Geospatial Health 6(2), 2012, pp. 221-231

Divergent spatial patterns in the prevalence of the human immunodeficiency virus (HIV) and syphilis in South African pregnant women

Samuel O.M. Manda[1,2], Carl J. Lombard[1], Thabang Mosala[3]

[1]Biostatistics Unit, South African Medical Research Council, South Africa; [2]School of Mathematics, Statistics and Computer Science, University of Kwazulu-Natal South Africa; [3]Epidemiology and Surveillance, National Department of Health, South Africa

Abstract. An analysis of the ecological association between the human immunodeficiency virus (HIV) and syphilis was undertaken using joint mapping modelling based on data from South African national HIV and syphilis sentinel surveillance surveys carried out between 2007 and 2009. The syphilis prevalence, taken as proxy for sexual behaviour and increased HIV transmission, was first used with health district-level deprivation and population density as a covariate in a HIV prevalence spatial regression model and, secondly, together with HIV as a bivariate outcome. HIV was more highly prevalent in deprived and populated urban areas than elsewhere, while syphilis had a high prevalence in less deprived and less populated rural areas. Spatially, the HIV prevalence was lowest in the southwestern and highest in the northeastern parts of the country. This was in discordance to the syphilis prevalence, which revealed negative correlations with the HIV prevalence. Considerable variations across the districts remained after adjusting for the contextual covariate factors. Divergent spatial patterns between HIV and syphilis were identified, regarding both observed and unobserved covariate effects. The differing disease-specific spatial prevalence patterns may point to inconsistent successes in interventions between the two diseases. Overall, the results emphasize the need to develop and test plausible aetiological hypotheses relating to ecological correlations and causes of the disease-specific interjectory between the districts.

Keywords: antenatal HIV and syphilis prevalence, spatial epidemiology, South Africa.

Introduction

With an estimated 22.4 million people living with the human immunodeficiency virus (HIV), the countries in the Sub-Saharan African (SSA) region are more heavily affected by the epidemic compared to anywhere else (UNAIDS, 2010). This means that two in every three people affected globally with HIV live in the SSA region. There were around 1.4 million deaths related to the acquired immune deficiency syndrome (AIDS) and 1.9 million new HIV infections in the region in 2010 (UNAIDS, 2010). In the same year, HIV prevalence among the adult population in the region was estimated to be 5.2%, whilst the global adult prevalence was estimated at 0.8%. However, the HIV prevalence in the region varies considerably between sub-regions. With HIV prevalence lower that 2%, western and central parts have relatively lower prevalence rates compared to the southern parts, with HIV prevalence between 15% and 30% (UNAIDS, 2010). Globally, countries in East Asia have a very low HIV prevalence, with under 0.1% of the population infected (UNAIDS, 2010).

The use of national HIV prevalence estimates for comparative purposes between countries might be misleading. Often substantial differences in the prevalence often exist between geographical areas in a country (Sandøy et al., 2006). South Africa, with an estimated adult HIV prevalence of around 17.9% and 5.6 million people living with HIV, has one of the largest HIV epidemics in the world (UNAIDS, 2010). However, these figures mask significant provincial variations in the HIV epidemic, with the adult (15-49 years) prevalence ranging from 5.3% and 9.0% in Western and Northern Cape provinces to 23.1% and 25.8% in Mpumalanga and Kwazulu-Natal provinces (Shisana et al., 2009). The national antenatal HIV prevalence is at 30.2%, but this specific HIV prevalence also shows considerable inter-province variations; ranging from a low of 18.4% in Northern Cape to a high of 39.5% in Kwazulu-Natal (Department of Health, 2011). The district-level antenatal HIV preva-

Corresponding author:
Samuel O.M. Manda
Biostatistics Unit
South African Medical Research Council
Private Bag X385
Pretoria 0001, South Africa
Tel: +27 12 338 8518; Fax +27 12 339 8582
Email: samuel.manda@mrc.ac.za

lence shows even higher order heterogeneity, with the prevalence ranging from 8.5% to 42.3% (Department of Health, 2010). These differences may be explained by demographic, socioeconomic, sexual behavioural and cultural factors at both individual and geographical levels (Cohen, 1998; Johnson et al., 2005; Pettifor et al., 2005; Sandøy et al., 2006; Kleinschmidt et al., 2007; Yahya-Malima et al., 2008; Shisana et al., 2009; Department of Health, 2011).

Previous studies in South Africa and elsewhere have only reported spatial heterogeneity in HIV prevalence using observed prevalence at sampling sites or districts or provinces (Sandøy et al., 2006; Montana et al., 2007; Shisana et al., 2009; Department of Health, 2011). Some have used HIV seroprevalence data collected in national household surveys to develop a predictive HIV prevalence model at the sub-provincial level in an effort to produce smooth maps of HIV prevalence (Montana et al., 2007). However, the observed and predicted prevalence rates and the resulting maps can be unstable and unreliable due to sparse data from small areas or from areas with very low prevalence rates. Thus, it becomes necessary to use smoothing techniques that borrow data from neighbouring areas, as areas that are closer in proximity may share similar factors affecting their responses. Spatial smoothing techniques also reduce the effect of administrative boundaries, which are unlikely to be related to disease incidence. In this way, any artefactual variation exhibited in the data by methods of data aggregation can be ameliorated (Kleinschmidt et al., 2007).

