Utilizing environmental, socioeconomic data and GIS techniques to estimate the risk for ascariasis and trichuriasis in Minas Gerais, Brazil

Ronaldo G.C. Scholte a,b,*, Corina C. Freitas c, Luciano V. Dutra c, Ricardo J.P.S. Guimaraes c, Sandra C. Drummond d, Guilherme Oliveira e, Omar S. Carvalho e

a Department of Epidemiology and Public Health, Swiss Tropical and Public Health Institute, P.O. Box, CH-4002 Basel, Switzerland
b University of Basel, P.O. Box, CH-4003 Basel, Switzerland
c National Spatial Research Institute/INPE São José dos Campos, SP, Brazil
d Health State Office of Minas Gerais, Belo Horizonte, MG, Brazil
e Research Center René Rachou/Fiocruz-MG, Av. Augusto de Lima, 1715 Barro Preto, CEP 30190-002 Belo Horizonte, MG, Brazil

ARTICLE INFO

Article history:
Received 12 July 2011
Received in revised form 11 October 2011
Accepted 15 October 2011
Available online 21 October 2011

Keywords:
Geohelminths
Ascariasis
Trichuriasis
Geographical information systems
Multiple regression

ABSTRACT

The impact of intestinal helminths on human health is well known among the population and health authorities because of their wide geographic distribution and the serious problems they cause. Geohelminths are highly prevalent and have a big impact on public health, mainly in underdeveloped and developing countries. Geohelminths are responsible for the high levels of debility found in the younger population and are often related to cases of chronic diarrhea and malnutrition, which put the physical and intellectual development of children at risk. These geohelminths have not been sufficiently studied. One obstacle in implementing a control program is the lack of knowledge of the prevalence and geographical distribution. Geographical information systems (GIS) and remote sensing (RS) have been utilized to improve understanding of infectious disease distribution and climatic patterns. In this study, GIS and RS technologies, as well as meteorological, social, and environmental variables were utilized for the modeling and prediction of ascariasis and trichuriasis. The GIS and RS technologies specifically used were those produced by orbital sensing including the Moderate Resolution Imaging Spectroradiometer (MODIS) and the Shuttle Radar Topography Mission (SRTM). The results of this study demonstrated important factors related to the transmission of ascariasis and trichuriasis and confirmed the key association between environmental variables and the poverty index, which enabled us to identify priority areas for intervention planning in the state of Minas Gerais in Brazil.

© 2011 Elsevier B.V. All rights reserved.

Helminths belong to one of the most important and largest zoological groups in the animal kingdom that includes humans, animals and plant parasites. They are called geohelminths because of their cycle in the environment. They are etiologic agents of human infectious disease usually caused by people consuming their eggs or via active worm penetration through the skin. They are especially prevalent in underdeveloped and developing countries, where the combination of climatic factors and poor hygienic conditions facilitates transmission (Bethony et al., 2006; Brooker et al., 2007). It is an important health problem (WHO, 2005).

Helminths, also known as soil-transmitted helminths (STHs), are responsible for physical and intellectual delays in children, mainly in early infancy, and interfere with school performance, attendance and productivity in children and adolescents (Bleakley, 2003; Miguel and Kremer, 2003). Although STHs have an important impact on education, economy and public health, they are still neglected diseases (Chan, 1997). The most severe infections occur primarily in children between the ages of 5 and 15 years. The intensity and frequency of the disease reduces as age increases (Galvani, 2005). More than one billion people globally are infected with at least one type of STH. Ascaris lumbricoides (A. lumbricoides) and Trichuris trichiura (T. trichiura) are the most important of the STHs (WHO, 2005; Bethony et al., 2006).

Climatic conditions have an important role in determining infection rates. The prevalence of the infection is relatively low in high altitude regions with low rainfall patterns. Hot and humid weather patterns are considered vital for the survival of the parasites especially during the embryonic stage of the eggs. Furthermore, because basic sanitation has been virtually nonexistent in areas of high prevalence, over many years a large population has contributed to an increase in the parasitic load and the pervasiveness of the infection (Massara and Enk, 2004).

* Corresponding author at: Department of Epidemiology and Public Health, Swiss Tropical and Public Health Institute, P.O. Box, CH-4002 Basel, Switzerland.
Tel.: +41 61 284 88226; fax: +41 61 284 8105.
E-mail address: ronaldo.scholte@unibas.ch (R.G.C. Scholte).

