Environmental and socio-economic risk modelling for Chagas disease in Bolivia

Paula Mischler¹, Michael Kearney¹, Jennifer C. McCarroll¹, Ronaldo G.C. Scholte²,³, Penelope Vounatsou²,¹, John B. Malone¹

¹Department of Pathobiological Sciences, School of Veterinary Medicine, Louisiana State University, Baton Rouge, LA 70803, USA; ²Department of Epidemiology and Public Health, Swiss Tropical and Public Health Institute, P.O. Box, CH-4002 Basel, Switzerland; ³University of Basel, P.O. Box, CH-4003 Basel, Switzerland

Abstract. Accurately defining disease distributions and calculating disease risk is an important step in the control and prevention of diseases. Geographical information systems (GIS) and remote sensing technologies, with maximum entropy (Maxent) ecological niche modelling computer software, were used to create predictive risk maps for Chagas disease in Bolivia. Prevalence rates were calculated from 2007 to 2009 household infection survey data for Bolivia, while environmental data were compiled from the Worldclim database and MODIS satellite imagery. Socio-economic data were obtained from the Bolivian National Institute of Statistics. Disease models identified altitudes at 500-3,500 m above the mean sea level (MSL), low annual precipitation (45-250 mm), and higher diurnal range of temperature (10-19 °C; peak 16 °C) as compatible with the biological requirements of the insect vectors. Socio-economic analyses demonstrated the importance of improved housing materials and water source. Home adobe wall materials and having to fetch drinking water from rivers or wells without pump were found to be highly related to distribution of the disease by the receiver operator characteristic (ROC) area under the curve (AUC) (0.69 AUC, 0.67 AUC and 0.62 AUC, respectively), while areas with hardwood floors demonstrated a direct negative relationship (-0.71 AUC). This study demonstrates that Maxent modelling can be used in disease prevalence and incidence studies to provide governmental agencies with an easily learned, understandable method to define areas as either high, moderate or low risk for the disease. This information may be used in resource planning, targeting and implementation. However, access to high-resolution, sub-municipality socio-economic data (e.g. census tracts) would facilitate elucidation of the relative influence of poverty-related factors on regional disease dynamics.

Keywords: Trypanosoma cruzi, Chagas disease, ecological niche model, risk maps, maximum entropy, geographical information system, remote sensing, Bolivia.

Introduction

Chagas disease is caused by the protozoan Trypanosoma cruzi and is transmitted primarily through the introduction of the organism by infected triatomine insect faeces into breaks in the skin or via mucosal membranes. In highly endemic areas, blood transfusions and congenital transmission can be additional important modes of transmission, which is particularly true in Bolivia (Torrico et al., 2004; Araujo-Jorge and Medrano-Mercado, 2009). Bolivia has the highest rate of Chagas infection in the Americas (Guillen et al., 1997; Guillen, 2002), despite implementation of control programmes since the mid-1980s. Seroprevalence rates in some communities can reach as high as 100% in the departments of Cochabamba, Chuquisaca, Tarija and Santa Cruz, where the disease has historically been a public health problem (Arata et al., 1994; Guillen et al., 1997). The existence of sylvatic and peridomestic populations of Triatoma spp., in addition to domestic populations of T. infestans, the primary vector in this area, currently complicate control efforts in this country (Noireau, 2009). Other factors such as rapid urbanization, seasonal migration, vector resistance to insecticides, and inconsistent monitoring and control activities, contribute to both re-infestation and the spread of the vector in Bolivia (Guillen et al., 1997; Araujo-Jorge and Medrano-Mercado, 2009; Briceno-Leon, 2009).

Geographical information systems (GIS) and ecological niche modeling (ENM) have been successfully used to define disease distribution for several diseases including schistosomiasis (Bavia et al., 1999; Malone et al., 2001, Scholte et al., 2012), malaria (Hay et al., 2000; Kiang et al., 2006) and leishmaniasis (Nieto et al., 2006; Rossi et al., 2007; Gil et al., 2010). In addition to defining disease distribution, these technologies can be used to predict risk by extrapolating to areas...
with similar microclimates where sampling data cannot, or have not, been obtained. Accurate predictive disease mapping is useful in the allocation of resources for disease surveillance, control and education by governments where resources are limited.

