Spatial epidemiology of bat-borne rabies in Colombia

To the Editor:

In their article, "Geographic distribution of wild rabies risk and evaluation of the factors associated with its incidence in Colombia, 1982–2010," Brito and colleagues (1) analyzed the spatial distribution of bat-borne rabies in Colombia and showed an association of occurrence of rabies outbreaks with climatic tendencies. The authors also identified risk clusters for rabies based on modeled potential distributions. While we appreciate this work, which fills knowledge gaps regarding the epidemiology of vampire-borne rabies in Latin America, we are also concerned because some of the data presented merits re-evaluation and correction.

The distribution of rabies in Colombia was modeled using an ecological niche modeling approach. The authors analyzed its distribution with respect to 15 climatic variables, using all available occurrences in model calibration. Model evaluation, a critical issue if modeling results are to be incorporated in public health decision making, was thus limited in scope and strength. Our goal herein is to evaluate the quality, quantity, and distribution of rabies occurrences data and effectiveness of the environmental data used by Brito and colleagues. We based our analyses on the data used in and provided by the article.

First, Brito and colleagues used all available occurrence points in model calibration. Generally, however, clusters of occurrences can lead to overfitting models to specific environments in specific areas (2). To avoid biases in model calibration, we therefore used single occurrences per cell—indeed, in some cases, multiple occurrences grouped within single cells in the environmental layers used by the authors. This step left 2 008 of the original 2 330 occurrences. This minor correction reduces overfitting, and increased area identified as suitable by > 4%.

Another major concern centers on use of climatic variables as environmental information in analyses across such restricted extents. Using principal components analysis, we found the climatic layers to be highly intercorrelated: indeed, just three principal components summarized > 99% of variance among the 15 climatic layers analyzed by Brito and colleagues. We compared models calibrated with respect to that same climatic information against those based on remotely sensed data (monthly Normalized Difference Vegetation Index [NDVI] data sets for October 1992 – September 1993 0.01° resolution), which provide finer spatial resolution and less need for interpolation. Indeed climate data showed environmental homogeneity across broad areas, with spatial lags of > 350 km, whereas remotely-sensed data offered a richer and more detailed environmental characterization: 11 principal components were required to summarize > 99% of overall variance, and spatial lags were only < 250 km (Figure 1). The contrast in detail available from these two data sources is evident in Figure 2.

Brito and colleagues subdivided available data at random for calibration versus evaluation of their model, and used approaches that weight presence and absence data equally (2). We evaluated model predictions more rigorously, with spatial subdivision of occurrences into calibration and evaluation areas, and using evaluation tools that emphasize presence information (2, 3). Models calibrated based on presence

FIGURE 1. Variogram plot of spatial autocorrelation based on 500 random points from the respective first principal component of climate and remotely-sensed environmental data sets. Vertical axis is semivariance for climate (black dots = 10^{-4}) and NVDI (gray dots = 10^{-1}); dashed lines represent sill (horizontal) and range (vertical), nugget = 0
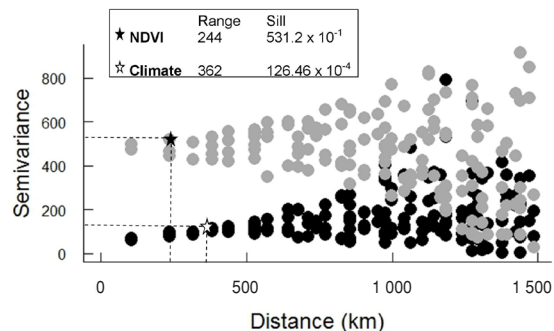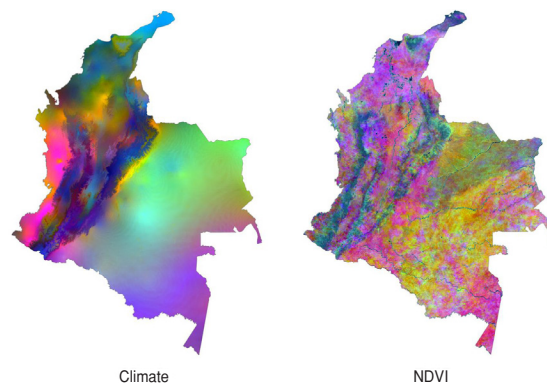

