Small-Area Spatial Analysis and Hierarchical Modeling of HIV-1 Infection: A Case Study among Pregnant Women in a Rural but Densely Populated Region of Southern Rwanda

DEMARC A. HICKSON¹,², LANCE A. WALLER¹, ANN CHAO³,⁴,⁷, ABEL DUSHIMIMANA⁵,⁶,⁷, MARC BULTERYS⁴,⁷

¹Department of Biostatistics, Rollins School of Public Health, Emory University, Atlanta, Georgia, 30322; ²Department of Medicine, School of Medicine, University of Mississippi Medical Center, Jackson, Mississippi, 39213; ³Global AIDS Program, National Center for HIV, STD, and TB Prevention (NCHSTP), Centers for Disease Control and Prevention (CDC), Atlanta, Georgia 30333; ⁴CDC – Zambia, Lusaka, Zambia; ⁵School of Medicine, National University of Rwanda, Butare, Rwanda; ⁶School of Public Health, National University of Rwanda, Kigali, Rwanda; ⁷National University of Rwanda – Johns Hopkins University AIDS Research Project, Butare, Rwanda (1989-1994)

SUMMARY

Background: Few studies have explored the association of regional/local risk factors with HIV-1 infection among women of childbearing age from a small-area spatial perspective, particularly outside the capital cities.

Methods: We use spatial analysis methods to investigate historical regional differences in risk factors associated with prevalent HIV-1 infection among 7,444 pregnant women residing in fifty-four geographic sectors within a twenty-five-kilometer radius of Butare in southern Rwanda in the period preceding the extensive political and ethnic upheaval of the mid 1990s.

Results: The spatial patterns suggest modifications to earlier models (e.g. omission/inclusion of covariates within small areas) and suggest the observed urban/rural gradient is not solely explained by local variations in the sociobehavioral covariates in the model. The analyses suggest that local standardized morbidity ratios (SMRs) were highest in the urban center and declined toward the rural boundaries of the study region. Specifically, the urban sectors within
the study area exhibit a five-fold increase in HIV-1 prevalence over that observed in the rural sectors.

Conclusions: While this study does not depict the current pattern of the epidemic, the analysis demonstrates the use of spatial analysis methodology in HIV studies and highlights historical differences in the risk factors associated with HIV-1 prevalence among pregnant women in the area surrounding the town of Butare in Southern Rwanda. These spatial models serve as a baseline for exploring spatially-varying coefficients in order to assess if covariate effects are the same and this study documents substantial small-area variations in HIV prevalence among pregnant Rwandan women.

Key Words:

Africa, Butare, Rwanda, hierarchical modeling, HIV-1 prevalence, spatial analysis, standardized morbidity ratio, young women
INTRODUCTION

According to the World Health Organization (WHO), the HIV epidemic entered into Rwanda in the late 1970s\(^1\). By the mid-1980s, HIV-1 prevalence was found to be 88% among commercial sex workers\(^2\), 29% among women of childbearing age in the capital city, Kigali\(^3\), and 1.3% among a national sample of rural inhabitants\(^4\). Before the 1994 Rwandan genocide, the HIV-1 infection pattern among pregnant women was similar to other sub-Saharan African countries with high rates of infection (>20%) in urban areas and generally low rates (1-5%) in rural areas\(^5\text{--7}\). Despite numerous HIV prevalence reports among various populations in sub-Saharan Africa\(^5\text{--16}\), few studies have explored in detail the association of regional/local risk factors with HIV-1 infection among women of childbearing age from a small-area spatial perspective, particularly outside the capital cities\(^8,13\).

We keep to earlier results presented by investigators who identify the following risk factors for HIV-1 prevalence among women seeking antenatal care at clinics in the vicinity of Butare, Rwanda: increased sexual partners other than husband/regular partner, history of at least one sexually transmitted disease (STD), not being legally married, high household income, and history of oral contraceptive use, history of sexual intercourse as a means of income, cigarette smoking, age at first pregnancy, low gravidity, and circumcision of partner\(^9\). A spatial analysis of these data is of interest for several reasons. First, the data were collected during a time when higher HIV-1 prevalence rates were transitioning from established urban foci to the surrounding peri-urban and rural areas. As a result, we may wish to see if observed associations between prevalence and risk factors vary between geographic regions (e.g. urban and rural areas). Second, the presence of omitted or unmeasured geographically-varying risk factors could induce
spatial correlation among model residuals which, in turn, could bias parameter estimates associated with those risk factors included in the model.