0001-706X/$ – see front matter © 2011 Elsevier B.V. All rights reserved.

Contents lists available at SciVerse ScienceDirect

Acta Tropica

journal homepage: www.elsevier.com/locate/actatropica
Since these diseases are caused by risk factors present through time and space, the geographic information system (GIS) and the remote sensing (RS) are powerful tools that can help elucidate the prevalence of infection and the distribution of risk factors. The joint use of GIS tools and statistic techniques allow for elaborate prediction and the creation of indication models of disease control strategies with better subsequent resource optimization (Beck et al., 1997, 2000; Bavia et al., 2001; Guimarães et al., 2006, 2008, 2009, 2010b). Brooker and Michael (2000) demonstrated in epidemiologic studies of helminthic infection control programs that the GIS/RS combination is a useful tool to better understand epidemiology aspects.

This study aims to create prediction models for cases of ascariasis and trichuriasis by utilizing social and environmental variables and to analyze them using GIS/RS techniques and multiple regressions.

1. Materials and methods

1.1. Study area

Minas Gerais is the fourth largest Brazilian state, approximately 590,000 km², and has the second highest population of all Brazilian states, estimated at 19 million inhabitants distributed through 853 municipalities. Fig. 1 shows the study area and its division into municipalities. The state is divided into five distinct ecological zones. For example, the north part of the state is arid, and the south is hilly and green (IBGE, 2011).

1.2. Ascariasis and trichuriasis prevalence data

Information regarding the prevalence (Pv) of ascariasis and trichuriasis in 449 districts (a dependent variable) in the state of Minas Gerais were obtained from annual reports (1997 to 2008) of the Schistosomiasis Control Program of the Health State Office of Minas Gerais.

1.3. Meteorological variables

The meteorological variables included total precipitation, minimum and maximum temperatures in the summer (between 01/17/2002 and 02/01/2002) and winter (between 06/28/2002 and 07/12/2002) and were obtained by the Center for Weather Forecast and Climate Studies (CPTEC/INPE). The year 2002 was chosen to represent meteorological data, which were considered typical with a small weather influence by the El Niño and La Niña weather systems.

1.4. Social and sanitation variables

Social variables, including basic sanitation (the water and sewage system), data from the Human Development Index (HDI) for the years of 1991 and 2000 and the rural population (Rural_Pop), were obtained from the Brazilian Urban Indicator System (BUIS) and the Brazilian Institute of Geography and Statistics (IBGE). These data are from the census and are updated every 10 years. The healthcare need index (HNI) was obtained from the João Pinheiro Foundation, which considers information from 49 variables related to the health, sanitation and the economy.

1.5. Moderate resolution imaging spectroradiometer (MODIS)

MODIS images were obtained for Minas Gerais during two seasons, summer (from January 17 through February 1 of 2002) and winter (from July 28 through August 12 of 2002). MODIS images with a 250 m spatial resolution of the normalized difference vegetation index (NDVI) and the enhanced vegetation index (EVI) were used in this study and ranged from blue-, red-, near-, to middle-infrared bands. Four images downloaded from MODIS’ website were used to cover the entire state.

![Fig. 1. The study area of Minas Gerais, Brazil.](image-url)
1.6. Shuttle radar topography mission (SRTM)

The Digital Elevation Model (DEM) was obtained by the Shuttle Radar Topography Mission (SRTM), creating the most complete and high-resolution digital Earth topography data. The SRTM was a modified radar system that flew with the Space Shuttle Endeavour during an 11-day mission in February of 2000. It has a spatial resolution of 90 m on the equator. The altitude was obtained by

---

Figure 2. Prevalence as a function of remote sensing variables: (a) observed prevalence, (c) estimated prevalence and (e) residuals for *Ascaris lumbricoides*; (b) observed prevalence, (d) estimated prevalence and (f) residuals for *Trichuris trichiura*. 
Table 1
Variables selected for the creation of the predictive model of ascariasis and its values.