FIGURE 2. Principal component analysis results from climatic (left) and NDVI (right) variables in Colombia. Colors represent environmental differences

Climate   NDVI

data only should not be evaluated using presence-absence methods (4), an issue of particular importance in public health. Our re-analysis of the original data generated models that were statistically robust in both cases ($P < 0.05$); however, models based on climate showed more omission error than models based on remotely-sensed data, in spite of climate-based models covering more area.

Making these adjustments to methodology, our final model corresponded to some degree to that of Brito and colleagues (Figure 3). However, our model identified potential transmission areas in finer detail: the Brito model covered 9.0×10^5 km^2 (79%) of Colombia, whereas our model covered 3.9×10^5 km^2 (49%). Curiously, while identifying areas in finer detail and covering less area, our map of risk areas included eight municipalities not identified in the Brito model (Alto Baudó, Bajo Baudó, La Chorrera, Lloró, Puerto Concordia, Puerto Santander, Quibdó, and Tadó).

Spatial epidemiology of rabies is still a topic in development, with focus on terrestrial reservoirs, but standard and tested methodologies are still lacking. The smaller, but broader areas identified in our analyses result from a combination of more careful methodology and interpreting model outputs in the framework of biogeography and ecology (2). Spatial epidemiology strives to anticipate the spatial behavior of disease transmission across landscapes. Ecological niche models are designed to anticipate geographic distributions of suitable areas for species (2), and thus have much to offer in public health (5). Nonetheless, application of these methods to disease mapping demands careful design, implementation, and testing, particularly in view of the serious nature of diseases such as rabies. We encourage future research in zoonotic disease mapping to progress via multidisciplinary teams that include ecologists, epidemiologists, and veterinarians.

Luis E. Escobar
Facultad de Ecología y Recursos Naturales
Universidad Andres Bello
Santiago, Chile
ecoguate2003@yahoo.com

A. Townsend Peterson
Biodiversity Institute
University of Kansas
Lawrence, Kansas, United States of America

FIGURE 3. **Potential distribution of bat-borne rabies in Colombia, according to Brito and colleagues**, **2013 (left). Revised map (right). Notice that rivers had strong associations with outbreak predictions. A threshold of** E **= 10% was used to generate binary maps**
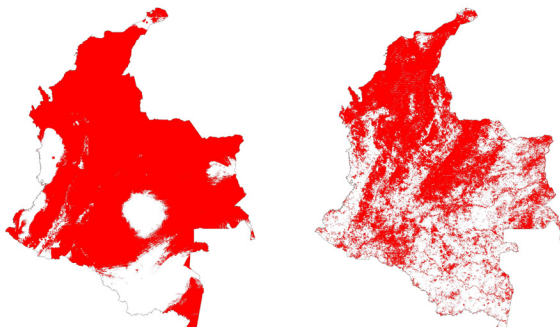
REFERENCES

1. Brito-Hoyos DM, Sierra EB, Álvarez V. Distribución geográfica del riesgo de rabia de origen silvestre y evaluación de los factores asociados con su incidencia. Rev Panam Salud Publica. 2013;33(1):8-14.
2. Peterson A, Soberón J, Pearson R, Anderson R, Martínez-Meyer E, Nakamura M, et al. Ecological niches and geographic distributions. New Jersey: Princeton University Press; 2011.
3. Peterson AT. Niche modeling: model evaluation. Biodiv Inform; 2012;8:41.
4. Lobo JM, Jiménez-Valverde A, Real R. AUC: a misleading measure of the performance of predictive distribution models. Glob Ecol Biogeogr. 2007;17:145–51.
5. Peterson AT. Biogeography of diseases: a framework for analysis. Naturwissenschaften. 2008;95(6):483-91.