Correlates of Sexually Transmitted Infections Among South African Women Using Individual- and Community-Level Factors: Results from Generalized Additive Mixed Models

Handan Wand1 · Natashia Morris2 · Reshmi Dassaye3 · Tarylee Reddy2 · Gita Ramjee2

Received: 30 November 2017 / Revised: 29 June 2018 / Accepted: 21 September 2018 / Published online: 14 February 2019
© Springer Science+Business Media, LLC, part of Springer Nature 2019

Abstract
South Africa has the highest burden of human immunodeficiency virus (HIV) infections in the world. There is also growing evidence that an individual’s risk of contracting HIV is increased by the presence of other sexually transmitted infections (STIs). The primary objective of this study was to examine the association between the prevalence of STIs in a cohort of South African women who enrolled in HIV prevention trials (2002–2012). The current study linked the individual factors with the community-level characteristics using geo-referencing. These multi-level data were analyzed in generalized additive mixed models settings. In the multivariate logistic regression model, younger age (odds ratio [OR] 4.30, 95% CI 3.20, 5.77 and OR 2.72, 95% CI 2.02, 3.66 for age < 25 and 25–29, respectively); being single/not cohabiting (OR 4.57, 95% CI 3.18, 6.53), two + sex partners (OR 1.46, 95% CI 1.18, 1.80); parity < 2 (OR 2.04, 95% CI 1.53, 2.72), parity = 2 (OR 1.85, 95% CI 1.37, 2.48), and using injectables (contraceptive) (OR 1.53, 95% CI 1.13, 2.06) were all significantly associated with increased prevalence of STIs. Women who resided in the communities with high proportions of female headed-households were also significantly at higher risk for STIs (OR 1.20, p = .0025). Because these factors may reflect characteristics of the larger groups who share similar cultural norms and social environments, they can provide considerable insight into the spread of STIs. Prevention strategies based on individual and community-level drivers of STIs are likely to be the most effective means of targeting and reaching those at greatest risk of infection. This strategy has the potential to play a significant role in the epidemic’s trajectory.

Keywords Generalized additive models · Individual-level and population-level risk factors · Sexually transmitted infection

Introduction
South Africa has the highest burden of human immunodeficiency virus (HIV) infections in the world, with more than 6 million people estimated to be infected (UNAIDS, 2016). Over the years, the HIV epidemic has been shown to have multiple individual-level drivers including certain socio-demographic characteristics and high-risk sexual behaviors (Tanser, Oliveira, Maheu-Giroux, & Bärnighausen, 2014). There is also growing evidence that an individual’s risk of contracting HIV is increased by the presence of other sexually transmitted infections (STIs) such as chlamydia, gonorrhrea and syphilis (Laga et al., 1993; Leon, Naidoo, Mathews, Lewin, & Lombard, 2010; Martin et al., 1999; McClelland et al., 2008). In addition to strong biological associations between these pathogens, they also share common high-risk sexual behaviors (McClelland et al., 2007; Zuma et al., 2005). Therefore, STI testing and treatment strategies are commonly included in HIV prevention packages (Skoler-Karpoff et al., 2008). KwaZulu-Natal is the most densely populated province in South Africa, accounting for 21% of the population, where 59% of them are sexually active. The HIV prevalence is estimated to be 40% among women who attended antenatal clinics; while prevalence of STIs is estimated to be more than 20% (Health NDo, 2013; Ramjee & Wand, 2014). These high rates contribute toward rising morbidity levels among young women in the region.

As the prevention still remains the primary strategy for controlling the epidemic, identifying and targeting
individuals who are at high risk of infection is crucial and a research priority. Concordant with most of the published literature, we previously focussed on demographic and socioeconomic factors as well as commonly collected sexual behaviors which are known to be associated with the infections (Wand et al., 2018). While these individual-level measurements may identify those at high risk of infections, they can only be relevant and have implications to the study population under investigation. Therefore, identifying community-level factors along with the individual-level factors can have a profound influence on our understanding of the spread of the STIs. These factors may reflect characteristics of larger groups who share similar cultural norms, religious beliefs and social environments at population level. Prevention strategies based on individual and community-level drivers of sexually transmitted infections are likely to be the most effective means of targeting communities as well as individuals at greatest risk of infection. Together, they will provide considerable insight into the epidemic by assisting to identify and reach the key populations in high-risk geographical areas.

The primary objective of this study was to examine the association between the prevalence of STIs and the individual as well as community-level factors in a cohort of women who enrolled in five HIV biomedical intervention trials from KwaZulu-Natal during the period of 2002–2012. We hypothesized that community-level factors may have significant impact on shaping the individual-level characteristics which have been shown to be linked to high rates of STI diagnosis. As the poor socioeconomic characteristics continue to have significant impact on increasing levels of HIV and STI diagnosis, identifying evidence-based benchmarks remains research priority. Consequently, results from the current study may guide policymakers on how to weigh all the empirical thresholds and their goals for reducing STIs to create cost-effective policies. This is particularly crucial to evaluate effectiveness of interventions to prevent new infections at the individual and community level.

We used generalized additive mixed models (GAMMs) in order to investigate significant features of the data with minimum statistical assumptions after accounting for potential geo-spatial correlations (Wood, 2004, 2006). These flexible modern statistical techniques can handle complex and nonlinear associations between a response variable and risk factor(s) of interest without imposing the restrictive linearity assumption between them, while adjusting for potential confounders. In addition, we also demonstrated how these models may be used in order to identify the significant threshold points of the community-level factors for STI diagnosis. This is particularly useful when the risk of a disease does not increase (or decrease) monotonically, and/or there is no clinically/epidemiologically relevant cut point(s).

Results from our study might expand and bring greater insights to the previous research in several ways, by including commonly measured individual risk factors as well as community-level factors which can be surrogate for unmeasured factors such as social/cultural norms, religious beliefs, and sexual networks. This strategy has the potential to inform the development of the most efficient and cost-effective HIV/STI prevention programs which may subsequently change the trajectory of the epidemic in the region. To our knowledge, this is one of the first studies to investigate individual and larger population-level characteristics in relation to STI prevalence among a large cohort of women using modern and flexible statistical techniques which can handle complex and nonlinear associations between a response and the predictors.