<table>
<thead>
<tr>
<th>Estimated parameters</th>
<th>Standard deviation</th>
<th>r</th>
<th>P value</th>
<th>Beta coefficients</th>
<th>Std. deviation Beta</th>
</tr>
</thead>
<tbody>
<tr>
<td>Intercepto</td>
<td>3.16</td>
<td>1.04</td>
<td>3.04</td>
<td>0.0026</td>
<td></td>
</tr>
<tr>
<td>IP,I</td>
<td>13.73</td>
<td>2.07</td>
<td>6.62</td>
<td>0.0000</td>
<td>0.34</td>
</tr>
<tr>
<td>δT,J</td>
<td>−0.18</td>
<td>0.03</td>
<td>−6.12</td>
<td>0.0000</td>
<td>−0.30</td>
</tr>
<tr>
<td>INS</td>
<td>1.05</td>
<td>0.21</td>
<td>4.86</td>
<td>0.0000</td>
<td>0.24</td>
</tr>
<tr>
<td>Tmin,V</td>
<td>−0.16</td>
<td>0.04</td>
<td>−3.80</td>
<td>0.0002</td>
<td>−0.20</td>
</tr>
</tbody>
</table>

the SRTM DEM for each pair of coordinates. The local declivity was derived from the SRTM DEM according to Guimarães et al. (2008).

1.7. Spectral linear mixing model (SLMM)

The SLMM is a process which creates an algorithm of images using fractions or proportions of each component (vegetation, soil and shade) inside the pixel, estimated by the minimization of the sum of the squared error (Shimabukuro and Smith, 1991). The response of the pixel in any spectral band is assumed to be a linear combination of responses of each individual response component (according to Guimarães et al., 2010a).

In this study, the vegetation, soil and shade image fractions were made by utilizing the MODIS data and the estimated values for the reflection components (Guimarães et al., 2010a).

Since the Pv data are provided by the municipalities of Minas Gerais, each variable was integrated inside municipality limits, utilizing the GIS programs (ArcGis 9.3, ESRI, Redlands, California, US) and exported to a standard excel spreadsheet for the statistical analysis and modeling.

1.8. Linear regression model for the estimation of the infection rate

In this study, a relationship between the Pv of ascariasis and trichuriasis and the aforementioned variables was established by utilizing multiple regression models.

The multiple linear regressions were employed based on the global model, in which a linear regression was established to estimate the disease rates throughout the state.

Homogeneous and contiguous regions were determined for the state using biological variables and the Skater algorithm (Assunção et al., 2006; Martins-Bedê et al., 2009; Guimarães et al., 2010b).

The model was built utilizing the data from the 449 municipalities, shown in Fig. 2a and b, where the Pv information was available. The fitted models were then used to build the risk map for the entire Minas Gerais state applying the model to the remaining municipalities.

1.9. Applied model

The values for the rates of ascariasis and trichuriasis infection, environmental and social variables, including some sanitation information and information from the human development index, served as input data to create the multiple regression models and estimate the infection rate. The relations between the dependable variable (Pv), the 69 independent social, meteorological, remote sensing and sanitation variables, were analyzed in terms of correlation and multicollinearity.

The logarithmical changing of the Pv variable improved the correlation with independent variables. The process of the variable selection first involved the variable reduction with a nonsignificant correlation with the Pv at a significance level of 5%. After this process, the variable selection was completed, utilizing the $R^2$ criterion, registering all possible models of regression (Neter et al., 1996). The selection criterion is the identification of a subgroup with a few variables and a coefficient of determination $R^2$. It is similar to what we have when all variables are used in the model. In the regression, $R^2$ is the closest statistical measure of the regression line, which is close to the real points. It measures the proportional reduction of the total variation related to the use of independent variables.

The final statistical model was built, utilizing the STATISCA 6.0 (StatSoft, Tulsa, Oklahoma, US) program based on 449 districts with some information of Pv. The adjusted model was utilized to create the risk map for the entire state of Minas Gerais using the application model for other districts.

2. Results

The relationship between the prevalence of ascariasis and trichuriasis, the social variables and the meteorological and remote sensing data of municipalities in the state of Minas Gerais was determined using multiple regression models, and these results were combined with the geographical information system data.

2.1. Ascaris lumbricoides

The matrix analysis of correlation has shown that some variables do not have meaningful correlations with Pv at a confidence level of 95%, and some predictive variables are correlated to themselves showing that the model can be simplified.

