Analysis and estimative of schistosomiasis prevalence for the state of Minas Gerais, Brazil, using multiple regression with social and environmental spatial data

Ricardo JPS Guimarães***, Corina C Freitas/+, Luciano V Dutra, Ana CM Moura*, Ronaldo S Amaral**, Sandra C Drummond***, Márcio Guerra******, Ronaldo GC Scholte****/*****, Charles R Freitas*, Omar S Carvalho******

Instituto Nacional de Pesquisas Espaciais, Caixa Postal 515, 12201-970 São José dos Campos, SP, Brasil *Universidade Federal de Minas Gerais/UFMG, Belo Horizonte, MG, Brasil **Secretaria de Vigilância em Saúde/MS, Brasíla, DF, Brasil ***Secretaria Estadual de Saúde, Belo Horizonte, MG, Brasil ****Centro de Pesquisas René Rachou-Fiocruz, Belo Horizonte, MG, Brasil *****Programa de Pós-Graduação em Clínica Médica/Biomedicina da Santa Casa de Misericórdia de Belo Horizonte, Belo Horizonte, MG, Brasil ******Companhia de Tecnologia da Informação do Estado de Minas Gerais, Belo Horizonte, MG, Brasil

The aim of this work is to establish a relationship between schistosomiasis prevalence and social-environmental variables, in the state of Minas Gerais, Brazil, through multiple linear regression. The final regression model was established, after a variables selection phase, with a set of spatial variables which contains the summer minimum temperature, human development index, and vegetation type variables. Based on this model, a schistosomiasis risk map was built for Minas Gerais.

Key words: schistosomiasis - risk map - geographical information system - multiple linear regression

Schistosomiasis is endemic in 74 tropical developing countries. It is estimated that 200 million people are already infected, and that 600 million people are at risk of becoming infected (WHO 1999). Disease prevalence is heterogeneous in vulnerable locales and tends to be worse in areas with extreme poverty, poor sanitation, increased freshwater irrigation usage, and with inadequacy or total lack of public health facilities.

The treatment for schistosomiasis is simple, due to the availability of fast action drugs prescribed in one single oral dose (Katz et al. 1989). However, the disease prevalence remains unchanged in endemic regions and it is expanding mainly in the urban centers periphery (Neves 2005).

The extensive distribution of the intermediate hosts, from *Biomphalaria* species, in Minas Gerais, Brazil gives to schistosomiasis an expansive characteristic, even for indene areas (Katz & Carvalho 1983, Carvalho et al. 1988, 1989). In the endemic areas, the high concentration of the hosts, associated to other risk factors, favors the existence of communities with high prevalence of schistosomiasis. The distribution of the schistosomiasis in the state of Minas Gerais is not regular, intercalating areas of high prevalence with others where the transmission is low or null. The disease is endemic in the north (comprising the Médio São Francisco and Itacambira zones), oriental, and center regions (zones of Alto Jequitinhonha, Metalúrgica, Oeste, and Alto São Francisco). The higher infection indices are found in the northeast and in the east part of the state, which includes the zones of Mucuri, Rio Doce, and Mata (Pellon & Teixeira 1950, Katz et al. 1978, Carvalho et al. 1987, Lambertucci et al. 1987).

Since schistosomiasis is a disease determined in space and time by risk factors, the Geographical Information System (GIS) is a very powerful tool that might be used for better understanding the disease prevalence and risk factors distributions. The use of GIS in the identification of environmental characteristics allows the determination and the delimitation of factors and risk areas, leading to the optimization of the resources and to the choice of better strategies for controlling the disease. The prediction of schistosomiasis using GIS was first attempted in the Philippines and the Caribbean by Cross et al. (1984). The influence of climate and environmental variables to the distribution of schistosomiasis was documented by Brown (1994) and Appleton (1978). The use of GIS for the study of schistosomiasis was also done in several other countries: in Asia (Cross et al. 1996), China (Zhou et al. 2001, Seto et al. 2002, Yang et al. 2005), Ethiopia (Kristensen et al. 2001, Malone et al. 2001), Egypt (Malone et al. 1994, 1997, Abdel-Rahman et al. 2001), Uganda (Kabatereine et al. 2004), Tanzania (Brooker et al. 2001), Chad (Beasley et al. 2002, Brooker et al. 2002). In Brazil, one of the first studies trying to correlate disease distribution with environmental variables was conducted by Bavia et al. (2001), in Bahia.

