CSP CADERNOS DE SAÚDE PÚBLICA
REPORTS IN PUBLIC HEALTH

ARTIGO
ARTICLE

Spatio-temporal GAMLSS modeling of the incidence of schistosomiasis in the central region of the State of Minas Gerais, Brazil

Modelagem GAMLSS espaçotemporal da incidência de esquistossomose na região central do Estado de Minas Gerais, Brasil

Modelado espaciotemporal GAMLSS para la incidencia de esquistosomiasis en la región central del estado de Minas Gerais, Brasil

Denismar Alves Nogueira [1]
Thelma Sáfadi [2]
Renato Ribeiro de Lima [2]
Angélica Sousa da Mata [2]
Miriam Monteiro de Castro Graciano [2]
Joziana Muniz de Paiva Barçante [2]
Thales Augusto Barçante [2]
Stela Márcia Pereira Dourado [2]

doi: 10.1590/0102-311XEN068822

Abstract

In Brazil, millions of people live in areas with risk of schistosomiasis, a neglected chronic disease with high morbidity. The Schistosoma mansoni *helminth is present in all macroregions of Brazil, including the State of Minas Gerais, one of the most endemic states. For this reason, the identification of potential foci is essential to support educational and prophylactic public policies to control this disease. This study aims to model schistosomiasis data based on spatial and temporal aspects and assess the importance of some exogenous socioeconomic variables and the presence of the main* Biomphalaria *species. Considering that, when working with incident cases, a discrete count variable requires an appropriate modeling, the GAMLSS modeling was chosen since it jointly considers a more appropriate distribution for the response variable due to zero inflation and spatial heteroscedasticity. Several municipalities presented high incidence values from 2010 to 2012, and a downward trend was observed until 2020. We also noticed that the distribution of incidence behaves differently in space and time. Municipalities with dams presented risk 2.25 times higher than municipalities without dams. The presence of* B. glabrata *was associated with the risk of schistosomiasis. On the other hand, the presence of* B. straminea *represented a lower risk of the disease. Thus, the control and monitoring of* B. glabrata *snails is essential to control and eliminate schistosomiasis; and the GAMLSS model was effective in the treatment and modeling of spatio-temporal data.*

Residence Characteristics; Biomphalaria*; Regression Analysis; Secondary Data Analysis*

Correspondence
D. A. Nogueira
Universidade Federal de Alfenas.
Rua Gabriel Monteiro da Silva 700, Alfenas, MG 37130-000, Brasil.
denismar.nogueira@unifal-mg.edu.br

[1] *Universidade Federal de Alfenas, Alfenas, Brasil.*
[2] *Universidade Federal de Lavras, Lavras, Brasil.*

Introduction

Schistosomiasis is a serious and neglected tropical disease caused by parasites of the genus *Schistosoma* and affects countries worldwide, leading to the death of up to 280,000 people per year [1]. According to the World Health Organization (WHO) [2], in 2019, more than 236 million people were affected by this disease. In Brazil, it is estimated that more than 25 million people live in risk areas; the parasite is present in all macro-regions of the country, with the State of Minas Gerais, in Southeast Region, being one of the most endemic states.

Several studies have reported that the infection occurs during agricultural and recreational activities and due to exposure to water contaminated with cercaria and, therefore, associated with the peripheries of cities where the lack of infrastructure and environments without sewage treatment is more common [3,4]. In Brazil, the endemic form is caused by the helminth *S. mansoni*, responsible for the hepatosplenic form of schistosomiasis, although there are reports of some cases from other species [5]. According to Souza et al. [6], 523 municipalities in Minas Gerais, or 61% of the total number of municipalities in the state, are considered endemic due to their high infection rates and because they contain a wide geographic distribution of snail species (*Biomphalaria* spp.), which is an intermediate host of the parasite. According to Massara et al. [7], the species *B. glabrata*, *B. tenagophila*, and *B. straminea*, which may or may not be infected with *S. mansoni*, are commonly found in Minas Gerais. Some disease control programs have been successfully implemented, and improvements in the health system with mass drug treatment have decreased the number of cases; nevertheless, the disease persists, maintaining its global relevance [8]. Moreover, according to the WHO, drug treatment in some regions has not shown satisfactory results, and a possible control of the intermediate host may be an important measure in the control of the disease [9].

Currently, control depends on municipal public policies [10]. Over 500 deaths have been verified in Brazil in the last 15 years. According to Silva da Paz [5] and Simões et al. [8], many cases have occurred in older adults (> 60 years), increasing the risk associated with chronic non-infectious diseases.

The State of Minas Gerais has been considered endemic for some years and studies, such as the one by Cardoso et al. [11], which evaluated the spatial and temporal aspect of deaths due to schistosomiasis, highlight the Vale do Aço, central, and northeastern regions of the state as the most worrisome. The identification of potential foci is essential for the planning of educational and prophylactic public policies for the control of schistosomiasis, especially in historically endemic regions, since this disease causes relevant impacts to those infected and, if untreated, can result in substantial morbidity or death [12,13].

Thus, the monitoring of schistosomiasis incidence and its associated variables is extremely important. Different forms of analysis and modeling have been used to approach this topic; the spatio-temporal analysis methodologies, however, have gained strength in recent years for their adequacy and efficiency in evaluating the effects of the disease considering space and time. Paz et al. [14] used this type of approach to evaluate the behavior of schistosomiasis in the Northeast Region of Brazil, enabling a better understanding of the problem and supporting public health decisions.

Notably, working with incidental cases, which is a discrete variable, requires a modeling of generalized methods since it allows for a more adequate distribution to the data. Unlike the generalized linear models (GLM) [15] and generalized additive models (GAM) [16], the generalized additive models for location, scale, and shape (GAMLSS) [17] are more flexible since they allow for the use of a wide variety of probability distributions for the dependent variable. This flexibility allows the modeling of the different distribution parameters, such as mean, variance, asymmetry, and kurtosis. Thus, one can associate different linear predictors to the different parameters of a given distribution, considering different binding functions. The terms of the linear predictors may contain parametric and smoothing (nonparametric) functions and may be of fixed or random effects. With the use of GAMLSS modeling we are able to assess not only how the mean of the dependent variable is influenced by the explanatory variables, but also how the variance and the other shape parameters of the distribution of the random (dependent) variable are influenced by the same or different explanatory variables.