The final model had four variables (Table 1): near infrared in the winter (NIR,W), the difference between the maximum and minimum winter temperature ($dT,W$), the minimum summer temperature ($Tmin,S$) and the healthcare need index (HNI) with a coefficient of determination ($R^2$) of 0.30 and statistical significance of $p<0.05$. The $dT,W$ variable was correlated negatively to the prevalence of ascariasis, while NIR,W, Tmin,S and HNI

Table 2
Variables selected for the creation of the predictive model of trichuriasis and its values.

<table>
<thead>
<tr>
<th>Estimated parameters</th>
<th>Standard deviation</th>
<th>r</th>
<th>P value</th>
<th>Beta coefficients</th>
<th>Std. deviation Beta</th>
</tr>
</thead>
<tbody>
<tr>
<td>Intercepto</td>
<td>1.84</td>
<td>0.60</td>
<td>3.03</td>
<td>0.003</td>
<td></td>
</tr>
<tr>
<td>Tmin,I</td>
<td>0.17</td>
<td>0.03</td>
<td>4.90</td>
<td>0.0000</td>
<td>0.40</td>
</tr>
<tr>
<td>Tmin,V</td>
<td>−0.21</td>
<td>0.04</td>
<td>−4.84</td>
<td>0.0000</td>
<td>−0.39</td>
</tr>
<tr>
<td>DEC</td>
<td>0.06</td>
<td>0.009</td>
<td>6.91</td>
<td>0.0000</td>
<td>0.35</td>
</tr>
<tr>
<td>SOMB,I</td>
<td>−0.03</td>
<td>0.01</td>
<td>−6.70</td>
<td>0.0000</td>
<td>−0.35</td>
</tr>
<tr>
<td>Pop,Rural</td>
<td>−0.007</td>
<td>0.001</td>
<td>−4.14</td>
<td>0.0000</td>
<td>−0.24</td>
</tr>
<tr>
<td>INS</td>
<td>0.59</td>
<td>0.2</td>
<td>2.95</td>
<td>0.003</td>
<td>0.20</td>
</tr>
</tbody>
</table>
demonstrated a positive correlation. After the variable determination, the final regression equation was as follows:

\[
P_V = e^{(3.17 + 13.73(NIR_W) - 0.18(dT_W) - 0.16(T_{\text{min}_S}) + 1.05(HNI))} - 1
\] 

This equation was utilized to estimate the \( P_V \) for all districts in the state of Minas Gerais (Fig. 2c). Fig. 2e shows the plot of the residuals from the difference among observed values (Fig. 2a) and the ascariasis \( P_V \) estimation (Fig. 2c). In this figure, the dark brown area shows the subestimates values, the light brown area shows the overestimated values and the blue area the municipalities with an accurate estimate.

2.2. Trichuris trichiura

For the trichuriasis \( P_V \), the final model had six variables (Table 2): declivity (DEC), winter shadow (SHADE\_W), rural population (Rural\_Pop), winter minimum temperature (Tmin\_W), summer minimum temperature (Tmin\_S) and the healthcare need index (HNI), with \( R^2 = 0.31 \). The SHADE\_W and the Rural\_Pop variables were correlated negatively with the trichuriasis IR, whereas the DEC, Tmin\_S, Tmin\_W and the HNI demonstrated a positive correlation. After the important variable determination for the model, the final regression equation was as follows:

\[
P_V = e^{(1.84 + 0.06(\text{DEC}) - 0.04(\text{SHADE\_W}) - 0.01(\text{Rural\_Pop}) + 0.17(T_{\text{min}_W}) - 0.22(T_{\text{min}_S}) - 0.59(HNI))} - 1
\]

This equation was used to estimate the \( T. trichiura \) for all districts. Fig. 2f shows the plot of the residuals from the difference among observed values (Fig. 2b) and the estimated trichuriasis \( P_V \) (Fig. 2d). In this figure, the dark brown areas show the municipalities with underestimated values and the blue areas show the municipalities with an accurate estimate.

3. Discussion

Variables from many sources (social, meteorology and remote sensing) were utilized in this study. The \( P_V \) estimate equation shows that meteorological (Tmin\_S) and social (HNI) variables are important because they are present in both disease models. These are the same variables (meteorological and social) which Guimarães et al. (2010b) suggested as important ones to use in creating a forecast map for Schistosoma mansoni in the state of Minas Gerais.