The objective of this paper is to provide a risk map for the state of Minas Gerais, by establishing a relationship between prevalence of schistosomiasis and social-environmental variables through multiple linear regression and

Financial support: Inpe, CNPq (309922/2003-8; 305546/2003-1; 380203/2004-9), Fapemig (EDT 1775/03; EDT 61775/03; CRA 0070/04)
+Corresponding author: corina@dpi.inpe.br
Received 25 May 2006
Accepted 26 June 2006

geographical information system. Also, we extend the methodology in several directions by introducing new explicative variables: climate variables and categorical data from standard biomes in Minas Gerais.

MATERIALS AND METHODS

Variables acquisition - The geo determination quality of a spatial model depends on how accurate the spatial positioning of involved variables is. Ideally it is desirable that each disease occurrence would be geo located by Geographic coordinates. Prevalence data, however, has been acquired for several years- even before the GPS technology became commonly available- being associated with the municipality center or some countryside village. Because of this limitation it was decided to integrate the environmental variables over the municipality territory to be used as an input to the modeling process. This is also coherent with the fact that the sociological indexes, also used here, are associated with the municipalities.

The schistosomiasis prevalence data (dependent variable) for 189 municipalities in Minas Gerais were obtained from Brazilian Health Ministry and from Health Secretariat of Minas Gerais State Annual Reports. The prevalence spatial distribution for Minas Gerais can be observed in Fig. 1.

Fourteen quantitative independent variables were used in the statistical analysis: three climatic variables (total precipitation, minimum and maximum temperature) in summer (from 17/Jan/2002 to 01/Feb/2002 period) and in winter (from 28/Jul/2002 to 12/Aug/2002 period) seasons, and four social variables [human development index (HDI), income, longevity and education indices] for the years of 1991 and 2000. Besides quantitative variables, two qualitative variables (binary) were also used, to represent three vegetation types: savanna, caatinga, and forest. The climatic variables were obtained from CPTEC/Inpe, the sociological ones from Brazilian Human Developing Atlas, and the qualitative variable from Geominas Project (Prodemge 1996). All independent and dependent data were given in a municipality base, with exception of the climatic data, which were given in a grid of 250 $\times$ 250 m. These data were, however, integrated over the municipality territory as explained.

Variables selection - The independent variables were used as input variables to establish the multiple regression model for prevalence risk. Since multicollinearity effects among the independent variables were detected, variables selection techniques were used in order to choose a set of variables (or transformations of them) that better explain the dependent variable.

A logarithmic transformation for the dependent variable (prevalence, denoted by *PREV*) was made as it improved the correlation with independent variables.

The variables selection was done by the R^2 criterion, using all possible regression procedure (Neter et al. 1996). This selection technique consists in the identification of a subset with few variables and a coefficient of determination R^2 sufficiently closed to that when all variables are used in the model. Interaction effects were also included in the model.

After choosing the model, the estimated regression function was applied to all municipalities to build a risk map for schistosomiasis prevalence.

RESULTS

To illustrate the variables selection technique, Fig. 2 contains a plot of the highest R^2 values against the number of variables in the model. In this figure the chosen model is highlighted in red, with $R^2 = 0.3774$, consisting of a model with five variables: summer precipitation (PC_s), summer minimum temperature (TN_s), 1991 Human Development Index (HDI_{91}) and two binary variables representing the vegetation types:

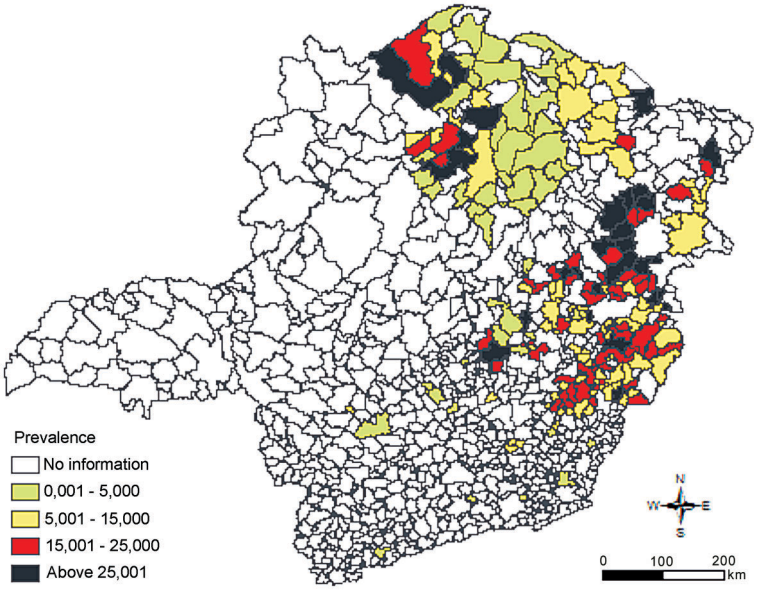


Fig. 1: schistosomiasis prevalence in Minas Gerais.

$$V_1 = \begin{cases} 1 & \text{if vegetation is savanna} \\ 0 & \text{otherwise} \end{cases}$$

and

$$V_2 = \begin{cases} 1 & \text{if vegetation is forest} \\ 0 & \text{otherwise} \end{cases}$$

The analysis of this model showed that the regression coefficient of the variable PC_s was not statistically significant at 5% level. Therefore, this variable was also discarded from the model. The model with four variables (TN_s, HDI_{91}, V_1, and V_2) has a coefficient of determination of 0.3569, and it is highlighted in green in Fig. 2.

After choosing the variables, the significance of several cross-product interaction effects were tested. The final selected model, with $R^2 = 0.3631$, consisted of the aforementioned variables and the interaction between HDI_{91}, and V_1, showing that the influence of HDI for the explanation of prevalence is different for savanna when compared to forest and caatinga.

The general estimated regression function is:

$$PREV = e^{(-0{,}2012 + 0{,}2113\ TN_s - 4{,}4012\ HDI_{91} + 1{,}1538\ V_2 + 1{,}1680\ HDI_{91} V_1)} - 1$$

This model for each vegetation type can be written as:

- Forest => $PREV = e^{(0{,}9526 + 0{,}2113\ TN_s - 4{,}4012\ HDI_{91})} - 1$
- Savanna => $PREV = e^{(-0{,}2012 + 0{,}2113\ TN_s - 3{,}2332\ \mathrm{HDI}_{91})} - 1$
- Caatinga => $PREV = e^{(-0{,}2012 + 0{,}2113\ TN_s - 4{,}4012\ HDI_{91})} - 1$

Fig. 3a shows the estimated prevalence and Fig. 3b its corresponding estimated standard deviation, for all municipalities of the state of Minas Gerais.

