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, LANCE A. WALLER1, ANN CHAO2,3*, ABEL DUSHIMIMANA4,5*, MARC BULTERYS3*

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

ABSTRACT

We investigate the regional pattern of 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. The spatial patterns in the random effects suggested that modifications (e.g. omission or inclusion of covariates) to the spatial model within small areas were needed and the observed urban/rural gradient was not solely explained by the sociobehavioral covariates in the model. The local standardized morbidity ratios (SMRs) were highest in the urban center and declined toward the rural boundaries of the study region. The SMRs in the urban sectors were approximately a five-fold increase over the SMRs in the rural sectors. While the choice of the prior distribution of the random effects had little impact on results, more ‘shrinkage’ in the estimates of the SMRs was made when considering local spatial clustering.

Manuscript Word Count
Abstract 143
Article: 3,424
INTRODUCTION

According to the World Health Organization (WHO), the HIV epidemic entered into Rwanda in the late 1970s. By the mid-1980s, HIV-1 prevalence was found to be 88% among commercial sex workers, 29% among women of childbearing age in the capital city, Kigali, and 1.3% among a national sample of rural inhabitants. 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. Despite numerous HIV prevalence reports among various populations in sub-Saharan Africa, 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.

In this analysis, we explore 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 were two-fold: to provide statistically stable estimates of local small-area HIV-1 prevalence and to assess whether model output (specifically posterior estimates of local random effects) could identify potential local improvements to the model. Hierarchical Bayesian methods allowed us to investigate spatial patterns associated with HIV-1 infection in this mixed rural and urban population and suggest where (in the study area) the inclusion/exclusion of various covariates might be needed. A small-area spatial analysis provided several advantages: first, the hierarchical Bayes approach “borrowed strength” across the data set to improve small-area HIV-1 prevalence estimation based on small local sample sizes. Second, the small-area spatial analysis compensated for any residual spatial correlation in the
model thereby adjusting for potential bias in parameter estimation from non-independent observations\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\(^9\). 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\(^9,10\).

Here, we concentrate on understanding the small-area geographical differences in risk factors among this population. Geographically, 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 focuses 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 allowed us to ‘borrow strength’ from both neighboring regions and the entire geographical region in order to stabilize estimates within small areas while maintaining geographic resolution. Empirical Bayes methods were first introduced to estimate area-specific relative risk\textsuperscript{17}. These methods were then extended to a fully Bayesian setting\textsuperscript{18} and to likelihood-based approximations within the context of generalized linear mixed models (GLMMs)\textsuperscript{19}. Methodological details have been described in detail elsewhere\textsuperscript{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 \(E_i\) represented the multiplicative estimate based on the number of pregnant women that sought antenatal care from sector \(i\) and the overall estimate of HIV prevalence among study subjects. The first-stage model was denoted as

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

where \(\eta_i\) was 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\textsuperscript{19} was written as

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

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

In this analysis, we evaluated the effects of ten risk factors on HIV-1 infection. These factors included 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 in past 5 years, engaged in sexual intercourse to support herself, or was currently infected with syphilis (positive syphilis serology by RPR). These risk factors were chosen for this analysis because of their previously described association with HIV infection among pregnant women in Rwanda and elsewhere in Africa.\textsuperscript{9-11} A woman’s age at first sexual intercourse and age at first pregnancy were coded in increasing years. A woman’s current age and monthly household income were considered as categorical: age (<19, 20--24, 25--29, and 30+) and income (<920, 920--4800, and 4800+ Rwandan Francs). Marital status was 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 were all coded yes or no. Risk factors that were not found to be significant (based on the coefficient estimate and 95% confidence interval/credible set) were removed from the model and any subsequent analyses.

\textit{Second-stage model.} The vector of model parameters $\beta$ was assigned a uniform prior distribution. We assumed the covariate effects were constant across the study area and captured any residual geographic correlation through the collection of random effects, $\theta$. We defined $\theta$ alternatively with both an exchangeable and a spatially correlated prior structure\textsuperscript{17,18,21,22}. First, we assigned $\theta$ a (exchangeable) Gaussian prior structure written as

$$\theta_i \sim N(0, \psi), \quad i = 1, \ldots, 54$$

(4)

where $\psi$ is the shared prior variance\textsuperscript{18}. This prior structure sampled each random effect from the same underlying distribution; implying similarities existed among all locations despite ‘apparent’
geographical differences. This resulted in ‘shrinkage’ of local rate estimates toward the global observed rate.