**Method**

**Participants**

The data used for this sub-study are drawn from a set of study sites located in the greater Durban area of KwaZulu-Natal, South Africa. Five trials included in these analyses were conducted between 2002 and 2012 at 6 sites, located at Verulam, Tongaat, Chatsworth, Umkomaas, Botha’s Hill, and Isipingo. Briefly, 12% (n = 781) of the study population were women who enrolled in the Methods for Improving Reproductive Health in Africa trial which tested the efficacy of diaphragm for HIV prevention (Padian et al., 2007); 21% (n = 1344) of the study population were women who enrolled in the Carraguard trial, a Phase 3, randomized, placebo-controlled, double-blind trial, which was designed to investigate the efficacy and long-term safety of a carrageenan-based gel formulation developed by the Population Council (New York, NY) for the prevention of HIV infection (Skoler-Karpoff et al., 2008); 26% (n = 1648) of the study population were women who enrolled in the Micobicide Development Programme 301 (McCormack et al., 2010) which was a Phase 3, randomized, double-blind, parallel-group trial; 11% (n = 635) of the women who enrolled in the HIV Prevention Trial Network 035 which was designed to evaluate the safety and effectiveness of two vaginal microbicides, Buffer Gel and PRO 2000/2005, in preventing the transmission of HIV (Abdool Karim et al., 2011); 30% (n = 2076) of the women were those who enrolled in the Vaginal and Oral Interventions to Control the Epidemic trial which was a randomized placebo-controlled trial that assessed the safety and effectiveness of oral tenofovir disoprophil fumarate (TDF), oral TDF plus emtricitabine (TDF-FTC), and vaginal 1% tenofovir (TFV) gel for HIV-1 prevention in women (Marruzzo et al., 2015). An interviewer-administered questionnaire on demographics and sexual behavior was completed. Participants received pretest counseling before HIV and STI testing. HIV diagnostic testing was conducted using two different rapid tests (Determine HIV-1/2 [Abbott Laboratories, Tokyo, Japan]...
and Oraquick [Orasure Technologies, Bethlehem, PA, USA]) on whole blood from either finger-prick or venepuncture. The current study only included data from 6,484 women who had geographical information through their residential addresses. Site specific sub-data were extracted and combined across the consistently measured characteristics; results were reported as overall and not study-specific.

**Primary Outcome Measure**

As part of the study protocols, women were tested for sexually transmitted infections including chlamydia, gonorrhea and syphilis at baseline. Individuals were classified as STI positive at baseline if they had at least one positive STI test result (i.e., chlamydia, gonorrhea or syphilis). Participants who tested positive for STIs were treated as per study protocols and local guidelines, free of charge.

**Individual-Level Measurements**

Individual-level data (from the combined trial populations) were geo-located at their place of residence (or nearest location point to residence) using global positioning system (GPS) coordinate data during enrollment of GPS satellite coordinate data. Our study considered a variety of individual-level factors if they were collected consistently in all five trials: women's age (< 25 years, 25–29 years and 30 + years), marital and/or cohabitation status (married/cohabiting vs. not), sex partner has another partner (no vs. yes, not sure), level of education (no education vs. some education); number of sexual partners in past three months (< 2 vs. 2 +); parity (null/prim parity (< 2 births), multi-parity: 2 births, multi-parity: 3 + birth), condom used in last sex (no vs. yes), contraceptive method (injectables, oral pills vs. others), and age at sexual debut (< 16 years, 16–19 years vs. 20 years). All these factors varied at Level 1 (i.e., community) and Level 2 (i.e., individual). Other factors considered but not available in all five trials included: language spoken at home, employment status (or regular income), partner’s circumcision status, and average number of sexual acts in past week.

**Community-Level Measurements**

We also considered a variety of community-level characteristics of the geographical areas where the clinical sites were located. Community-level data from the 2011 South African Census Statistics recruitment area boundaries were developed using census delineations. A total of 43 recruitment areas were developed consolidating participant data across all studies.

This information was extracted and linked to the individual level combined trial data through the geographical components (i.e., latitude and longitude) (i.e., spatial aggregation). As shown in Table 1, we have focused on several community-level characteristics. These particular factors have been selected because of their potential to impact the epidemic directly or indirectly. For example, “% population female aged 18–25 years” and “% adult population unmarried/not cohabiting” have the potential to be surrogate for partnership and sexual networking; while “% female-headed households” “% population unemployed,” and “% population with no schooling” may reflect socioeconomic conditions in the region. We have also calculated age-standardized HIV incidence rates during the trials. Although these rates were calculated using the combined individual-level data, final estimates were aggregated by the geographical units. Therefore, we were able to use these rates as a community-level estimate. The following distal, proximal and structural characteristics have also been considered but were not included in the current study due to lack of associations with the primary outcome of interest: “average distance to a hospital,” “average distance to a

**Table 1** Community- and individual-level measurements collected

| Community-level factors a | Individual-level factors b |
|---------------------------|---------------------------|
| % population with:        | Study participants:       |
| Females 18-25 years old   | Diagnosed with STIs (response) |
| Adults single/not-cohabiting | Age (years) |
| Female headed households   | Education                |
| HIV incidence rates c     | Age at sexual debut       |
| Unemployed                | Single/not cohabiting     |
| No education              | Number of sex partners d  |
|                           | Partner has another partner|
|                           | Contraceptive use/type    |
|                           | Parity                    |

a Other factors considered but not included in the current analysis are: average distance to a hospital; average distance to a clinic; % population without pipe/running water; % households (HH) with no electricity; %HH traditional informal living; bCondom used in last sex; cAge standardized; dPast 3 months
Geographical Data

Individual-level data (from the combined trial populations) were geo-located at individuals’ place of residence (or nearest location point to residence) using GPS coordinate data during enrollment of GPS satellite coordinate data (i.e., latitude, longitude) and were collected using Garmin™ Nüvi (model no: 2360) handheld devices, downloaded into a Microsoft Access 2013 database and plotted spatially using the ArcGIS software package (version 10.4, CA), while for the community-level data extracted from the 2011 South African Census Statistics, recruitment area boundaries were developed using census delineations. A total of 43 recruitment areas (average of 157) were developed consolidating participant data across all studies.