These variables were selected based on the data about the biology of \( A. lumbricoides \) and \( T. trichiura \). The factors that facilitate disease transmission are an ideal temperature for the egg development that varies from 28 to 35 °C and the living conditions of low-income populations (Crompton and Pawlowski, 1985; Beer, 1971, 1973, 1976).

- **Ascaris lumbricoides**

For a better understanding of these results, we researched the biology of helminths and found that the ideal temperature for the egg development is between 28 and 32 °C. Hence when the dT\_W is high, the temperature variation affects egg development such that the positive rate decreases when the Tmin\_S is high. An increase in humidity makes an ideal environment for egg development, which increases the infection rate. The NIR\_W variable is related to vegetation. Areas of vegetation in less sandy and more humid soil facilitate egg development and increase the helminth infection rate. Alternatively, the HNI is directly related to the life conditions of the population: the more healthcare that is needed, the poorer life conditions the population will have, and therefore the infection probability increases (Brooker and Michael, 2000; Hotez et al., 2003).

- **Trichuris trichiura**

The SHADE\_W variable is correlated with the presence of water collection. The \( T. trichiura \) eggs are not resistant to water immersion (Brooker and Michael, 2000) making egg development difficult and reducing the possibility of transmission.

The Rural\_Pop variable has shown a negative correlation with trichuriasis because the rural population is diffuse and the presence of parasites is related to population concentration (Crompton and Savioi, 1993; Phiri et al., 2000). In summary, in the areas with a large rural population, the trichuriasis infection rate is reduced.

High and low temperatures limit \( T. trichiura \) egg evolution, as the ideal temperature for development is between 28 and 35 °C. High temperature kills eggs, while low temperature postpones their development. Therefore, a high Tmin\_S variable increases the humidity, making the hot and humid environment perfect for the egg development, and therefore increasing the infection rate. Alternatively, when the Tmin\_W is low, the egg evolution is slow, which increases the parasite’s life cycle and consequently reduces the infection rate. (Seamster, 1950; WHO, 1967; Appleton and Gouws, 1996; Appleton et al., 1999; Brooker and Michael, 2000; Hotez et al., 2003).

The DEC variable is related to the water dispersion speed and the rainfall accumulation. Lower declivity lowers the probability of infection, perhaps because these conditions make the eggs not as viable.

The \( T. trichiura \) infection has fertile territory in the poorest communities with poor living conditions, so the HNI is a good indicator for the occurrence of the parasite. A higher index number correlates with a higher trichuriasis infection rate.

Some questions related to the nature and the precision of \( P_v \) must be considered when the results are observed. The majority of the \( P_v \) data were obtained before GPS equipment was available for this purpose. The information is on a district scale, and most districts are in the north and northeast of the state. Therefore, the data do not show the spatial distribution of the parasites and the exact geographical locations. These issues were discussed by Guimarães et al. (2006) and they may have affected and masked the infection rate correlations with the independent variables.

First, the results of this study reveal important factors related to the disease transmission and confirm the relevance of basic sanitation (proper treatment of effluents and potable water supply) and health education as measures which could reduce the probability of infection. (Massara and Enk, 2004). Furthermore, the approach here can be useful in the identification of priority areas for intervention, aiding in making proper decisions and more efficiently using the resources of the economy. Therefore, even with a low coefficient of determination, the combination of the GIS/RS and statistical techniques, used with socioeconomic, meteorological and remote sensing data allows the identification of regional factors related to ascariasis and trichuriasis transmission and facilitate the delimitation and classification of areas according to the transmission probability of these parasites. It is important to conduct georeferencing studies of the prevalence of geohelminths in districts where there are no such data, to improve the modeling precision for more accurate estimates and use the application in the identification of priority areas on a larger scale.

The database created by this project allows it to be used in other modeling studies of STHs or for other diseases.

The current priority in this technique is the improvement of accuracy in geolocation and in the accuracy of risk estimation. The method of building the field data registry has been optimized with the introduction of cutting-edge geographical location equipment making a more precise determination of a possible case event. It is a great resource for the development of risk models with a better spatial resolution.
Acknowledgements

CNPq (300679/2011-4, 384571/2010-7, 302966/2009-9, 308253/2008-6), FAPEMIG (EDP 1775/03, EDT 61775/03, CRA 0070/04), NIH-Fogarty (5D43TW007012), FIC Training Grant (1D43TW006580).

References