The HIV and syphilis prevalence data among pregnant women in South Africa constitute a good example of how to apply novel multivariate spatial models for multiple disease outcomes. Our approach is similar to that of Kleinschmidt et al. (2007) where the spatial HIV prevalence in South Africa was analysed using a geo-statistical model (Diggle et al., 1998). In here we use relative positions of the areas to describe neighbourhood set of areas in describing spatial dependence in HIV prevalence (Besag et al., 1992; Browne et al., 2001). Furthermore, we extend the methodology by using multivariate spatial models by modelling HIV and syphilis prevalence rates as a bivariate outcome; the syphilis prevalence is taken as a proxy for risky sexual behaviour, which is not directly measured in the surveys. We are not aware of any previous work based on this approach to estimate the geographical distribution of HIV prevalence across South Africa or in any other country in SSA. In particular, a Bayesian bivariate binomial spatial model was used to model jointly HIV and syphilis prevalence data across health districts in South Africa.

A primary motivation for this study was to investigate ecological associations between HIV and syphilis in South Africa using joint spatial modelling for multiple disease outcomes. We account for district level social and material deprivation and population density. We estimate and map the resulting smoothed HIV and syphilis prevalence rates before and after adjusting for these two spatially relevant covariates. The modelling and estimation procedures are carried out within a full Bayesian hierarchical framework using a bivariate binomial, model.

Materials and methods

South African national antenatal-based sentinel surveillance

South Africa conducts high-quality HIV and syphilis sentinel surveillance surveys among antenatal clinic attendees, one of the key priorities in public health (Shisana et al., 2009; Department of Health, 2011). The surveys currently cover a total of 1,457 sites, targeting 36,000 pregnant women aged 15-49 years annually in all of the 52 health districts of the country. The sampled women have their blood samples tested for HIV and active syphilis with the rapid plasma regain (RPR) test (Department of Health, 2011). The demographic details are anonymously linked to the test results. We use data collected over a 3-year period, 2007, 2008 and 2009, in which 33,684, 33,927 and 33,861 pregnant women, respectively, were sampled. By using the more recent rounds of surveillance data, the current situation would be better assessed.

The analyses were done at the health district level, the basic unit, through which the delivery of primary health care is undertaken in South Africa. For each district, the following two contextual variables were extracted from the district health barometer database maintained by the health service trust (Health Systems Trust, 2009):

(i) the deprivation index, derived as composite set of demographic and socioeconomic variables that are indicative of material and social deprivation (the variables that were used to construct the index are listed in Appendix A, where also the method used to derive the index is described); and

(ii) the population density, defined as number of people per square kilometre, used as a proxy for social network and interactions since there is evidence that more densely populated areas have higher HIV and STI prevalence rates, possibly due to increased economic activity (Meidany et al., 2000).

Each of these contextual covariates was partitioned into quintiles, with the lowest, e.g. the least deprived, taken as the reference group. This categorisation enables effects to be detected at the extremes of the range.

A joint spatial model for HIV and syphilis prevalence data

In spatial modelling of observed responses across geographical areas, both the effects of the area under investigation and the surrounding areas are modelled, where the relative positions of the areas are used to define neighbourhood sets (Besag et al., 1991) . We use the recently developed multivariate disease models for multiple outcomes to jointly investigate geographic comorbidity between HIV and syphilis among antenatal attendees in South Africa (Leyland et al., 2000; Feltbower et al., 2005; Manda et al., 2009).

For our application, there are $I=52$ contiguous health districts and the data were collected at three successive years. *(T=1, 2007; T=2, 2008; T=3, 2009).* Now let Y_{it1} and Y_{it2} represent the number of subjects who tested positively for HIV and syphilis, respectively, in district i for year t out of the corresponding N_{it} antenatal clinic attendees sampled. The observed positive indications Y_{itj} aare conditionally independent binomial variables with spatial prevalence probability π_{ijt}; *(j=1,2)*. The spatial prevalence probability depends on the district observed covariates and unobserved spatial process by means of the logit link function $logit\ \pi_{ijt} = \alpha_{0jt} + \beta_j^T X_i + \theta_{ijt}$, where α_{0jt} is the disease-specific overall mean prevalence on the log-odds scale at time-point t and β_j represent the disease-specific regression coefficients for a set of known and measured covariates X_i for district i. The terms θ_{i1t} and θ_{i2t} represent the spatially varying log-odds of HIV and syphilis prevalence, respectively, at time-point t. Thus, this model describes the spatial variation in the overall mean prevalence attributable to both the effects of the covariates observed as well as the effects of unobserved and unmeasured covariates, which may induce dependences in the observed prevalence rates.

The random spatial effect θ_{ijt} is split into two components, u_{ijt} and v_{ijt}, which represent the unstructured and the spatially structured random effects, respectively, in area (Besag et al., 1991). We adopt a multiple memberships multiple classification (MMMC) model to characterise the spatially correlated random effect v_{ijt} (Browne et al., 2001) and used two classifications for the two variations:district classification to capture the unstructured variation (classification level 2), and neighbour classification to capture the structured variation (classification level 3), i.e. the effects due to neighbouring areas. A multivariate version of the MMMC spatial model, proposed by Leyland et al. (2000) was adopted within a Bayesian framework as done by Feltbower et al. (2005) and by Manda and Leyland (2007). The details for formulating the spatial random effects are explained in the Appendix B, which also describes the characterisation of the correlation structure between the various random effects and prior specification of both regression and variance parameters.

We also empirically computed relative contributions of unstructured and structured variation to the total variation of random effects. If the spatial fraction is close to 1 then the spatial heterogeneity dominates, otherwise the unstructured heterogeneity dominates. Overall, we were interested in comparing variations attributable to spatial and unstructured heterogeneity random effects before and after adjusting for the relevant contextual covariates.

Modelling and mapping strategies

We initially fitted a univariate spatial binomial model for the HIV prevalence where syphilis was used as a covariate in addition to the two other contextual factors, and we also examined the distribution of spatially smoothed estimates of the log-odds of HIV prevalence. We then modeled the HIV and syphilis prevalence rates jointly using a bivariate binomial model. We calculated unadjusted and covariate-adjusted posterior probabilities that a district log-odds exceeds 0. Finally, the effect on the degree of spatial correlation between the two diseases was examined before and after allowing for the two contextual covariate factors.