All the spatially aggregated community-level factors were linked to the individual-level combined trial data to create multi-level data structure. Therefore, individual-level measurements (including the response variable) had a two-level structure from 43 geographical locations (Level 1) with 6484 observations at the participant level (Level 2). Therefore, they all varied at Level 1 and Level 2, while community factors varied at Level 1 (i.e., at group units) only (Table 1).

Statistical Analysis

We conducted statistical analysis in several steps: first, we used GAMMs to investigate the potential nonlinear associations between various geographical-level characteristics and STI diagnosis. This part of the analysis was exploratory in nature and conducted in several steps. All the measurements were considered as continuous variables and included in the models as a smooth function of splines one by one after accounting for geographical clustering (Wood, 2006); GAMMs were constructed using the “logit” link function for the ith group (i.e., geographical unit) (Level 1) and jth individual measurement (Level 2) (Wood, 2006):

\[ \logit(P(\text{STI}_{ij} = 1)) = s(x_i) + U_{ij} + \varepsilon_{ij} \]

where STI_{ij} corresponds to the STI status of the ith individual at jth geographical location; and

\[ s(x_i) = \beta_0 x + \sum_{k=1}^{K} \phi_k z_k(x_i) \]

is a penalized spline function for each community-level measurements (x_i) (for the ith geographical-level measurement); z_1(.), …, z_K(.) is a set of spline function and \( \sigma^2 \) is the penalization of the spline coefficients \( s_1(.) \ldots, s_k(.) \) and 

\[ U_{ij} \sim N(0, \sigma^2_u), \text{ random intercept due to the geographical clustering and model error: } \varepsilon_{ij} \sim N(0, \sigma^2). \]

Evidence for nonlinear associations was inferred using estimated “degrees of freedom” (“df”) and their respective p values. The “df” can be interpreted as the number of independent pieces of information in model fitting and can be considered as the total impact of all observations. Therefore, a higher “df” provides evidence for stronger influence and departure from linearity between a response variable and a predictor. After fitting the GAMMs for each measurement, we visually examined significant cut points based on the curves from the fitted models. The primary objective of this aspect of the analysis was to investigate the degree to which community-level factors are associated with increased/decreased prevalence of STIs. Visual and quantitative assessments are presented in Fig. 1a and Table 2a–f, respectively (Chaudhuri & Marron, 1999).

Model performances were assessed using percent variations explained by the models (i.e., intra-class correlation coefficient [ICC]). Adjusted odds ratios (aORs) and 95% confidence intervals (CIs) were reported for each individual and community-level factors identified as significant (<.05) in the final model.

We used the software package Multivariate Generalized Cross Validation (mgcv) (R 3.13 (http://cran.r-project.org/)) (Wood, 2006) with a “logit” link function as appropriate for our “binary response variable” (i.e., presence of STI vs. not).

Ethical Statement

All protocols and informed consent forms were approved by the Biomedical Research Ethics Committee at the University of KwaZulu-Natal in Durban.

Results

The current study included 6484 data points across 43 geographical locations. Overall, 45% of the study population was younger than 25 years of age, just over 50% of the study population did not have any schooling, while > 80% had their first sexual debut when they were younger than 20 years of age. In terms of the partnership factors, 18% of the women reported that their partner had another sex partner, while more than 50% of them were not sure if their partner had another partner or not. The vast majority of the study population (86%) was either unmarried and/or not cohabiting with their sexual partners, while 13% of the women reported at least two or more sexual partners in the past three months. Approximately one-third of the study population reported that they did not use a condom during their last vaginal sex encounter. Forty-five percent of the women had null or prim parity. Injectable were the most commonly (51%) used contraceptive method at baseline.
Fig. 1  a Smooth function of % female population 18–25 for STI diagnosis.  

b Smooth function of % adults single/not cohabiting for STI diagnosis.  
c Smooth function of % female-headed households for STI diagnosis.  
d Smooth function of age-standardized HIV incidence rate for STI diagnosis.  
e Smooth function of % population unemployed for STI diagnosis.  
f Smooth function of % Population no education for STI diagnosis.
Community-Level Characteristics

Results from the GAMMs for each community-level factor and STI diagnosis when they were fitted as continuous and dichotomized are presented in Table 2a–f. Their visual assessments with smoothing splines are depicted in Fig. 1a–f.

The association between STI diagnosis and population with “% population female aged 18–25 years old” is plotted in Fig. 1a. According to this, STI prevalence was increased among women who resided in the geographical locations where more than 8% of the population was estimated to be young females aged 18–25 years old. The “df” for this association was 8.351 ($p < .001$) and provided further evidence for this nonlinear association. The significant zero-crossing methodology identified “8%” to be the crucial cut point, and it was used to dichotomise this factor. After accounting for any geographical clustering effects, women who resided in the areas where more than 8% of the population was estimated to be young females were 25% more likely to be STI positive (OR 1.25, 95% CI 1.08, 1.45, $p < .001$) (Table 2a).

Prevalence of STIs was also associated with the levels of “% adults unmarried/not cohabiting” (Fig. 1b). Estimated “df” of 8.93 ($p < .001$) provided further evidence for strong nonlinear influence of this characteristic on STI diagnosis rates (Table 2b). Below/above 70% was identified as the significant cut point and was used to dichotomise this community-level characteristic. Results from this analysis indicated that women who resided in areas where more than 70% of the population was estimated to be unmarried/not cohabiting were more likely to be diagnosed with STIs (OR 1.83, 95% CI 1.05, 3.18, $p = .032$).

Although it was not as strong as the previous two community-level characteristics, prevalence of STIs was also associated with the levels of “% female-headed households (HH)” (Fig. 1c) with estimated “df” of 4.317, ($p = .0065$); women who resided in geographical units with higher proportions (≥ 40%) of study population estimated to have had female-headed households (HH) (OR 1.20, 95% CI 1.04, 1.40, $p = .018$) were also more likely to be diagnosed with STIs (Table 2c).

