

Localized spatial clustering of HIV infections in a widely disseminated rural South African epidemic

Frank Tanser,^{1*} Till Bärnighausen,^{1,2} Graham S Cooke^{1,3} and Marie-Louise Newell^{1,4}

Accepted 27 January 2009

Background South Africa contains more than one in seven of the world's HIV-positive population. Knowledge of local variation in levels of HIV infection is important for prioritization of areas for intervention. We apply two spatial analytical techniques to investigate the micro-geographical patterns and clustering of HIV infections in a high prevalence, rural population in KwaZulu-Natal, South Africa.

Methods All 12 221 participants who consented to an HIV test in a population under continuous demographical surveillance were linked to their homesteads and geo-located in a geographical information system (accuracy of <2 m). We then used a two-dimensional Gaussian kernel of radius 3 km to produce robust estimates of HIV prevalence that vary across continuous geographical space. We also applied a Kulldorff spatial scan statistic (Bernoulli model) to formally identify clusters of infections ($P < 0.05$).

Results The results reveal considerable geographical variation in local HIV prevalence (range = 6–36%) within this relatively homogenous population and provide clear empirical evidence for the localized clustering of HIV infections. Three high-risk, overlapping spatial clusters [Relative Risk (RR) = 1.34–1.62] were identified by the Kulldorff statistic along the National Road ($P \leq 0.01$), whereas three low risk clusters (RR = 0.2–0.38) were identified elsewhere in the study area ($P \leq 0.017$).

Conclusions The findings show the existence of several localized HIV epidemics of varying intensity that are partly contained within geographically defined communities. Despite the overall high prevalence of HIV in many rural South African settings, the results support the need for interventions that target socio-geographic spaces (communities) at greatest risk to supplement measures aimed at the general population.

Keywords HIV seroprevalence, geographical information systems, disease clustering, small-area variation

¹ Africa Centre for Health and Population Studies, University of KwaZulu-Natal, Mtubatuba, South Africa.

² Department of Global Health and Population, Harvard School of Public Health, Boston, USA.

³ Imperial College, London, UK.

⁴ Centre for Paediatric Epidemiology and Biostatistics, Institute of Child Health, University College London, UK.

* Corresponding author. Africa Centre for Health and Population Studies, University of KwaZulu-Natal, PO Box 198, Mtubatuba 3935, South Africa.

E-mail: tanserf@africacentre.ac.za

Introduction

South Africa has experienced one of the fastest growing HIV epidemics in the world and is currently home to more than one in seven of the world's HIV-positive population.¹ HIV prevalence varies greatly at a provincial level² and is highest in the province of KwaZulu-Natal where 39% of antenatal attendees are HIV positive.³ Efforts to curtail the devastating impacts of the epidemic have been disappointing and HIV incidence remains high.⁴ Consequently, new strategies are required that make better use of the limited resources available.

The spatial structure of sexually transmitted infections (STI) epidemics can have a profound impact on the epidemic dynamics, future spread, persistence and the nature and success of interventions.⁵ Most risk and health-promoting behaviours are not distributed uniformly across populations, but tend to cluster in specific communities.⁶ Behaviour is patterned by the social structure in which individuals are positioned, placing individuals in some communities 'at risk of risks'.⁷ Research has demonstrated that many STIs are not equally distributed across geographic areas or communities.⁸⁻¹¹ As such, the local community is increasingly being seen as critical to understanding the spread of HIV and key to prevention efforts.¹²⁻¹⁶ Given this background and the problems associated with the identification of the most active individuals and groups, geographic approaches to HIV prevention have received increasing attention.^{17,18} However, it has been unclear whether localized HIV interventions targeting specific communities could be effective in a rural South African context where prevalence levels are amongst the highest in the world. Consequently, no such interventions have been attempted in such a setting.

A recent editorial commented that focusing scale-up of services where they are needed requires 'knowing your epidemic', globally and locally.¹⁹ However, because of the difficulty of obtaining spatially referenced HIV data at the local level, aggregated HIV outcomes are frequently available only by large geographic units. These units are often too coarse to capture the true scale of the processes under study. There are potentially numerous disadvantages associated with this type of analytical approach that make assumptions about the size and characteristics of the 'community' and location of participants and can mask important sub-area variation and cause misinterpretation of true underlying spatial patterns.^{20,21} In addition, it is important to distinguish 'real' from 'apparent' excess number of cases (particularly in large, sparsely populated areas) and to deal appropriately with elements of chance including the problem of multiple significance testing.²² Here we present results from a population-based HIV sero-survey in rural KwaZulu-Natal, South Africa in which we geolocate more than 12 000 study participants to an accuracy of <2 m. We apply two innovative spatial techniques to investigate the micro-geographical variation

in HIV prevalence and clustering of HIV infections at a truly local level.

Methods

Study area

The study area (Figure 1) is located in Hlabisa sub-district (one of five sub-districts in the rural district of Umkhanyakude in northern KwaZulu-Natal), ~250 km north of the city of Durban. It covers 438 km² (approximately one-third of the sub-district) and contains a total population of about 87 000 Zulu-speaking people. The majority of the population live in scattered homesteads that are not concentrated into villages or compounds. Running along the eastern boundary of the study area is the National Road, which links the city of Durban with Mozambique, Swaziland and the province of Mpumalanga. The study area is typical of many rural South African settings in that while it is largely rural it also contains a formal urban township and a series of high-density settlements located predominantly along the major transport routes.