The model parameters were estimated using Bayesian computer software (Lunn et al., 2000). Two parallel Gibbs sampler chains were run for 20,000 iterations from independent starting positions. Using German and Rubin (1992) convergence diagnosis tools, satisfactory convergence was already observed after 5,000 iterations in each case. Posterior summaries were based on a combined sample of the remaining 30,000 iterations needed to complete the cycles. We compared performance of various models using deviance information criterion (DIC), which is the sum of model fit and complexity, both of which must be considered to identify a model that supports the data better (Spiegelhalter et al., 2002). The model with a smaller DIC is better supported by the data. An abridged copy of the WinBUGS code used for fitting the bivariate, binomial, spatial model is provided at the end of Appendix B.

Results

Observed data

The mean, median and range of the district prevalence of HIV and syphilis by year are shown in Table 1. Across the 3-year period, the minimum number of antenatal clinic attendees sampled per district was 51 and the maximum was 2,627. The prevalence of HIV averaged 27.8%, 26.6% and 27.4% per district over the three years studied, with ranges of 7.3% to 41.6%, 2.3% to 45.7% and 0% to 46.4%, respectively, whereas the corresponding averages for syphilis prevalence were 3.2% (0.2% to 11.2%), 2.3% (0% to 12.6%) and 2.2% (0.1% to 9.7%) per district. This shows that there was a great deal of variation in both HIV and syphilis prevalence between the districts. However, some of the estimates, especially for syphilis, were based on comparatively small case and sample sizes, so it is difficult to clearly see the geographical patterns of risk.

Table 1. Health district level summaries by year of the prevalence survey, 2007 to 2009.

		2007	2008	2009
Total sample (N)		33,684	34,927	33,861
Sample size	Mean	647.5	650	631.9
	Median	522	513	492
	Minimum	55	54	51
	Maximum	2627	2536	2489
HIV prevalence	Mean	27.8	26.6	27.4
	Median	29.3	28.2	28.1
	Minimum	7.3	2.3	0
	Maximum	41.6	45.7	46.4
Syphilis prevalence	Mean	3.2	2.3	2.2
	Median	2.6	1.6	1.7
	Minimum	0.2	0	0.1
	Maximum	11.2	12.6	9.7

The spatial distributions of the observed HIV and syphilis prevalence rates for each year showed similar patterns (results not shown). Thus, we present the average prevalence rates over the 3-year period in Fig. 1, which also shows the distribution of the two contextual factors. High HIV rates were found to be concentrated in the northeast, covering the provinces of Kwazulu-Natal, Free State, Mpumalanga and Limpopo. Low HIV rates were mainly seen in the southwest, covering Northern and Western Cape provinces. The more rural and least populated areas

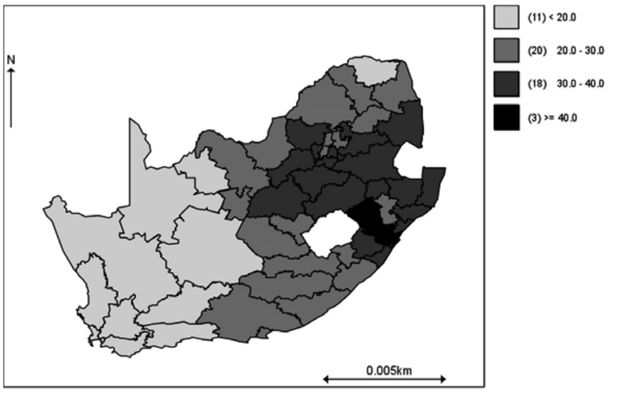

Fig. 1a. Antenatal HIV prevalence by health district in South Africa, 2007-2009.

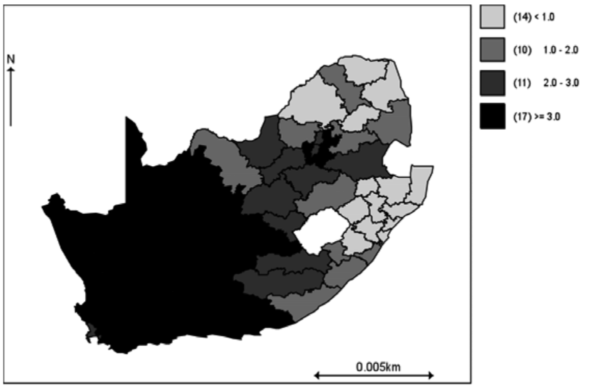

Fig. 1b. Antenatal syphilis prevalence by health district in South Africa, 2007-2009.

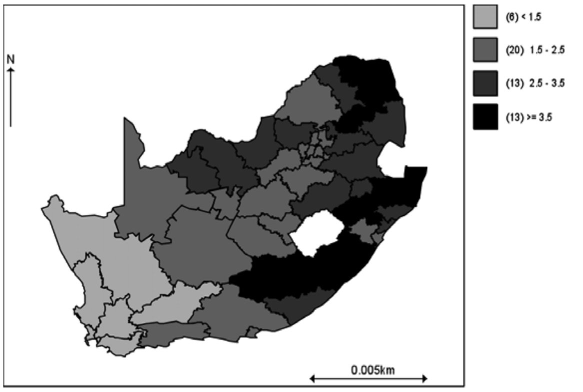

Fig. 1c. Average level of material and social deprivation by health district in South Africa, 2005-2007.