Next, we assigned $\Theta$ a conditionally autoregressive (CAR) Gaussian prior structure. This prior took into account the local ‘clustering’ of risk among neighboring regions and provided a spatial dependency commonly used in disease mapping\textsuperscript{17,20–22}. One formulation of this model can be given as

$$
\theta_i \mid \theta_{j,i \neq i} \sim N \left( \frac{\sum_{j \neq i} w_{ij} \theta_j}{\sum_{j \neq i} w_{ij} + \tau \sum_{j \neq i} w_{ij}}, \frac{1}{\tau \sum_{j \neq i} w_{ij}} \right), \tag{5}
$$

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

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

$$
\frac{1}{\omega} \sim \text{Gamma}(\alpha, \delta), \tag{6}
$$

where $\omega = \psi$ or $\tau$. The chosen parametric model served as either an informative or weakly informative prior and we explored various options for values of $\alpha$ and $\delta$\textsuperscript{21–23}.

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

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 the different prior specifications for the random effects. The first MCMC model (Model II) incorporated exchangeable random effects while the second MCMC model (Model III) incorporated spatially correlated random effects. Convergence of the parameter estimates in Models II and III were evaluated by the Gelman-Rubin (G-R) statistic\textsuperscript{26}. There was no diagnostic tool in the statistical literature to make model comparisons across all three models, however, the deviance information criterion (DIC) allowed for the comparison of model fit between Models II and III (see Appendix)\textsuperscript{27}.

**Missing data.** Small-area spatial residency information was of particular interest because it allowed us to identify regional model fit patterns and map HIV-1 seroprevalence estimates. A woman missing residency information was not included in this analysis. Second, an independent draw from the MCMC posterior predictive distribution provided imputation of missing covariate data for a given individual. This allowed us to obtain a complete dataset to fit the hierarchical models in both SAS and WinBUGS.
Computation of OR, SMR, CI, and CS. An odds ratio (OR) for each covariate effect along with a standardized morbidity ratio (SMR) associated with residence in sector $i$ were calculated. The coefficient estimate of each respective covariate effect from the three models above were exponentiated in order to compute an odds ratio. The sector-specific SMRs were calculated by exponentiating the coefficient estimate of the respective random effect. Confidence intervals (CI) were computed for estimates obtained from SAS and credible sets (CS) were computed for estimates obtained from WinBUGS. The 95% CIs were computed by adding and subtracting the multiplicative result of the coefficient’s standard deviation and $Z$-score associated with a 5% significance level from the coefficient estimate of the covariate and random effects. The 95% CSs were computed by extracting the 2.5th and 97.5th percentile estimates of 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 that resided in the urban sectors with the lowest prevalence (3.6%) observed among women that 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, too yielded a high crude estimate of HIV-1 prevalence (18.8 per 100 women) and too may be influence by sampling variability. The estimates of HIV-1 prevalence did not appear constant over the entire map, but this was difficult to assess because of such wide variation in the local estimates.

**Model fitting.** The odds ratios (OR) were similar across all three models, providing some consistency across the three hierarchical models (Table 1). The DIC for Model II (4061.3) and Model III (4058.3) were strikingly similar. By definition of the DIC, Model III fit the data slightly better. Thus, the spatial adjustments to the random effects (Model III) offered marginally better model fit and the results (OR (95% CS)) presented below are from Model III (Table 1).

There was an increased risk of HIV-1 infection among pregnant women who reported a history of STD (OR = 2.2, 95% CS = 1.8--2.7), ever having had sex to support herself (OR = 1.8, 95% CS = 1.3--2.3), were infected with syphilis (OR = 1.7, 95% CS = 1.1-2.6), or who reported male partner was circumcised (OR = 1.8, 95% CS = 1.4--2.3) (Table 1). Second, married women and women who reported a lower monthly household income were at a lower risk for HIV-1 infection. Furthermore, women who were less than 30 years of age were more likely to be infected (<20: OR = 1.2, 95% CS = 0.8--1.7; 20--24: OR = 1.8, 95% CS = 1.4--2.3; 25--29: OR = 2.0, 95% CS = 1.6--2.5, respectively) with the highest odds observed among women between the ages of 25 and 29 years.
Spatial analysis. The geographic patterns in the estimates of the random effects indicated where in the study area the model underestimated or overestimated the observed data, controlling for a woman’s current age, monthly household income, any sexual partner(s) other than husband, history of STD, having had sex to support herself, partner circumcision, syphilis infection, and marital status. Positive estimates of $\theta_i$ corresponded to under-prediction of the outcome while negative estimates corresponded to over-prediction of the outcome. Model I and Model II identified 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 identified the same sectors with extreme and non-extreme random effects as Model I, controlling for the covariate effects in the model, but also estimated 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). Furthermore, the models under-predicted the outcome in all of the urban and peri-urban sectors (Figure 2). The models appeared to have over- and under-predicted the outcome in an equal number of semi-rural sectors (Figure 2). Lastly, the models over-predicted the outcome in a majority of the rural sectors (Figure 2).

