كلية طب وايل كورنيل في قطر
Weill Cornell Medical College in Qatar

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

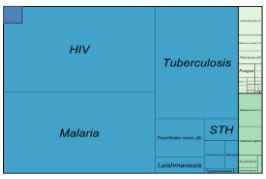
Introduction: HIV

- The '***Know your epidemic, know your response***', a Joint United Nation Programme on HIV/AIDS (UNAIDS), has become one of the first calls to modify the current strategy by recognition of the fact that **there is not a single global HIV epidemic**
- This strategy also highlights the significant role that **geographical space** plays in the identification of populations at **higher risk**

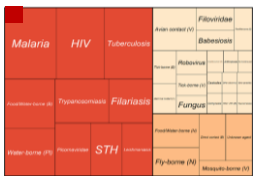
Introduction: Medical geography

- 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 all. Philosophical Transaction of the Royal Society B, 2013*)

Introduction: Medical geography



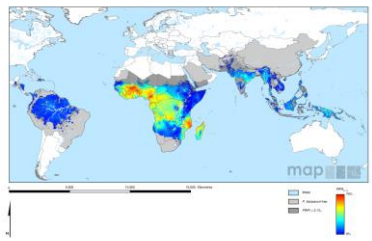
- **Area of each section is determined by the total DALY contribution. Blue indicates a cluster contributing to the top ten clusters to be prioritized**



- **Area of each section is determined by the total policy interest score. Red indicates a cluster within the top ten to be prioritized**

- Malaria, HIV and tuberculosis have the highest priority for mapping due to their considerable burden (*Pigott et al. Plos Neglected Tropical Diseases, 2015*)

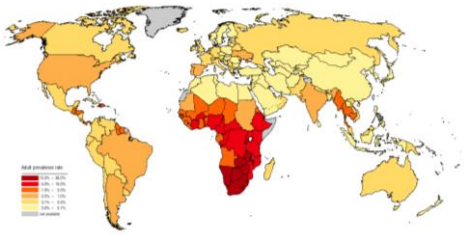
Introduction: Medical geography



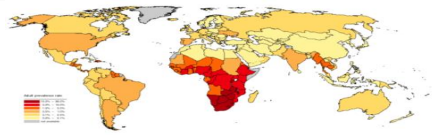
Malaria Atlas Project

The spatial distribution of Plasmodium falciparum malaria endemicity map in 2010 globally

Introduction: HIV

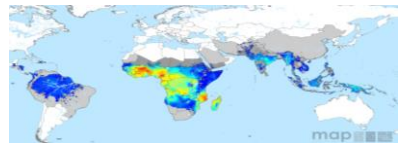


Introduction: HIV



HIV

World Health Organization (2007)

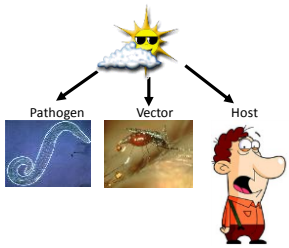


Malaria

Malaria Atlas Project

Introduction: Medical geography

Vector-borne diseases

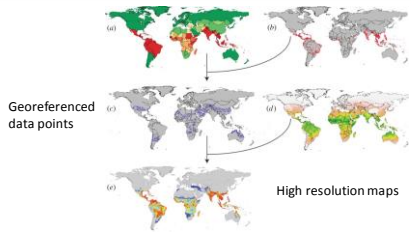


Malaria Cofactors

Environmental variables:

- Normalized Difference Vegetation Index (NDVI)
- Temperature
- Precipitation
- Evapotranspiration
- Elevation

Introduction: Medical geography



Georeferenced data points

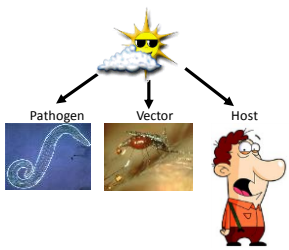
Environmental cofactors

High resolution maps

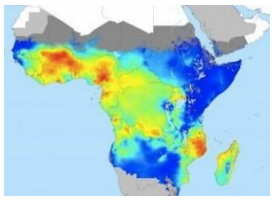
High resolution maps of the geographic distribution of a disease. The use of the survey data from a sample of locations to predict continuous surfaces of risk, informed by environmental and demographic covariates