In this analysis, we use spatial analytic techniques to highlight differences in risk factors associated with prevalent HIV-1 infection among pregnant women by geographic region in and around the town of Butare in southern Rwanda. Our goals are two-fold: to assess whether model output (specifically posterior estimates of local random effects) can identify potential local improvements to the model and to provide statistically stable estimates of local small-area HIV-1 prevalence. Hierarchical Bayesian methods allow us to investigate spatial patterns associated with HIV-1 infection in this mixed rural and urban population and suggest where (within the study area) the inclusion/exclusion of various covariates might be needed. A small-area spatial approach provides several advantages over simple mapping of observed crude prevalence rates: first, the hierarchical Bayes approach “borrows strength” across the data set to improve small-area HIV-1 prevalence estimation based on small local sample sizes\textsuperscript{17–19}. Second, the small-area spatial analysis compensates for any residual spatial correlation in the model thereby adjusting for potential bias in parameter estimation\textsuperscript{18}.

**MATERIALS**

**Description of study population and study area.** A large cross-sectional survey was conducted among pregnant women residing in the region around the town of Butare in southern Rwanda to estimate HIV-1 prevalence and identify associated sociodemographic and sexual history risk factors. Between October 31, 1989, and February 28, 1993, pregnant women who presented at one of five antenatal clinics in the area, were offered a voluntary HIV test after informed consent and pre-test counseling. An estimated 94% of pregnant women who were offered HIV screening accepted\textsuperscript{9}. Each of the 7,444 participants provided demographic and
health-related information, sexual and reproductive history, as well as geographic residency information. The results regarding risk factors associated with prevalent and incident HIV-1 infection in this study population have been presented previously\(^9,10\). In the sections below, we concentrate on identifying possible small-area geographical differences in risk factors among this population.

At the time of the original study, Rwanda consisted of 12 Provinces and 116 districts and municipalities that were further partitioned into smaller regions, which we refer to as sectors. Our spatial analysis focused on pregnant women who resided in 54 sectors within a 25-kilometer radius around the town of Butare where antenatal coverage was high (80% or higher). Geographic sectors were categorized further into urban (5), peri-urban (11), semi-rural (22), or strictly rural (16) as illustrated in Figure 1(a). In this sample of participants, a total of 1,092 pregnant women (14.7%) reported residency in an urban sector, followed by 2,076 (27.9%) in a peri-urban sector, 2,675 (35.9%) in a semi-rural sector, and 1,601 (21.5%) in a strictly rural sector.

**METHODS**

Hierarchical Bayesian modeling. Hierarchical models allow us to ‘borrow strength’ from both neighboring regions and the entire geographical region in order to stabilize estimates based on small local sample sizes within sectors. Such methods were introduced to disease mapping in an empirical Bayes setting\(^{17}\) then extended to a fully Bayesian formulation\(^{18}\) and placed within the general framework of generalized linear mixed models (GLMMs)\(^{19}\). Methodological details have been described in detail elsewhere\(^{18–21}\) and are briefly outlined below.
**First-stage model.** We created pairs of observed and expected observations \((Y_i, E_i)\) using the Butare HIV data in the \(i = 1, \ldots, 54\) sectors, where \(E_i\) represents a baseline multiplicative estimate of prevalence based on the number of pregnant women seeking antenatal care within sector \(i\) and the overall estimate of HIV prevalence among study subjects. The first-stage model assumes

\[
Y_i | \eta_i \sim \text{Poisson}(E_i \times \exp(\eta_i)), \quad (1)
\]

where \(\eta_i\) denotes the logarithm of the relative risk associated with residence in sector \(i\). A similar model, in the context of GLMMs, to allow for subject-based risk factors\(^{19}\), is given by

\[
\eta_i = \sum_{j=1}^{n_i} X_{ij} \beta + \theta_i, \quad i = 1, \ldots, 54, \quad (3)
\]

where, in our context, \(X_{ij}\) represents the design matrix of observed covariate values for individual \(j\) in sector \(i\), \(\beta\) is the corresponding vector of unknown model parameters, and \(\theta_i\) are the sector-specific random effects.