Graphical representations illustrated 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 that was significantly different from zero. The rural and semi-rural sectors with an extreme random effect (e.g. Sector 167) were located 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) resulted from local, rather than global, random effect adjustments. The estimate for Sector 222 was smoothed, or drawn upward, in Figure 3(c) as compared to that in
Figure 3(b). In fact, this sector was adjacent to one urban sector, two peri-urban sectors, and two semi-rural sectors with higher crude estimates of HIV-1 prevalence. There was supportive evidence of spatial smoothing with several urban and peri-urban sectors being smoothed either upward or semi-rural and rural sectors being smoothed downward to more closely mirror rates in neighboring sectors (Figure 3(c)).

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, illustrated 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 were 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 being drawn toward a more elevated SMR (e.g. estimates in Figure 4(b) are from Model III). More ‘shrinkage’ in the SMR estimates (toward local estimates) was observed when considering local spatial dependency (Figure 4(b)). The SMRs were the highest in the urban and surrounding sectors (e.g. Sector 01, Sector 185 and Sector 187) and decreased gradually toward the edges of the study region.

DISCUSSION

This study investigated the small-area patterns of sociodemographic and sexual history risk factors associated with HIV-1 seroprevalence among pregnant women in the mostly rural but densely populated area surrounding the town of Butare in Southern Rwanda. Analysis of the
cross-sectional survey data confirmed the remarkable urban/rural HIV-1 prevalence gradient seen among pregnant women in a number of African countries early in the HIV epidemic. We found a high concentration of HIV-1 seroprevalence in the urban center with a gradual decrease as one moved 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 general patterns in the smoothed maps clarify patterns observed in the crude values, with added statistical stability to the local risk estimates. We found that the SMRs were highest in the urban areas and declined in the rural sectors located around the boundaries of the study region. In addition, the highest SMRs generally corresponded with the highest crude rates. Furthermore, the ‘hotspot’ located in the northeastern region of the crude estimate map was smoothed toward its local mean and appeared to be the result of sampling variation due to the small local sample size. In addition to the maps of local estimated risk ratios (SMRs), we also examined local model fit via maps of estimated random effects. Such maps revealed “residual” variation suggesting additional urban/rural variation unaccounted for by the demographic and behavioral covariates included in the model.

The random effects identified sectors where the model had a tendency to over/under-predict the outcome, which may suggest the need for further improvements to the model. The small-area spatial patterns in the random effects suggested possible regional covariates omitted from the model (e.g., additional factors associated with decreasing risk in the urban areas or perhaps increasing risk in rural areas) and that the observed urban/rural gradient was not solely due to the behavioral covariates included in the model.
Our investigation of the regional/small area patterns associated with HIV-1 among this cohort of pregnant women allowed us to investigate the GLMM PQL and hierarchical Bayesian model fitting techniques and different approaches to “borrowing strength17,18,20,21.” Although the results were not identical, results were similar across models, as expected due to the structural similarity of the underlying models. In particular, while 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), broader trends remained.

In conclusion, hierarchical Bayes methods offer an attractive method to stabilize small-area estimates while maintaining a high level of geographic precision22. In addition, these models could serve as a baseline for exploring spatially-varying coefficients in order to assess if covariate effects are the same everywhere28 and if additional analyses would benefit from including potential location-by-covariate interactions. This study also documents very 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. Furthermore, our results have implications for HIV sentinel surveillance among pregnant women, as sentinel sites are often located in high transmission areas29. Noteworthy declines in HIV prevalence have recently been reported from a number of sub-Saharan African countries, including urban Rwanda30,31. However, HIV prevention programs should continue unabated and special efforts are especially needed to arrest the epidemic in 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 is 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).