The Africa Centre Demographic Information system (ACDIS) and population-based HIV surveillance

The study forms part of the population-based household and HIV surveillance that is linked to the Africa Centre Demographic Information System.²³ Every 6 months the surveillance system collects demographic, socio-economic and behavioural data on individuals resident at one of the 11 000 homesteads in the study area. The annual HIV surveillance among residents included all women aged 15-49 years and

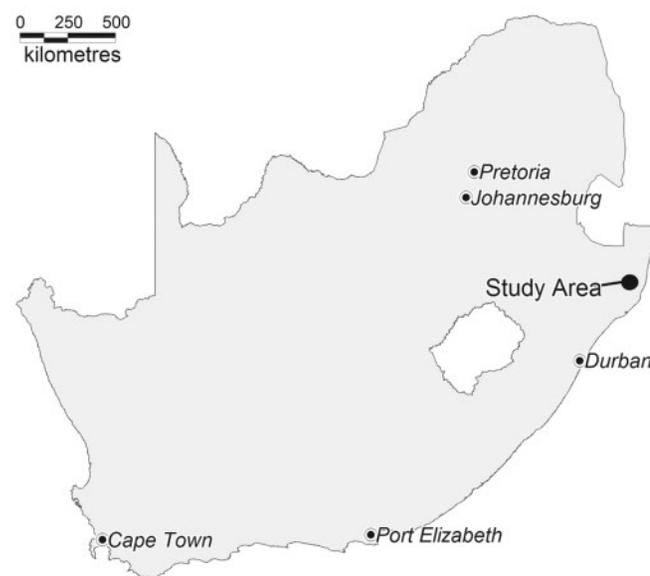


Figure 1 Location of the study area in South Africa

men 15–54 years and was conducted from 2003 to 2004. The demographic and socio-economic characteristics of the cohort have been described in detail previously.²⁴ Dried blood spot samples were taken during home visits and HIV serology performed using an ELISA methodology. Informed consent was sought and ethical approval received from the University of KwaZulu Natal (E029/2003). The consent rate among those contacted was 58%.²⁴ The age-adjusted HIV prevalence was 27% for women and 14% for men. Prevalence was highest among females aged 25–29 years (51%) and in men aged 30–34 years (44%).²⁴ The average incidence of HIV infection from June 2003 to December 2005 was 3.2 per 100 person-years (95% CI: 2.8–3.8), peaking in the age group of 25–29 years for both men and women.⁴ All homesteads in the surveillance area have been positioned to an accuracy of <2 m using differential global positioning systems as part of the surveillance programme.²⁵

Spatial analysis

All 12 221 adult residents (women aged 15–49 years and men 15–54 years) who consented to an HIV test were geo-located to their homestead of residence (Figure 2). For the purpose of protecting participant confidentiality, a small random error has been incorporated into the geographical position of each participant. This also serves to create an accurate visual representation of the distribution of all participants in the survey due to multiple participants sometimes being resident in a single homestead.

To produce robust estimates of HIV prevalence that vary across continuous geographical space we used a Gaussian kernel methodology.²⁶ The Gaussian kernel does not impose any static geographical boundaries on the data and produces prevalence estimates that are sensitive to local variation whilst at the same time being relatively robust to the effects of random noise.

The HIV status of each participant was superimposed onto a 30 × 30 m grid (representing the study area) at the location of individual's place of residence in Idrisi Andes (Clark labs, Clark University, MA, USA). We then passed a two-dimensional standard Gaussian kernel of radius 3 km over the grid. The kernel moves systematically across the grid and calculates a Gaussian-weighted prevalence estimate for every cell's neighbourhood. The resulting prevalence estimate is placed onto a new map at the same location as the central cell. The kernel is then moved one cell to the right (and then down one row at the end of the row) and the process repeated. The size of the kernel was determined on the basis of a spatial variogram²⁷ (constructed using HIV prevalence estimates aggregated by administrative ward), which showed there to be clear spatial dependence in the resulting HIV prevalence estimates within a distance of 3 km. A median of 1134 participants of known HIV sero-status was evaluated for each cell's unique neighbourhood.

The resulting continuous HIV prevalence estimates were then converted into a contour map showing lines of equal prevalence at 5% intervals.

Cluster detection

To formally identify clusters of infection (either high or low numbers of infections) we applied a Kulldorff spatial scan statistic (implemented within the SaTScan spatial cluster detection programme²⁸). A spatial scan statistic is a cluster detection test that is able to both detect the location of clusters and evaluate their statistical significance without the problems associated with multiple testing. This is done by gradually scanning a window across space. The general statistical theory behind the spatial scan statistics is described in detail elsewhere.²⁹ The scan statistic adjusts for the uneven geographical density of a background population and the analyses are conditioned on the total number of cases observed.