In this analysis, we select a subset of factors previously described to be associated with HIV infection among pregnant women in Rwanda and elsewhere in Africa\(^{9-11}\) as an illustration of spatial analysis methods in HIV studies. The selected factors include each woman's current age, marital status, partner's circumcision status as reported by the woman, monthly household income, age at first sexual intercourse, age at first pregnancy, and history of sexually transmitted diseases (STDs), as well as whether the woman had any sexual partner(s) other than her husband/regular partner in the past 5 years, engaged in sexual intercourse to support herself, or was currently infected with syphilis (positive syphilis serology by RPR). We record a woman’s age at first sexual intercourse and age at first pregnancy in increasing years. We treat a woman’s current age and monthly household income as categorical variables: age \((<19, 20--24, 25--29,\)
and 30+) and income (<920, 920--4800, and 4800+ Rwandan Francs), respectively. Marital status is coded as married (common law or civil union) or not married (single, divorced, separated, or widowed). Lastly, a woman’s partner’s circumcision status as reported by the woman, having had sexual partners other than husband, history of sexually transmitted diseases (STDs), having had sex to support herself, and current infection with syphilis are all coded yes or no. We remove non-significant risk factors (based on the 95% confidence interval/credible set and corresponding model fit criterion) from the model and any subsequent analyses.

**Second-stage model.** We assign a uniform prior distribution to the vector of model parameters $\beta$. We assume the covariate effects are constant across the study area and capture any residual geographic correlation through the collection of random effects, $\theta$. Positive estimates of $\theta_i$ correspond to under-prediction of the outcome by local covariate values (in sector $i$) while negative estimates correspond to over-prediction of the outcome.

We define $\theta$ alternatively with an exchangeable and a spatially correlated prior structure\textsuperscript{17,18,21,22}. First, we assign $\theta$ an exchangeable Gaussian prior structure written as

$$\theta_i \sim \text{N}(0, \psi), \quad i = 1, \ldots, 54$$  \hspace{1cm} (4)

where $\psi$ is the shared prior variance\textsuperscript{18}. This prior structure samples each random effect from the same underlying distribution; implying similarities among all locations despite ‘apparent’ geographical differences. This results in ‘shrinkage’ of local rate estimates toward the global observed rate.
For comparison, we also assign $\theta$ a spatially conditionally autoregressive (CAR) Gaussian prior structure. This prior takes into account the local ‘clustering’ of risk among neighboring regions and provides a spatial dependency commonly used in disease mapping\textsuperscript{17,20--22}. One formulation of this model is

$$
\theta_i | \theta_{j \neq i} \sim N \left( \sum_{j \neq i} w_{ij} \theta_j, \frac{1}{\tau \sum_{j \neq i} w_{ij}} \right),
$$

where $w_{ij}$ denotes a set of weights indicating the adjacency of neighboring sectors and $\tau$ denotes a variance hyperparameter. For our purposes, if two sectors share a common border $w_{ij} = 1$, and $w_{ij} = 0$ otherwise. By convention, $w_{ii} = 0$. As a result, $\sum_{j \neq i} w_{ij}$ indicates the number of sectors neighboring sector $i$.

**Third-stage model.** Hyperparameters capture the sampling variability and spatial dependency of the random effects and complete the model. The Gamma distribution provides a commonly used conjugate prior distribution for the shared variance ($\psi$) in (4) and conditional variance ($\tau$) in (5) and can be written as

$$
\frac{1}{\omega} \sim \text{Gamma}(\alpha, \delta),
$$

where $\omega = \psi$ or $\tau$. The chosen parametric model serves as either an informative or weakly informative prior and we explore a range of values for $\alpha$ and $\delta$\textsuperscript{21--23} in our analyses.

**Fitting the hierarchical model: likelihood-based and MCMC approximation methods.** We compare two analytical approximation approaches to fit the hierarchical models: maximization of the penalized quasi-likelihood (PQL)\textsuperscript{19} and Markov chain Monte Carlo (MCMC) methods\textsuperscript{24}. The primary theoretical difference involve the latter’s inclusion and estimation of the third stage model, i.e., the prior distribution of $\tau$ and $\psi$. Algorithmically, the
two methods make different assumptions and implement different coefficient approximation methods, but ultimately fit the same model\textsuperscript{25}. We utilize both approaches to fit the hierarchical models and to assess the stability and reliability of the results across fitting techniques.

We fit the hierarchical model with the exchangeable random effects in SAS Version 8.2 using the \texttt{\%GLIMMIX} Macro (Model I), which implements PQL estimation methods. We fit two separate models in WinBUGS 1.4 with two independent chains of an MCMC sampler for 1,000 iterations, following a ‘burn-in’ period of 1,000 iterations using each of the prior specifications for the random effects. The first MCMC model (Model II) incorporates exchangeable random effects while the second MCMC model (Model III) incorporates spatially correlated random effects. We assess convergence of the parameter estimates in Models II and III by the Gelman-Rubin (G-R) statistic\textsuperscript{26}. There is currently no general diagnostic tool in the statistical literature to make model comparisons across all three models, however, the deviance information criterion (DIC) allows for comparison of model fit between Models II and III (see Appendix)\textsuperscript{27}.