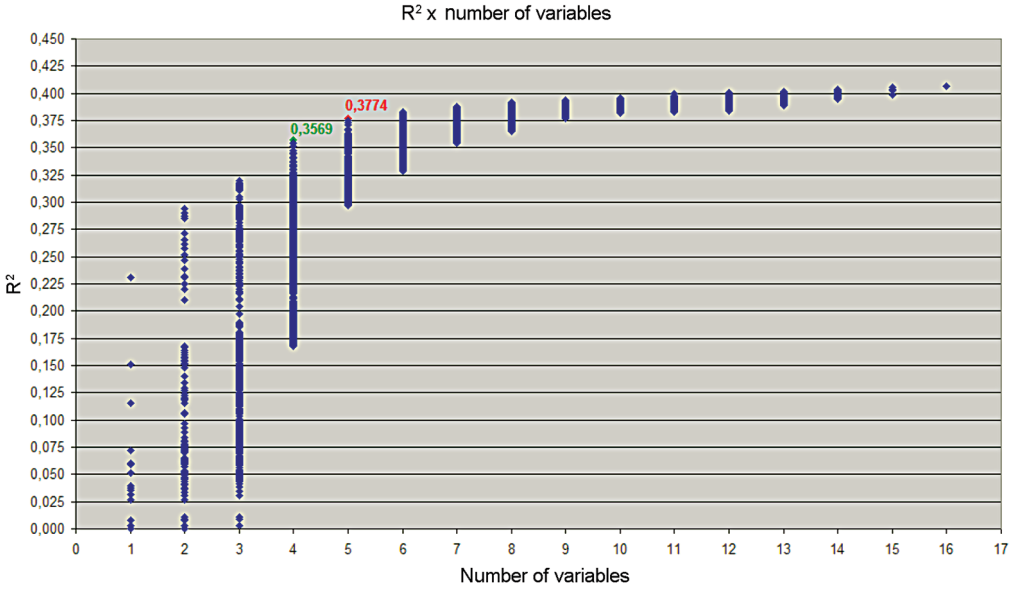


Fig. 2: R² x number of variables

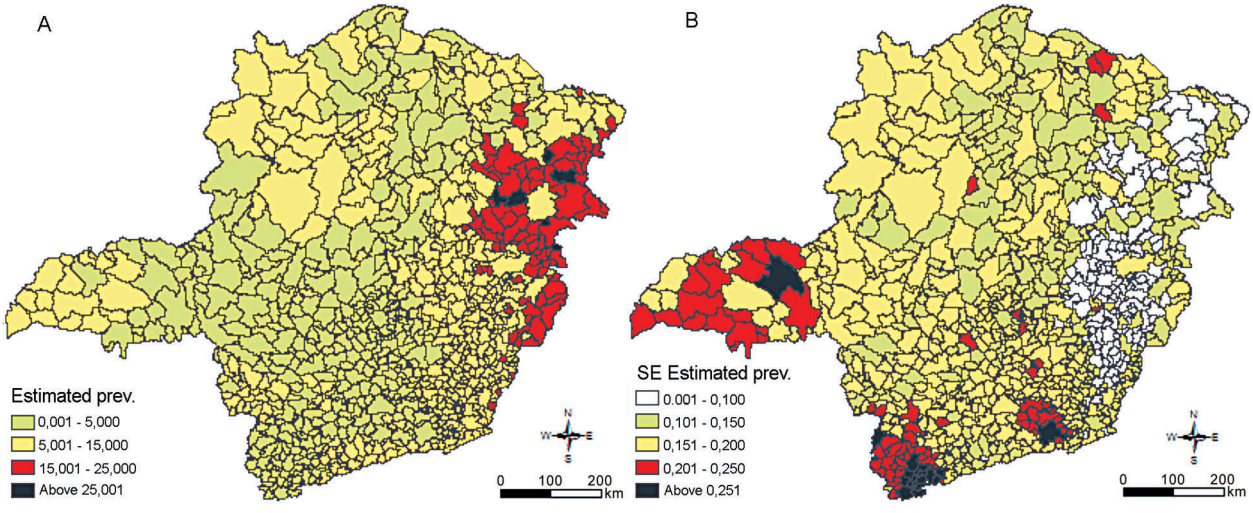


Fig. 3A: estimated schistosomiasis prevalence; B: corresponding estimated standard deviation for the state of Minas Gerais.

Fig. 4 shows the plot of the residuals, resulting from the difference between observed (Fig. 1) and estimated (Fig. 3A) schistosomiasis prevalence. In this Figure, dark colors (red and blue) represent overestimated values, light colors (cyan and magenta) underestimated ones, and in white are the municipalities with good estimative.

DISCUSSION

The spatial distributions of the observed prevalence and of the selected variables for the regression model are illustrated in Fig. 5. It can be seen that summer minimum temperature, 1991 human development index and vegetation types are most related with schistosomiasis prevalence. During the summer season the risk of contracting schistosomiasis increases, due to high concentrations of the snails in the drainage caused mainly by lack of sanitation, small amount of rain, high temperature, among other factors, and by the population searching for water bodies, either for drinking or as a form of minimizing the warmth.