The spatial scan statistic imposes a circular window on a map and it allows the centre of the circle to move across the study region. For any given position of the centre, the radius of the circle changes continuously so that it can take any value from zero up to a specified maximum value. For each potential cluster, a likelihood ratio test statistic was used to determine if the number of HIV cases within the potential cluster was higher than expected. Expected numbers of cases were calculated on the basis of the null hypothesis of complete spatial randomness by assuming that the number of HIV infected individuals cases in each circle is an independent Bernoulli random variable with a constant prevalence. The circle with the maximum likelihood is defined as the most likely cluster, implying that it is least likely to have occurred by chance. The maximum observed value of the test statistic for each possible cluster is then compared with the overall distribution of maximum values. The *P*-value of the statistic is obtained through Monte Carlo hypothesis testing (9999 iterations), where the null hypothesis of no clustering is rejected if the simulated *P*-value is <0.05. We allowed the clusters to overlap by <50% and set the maximum search radius of the circle to be 3 km to facilitate comparison with the output generated from the Gaussian kernel smoothing approach. The resulting relative risk estimates were adjusted by sex and 5-year age band.

Characterization of clusters

We used data from the Africa Centre's household socio-economic survey (comprising a total of 29 217 individuals >18 years of age and 10 856 households) undertaken during the same period as the HIV survey and described previously²³ to briefly characterize any high- and low-risk clusters identified by the spatial scan statistic in terms of levels of education, household wealth (as measured by numbers of assets owned by the household), marriage, numbers of

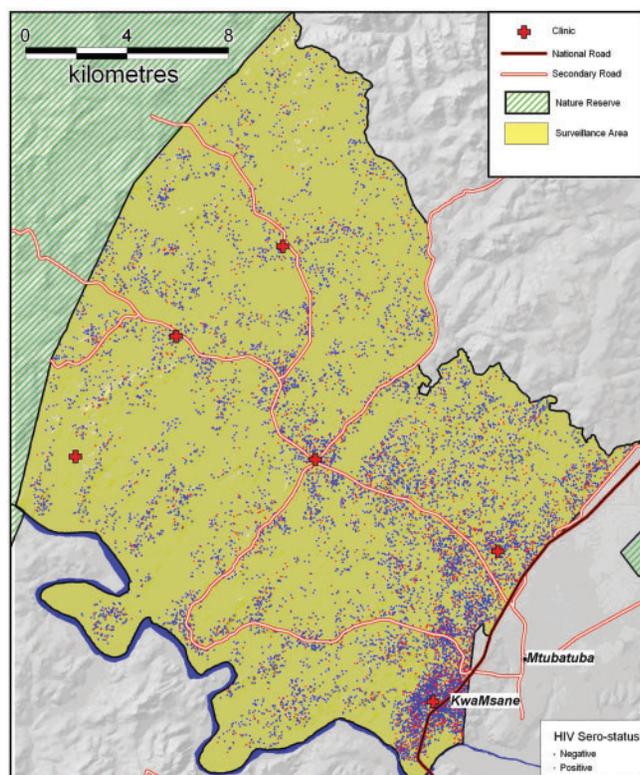


Figure 2 Map of the study area showing the approximate location (incorporating an intentional random error) of all 12 221 participants coded by HIV status

non-resident household members ('migrants') and levels of employment.

Results

The results of the Gaussian kernel analysis reveal considerable spatial heterogeneity in HIV prevalence (range = 6–36%) across the study area (Figure 3). In general, high-density settlements in the south-east of the study area near the National Road have the highest HIV prevalence (>35%) whilst the more inaccessible rural areas near the western boundary have the lowest HIV prevalence (<10%). HIV cases are heavily concentrated in the areas of the urban township and high-density settlements near the National Road where the highest population density coincides with highest HIV prevalence. The results show that 40% of infected individuals live within 1 km of the National Road (Figure 4) and there is a steep fall-off in numbers of infections with increasing distance from the road. The estimated density of HIV-infected individuals (total HIV cases per square kilometre) living within 1 km of the road is 15.7 times higher than the mean density of infected individuals across the remainder of the study area. The exclusion of the township from the analysis had little impact on these results (Figure 4).

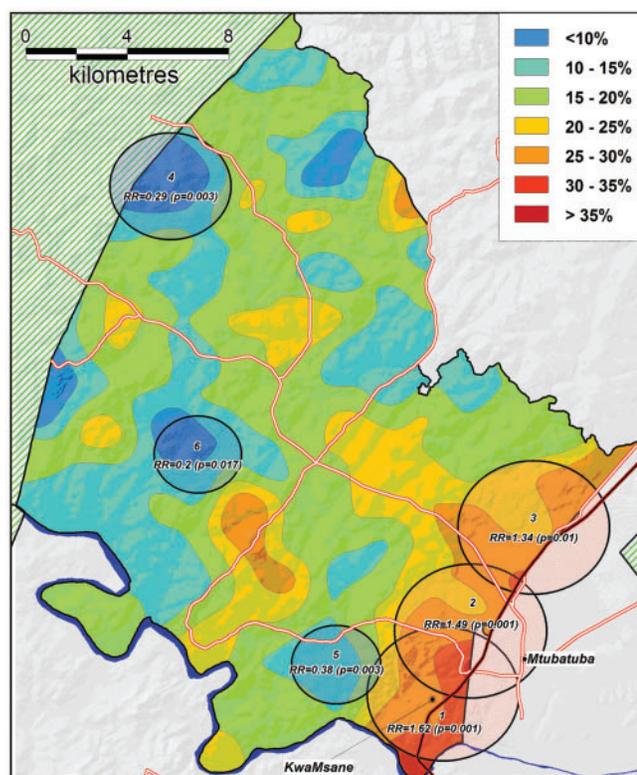


Figure 3 Geographical variations in HIV prevalence by 5% intervals across the surveillance area (for all women aged 15–49 years and men 15–54 years) obtained by a 3-km standard Gaussian kernel (mean = 21.7%, 95% CI 20.1–22.5). Superimposed on the map are the clusters independently identified by the Kulldorff spatial scan statistic (blue shading, low relative-risk; red, high relative-risk). The National Road can be seen next to KwaMsane Township continuing along the eastern boundary of the surveillance area towards Mozambique