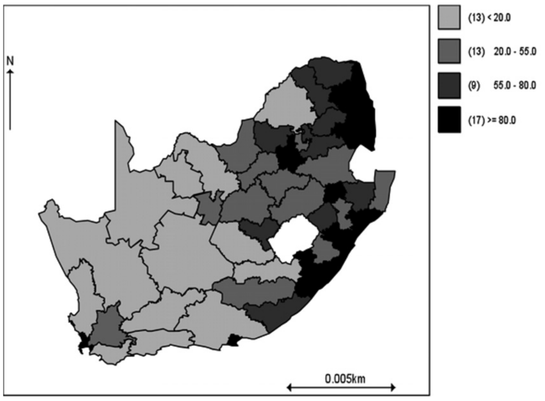

Fig. 1d. Population density by health district in South Africa, 2007.

were found to have lower HIV rates in general. In contrast, the syphilis prevalence was found to be high in the western parts, and low in the east. Most of the deprived districts are in the north-eastern corridor, and the least deprived districts in the south-western parts (Fig. 1c). A further investigation (results not shown), revealed that most rural districts are among the most deprived districts. High population density levels are found in the north-eastern corridor (Fig. 1d).

Results using the fitted models

The estimated covariate effects are presented on the odds ratio scale and are shown in Table 2 for a variety of models. Higher prevalence rates of HIV were found in the most deprived districts and with high population densities. There was evidence of negative associations between syphilis and both factors: population density and deprivation. Using the unadjusted HIV ecological model, the spatial component

Table 2. Posterior median (95% CI) estimates for HIV and syphilis prevalence parameters, South Africa, 2007-2009.

		Unadjusted model	Adjusted model	Unadjusted bivariate model		Adjusted bivariate model	
Covariate effects		HIV	HIV	HIV	Syphilis	HIV	Syphilis
Yearly prevalence (%)	2007	26.6 (22.6, 31.7)	25.3 (18.6, 32.6)	27.0 (21.9, 33.2)	2.4 (1.7, 3.7)	16.0 (10.3, 24.1)	5.6 (3.1, 9.5)
	2008	26.5 (22.6, 31.7)	25.3 (18.5, 32.5)	27.0 (21.9, 33.1)	1.7 (1.1, 2.5)	15.9 (10.3, 24.1)	3.9 (2.2, 6.6)
	2009	26.6 (22.6, 31.7)	25.4 (18.6, 32.6)	27.0 (21.9, 33.2)	1.6 (1.1, 2.5)	16.0 (10.3, 24.2)	3.8 (2.1, 6.5)
Deprivation	I (Lowest)		1.00			1.00	1.00
	II		1.29 (1.05, 1.64)			1.43 (0.93, 2.13)	0.86 (0.52, 1.50)
	III		1.53 (1.21, 2.03)			1.86 (1.26, 2.84)	0.58 (0.33, 1.03)
	IV		1.13 (0.84, 1.44)			1.47 (0.99, 2.26)	0.41 (0.23, 0.78)
	V (Highest)		1.14 (0.83, 1.61)			1.69 (1.02, 2.81)	0.33 (0.18, 0.64)
Population density	I (Lowest)		1.00			1.00	1.00
	II		1.31 (1.05, 1.66)			1.35 (0.89, 2.16)	0.87 (0.55, 1.42)
	III		1.20 (0.93, 1.51)			1.23 (0.76, 2.04)	0.75 (0.45, 1.35)
	IV		1.19 (0.89, 1.53)			1.34 (0.84, 2.17)	0.55 (0.31, 1.03)
	V (Highest)		1.36 (0.87, 2.02)			1.72 (1.07, 2.77)	0.59 (0.34, 1.07)
Syphilis prevalence	I (Lowest)		1.00				
	II		0.84 (0.66, 1.06)				
	III		0.81 (0.61, 1.07)				
	IV		0.71 (0.52, 1.06)				
	V (Highest)		0.34 (0.49, 0.73)				
Random effects							
Unstructured standard deviation		0.16 (0.10, 0.24)	0.17 (0.10, 0.27)	0.33 (0.25, 0.43)	0.42 (0.28, 0.60)	0.31 (0.23, 0.41)	0.39 (0.27, 0.58)
Structured standard deviation		0.78 (0.59, 1.05)	0.50 (0.26, 0.79)	0.69 (0.49, 0.96)	1.11 (0.77, 1.56)	0.54 (0.35, 0.85)	0.79 (0.46, 1.26)
Proportion structured		0.90 (0.78, 0.96)	0.71 (0.04, 0.98)	0.64 (0.45, 0.79)	0.75 (0.49, 0.89)	0.50 (0.22, 0.73)	0.55 (0.19, 0.83)
Spatially unstructured correlation				-0.26 (-0.59, 0.26)		-0.16 (-0.55, 0.31)	
Spatially ctructured correlation				-0.50 (-0.79, -0.02)		-0.31 (-0.77, 0.36)	
Model comparison (DIC)		1204.2	1203.4	2108.0		2109.3	

accounted for 90% of the variation. However, this contribution was reduced to 71% after adjustments for syphilis, deprivation and population density. In the unadjusted bivariate model, 64% and 75% of the HIV and syphilis variation was explained through the spatially structured component, respectively. A significant negative correlation was found between the two diseases keep spatially structured effects (correlation -0.50; 95% CI: -0.79, -0.02), while for the unstructured effects, the correlation was also negative but not significant (-0.26; 95% CI: -0.59, 0.26). The spatially structured contributions fell to 50% and 55% for HIV and syphilis, respectively, when adjustments were made for the two factors; even the negative correlation between the two structured effects fell slightly and was no longer significant (correlation = -0.31; 95% CI: -0.77, 0.36). The spatially dependent variance (but not the unstructured one) was affected, meaning that deprivation and population density partially account for the spatial variation in both diseases.

Smoothed maps of HIV risk from fitting univariate HIV mapping models were generally similar to the raw prevalence maps (Fig. 1a), only that now they are more evenly distributed (maps not shown). Figs. 2a and 2b show covariates-adjusted maps of the probability that log-odds of the disease risk is above 0 using the HIV and syphilis bivariate mapping. The concentrations of high HIV and syphilis districts in the northeastern and south-western corridors, respectively, are now more apparent. The number of high-risk districts for either of the two diseases has fallen.