Introduction: Medical geography

Vector-borne diseases

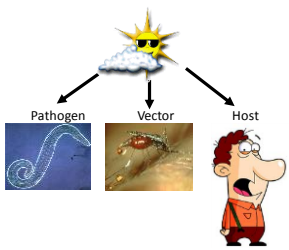


Malaria atlas project



Introduction: Medical geography

Vector-borne diseases



Sexually transmitted diseases (STD)

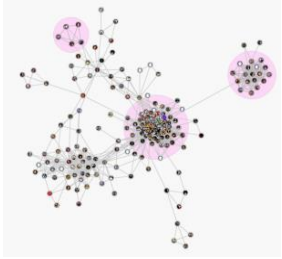
Introduction: Sexually transmitted diseases

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

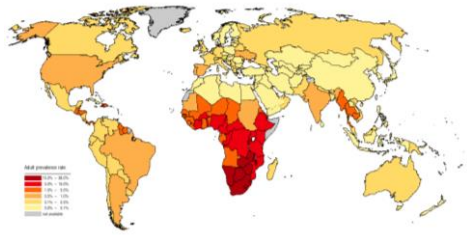
Sexual networks: groups of persons Connected to one and other sexually

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

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



Introduction: Mapping HIV

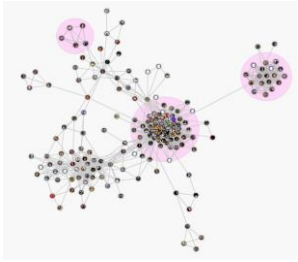


Rationale: Mapping HIV

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

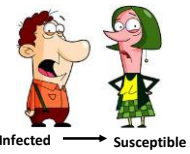


Rationale: Mapping HIV

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

Introduction: Mapping HIV

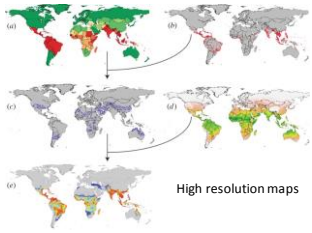
Sexually transmitted diseases (STD)



HIV Cofactors

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

Introduction: Mapping HIV



Georeferenced data points

Environmental cofactors

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

Introduction

Research question:

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

Introduction

Three countries
Tanzania, Kenya, and Malawi

Environmental (geographical) factors:
- Normalized Difference Vegetation Index (NDVI)
- Population
- Distance to main roads

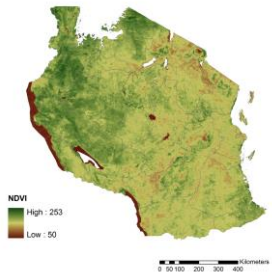
Demographic and Health Survey (DHS)
Socio-behavioral factors:
- Wealth index
- Male circumcision
- Lifetime sexual partners
- Education
- Ever been tested for HIV
- Condom use

Methods: Environmental factors

Normalized Difference Vegetation Index (NDVI)

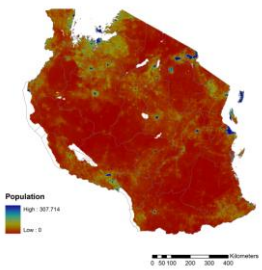
Measure of the density of green vegetation over the Earth created by measuring the wavelengths and intensity of visible and near-infrared light reflected by the land surface back up into space. **It quantifies the concentrations of green leaf vegetation**

(*NASA's Earth Observatory Group*)



