Spatial analysis and risk mapping of soil-transmitted helminth infections in Brazil, using Bayesian geostatistical models

Ronaldo G. C. Scholte[1,2,3], Nadine Schur[1,2], Maria E. Bavia[4], Edgar M. Carvalho[5], Frédérique Chammartin[1,2], Jürg Utzinger[1,2], Penelope Vounatsou[1,2]

[1]Department of Epidemiology and Public Health, Swiss Tropical and Public Health Institute, Basel, Switzerland; [2]University of Basel, Basel, Switzerland; [3]Laboratory of Helmintology and Medical Malacology - René Rachou Research Center/Fiocruz, Belo Horizonte, Brazil; [4]Laboratory of Monitoring Diseases by Geographic Information System, School of Veterinary Medicine, Federal University of Bahia, Salvador, Brazil; [5]Department of Preventive Medicine, Federal University of Bahia, Salvador, Brazil

Abstract. Soil-transmitted helminths (*Ascaris lumbricoides*, *Trichuris trichiura* and hookworm) negatively impact the health and wellbeing of hundreds of millions of people, particularly in tropical and subtropical countries, including Brazil. Reliable maps of the spatial distribution and estimates of the number of infected people are required for the control and eventual elimination of soil-transmitted helminthiasis. We used advanced Bayesian geostatistical modelling, coupled with geographical information systems and remote sensing to visualize the distribution of the three soil-transmitted helminth species in Brazil. Remotely sensed climatic and environmental data, along with socioeconomic variables from readily available databases were employed as predictors. Our models provided mean prevalence estimates for *A. lumbricoides*, *T. trichiura* and hookworm of 15.6%, 10.1% and 2.5%, respectively. By considering infection risk and population numbers at the unit of the municipality, we estimate that 29.7 million Brazilians are infected with *A. lumbricoides*, 19.2 million with *T. trichiura* and 4.7 million with hookworm. Our model-based maps identified important risk factors related to the transmission of soil-transmitted helminths and confirm that environmental variables are closely associated with indices of poverty. Our smoothed risk maps, including uncertainty, highlight areas where soil-transmitted helminthiasis control interventions are most urgently required, namely in the North and along most of the coastal areas of Brazil. We believe that our predictive risk maps are useful for disease control managers for prioritising control interventions and for providing a tool for more efficient surveillance-response mechanisms.

Keywords: Bayesian modelling, geographical information system, remote sensing, soil-transmitted helminths, variable selection, Brazil.

Introduction

The soil-transmitted helminths are a group of nematode parasites that cause human infection when their eggs are ingested (*Ascaris lumbricoides* and *Trichuris trichiura*) or larvae penetrate the skin (the two hookworm species *Ancylostoma duodenale* and *Necator americanus*) (Bethony et al., 2006; Hotez et al., 2008; Knopp et al., 2012). Recent climate suitability modelling suggests that more than 5.3 billion people were at risk of soil-transmitted helminthiasis in 2010 with over 1 billion people infected with at least one of the aforementioned soil-transmitted helminth species (Pullan and Brooker, 2012). The distribution of infec-

tions is governed by social-ecological systems with poverty playing a key role in maintaining high levels of endemicity. Indeed, the highest rates and intensity of infection and public health burden are concentrated in poor communities in tropical and subtropical areas (de Silva et al., 2003; Tchuem Tchuenté, 2011; Lustigman et al., 2012; Ziegelbauer et al., 2012). The highest infection intensities are usually observed in school-aged children, whilst there is a decline in prevalence and intensity of infection with age (Anderson and May, 1985; de Silva et al., 2003; Bethony et al., 2006; Brooker et al., 2007). Hence, the global strategy to control soil-transmitted helminthiasis (and other helminth infections such as schistosomiasis) targets the school-aged population, and this is done by periodic administration of anthelminthic drugs, usually without prior diagnosis (WHO, 2002, 2006; Hotez et al., 2007; Utzinger et al., 2009).

