Geospatial Analysis contributions to the challenge of HIV: Results from a South African National HIV Prevalence, Incidence, Behaviour Survey

Njeri Wabiri, Human Sciences Research Council, South Africa
Outline

- Context and Rationale
- Data and Analytical Methods
- Results
- Conclusions
More than 6.4 million people live with HIV in South Africa.

Distribution of HIV is heterogeneous in South Africa.

Example, based on Shisana et al., 2008; 2012, population based HIV surveys.
HIV prevalence Varies by district, SA, 2008

similar spatial pattern in 2012
While there are ....

- studies* that have looked at clustering and the spatial heterogeneity of HIV e.g., among for youth, and/or selected communities

- Limited studies focusing on how relationships of HIV risk factors associates with HIV prevalence from location to location

More essential to...

- Predicted levels of change in HIV prevalence with changes in the risk factors from location to location.

- **Example:** Establish how change in percent female and/or single population significantly associates with HIV prevalence over space?

Rationale: Inform the focus of interventions to areas of particular need taking into account the risks.

Outline

• Background and Rationale
• Data and Analytic methods
• Results
• Conclusions
Data sources

- National population-based household survey data
- Multistage cluster surveillance survey design
  - **Primary sampling unit**: 1000 EAs from 86000 EAs
  - **Secondary sampling unit**: Cluster of 15 households per EA
  - **Final sampling unit**: 4 eligible individuals per household four mutually exclusive age groups (< 2 years, 2-14 years, 15-24 years and ≥ 25 years), latest survey-all people in the household.
  - Geographic identifiers-EA’s linked to external datasets
- Linked anonymous HIV testing: dried blood spot specimens with behavioural data for participants
- District health barometer- socio-economic measures

Analytic methods

• Descriptive analysis
• Non-spatial and Spatial regression models
• Spatial mapping unit (n=52):
  • The “District Municipality” in South Africa
  • District Health System - Primary Health Care delivery system*

Wabiri et al., Spatial and Spatio-temporal Epidemiology 2016; Shisana et al., 2014; *Monticelli et al. 2010
Descriptive analysis

- Pearson correlations between risk factors and HIV prevalence
- Risk factors not associated with HIV prevalence not included in the models
- Only one of highly correlated risks is included in the regression models to avoid model redundancy and multicollinearity.
- Additional explanatory approach used to select risk factors
  - Thematic maps of district HIV prevalence and the covariates
Modelling

Non-spatial regression

• Fit Global model

\[ y_i = a_0 + \sum_{k=1}^{n} a_k x_{ik} + \varepsilon_i \]

• \(a_k\), the value of \(k^{th}\) parameter of independent variable \(k\)
• \(y_i\) is \(i^{th}\) observation of the dependent variable
• \(x_{ik}\) is the \(i^{th}\) observation of the \(k^{th}\) independent variable
• Explores significant relationship to explore with GWR

Spatial regression (Geographically weighted Regression)

• Localized multivariate regression
• Estimate parameters at each location

\[ y_i = a_{i0} + \sum_{k=1}^{n} a_{ik} x_{ik} + \varepsilon_i \]

• \(a_{ik}\) is the value of \(k^{th}\) parameter at location \(i\)
• \(y_i\) is the HIV prevalence at location \(i\)
• \(x_{ik}\) is the \(i^{th}\) observation of the \(k^{th}\) independent variable

Test spatial independence of residuals

• Moran’s I*

\[ I = \frac{n \sum_{i=1}^{n} \sum_{j=1}^{n} w_{ij} (y_i - \bar{y}) (y_j - \bar{y})}{(\sum_{i \neq j} \sum w_{ij}) (\sum_{i=1}^{n} (y_i - \bar{y})^2)} \]

• \( W_{ij} \) is a measure of spatial proximity pairs of i and j.
• Significant (\( P<0.05 \)) Moran’s I -clustering of residuals not chance.
• Also confirms the models included all the key covariates.
Analytical models used