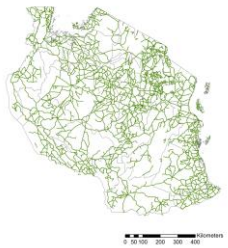
Methods: Environmental factors

Population



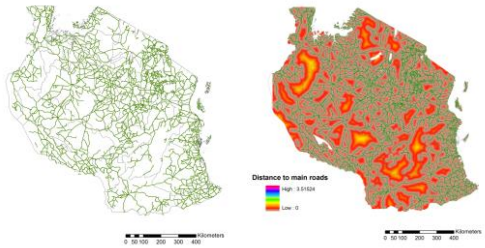
Methods: Environmental factors

Distance to main roads



Methods: Environmental factors

Distance to main roads



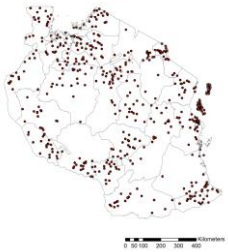
Methods: Demographic and Health Survey (DHS)

- **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



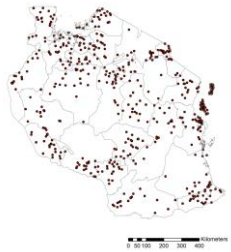
Methods: Socio-behavioral factors

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

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

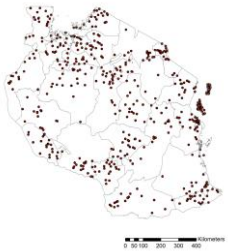
Wealth index

Semivariograms were used to observe the spatial pattern of the data.

The semivariogram uses the semi-variance $\gamma(h)$ as a measure of half the average squared difference between pairs of data values separated by the distance h in the form;

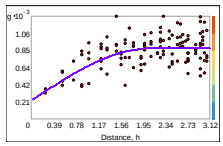
$$\gamma(h) = \frac{1}{2|N(h)|}\sum_{i}^{N(h)}(yi - yj)^2$$

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

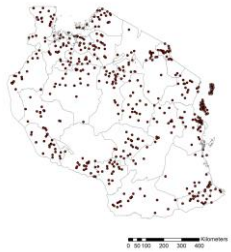


Methods: Socio-behavioral factors

Wealth index

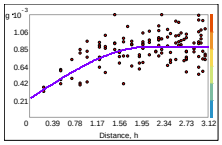


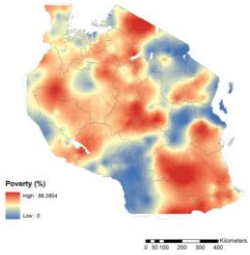
- The semivariogram model generated was used for interpolation by the technique called **Kriging**
- The interpolation is based on the assumption that covariance between points is entirely a function of the distance between them as modeled in the semivariogram



Methods: Socio-behavioral factors

Wealth index



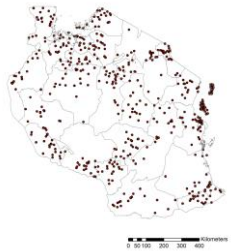


- The semivariogram model generated was used for interpolation by the technique called **Kriging**
- The interpolation is based on the assumption that covariance between points is entirely a function of the distance between them as modeled in the semivariogram

Methods: Socio-behavioral factors

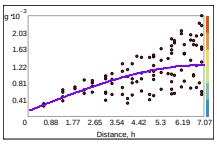
Male circumcision

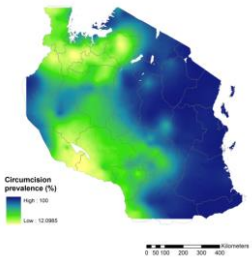
Percentage of circumcised males were estimated in each data point



Methods: Socio-behavioral factors

Male circumcision





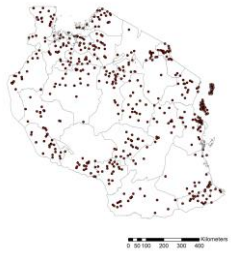
Methods: Socio-behavioral factors

Lifetime number of sexual partners

Dummy variable:

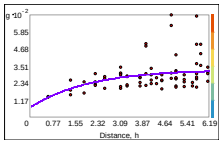
1, 2, 3 = 0
> 3 = 1

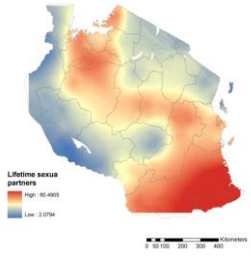
Percentage of individuals with more than three lifetime sexual partners were estimated for each data point