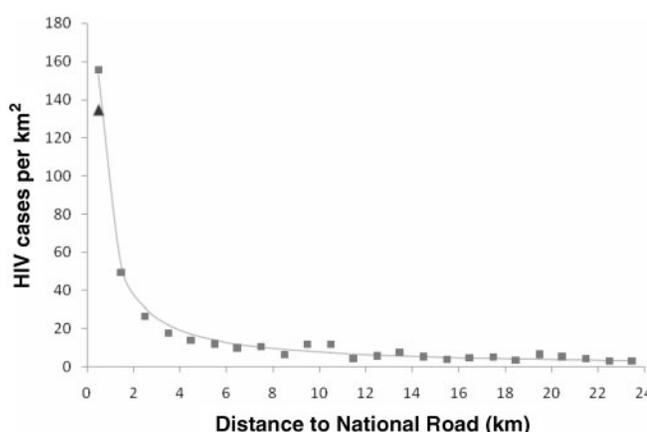


Figure 4 Graph showing the inverse relationship between distance to National Road and the estimated number of HIV-positive residents per square kilometre (for all resident women aged 15–49 years and men 15–54 years). The triangle in the graph indicates the datapoint generated after the exclusion of the population in the township (which is fully contained within 1 km of the National Road) from the analysis

Table 1 Clusters of infections identified by the Kulldorff spatial scan statistic ($P < 0.05$) across the surveillance area (Figure 3)

	Cluster number	Radius (km)	Prevalence (%)	Total tested	RR	Log likelihood	P-value	Description
High RR	1	3.0	30.5	2754	1.62	83.11	0.001	Urban/high-density settlement
	2	3.0	29.3	2051	1.49	44.25	0.001	High-density settlement
	3	3.0	27.6	1099	1.34	14.00	0.01	High-density settlement
Low RR	4	2.39	6.2	178	0.29	15.99	0.003	Deep-rural
	5	1.76	8.4	237	0.38	15.48	0.003	Semi-rural
	6	1.73	4.4	114	0.20	13.58	0.017	Rural

Table 2 Summary table showing the basic demographic and socio-economic characteristics^a of the high- and low-prevalence clusters respectively identified by the Kulldorff spatial scan statistic

	High-prevalence clusters	Low-prevalence clusters	P-value of difference
HIV prevalence (%)	28.67	6.81	
Total tested	4600	529	
Distance from the National Road ^b (km)	0.65	13.30	<0.0001
Non-resident household members (%)	32.3	38.1	<0.0001
Employment (%)	40.94	20.91	<0.0001
Currently married (%)	19.9	24.2	0.001
Mean household assets ^c	7.73	6.66	<0.0001
Mean number of years of education completed ^d	7.98	6.12	<0.0001

^aThe data comprise a total of 12 652 individuals (>18 years of age) and 4943 households who were resident in the six clusters and answered the household socio-economic questionnaire.

^bMeasured from cluster centre to the National Road.

^cOut of a total number of 27 household assets listed in the household socio-economic questionnaire.

^dSuccessful completion of primary school = 7 years.

The location of all clusters identified by the Kulldorff spatial scan statistic ($P < 0.05$) corresponded well with the spatial distribution in HIV prevalence generated by the Gaussian Kernel smoothing approach (Figure 3). The scan statistic identified three clusters with large excess numbers of HIV infections (RR = 1.34–1.62) in communities along the National Road. Three other communities were identified containing substantially smaller numbers of infections (RR = 0.2–0.38) relative to expectation (Table 1). These variations are unlikely to have arisen by chance alone (P -value range = 0.001–0.017). Two of the low-risk clusters identified (Clusters 5, 6) are rural but relatively close to high-density settlements whereas cluster 4 is located in a 'deep-rural' area. One of the low-prevalence clusters (Cluster 5) is juxta-posed against a high-prevalence cluster (equating to a four-fold difference in risk of infection between the two adjacent clusters) suggesting considerable constraining of sexual mixing patterns within the low-prevalence community. Some high- and low-prevalence areas identified by the Gaussian kernel approach were not characterized as clusters by the scan statistic because the role of chance in creating the variations could not be entirely discounted (at the $P < 0.05$ significance level).

The high-prevalence communities identified by the spatial scan statistic were characterized by higher aggregate levels of education, household wealth and nearly double the rate of employment, but had lower levels of marriage and proportion of non-resident household members ('migrants') in comparison to their low-prevalence counterparts (Table 2).