In Brazil, soil-transmitted helminth infections are of considerable public health importance, particularly through their association with anaemia, diarrhoea and

Corresponding author:
Penelope Vounatsou
Department of Epidemiology and Public Health
Swiss Tropical and Public Health Institute
P.O. Box, CH-4002 Basel, Switzerland
Tel. +41 61 284-8109; Fax +41 61 284-8105
E-mail: penelope.vounatsou@unibas.ch

malnutrition, which impairs the physical, psychosomatic and social development of school-aged children and adolescents (Pedrazzani et al., 1988; Nascimento and Moitinho, 2005; Brooker et al., 2007; Jardim-Botelho et al., 2008). In 2009, the Pan American Health Organization (PAHO) put forward an agenda for the control and elimination of neglected diseases and other poverty-related infections in Latin America and the Caribbean (PAHO, 2009; Schneider et al., 2011; Colston and Saboyá, 2013). Interventions that focus on priority areas are required to improve the cost-effectiveness of control programmes and, among the proposed deliverables, reliable maps of the distribution of neglected tropical diseases along with estimates of the numbers infected play an important role (Carvalho et al., 2010; Guimarães et al., 2010; Ault et al., 2012; Scholte et al., 2012a; Colston and Saboyá, 2013).

Bayesian geostatistical models have been widely used for predicting the spatial distribution of helminth infections at different scales and for different parts of the world. Successful applications include risk profiling of schistosomiasis, soil-transmitted helminthiasis and co-infections at the village, district, national and regional level in Africa and Asia (Clements et al., 2006; Raso et al., 2006; Steinmann et al., 2007; Schur et al., 2011a; Chammartin et al., 2013a). In Latin America, the paucity of data has considerably delayed the mapping, and hence control, of soil-transmitted helminthiasis (Sabin Vaccine Institute/Inter-American Development Bank/PAHO, 2011). Recent work attempted to fill this gap. Indeed, the geographical distribution of the three major soil-transmitted helminth species has been modelled based on historical data from South America using different approaches (Chammartin et al., 2013b; Colston and Saboyá, 2013). Another recent study assessed the regional risk of ascariasis and trichuriasis in relation to climatic, environmental and social variables in the Brazilian State of Minas Gerais (Scholte et al., 2012b) based on data from the schistosomiasis national control programme (NCP). However, the analyses by Scholte et al. (2012b) and Colston and Saboyá (2013) used standard statistical models that ignore spatial correlation of the data, and hence estimates have to be interpreted with caution.

In the study presented here, we analyse prevalence data of *A. lumbricoides*, *T. trichiura* and hookworm collected by the NCP from 2005 to 2009 and produce nationwide high-resolution risk maps for the three soil-transmitted helminth species. Superimposing population density, we estimated the number of infected

people stratified by state. A Bayesian geostatistical modelling framework was employed with aligned data collected over large number of locations. As predictors, we used remotely sensed climatic and environmental data and socioeconomic proxies. Our predictive maps aim to inform local and national authorities to better target limited resources to control and eventually eliminate soil-transmitted helminthiasis in Brazil.

Materials and methods

Soil-transmitted helminth infection prevalence data

Prevalence data for *A. lumbricoides*, *T. trichiura* and hookworm were obtained from 1,020 municipalities across Brazil surveyed in the years 2005-2009 within the NCP framework. The diagnostic method employed to detect *Schistosoma mansoni* infection was the Kato-Katz technique (Katz et al., 1972), which is also widely used for detection and quantification of soil-transmitted helminth infections (Speich et al., 2010; Knopp et al., 2012). A single Kato-Katz thick smear was prepared from each participant. Fig. 1 shows the localities of the surveyed municipalities.