Methods: Socio-behavioral factors

Lifetime number of sexual partners





Methods: Socio-behavioral factors

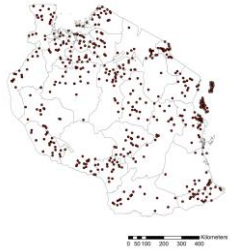
Education

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

Dummy variable:

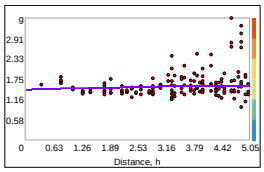
No education, primary education = 0
secondary education and higher education = 1

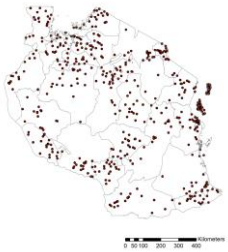
Percentage of individuals with secondary education and higher education were estimated for each data point



Methods: Socio-behavioral factors

Education



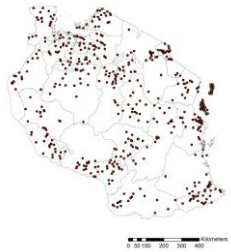


Methods: Socio-behavioral factors

Condom use

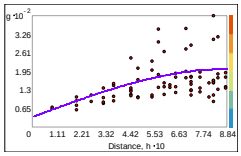
Condom use last time have sex

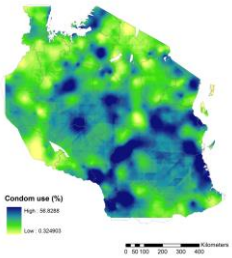
Percentage of individuals who used condom last time have sex were estimated for each data point



Methods: Socio-behavioral factors

Condom use

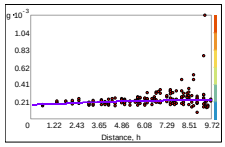


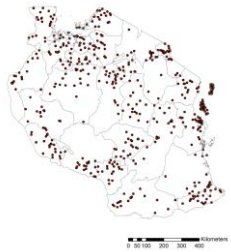


Methods: Socio-behavioral factors

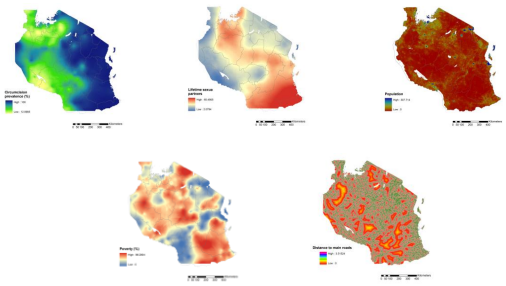
Ever been tested for HIV

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

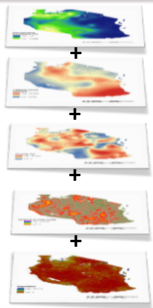




Methods: Maps of cofactors



Methods: Maps of cofactors

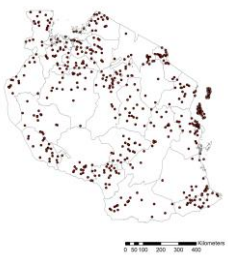


= High Resolution HIV prevalence map

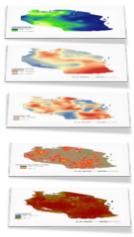
Methods: Non-spatial logistic regression model

Logit (p) = $\alpha + \beta_1X_1 + \beta_2X_2 + \beta_3X_3 + \beta_4X_4 + \beta_5X_5$ …

Where;
p = HIV infection probability
α = constant
X_1 = NDVI
X_2 = Distance to main roads
X_3 = Poverty
X_4 = Male circumcision
X_5 = Education
.
.
.



Methods: Non-spatial logistic regression model



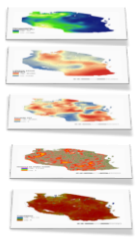
X1 = condom use
X2 = Male circumcision
X3 = Population
X4 = Distance to roads
X5 = Poverty
.
.
.