Discussion

We apply for the first time two complimentary spatial analytical techniques (kernel smoothing and spatial cluster detection using the Kulldorff spatial scan statistic) to investigate the geographical patterns and clustering of HIV infections at a truly local level. Despite the stage and severity of the epidemic there is remarkable geographical variation in HIV prevalence within this relatively homogenous population that has been exposed to HIV for more than two decades. We find strong evidence for the clustering of HIV infections in communities along one of South Africa's National Roads. This demonstrates that risks for HIV are associated with specific socio-geographic spaces (communities), and provide an opportunity for

targeted interventions to supplement existing measures aimed at the general population. The results challenge the prevailing paradigm of a ubiquitous 'generalized' rural epidemic. Rather our findings reveal the existence of several localized epidemics of varying intensity that are partly contained within geographically defined communities. Consequently, a one-size-fits-all intervention strategy may not be effective in such a setting where such marked variations in epidemiological and socio-geographical context exist.³⁰

The large (and sometimes abrupt) spatial variations in HIV prevalence further highlight the importance of using a spatial analytical approach that does not rely on aggregation of infections by coarse administrative unit. Unless the administrative unit's boundaries happen to coincide with the 'boundary' of the cluster, aggregation will mask important sub-area variation and may result in misinterpretation of true underlying spatial patterns. However, the spatial methods used do have some limitations. Most importantly, the choice of the size of the radius used in the Gaussian kernel can impact significantly on the final result. The appropriate choice of the size of the kernel is described elsewhere³¹ and always depends on the purpose intended for the smoothed estimate. The smaller the radius used, the greater the range in prevalence estimates obtained and the greater the sensitivity to local variation. The use of a larger kernel will result in smoothing towards the mean and important variation in prevalence may be lost. The choice of the 3-km radius of the kernel used in this study was informed largely by empirical data from spatial dependence in ward-level HIV prevalence. By using a relatively large kernel, we have produced stable estimates that reliably demonstrate the large variations in prevalence in the study area. Although the precise mathematical form of the kernel can also impact on the final result, this effect is relatively modest in comparison to the size of the kernel.²⁶ A limitation of the Kulldorff spatial scan statistic is that clusters are arbitrarily defined as circles. Given the highly heterogeneous spatial distribution of the population and HIV prevalence distribution, an elliptical scan window (with long axis orientated parallel to the National Road) might be appropriate for the detection of excess infections in this setting.

Although widely applied to the detection of clustering of non-communicable diseases, especially cancer,³²⁻³⁴ the application of spatial scan statistics to sexually transmitted diseases has been limited and occurred only in the developed world. The methods have been applied to the clustering of gonorrhoea in the USA⁸ and Chlamydia trachomatis in Canada³⁵ and the clustering of three bacterial STI infections (gonorrhoea, chlamydia and syphilis) in Australia.³⁶ In the developing world, the methods have been used to investigate spatial patterns of childhood mortality in Burkina Faso.³⁷ Common to all these studies

was the use of disease data aggregated by area such as census tract or administrative unit. We could not locate any studies that had applied kernel density smoothing methods to analysing the spatial distribution of a sexually transmitted disease. However, examples of its use in other diseases include cancer,^{38,39} severe acute respiratory syndrome,⁴⁰ dengue⁴¹ and tick-borne encephalitis and Lyme borreliosis.⁴²

Whereas ideally the rate of new HIV infections (incidence) should be used to inform prevention activities, we analyse the distribution of existing HIV infections (prevalence) and as such the results may have some limitations. However, one would expect an individual's risk of acquisition of HIV infection to be strongly related to HIV prevalence in the surrounding community. Indeed, through subsequent rounds of HIV surveillance data we have started to analyse spatial patterns of HIV incidence in the study area and find that they are remarkably similar to prevalence in spatial distribution.

Our findings are consistent with previous work in this population. For example, we found at the level of the individual, that after controlling for age, sex, setting (urban/rural) and educational attainment there is a strong negative association in HIV infection risk with increasing distance to the National Road (F Tanser *et al.*, in preparation). Those living within 1km of the National Road are twice as likely to be HIV-infected (adjusted odds ratio = 1.96, 95% CI: 1.5-2.5) in comparison to someone living >15km away. Likewise, for incidence, we found in multiple regression analysis that the hazard of HIV acquisition decreased with the logarithmically transformed distance to the National Road (adjusted hazard ratio = 0.856, $P = 0.002$).⁴ These findings support ecological observations made in this area⁴³ and other parts of Africa⁴⁴⁻⁴⁶ and suggest that individuals living in communities with better access to transport and transport routes are at higher risk of infection. A recent study in Uganda showed an association between distance to main roads and increase in both HIV incidence and genetic complexity of the virus.⁴⁶ Individuals from communities located near a main road were twice as likely to be infected with an HIV-1 recombinant strain in comparison with those living near a secondary road.

It is widely accepted that spatial heterogeneity and spatial processes (such as circular migration patterns) during the early phase of the epidemic led to the explosive spread of HIV in South Africa.⁴⁷ However, given the advanced stage of the epidemic, the apparent widespread high HIV prevalence and long duration of infection, it might be expected that a more even spatial distribution would result at this stage of the epidemic, at least within a small, circumscribed rural population. The fact that such marked spatial variation persists suggests that spatial processes, complex sexual mixing patterns and community effects

continue to impact on transmission dynamics at this advanced stage of the epidemic.