Age-standardized HIV incidence rates were also significantly associated with increased prevalence of STI diagnosis among women (Fig. 1d). This association was also identified as nonlinear with estimated “df” of 4.864, ($p < .001$) (Table 2d); 6 per 100 person-year (PY) was identified as the significant cut point, and it was used to dichotomize the HIV incidence rates across the geographical units. After re-fitting GAMM, women who enrolled from areas with high HIV incidence rates (≥ 6 per 100 PY) were also more likely to be diagnosed with STIs (OR 1.25, 95% CI 1.08, 1.45, $p = .003$) (Table 2d).

Although it was not statistically significant, women who resided in areas where more than 25% of the population was
estimated to be unemployed (or have no regular income) (with estimated “df” of 2.73, \( p = .119 \)) (Fig. 1e, Table 3e) were more likely to be diagnosed with STIs (OR 1.17, 95% CI 1.00, 1.37, \( p = .058 \)).

High STI prevalences were strongly associated with the geographical units with a high proportion of the population having no education (Fig. 1f). Degrees of freedom for this association was estimated at 1.429 (\( p = .343 \)) which provided

---

Table 2: Community-level factors: analyzing community-level characteristics using logistic regression and generalized additive models

| Community-level factors | Generalized additive mixed model |
|-------------------------|----------------------------------|
|                         | Degrees of freedom | \( p \) value | Random intercept\(^a\) (95% CI) |
| (a) % Female population 18–25 years | | | |
| Model\(^a\) | | | |
| Per 1% increase | 8.351 | \(< .001\) | 0.14 (0.05, 0.39) |
| Model 1\(^b\) | aOR (95% CI) | | |
| \(< 8\%) | 1 (reference) | | |
| \(8\% +\) | 1.25 (1.08, 1.45) | \(0.03\) |
| (b) % Adults unmarried/not cohabiting | | | |
| Model 2\(^a\) | | | |
| Per 1% increase | 8.935 | \(< .001\) | 0.25 (0.16, 38.4) |
| Model 2\(^b\) | aOR (95% CI) | | |
| \(< 70\%) | 1 (reference) | | |
| \(70\% +\) | 1.83 (1.05, 3.18) | \(0.032\) |
| (c) % HH female head of house 25–34 years | | | |
| Model 3\(^a\) | | | |
| Per 1% increase | 4.317 | \(0.005\) | 0.24 (0.15, 0.38) |
| Model 3\(^b\) | aOR (95% CI) | | |
| \(< 40\%) | 1 (reference) | | |
| \(40\% +\) | 1.20 (1.04, 1.40) | \(0.018\) |
| (d) Age-standardized HIV rate | | | |
| Model 4\(^a\) | | | |
| Per 1% increase | 4.864 | \(< .001\) | 0.23 (0.13, 0.37) |
| Model 4\(^b\) | aOR (95% CI) | | |
| \(< 6\ per 100 PY | 1 (reference) | | |
| \(6 +\ per 100 PY | 1.25 (1.08, 1.45) | \(0.003\) |
| (e) % Unemployed (not regular income) | | | |
| Model 5\(^a\) | | | |
| Per 1% increase | 2.73 | \(0.119\) | 0.24 (0.16, 0.38) |
| Model 5\(^b\) | aOR (95% CI) | | |
| \(< 25\%) | 1 (reference) | | |
| \(25\% +\) | 1.17 (1.00, 1.37) | \(0.057\) |
| (f) % No education | | | |
| Model 6\(^a\) | | | |
| Per 1% increase | 1.429 | \(0.002\) | 0.22 (0.14, 0.37) |
| Model 6\(^b\) | aOR (95% CI) | | |
| \(< 50\%) | 1 (reference) | | |
| \(50\% +\) | 1.27 (1.12, 1.48) | \(0.001\) |

Other population-level factors considered but did not show any associations with the STIs were average distance to a hospital: median: 6.5 km, IQR 4.1–13.6); average distance to a clinic: median: 1.5 km, IQR 8.8, 19.0); % population without pipe/running water median: 20.8%, IQR 12%, 40%); % percent households (HH) with no electricity: median: 3%, IQR 2%, 4%; %HH traditional informal living: median: 32.8%, IQR 23%, 54.7%

\(^a\)Male/female condoms, IUD
further evidence for a linear increasing trend between the populations with no education and STI diagnosis. Based on the significant cut point estimation, women who resided in areas where more than 50% of the study population had no education were significantly correlated with high STI prevalence (OR 1.27, 95% CI 1.12, 1.48, \( p = .001 \)) (Table 2f).

### Individual and Community-Level Correlates of STI Positivity

The associations between the individual and community-level factors and STI positivity are presented in Table 3. In the multivariable generalized additive model, younger age (<25 years, 25–29 years vs. 30+ years), being single and/or not cohabiting (vs. married/cohabitating), two or more sexual partners in past 3 months (vs. < two sexual partners), 0/1 child, 2 children (vs. three or more children), and using injectables as a contraceptive (vs. no contraceptives) were all significantly associated with increased prevalence of STI diagnosis after adjusting for potential clustering effects due to the geographical sites (<.05). Among the community-level factors, due to the high levels of collinearity/multicollinearity, several variables were omitted from the model. The final model identified six individual-level (age, partnership factors, parity, and contraceptive use) and two community-level factors (% population with no education and “% female-headed households (HH)”\(^{ab}\)) as the independent correlates of STI diagnosis.

### Discussion

Our study expands and brings greater insights to the previous research based on both individual and community-level risk factors for STI prevalence among a large cohort of women who enrolled in various HIV prevention trials. Besides well-established socio-demographic factors and sexual behaviors, we also reported significant correlations between STI diagnosis and characteristics of the communities. Our study also demonstrated that to have a significant effect on STI prevention, a range of individual- and community-level factors needed to be targeted.

Results from the current study highlighted the strong association between STI diagnosis and younger age (<25 years). Consistent with this individual-level finding, at population level, women living in the areas with greater proportions of young females were also identified as being at higher risk of STI diagnosis. Our findings are concurrent with previous studies (Ramjee et al., 2012; Santelli et al., 2013). In South Africa, young women continue to be the most vulnerable group for HIV and other sexually transmitted infections (Wand & Ramjee, 2015). Despite intensive prevention and treatment efforts these rates have steadily increased over the years (Frolich, Abdool Karim, Mashego, Sturm, & Abdool Karim, 2007; Hoque, 2011; Wilkinson, Connolly, Harrison, Lurie, & Abdool, 1998). There are several factors that may increase a young women’s risk of HIV and other STIs. Besides biological factors, high-risk sexual behaviors such as early age at sexual debut, inconsistent condom use, multiple/concurrent sex partners, age–disparate relationships and transactional sex have been reported to be more prevalent among young women (Dellar, Dlamini, & Karim, 2015; Street, Reddy, & Ramjee, 2016).

