Spatial and Temporal Malaria Risk Profiles

MARCIA CASTRO
Department of Global Health and Population
Harvard School of Public Health
Research Questions

- What is the local pattern of malaria transmission (spatial and temporal)?
- Does this pattern allow the identification of sub-areas of risk, and how can they be described?
- Which are the local determinants of transmission? Do they change over time and across space? If so, how?
- How interventions for malaria control could be improved?
Defeating Malaria From the Genes to the Globe
defeatingmalaria.harvard.edu
Frontier Malaria (Brazilian Amazon)

- Distinct definition
  - 3 scales
  - Varies by type of human settlement

- Distinct time path
Amazon

- Historical trend
  - Endemic Malaria

- 1970s-present
  - Opening of roads and settlement projects
    - In 1970 - only 52.5 th. cases of malaria in Brazil
    - In 2000 there were more than 600 th.
      - 99.9% in the Amazon
      - 60% in settlement areas
    - Currently - nearly 260 th. cases reported each year
Amazon: from forest ...
... to city
Data

- Field surveys
- Variables created with the aid of GIS
  - Distance, nearest neighbors, buffers
- Landsat 5 TM images (30m of spatial resolution)
  - Risk factors not available in field survey data
RS2 - Illegal deforestation

Landsat-5/TM - July 15, 1994

Landsat-5/TM – April 31, 1995
Data

- Field surveys
- Variables created with the aid of GIS
  - Distance, nearest neighbors, buffers)
- Landsat 5 TM images (30m of spatial resolution)
  - Risk factors not available in field survey data
- Qualitative
  - Focus groups
  - Ethnography
Ethnography

- Presence of rubber tappers
- Areas with intense human mobility
- Severe disturbances to the environment
- Isolated areas
Methodological Approach

A. Spatial Clustering Analysis

B. Spatial Estimation

- Spatial pattern of malaria transmission
- Possible identification of sub-areas of risk

Sub-areas of risk
Methodological Approach

A. Spatial Clustering Analysis → Sub-areas of risk
B. Spatial Estimation
C. Grade of Membership (GoM) → Risk profiles

- Describing sub-areas of risk
- Assessing the most important determinants of transmission
- Targeted interventions for malaria control
A. Spatial Clustering Analysis

- Local Indicators of Spatial Association – LISA
  - $G_i^*(d)$ statistic
    - Identify clusters of high and low values surrounding a particular location $i$ within a distance $d$ from $i$
      - $d = 3,500$ meters
    - Tobler’s First Law of Geography: “everything is related to everything else, but near things are more related than distant things” (Tobler, 1979)
  - Correction for multiple and dependent tests
    - False Discovery Rate (FDR)
      - Proportion of null hypotheses incorrectly rejected among all those that were rejected
$G_i^*(d)$

Distance = 3,500 meters

FDR approach to correct for multiple testing

Significant clusters of low malaria rates
No significant clusters
No significant clusters
Significant clusters of high malaria rates
B. Spatial Estimation

- Kriging - Optimal spatial estimation

- 2-step process:
  - Quantify the spatial structure of the data (semivariogram)
  - Produce an estimation
Kriging

- Low malaria rates
- Medium malaria rates
- High malaria rates
Sub-areas

Malaria Rates (%)

1985
- Sub-area 1 = 7.4
- Sub-area 2 = 35.9
- Sub-area 3 = 24.8

1986
- Sub-area 1 = 26.3
- Sub-area 2 = 27.9
- Sub-area 3 = 17.0
- Sub-area 4 = 43.8

1987
- Sub-area 1 = 27.5
- Sub-area 2 = 24.3
- Sub-area 3 = 14.9
- Sub-area 4 = 15.3
- Sub-area 5 = 38.5

1995
- Sub-area 1 = 11.3
- Sub-area 2 = 2.9
- Sub-area 3 = 11.8
C. Grade of membership - GoM

- Describe heterogeneous data
- Assume that the original variables are independent, conditional on the GoM scores
- Estimate a GoM score that relates an individual to a fuzzy partition (profile)
Understanding the profiles

2 Profiles

- Profile 1
- Profile 2

Share characteristics of both profiles

3 Profiles

- Profile 1
- Profile 2
- Profile 3

Share characteristics of profiles 1 and 3
Share characteristics of profiles 2 and 3
Share characteristics of profiles 1 and 2
Share prof. 1, 2 and 3

4 Profiles

- Profile 1
- Profile 2
- Profile 3
- Profile 4

Vertices – Match one profile
Edges – Share 2 profiles
Faces – Share 3 profiles
Interior – Share all 4 profiles
C. Grade of membership - GoM

- Profile definition given by levels of response of variables
  - Pure-type probabilities - probability in profile \( k \) of observing level \( l_j \) on variable \( X_j \).
    - Establish a cutoff for low and high prevalence (marginal distributions)
      - Below the cutoff – consider those much more frequent than observed in the population
      - Above the cutoff – consider a percentage change in the observed condition
Table 7.2 Description of high and low malaria risk profile in subarea 4, Machadinho, 1986.