The development of the maximum entropy (Maxent) ENM software provides an easily learned, understandable method for identifying environmental and socio-economic variables related to neglected tropical diseases (NTDs), which include Chagas disease. The objective of this study was to identify environmental and socio-economic variables related to Chagas risk in Bolivia and to create a predictive risk map from this information using Maxent ENM software in an analysis of disease based on “presence-only” data (Philips et al., 2004).

**Materials and methods**

**Data collection**

Bolivia is divided into nine departments, 112 provinces and 337 municipalities. The country has an estimated population of over 10 million people (2010) and an area of 1,098,581 km². The prevalence of Chagas in the country was calculated from household surveillance data compiled by municipality for 2007, 2008 and 2009 by the Bolivian Ministry of Health. An average prevalence was calculated for each municipality from the years provided. To improve the quality of the data, municipalities with less than 100 households in a given year were not considered. Additionally, data from the municipality of Reyes in the department of El Beni was not evaluated because the validity of data from that municipality is questionable. Although the literature suggests that this area is non-endemic for Chagas, the only year surveyed in this municipality showed a very high prevalence rate (>60%).

Socio-economic data at the municipality level was obtained from the Bolivian National Institute of Statistics (INE) (Table 1). Environmental data used in the analysis was compiled from the United States Geological Survey (http://www.usgs.gov/), the Moderate Resolution Imaging Spectroradiometer (MODIS) (http://modis.gsfc.nasa.gov/) and the Worldclim database (BIO1-19, altitude) (http://www.worldclim.org) (Table 2). Environmental data parameters considered included bioclimatic variables (Biolim), Shuttle Radar Topography Mission (SRTM) data on elevation, Worldclim data on the long-term normal average temperature and precipitation, MODIS remotely sensed data for the normalized difference vegetation index (NDVI) as well as daytime and nighttime land surface temperature (LST). Biolim variables were derived from monthly temperature and rainfall data to provide more biologically meaningful variables such as temperature seasonality, isothermality (a measure of temperature evenness within an area), and precipitation in the wettest quarter (http://worldclim.org/biolim). Socio-economic data considered were the human development index (HDI), infant mortality rate (IMR) and 40 “unsatisfied basic needs” (UBNs) variables relating to housing, water supply, education and income level. Environmental and socio-economic data records were maintained in separate databases for the statistical analysis. This was due to the inherent differences in the data types as well as to facilitate multiple regression modelling evaluations.

| Home with brick, block or concrete walls | Home with electricity |
| Home with adobe wall material | Home without or with inadequate sanitation |
| Home with mud mixed with straw or cane | Public tap water source (non-household source) |
| Home with stone or rock wall material | Drinking water from well with pump |
| Home with cane, palm or trunk wall material | Drinking water from well without pump |
| Home with dirt floors | Drinking water from river/stream |
| Home with hardwood floors | Drinking water from lagoon/lake |
| Home with wood paneled floors | Unemployed |
| Home with carpet floors | Human development index (HDI) 2005 |
| Home with cement floors | Infant mortality rate (IMR) |
| Home with tiled/ceiling floors | Living under overcrowded conditions |
| Home with corrugate metal roofing | Living a subsistence way of life |
| Home with cement or clay roofing | Having a high quality of life |
| Home with concrete slab roofing | Having a medium quality of life |
| Home with straw, reed or palm roofing | Having a low quality of life |
| Home with inadequate housing materials | Life with 1 unsatisfied basic need (UBN) |
| Home with indoor plumbing | Life with 2 UBNs |
| Home with piped drinking water | Life with 3 UBNs |
| Home with access to drinking water | Life with 4 UBNs |
| Home with indoor toilet | Educational UBNs |
Each database was analysed statistically through multiple stepwise regression modelling. Multiple regression was chosen to reduce the number of variables used in the final modelling process as both the environmental and socio-economic variable databases contained a large number of variables. The models were selected through regression analysis, F statistic, Mallows Cp and R² value. Stepwise regression modelling was done on the municipalities that had a Chagas prevalence of 1% or greater. To eliminate multicollinearity between variables, variance inflation factors (VIF) were calculated for each variable in the model and those with VIFs greater than ten were sequentially removed until all remaining variable VIF values were less than 10.