Socio-behavioral risks GWR model

\[ \text{HIV}_i(z) = \]
\[ \beta_0i(z) + \beta_1i \text{PropSingle}_i(z) + \beta_2i \text{PropOldSexPartner}_i(z) + \beta_3i \text{PropNonregCodm}_i(z) + \epsilon_i \]

Background/ Demographic risks GWR model

\[ \text{HIV}_i(z) = \]
\[ \beta_0i(z) + \beta_1i \text{SEQ}_i(z) + \beta_2i \text{PropFemale}_i(z) + \beta_3i \text{PropAfrican}_i(z) + \beta_4i \text{Prop25-49}_i(z) + \epsilon_i \]

- \((z)\) indicate the parameters, \(\beta\), estimated at each District
- vector \(z\) is the centroid coordinates for each district, \((i=1,2,\ldots,52)\) districts
- \(\epsilon_i\) residual at district \(i\).
- Fixed kernel to calibrate the model- spatial structure.
Model fitness

- A difference of greater than 3 in the AIC between the non-spatial (OLS) and spatial (GWR) model signifies better model fit.

- The variance inflation factor VIF >10 indicate multicollinearity.

- \(-1.96 \leq Pseudo-t \leq 1.96\) to test significance of the local parameters.

- Overlay smooth map of Pseudo-t values on map of local parameter to visualize significant parameters.
Descriptive Results
## Correlations: HIV and associated risks

<table>
<thead>
<tr>
<th></th>
<th>HIV prevalence</th>
<th>Rural informal</th>
<th>Urban formal</th>
<th>Female</th>
<th>African</th>
<th>25–49 yrs.</th>
<th>Single</th>
<th>Older partner &gt; 5 yrs</th>
<th>Young partner &lt; 5 yrs</th>
<th>Condom</th>
</tr>
</thead>
<tbody>
<tr>
<td>Rural Informal</td>
<td>0.34*</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Urban Formal</td>
<td>-0.36*</td>
<td>-0.92*</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Female</td>
<td>0.55*</td>
<td>0.56*</td>
<td>-0.49*</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>African</td>
<td>0.49*</td>
<td>0.64*</td>
<td>-0.55*</td>
<td>0.43*</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>25–49 years</td>
<td>0.30*</td>
<td>-0.22</td>
<td>0.09</td>
<td>0.07</td>
<td>-0.22</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Single</td>
<td>0.25</td>
<td>0.19</td>
<td>-0.16</td>
<td>0.15</td>
<td>0.41*</td>
<td>-0.15</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Older partner &gt; 5 yrs.</td>
<td>0.28*</td>
<td>0.50*</td>
<td>-0.51*</td>
<td>0.62*</td>
<td>0.15</td>
<td>0.15</td>
<td>-0.13</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Young partner &lt; 5 yrs.</td>
<td>-0.33*</td>
<td>-0.20</td>
<td>0.18</td>
<td>-0.56*</td>
<td>-0.07</td>
<td>-0.20</td>
<td>-0.21</td>
<td>-0.31*</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Condom</td>
<td>0.29*</td>
<td>0.36*</td>
<td>-0.27</td>
<td>0.15</td>
<td>0.67*</td>
<td>-0.34*</td>
<td>0.41*</td>
<td>-0.05</td>
<td>0.21</td>
<td></td>
</tr>
<tr>
<td>SEQ</td>
<td>-0.44*</td>
<td>-0.86*</td>
<td>0.86*</td>
<td>-0.56*</td>
<td>-0.73*</td>
<td>0.14</td>
<td>-0.32*</td>
<td>-0.45*</td>
<td>0.21</td>
<td>0.44*</td>
</tr>
</tbody>
</table>

* significant at 5%; n = 52.
HIV prevalence associates with regions with high females proportions.
High HIV prevalence in Black African dominated districts
HIV prevalence geographically associates with Ruralness.
High Proportions of 25+ in high prevalence districts
Deprivation associated with HIV prevalence

Socio-economic Quintile (%):
- Q1 - Most deprived
- QII
- QIII
- QIV
- QV - Least deprived