The analyses of the estimated regression function and

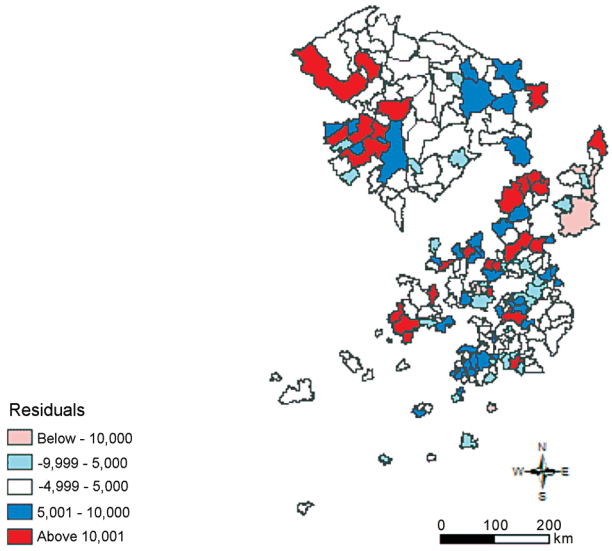


Fig. 4: residuals.

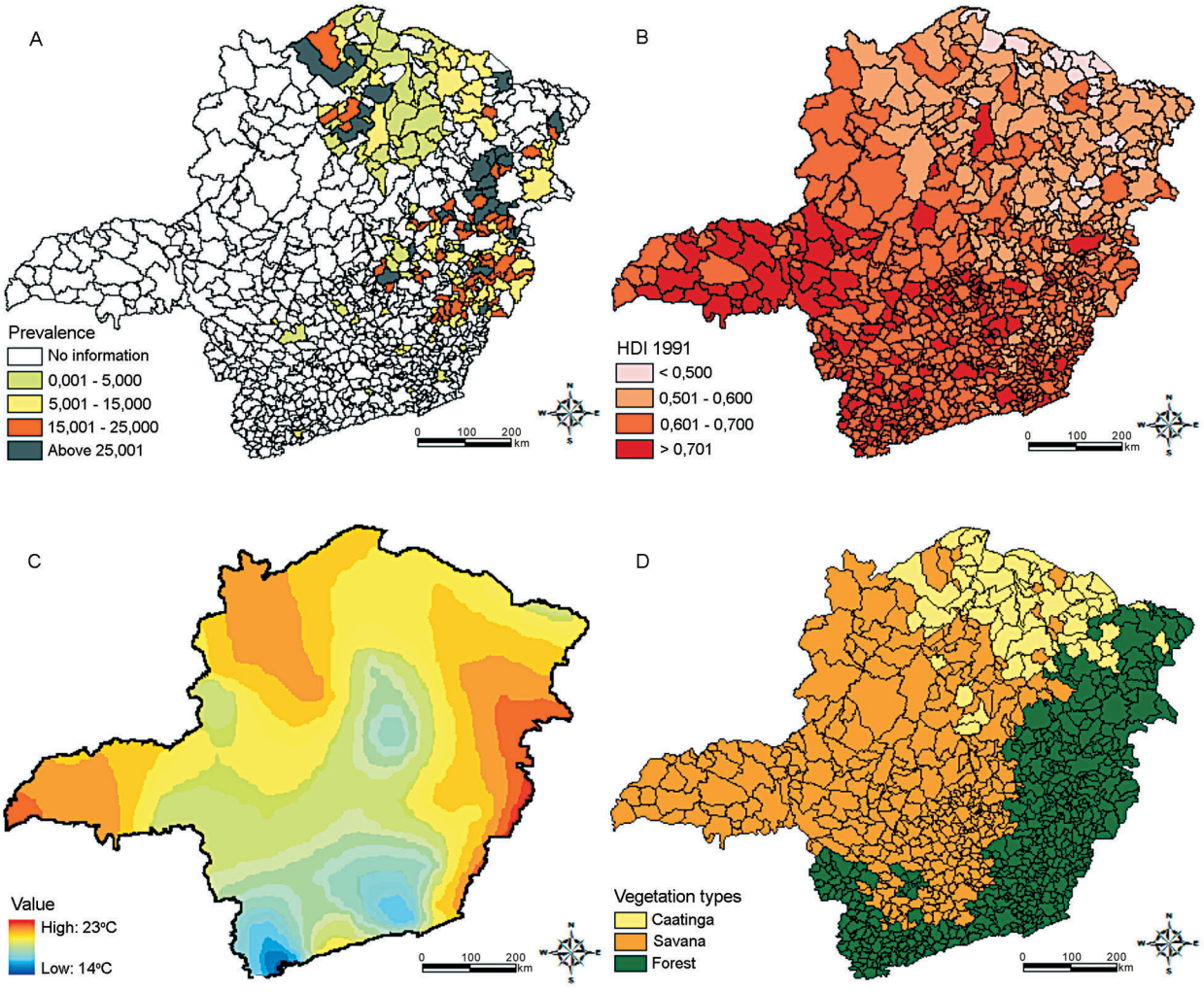


Fig. 5A: schistosomiasis prevalence; B: 1991 HDI; C: summer minimum temperature; D: vegetation type.

Figs 3, 4, and 5 shows that: (i) summer minimum temperature has positive correlation with schistosomiasis prevalence, while human development index has negative correlation; (ii) the effect of summer minimum temperature over schistosomiasis prevalence is the same for all vegetation types; (iii) the effect of 1991 human development index is lower for savanna than for forest and caatinga biomes; (iv) for fixed values of summer minimum temperature and 1991 human development index, the regression models for caatinga and forest differ approximately by a factor, being the schistosomiasis in forest regions about three times higher than that for caatinga; (v) the precision (standard deviation) of the estimative are quite good for the municipalities where the prevalence data were available, but there is a tendency of decreasing when the estimative are made for far away municipalities.

Therefore, even with a low coefficient of determination, it might be concluded that the joint use of geographical information systems and statistical techniques allows the determination of related factors and the delimitation of risk areas for schistosomiasis.

Several other variables related to the use of the water, such as sanitation, water quality, water retention by the soil, existence of intermediate hosts, and remote sensing variables, might be tested as explanatory variables to improve the model.

REFERENCES