ENMs were developed to define the relationship between the geographic distribution of Chagas in Bolivia, and those environmental and socioeconomic factors that may be related to the disease. Maxent is a maximum entropy approach to presence-only distribution modelling that has shown a high predictive power for both large and very small sample sizes (Hernandez et al., 2006; Phillips et al., 2006; Phillips and Dudík, 2008). AUC values range from 0.5 to 1, where the former value indicates a model no different from chance and the latter gives increasingly good discrimination as it bears 1, i.e. 0.5-0.6 = no discrimination; 0.6-0.7 = discrimination; 0.7-0.8 = acceptable; 0.8-0.9 = excellent; 0.9-1.0 = outstanding (Phillips et al., 2006).

Results and Discussion

A total of 99 Bolivian municipalities had a prevalence of 1% or greater and were evaluated by multiple linear regression analysis. From this preliminary modelling step, the following environmental variables were identified as being significantly associated with disease: altitude, mean diurnal range (BIO 2) and isothermality (BIO 3) from the Bioclim database, and precipitation in the month of May, from the Worldclim long-term-normal monthly climate database. Socio-economic variables selected by regression modeling included living in a home with either adobe walls, wooden floors or a corrugated metal roof and the use of water from a well without pump, the use of piped water, the use of river water and/or subsistence living.

Socio-economic and environmental variables were combined as Maxent input data and evaluated with point records of the prevalence reported for Chagas at the administrative city centre of the 99 Bolivian municipalities. Probability risk maps generated by Maxent are shown in Fig. 1. The predicted risk maps for Chagas in Bolivia generated by Maxent analysis are expressed as probability percentage (Fig. 1a) and
low, moderate and high risk category (Fig. 1b). Table
3 shows the results of regression analysis and the per-
cent contribution of each variable to the Maxent
model. The relative importance of each variable in the
Maxent model was evaluated by jackknife plots of
AUC (Fig. 2).

The average test AUC for the 10 replicate runs of the
model was 0.87 with a standard deviation of 0.03
(Fig. 1). The probability map output for Chagas in
Bolivia generated by Maxent is presented in Fig. 1a.
Thresholds were then determined using average “max-
imum training sensitivity plus specificity logistic
threshold” and “balance training omission, predicted
area and threshold value” over the 10 model iterations
(Cantor et al., 1999; Cramer, 2003; Liu et al., 2005;
Wei et al., 2011). The results gave thresholds of 0.20
and 0.08, respectively which indicate the various grades
of area risk, i.e. high (dark brown), moderate (light
brown), and low (tan) demonstrated in Fig. 1b.

Altitude was found to be the single most important
factor influencing the model by both the jackknife
analysis and percent contribution to the model (Fig. 2,
Table 3). This particular variable had the highest gain
when used in isolation and decreased the gain the most
when omitted from the model, indicating that it was
more informative than the other variables. The
response curve shows an increase from 500 m above
the mean sea level (MSL) to 3,500 m above MSL with
a peak around 3,000 m above MSL (Fig. 4a). Other
reports have documented that T. infestans, the pri-
mary vector in Bolivia, is found concurrent with
human disease from 330 meters above MSL to 3,500
m above MSL (Borda-Pisterna, 1985; Guillen et al.,
1997; Zuna et al., 1991).

The Worldcim variable May precipitation had the
second largest contribution to the model both in terms
of percent contribution and in the jackknife rate of
gain analysis (Table 3, Fig. 4b). The response curve
increased with precipitation from 45 mm, peaking at
125 mm and then decreased to 250 mm. It has been
reported that the vector species, T. infestans is found
in dry areas where the annual relative humidity does
not exceed 60%, supporting the importance of this
variable to the model. Humidity acts as an ecological

Fig. 1. (a) Maxent predicted risk map expressed as percent probability of Chagas occurrence in Bolivia at municipality level; (b)
Maxent model for Chagas in Bolivia showing low, moderate and high risk areas.
barrier to this triatomine species and is reported to limit its dispersal and reproductive capacity (Borda-Pisterna, 1985).