**Missing data.** Small-area spatial residency information is of particular interest because it allows us to identify regional model fit patterns and map HIV-1 prevalence estimates. A woman missing residency information was not included in this analysis. However, an independent draw from the MCMC posterior predictive distribution provides imputation of missing covariate data for any given individual. This allows us to obtain a complete dataset to fit the hierarchical models in both SAS and WinBUGS.

**Computation of OR, SMR, CI, and CS.** We calculate an odds ratio (OR) for each covariate effect along with a standardized morbidity ratio (SMR) associated with residence in sector $i$. Exponentiating the coefficient estimate of each respective covariate effect from the three
models above provides an associated odds ratio. Sector-specific SMRs result from exponentiating the estimate of the respective random effect. SAS provides confidence intervals (CI) and WinBUGS provides credible sets (CS) for each parameter in the model. More specifically, the 95% CIs result from adding and subtracting the multiplicative result of the coefficient’s standard deviation and Z-score associated with a 2.5% significance level from the coefficient estimate of the covariate and random effects. The 95% CSs result from extracting the 2.5\textsuperscript{th} and 97.5\textsuperscript{th} percentile estimates of the MCMC samples from the posterior distribution of the respective covariate and random effects.

RESULTS

Of the 7,444 women who participated in the study and reported residency information, 704 tested positive (overall HIV-1 prevalence = 9.5%) for HIV-1 with the largest number (72) of HIV-1 infected women from Sector 01, an urban sector. No pregnant woman tested positive for HIV-1 from Sector 162, a rural sector. A large percentage (67.1%) of women infected with HIV-1 resided within the urban and peri-urban sectors of the study region. The highest HIV-1 prevalence (21.1%) was observed among women who resided in the urban sectors and the lowest prevalence (3.6%) observed among women who resided in the rural sectors.

Crude HIV-1 infection mapping. Estimates of HIV-1 infection in the northeastern and southern rural sectors of the study area were relatively low, similar to a few eastern and northwestern sectors (Figure 1(b)). HIV-1 prevalence in the urban and peri-urban sectors was, on average, much higher than in the semi-rural and rural sectors (Figure 1(b)). In fact, urban Sectors 01 and 00, yielded the highest crude estimates of HIV-1 infection (42.9 per 100 women and 20.0 per 100 women, respectively). Sector 202, a rural sector, also yielded a high crude estimate of HIV-1 prevalence (17.7 per 100 women) but this was likely influenced by high local sampling
variability due to the relatively large observed number of infected women (3) among only 17 women who presented at one of the antenatal clinics. Lastly, Sector 225, a peri-urban sector, also yielded a high crude estimate of HIV-1 prevalence (18.8 per 100 women) and also may be influenced by sampling variability.

**Model fitting.** We find the estimated odds ratios (OR) associated with risk factors to be similar across all three models, providing some consistency across the approaches (Table 1). The DIC for Model II (4019.9) and Model III (4016.0) are strikingly similar, with Model III providing slightly better fit. Thus, incorporating spatial random effects (Model III) offer marginally better model fit over a model with non-spatial random effects. We present the results (OR (95% CS)) from Model III below and compare them to the traditional (i.e. non-spatial) logistical regression results$^9$ (Table 1).