Climatic and environmental data

Table 1 summarises the different data sources utilised to derive climatic and environmental data. In brief, climate data were extracted from Worldclim Global Climate Data (Hijmans et al., 2005). These data consist of 19 bioclimatic variables. Environmental data were obtained from the Moderate Resolution Imaging Spectroradiometer (MODIS) onboard the Terra satellite. Land surface temperature (LST) data were used as proxy for day and night temperatures. The normalised difference vegetation index (NDVI) and enhanced vegetation index (EVI) were used as proxies for moisture and vegetation, respectively. A digital elevation model (DEM) was employed to extract altitude data.

Socioeconomic data

Table 2 summarises the socioeconomic indicators used in our study. In brief, population data for the year 2010 were stratified into rural and urban. Human development index (HDI) data for the year 2000 were provided by the Instituto Brasileiro de Geografia e Estatística (IBGE). Unsatisfied basic needs
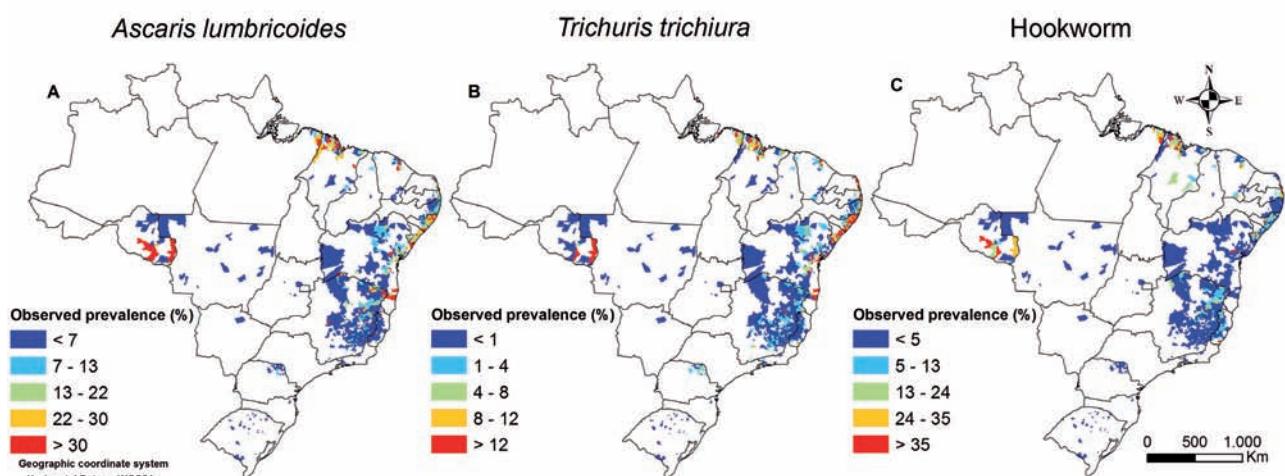
Fig. 1. Observed prevalence of soil-transmitted helminth infections in Brazil from 2005-2009 (data obtained from the schistoso-miasis NCP).

(UBN) data for the year 2000 were provided by PAHO. Infant mortality rate (IMR), which has been used as a poverty proxy by the Center for International Earth Science Information Network (CIESIN), for the year 2000, and human influence index (HII) for the year 2005, were provided by the CIESIN.

Statistical analysis

For each soil-transmitted helminth species, univariate logistic regressions for the continuous environmental and socioeconomic factors were performed in order to assess potential non-linear trends in relation to the logit scale of the infection risk. Covariates indicating

Table 1. Data sources and properties of the climatic and other environmental covariates used for Bayesian geostatistical modelling (spatial resolution = 1 km).