Additional work is required to fully understand the reasons for these large spatial variations in HIV prevalence in this population and important insights are likely to be gained by further in-depth study of the communities indentified in this research. A combination of individual, household, community, structural and geographical risk factors are likely responsible for the substantial heterogeneity observed. These include the possibility of higher rates of sex work along the National Road,⁴⁵ hierarchical diffusion processes of the virus along the transport routes during the early stages of the epidemic,⁴⁸ higher population mobility and migration rates in populations living near the roads,⁴⁹ geographical clustering of the most sexually active groups in high-density communities, more dissortative sexual mixing patterns within the urban township and high-density settlement populations,⁵⁰ possibility of higher prevalence of other STIs (many of which are linked more directly to the behaviour of high-risk groups⁵¹) in some areas⁵² and systematic geographic variations in other individual, household and community level risk factors.⁵³

Research and theory point to the need for HIV prevention interventions that focus both on persons at greatest risk as well as the general public.⁵⁴ Targeting efforts at settings where HIV transmission is most intense is crucial. A decrease in HIV incidence in these communities, more extensive coverage of antiretroviral treatment (ART), increased knowledge of infection status, an increase in condom usage and introduction of venue-based interventions could therefore have a disproportionately large impact on

the general population as a whole. Despite the overall high-prevalence levels in many rural South African settings, our results call for a 'place-based'¹⁷ approach centred on the identification of high-risk communities to supplement prevention efforts aimed at the general population.

Funding

IR01-HD058482-01 from the National Institute of Child Health and Human Development. Core funding for the Africa Centre's Demographic Surveillance Information System (GR065377/Z/01/H) and Population-based HIV Survey (GR065377/Z/01/B) was received from the Wellcome Trust, UK. The funding organizations had no role in the design and conduct of the study, in the collection, analysis, and interpretation of the data, or in the preparation, review or approval of the manuscript.

Acknowledgements

This article uses data generated by the Africa Centre Demographic Information System and population-based HIV survey. We thank the field staff at the Africa Centre for Health and Population Studies for their work in collecting the data used in this study and the communities in the Africa Centre demographic surveillance area for their support and participation in this study.

Conflict of interest: None declared.

KEY MESSAGES

- The results show considerable geographical variation in HIV prevalence and localized clustering of infections in a high prevalence, rural South African population.
- The findings challenge the prevailing paradigm of a ubiquitous 'generalized' rural epidemic and reveal the existence of several localized epidemics of varying intensity that are partly contained within geographically defined communities.
- Despite the overall high prevalence of HIV in many rural South African settings, the results support the need for interventions that target socio-geographic spaces (communities) at greatest risk to supplement prevention efforts aimed at the general population.

References

- ¹ UNAIDS. *2008 Report on the Global AIDS Epidemic*. UNAIDS, the Joint United Nations Programme on HIV/AIDS.
- ² Kleinschmidt I, Pettifor A, Morris N, MacPhail C, Rees H. Geographic distribution of human immunodeficiency virus in South Africa. *Am J Trop Med Hyg* 2007;**77**: 1163–69.
- ³ DOH. *National HIV and Syphilis Antenatal Prevalence Survey, South Africa 2006*. Pretoria. Pretoria: Department of Health South Africa, 2007.
- ⁴ Bärnighausen T, Tanser F, Gqwede Z, Mbizana C, Herbst K, Newell ML. High HIV incidence in a community with high HIV prevalence in rural South Africa: findings from a prospective population-based study. *AIDS* 2008;**22**:139–44.