Abdel-Rahman MS, El-Bahy MM, Malone JB, Thompson RA, El Bahy NM 2001. Geographic information systems as a tool for control program management for schistosomiasis in Egypt. *Acta Trop 79*: 49-57.

Appleton CC 1978. Review of literature on abiotic factors influencing the distribution and life cycles of bilharziasis intermediate host snails. *Malacol Rev 11*: 1-25.

Bavia ME, Malone JB, Hale L, Dantas A, Marroni L, Reis R 2001. Use of thermal and vegetation index data from earth observing satellites to evaluate the risk of schistosomiasis in Bahia, Brazil. *Acta Trop 79*: 79-85.

Beasley M, Brooker S, Ndinaromtan M, Madjiouroum EM, Baboguel M, Djenguinabe E, Bundy DAP 2002. First nationwide survey of the health of schoolchildren in Chad. *Trop Med Int Health 7*: 625-630.

Brooker S, Beasley M, Ndinaromtan M, Madjiouroum EM, Baboguel M, Djenguinabe E, Hay SI, Bundy DAP 2002. Use of remote sensing and a geographic information system in a national helminth control programme in Chad. *Bull 80*: 783-789.

Brooker S, Hay SI, Issae W, Hall A, Kihamia CM, Lwambo NJ, Wint W, Rogers DJ, Bundy DA 2001. Predicting the distribution of urinary schistosomiasis in Tanzania using satellite sensor data. *Trop Med Int Health 6*: 998-1007.

Brown DS 1994. *Freshwater Snails of Africa and their Medical Importance*, 2nd. ed., Taylor & Francis Ltd, London, 609 pp.

Carvalho OS, Massara CL, Rocha RS, Katz N 1989. Esquistossomose mansoni no sudoeste do Estado de Minas Gerais (Brasil). *Rev Saúde Públ São Paulo 23*: 341-344.

Carvalho OS, Rocha RS, Massara CL, Katz N 1987. Expansão da esquistossomose mansoni em Minas Gerais. *Mem Inst Oswaldo Cruz 82*: 295-298.

Carvalho OS, Rocha RS, Massara CL, Katz N 1988. Primeiros casos autóctones de esquistossomose mansonica em região do noroeste do Estado de Minas Gerais (Brasil). *Rev Saúde Públ São Paulo 22*: 237-239.