Source	Type of data	Period	Temporal resolution
Shuttle Radar Topography Mission (SRTM)	Digital elevation model (DEM)	2000	Once
Moderate Resolution Imaging Spectroradiometer (MODIS)/Terra	Land surface temperature (LST) for day and night	2005-2009	8 days
	Normalised difference vegetation index (NDVI)	2005-2009	16 days
Worldclim Global Climate Data	BIO1 (annual mean temperature)	1950-2000	Once
	BIO2 (mean diurnal range (mean of monthly (max temp - min temp)))	1950-2000	Once
	BIO3 (isothermality (BIO2/BIO7) (*100))	1950-2000	Once
	BIO4 (temperature seasonality (standard deviation*100))	1950-2000	Once
	BIO5 (max temperature of warmest month)	1950-2000	Once
	BIO6 (min temperature of coldest month)	1950-2000	Once
	BIO7 (temperature annual range (BIO5-BIO6))	1950-2000	Once
	BIO8 (mean temperature of wettest quarter)	1950-2000	Once
	BIO9 (mean temperature of driest quarter)	1950-2000	Once
	BIO10 (mean temperature of warmest quarter)	1950-2000	Once
	BIO11 (mean temperature of coldest quarter)	1950-2000	Once
	BIO12 (annual precipitation)	1950-2000	Once
	BIO13 (precipitation of wettest month)	1950-2000	Once
	BIO14 (precipitation of driest month)	1950-2000	Once
	BIO15 (precipitation seasonality (coefficient of variation))	1950-2000	Once
	BIO16 (precipitation of wettest quarter)	1950-2000	Once
	BIO17 (precipitation of driest quarter)	1950-2000	Once
	BIO18 (precipitation of warmest quarter)	1950-2000	Once
	BIO19 (precipitation of coldest quarter)	1950-2000	Once

Table 2. Data sources and properties of the socioeconomic covariates used in our Bayesian geostatistical model to predict soil-transmitted helminth infection prevalence in Brazil (socioeconomic data are at the unit of the municipality).

Source	Type of data	Year
Instituto Brasileiro de Geografia e Estatística (IBGE) (census data)	Population data	2010
	Human development index (HDI)	2000
	Rural population	2000
Pan American Health Organization (unsatisfied basic needs) (census data)	Bras0_3 (% of pupils enrolled in primary school)	2000
	Bras0_4 (% of pupils completing primary school)	2000
	Bras0_5 (rate literacy 15 to 24 years)	2000
	Bras0_6 (girls and boys, primary school)	2000
	Bras0_7 (girls and boys, high school)	2000
	Bras0_8 (girls and boys, undergraduate school)	2000
	Bras0_9 (relation literacy women and men 15 to 24 years)	2000
	Bras0_10 (% women work, non-farm)	2000
	Bras0_11 (% people with potable water at house)	2000
	Bras0_12 (% people with sanitation at house)	2000
	Bras0_13 (% people with energy at house)	2000
	Bras0_14 (% own x rent house)	2000
	Bras0_15 (index secure tenure house)	2000
	Bras0_16 (unemployment rate)	2000
	Bras0_17 (% of houses with phone)	2000
	Bras0_18 (% of houses with computer)	2000
	Bras2_11 (% of people overcrowding)	2000
	Bras2_15 (% of people subsistence)	2000
	Infant mortality rate (IMR)	2000
Center for International Earth Science Information Network (CIESIN)	Human influence index (HII)	2005

non-linear trends were categorised into three or four groups, based on inspection of scatter-plots. The deviance information criterion (DIC) was employed to detect whether univariate logistic regressions on linear or categorised covariates have smaller DIC and therefore better fit the outcome.

Parsimonious sets of environmental and socioeconomic covariates were obtained for each infection using Gibbs variable selection (George and McCulloch, 1993). Variable selection was carried out in a Bayesian framework, using non-spatial multivariate logistic regression models with standardised linear covariates. Bayesian geostatistical logistic regression models were fitted on the reduced set of covariates with location-specific random effect parameters to account for potential spatial correlation. The model was developed assuming that the random effects follow a multivariate normal distribution with variance-covariance matrix related to an exponential correlation function between any pair of locations. Markov chain Monte Carlo (MCMC) simulation was employed to estimate all model parameters.