This study aims to model schistosomiasis data in relation to spatial and temporal aspects, in addition to evaluating the risk and importance of some socioeconomic exogenous variables considering the presence of the main species of *Biomphalaria* in the central region of the State of Minas Gerais.

Methodology

This ecological study uses secondary data from the public domain of the Brazilian Health Informatics Departments (DATASUS) regarding the occurrence of schistosomiasis in the State of Minas Gerais, from the Brazilian Information System for Notifiable Diseases (SINAN). The area in focus refers to the residents in the central region of the state, comprising the municipalities that border the Paraopeba River and its primary and secondary neighboring municipalities. The outcome variable refers to the sum of cases in the 103 municipalities, from 2007 to 2020, available in the SINAN database. The population of the region is of 6,182,086 inhabitants, according to the 2010 census conducted by the Brazilian Institute of Geography and Statistics (IBGE) [18]. The exogenous variables include the M-HDI (Municipal Human Development Index) [18]; flood events in the reports of emergency situation (Brazilian National Water Resources Information System – SNIRH of the Brazilian National Water and Sanitation Agency – ANA) [19]; percentage of the population who does not have treated water and sewage (ANA) [19]; percentage of the population with inadequate sanitation; rate of illiteracy among the population; percentage of the population with open sewage; percentage of the population living on 1/4 of the minimum wage per capita; rate of households without a bathroom (2010 IBGE census) [18]; proximity to the Paraopeba River (municipalities that are in direct contact with the banks of the river; second-order and third-order neighbor municipalities); presence of mining residue dams in the municipality (Brazilian National Mining Agency – ANM) [20]; presence of snails from the three intermediate host species, according to the report issued by the technical advisory committee of the schistosomiasis program of the Brazilian Ministry of Health (2008) [21]; and number of inhabitants (offset). Other variables such as year, municipalities, and first and second order lags (Y_{i-1} for one year before and Y_{i-2} for two years before within each municipality) were also considered in the modeling to assess spatial and temporal dependence. The study uses secondary sources in the public domain and therefore respects ethical principles.

We chose to use GAMLSS modeling, as proposed by Rigby & Stasinopoulos [17]. This methodology allows working with different probability distributions (not just of the exponential family), or with a mixture thereof, in the adjustment of regression and modeling of space and time. According to Rigby & Stasinopoulos [17], the GAMLSS regression for n independent observations, which is not the case of our study, are structured in the random effects and in the covariance matrix, leading to the understanding that the joint distribution will be conditional to these past values. It is normal to assume that the distribution or density function can present four (k = 1,…, 4; or more) parameters – location (μ), scale (σ), form (ν and τ) – when writing the bond function $g_k(.)$ as a known monotone, relating $\theta_k = (\theta_{k1}, ..., \theta_{kn})$ to the predictor variables and random effects through:

$$g_k(\theta_k) = X_k\beta_k + \sum_{j=1}^{J_k} h_{jk}(x_{jk})$$

in which X_k is an array of covariates, β_k is the vector with the parameters associated with the covariates of each of the k parameters; h_{jk} is a nonparametric smoothing function applied to some of the continuous exogenous variables, considering that J_k are the functions applied to each of the parameters. The additive part, on the right side of the equation, can be replaced by $\Sigma Z_{jk}\gamma_{jk}$ (or added) when the random effects are included in the model, in which $\gamma_{jk} \sim N_{qjk}(0, \lambda_{jk}^{-1} G_{jk}^{-1})$ with G_{jk}^{-1} is the inverse (generalized) of the symmetric matrix (q_{ij} x q_{ij}), considering that qij is the number of random variables, which in this study represents the number of municipalities or areas. According to De Bastiani et al. [22], random effects can be characterized as a Gaussian Markov random field (GMRF) and, therefore, it is possible to assume the matrix G as a weight or neighborhood structure in spatial modeling, thus obtaining the IAR (intrinsic autoregressive) models, coming from the CAR (conditional autoregressive) models, since G is singular.

The estimation of the parameters of the model is given by penalized maximum likelihood, in which the β, γ are estimated. The λ (hyperparameters present in the likelihood penalty function) can be fixed or estimated by the penalized quasi-likelihood. Stasinopoulos et al. [23] detail the processes and algorithms for estimation. For the analysis, we used the family of GAMLSS R packages (http://www.r-project.org) and others such as *mgcv* [24] and *spdep* [25].

To define the variables that make up the model, as recommended by De Bastiani et al. [22], the stepwise procedure and the generalized Akaike information criterion (GAIC) were considered, which can be reduced to both criteria Akaike information criterion (AIC) [26] (when the penalty is equal to 2) and bayesian information criterion (BIC) [27] (when the penalty is ln(n)). The selection of the best model was performed after assessing the quality of adjustment by means of normalized randomized quantile residuals [28], evaluated by Q [29] statistics and with the confirmation of independence with the use of correlograms.

To evaluate the distribution that best adheres to the data, a study was conducted with several discrete distributions that could encompass the excess of zeros and the long tail. Distributions such as zero-inflated poisson (ZIP), zero-inflated negative binomial (ZINBI), poisson inverse gaussian (PIG), and Sichel distribution were performed in addition to the variations of these distributions adjusted or altered to model the excess of zeros. A total of 16 distributions were evaluated and the one that best adhered to it was chosen, according to the GAIC criterion. Further details on the distributions can be found in Rigby et al. [30].

To avoid collinearity problems, exogenous variables that presented a correlation higher than |0.6| were eliminated; variables such as illiteracy rate, percentage of the population with open sewage, percentage of the population living with 1/4 of the minimum wage per capita, use of cesspools were represented by the M-HDI. The variables proximity (1 – closer, 2, and 3), presence of dam (0.1), and presence of the snail of the three species (*B. glabrata*, *B. straminea*, and *B. tenagophila*) (0.1) were considered as factors. As offset, we used the logarithm of the total population of each municipality and considered it constant in the study period.

Map and region of interest

With the development of digital map creation technologies and the emergence of geoprocessing, it is possible to collect, display, and process georeferenced information. Statistical methodologies have considered this geographic information system (GIS), incorporating the location of the observation into the analyses. In this study, the occurrence of cases was referred to the municipalities of residence. For the construction of the maps, the cartographic base of Minas Gerais was used through the IBGE website (https://portaldemapas.ibge.gov.br/portal.php#mapa222139). The maps were built in ArcMap v.10.2.2 (https://www.esri.com) and analyzed in the R programming language. GIS allows the database to be connected to geographic features and to the construction of the neighborhood matrix G.