We observe significantly increased risk of HIV-1 infection among pregnant women who reported a history of STD (OR = 2.2, 95% CS = 1.8--2.6), ever having had sex to support herself (OR = 1.7, 95% CS = 1.2--2.2), or who reported male partner was circumcised (OR = 1.7, 95% CS = 1.3--2.3) (Table 1). Second, we find higher HIV prevalence among women who were not legally married and reported any sexual partner other than husband/regular partner and a higher monthly household income. These OR estimates (and corresponding 95% (CSs)) are similar to the logistic regression results. Furthermore, we observe an increased risk of HIV-1 infection among pregnant women who were infected with syphilis at the screening visit (OR = 1.7, 95% CS = 1.1--2.6): this covariate was not included in the logistic regression analysis$^9$. However, when controlling for all variables, a woman’s current age, age at first sexual intercourse and age at first pregnancy were not important risk factors for HIV-1 infection among this cohort of pregnant women.
Spatial analysis. The geographic patterns in the estimates of the random effects indicate where in the study area the model underestimate or overestimate the observed data, controlling for a woman’s monthly household income, any sexual partner(s) other than husband/regular partner, history of STD, having had sex to support herself, partner circumcision, syphilis infection, and marital status. The map provides a valuable visual diagnostic of local model fit, similar to a map or plot of model residuals in linear regression. Model I and Model II identify the same sectors with extreme (i.e. statistically different from zero) random effects, controlling for the covariate effects in the respective model, with the exception of Sector 222 (Table 2). Model III identifies the same sectors with extreme and non-extreme random effects as Model I, controlling for the covariate effects in the model, but also identifies the random effects for Sector 0, Sector 105 and Sector 187 to be statistically different from zero (i.e. 95% CS does not contain zero) due to their proximity to sectors with extreme values. The random effects reveal that all models considered under-predict the outcome in all of the urban and peri-urban sectors (Figure 2). The models appeared to have over- and under-predict the outcome in an equal number of semi-rural sectors (Figure 2). Lastly, the models over-predict the outcome in a majority of the rural, or outlying, sectors (Figure 2).

Graphical representations illustrate that the model tended to under-predict the outcome in the center of the study region, or within the urban and peri-urban sectors, and over-predict the outcome in the outlying rural and semi-rural sectors (Figure 3). The sectors highlighted in Figure 3 represent those with a random effect significantly different from zero. The rural and semi-rural sectors with an extreme random effect (e.g. Sector 167) appear in the southern region of the study region and sectors surrounding the urban and peri-urban sectors (Figures 3(a) and 3(b)). Figure 3(c) result from local, rather than global, random effect adjustments. We observe that the
estimate for Sector 222 is smoothed, or drawn upward, in Figure 3(c) as compared to Figure 3(b). In fact, this sector is adjacent to one urban sector, two peri-urban sectors, and two semi-rural sectors each with higher crude estimates of HIV-1 prevalence.

The smoothed means (Model I) and posterior medians (Model III) of the SMRs of HIV-1 for each respective sector, controlling for the covariate effects in the model, illustrate slightly different geographical results (Figure 4). The estimates of the SMRs from Model I are smoothed toward a global mean with the majority of sectors possessing an SMR in the range of 0.60--1.20 (Figure 4(a)). A few rural and semi-rural sectors in the northern and northwestern periphery have relatively high SMRs (1.20--1.80) indicating possible ‘local clustering’ of HIV-1 infections (Figure 4(a)).

More SMR estimates are drawn toward a local mean in the southwestern, eastern and northeastern portion of the geographical region (e.g. Sector 202, Sector 224, and Sector 105) with more sectors in the center of the geographical region are drawn toward a more elevated SMR (e.g. estimates in Figure 4(b) are from Model III). We observe more ‘shrinkage’ in the SMR estimates (toward local estimates) when considering local spatial dependency in random effects (Figure 4(b)). The estimated SMRs are highest in the urban and surrounding sectors (e.g. Sector 01, Sector 185 and Sector 187) and decrease gradually toward the edges of the study region.

DISCUSSION

While this study does not depict the current pattern of the epidemic, the analysis demonstrates the use of spatial analysis methodology in HIV studies and highlights historical differences in the risk factors associated with HIV-1 prevalence among pregnant women in the mostly rural but densely populated area surrounding the town of Butare in Southern Rwanda.
Our spatial analysis of the cross-sectional survey data confirms the remarkable urban/rural HIV-1 prevalence gradient seen among pregnant women in a number of African countries early in the HIV epidemic\textsuperscript{4-8} and indicates that this gradient is not solely due to urban/rural differences in the sociobehavioral factors currently included in the model. We observe a high concentration of HIV-1 prevalence in the urban center with a gradual decrease as one moves toward the increasingly rural outer edges of the study area despite substantial interaction and movement between people who were living in this relatively small geographic region at the time of the study.

The crude estimates of HIV-1 prevalence do not appear constant over the entire study region, but this was difficult to assess because of such wide variation in the local estimates. Our analysis smooths these crude estimates by adding statistical stability to the local risk estimates, and clarifies the observed risk pattern\textsuperscript{18}. We observe high prevalence of HIV-1 infection in the urban areas with a much lower prevalence in the rural sectors located around the boundaries of the study region. Also, the spatial model smoothes the ‘hotspot’ located in the northeastern region of the crude estimate map toward its local mean. This locally elevated disease prevalence appears to be the result of sampling variation due to the small local sample size, but is stabilized when we ‘borrow strength’ from neighboring sectors\textsuperscript{17,18,20,21}.