We predicted species-specific soil-transmitted helminth infection prevalence based on the estimated associations of the reduced set of covariates on the outcome and spatial parameters via Bayesian kriging, using marginal predictive posterior distributions. Prediction was carried out using a grid with a spatial resolution of

5 x 5 km that resulted in over 280,000 locations for Brazil. A random sample of 820 training locations was selected to fit the geostatistical models. The predictive ability of each model was then assessed by comparing the model-based predictions at the remaining 200 test locations with the observed outcomes at these locations using (i) mean absolute errors (MAE) and (ii) proportion of location correctly predicted within a 95% Bayesian credible interval (BCI). The statistical analyses were carried out in STATA/IC version 10.1 (Stata Corp. LP; College Station, USA), WinBUGS 1.4.3 (Imperial College and Medical Research Council, London, UK) and Fortran 95 (Digital Equipment Corp.; Maynard, USA).

Ethical considerations

Soil-transmitted helminth infection prevalence data were obtained from the schistosomiasis NCP, approved by the Ministry of Health in Brazil. No specific ethical approval was required for the secondary analysis presented in this article, which focuses on risk profiling of soil-transmitted helminth infection.

Results

The georeferenced soil-transmitted helminth species-specific infection prevalence data obtained from the

schistosomiasis NCP consisted of 1,020 unique survey locations. Taken together, the observed prevalence ranged from nil to 64.3% with a mean prevalence of 10.3% (median = 5.0% and standard deviation (SD) = 13.4%) for *A. lumbricoides*, from nil to 42.5% with a mean prevalence of 4.9% (median = 0.8% and SD = 9.1%) for hookworm, and from nil to 33.4% with a mean prevalence of 3.7% (median = 0.8% and SD = 6.7%) for *T. trichiura* infection. The spatial distribution of the survey locations and the prevalence of infection with *A. lumbricoides*, hookworm and *T. trichiura* are shown in Figs. 1a, 1b and 1c, respectively.

In our initial analyses, we included a set of 45 climatic, environmental and socioeconomic covariates. However, after careful variable selection, between 29 and 33 of these covariates were removed depending on the soil-transmitted helminth species in question. Hence, for the final geostatistical analyses, a total of 16, 15 and 12 covariates were used for the prediction of risk with hookworm, *A. lumbricoides* and *T. trichiura*, respectively. Model validation indicated that the MAE of the *A. lumbricoides* was 7.3, meaning that, on average, our predicted prevalence at the test locations for *A. lumbricoides* in Brazil was about 7% higher or lower than the observed prevalence data. The hookworm and *T. trichiura* models showed better predictive performance with MAEs of 4.4 and 3.0, respectively. Additionally, the latter two soil-transmitted helminth species were correctly predicted within 95% BCIs for 98% of the test locations. Even though the model for *A. lumbricoides* ranked third, its predictive ability was high as it correctly predicted 94% of the test locations within a 95% BCI. Tables 3, 4 and 5 show the results of the Bayesian variable selection and parameter estimates of the final geostatistical models for hookworm, *A. lumbricoides* and *T. trichiura*, respectively.

For *A. lumbricoides*, the most suitable climate conditions appear to be high rainfall and temperatures below 26 °C. In addition, we found a negative association between *A. lumbricoides* and altitude. The socioeconomic indicators suggest that the higher the HDI and the proportion of people living in houses with sanitation facilities and electricity, the lower the odds of *A. lumbricoides* infection. Municipalities with higher tenure housing index were at lower risk of *A. lumbricoides*, while HII was positively correlated with *A. lumbricoides* infection.

For *T. trichiura*, the socioeconomic indicators suggest that the higher the HDI and percentage of self-subsistence people, the lower the odds of an infection with this helminth species. On the other hand, the higher the HII and the percentage of people living under crowded conditions, the higher the odds of *T. trichiura* infection. The climatic factors that were found to be significantly associated with *T. trichiura* infection were high annual rainfall and temperatures below 26 °C. However, altitude showed a negative relation.

The hookworm model suggested that high vegetation index and precipitation are suitable for transmission, while altitude was negatively associated. The socioeconomic variables suggested that the higher the percentage of urban population and HDI, the lower the odds for hookworm infection. Municipalities with high IMR had lower odds for both hookworm and *T. trichiura* infection. This negative relation has been estimated both by non-spatial bivariate logistic regression (data not shown) and multivariate geostatistical models.