Our study also identified significant associations between STI diagnosis and various relationship factors including higher number of sex partners, being single/not cohabiting, and partner having another partner. Consistent with these results, at the community level, women living in areas with

### Table 3 Correlates of STI (chlamydia, gonorrhea or syphilis) diagnosis: individual- and community-level factors

| Individual level                      | %  | Adjusted odds ratio\(^{ab}\) | \( p \) value |
|---------------------------------------|----|-----------------------------|--------------|
| Age                                   |    |                             |              |
| <25 years                             | 45 | 4.30 (3.20, 5.77)           | <.001        |
| 25–29 years                           | 21 | 2.72 (2.02, 3.66)           | <.001        |
| 30+ years                             | 34 | 1                           |              |
| Partner has another sex partner       |    |                             |              |
| No                                    | 25 | 1                           |              |
| Yes                                   | 18 | 1.36 (1.11, 1.67)           | .004         |
| Don’t know                            | 57 | 1.91 (1.49, 2.44)           | <.001        |
| Number of sex partners                |    |                             |              |
| <2                                     | 87 | 1                           |              |
| 2+                                     | 13 | 1.46 (1.18, 1.80)           | <.001        |
| Parity                                |    |                             |              |
| 0/1 Child                             | 45 | 2.04 (1.53, 2.72)           | <.001        |
| 2 Children                            | 23 | 1.85 (1.37, 2.48)           | <.001        |
| 3+ children                           | 22 | 1                           |              |
| Contraceptive use                     |    |                             |              |
| No contraceptive                      | 14 | 1                           |              |
| Injectable                             | 51 | 1.53 (1.13, 2.06)           | <.001        |
| Oral contraceptive                   | 10 | 1.02 (0.69, 1.47)           | .982         |
| Others                                | 25 | 1.07 (0.75, 1.54)           | .694         |
| Community level                       |    |                             |              |
| % No education                        |    |                             |              |
| <50%                                  | 43 | 1                           |              |
| 50% +                                 | 57 | 1.26 (1.08, 1.45)           | .002         |
| % Female-headed HH                    |    |                             |              |
| <40%                                  | 26 | 1                           |              |
| 40% +                                 | 74 | 1.17 (1.02, 1.38)           | .025         |

\(^a\)Results were further adjusted for individual trials; \(^b\)intra-class coefficient (ICC) = 19.56%
a higher proportion of single/not cohabiting adults (men and women) were also more likely to be at risk of STI diagnosis. Partnership factors, particularly lack of formal marriage, have been consistently associated with HIV and STIs (Ramjee, Moonsamy, Abbai, & Wand, 2016). Given the high prevalence of STIs in the region, sexual encounters of uninfected and infected individuals will potentially increase the number of new infections. Therefore, such individuals may have higher exposure to sexually transmitted infections due to this sexual mixing. These results are also consistent with the published literature (Street et al., 2016). Partnership/relationship factors and sexual mixing have been frequently linked to HIV and other STIs (Jewkes, Flood, & Lang, 2014; Mlisana et al., 2012). Null/prim parity also correlated with increased prevalence of STI diagnosis. This is primarily due to younger age among those who had fewer children which is also consistent with our earlier findings regarding the strong association between younger age and STI prevalence. In age specific descriptive analysis, vast majority of the younger women (< 25 years) had null/prim parity compared to the older women (84% vs. 51% and 16% for 25–29, 30 + years, respectively) (data not shown).

Consistent with these findings, women living in areas with high rates of HIV infections (> 6 per 100 PY) were also identified as at increased risk of STIs. This finding is not surprising since strong associations between HIV and STIs have been widely reported previously (Mlisana et al., 2012; Naidoo, Wand, Abbai, & Ramjee, 2014; Wand & Ramjee, 2015). Several bacterial and viral STIs have been shown to increase women’s susceptibility to HIV by increasing the quantity of HIV cells due to disruption of the genital epithelium (Shattock & Moore, 2003). In addition to the biological relationships between these pathogens, they also share a simple high-risk sexual behavior: condomless sex.

In our study, 66% of the women reported that they had used condom in last sexual act. However, STI positivity and incidence of pregnancy rates were quite high (20% and 11 per 100 person-year, respectively) (data not shown). Since HIV and STI acquisition occur with unprotected sexual encounter, these findings were interpreted as strong evidence for much lower levels of condom use than it was reported. Despite intensive condom counseling, many women in southern Africa are unable to negotiate safe sex (Pool et al., 2010). Additional challenges for women may include cultural norms and values could influence the ability to achieve correct and consistent use of condoms. In addition to this, collecting sensitive information such as sexual behaviors have been proven to be challenging in the trials conducted in South Africa due to intentional or unintentional misreporting (Ramjee et al., 2012). For example, self-reported condom use was primarily collected by trained interviewers or audio computer assisted technology, which might be subject to misreporting/over-reporting in order to avoid additional counseling (Pool et al., 2010; Wand & Ramjee, 2011).

However, more objective measures such as using hormonal injectables (> 50%) as a contraceptive method were significantly associated with increased prevalence of STI diagnosis. Although these effective long-term methods play a crucial role in preventing unintended pregnancies, they do not protect against HIV and other STIs (Heffron et al., 2012; Wand & Ramjee, 2012).