For the spatial analysis, a central region of the State of Minas Gerais, was selected, which includes 103 municipalities around the Paraopeba River basin. Of these, 20 are in direct contact with the bank of the river (proximity 1), 40 are at proximity 2, and 43 at proximity 3. Figure 1 illustrates the studied region, highlighting the municipalities.

The region under study includes municipalities with populations of varying sizes, including the capital of Minas Gerais. The climate of the region is of the subhumid temperate type (the capital is close to 853 meters from sea level). The classification according to Köppen for the region is tropical savannah, bordering humid temperate. In the northern part of the region, the prevailing classification is Aw, and the southern part is Cwa [31].

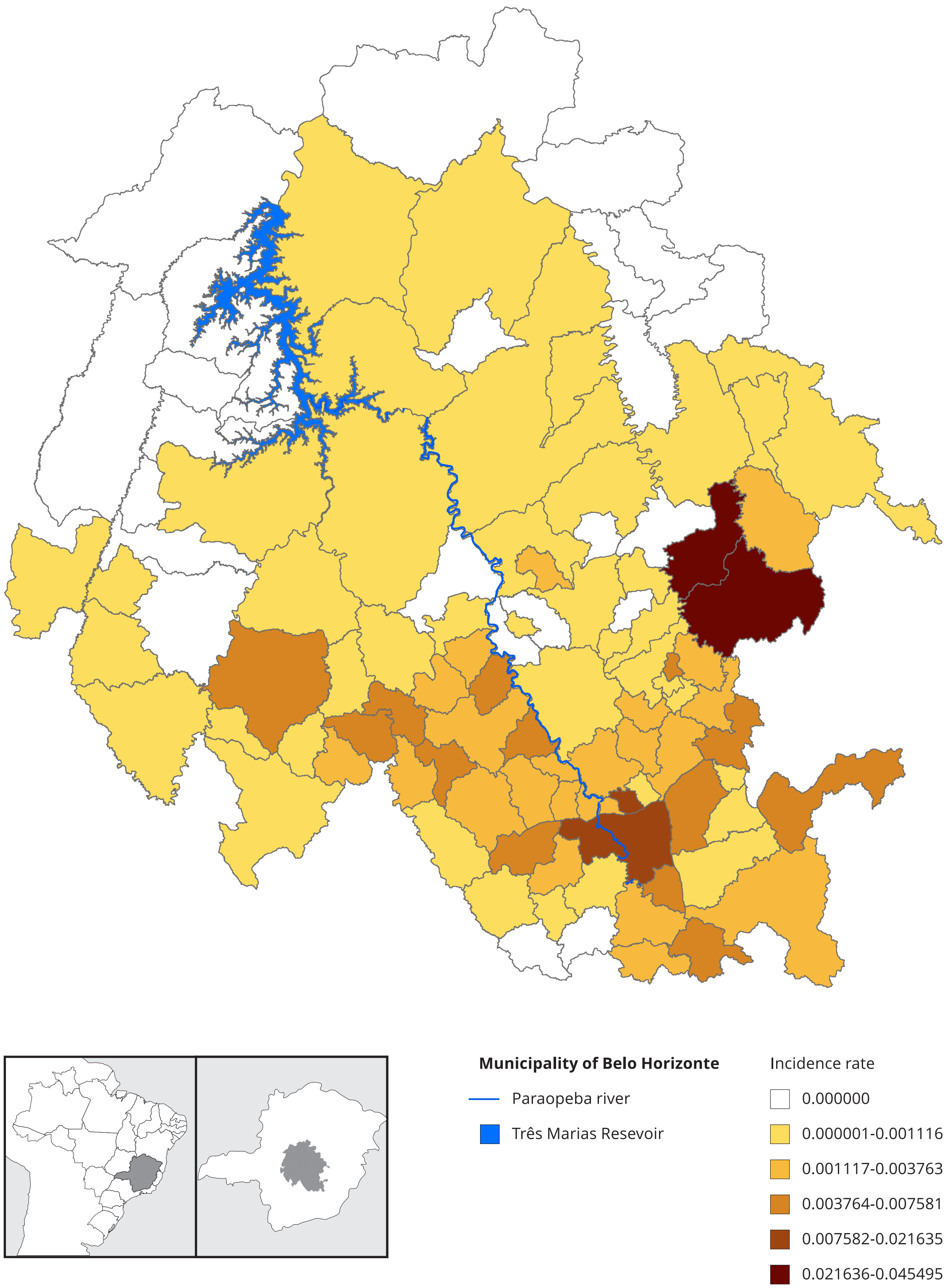
Results and discussion

Periodically, annual data were observed, from 2007 to 2020, in each of the 103 municipalities investigated. The average rates observed per 100,000 inhabitants (total number of cases) between the years 2007 and 2020 were: 7.134 (458), 24.637 (1,430), 44.15 (1,730), 69.478 (2,526), 43.727 (1,649), 12.384 (595), 9.849 (659), 6.386 (574), 7.535 (594), 5.747 (580), 7.981 (507), 7.995 (584), 10.123 (491), and 3.11 (265), respectively. In this period, the total number of cases was 12,642, with an annual average of 903 cases and an average annual rate of 18.588 per 100,000 inhabitants. The cumulative incidence in the period was 204.494 per 100,000 inhabitants.

Generally, high rates were verified in 2010, 2011, and 2012, followed by a downward trend and stabilization, from 2013 to 2019, and a reduction, in 2020. Such findings may be related to a reduction

Figure 1

Studied region with the cumulative incidence rate of schistosomiasis, in 103 municipalities in the Minas Gerais State, Brazil, 2007-2020.



Source: elaborated by the authors.

in the number of tests performed over time. Additionally, the implementation of control and treatment policies may have led to a decrease in positive cases observed over time [32]. It has been verified that the Southeast Region of the country has been presenting a decrease in cases since 2013, mainly in the State of Minas Gerais, despite it still being the state with the highest number of absolute cases and incidence. This is not the case with the Northeast Region, which showed an increase in the number of cases over this period. Thus, the possible decrease in rates may be due to the alteration of the infection dynamics, observing a different profile for the infected and adaptations of the host to new sites [33].