Discussion

Novel spatial epidemiological methods were used to investigate bivariate district-level distribution of HIV and syphilis prevalence among pregnant women in South Africa. Mapping of the spatial distribution of the diseases using Bayesian smoothing techniques helped us to visualise and assess the level of disease-specific spatial variation across the districts. We were also able to test for common and uncommon environmental aetiology between HIV and syphilis within a bivariate binomial model using a hierarchical Bayesian framework. The modelling approach based on spatial smoothing techniques nullified the effects of arbitrary administrative boundaries. Thus, any artefactual variation that may have been revealed through data aggregation was smoothed out. The result is that bordering provinces show smooth gradient in the risks of the diseases (Kleinschmidt et al., 2007). The results showed that the effects of deprivation and population density on HIV and syphilis were different. The effects were positively associated with HIV but negatively with syphilis. We also found evidence of discordant spatial distributions between two diseases. Higher HIV prevalence rates were more concentrated in the northeast than in the west while syphilis was highly concentrated in the western parts than elsewhere.

The discordant ecological and spatial effects between HIV and syphilis may suggest that the syphilis prevalence is not suited as a predictor of HIV prevalence. Thus, it cannot be used to investigate ecological differentials in risky sexual behaviour that exposes women to HIV infection. This was reinforced when we used individual data to assess the association between HIV and syphilis using binary logistic regression. Syphilis was found not to be linearly associated with HIV; or rather syphilis was not a predictor of HIV using individual-level data. There were also temporal differences, with HIV showing an increasing trend, whilst for syphilis; the trend was declining over the last decade. The differential temporal trends may indicate that the syphilis control and prevention programmes have been successful.

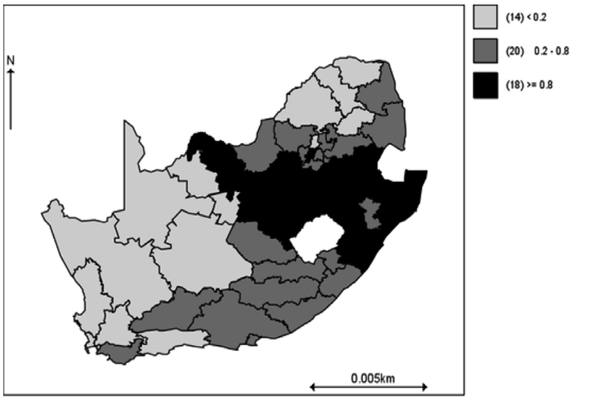

Fig. 2a. Map of the posterior probability that the smoothed covariate-adjusted log-odds for HIV prevalence in South Africa exceed 0, using the bivariate spatial model.

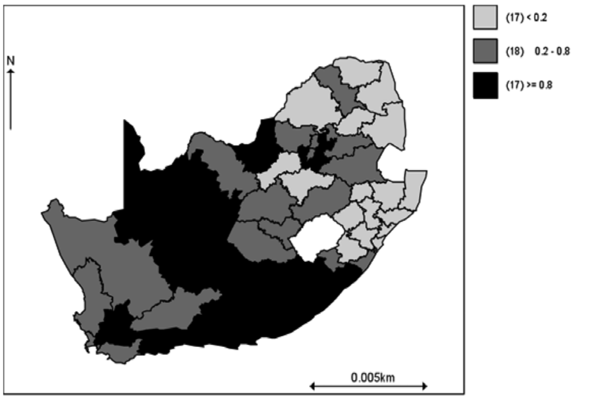

Fig. 2b. Map of the posterior probability that the smoothed covariate-adjusted log-odds for syphilis prevalence in South Africa exceed 0, using the bivariate spatial model.

The observed and estimated spatial patterns of HIV prevalence may merely reflect racial and age distributions in the different parts of the country. Younger women, who are predominantly in the urban areas, have higher HIV prevalence than older women; the rates are even higher among African women (Kleinschmidt et al., 2007). We postulate that the general low prevalence of HIV in the rural western parts may be due to reduced mobility associated with remoteness and economic inactivity, which are some of factors known to be protective against HIV infection (Kishamawe et al., 2006; Yahya-Malima et al., 2008). However, improved infrastructure in the rural areas has increased mobility, and within this context, it is conceivable that the HIV has not yet reached epidemic levels, or established itself, in the rural areas. Indeed, there is a shift in HIV infection from urban sites to rural areas as a result of oscillatory migrant labour when men spending long periods separated from their families (Johnson et al., 2005; Kleinschmidt et al., 2007; Hargrove, 2008). Thus, we might only conclude that HIV is still spreading to the rural areas that are now being exposed to an influx of residents that are relatively more mobile. The high prevalence rates of syphilis among the more rural north-western women compared to their urban counterparts in the north-eastern part might be attributable to low education levels, limited access to health care and high levels of poverty (Yahya-Malima et al., 2008). As a result, the rural areas might lag behind urban areas in terms of treatment adherence as many rural, pregnant women attend late for antenatal care (Mullick et al., 2005).

In using an adjacency matrix for the neighbourhood classification, we did not include data from districts in bordering countries in the classification. In particular, countries such as Swaziland (42.2%) in the southeast of the border with Mozambique, Botswana (38.5%) in the northwest and the Lesotho (28.4%) enclave have some of the highest antenatal HIV prevalence (UNAIDS, 2011). Thus, the border districts had missing neighbours, resulting in estimates that suffer from over- or under-estimation, i.e. the edge effects phenomenon. Nonetheless, the flexibility of Bayesian hierarchical smoothing has partially compensated for this as evidenced by high estimated HIV prevalence in districts situated on the borders of Swaziland, Botswana and Lesotho.