Socioeconomic factors, including education and employment opportunities, have been shown to have a profound impact on the spread of HIV and STIs (Hunter, 2007). Besides high HIV and STI rates, KwaZulu-Natal has also some of the highest unemployment rates in the country (Census, 2011). The linkage between HIV/STIs and low socioeconomic conditions, particularly high unemployment and lack of education rates, has been well established (Lurie et al., 2008). Our study also reported concurrent associations between the high prevalence of STIs and lack of schooling at individual and community level. Employment status of the women was not collected in all trials therefore; we were not able to present the results at individual level. However, at the community level, women living in areas with higher unemployment rates were at increased risk of STIs. In an additional analysis, we also identified these areas as being the areas with high proportion of lack of education (Spearman correlation coefficient = 54%, p < .001) (data not shown). Studies also reported associations between poverty and female-headed households (Posel & Rogan, 2009). According to the South African Department of Health, Medical Research Council, almost 50% of all households are headed by women (Census, 2011). This is primarily due to male labor migration, lack of formal marriage and HIV/AIDS related deaths (Leah, 2016; Posel, 2001). Therefore, female-headed households may be more vulnerable to income poverty than male-headed households (Posel & Rogan, 2009). Concurrent with the previous research, our results also showed that women who resided in communities with a greater proportion of female-headed households were also significantly associated with increased prevalence of STIs. Taken together, these data indicate that women who lived in areas with high proportion of female-headed households may more likely to expose to transactional sex for survival which may compromise women’s sexual negotiating power for condoms.

Limitations

Our study had limitations; therefore, the results should be interpreted carefully. First, our primary endpoint was diagnosis of STIs at baseline; therefore, we analyzed the data as cross sectional. Therefore, this study has the same limitations of these types of study designs. We considered a wide range of measurements if they were collected consistently across the trials. Although this may potentially limit our ability to identify the impact of unmeasured factors on STI diagnosis, because of the nature of the research conducted in these.
trials, factors included in this study have been well established and consistently reported to have significant impacts on increasing prevalence for STIs (Dellar et al., 2015; Jewkes et al., 2014; Milicana et al., 2012; Naidoo et al., 2014; Street et al., 2016; Wand & Ramjee, 2015). Due to the variations in questionnaires across the trials, we were restricted in categorization of some of the variables (e.g., level of education). We have also considered wide range of community-level factors using the 2011 South African Census Statistics. We particularly focused on the characteristics because of their potential impact on the epidemic directly or indirectly. We hypothesized that the community-level factors considered in this study may be influential enough to shape individual-level characteristics which were known to have significant impact with STI diagnosis in the region. Particularly, those related to the partnership factors which have been consistently shown to be associated with STI diagnosis (Zuma et al., 2005). For example, “% population female aged 18–25 years” and “% adult population unmarried/not cohabiting” were included in our analysis because, they have the potential to be surrogate for partnership and sexual networking. We hypothesized that women who lived in the areas with high proportion of young females and single adults were more likely to have higher number of sexual partners which has been linked to increase risk of STIs; therefore, they potentially create a sexual network. Given the high prevalence of STI in the region, their risk of STIs will increase (Jewkes et al., 2014; Milicana et al., 2012); while “% female-headed households,” “% population unemployed,” and “% population with no schooling” may reflect poor socioeconomic conditions in the region. However, we still cannot rule out the potential impact of unmeasured characteristics such as poverty, social norms, cultural beliefs as well as traditional sexual practices. No data were available for the male partners. Despite these limitations, results from our study demonstrated crucial individual- and community-level information based on a large cohort of women. We did not have data concerning “sexual mixing”; therefore, we were not able to examine its impact on STI diagnosis directly. However, our study presented significant associations between STI diagnosis and partnership factors including women’s marital status and number of sexual partners. Based on the published research (modeling and empirical), we hypothesized that these factors (collectively) may create a sexual mixing due to high partner change rate (Johnson, Dorrington, Bradshaw, Pillay-Van Wyk, & Rehle, 2009; Morris, Kurth, Hamilton, Moody, & Wakefield, 2009). Therefore, they may likely to play significant role in shaping other (measured/unmeasured) sexual behaviors that increase women’s risk of STIs (Ghani, Swinton, & Garnett, 1997). These results were also empirically confirmed in our study too. For example, higher number of sexual partners (2 + past three months) were more common among women who were not married (15% vs. 5%, p < .001) (data not shown); in addition, the vast majority of the unmarried women reported that their partner had another partner (or not sure whether they had or not) (88% vs. 37%, p < .001). These results may have broader implications due to the additional challenges for women may include poverty and unemployment. We have reasons to believe that these women might have been involved in unprotected sexual contacts which can be characterized as transactional and/or inter-generational sex which would increase women’s vulnerability for STIs (Dunkle, 2004; Hunter, 2002; Luke, 2003). Although we were not able to collect sexual behavior data from male partners of the women, our findings collectively suggested that they were likely to be high-risk men with multiple/concurrent partners.

Results from our study also suggest that the GAMMs may be a useful tool for informing these efforts by identifying cut points on community-level measures that distinguish ranges on the scale that are most strongly associated with STI positivity. As the poor socioeconomic characteristics continue to have significant impact on increasing levels of HIV and STI diagnosis, determining meaningful benchmarks remains research priority. This is particularly crucial to evaluate effectiveness of interventions to prevent new infections at community level. Existent research provides little guidance to policy makers where these thresholds should be set. Our findings may potentially have important policy implications both in terms of identifying individual and community-level vulnerabilities as well as setting meaningful benchmarks using statistically sound techniques. Consequently, the GAMMs could be used to identify statistically meaningful, evidence-based thresholds for the community-level characteristics which may guide to policymakers how to weigh all the empirical thresholds and their goals for reducing STIs to create cost-effective policies.

**Conclusion**

This is one of the first studies to present results by combining various individual-level data sources and linking them to spatially aggregated community-level data using geo-referencing techniques. Results from our study showed that besides established individual-level factors including socio-demographic measurements and high-risk sexual behaviors, prevalence of STIs may also be influenced by the level of exposure to the geographical-level characteristics. Intuitively, higher-level exposures to certain characteristics may have a profound influence on the dynamic of the STI diagnosis. Future prevention strategies based on identifying individual- and community-level drivers of sexually transmitted infections are likely to constitute the most effective means of targeting communities as well as individuals at highest risk of sexually transmitted infections. This strategy has the potential to play a significant role in the epidemic’s trajectory.
References

Abdool Karim, S. S., Richardson, B. A., Ramjee, G., Hoffman, I. F., Chirench, Z. M., Taha, T., et al. (2011). Safety and effectiveness of buffer gel and 1% PRO2000 gel for the prevention of HIV infection in women. *AIDS*, 25(7), 957–966.