Figure 2 shows the spatio-temporal variation with decreased incidence and persistence of the focus in the southern region of the studied area. Although the incidence is decreasing, the endemic region persists.

When evaluating the exogenous variables, we observed a relationship that may be nonlinear with the parameters, with the presence of asymmetry in the distribution and heteroscedasticity of the dependent variable. An important fact to consider is the distribution of data for each municipality. In this case, the variable presents distinct destructions for each municipality, evidencing the need to model the scale and form of the distribution due to the presence of zeros in some municipalities and extreme values in others. These behaviors need to be modeled with appropriateness and with a methodology that can deal with these characteristics. In addition, the mean incidence can be influenced by spatial variation. Spatial correlation test such as Global Moran's Index and per year were evaluated; a significance ($p < 0.05$) was found for some of the years evaluated. Another statistic also evaluated was the scan for the presence of spatial and temporal clusters with the SatScan software tool (http://www.satscan.org) [34,35], being significant for at least 3 spatio-temporal clusters.

Moreover, it was verified that the distribution of incidence behaves differently in space and time, which makes it difficult to choose a methodology that can portray the problem. Several studies have sought to use methodologies that are able to model this type of data in the best way.

Authors such as Wood et al. [9] used models with mixed distributions to model the excess of zeros and overdispersion in the data for snail occurrence, in an attempt to determine the spatial-temporal behavior of the species. In the study by Scholte et al. [36], the Bayesian geostatistical model was used to predict the disease risk map.

Simões et al. [8] also use a methodology that considers spatial variation and Bayesian inference. The authors argue that this methodology allows for the use of a more appropriate distribution for likelihood and priors, with the incorporation of random effects and additional structures for time and space. These studies, despite their different objectives, used different methodologies to deal with the incidence of schistosomiasis. We chose to use GAMLSS models with spatial structure, modeling the temporal effect and seeking the best distribution to deal with the excess of zeros and asymmetry presented by the random variable of interest.

To adjust the number of cases of schistosomiasis, it was necessary to use a Sichel [37] mixture distribution to deal with the excess of zeros and long tail. The distribution was selected according to the criteria established by better adjusting the variable of interest. The use of this distribution enabled an adequate representation of the empirical distribution (Figure 3). It is possible to verify the complexity of the modeling due to the presence of more than 60% of the data being zero and a prominent tail on the right. The median of the cases is 0, the third quartile is 1, with a maximum of 477 cases, which characterizes a data set very concentrated in these values and an asymmetric associated distribution. To study the effect of time, the time series of average incidence per year was evaluated, totaling 14 years of registration. We were still able to verify, by the autocorrelation function, the presence of significant *lags*, which characterizes temporal dependence. The Ljung-Box test [38] was performed and confirmed significance for *lag* 1 ($p = 0.006$) and *lag* 2 ($p = 0.016$). This result justifies the presence of the variables of first and second order lags in the modeling (Supplementary Material: https://cadernos.ensp.fiocruz.br/static//arquivo/suppl-e00068822-eng_2815.pdf).

The Sichel distribution is a mixture of the Poisson distribution and the generalized inverse normal. This distribution presents three y~Sichel parameters: mean (μ), variability (σ), and form (υ). When σ tends to infinity, Sichel tends to the Poisson distribution (μ); and when σ tends to infinity and $\upsilon > 0$, it tends to negative binomial. Logarithmic link functions for μ and σ and identity for υ [30] were used for the analysis.

Figure 2

Studied region with the incidence of schistosomiasis per year, in 103 municipalities in the Minas Gerais State, Brazil, 2007-2020.

Source: elaborated by the authors.

After defining the distribution that would be used, the adjustment process began by using the model selection strategy presented by De Bastiani et al. [22], in which two strategic stages were made, adjustment and subsequent adaptation of the model using normalized randomized quantile residuals [28], worm-plot, and Q statistics [29], in addition to the evaluation of independence. The selection of variables to compose the final model was performed by stepwise GAIC.

Table 1 presents the final model with all significant variables at 5%. The exogenous variables that are not composing the model were removed since they did not present statistical significance, according to the stepwise criterion [39]. Table 1 shows the GMRF, which characterizes the spatial smoothing function or random function considering IAR model and weight matrix of the nearest neighbor type, as a GMRF. In the modeling, interactions between variables were not considered and the use of offset allowed us to work with incidence and, therefore, allowing for the estimation of the risk of occurrence of schistosomiasis in a unit of time/municipality.

Figure 3

Graphic representation of the frequency of occurrence of schistosomiasis and the adjusted Sichel distribution.

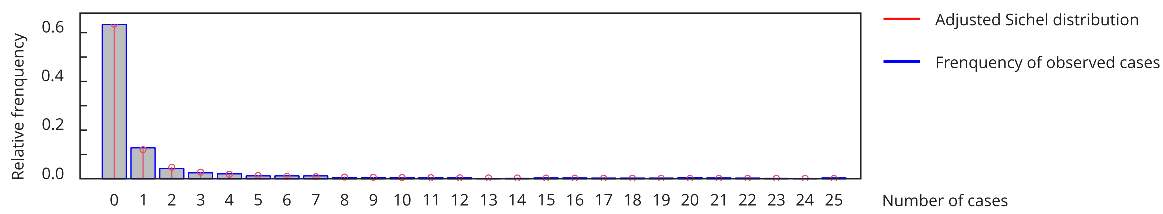


Table 1

GAMLSS model estimates with parametric and nonparametric functions and random spatial effect.

Parameter	Predictive variables	Estimates	Standard errors	Exponential of the estimate	p-value
μ	Intercept	-8.8295	0.1282	0.0001	< 0.001
	Y_{i-1}	0.0248	0.0011	1.0251	< 0.001
	$h_{11}(Y_{i-2})$	-			
	h_{21}(year)	-			
	GMRF(municipality)	-			
	B. glabrata = 1	0.4478	0.1333	1.5649	< 0.001
	B. straminea = 1	-0.9811	0.1102	0.3749	< 0.001
	Barragem = 1	0.8128	0.1063	2.2542	< 0.001
σ	Intercept	0.4228	0.1159	1.5262	< 0.001
	Y_{i-1}	0.0201	0.0018	1.0203	< 0.001
υ	Intercept	-1.0182	0.1055	0.3612	< 0.001
	H_{13} (Y_{i-1})	-			

GAMLSS: generalized additive models for location, scale and shape; GMRF: Gaussian Markov random field (spatial function); h: p-splines smoothing function. For the year, a cubic spline function with 9g.l was used. The offset (log(populationi)) was also used in the modeling.