Although we did not consider all of the risk factors in earlier studies, our results, for the most part, agree with the estimates from the non-spatial logistic regression analysis\textsuperscript{9}. We report an increased risk of HIV-1 infection for pregnant women who report never being legally married, an increased monthly household income, any sexual partner(s) other than husband/regular partner, history of STD, partner circumcision, having had sexual intercourse to support herself, and testing positive for syphilis at the screening visit. A notable difference includes our results
showing no significant association between HIV prevalence and a woman’s age at first pregnancy. An earlier study found an association (with a woman’s age at first pregnancy) and reports a higher HIV prevalence among women who reported their first pregnancy at less than 17 years of age\textsuperscript{9}. Our results do not necessarily contradict those previously reported\textsuperscript{9}, but indicate that this factor is no longer a significant predictor of HIV infection when we consider spatial effects.

The spatial results were similar across models as expected because of the structural similarities in models. In particular, the choice of distribution of the random effects did not appear to matter (e.g. spatially correlated random effects did not offer markedly better model fit). In this analysis, the primary advantage of our spatial analysis over a non-spatial analysis is the model-based approach to assess local model fit via maps of estimated random effects. Such maps reveal “residual” variation suggesting additional urban/rural variation unaccounted for by the demographic and behavioral covariates included in the model. The random effects identify sectors where the model had a tendency to over/under-predict the outcome, which may suggest further improvements to the model (i.e. regional covariates currently omitted from the model). In other word, the model results suggest that additional factors associated with decreasing risk in the urban areas or perhaps increasing risk in rural areas need to be considered. Possible additions include a woman’s ability to travel (i.e., mobility) within the study region, especially within the urban center where HIV prevalence is highest, which may increase the risk of HIV infection for women who report rural residency. In addition, a woman’s age at first pregnancy may increase prevalence of HIV infection for women in rural areas of the study region.

These spatial models serve as a baseline for exploring spatially-varying coefficients in order to assess if covariate effects are the same everywhere\textsuperscript{28} and if additional analyses would
benefit from including potential location-by-covariate interactions. Hierarchical Bayes methods offer an attractive method for such models and a mechanism to stabilize small-area estimates while maintaining a high level of geographic precision\textsuperscript{22}. This study documents substantial small-area variations in HIV prevalence among pregnant Rwandan women seeking care within a defined geographic area and emphasizes the importance of concentrating HIV prevention efforts especially on high transmission areas. The results reinforce conclusions that HIV prevention programs should continue unabated and special efforts are especially needed to better understand and possibly arrest the spread of the epidemic from urban to rural populations.
Acknowledgements: The authors would like to thank the National University of Rwanda, and all of the National University of Rwanda-Johns Hopkins University AIDS Project staff members who participated in study design and data collection. Unfortunately, many of them have died in the civil war and genocide of 1994. We gratefully acknowledge all of the women who participated in the study and especially acknowledge Mr. Jules Kajugiro for his efforts in coding the geographic areas of the participants. Drs. Alfred Saah, Phocas Habimana, and Donald Hoover provided invaluable assistance.

Financial support: DeMarc A. Hickson, at the time this study began, was a David and Lucille Packard Graduate Fellowship Scholar; Lance A. Waller is supported by the National Institutes of Environmental Health Sciences (LAW) - R01-ES07750 grant. Data collection for this study in Rwanda (1989-1994) was supported by the U.S. National Institute of Child Health and Human Development (HD22496 and HD25785).

Corresponding Author: Lance A. Waller, Department of Biostatistics, Rollins School of Public Health, Emory University 1518 Clifton Road, NE, Atlanta, GA 30333, Telephone: 404.727.1057, Fax: 404.727.1370, E-mail: lwaller@sph.emory.edu.
REFERENCES


APPENDIX

**Deviance Information Criterion.** The deviance information criterion (DIC) is a Bayesian model fit criterion that combines model fit and complexity. Intuitively, as the complexity of a model increases, so does the fit of the model; hence, there should be a tradeoff between the two\(^{29}\). The DIC is defined as a classical Bayesian measure of fit or ‘adequacy’ penalized by twice the number of effective parameters and can be written as

\[
\text{DIC} = D(\bar{\theta}) + 2p_D
\]

\[
= D(\theta) + p_D.
\]