Census. (2011). *Statistical release. Statistics South Africa*. 2012 (p. 14) [http://www.statssa.gov.za]. Retrieved November 24, 2017.

Chaudhuri, P., & Marron, J. S. (1999). SiZer for exploration of structures in curves. *Journal of American Statistical Association*, 94, 807–823.

Dellar, R., Dlamini, S., & Karim, Q. (2015). Adolescent girls and young women: Key populations for HIV epidemic control. *Journal of the International AIDS Society*, 18, 64–70.

Dunkle, K. (2004). Transactional sex among women in Soweto, South Africa: Prevalence, risk factors and association with HIV infection. *Social Science and Medicine*, 59(8), 1581–1592.

Frollich, J. A., Abdool Karim, Q., Mashego, M. M., Sturm, A. W., & Abdool Karim, S. S. (2007). Opportunities for treating sexually transmitted infections and reducing HIV risk in rural South Africa. *Journal of Advanced Nursing*, 60(4), 377–383. [https://doi.org/10.1111/j.1365-2648.2007.04045.x].

Ghani, A. C., Swinton, J., & Garnett, G. P. (1997). The role of sexual partnership networks in the epidemiology of gonorrhea. *Sexually Transmitted Diseases*, 24, 45–56.

Health NDo. (2013). *The 2013 national antenatal sentinel HIV and syphilis prevalence survey, South Africa*. Retrieved November 25, 2017.

Heffron, R., Donnell, D., Rees, H., Celum, C., Mugo, N., Were, E., et al. (2012). Use of hormonal contraceptives and risk of HIV-1 transmission: A prospective cohort study. *The Lancet Infectious Disease*, 12, 19–26. [https://doi.org/10.1016/S1473-3099(11)70247-X].

Hogue, M. E. (2011). Reported risky sexual practices amongst female undergraduate students in KwaZulu-Natal, South Africa. *African Journal of Primary Health Care & Family Medicine*, 3(1), Art. #281. [http://dx.doi.org/10.4102/phcfm.v3i1.281].

Hunter, M. (2002). The materiality of sex: Thinking beyond prostitution. *African Studies*, 61(1), 99–120.

Hunter, M. (2007). The changing political economy of sex in South Africa: The significance of unemployment and inequalities to the scale of the AIDS pandemic. *Social Science & Medicine*, 64, 689–700.

Jewkes, R., Flood, M., & Lang, J. (2014). From work with men and boys to changes of social norms and reduction of inequities in gender relations: A conceptual shift in prevention of violence against women and girls. *The Lancet*, 385, 1580–1589.

Johnson, L. F., Dorrington, R. E., Bradshaw, D., Pillay-Van Wyk, V., & Rehle, T. M. (2009). Sexual behaviour patterns in South Africa and their association with the spread of HIV: Insights from a mathematical model. *Demographic Research*, 21, 289–340.

Laga, M., Manoka, A., Kivuna, M., Malele, B., Tuliza, M., & Nzila, N. (1993). Non-ulcerative sexually transmitted diseases as risk factors for HIV-1 transmission in women: Results from a cohort study. *AIDS*, 7, 95–102.

Leah, G. (2016). The mercurial piece of the puzzle: Understanding stigma and HIV/AIDS in South Africa. *SAHARA-J: Journal of Social Aspects of HIV/AIDS*, 7, 95–102. [https://doi.org/10.7199/SAHARA-J.2015.0130644].

Leon, N., Naidoo, P., Mathews, C., Lewin, S., & Lombard, C. (2010). The impact of provider-initiated (opt-out) HIV testing and counseling of patients with sexually transmitted infection in Cape Town, South Africa: A controlled trial. *Implementation Science*, 5, 8. [https://doi.org/10.1186/1748-9098-5-8].

Luke, N. (2003). Age and economic asymmetries in the sexual relationships of adolescent girls in sub-Saharan Africa. *Studies in Family Planning*, 34(2), 67–86.

Lurie, M., Pronyk, P., de Moor, E., Heyer, A., de Bruyn, G., & Struthers, H. (2008). Sexual behavior and reproductive health among HIV-infected patients in urban and rural South Africa. *Journal of Acquired Immune Deficiency Syndromes*, 47(4), 484–493.

Marrazzo, J. M., Ramjee, G., Richardson, B. A., Gomez, K., Mgodi, N., & Nair, G. (2015). Tenofovir-based preexposure prophylaxis for HIV infection among African women. *New England Journal of Medicine*, 372, 509–518. [https://doi.org/10.1056/NEJMa1402269].

Martin, H., Richardson, B., Nyange, P., Lavreys, L., Hillier, S., Chohan, B., et al. (1999). Vaginal lactobacilli, microbial flora, and risk of human immunodeficiency virus type 1 and sexually transmitted disease acquisition. *Journal of Infectious Diseases*, 180, 1863–1868.

McClelland, R., Richardson, B., Hassan, W., Chohan, V., Lavreys, L., Mandalaya, K., et al. (2008). Improvement of vaginal health for Kenyan women at risk for acquisition of human immunodeficiency virus type 1: Results of a randomized trial. *Journal of Infectious Diseases*, 197, 1361–1368.

McClelland, R., Sangare, L., Hassan, W., Lavreys, L., Mandalaya, K., Kiarie, J., et al. (2007). Infection with *Trichomonas vaginalis* increases the risk of HIV-1 acquisition. *Journal of Infectious Diseases*, 195, 698–702.

McCormack, S., Ramjee, G., Kamali, A., Rees, H., Crook, A. M., Gafos, M., et al. (2010). PRO2000 vaginal gel for prevention of HIV-1 infection (Microbicides Development Programme 301): A phase 3, randomised, double-blind, parallel-group trial. *Lancet*, 376, 1329–1337. [https://doi.org/10.1016/S0140-6736(10)60339-0].

Misana, K., Naicker, N., Werner, L., Roberts, L., van Loggerenberg, F., Baxter, C., et al. (2012). Symptomatic vaginal discharge is a poor predictor of sexually transmitted infections and genital tract inflammation in high-risk women in South Africa. *Journal of Infectious Diseases*, 206, 6–14. [https://doi.org/10.1093/infdis/jit298].