The use of variables that depict the regions where the intermediate hosts were found indicated an important relationship for the follow-up of the disease, especially regarding the species *B. glabrata* and *B. straminea*. According to the estimated model, keeping the other variables constant, the presence of *B. glabrata* is related to higher risk, being 1.56 times higher, when compared to municipalities where this species was not found. For *B. straminea*, this represents an inverse relationship to incidence, and its presence has a 62.51% lower risk. It was expected that the presence of *B. glabrata* was actually associated with the occurrence of schistosomiasis [40], considering that this is the most important species in the transmission, and it is adapted to the region. The result for the presence of *B. straminea* was interesting. This species is found in several regions of the country because it adapts well to different types of climates [40], however, this species is responsible for higher risk of occurrence in the Northeast Region of the country. In Minas Gerais, *B. straminea* is less susceptible than *B. Glabrata* [10]. Of the area under study, 27.18% of the municipalities have snail species, of which 37.86% confirmed the presence of only *B. glabrata* and 33.98% of only *B. straminea*. The presence of *B. tenagophila* was verified only in

13.59% of the municipalities, and its statistical significance was not observed. This can be explained by the fact that this species is not associated with reports of importance in the transmission of the disease in Minas Gerais, being more present in the Southern Brazil and associated with the disease in the State of São Paulo [10].

The three species of *Biomphalaria* have been progressively found in new municipalities, which gives an expansive character to schistosomiasis, including in unaffected areas [40], despite positivity being verified in only 1% of the municipalities studied. In the municipality of Belo Horizonte, even after four decades, the three species continue to be found [7] in the parks, reinforcing the relationship with the disease. Oliveira et al. [41] showed that social inequality is a relevant factor in the incidence of the disease in the State of Minas Gerais. In our study, however, the variables that were associated with social factors, such as M-HDI, sanitation, low income, lack of infrastructure did not present significance. This result can be explained by the region under study, where there is no evident socio-economic differentiation. Another point that draws attention is that municipalities such as Belo Horizonte and Brumadinho, which presented high numbers of cases, showed satisfactory development indices; therefore, the use of these indicators do not faithfully represent the marginalized population, who are probably more exposed. Another hypothesis to be considered is that these two municipalities may have a more structured epidemiological surveillance system than the others, which could lead to an information bias due to an apparently higher number of cases, but which is, in fact, due to a better notification system and not an actual increase in cases.

Understanding the occurrence of incidence motivated the search for variables that could better characterize the problem. One hypothesis is the possibility of the high incidence in the south of the studied area being associated with the importation of cases, due, for example, to the migration of workers from other regions of Brazil, such as the State of São Paulo and the Northeast Region [33]. We also observed that a section of the studied area presents a concentration of dams with mining residue. This variable was treated as dichotomous and, according to the model, municipalities with the presence of a dam presented a risk 2.25 times higher than in municipalities without it. This result may indicate a possible migration of positive individuals in these regions due to a greater supply of work and more structured health services.

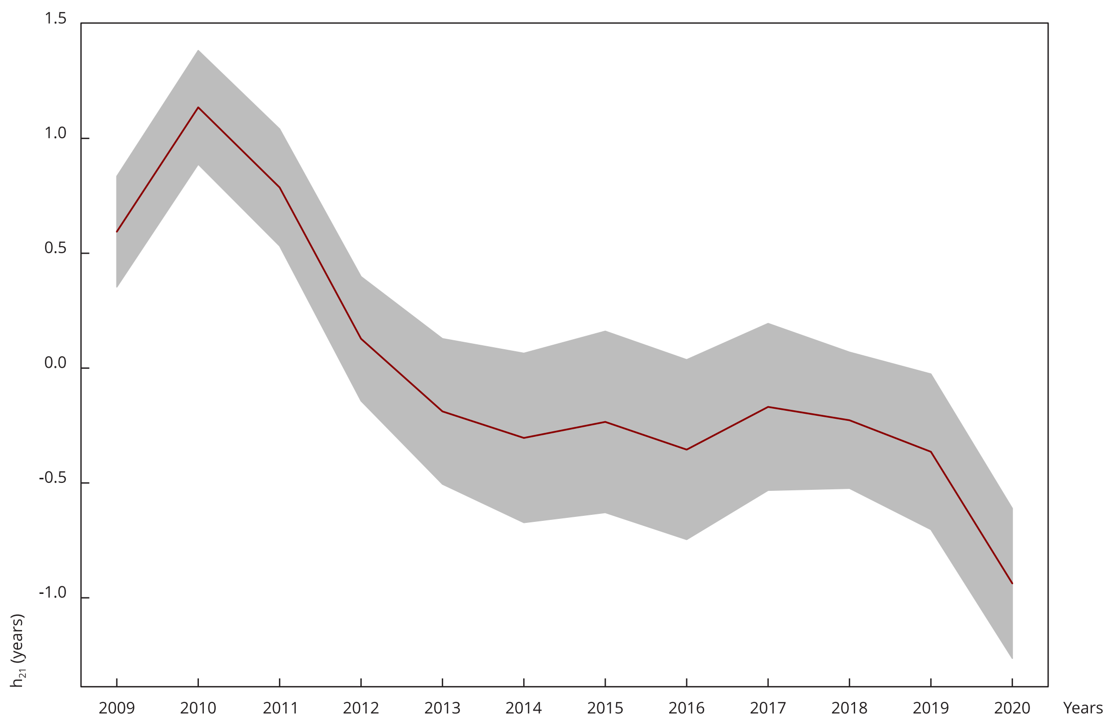
Figure 4 shows the adjustment of the smoothing function for the temporal effect, evidencing the year 2010 as the one with the highest risk, reaching 4 times more risk than the baseline year of 2012. Another point of note is that the year that presented the lowest rate was 2020. Most notably, from 2012 to 2019 the incidences behaved in a constant way, with little spatial variation.

