	0.012	0.008	0.038	0.009
Variability of the standard incidence estimate ($\widehat{SE}(\hat{\lambda}^*_{fixed})$)	0.0012	0.0011	0.0025	0.0011
Overall incidence by the proposed model ($\hat{\lambda}^*_{random}$)	0.011	0.007	0.034	0.008
Variability of the proposed incidence estimate ($\widehat{SE}(\hat{\lambda}^*_{random})$)	0.0030	0.0010	0.0027	0.0013
Permutation test p-value	0.005	>0.99	0.27	0.17
Coefficient of variation ($\widehat{CV}$)	0.7420	0.0001	0.0631	0.1872
Variability of $\widehat{CV}$ by delta method ($\widehat{SE}_{delta}(\widehat{CV})$)	0.310	——	0.157	0.207
Variability of $\widehat{CV}$ by cluster bootstrap ($\widehat{SE}_{cluster}(\widehat{CV})$)	0.282	0.076	0.072	0.142
Variability of $\widehat{CV}$ by subject bootstrap ($\widehat{SE}_{subject}(\widehat{CV})$)	0.237	0.208	0.098	0.180

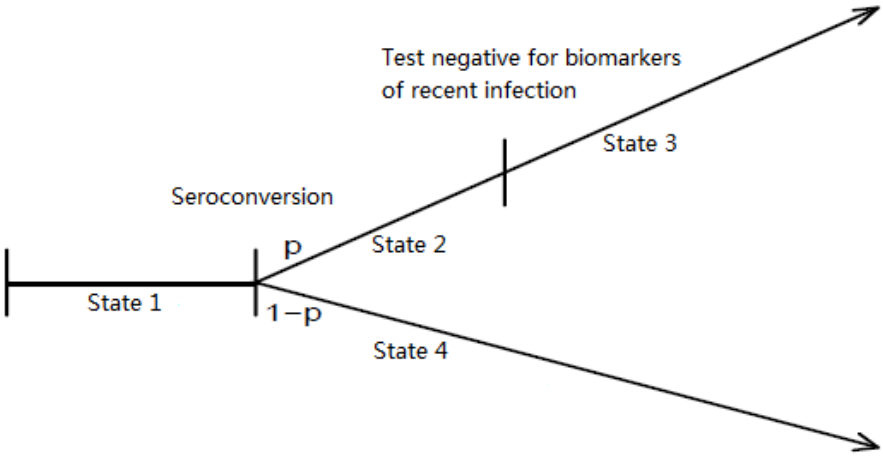
Supplementary Materials for "Cross-sectional HIV Incidence Estimation Accounting for Heterogeneity Across Communities"

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

Supplementary Figure and Table



(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 longitudinal models.

S-Table 1: *Simulation results for the mis-specification of the random effect distribution, with $\mu = 0.5$, $p = 1$, $\phi = 0.75$, and varying incidences. Data generated under Weibull or Gamma, while estimation based on assuming the lognormal random effect. $\hat{E}(\hat{\lambda}^*)$ 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 λ^* is contained in the nominal 95% confidence interval (CI), and width refers to the mean width of the nominal 95% CI.*

		Weibull					Gamma				
λ^*	τ	$\hat{E}(\hat{\lambda}^*)$	$\widehat{SE}(\hat{\lambda}^*)$	$\hat{E}(\hat{s})$	Coverage	Width	$\hat{E}(\hat{\lambda}^*)$	$\widehat{SE}(\hat{\lambda}^*)$	$\hat{E}(\hat{s})$	Coverage	Width
					$M = 30, N_i = 500$						
0.010	0.001	0.010	0.0014	0.0015	0.968	0.006	0.010	0.0014	0.0016	0.962	0.006
0.011	0.003	0.010	0.0015	0.0017	0.955	0.007	0.010	0.0015	0.0017	0.952	0.007
0.011	0.006	0.011	0.0018	0.0022	0.946	0.008	0.011	0.0018	0.0021	0.949	0.008
0.030	0.003	0.030	0.0024	0.0026	0.948	0.010	0.030	0.0023	0.0026	0.968	0.010
0.031	0.010	0.031	0.0029	0.0033	0.957	0.013	0.031	0.0029	0.0033	0.958	0.013
0.034	0.018	0.033	0.0040	0.0047	0.958	0.018	0.034	0.0040	0.0046	0.960	0.018
0.050	0.005	0.050	0.0030	0.0033	0.960	0.013	0.050	0.0032	0.0033	0.955	0.013
0.052	0.016	0.052	0.0043	0.0047	0.956	0.019	0.051	0.0043	0.0046	0.945	0.018
0.057	0.030	0.056	0.0063	0.0073	0.968	0.029	0.056	0.0064	0.0070	0.951	0.027
					$M = 20, N_i = 300$						
0.010	0.001	0.010	0.0021	0.0025	0.954	0.010	0.010	0.0022	0.0025	0.960	0.010
0.011	0.003	0.010	0.0023	0.0027	0.953	0.011	0.010	0.0023	0.0027	0.950	0.011
0.011	0.006	0.011	0.0025	0.0032	0.958	0.013	0.011	0.0026	0.0032	0.968	0.013
0.030	0.003	0.030	0.0037	0.0041	0.968	0.016	0.030	0.0036	0.0042	0.961	0.016
0.031	0.010	0.031	0.0044	0.0048	0.948	0.019	0.031	0.0043	0.0048	0.944	0.019
0.034	0.018	0.033	0.0055	0.0065	0.946	0.025	0.033	0.0055	0.0063	0.949	0.025
0.050	0.005	0.050	0.0047	0.0053	0.967	0.021	0.050	0.0049	0.0053	0.958	0.021
0.052	0.016	0.052	0.0061	0.0066	0.950	0.026	0.052	0.0061	0.0065	0.944	0.026
0.057	0.030	0.055	0.0083	0.0094	0.960	0.037	0.056	0.0080	0.0092	0.957	0.036