Authors’ addresses: DeMarc A. Hickson, MS, PhD, Department of Biostatistics, Rollins School of Public Health, Emory University 1518 Clifton Road, NE, Atlanta, GA 30333, Telephone: 404.712.9575, Fax: 404.727.1370, email: dhickso@sph.emory.edu. Lance A. Waller, PhD, 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. Ann Chao, PhD, CDC-Zambia, US Embassy, P.O. Box 31617, Lusaka, Zambia; Telephone: (260) 1-250-955, Fax: (260) 1-251-142, Email: chaoa@cdczm.org; Dr Abel Dushimimana, School of Public Health, B.P. 5229 Kigali, Rwanda; and Marc Bulterys, MD, PhD, Director, CDC-Zambia, US Embassy, P.O. Box 31617, Lusaka, Zambia; Telephone: (260) 1-250-955, Fax: (260) 1-251-142, Email: bulterysm@cdczm.org.
Reprint requests: 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. 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

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

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

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

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 \( D(\bar{\theta}) \) 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.
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>Coefficient</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>
</tr>
</thead>
<tbody>
<tr>
<td>Current age (years)</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>&lt; 20</td>
<td>49 / 448</td>
<td>1.27 (0.95, 1.58)</td>
<td>1.17 (0.80, 1.69)</td>
<td>1.19 (0.81, 1.74)</td>
</tr>
<tr>
<td>20 -- 24</td>
<td>257 / 2,072</td>
<td>1.68 (1.47, 1.89)</td>
<td>1.75 (1.38, 2.22)</td>
<td>1.78 (1.42, 2.33)</td>
</tr>
<tr>
<td>25 -- 29</td>
<td>287 / 2,645</td>
<td>1.80 (1.07, 1.47)</td>
<td>1.96 (1.55, 2.50)</td>
<td>1.98 (1.59, 2.54)</td>
</tr>
<tr>
<td>30 +</td>
<td>111 / 2,279</td>
<td>1.0</td>
<td>1.0</td>
<td>1.0</td>
</tr>
<tr>
<td>Monthly household income</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>(Rwandan franc)</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>&lt; 920</td>
<td>298 / 3,961</td>
<td>0.50 (0.31, 0.68)</td>
<td>0.44 (0.35, 0.57)</td>
<td>0.46 (0.37, 0.57)</td>
</tr>
<tr>
<td>920 -- 4800</td>
<td>223 / 2,471</td>
<td>0.61 (0.43, 0.80)</td>
<td>0.53 (0.42, 0.69)</td>
<td>0.55 (0.44, 0.68)</td>
</tr>
<tr>
<td>4800 +</td>
<td>176 / 1,022</td>
<td>1.0</td>
<td>1.0</td>
<td>1.0</td>
</tr>
<tr>
<td>Any sex partner(s) other than husband</td>
<td>265 / 1,194</td>
<td>1.88 (1.72, 2.05)</td>
<td>2.13 (1.73, 2.61)</td>
<td>2.14 (1.72, 2.61)</td>
</tr>
<tr>
<td>History of STD</td>
<td>242 / 1,169</td>
<td>1.79 (1.64, 1.96)</td>
<td>2.21 (1.79, 2.71)</td>
<td>2.20 (1.78, 2.66)</td>
</tr>
<tr>
<td>Partner circumcised</td>
<td>105 / 424</td>
<td>1.64 (1.43, 1.84)</td>
<td>1.80 (1.38, 2.35)</td>
<td>1.77 (1.35, 2.29)</td>
</tr>
<tr>
<td>Has had sex to support herself</td>
<td>111 / 390</td>
<td>1.37 (1.15, 1.59)</td>
<td>1.75 (1.32, 2.36)</td>
<td>1.75 (1.31, 2.33)</td>
</tr>
<tr>
<td>Syphilis (RPR+ at screening visit)</td>
<td>46 / 170</td>
<td>1.44 (1.15, 1.74)</td>
<td>1.73 (1.14, 2.57)</td>
<td>1.73 (1.13, 2.60)</td>
</tr>
<tr>
<td>Marital Status (Married vs. Not Married)</td>
<td>140 / 2,927</td>
<td>0.50 (0.32, 0.68)</td>
<td>0.45 (0.36, 0.56)</td>
<td>0.46 (0.37, 0.56)</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.
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.