The three parameters of the Sichel distribution were modeled with temporal delay variables, which shows that the distribution changes its shape and variability over time, presenting a dependence with the previous period (year). The spatial dependence was verified only in the mean parameter. Figure 5 shows the adjusted spatial effect, the lighter colors characterize a higher risk ($\exp(\mu)$); thus allowing us to observe regions further south and east of the studied area that presented higher risk in relation to the northern region, possibly reaching over 20 times higher, in some municipalities, when compared to the basal (municipality with value 0 on the map scale).

Nevertheless, we must consider the presence of informational bias due to the underreporting of cases, and due to the study's ecological approach, at the municipal level, which makes important details unfeasible in a possibly heterogeneous society.

Figure 4

Effects estimated using p-splines smoothing function, associated with the mean parameter with relation to time.



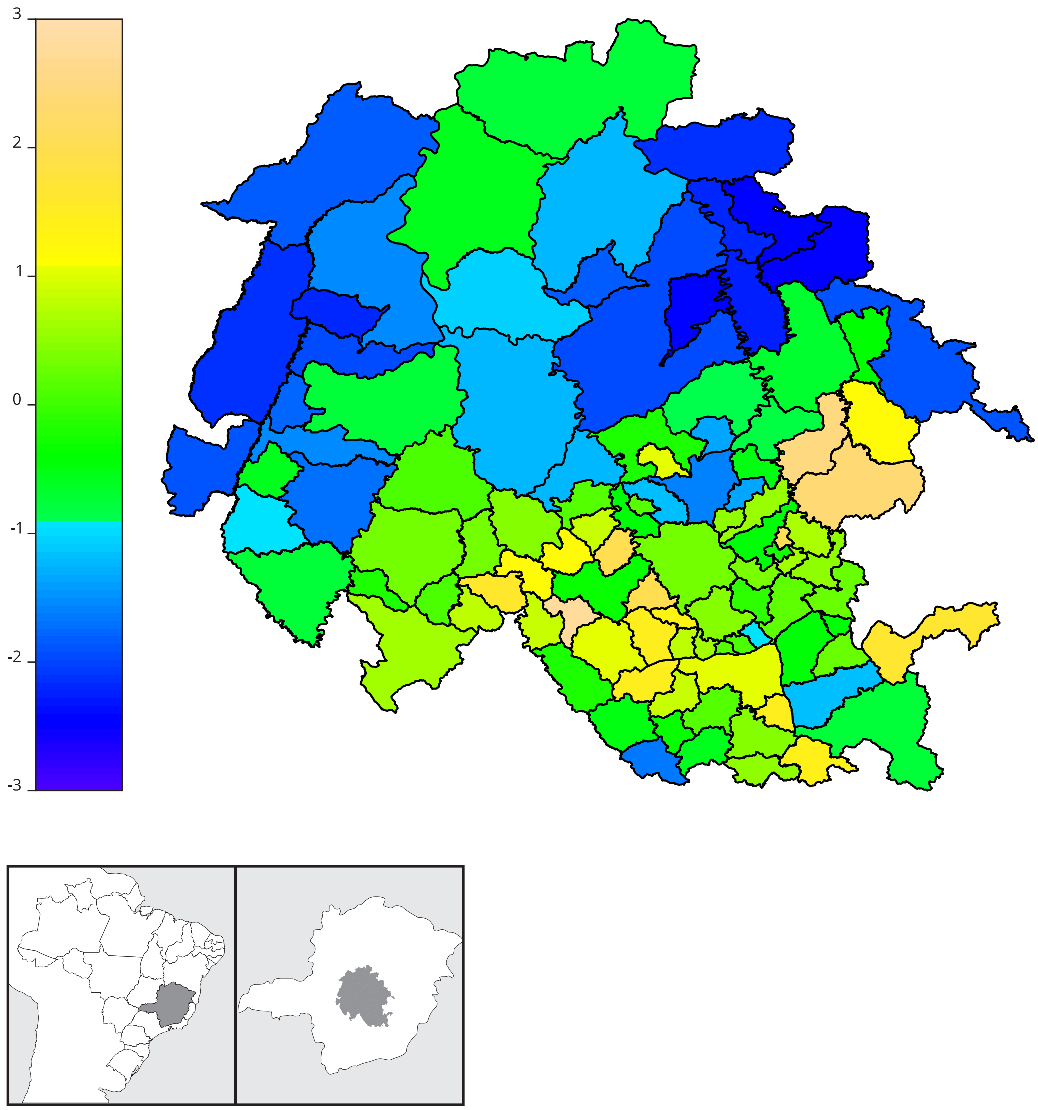
Conclusions

The GAMLSS model enabled the treatment and modeling of spatio-temporal data, with the use of a Gaussian Markov random field for the treatment of area data, using structured spatial effect, from a contiguity neighborhood matrix. For the temporal adjustment, dependence of orders 1 and 2 and cubic spline function were considered to model the trend.

Our study showed that the control and follow-up of *B. glabrata* snails may be fundamental for the control of schistosomiasis in the studied area. The presence of *B. straminea* snails was inversely associated with the incidence of schistosomiasis in the studied area, and the presence of *B. tenagophila* was not relevant. Another point that deserves a more detailed analysis is the relationship of municipalities with the presence of mining dams and the possible migration of positive individuals, which would need to be better investigated.

Figure 5

Representation of spatial risk adjusted for the mean parameter. Municipalities of the Minas Gerais State, Brazil, 2007-2020.



Contributors

D. A. Nogueira contributed to the elaboration of the project, data analysis, writing, and revision; and approved the final version. T. Sáfadi contributed to the design and revision; and approved the final version. R. R. Lima contributed to the preparation of the project, writing, and revision; and approved the final version. A. S. Mata contributed to the design and review; and approved the final version. M. M. C. Graciano contributed to the elaboration of the project and revision; and approved the final version. J. M. P. Barçante contributed to the revision; and approved the final version. T. A. Barçante contributed to the design and revision; and approved the final version. S. M. P. Dourado contributed to the elaboration of the project and revision; and approved the final version.

Additional information

ORCID: Denismar Alves Nogueira (0000-0003-2285-8764); Thelma Sáfadi (0000-0002-4918-300X); Renato Ribeiro de Lima (0000-0003-4607-4964); Angélica Sousa da Mata (0000-0002-3892-5274); Miriam Monteiro de Castro Graciano (0000-0003-2764-6891); Joziana Muniz de Paiva Barçante (0000-0002-7963-5775); Thales Augusto Barçante (0000-0002-2506-4065); Stela Márcia Pereira Dourado (0000-0001-7072-175X).