HIV Prevalence (%):
- 1 - 8
- 9 - 12
- 13 - 15
- 16 - 19
- 20 - 24
- 25 - 56
Social-Behaviour risk factors
High HIV prevalence spatially associates with intergeneration sex.
Spatial association of HIV prevalence with singlehood
High HIV prevalence; mixed levels of condom use
Analytical Results
## Non-spatial model results

<table>
<thead>
<tr>
<th>Risk factors</th>
<th>Parameter</th>
<th>Std. Error</th>
<th>R²</th>
<th>Adjusted R²</th>
<th>AICc</th>
<th>Moran’s I</th>
</tr>
</thead>
<tbody>
<tr>
<td>Rural Informal</td>
<td>0.10*</td>
<td>0.04</td>
<td>11.44</td>
<td>9.67</td>
<td>383.8</td>
<td>0.23*</td>
</tr>
<tr>
<td>Urban Formal</td>
<td>-0.12*</td>
<td>0.04</td>
<td>12.90</td>
<td>11.50</td>
<td>382.9</td>
<td>0.22*</td>
</tr>
<tr>
<td>Female</td>
<td>0.56***</td>
<td>0.12</td>
<td>30.27</td>
<td>28.85</td>
<td>371.0</td>
<td>0.07</td>
</tr>
<tr>
<td>Black African</td>
<td>0.17***</td>
<td>0.04</td>
<td>23.90</td>
<td>22.46</td>
<td>375.9</td>
<td>0.20*</td>
</tr>
<tr>
<td>25-49 Years</td>
<td>0.27*</td>
<td>0.12</td>
<td>8.91</td>
<td>7.09</td>
<td>385.3</td>
<td>0.34*</td>
</tr>
<tr>
<td>Single</td>
<td>0.22</td>
<td>0.12</td>
<td>6.28</td>
<td>4.41</td>
<td>386.8</td>
<td>0.26*</td>
</tr>
<tr>
<td>Older Sexual Partner</td>
<td>0.32*</td>
<td>0.16</td>
<td>7.73</td>
<td>5.89</td>
<td>385.9</td>
<td>0.26*</td>
</tr>
<tr>
<td>Non-regular Partner Condom Use</td>
<td>0.01</td>
<td>0.03</td>
<td>0.1</td>
<td>-1.80</td>
<td>384.0</td>
<td></td>
</tr>
<tr>
<td>Younger Sexual Partners</td>
<td>-0.40*</td>
<td>0.16</td>
<td>11.08</td>
<td>9.31</td>
<td>383.0</td>
<td>0.24*</td>
</tr>
<tr>
<td>Social Economic Quintile(SEQ)</td>
<td>-2.98**</td>
<td>0.87</td>
<td>19.01</td>
<td>17.40</td>
<td>379.2</td>
<td>0.16*</td>
</tr>
</tbody>
</table>

*significant at 5% level; **significant at 1%; ***significant at 0.1%
## Demographic: Spatial vs. non-spatial

<table>
<thead>
<tr>
<th>Demographics Risks</th>
<th>Estimate</th>
<th>Standard Error</th>
<th>VIF</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Intercept</strong></td>
<td>-32.15***</td>
<td>7.88</td>
<td></td>
</tr>
<tr>
<td>Female</td>
<td>0.40**</td>
<td>0.13</td>
<td>1.63</td>
</tr>
<tr>
<td>Black African</td>
<td>0.15**</td>
<td>0.05</td>
<td>2.32</td>
</tr>
<tr>
<td>25-49 years</td>
<td>0.33**</td>
<td>0.09</td>
<td>1.09</td>
</tr>
<tr>
<td><strong>Social-Economic Quintile</strong></td>
<td>0.21</td>
<td>1.11</td>
<td>2.53</td>
</tr>
</tbody>
</table>

### Adjusted R²
- Non-spatial regression (OLS): 46.5
- Spatial Regression (GWR): 52.3

### Condition Number
- Non-spatial regression (OLS): 11.0
- Spatial Regression (GWR): 6.75

### AICc
- Non-spatial regression (OLS): 360.7
- Spatial Regression (GWR): 357.4

### Moran's I
- Non-spatial regression (OLS): 0.007
- Spatial Regression (GWR): -0.05

*significant at 5% level; **significant at 1%; ***significant at 0.1%

A condition number less than 30 suggests lack of multicollinearity.
## Socio-behavior risks: Spatial vs. non-spatial