We used data from national antenatal sentinel HIV and syphilis prevalence surveys on pregnant women attending public health clinics. These data may not be representative for all pregnant women in a country or indeed the general adult population. HIV prevalence estimates from such surveys tend to be higher than estimates from population-based surveys (Montana et al., 2008). Thus, prevalence data emanating from antenatal clinics should be used with caution and preferably only used as an indicator of the potential number of HIV exposed and infected infants and for assessing needs for mother-to-child HIV transmission intervention programmes (UNAIDS/WHO, 2003). However, in a generalised HIV epidemic as in South Africa, estimates based on antenatal surveillance data are usually reliable. Moreover, these data provide the best available estimates of HIV infection among the South African females accessing antenatal services.

Conclusion

Our modelling and analysis approaches have provided an insight into similarities and differences in the ecological and geographical associations between HIV and syphilis prevalence in South Africa. The maps generated increase the visibility regarding differentials in material, social and health deprivation as well as risky sexual behaviour. Thus, they are useful for public health agencies and researchers with respect to prevention and management of HIV and syphilis among pregnant women in South Africa. In particular, preventive efforts should seek to address the possibility of the spread of the HIV to the rural areas. The results should also help in generating in-depth epidemiological investigations on what could be causing the interjectory between the districts. There is a huge demand for joint spatial modelling methods for multiple disease outcomes in the epidemiological field, and this study has demonstrated the feasibility and utility of such a model to an important application in public health.

Disclaimer

Thabang Mosala works for the National Department of Health, which conducts the South African antennal sentinel HIV and syphilis prevalence surveys in public health clinics. However, the findings and conclusions in this study are those of the authors and do not necessarily represent the views of the department. All authors declare that they have no competing interests.

Appendix A

The deprivation index was constructed from variables that were indicative of material and social deprivation. The variables were selected from the South African 2007 community survey and the 2005 and 2006 general household surveys. In particular, the following variables: the proportion of the population in the district that were below the age of 5 years; the proportion of the population of the district that were black Africans; the proportion of the population of the district that were from a household that was headed by a female; the proportion of the population of the district whose household heads had no schooling, the proportion of adults aged between 25 and 59 years in the district who were classified as unemployed; the proportion of the population in the district that lived in a traditional dwelling (informal shack or tent); the proportion of households in the district that had no piped water in their house or on site; the proportion of households in the district with a pit or bucket toilet or no form of toilet; the proportion of households that had no access to electricity or solar power for lighting, heating or cooking; were linearly combined using the first component from a principal component analysis (PCA). Higher values of the index denote higher levels of deprivation and *viceversa* (Health Systems Trust, 2009).

Appendix B

A bivariate binomial spatial model

Suppose $m_i$districts are neighbours of district*i*. The following bivariate, logit-link, spatial specification was adopted to model the bivariate for HIV and syphilis prevalence among pregnant women in South Africa.

$$\text{logit}\ (\pi_{i1t}) = \alpha_{01t} + \beta_1^T X_i + u^{(2)}_{1district(i)} + \sum_{k \in Neighbours(i)} w_{i,k}\ u^{(3)}_{1k}$$

$$\text{logit}\ (\pi_{i2t}) = \alpha_{02t} + \beta_2^T X_i + u^{(2)}_{2district(i)} + \sum_{k \in Neighbours(i)} w_{i,k}\ u^{(3)}_{2k} \tag{1}$$

where $u_{ij} = u^{(2)}_{jdistrict(i)}$ $(district(i) \in (1, 2, ..., I))$ are the disease-specific unstructured random effects and $v_{ij} = \Sigma_{k \in Neighbours(i)}\ w_{i,k}\ u^{(3)}_{jk} (Neighbours \in (1, 2, ..., m_i))$ are the spatially structured random effects, where, for both, time dependence has been dropped to concentrate only on spatial effects averaged over the three years. However, flexibility is still maintained by using disease-specific time-varying intercepts. The error $u^{(3)}_{jk}$ represents the effect of the k^{th} district upon other districts' prevalence rates for disease *j*. The effect is weighted by $w_{i,k}$, indicating the relevance of the k^{th} district to the i^{th} district (Browne et al., 2001). The weights could be a scaled representation of distance between two districts. The simplest weight model has $w^{(3)}_{i,k} = {}^1/m_i$ if district *i* and *k* are neighbours and 0 otherwise. Thus, all districts that border a particular district are part of neighbour classification for that district. The direct district effects $u^{(2)}_{jdistrict(i)}$ are modeled as $u^{(2)}_{jdistrict(i)} \sim Normal\ (0,\ \sigma^2_{uj(2)})$ and the neighbour district effects $u^{(3)}_{jk}$ by $u^{(3)}_{jk} \sim Normal\ (0,\ \sigma^2_{uj(3)})$. Thus the spatial structured effect v_{ij} is normally distributed with mean *0* and variance $\sigma^2_{uj(3)}/m_i$. Thus, the between district spatial dispersion is inversely proportional to the number of neighbours a district has. We model the disease-specific time-intercepts α_{0jt} with a random walk prior $\alpha_{0j1} \sim Normal\ (0\ \sigma^2_\alpha)$; $\alpha_{0jt} \sim Normal\ (\alpha_{0j(t-1)},$ $\sigma^2_\alpha)$ for t >1. The model presented in (1) is a simplified version of the spatio-temporal model presented in Manda et al. (2009), where space-temporal interactions were included. In the present study, we have very few time points to ascertain any prevailing trends.