Cross ER, Newcomb WW, Tucker CJ 1996. Use of weather data and remote sensing to predict the geographic and seasonal distribution of *Phlebotomus papatasi* in Southwest Asia. *Am J Trop Med Hyg 54:* 530-536.

Cross ER, Perrine R, Sheffield C, Pazzaglia LTG 1984. Predicting areas endemic for schistosomiasis using weather variables and a Landsat data base. *Military Medicine 149*: 542-544.

Kabatereine NB, Brooker S, Tukahebwa EM, Kazibwe F, Onapa AW 2004. Epidemiology and geography of *Schistosoma mansoni* in Uganda: implications for planning control. *Trop Med Int Health 9*: 372-380.

Katz N, Carvalho OS 1983. Introdução recente da esquistossomose mansoni no sul do estado de Minas Gerais, Brasil. *Mem Inst Oswaldo Cruz* 78: 281-284.

Katz N, Dias EP, Souza CP, Bruce JI, Coles GC 1989. Rate of action of schistosomicides in mice infected with *Schistosoma mansoni*. *Rev Soc Bras Med Trop 22*: 183-186.

Katz N, Motta E, Oliveira UB, Carvalho EF 1978. Prevalência da esquistossomose em escolares do estado de Minas Gerais. Resumos do XVI Congresso da Sociedade Brasileira de Medicina Tropical, João Pessoa, p. 102.

Kristensen TK, Malone JB, McCarroll JC 2001. Use of satellite remote sensing and geographic information systems to model the distribution and abundance of snail intermediate hosts in Africa: a preliminary model for *Biomphalaria pfeifferi* in Ethiopia. *Acta Trop 79*: 73-78.

Lambertucci JR, Rocha RS, Carvalho OS, Katz N 1987. A esquistossomose mansoni em Minas Gerais. *Rev Soc Bras Med Trop 20*: 47-52.

Malone JB, Abdel-Rahman MS, El Bahy MM, Huh OK, Shafik M, Bavia M 1997. Geographic information systems and the distribution of Schistosoma mansoni in the Nile delta. *Parasitol Today 13*: 112-119.

Malone JB, Huh OK, Fehler DP, Wilson PA, Wilensky DE, Holmes RA, Elmagdoub AL 1994. Temperature data from satellite imagery and distribution of schistosomiasis in Egypt. *Am J Trop Med Hyg 51*: 714-722.

Malone JB, Yilma JM, McCarroll JC, Erko B, Mukaratirwa S, Zhou X 2001. Satellite climatology and the environmental risk of Schistosoma mansoni in Ethiopia and east Africa. *Acta Trop 79*: 59-72.

Neter J, Kutner MH, Nachtssheim CJ, Wasserman W 1996. *Applied Linear Statistical Models,* 4 ed., WCB/McGraw-Hill, Boston.

Neves DP 2005. *Parasitologia Humana,* 10th ed., Atheneu, São Paulo.

Pellon B, Teixeira I 1950. *Distribuição Geográfica da Esquistossomose Mansônica no Brasil*, Divisão de Orga-nização Sanitária, Ministério da Saúde, Rio de Janeiro, 24 pp.

Prodemge 1996. Geominas. Prodemge. Disponível em: http://www.geominas.mg.gov.br. Acesso em: 14 de fev. de 2006.

Seto E, Xu B, Liang S, Gong P, Wu W, Davis GM, Qui D, Gu X, Spear R 2002. The use of remote sensing for predictive modeling of schistosomiasis in China. *Photogram Eng Rem S 68*: 167-174.

WHO 1999. Schistosomiasis. W.H.O. Disponível em: http://www.tulane.edu/~dmsander/WWW/224/Schisto.html. Acess: 31 Jan. 2006.

Yang G-J, Vounatsou P, Xiao-Nong Z, Utzinger J, Tanner M 2005. A review of geographic information system and remote sensing with applications to the epidemiology and control of schistosomiasis in China. *Acta Trop 96*: 117-129.

Zhou XN, Malone JB, Kristensen TK, Bergquist NR 2001. Application of geographic information systems and remote sensing to schistosomiasis control in China. *Acta Trop 79*: 97-106.