<table>
<thead>
<tr>
<th>High malaria risk</th>
<th>Low malaria risk</th>
</tr>
</thead>
<tbody>
<tr>
<td>The settler was not working one year before he or she came to Machadinho,</td>
<td>The settler was engaged in rural activities one year before he or she came to Machadinho,</td>
</tr>
<tr>
<td>arrived in Machadinho in 1986, and the region of residence one year before he</td>
<td>was North. The household owns more than five goods, including a planter. The distance from to</td>
</tr>
<tr>
<td>or she came to Machadinho was either south or center-west. The house has poor</td>
<td>the house to the health post in the urban area is less than 20 km. The plot is not within a zone</td>
</tr>
<tr>
<td>quality walls and roof, has only one or two rooms, and is located in an area with</td>
<td>of 900 meters from a protected forest reserve, and there is no source of permanent water near the</td>
</tr>
<tr>
<td>low elevation. The household owns less than five goods, and lacks a chainsaw.</td>
<td>house. The planted area in the plot is between 16 and 45 ha, and different crops are cultivated.</td>
</tr>
<tr>
<td>The distance to the nearest neighbor is greater than 1 km. The cleared area in</td>
<td>The settler has chickens and pigs in the plot, and the house was sprayed with DDT.</td>
</tr>
<tr>
<td>the plot is smaller than 16 ha, no major crops are cultivated, and the settler</td>
<td></td>
</tr>
<tr>
<td>does not have chickens and pigs in the plot.</td>
<td></td>
</tr>
</tbody>
</table>
What was learned?

- Determinants of transmission:
  - Vary across space and over time

- 2 scenarios in a settlement area:
  - 1st – initial stage; low social and economic diversity; large differences in transmission; **natural and man-made environment**
  - 2nd – later stages; significant social and economic diversity; stable malaria at lower levels; **personal characteristics and behavior**
What was learned?

- Strategies for control should be targeted in time and space
  - The environment and the habitat conditions should be the focus of interventions in the early years of settlement
    - Environmental management
    - Rapid and complete clearance near the house
    - Good quality housing
    - Good quality clearing (chainsaw)
    - Do not leave bare soil exposed for long
  - In 2000 the Brazilian government started a control program that targeted 32% of the municipalities in the Amazon, which accounted for 93.6% of the cases
Imagine their future without the threat of malaria...
$G_i^*(d)$ Statistic
$G_i^*(d)$ Statistic
$G_i^*(d)$ Statistic

- **Multiple and dependent tests**
  - Two sources of spatial dependence
    - Geometric
    - Between the values of nearby locations

- **Multiple comparison correction**
  - Conservative – Bonferroni, Sidak
    - Probability that a true null hypothesis is incorrectly rejected - Type I error
  - False Discovery Rate
    - Proportion of null hypotheses incorrectly rejected among all those that were rejected
      - Proportion of null hypotheses incorrectly rejected among all those that were rejected
GoM formal representation

An individual’s response vector of variables can be written as \( \mathbf{X}^{(g)} = (X_{1}^{(g)}, \ldots, X_{j}^{(g)}) \), with distribution \( \Pr(\mathbf{X}^{(g)} = l) \). Assuming that the original variables are independent, conditional on the GOM scores, the distribution of \( \mathbf{X}^{(g)} \) can be written as (Singer, 1989)

\[
\Pr(\mathbf{X}^{(g)} = l) = \int_{S_{K}} \prod_{j=1}^{J} \Pr(X_{j}^{(g)} = l_{j} | g = \gamma) d\mu(\gamma)
\]

\[
= \int_{S_{K}} \prod_{j=1}^{J} \Pr(X_{j}^{(g)} = l_{j} | g = \gamma) d\mu(\gamma),
\]

where \( \mathbf{X}^{(g)} \) is the distribution of GOM scores and \( S_{K} = \{ \mathbf{y} = (y_{1}, \ldots, y_{K}) : y_{k} \geq 0, \sum_{k} y_{k} = 1 \} \). The conditional probabilities can be written as

\[
\Pr(\mathbf{X}^{(g)} = l_{j} | g = \gamma) = \sum_{k=1}^{K} \gamma_{k} \lambda_{k,j,l_{j}}
\]

where \( \lambda_{k,j,l_{j}} \) (called pure-type probabilities) can be defined as the probability in profile \( k \) of observing level \( l_{j} \) on variable \( X_{j} \).
Grade of membership scores act as weights that relate each element of the fuzzy set to a particular profile. The probability model associated with GOM models is the multinomial. Under this distribution, the likelihood equation is

\[
L(\delta) = \prod_{i=1}^{I} \prod_{j=1}^{J} \prod_{l=1}^{L_j} \left( \sum_{k=1}^{K} \gamma_{i,k} \lambda_{k,j,l} \right)^{\delta_{i,l,j}}
\]

where \(I\) is the total number of elements (individuals) in the fuzzy set, \(\gamma_{i,k}\) are the GOM scores, \(\lambda_{k,j,l}\) are the pure-type probabilities, and \(\delta_{i,l,j}\) assumes values equal to one or zero according to the following conditions (Singer, 1989)

\[
\delta_{i,l,j} = \begin{cases} 
1 & \text{if } X_{j}^{(i)} = l_j \\
0 & \text{otherwise}
\end{cases}
\]