There are both statistical and epidemiological advantages of combining multiple disease outcomes using a joint disease modeling. By incorporating information from several related outcomes, the efficiency of the resulting fixed and random parameter estimates improve (Manda and Leyland, 2007). From an epidemiological perspective, a joint model allows inferences to be made on the similarities as well as the differences with respect to the effect of risk factors (Manda et al., 2009). In particular, any discordant patterns in the spatial distributions of observed and unobserved HIV and syphilis in districts would reflect differences in risk factors beyond those that are well-known. In using the bivariate disease prevalence model, our primary joint model is similar to that in Crainiceanu et al. (2006), where, instead, Bayesian geostatistical model was employed to construct spatial dependence (Diggle et al., 1998).

Correlation model specification

We might expect the four random effects to be correlated. This is achieved by modeling the four effects as arising from a multivariate normal prior distribution (Leyland et al., 2000; Manda and Leyland, 2006). For simplicity, suppose ϕ_i is the overall district-level spatial vector with elements $\phi_i = (u^{(2)}_1,\ u^{(2)}_2,\ u^{(3)}_1,\ u^{(3)}_2)$, a vector of unstructured direct district effect for diseases 1 and 2, and neighbour district effect for diseases 1 and 2, respectively. The random effects vector ϕ_i is set to have a mean vector $\theta_\phi = (0, 0, 0, 0)$ and covariance matrix Σ_{ϕ}, having diagonal elements ($\sigma^2_{u1(2)}$, $\sigma^2_{u2(2)}$, $\sigma^2_{u1(3)}$, $\sigma^2_{u2(3)}$) and (upper) off-diagonal elements ($\sigma_{u1(2)u2(2)}$, $\sigma_{u1(2)u1(3)}$, $\sigma_{u1(2)u2(3)}$, $\sigma_{u2(2)u1(3)}$, $\sigma_{u2(2)u2(3)}$, $\sigma_{u1(3)u2(3)}$). This specification provides interpretation for variances and covariances between disease risk profiles within and between districts (Leyland et al., 2000).

Other prior specification

In the analyses, the fixed effect parameters in β were assigned independent Normal *(0, 1000)* priors. All scalar precision parameters were assigned independent hyper prior Gamma (0.5, 0.0005) distributions, and 4-way covariance matrices were appropriately assigned Wishart (Q, 4) prior distribution, where the scale parameter Q had entries of 1s on the diagonal and 0s off-diagonal.

Code

The WinBUGS code for the main model, involving the bivariate HIV and syphilis prevalence rates with the two covariate effects and the four random effects, is provided below for reference. The "data node adj" contains a set of adjacent districts for each district, and "num" contains a cumulative total count of the number of neighbours for each district, e.g. if district 1 and 2 have six and three neighbours, then num=c(0, 6, 9, .., NN).

Model

```
{
        for (t in 1 : T) {# no of time points
                for(i in 1 : Nareas) {# no of areas
                for (j in 1:D){ #no of diseases
                y[i,t,j] ~ dbin(pi[i,t,j], N[i,t]) # bino-
mial model/likelihood
         #
         # logit link function with fixed and two spa-
tial (phi and psi) effects effects
                                        logit(pi[i,t,j])<-
alpha0[t,j]+beta[1,j]*pdn2[i]+beta[2,j]*pdn3[i]+

beta[3,j]*pdn4[i]+beta[4,j]*pdn5[i]+beta[5,j]*deprv2[
i]+
             beta[6,j]*deprv3[i]+beta[7,j]*deprv4[i] +
beta[8,j]*deprv5[i]+
                phi[i,j]+psi[i,j]

                }
                }
                }

                # defining unstructured and struc-
tured error terms

                for (i in 1:Nareas) {
                phi[i,1:M]          ~dmnorm(zero[],
omega[,])
# M is number of size of random effect; here M=4
        psi[i,1]<-mean(W1[num[i]+1:num[i+1]])
        psi[i,2]<-mean(W2[num[i]+1:num[i+1]])
                        }

# determining effects of area j on the other areas,
whose
weighted mean gives spatial effects for area i
                for (j in 1:NN) { W1[i]<-phi[adj[i],3]
                              W2[i]<-phi[adj[i],4]
}

                # prior for the intercepts
                alpha0[1,1]       ~        dnorm(0,
prec.alpha0[1])
                alpha0[1,2]       ~        dnorm(0,
prec.alpha0[2])
            logit(HIV_prev0[1,1])<-alpha0[1,1]
                logit(HIV_prev0[1,2])<-alpha0[1,2]
                for (t in 2:T) {
                for (j in 1:D){
                        alpha0[t,j] ~dnorm(alpha0[t-
1,j],prec.alpha0[j])

logit(HIV_prev0[t,j])<-alpha0[t,j]
                        }
                        }
                # prior for fixed effects
                 for (j in 1:lbeta) { for (k in 1:D)
{beta[j,k] ~ dnorm(0,0.001)}}
                 for (j in 1:lbeta) { for (k in 1:D)
{ORbeta[j,k] <-exp(beta[j,k]) }}

                # prior for univariate precisions
                for (j in 1:D){
                prec.alpha0[j]~ dgamma(1,1)
                sigma.alpha0[j]<-
sqrt(1/prec.alpha0[j])
                }
                # prior for the precision/covariance
matrix for MVMM model
                omega[1:M, 1:M] ~ dwish(R[ , ], M)
        # Precision matrix of MVMM
```

```
                    sigma[1:M, 1:M] <- inverse(omega[ ,
])                                    #       #
within-area conditional correlation between spatial



                    #
            for (i in 1:Nareas){
                                # prob log-odds HIV > 0 (the
average log-odds=0)
                                # prob log-odds Syphilis > 0
(the average log-odds=0)

                                    total_logOdds[i,1]     <-
phi[i,1]+psi[i,1]
                                    total_logOdds[i,2]     <-
phi[i,2]+psi[i,2]


                                                ProbHIV[i]<-
step(total_logOdds[i,1]-0.0000001)
                                                 ProbSyp[i]<-
step(total_logOdds[i,2]-0.0000001)

                                }

}
```

References

Besag J, York J, Mollie A, 1992. Bayesian image restoration, with two applications in spatial statistics (with discussion). Ann Inst Statist Math 43, 1-59.