<table>
<thead>
<tr>
<th>Socio-behavior risks</th>
<th>Estimate</th>
<th>Standard Error</th>
<th>VIF</th>
<th>Spatial Regression (GWR)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Intercept</td>
<td>-3.64</td>
<td>7.07</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Single</td>
<td>0.26*</td>
<td>0.11</td>
<td>2</td>
<td></td>
</tr>
<tr>
<td>Older Sex partner</td>
<td>0.36*</td>
<td>0.15</td>
<td>2</td>
<td></td>
</tr>
<tr>
<td>Non-Regular partner and Condom</td>
<td>0.01</td>
<td>0.03</td>
<td>0</td>
<td></td>
</tr>
<tr>
<td>Adjusted $R^2$</td>
<td>10.88</td>
<td></td>
<td></td>
<td>28.10</td>
</tr>
<tr>
<td>Condition Number</td>
<td>1.24</td>
<td></td>
<td></td>
<td>8.28</td>
</tr>
<tr>
<td>AICc</td>
<td>385.80</td>
<td></td>
<td></td>
<td>378.6</td>
</tr>
<tr>
<td>Moran's I</td>
<td>0.22*</td>
<td></td>
<td></td>
<td>0.05</td>
</tr>
</tbody>
</table>

*significant at 5% level; **significant at 1%; ***significant at 0.1%

A condition number less than 30 suggests lack of multicollinearity.
Message: A percent increase in old sexual partnership is associated with (0·14-0·30) average increase in HIV Prevalence
Message: A percent increase in single population is associated with (0.11-0.17) average increase in HIV Prevalence

Marriage rates high in Western cape province, know to have low HIV prevalence
A percent increase in female population is associated with 0.21 average increase in HIV Prevalence.

A percent increase in those aged 25-49 years population is associated with 0.15 average increase in HIV Prevalence.
Marginal average increase of (0.05-0.08) in HIV prevalence associated Black African population
Message: High intercept estimates: residual variation explained by location
Alignment of public health care utilisation and local patterns of HIV
**Note: High HIV Prevalence spatially associates with PHC utilization rates**

The **PHC utilization rate** is the PHC total annual headcount divided by the total catchment population. It is a measure of the average number of primary health care visits per person per year to a public PHC facility.

*District health barometer/; District Health information systems*
Conclusions

• Relationship between HIV risks and HIV prevalence is non-stationary with covariates causing different levels of prevalence in different districts.

• Targeting the who and where with a good understanding of variation in HIV risks will make every rand count.
Conclusions

- High HIV prevalence spatially associates with:
  - Black African origin,
  - unfavourable sex ratio (high proportion of females),
  - being single or low marriage rates
  - Intergenerational sex
  - Deprivation

- Intergenerational sex compounds the risk of acquiring HIV infection for females in deprived districts
Conclusions

• Need for additional research to ascertain other HIV infection risks
  • How ARV rates at the district level relate to changes in prevalence rates?
  • What are the major differences in behaviour patterns in rural and informal settlements in areas where HIV prevalence remains low in spite of high numbers of settlements?
• Observed geographical patterns of HIV prevalence parallels that observed using the National Antenatal Sentinel HIV & Herpes Simplex Type-2 Prevalence Survey in South Africa (Department of Health 2011, 2013)
• Distribution of the covariates: race, sex, SEQ, and inter-generational sex have predominantly remained the same over the years
• The district Health System is basic channel through which the delivery of Primary Health Care is undertaken in South Africa
Acknowledgements

Olive Shisana
Evidence Based Solutions, Cape Town, South Africa &
Department of Psychiatry and Mental Health, University
of Cape Town, South Africa.

Khangelani Zuma (HSRC)
Jeffrey Freeman (Emory University)
HSRC survey team
Survey funding: CDC