Analysis of Maximum Likelihood Estimates		
Parameter	β	Pr > ChiSq
Intercept	-1.9657	<.0001
Condom use	0.0296	<.0001
Male Circumcision	-0.0151	<.0001
Population	-0.00306	0.0315
Distance to main roads	-3.8668	0.0003
NDVI	-0.00352	0.0527
Poverty	-0.0117	<.0001
Lifetime sexual partners	0.0148	<.0001

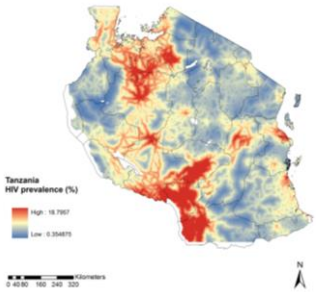
HIV prevalence prediction =

$$\frac{e^{\alpha + \beta 1X1 + \beta 2X2 + \beta 3X3 + \beta 4X4 + \beta 5X5 + \beta 6X6}}{1 + e^{\alpha + \beta 1X1 + \beta 2X2 + \beta 3X3 + \beta 4X4 + \beta 5X5 + \beta 6X6}}$$

Results: High resolution map of HIV in Tanzania



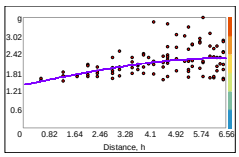
=

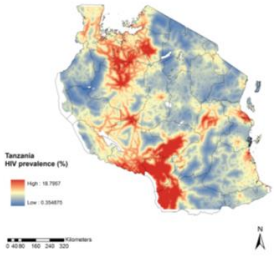


.
.
.

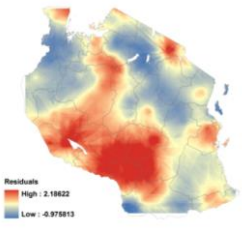
Results: Residuals

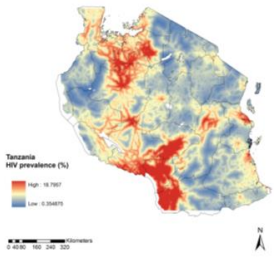
Residuals



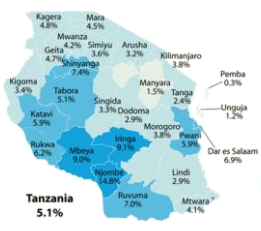


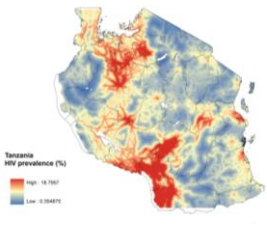
Results: Residuals



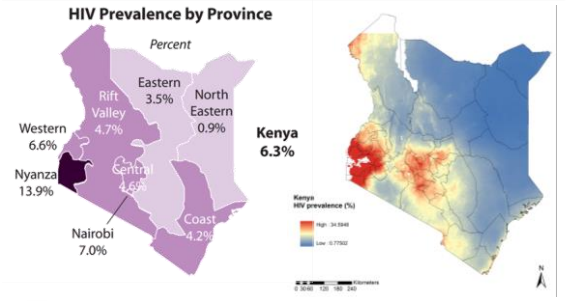


Results: Mapping HIV in Tanzania

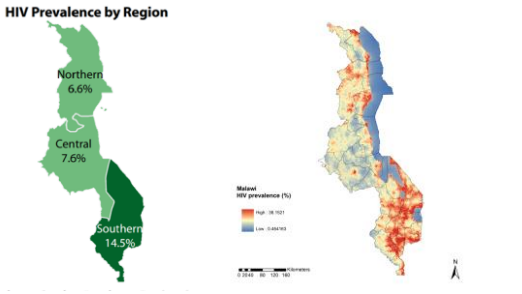




Results: Mapping HIV in Tanzania



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.

Qatar National Research Fund
Member of Qatar Foundation