- ⁵ Anderson R, May R. *Infectious Diseases of Humans: Dynamics and Control*. New York: Oxford University Press, 1991.
- ⁶ Aral SO, Padian NS, Holmes KK. Advances in multilevel approaches to understanding the epidemiology and prevention of sexually transmitted infections and HIV: an overview. *J Infect Dis* 2005;**191** (Suppl 1):S1–S6.
- ⁷ Link BG, Phelan J. Social conditions as fundamental causes of disease. *J Health Soc Behav* 1995;**35**:80–94.
- ⁸ Jennings JM, Curriero FC, Celentano D, Ellen JM. Geographic identification of high gonorrhoea transmission areas in Baltimore, Maryland. *Am J Epidemiol* 2005;**161**: 73–80.
- ⁹ Becker KM, Glass GE, Brathwaite W, Zenilman JM. Geographic epidemiology of gonorrhoea in Baltimore, Maryland, using a geographic information system. *Am J Epidemiol* 1998;**147**:709–16.
- ¹⁰ Bernstein KT, Curriero FC, Jennings JM, Olthoff G, Erbdelding EJ, Zenilman J. Defining core gonorrhoea transmission utilizing spatial data. *Am J Epidemiol* 2004;**160**: 51–58.
- ¹¹ Ellen JM, Hessol NA, Kohn RP, Bolan GA. An investigation of geographic clustering of repeat cases of gonorrhoea and chlamydial infection in San Francisco, 1989–1993: evidence for core groups. *J Infect Dis* 1997;**175**:1519–22.
- ¹² Padian NS, Aral SO, Holmes KK. Individual, group and population approaches to STD/HIV prevention. In: Wasserheit J (ed.). *Sexually Transmitted Diseases*. 3rd edn. New York: McGraw-Hill, 1999. pp. 1231–38.
- ¹³ Maticka-Tyndale E, Brouillard-Coylea C. The effectiveness of community interventions targeting HIV and AIDS prevention at young people in developing countries. *World Health Organ Tech Rep Ser* 2006;**938**:243–85; discussion 317–41.
- ¹⁴ Wegbreit J, Bertozzi S, DeMaria LM, Padian NS. Effectiveness of HIV prevention strategies in resource-poor countries: tailoring the intervention to the context. *AIDS* 2006;**20**:1217–35.
- ¹⁵ Auerbach JD, Hayes RJ, Kandathil SM. Overview of effective and promising interventions to prevent HIV infection. *World Health Organ Tech Rep Ser* 2006;**938**:43–78; discussion 317–41.
- ¹⁶ Fichtenberg CM, Ellen JM. Moving from core groups to risk spaces. *Sex Transm Dis* 2003;**30**:825–26.
- ¹⁷ Weir SS, Pailman C, Mahlalela X, Coetzee N, Meidany F, Boerma JT. From people to places: focusing AIDS prevention efforts where it matters most. *AIDS* 2003;**17**:895–903.
- ¹⁸ Msisha WM, Kapiga SH, Earls FJ, Subramanian SV. Place matters: multilevel investigation of HIV distribution in Tanzania. *AIDS* 2008;**22**:741–48.
- ¹⁹ De Cock K, De Lay P. HIV/AIDS estimates and the quest for universal access. *Lancet* 2008;**6736**:60732–1.
- ²⁰ Openshaw S. *The Modifiable Areal Unit Problem*. Norwich, England: Geo Books, 1984.
- ²¹ Tukey J. Statistical mapping: What should not be mapped. *Collected Works of John W Tukey*. Belmont, CA: Wadsworth, 1988. pp. 109–21.
- ²² Elliott P, Wakefield J, Best N, Briggs D. *Spatial Epidemiology*. New York: Oxford University Press, 2000.
- ²³ Tanser F, Hosegood V, Barnighausen T *et al*. Cohort Profile: Africa Centre Demographic Information System (ACDIS) and population-based HIV survey. *Int J Epidemiol* 2008;**37**:956–62.
- ²⁴ Welz T, Hosegood V, Jaffar S, Batzing-Feigenbaum J, Herbst K, Newell ML. Continued very high prevalence of HIV infection in rural KwaZulu-Natal, South Africa: a population-based longitudinal study. *AIDS* 2007;**21**: 1467–72.
- ²⁵ Tanser F, Hosegood V, Benzler J, Solarsh G. New approaches to spatially analyse primary health care usage patterns in rural South Africa. *Trop Med Int Health* 2001;**6**:826–38.
- ²⁶ Waller LA, Gotway CA. *Applied Spatial Statistics for Public Health Data*. Hoboken, NJ: Wiley-Interscience, 2004.
- ²⁷ Rossi RE, Mulla DJ, Journel AG, Franz EH. Geostatistical tools for modeling and interpreting ecological spatial dependence. *Ecol Monogr* 1992;**62**:277–314.
- ²⁸ Kulldorff M. SaTScan v7.0: software for the spatial and space-time statistics. National Cancer Institute, 2006.
- ²⁹ Kulldorff M, Heffernan R, Hartman J, Assuncao R, Mostashari F. A space-time permutation scan statistic for disease outbreak detection. *PLoS Med* 2005;**2**:e59.
- ³⁰ Grassly NC, Garnett GP, Schwartlander B, Gregson S, Anderson RM. The effectiveness of HIV prevention and the epidemiological context. *Bull World Health Organ* 2001; **79**:1121–32.
- ³¹ Silverman BW. *Density Estimation for Statistics and Data Analysis*. London: Chapman & Hall/CRC, 1986.
- ³² Jemal A, Devesa S, Kulldorff M, Hayes R, Fraumeni J. Geographic variation in prostate cancer mortality rates among white males in the united states. *Ann Epidemiol* 2000;**10**:470.
- ³³ Kulldorff M, Feuer EJ, Miller BA, Freedman LS. Breast cancer clusters in the northeast United States: a geographic analysis. *Am J Epidemiol* 1997;**146**:161–70.
- ³⁴ Hjalmars U, Kulldorff M, Gustafsson G, Nagarwalla N. Childhood leukaemia in Sweden: using GIS and a spatial scan statistic for cluster detection. *Stat Med* 1996;**15**: 707–15.
- ³⁵ Wylie JL, Cabral T, Jolly AM. Identification of networks of sexually transmitted infection: a molecular, geographic, and social network analysis. *J Infect Dis* 2005; **191**:899–906.
- ³⁶ Schleihauf E, Watkins RE, Plant AJ. Heterogeneity in the spatial distribution of bacterial STIs. *Sex Transm Infect* 2009;**85**:45–49.
- ³⁷ Sankoh OA, Ye Y, Sauerborn R, Muller O, Becher H. Clustering of childhood mortality in rural Burkina Faso. *Int J Epidemiol* 2001;**30**:485–92.
- ³⁸ Bithell JF. An application of density estimation to geographical epidemiology. *Stat Med* 1990;**9**:691–701.
- ³⁹ James L, Matthews I, Nix B. Spatial contouring of risk: a tool for environmental epidemiology. *Epidemiology* 2004; **15**:287–92.
- ⁴⁰ Lai PC, Wong CM, Hedley AJ *et al*. Understanding the spatial clustering of severe acute respiratory syndrome (SARS) in Hong Kong. *Environ Health Perspect* 2004;**112**: 1550–56.
- ⁴¹ Barreto FR, Teixeira MG, Costa Mda C, Carvalho MS, Barreto ML. Spread pattern of the first dengue epidemic in the city of Salvador, Brazil. *BMC Public Health* 2008;**8**:51.
- ⁴² Zeman P. Objective assessment of risk maps of tick-borne encephalitis and Lyme borreliosis based on spatial patterns of located cases. *Int J Epidemiol* 1997;**26**:1121–29.

