Incorporating spatial variability to generate sub-national estimates of HIV prevalence in SSA

Diego Cuadros PhD
Laith Abu-Raddad PhD
Infectious Disease Epidemiology Group

World STI & HIV Congress
September 2015

Introduction: HIV

- Sub-Saharan Africa (SSA) has by far the largest HIV epidemic in the world, with an estimated 25 million infected individuals over the past two decades
- The complex epidemiological context in SSA has prevented the elucidation of the drivers of such epidemic
- The necessity of a thorough change of perspective to better understand the epidemic is imperative

Introduction: Medical geography

- Significant development advanced spatial statistics and the increasing availability of computerized geographic information system technology have occurred over the last few decades
- Despite these advances, a recent review of 355 clinically-significant infectious diseases indicated that only 2% had been comprehensively mapped (Hay et al. Philosophical Transaction of the Royal Society B, 2013)

Introduction: Medical geography

- Area of each section is determined by the total DALY contribution. Blue indicates a cluster contributing to the top ten clusters to be prioritised
- Area of each section is determined by the total policy interest score. Red indicates a cluster within the top ten to be prioritised
- Malaria, HIV and tuberculosis have the highest priority for mapping due to their considerable burden (Pigott et al. Plos Neglected Tropical Diseases, 2015)
Weill Cornell Medical College in Qatar
Infectious Disease Epidemiology Group

Introduction: HIV

Malaria
Malaria Atlas Project
World Health Organization (2007)

Introduction: Medical geography

Vector-borne diseases
Pathogen
Vector
Host

Malaria Cofactors
Environmental variables:
- Normalized Difference Vegetation Index (NDVI)
- Temperature
- Precipitation
- Evapotranspiration
- Elevation

High resolution maps of the geographic distribution of a disease. The use of the survey data from a sample of locations to predict continuous surfaces of risk, informed by environmental and demographic covariates

Introduction: Medical geography

Vector-borne diseases
Pathogen
Vector
Host

Malaria atlas project

High resolution maps
Environmentally referenced data points

Introduction: Medical geography

Vector-borne diseases
Pathogen
Vector
Host

Sexually transmitted diseases (STD)
Infected
Susceptible
**Introduction:** Sexually transmitted diseases

The study of sexually transmitted infections such as HIV has focused on social space.

**Sexual networks:** groups of persons connected to one and other sexually

**Characteristics:**
- Number of partners (links)
- Serial monogamy
- Concurrent relationships

**Core groups:** members that have high levels of risk behavior and can fuel sustained transmission.

---

**Rationale:** Mapping HIV

**Geographical factors** could be playing an important role in the distribution of the HIV infection in SSA.

---

**HIV Cofactors**

Sexually transmitted diseases (STDs)
- Wealth index
- Male circumcision
- Lifetime sexual partners
- Education
- Ever been tested for HIV
- Condom use

---

**Introduction:** Mapping HIV

Geographical factors could be playing an important role in the distribution of the HIV infection in SSA.

---

**Rationale:** Mapping HIV

**Geographical factors** could be playing an important role in the distribution of the HIV infection in SSA.
**Introduction: Mapping HIV**

Georeferenced data points and Environmental cofactors lead to High resolution maps.

High resolution maps of the geographic distribution of a disease. The use of the survey data from a sample of locations to predict continuous surfaces of risk, informed by environmental and demographic covariates.

**Research question:**

Could “environmental” and socio-behavioral factors be used to generate HIV prevalence prediction maps in sub-Saharan Africa?

**Methods: Environmental factors**

**Normalized Difference Vegetation Index (NDVI)**

Measure of the density of green vegetation over the Earth created by measuring the wavelengths and intensity of visible and near-infrared light reflected by the land surface back up into space. It quantifies the concentrations of green leaf vegetation. (NASA’s Earth Observatory Group)

**Population**

**Distance to main roads**
## Methods: Environmental factors

- Distance to main roads

## Methods: Demographic and Health Survey (DHS)

- **Standard DHS surveys**: national representative household surveys that provide data in the areas of socio-economic and demographic indicators
- **Biomarker collection**: dichotomous HIV serostatus for several individuals included in the survey
- **Geographical database**: Collection of geographical coordinates corresponding to the locations of the communities that participate in the survey

## Methods: Socio-behavioral factors

### DHS data

- Wealth index
- Male circumcision
- Lifetime sexual partners
- Ever been tested for HIV
- Education
- Condom use