Spatial parameter estimates were similar for all three soil-transmitted helminth species. The spatial range was estimated to be 4.9 km (95% BCI: 3.1-12.7 km), 5.1 km (95% BCI: 3.1-14.7 km) and 5.1 km (95% BCI: 3.1-14.2 km) for *A. lumbricoides*, *T. trichiura* and hookworm infection, respectively. The residual spatial variation was overall very low with the largest variance of 0.03 (95% BCI: 0.01-0.04) observed for the *A. lumbricoides* model. In comparison, the non-spatial variance was between 160 (*A. lumbricoides*) and 510 (hookworm) times smaller than the spatial variation. These results suggest that the remaining spatial correlation was low after adjusting for the climatic and socioeconomic predictors.

The predictive spatial distribution of the risk for *A. lumbricoides*, *T. trichiura* and hookworm in Brazil, based on the estimated outcome-predictor associations and other modelling parameters, are shown in Fig. 2. The highest risk is in the North, the eastern part of Northeast (the coast line), the South and along the south-eastern coastline. While there is low risk for *A. lumbricoides* infection in the south-eastern and central regions, the risk of infection with this helminth species is considerable higher elsewhere in Brazil. The highest risk for *T. trichiura* is in the North and along the north-eastern coast line, whereas the low risk areas are concentrated in the South, the Central and the Southeast regions. High-risk areas for hookworm are in the North and inland with the low-risk areas mainly found in the South, Southeast and Northeast. Fig. 3 shows the corresponding prediction uncertainty maps, as indicated by the SDs of the model prediction errors.

By considering municipality-specific population sizes, we were able to convert model-based prevalence esti-

R.G.C. Scholte et al. - Geospatial Health 8(1), 2013, pp. 97-110

Table 3. Parameter estimates of Bayesian geostatistical logistic regression models relating hookworm risk in Brazil with climatic, environmental and socioeconomic predictors.

Variable	Geostatistical model	
	OR[a]	95% BCI[b]
Normalised difference vegetation index (NDVI)	1.72*	1.60, 1.77
Mean diurnal range		
<8.0 °C	1.00	
8.0-9.9 °C	1.38*	1.20, 1.56
10.0-11.9 °C	1.39	0.96, 1.51
≥12.0 °C	1.16	0.94, 1.44
Precipitation wettest month	1.75*	1.64, 1.84
Precipitation driest month	0.70*	0.68, 0.73
Precipitation warmest quarter		
<200 mm	1.00	
200-299 mm	0.67*	0.40, 0.75
300-449 mm	0.64*	0.59, 0.74
≥450 mm	0.19*	0.14, 0.26
Altitude	0.36*	0.35, 0.37
% urban population		
<35.0	1.00	
35.0-54.9	0.55*	0.45, 0.61
55.0-74.9	0.54*	0.42, 0.59
≥75.0	0.49*	0.29, 0.55
Infant mortality rate (IMR)		
<0.3	1.00	
0.3-0.39	0.45*	0.39, 0.54
0.4-0.49	0.35*	0.31, 0.36
≥0.5	0.10*	0.07, 0.11
Human development index (HDI)	0.30*	0.22, 0.32
Human influence index (HII)		
<19.0	1.00	
19.0-20.9	1.10	0.78, 1.26
21.0-23.9	1.93*	1.02, 2.12
≥24.0	1.31	0.51, 1.48
% people potable water at house	1.85*	1.43, 1.97
% people sanitation at house		
<1.0	1.00	
1.0-14.9	1.18	0.97, 1.33
15.0-39.9	2.69*	2.25, 2.90
≥40.0	3.74*	2.08, 4.39
% people energy at house		
<85.0	1.00	
85.0-94.9	1.20	0.91, 1.28
≥95.0	1.98*	1.26, 2.44
Index secure tenure house	0.52*	0.42, 0.54
Unemployment rate		
<15.0	1.00	
15.0-24.9	0.87	0.76, 1.00
≥25.0	0.59*	0.45, 0.63
% house with phone	1.66*	1.48, 1.79
σ^2	0.01	0.01, 0.02
τ^2	5.36	4.82, 5.98
Range (km)	5.12	3.05, 14.20