In the latter equation, the DIC is a summary of how well the data fits the model \(D(\bar{\theta})\) and a penalty for the complexity of the model \(p_D\). \(p_D\), the number of effective parameters is written as

\[
p_D = D(\theta) - D(\bar{\theta}),
\]

where \(D(\theta) = -2 \log p(y|\theta) + 2 \log f(y)\) and is the Bayesian deviance. In practice, \(p_D\) is interpreted as the mean deviance \(D(\bar{\theta})\) minus the deviance of the means \(D(\theta)\) where the mean deviance is oftentimes referred to as a Bayesian measure of fit and the deviance of the means represents a classical ‘plug-in' measure of fit\(^{27}\).
Table 1: Prevalence odds ratios (OR) and 95% confidence interval (CI) and ‘credible set (CS)’ estimates from hierarchical models of HIV infection among pregnant women in the province of Butare, Rwanda, 1989--1993.

<table>
<thead>
<tr>
<th>Variable</th>
<th>Number of HIV+ women/Total number of women</th>
<th>Model I OR (95% CI*)</th>
<th>Model II OR (95% CS†)</th>
<th>Model III OR (95% CS†)</th>
<th>Logistic Regression OR (95% CI*)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Monthly household income (Rwandan franc)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>&lt;920</td>
<td>298 / 3,961</td>
<td>1.0</td>
<td>1.0</td>
<td>1.0</td>
<td>&lt;2500: 1.0</td>
</tr>
<tr>
<td>920 -- 4800</td>
<td>223 / 2,471</td>
<td>1.7 (1.3, 2.1)</td>
<td>1.8 (1.5, 2.4)</td>
<td>1.9 (1.5, 2.4)</td>
<td>≥2500: 2.5 (1.9, 3.1)</td>
</tr>
<tr>
<td>≥4800</td>
<td>176 / 1,022</td>
<td>2.2 (1.7, 2.9)</td>
<td>2.3 (1.9, 3.0)</td>
<td>2.3 (1.8, 2.9)</td>
<td></td>
</tr>
<tr>
<td>Any sex partner(s) other than husband/regular partner</td>
<td>265 / 1,194</td>
<td>1.9 (1.7, 2.1)</td>
<td>2.1 (1.7, 2.6)</td>
<td>2.1 (1.7, 2.6)</td>
<td>None: 1.0</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>1 Other: 1.6 (1.2, 2.1)</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>2+: 3.0 (2.1, 4.3)</td>
</tr>
<tr>
<td>History of STD</td>
<td>242 / 1,169</td>
<td>1.8 (1.6, 2.0)</td>
<td>2.2 (1.8, 2.7)</td>
<td>2.2 (1.8, 2.6)</td>
<td>2.3 (1.8, 2.9)</td>
</tr>
<tr>
<td>Partner circumcised</td>
<td>105 / 424</td>
<td>1.6 (1.4, 1.9)</td>
<td>1.8 (1.3, 2.3)</td>
<td>1.7 (1.3, 2.3)</td>
<td>2.1 (1.5, 2.9)</td>
</tr>
<tr>
<td>Has had sex to support herself</td>
<td>111 / 390</td>
<td>1.4 (1.2, 1.6)</td>
<td>1.7 (1.3, 2.3)</td>
<td>1.7 (1.2, 2.2)</td>
<td>1.6 (1.1, 2.3)</td>
</tr>
<tr>
<td>Syphilis (RPR+ at screening visit)</td>
<td>46 / 170</td>
<td>1.4 (1.2, 1.7)</td>
<td>1.7 (1.1, 2.6)</td>
<td>1.7 (1.1, 2.6)</td>
<td>Not Included</td>
</tr>
<tr>
<td>Marital Status (Married vs. Not Married)</td>
<td>140 / 2,927</td>
<td>2.0 (1.5, 3.1)</td>
<td>2.4 (1.9, 2.9)</td>
<td>2.3 (1.9, 2.9)</td>
<td>2.4 (1.9, 3.1)</td>
</tr>
</tbody>
</table>

* CI = Confidence Interval; † CS = Credible Set
Table 2: Coefficient estimates of the ‘extreme’ random effects and 95% confidence interval (CI) and ‘credible set (CS)’ from the hierarchical models fit to analyze risk factors for HIV infection among pregnant women in the province of Butare, Rwanda, 1989--1993.**