References

1. Becker JM, Ganatra AA, Kandie F, Mühlbauer L, Ahlheim J, Brack W, et al. Pesticide pollution in freshwater paves the way for schistosomiasis transmission. Sci Rep 2020; 10:3650.
2. World Health Organization. Field use of molluscicides in schistosomiasis control programmes: an operational manual for programme managers. Geneva: World Health Organization; 2017.
3. Santos AD, Santos MB, Santos PGR, Barreto AS, Araújo KCGM. Spatial analysis and epidemiological characteristics of cases of schistosomiasis in the municipality of Simão Dias, northeast of Brazil. Rev Patol Trop 2016; 45:99-114.
4. Cruz JIN, Salazar GO, Corte RL. Setback of the schistosomiasis control program in the Brazilian state with the highest prevalence of the disease. Rev Pan-Amazônica Saúde 2020; 11:e202000567.
5. Silva da Paz W, dos Santos Reis E, Leal IB, Barbosa YM, de Araújo RJ, de Sousa AR, et al. Basic and associated causes of schistosomiasis-related mortality in Brazil: a population-based study and a 20-year time series of a disease still neglected. J Glob Health 2021; 11:04061.
6. Souza RLM, Gargioni C, Siqueira RV, Silva RM, Pinto PLS, Kanamura HY. Epidemiological aspects of schistosomiasis in area at the southwest of Minas Gerais, Brazil. Rev Inst Adolfo Lutz 2017; 76:e1730.
7. Massara CL, Enk MJ, Caldeira RL, Mendonça CLF, Scholte RGC, Carvalho OS. Occurrence of mollusks, genus *Biomphalaria*, in parks of the city of Belo Horizonte, Minas Gerais, Brazil. Rev Patol Trop 2012; 41:471-9.
8. Simões TC, Sena R, Meira KC. The influence of the age-period-cohort effects on the temporal trend mortality from schistosomiasis in Brazil from 1980 to 2014. PLoS One 2020; 15:e0231874.
9. Wood CL, Sokolow S, Jones I, Chamberlin A, Lafferty KD, Kuris AM, et al. Precision mapping of snail habitat provides a powerful indicator of human schistosomiasis transmission. Proc Natl Acad Sci U S A 2019; 116:23182-91.
10. Coordenação Geral de Doenças em Eliminação, Departamento de Vigilância das Doenças Transmissíveis, Secretaria de Vigilância em Saúde, Ministério da Saúde. Vigilância da esquistossomose mansoni: diretrizes técnicas. 4th Ed. Brasília: Ministério da Saúde; 2014.
11. Cardoso DM, Araújo AF, Gonçalves SA. Spatial, socio-demographic, clinical and temporal aspects of schistosomiasis in the state of Minas Gerais between the years of 2011 and 2020. Brazilian Journal of Development 2021; 7:78130-43.
12. McManus DP, Dunne DW, Sacko M, Utzinger J, Vennervald BJ, Zhou XN. Schistosomiasis. Nat Rev Dis Primers 2018; 4:13.

13. Anderson TJC, Enabulele EE. *Schistosoma mansoni*. Trends Parasitol 2021; 37:176-7.
14. Paz WS, Gomes DS, Ramos RES, Cirilo TM, Santos IGA, Ribeiro CJN, et al. Spatiotemporal clusters of schistosomiasis mortality and association with social determinants of health in the Northeast region of Brazil (1980-2017). Acta Trop 2020; 212:105668.
15. Nelder JA, Wedderburn RWM. Generalized linear models. J R Stat Soc Ser A 1972; 135:370-84.
16. Hastie TJ, Tibshirani RJ. Generalized additive models. New York: Chapman & Hall/Boca Raton: CRC Press; 1990.
17. Rigby RA, Stasinopoulos DM. Generalized additive models for location, scale and shape. J R Stat Soc Ser C Appl Stat 2005; 54:507-54.
18. Instituto Brasileiro de Geografia e Estatística. Cidades@. https://cidades.ibge.gov.br/ (accessed on 05/Jun/2021).
19. Agência Nacional de Águas. HidroWeb. https://www.snirh.gov.br/hidroweb/ (accessed on 09/Jun/2021).
20. Agência Nacional de Mineração. Sistema Integrado de Gestão de Barragens de Mineração (SIGBM) versão pública. https://www.gov.br/anm/pt-br/assuntos/acesso-a-sistemas/sistema-integrado-de-gestao-de-barragens-de-mineracao-sigbm-versao-publica (accessed on 11/Jun/2021).
21. Departamento de Vigilância Epidemiológica, Secretaria de Vigilância em Saúde, Ministério da Saúde. Vigilância e controle de moluscos de importância epidemiológica. Diretrizes técnicas: Programa de Vigilância e Controle da Esquistossomose (PCE). 2nd Ed. Brasília: Ministério da Saúde; 2008. (Série A. Normas e Manuais Técnicos).
22. De Bastiani F, Rigby RA, Stasinopoulos DM, Cysneiros AHMA, Uribe-Opazo MA. Gaussian Markov random field spatial models in GAMLSS. J Appl Stat 2018; 45:168-86.
23. Stasinopoulos DM, Rigby RA, Heller GZ, Voudouris V, De Bastiani F. Flexible regression and smoothing: using GAMLSS in R. New York: Chapman & Hall/CRC Press; 2017.
24. Wood SN. Generalized additive models: an introduction with R. 2nd Ed. Boca Raton: Chapman & Hall/CRC Press; 2017.
25. Bivand RS, Wong DWS. Comparing implementations of global and local indicators of spatial association. TEST 2018; 27:716-48.
26. Akaike H. A new look at the statistical model identification. IEEE Trans Automat Control 1974; 19:716-23.
27. Schwarz G. Estimating the dimension of a model. Ann Stat 1978; 6:461-4.
28. Dunn PK, Smyth GK. Randomized quantile residuals. J Comput Graph Stat 1996; 5:236-44.
29. Royston P, Wright EM. Goodness-of-fit statistics for age-specific reference intervals. Stat Med 2000; 19:2943-62.
30. Rigby RA, Stasinopoulos MD, Heller GZ, De Bastiani F. Distributions for modeling location, scale, and shape: using GAMLSS in R. New York: Chapman & Hall/CRC Press; 2019.
31. de Sá Júnior A, De Carvalho LG, Da Silva FF, Alves MC. Application of the Köppen classification for climatic zoning in the state of Minas Gerais, Brazil. Theor Appl Climatol 2012; 108:1-7.
32. Barreto AVMS, Melo ND, Ventura JVT, Santiago RT, Silva MBA. Analysis of schistosomiasis mansoni positivity in endemic health regions in the State of Pernambuco, Brazil, 2005-2010. Epidemiol Serv Saúde 2015; 24:87-96.
33. Barreto Costa JV, Marques da Silva Filho J. Esquistossomose mansônica: uma análise do perfil epidemiológico na região sudeste. Revista Saúde.com 2021; 17:2226-34.
34. Kulldorff M, Heffernan R, Hartman J, Assuncao R, Mostashari F. A space-time permutation scan statistic for disease outbreak detection. PLoS Med 2005; 2:e59.
35. Xu F, Beard K. A comparison of prospective space-time scan statistics and spatiotemporal event sequence based clustering for COVID-19 surveillance. PLoS One 2021; 16:e0252990.
36. Scholte RGC, Gosoniu L, Malone JB, Chammartin F, Utzinger J, Vounatsou P. Predictive risk mapping of schistosomiasis in Brazil using Bayesian geostatistical models. Acta Trop 2014; 132:57-63.
37. Rigby RA, Stasinopoulos DM, Akantziliotou C. A framework for modeling overdispersed count data, including the Poisson-shifted generalized inverse Gaussian distribution. Comput Stat Data Anal 2008; 53:381-93.
38. Ljung GM, Box GEP. On a measure of lack of fit in time series models. Biometrika 1978; 65:297-303.
39. Draper N, Smith H. Applied regression analysis. 2nd Ed. New York: John Wiley & Sons; 1981.
40. Carvalho OS, Mendonça CLF, Marcelino JMR, Passos LKJ, Fernandez MA, Leal RS, et al. Geographical distribution of intermediate hosts of *Schistosoma mansoni* in the states of Paraná, Minas Gerais, Bahia, Pernambuco and Rio Grande do Norte, Brazil, 2012-2014. Epidemiol Serv Saúde 2018; 27:e2017343.
41. Oliveira TD, Amaral OV, Braga LMV. Occurrence and spatial analysis of schistosomosis in the microregion of Caratinga, Minas Gerais, in the period 2011-2015. Braz J Surg Clin Res 2018; 22:7-13.