[a]OR, odds ratio; [b]Bayesian credible interval; *statistical significance based on 95% BCIs

Table 4. Parameter estimates of Bayesian geostatistical logistic regression models relating *A. lumbricoides* risk in Brazil with climatic, environmental and socioeconomic predictors.

Variable	Geostatistical model	
	OR[a]	95% BCI[b]
Land surface temperature (LST) day		
<26.0 °C	1.00	
26.0-27.4 °C	0.56*	0.53, 0.58
27.5-28.9 °C	0.52*	0.41, 0.55
≥29.0 °C	0.35*	0.22, 0.42
Temperature annual range		
<12.0 °C	1.00	
12.0-14.9 °C	0.76*	0.68, 0.86
15.0-17.9 °C	1.11*	1.02, 1.22
≥18.0 °C	1.20	0.76, 1.35
Annual precipitation	1.32*	1.22, 1.35
Precipitation driest month		
<10.0 mm	1.00	
10.0-19.9 mm	1.84*	1.33, 2.08
20.0-29.9 mm	1.93*	1.84, 2.03
≥30.0 mm	2.37*	1.70, 2.59
Precipitation warmest quarter	0.72*	0.71, 0.76
Altitude		
<150 m	1.00	
150-390 m	0.72*	0.48, 0.78
400-690 m	0.37*	0.32, 0.40
≥700 m	0.30*	0.23, 0.35
Infant mortality rate (IMR)		
<0.3	1.00	
0.3-0.39	0.90*	0.73, 0.98
0.4-0.49	1.11	0.76, 1.32
≥0.5	1.19	0.71, 1.25
Human development index (HDI)		
<0.6	1.00	
0.6-0.64	0.52*	0.50, 0.55
0.65-0.69	0.57*	0.43, 0.66
≥0.7	0.24*	0.23, 0.25
Human influence index (HII)	1.41*	1.38, 1.44
% people sanitation in house	0.85*	0.72, 0.89
% people energy in house		
<80.0	1.00	
80.0-89.9	0.89	0.64, 1.01
90.0-94.9	0.69*	0.63, 0.74
≥95.0	0.56*	0.42, 0.63
% own x rent house		
<80.0	1.00	
80.0-84.9	1.35*	1.20, 1.45
85.0-89.9	1.02	0.76, 1.19
≥90.0	1.13	0.92, 1.19
Index secure tenure house	0.88*	0.85, 0.91
% house with phone		
<4.0	1.00	
4.0-5.9	1.21	0.96, 1.38
6.0-11.9	1.33*	1.19, 1.42
≥12.0	1.71*	1.52, 1.86
% house with computer		
<0.6	1.00	
0.6-1.5	1.12*	1.05, 1.18
1.6-2.5	1.03	0.98, 1.13
≥2.6	1.21	0.91, 1.42
σ^2	0.03	0.01, 0.04
τ^2	2.89	2.63, 3.20
Range (km)	4.86	3.05, 12.70

[a]OR, odds ratio; [b]Bayesian credible interval; *statistical significance based on 95% BCIs

Table 5. Parameter estimates of Bayesian geostatistical logistic regression models relating *T. trichiura* risk in Brazil with climatic, environmental and socioeconomic predictors.