- ⁴³ Tanser F, Lesueur D, Solarsh G, Wilkinson D. HIV heterogeneity and proximity of homestead to roads in rural South Africa: an exploration using a geographical information system. *Trop Med Int Health* 2000;**5**:40–46.
- ⁴⁴ Wawer MJ, Sewankambo NK, Berkley S *et al*. Incidence of HIV-1 infection in a rural region of Uganda. *Br Med J* 1994;**308**:171–73.
- ⁴⁵ Ferguson AG, Morris CN. Mapping transactional sex on the Northern Corridor highway in Kenya. *Health Place* 2007;**13**:504–19.
- ⁴⁶ Arroyo MA, Sateren WB, Serwadda D *et al*. Higher HIV-1 incidence and genetic complexity along main roads in Rakai District, Uganda. *J Acquir Immune Defic Syndr* 2006;**43**:440–45.
- ⁴⁷ Williams BG, Gouws E. The epidemiology of human immunodeficiency virus in South Africa. *Philos Trans R Soc Lond B Biol Sci* 2001;**356**:1077–86.
- ⁴⁸ Gould P, Wallace R. *Spatial Structures and Scientific Paradoxes in the AIDS Pandemic*. JSTOR, 1994. pp. 105–16.
- ⁴⁹ Lurie MN, Williams BG, Zuma K *et al*. The impact of migration on HIV-1 transmission in South Africa: a study of migrant and nonmigrant men and their partners. *Sex Transm Dis* 2003;**30**:149–56.
- ⁵⁰ Pickering H, Okongo M, Ojwiya A, Yirrell D, Whitworth J. Sexual networks in Uganda: mixing patterns between a trading town, its rural hinterland and a nearby fishing village. *Int J STD AIDS* 1997;**8**:495–500.
- ⁵¹ Aral SO. Behavioral aspects of sexually transmitted diseases: core groups and bridge populations. *Sex Transm Dis* 2000;**27**:327–28.
- ⁵² Moodley P, Connolly C, Sturm AW. Interrelationships among human immunodeficiency virus type 1 infection, bacterial vaginosis, trichomoniasis, and the presence of yeasts. *J Infect Dis* 2002;**185**:69–73.
- ⁵³ Manhart LE, Holmes KK. Randomized controlled trials of individual-level, population-level, and multilevel interventions for preventing sexually transmitted infections: what has worked? *J Infect Dis* 2005;**191** (Suppl 1):S7–S24.
- ⁵⁴ Sumartojo E, Carey JW, Doll LS, Gayle H. Targeted and general population interventions for HIV prevention: towards a comprehensive approach. *AIDS* 1997;**11**:1201–9.

Published by Oxford University Press on behalf of the International Epidemiological Association
© The Author 2009; all rights reserved. Advance Access publication 30 April 2009

International Journal of Epidemiology 2009;**38**:1016–1017
doi:10.1093/ije/dyp201

Commentary: Spatial clustering of HIV infection: providing clues for effective HIV prevention

Salim S Abdool Karim^{1,2}

Accepted 30 March 2009

‘Know your epidemic’ is the Joint United Nations Programme on HIV/AIDS (UNAIDS) mantra for HIV prevention.¹ Knowledge of the nature of the local HIV epidemic is critical for HIV prevention programmes to be effective. Overall HIV prevalence rates hide the true complex mosaic of the HIV epidemic. HIV infection, like most other infectious diseases, is known to cluster in relation to risk factors, especially substantial risk factors like migration and mobility. Tanser and colleagues² identified clusters of high HIV prevalence close to the national roads in rural KwaZulu-Natal,

South Africa, reiterating the spatial observation made >15 years earlier in this area along the same national road.³

This observation is significant because community-based HIV outreach programmes in hard-to-reach rural areas, which are distant from the main roads, are expensive and labour intensive. These findings provide grounds for targeting HIV prevention efforts in the identified high HIV prevalence geographical regions close to the national road. Knowing where to target interventions is a useful first step. The next step is to establish why HIV is clustering along this national road. There are at least four potential explanations that deserve further investigation.

The first consideration is age. Remote rural areas tend to have higher proportions of toddlers and the elderly. Hence, it is important to establish whether those living close to the national road were generally

¹ Centre for the AIDS Programme of Research in South Africa – CAPRISA, University of KwaZulu-Natal, Congella, South Africa.

² Department of Epidemiology, Columbia University, New York, USA. E-mail: karims1@ukzn.ac.za