<table>
<thead>
<tr>
<th>Sector</th>
<th>Type</th>
<th>Model I $\theta_i$ (95% CI*)</th>
<th>Model II $\theta_i$ (95% CS†)</th>
<th>Model III $\theta_i$ (95% CS†)</th>
</tr>
</thead>
<tbody>
<tr>
<td>0</td>
<td>Urban</td>
<td>0.26 (-0.20, 0.73)‡</td>
<td>0.25 (-0.34, 0.86)‡</td>
<td>0.46 (0.03, 0.88)</td>
</tr>
<tr>
<td>1</td>
<td>Urban</td>
<td>0.88 (0.61, 1.15)</td>
<td>1.41 (1.02, 1.80)</td>
<td>1.51 (1.16, 1.88)</td>
</tr>
<tr>
<td>2</td>
<td>Urban</td>
<td>0.53 (0.23, 0.85)</td>
<td>0.60 (0.20, 1.03)</td>
<td>0.71 (0.32, 1.06)</td>
</tr>
<tr>
<td>3</td>
<td>Urban</td>
<td>0.38 (0.09, 0.67)</td>
<td>0.45 (0.10, 0.83)</td>
<td>0.50 (0.17, 0.82)</td>
</tr>
<tr>
<td>5</td>
<td>Urban</td>
<td>0.44 (0.18, 0.70)</td>
<td>0.58 (0.25, 0.92)</td>
<td>0.61 (0.31, 0.89)</td>
</tr>
<tr>
<td>112</td>
<td>Peri-Urban</td>
<td>0.55 (0.24, 0.86)</td>
<td>0.73 (0.34, 1.08)</td>
<td>0.72 (0.35, 1.07)</td>
</tr>
<tr>
<td>187</td>
<td>Peri-Urban</td>
<td>0.25 (-0.08, 0.57)‡</td>
<td>0.37 (-0.02, 0.77)‡</td>
<td>0.47 (0.15, 0.79)</td>
</tr>
<tr>
<td>221</td>
<td>Peri-Urban</td>
<td>0.36 (0.01, 0.70)</td>
<td>0.50 (0.07, 0.92)</td>
<td>0.51 (0.15, 0.88)</td>
</tr>
<tr>
<td>121</td>
<td>Semi-Rural</td>
<td>-0.59 (-1.09, -0.09)</td>
<td>-0.71 (-1.35, -0.15)</td>
<td>-0.59 (-1.18, -0.10)</td>
</tr>
<tr>
<td>222</td>
<td>Semi-Rural</td>
<td>-0.43 (-0.89, 0.04)‡</td>
<td>-0.54 (-1.15, -0.01)</td>
<td>-0.30 (-0.81, 0.16)‡</td>
</tr>
<tr>
<td>105</td>
<td>Rural</td>
<td>-0.46 (-1.08, 0.16)‡</td>
<td>-0.63 (-1.41, 0.08)‡</td>
<td>-0.76 (-1.54, -0.07)</td>
</tr>
<tr>
<td>167</td>
<td>Rural</td>
<td>-0.54 (-1.06, -0.01)</td>
<td>-0.67 (-1.38, -0.04)</td>
<td>-0.77 (-1.47, -0.13)</td>
</tr>
</tbody>
</table>

* CI = Confidence Interval
† CS = Credible Set
‡ Non-significant Random Effect
** Controlling for a woman’s current age, monthly household income, sexual partner(s) other than husband, history of STD, having had sex to support herself, and (woman reported) partner circumcision, syphilis infection, and marital status.
FIGURES

Figure 1(a): Small-area urban/rural assignment of sectors in the province of Butare, Rwanda, 1989-1993.

Figure 1(b): Crude HIV-1 prevalence among pregnant women in the province of Butare, Rwanda, 1989 – 93. Numerical labels indicate sectors of particular interest (See text).

Note: Maps are oriented in the traditional manner with North at the top.
Figure 2: Estimates of the random effects and corresponding 95% confidence intervals and credible sets while controlling for covariate effects, ordered and stratified by type of sector. Circle represents the random effect with a CAR prior distribution (Model III), triangle and diamond represent the random effects with an exchangeable prior (Model II and Model I, respectively).
Figure 3: Estimates of the random effects from the three hierarchical models. Two different distributions for the random effects. Map (a) illustrates the coefficient estimates from Model I; Map (b) illustrates the coefficient estimates from Model II; and Map (c) illustrates the coefficient estimates from Model III. Numerical labels indicate sectors of particular interest (See Table 2).

Note: Maps are oriented in the traditional manner with North at the top.
Figure 4: Estimates of the standardized morbidity ratio (SMR) of HIV-1 for pregnant women in the province of Butare, Rwanda, 1989–93. Map (a) illustrates the SMR estimates from Model I; Map (b) illustrates the SMR estimates from Model III. 
Note: Maps are oriented in the traditional manner with North at the top.