Other environmental factors that significantly influenced the model were mean diurnal range and isothermality. Mean diurnal range provides information about the difference in day to night temperature and can be affected by relative humidity and cloudiness. Isothermality relates the mean diurnal range to the annual temperature range, demonstrating the daily to yearly variation in temperature. The response curves for these variables (Figs. 4c and 4d) indicated an increased suitability in areas with a moderate isothermality and a moderate mean diurnal range, which may explain why northern Bolivia (high isothermality and low mean diurnal range) and far western Bolivia (low isothermality and high mean diurnal range) were predicted as low risk. Northern Bolivia is very hot and humid and is located at a relatively low elevation. Far western Bolivia is mountainous and has a dry, cold climate.

Living in a home with adobe walls and/or wooden floors were the two most influential predictors among the socio-economic variables. The response curve for adobe wall homes shows a correlation between the percentage of people with this type of wall material and increased likelihood of disease occurrence (Fig. 5a). It is well documented in the liter-
ature that adobe walls provide the insect with shelter and access to human blood meals (Albarracin-Veizaga et al., 1999; Dias and Schofield, 1999). There was a negative correlation observed between hardwood floors and disease occurrence (Fig. 5b). Hardwood flooring in Bolivia is rare and typically associated with higher socio-economic conditions. The presence of corrugated metal roofing material was also correlated with disease (Fig. 5c); however it has been observed in a study in Cochabamba, Bolivia that 86.7% of the houses surveyed have this type of roof (Albarracin-Veizaga et al., 1999), and the correlation to corrugated metal roofing could be due to the predominance of this roof type in endemic areas in Bolivia. Subsistence living was also related to disease occurrence and showed a positive correlation (Fig. 5d). Drinking water sources contributed to the model, though at lower level of importance. Fetching water from a well without pump showed a negative relationship with disease occurrence, while piped water showed the opposite relationship (Figs. 5e, 5f). Collection of river water for drinking increased with disease occurrence up to 50% and then decreased (Fig. 5g). The relationship between piped drinking water might be related to the movement of the disease from rural areas into urban environments and would therefore be indicative of the increased number of people with this access and not necessarily disease.

The predicted risk maps (Figs. 1a, 1b) are consistent with literature case reports of the distribution of both the disease and the primary vector in Bolivia (T. infestans). It is anticipated that the model can be further validated and strengthened by obtaining disease surveillance data for additional years and comparing both statistical and Maxent ENM analysis annually. Future systematic surveys of Chagas in Bolivia may, moreover, reveal unknown gaps on Chagas leading to improve accuracy of Maxent model geospatial extrapolations. Finally, regression modelling was able to identify a strong socio-economic model for Chagas, but some of the relationships between variables remain weak and unclear. A solution for clarifying the role of socio-economic factors would be to obtain measures of the same variables at a finer level other than municipality, within which the value of a single variable can fluctuate widely in income level and poverty within different zones of a single municipality. Census tract level data for socio-economic variables would give a more accurate estimate of the risk attributable to each indicator of poverty and refine the accuracy of current Maxent predictive risk models.

Fig. 5 a-g. Response curves for socio-economic variables related to the distribution of Chagas. The red line is the mean value for the 10 Maxent runs and the blue bar represents ± 1 standard deviation.
Conclusions

Results of these studies indicate GIS and Maxent ENM are useful tools for defining Chagas distributions in Bolivia and extrapolating results to areas with similar environmental and socio-economic risk factors and that ENM models are suitable for making health planning decisions and monitoring progress on a national level. This is particularly important in data-scarce environments where resources and health data are limited, and effective resource allocation and planning are needed to have a significant impact on sustained disease control programmes.

References


Noireau F, 2009. Wild Triatoma infestans, a potential threat that needs to be monitored. Mem Inst Oswaldo Cruz 104, 60-64.


Scholte RGC, Carvalho OS, Malone JB, Utzinger J, Yousatsou P, 2012. Spatial distribution of Biomphalaria spp, the inter-
mediate host snails of *Schistosoma mansoni* in Brazil. Geospat Health 6, S95-S101.