### Wealth index

Wealth index is an ordinal variable that characterizes standard of living as determined by material possessions. The resulting asset scores were then used to define wealth quintiles: poorest, poorer, middle, richer and richest

<table>
<thead>
<tr>
<th>Distance, h</th>
<th>y(h)</th>
</tr>
</thead>
<tbody>
<tr>
<td>0</td>
<td>0.21</td>
</tr>
<tr>
<td>0.39</td>
<td>0.42</td>
</tr>
<tr>
<td>0.78</td>
<td>0.64</td>
</tr>
<tr>
<td>1.17</td>
<td>0.85</td>
</tr>
<tr>
<td>1.56</td>
<td>1.06</td>
</tr>
</tbody>
</table>

Dummy variable:
- Poorest, poorer = 1
- Middle, richer and richest = 0

Percentage of poorest and poorer people was estimated for each data point

## Methods: Socio-behavioral factors

- Semivariograms were used to observe the spatial pattern of the data.
- The semivariogram uses the semi-variance \( y(h) \) as a measure of half the average squared difference between pairs of data values separated by the distance \( h \) in the form:
  \[
  y(h) = \frac{1}{2N(h)} \sum_{i=1}^{N(h)} (y_i - y_j)^2
  \]
  Where \( N(h) \) is the number of distinct pairs of observed data that are separated by \( h \) and \( |N(h)| \) is the number of pairs in that set

- The semivariogram model generated was used for interpolation by the technique called **Kriging**
  - The interpolation is based on the assumption that covariance between points is entirely a function of the distance between them as modeled in the semivariogram
**Methods: Socio-behavioral factors**

**Wealth index**

- The semivariogram model generated was used for interpolation by the technique called Kriging.
- The interpolation is based on the assumption that covariance between points is entirely a function of the distance between them as modeled in the semivariogram.

**Male circumcision**

Percentage of circumcised males were estimated in each data point.

**Lifetime number of sexual partners**

Dummy variable:
- $1, 2, 3 = 0$
- $> 3 = 1$

Percentage of individuals with more than three lifetime sexual partners were estimated for each data point.

**Education**

Education level was evaluated as a categorical variable with four levels: no education, primary education, secondary education and higher education.

Dummy variable:
- No education, primary education = 0
- secondary education and higher education = 1

Percentage of individuals with secondary education and higher education were estimated for each data point.
Methods: Socio-behavioral factors

Education

Condom use

Condom use last time have sex
Percentage of individuals who used condom last time have sex were estimated for each data point

Condom use

Condom use (%)

Distance, h

0.63 1.26 1.89 2.53 3.16 3.79 4.42 5.05

0.58 1.16 1.75 2.33 2.91

Methods: Socio-behavioral factors

Percentage of individuals who have ever been tested for HIV were estimated for each data point

Ever been tested for HIV

Distance, h

0.21 0.41 0.62 0.83 1.04

Methods: Maps of cofactors

Methods: Maps of cofactors

High Resolution HIV prevalence map
Methods: Non-spatial logistic regression model

Logit (p) = α + β1X1 + β2X2 + β3X3 + β4X4 + β5X5 + ...

Where:
- p = HIV infection probability
- α = constant
- X1 = NDVI
- X2 = Distance to main roads
- X3 = Poverty
- X4 = Male circumcision
- X5 = Education

Results: High resolution map of HIV in Tanzania

Results: Residuals

Results: Mapping HIV in Tanzania

Results: Residuals

HIV prevalence prediction =

\[ \alpha + \beta_1 X_1 + \beta_2 X_2 + \beta_3 X_3 + \beta_4 X_4 + \beta_5 X_5 + \ldots \]
Results: Mapping HIV in Tanzania

Limitations

- Bias inherent to self-reported behavior
- Assumption that individuals that live close tend to behave similar
- Some spatial structure still missing in the model
- Several biological and behavioral factors not included in the model
  - Other sexually transmitted infections (e.g. HSV-2)
  - Concurrency
  - Commercial sex

Conclusions: Mapping HIV in Tanzania

- Our results highlight the stark spatial disparities in the epidemic within a country, and localize areas where both the burden and drivers of the HIV epidemic are concentrated
- HIV cofactors could be used to generate high resolution maps of HIV prevalence. These maps delineate the high disease-burden areas where spatially-targeted prevention strategies should be implemented

Funding Acknowledgement

This work was made possible by JSREP grant number [JSREP 3-014-3-007 ] from the Qatar National Research Fund (a member of Qatar Foundation). The statements made herein are solely the responsibility of the authors.