Browne WJ, Goldstein H, Rasbash J, 2001. Multiple membership multiple classification (MMMC). Stat Model 1, 103-124.

Cohen D, 1998. Socioeconomic causes and consequences of the HIV epidemic in Southern Africa: the case of Namibia, UNDP Issues Paper No. 31, 1998.

Crainiceanu CM, Diggle PJ, Rowlingson B, 2006. Bivariate binomial spatial modelling of *Loa loa* prevalence in tropical Africa. Johns Hopkins University; Dept. of Biostatistics Working Papers, Paper 103.

Department of Health, 2011. National antenatal sentinel HIV and syphilis prevalence survey in South Africa 2010 Report. Department of Health, Pretoria, South Africa.

Diggle PJ, Moyeed RA, Tawn JA, 1998. Model based geostatistics (with discussions). J Roy Stat Soc-App 47, 299-326.

Feltbower RG, Manda SOM, Gilthorpe MS, 2005. Detecting small area similarities in the epidemiology of childhood acute lymphoblastic leukaemia and type 1 diabetes: a Bayesian approach. Am J Epidemiol 161, 1168-1180.

Gelman A, Rubin DB, 1992. Inference from iterative simulation using multiple sequences. Stat Sci 7, 457-472.

Hargrove J, 2008. Migration, mines and mores: the HIV epidemic in Southern Africa. South Afr J Sci, 104, 53-61.

Health Systems Trust, 2009. The district health barometer 2008/09. Health Systems Trust, Durban.

Johnson LF, Coetzee DJ, Dorrington RE, 2005. Sentinel surveillance of sexually transmitted infections in South Africa: a review. Sex Transm Dis 81, 287-293.

Kleinschmidt I, Pettifor A, Morris N, MacPhail C, Rees H, 2007. Geographic distribution of human immunodeficiency virus in South Africa. Am J Trop Med Hyg 77, 1163-1169.

Kishamawe C, Vissers DC, Urassa M, Isingo R, Mwaluko G, Borsboom GJ, Voeten HA, Zaba B, Habbema JD, de Vlas SJ, 2006. Mobility and HIV in Tanzanian couples: both mobile persons and their partners show increased risk. Aids 20, 601-608.

Leyland AH, Langford IH, Rasbash J, Goldstein H, 2000. Multivariate spatial models for event data. Stat Med 19, 2469-2478.

Lunn DJ, Thomas A, Best N, Spiegelhalter D, 2000. WinBUGS – a Bayesian modelling framework: concepts, structure, and extensibility. Stat Comp 10, 325-337.

Meidany F, Horikoshi Y, Rohde J, 2000. HIV prevalence rate and population density: Eastern Cape experience, South Africa International Conference on AIDS.

Manda SOM, Leyland A, 2007. An empirical comparison of maximum likelihood and Bayesian estimation methods for multivariate spatial disease model. South Afr Stat J 41, 1-21.

Manda SOM, Feltbower RG, Gilthorpe MS, 2009. Investigating spatial-temporal similarities in the epidemiology of childhood leukaemia and diabetes. Eur J Epidemiol 24, 743-752.

Montana L, Neuman M, Mishra V, 2007. Spatial modelling of HIV prevalence in Kenya. DHS Working Papers. No 27, Calverton.

Montana, LS, Mishra V, Hong R, 2008. Comparison of HIV prevalence estimates from antenatal. Sex Trans Infect 84, i78-i84.

Mullick S, Beksinksa M, Msomi S, 2005. Treatment for syphilis in antenatal care: compliance with the three dose standard treatment regimen. Sex Trans Infect 81, 220-222.

Pettifor AE, Kleinschmidt I, Levin J, Rees HV, MacPhail C, Madikizela-Hlongwa L, Vermaak K, Napier G, Stevens W, Padian, NS, 2005. A community-based study to examine the effect of a youth HIV prevention intervention on young people aged 15–24 in South Africa: results of the baseline survey. Trop Med Int Health 10, 971-980.

Sandøy IF, Gunnar KG, Michelo C, Fylkesneset K, 2006. Antenatal clinic-based HIV prevalence in Zambia: declining trends but sharp local contrasts in young women. Trop Med Int Health 11, 917-928.

Shisana O, Rehle T, Simbayi LC, Zuma K, Jooste S, Pillay-van-Wyk V, Mbelle N, Van Zyl J, Parker W, Zungu NP, Pezi S, the SABSSM III Implementation Team, 2009. South African national HIV prevalence, incidence, behaviour and communication survey 2008: a turning tide among teenagers? HSRC

Press, Cape Town, South Africa.

Spiegelhater DJ, Best NG, Carlin BP, van der Linde A, 2002. Bayesian measures of model complexity and fit. J Roy Stat Soc Ser B 64, 583-639.

UNAIDS/WHO, 2003. Guidelines for conducting HIV sentinel serosurveys among pregnant women and other groups / UNAIDS/WHO Working Group on Global HIV/AIDS and STI Surveillance. WHO Library Cataloguing-in-Publication Data. Geneva, Switzerland.

UNAIDS, 2010. Global report: UNAIDS report on the global AIDS epidemic: 2010: Available at: http://www.unaids.org/globalreport/Global_report.htm (accessed on January 2012).

Yahya-Malima KI, Evjen-Olsen B, Matee MI, Fylkesnes K, Haarr L, 2008. HIV-I, HSV-2 and syphilis among pregnant women in a rural area of Tanzania: prevalence and risk factors. BMC Infect Dis 8, 75.