Resumo

No Brasil, milhões de pessoas vivem em áreas de risco para a esquistossomose, uma doença negligenciada, de caráter crônico e com elevada morbidade. O helminto Schistosoma mansoni *está presente em todas as macrorregiões, incluindo o Estado de Minas Gerais, um dos mais endêmicos. Por essa razão, a identificação de potenciais focos é fundamental para subsidiar políticas públicas de cunho educativo e profilático no controle desse desfecho. Nesse contexto, o objetivo do trabalho consiste em modelar dados de esquistossomose em relação aos aspectos espaciais e temporais, além de avaliar a importância de algumas variáveis exógenas socioeconômicas e a presença das principais espécies de* Biomphalaria*. Como trabalhar com casos incidentes, uma variável discreta de contagem, exige uma modelagem apropriada, foi escolhida a modelagem GAMLSS por considerar conjuntamente uma distribuição mais adequada à variável resposta devido à inflação de zeros e à heterocedasticidade espacial. Verificaram-se valores elevados de incidência em diversos municípios de 2010 a 2012 e uma tendência de queda até 2020. Também foi identificado que a distribuição da incidência se comporta de maneira diferente no espaço e no tempo. Municípios com barragem apresentaram risco 2,25 vezes maior do que os que não a continham. A presença de* B. glabrata *foi relacionada ao risco de ocorrência da doença. Por outro lado, a presença de* B. straminea *refletiu em menor risco de ocorrência da esquistossomose. Conclui-se que o controle e o acompanhamento dos caramujos da* B. glabrata *podem ser fundamentais para a contenção e a eliminação da esquistossomose e o modelo GAMLSS foi eficaz para tratamento e modelagem de dados espaçotemporais.*

Distribuição Espacial; Biomphalaria*; Análise de Regressão; Análise de Dados Secundários*

Resumen

En Brasil, millones de personas viven en áreas de riesgo de esquistosomiasis, una enfermedad crónica desatendida y con alta morbilidad. El helminto Schistosoma mansoni *está presente en todas las macrorregiones, incluido el estado de Minas Gerais, uno de los más endémicos del país. Por ello, la identificación de potenciales brotes es fundamental para promover políticas públicas de carácter educativo y profiláctico en el control de este desenlace. En este contexto, el objetivo de este trabajo es modelar datos sobre esquistosomiasis con respecto a aspectos espaciotemporales, además de evaluar la importancia de algunas variables socioeconómicas exógenas y la presencia de las principales especies de* Biomphalaria*. Dado que en el trabajo con casos incidentes una variable de conteo discreta requiere un adecuado modelado, se eligió el modelo GAMLSS, ya que en conjunto considera una distribución más adecuada para la variable de respuesta debido a la inflación de ceros y la heterocedasticidad espacial. Se encontraron valores de alta incidencia en varios municipios en el periodo evaluado de 2010 a 2012 y una tendencia a descenso hasta 2020. También se verificó que existe una distribución de incidencia de manera diferente en el espacio y el tiempo. Los municipios con represas presentaban 2,25 veces más riesgo que los que no las tenían. La presencia de* B. glabrata *estuvo relacionada con el riesgo de la enfermedad. Por otro lado, la presencia de* B. straminea *ocasionaba un menor riesgo de padecer la enfermedad. Se concluye que el control y seguimiento de caracoles* B. glabrata *puede ser fundamental para el control y eliminación de la esquistosomiasis y que el modelo GAMLSS resultó ser efectivo para el tratamiento y modelado de datos espaciotemporales.*

Distibución Espacial; Biomphalaria*; Análisis de Regresión; Análisis de Datos Secundarios*

Submitted on 14/Apr/2022
Final version resubmitted on 23/Jan/2023
Approved on 09/Mar/2023