Morris, M., Kuth, A. E., Hamilton, D. T., Moody, J., & Wakefield, S. (2009). Concurrent partner-ship and HIV prevalence disparities by race: Linking science and public health practice. *American Journal of Public Health*, 99, 1023–1031.

Naidoo, S., Wand, H., Abbia, N. S., & Ramjee, G. (2014). High prevalence and incidence of sexually transmitted infections among women living in KwaZulu-Natal, South Africa. *AIDS Research and Therapy*, 11, 31. [https://doi.org/10.1186/1742-6405-11-31].

Padian, N. S., van der Straten, A., Ramjee, G., Chipato, T., de Bruyn, G., Blanchard, K., et al. (2007). Diaphragm and lubricant gel for prevention of HIV in southern African women: A randomised controlled trial. *Lancet*, 370, 251–261. [https://doi.org/10.1016/S0140-6736(07)60950-7].

Pool, R., Montgomery, C. M., Morar, N. S., Mweebma, O., Szali, A., & Gafos, M. (2010). A mixed methods and triangulation model for increasing the accuracy of adherence and sexual behaviour data: The Microbicides Development Programme. *PLoS ONE*, 5(7), e11600.

Posel, D. R. (2001). Intra-family transfers and income-pooling a study of remittances in KwaZulu-Natal. *South African Journal of Economics*, 69, 3. [https://doi.org/10.1111/j.1312-8692.2001.tb00023.x].

Posel, D., & Rogan, M. (2009). Women, income and poverty: Gendered access to resources in post-apartheid South Africa. *Agenda*, 81, 25–34.

Ramjee, G., Moonsamy, S., Abbia, N. S., & Wand, H. (2016). Individual and population level impact of key HIV risk factors on HIV incidence rates in Durban, South Africa. *PLoS ONE*, 11(4), e0153969.

Ramjee, H., & Wand, H. (2014). Geographical clustering of high risk sexual behaviors in “hot spots” for HIV and sexually transmitted infections in KwaZulu-Natal, South Africa. *AIDS and Behavior*, 18, 317–322. [https://doi.org/10.1007/s10461-013-0578-x].

Ramjee, G., Wand, H., Whitaker, C., McCormack, S., Padian, N., Kelly, C., et al. (2012). HIV incidence among non-pregnant women living
in selected rural, semi-rural and urban areas in Kwazulu-Natal, South Africa. *AIDS and Behavior, 16*, 2062–2071.

Santelli, J. S., Edelstein, Z. R., Mathur, S., Wei, Y., Zhang, W., Orr, M. G., et al. (2013). Behavioral, biological, and demographic risk and protective factors for new HIV infections among youth, Rakai, Uganda. *Journal of Acquired Immune Deficiency Syndromes, 63*, 393–400. https://doi.org/10.1097/qai.0b013e3182926795.

Shattock, R. J., & Moore, J. P. (2003). Inhibiting sexual transmission of HIV-1 infection. *Nature Reviews Microbiology, 1*(1), 25–34.

Skoler-Karpoff, S., Ramjee, G., Ahmed, K., Altini, L., Plagianos, M., Friedland, B., et al. (2008). Efficacy of Carraguard for prevention of HIV infection in women in South Africa: A randomised, double-blind, placebo-controlled trial. *Lancet, 372*, 1977–1987. https://doi.org/10.1016/S0140-6736(08)61842-5.

Street, R., Reddy, T., & Ramjee, G. (2016). The generational effect on age disparate partnerships and the risk for human immunodeficiency virus and sexually transmitted infections acquisition. *International Journal of STD and AIDS, 27*, 9. https://doi.org/10.1177/095646241559232

Tanser, F., Oliveira, T., Maheu-Giroux, M., & Bärnighausen, T. (2014). Concentrated HIV sub-epidemics in generalized epidemic settings. *Current Opinion in HIV and AIDS, 9*(2), 115–125. https://doi.org/10.1097/COH.0000000000000034.

UNAIDS. (2016) *Global AIDS update*. Geneva, Switzerland: Joint United Nations Program on HIV/AIDS.

Wand, H., & Ramjee, G. (2011). Combined impact of sexual risk behaviors for HIV seroconversion among women in Durban, South Africa: Implications for prevention policy and planning. *AIDS and Behavior, 15*, 479–486. https://doi.org/10.1007/s10461-010-9845-2.

Wand, H., & Ramjee, R. (2012). The effects of injectable hormonal contraceptives on HIV seroconversion and on sexually transmitted infections. *AIDS, 26*, 375–380. https://doi.org/10.1097/QAD.0b013e32834f990f.

Wand, H., & Ramjee, G. (2015). Biological impact of recurrent sexually transmitted infections on HIV seroconversion among women in South Africa: Results from frailty models. *Journal of the International AIDS Society, 18*, 19866. https://doi.org/10.7448/IAS.18.1.19866.

Wand, H., Reddy, T., Naidoo, S., Moonsamy, S., Siva, S., Morar, N. S., et al. (2018). A simple risk prediction algorithm for HIV transmission: Results from HIV prevention trials in KwaZulu Natal, South Africa (2002–2012). *AIDS and Behavior, 22*, 325–336. https://doi.org/10.1007/s10461-017-1785-7.

Wilkinson, D., Connolly, A. M., Harrison, A., Lurie, M., & Abdool Karim, S. S. (1998). Sexually transmitted disease syndromes in rural South Africa: Results from health facility surveillance. *Sexually Transmitted Diseases, 25*(1), 20–23. https://doi.org/10.1097/00007435-199801000-00005.

Wood, S. N. (2004). Stable and efficient multiple smoothing parameter estimation for generalized additive models. *Journal of the American Statistical Association, 99*, 673–686.

Wood, S. N. (2006). Low rank scale invariant tensor product smooths for generalized additive mixed models. *Biometrics, 62*(4), 1025–1036.

Zuma, K., Lurie, M., Williams, B., Mkaya-Mwamburi, D., Garnett, G., & Sturm, A. (2005). Risk factors of sexually transmitted infections among migrant and non-migrant sexual partnerships from rural South Africa. *Epidemiology and Infection, 133*(3), 421–428.