Variable	Geostatistical model	
	OR[a]	95% BCI[b]
Land surface temperature (LST) day		
<26.0 °C	1.00	
26.0-27.4 °C	0.73*	0.39, 0.83
27.5-28.9 °C	0.95	0.50, 1.01
≥29.0 °C	0.58*	0.32, 0.69
Temperature annual range	0.50*	0.35, 0.52
Annual precipitation		
<1,000 mm	1.00	
1,000-1,199 mm	1.22*	1.10, 1.43
1,200-1,399 mm	2.51*	1.95, 2.79
≥1,400 mm	2.04*	1.82, 2.25
Precipitation warmest quarter		
<200 mm	1.00	
200-299 mm	0.87*	0.79, 0.90
300-449 mm	0.76*	0.70, 0.84
≥450 mm	1.38	0.59, 1.97
Altitude	0.55*	0.54, 0.59
Infant mortality rate (IMR)		
<0.3	1.00	
0.3-0.39	0.61*	0.52, 0.68
0.4-0.49	0.74*	0.37, 0.84
≥0.5	0.57*	0.32, 0.64
Human development index (HDI)		
<0.6	1.00	
0.6-0.64	0.65*	0.53, 0.72
0.65-0.69	0.46*	0.40, 0.51
≥0.7	0.20*	0.15, 0.23
Human influence index (HII)	1.63*	1.46, 1.70
% own x rent house		
<80.0	1.00	
80.0-84.9	1.34*	1.12, 1.49
85.0-89.9	1.29	0.96, 1.52
≥90.0	1.09	0.85, 1.14
% house with computer		
<0.6	1.00	
0.6-1.5	1.11	0.92, 1.19
1.6-2.5	1.06	0.94, 1.24
≥2.6	1.83*	1.74, 1.98
% people overcrowding		
<45.0	1.00	
45.0-54.9	1.24	0.97, 1.40
55.0-64.9	1.41*	1.26, 1.48
≥65.0	1.88*	1.70, 2.29
% people subsistence	0.72*	0.70, 0.77
σ^2	0.02	0.01, 0.07
τ^2	4.37	3.92, 4.90
Range (km)	5.12	3.06, 14.7

[a]OR, odds ratio; [b]Bayesian credible interval; *statistical significance based on 95% BCIs
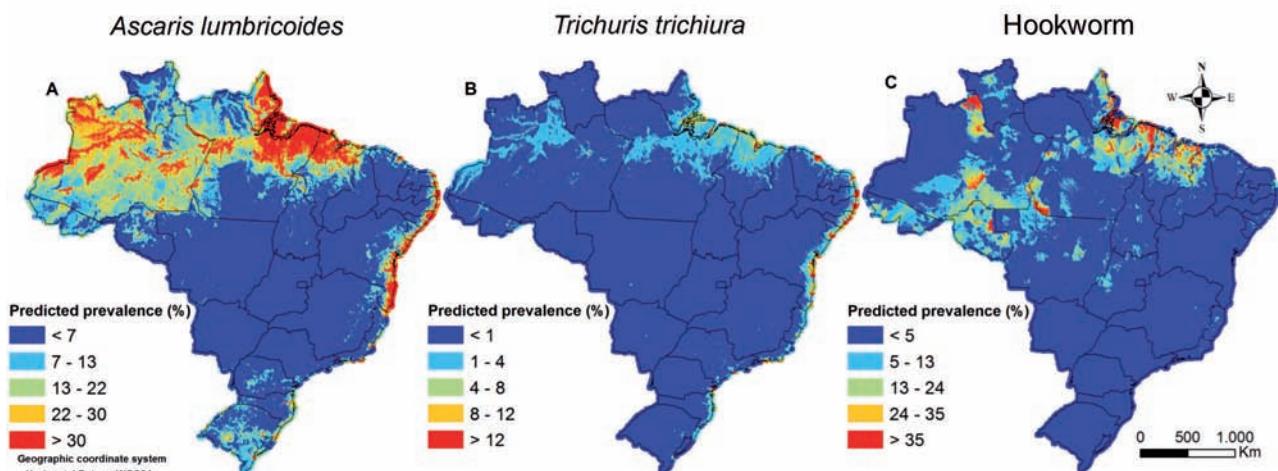
Fig. 2. Estimates of the spatial distribution of soil-transmitted helminth infection prevalence in Brazil.

mates into number of infected people at the unit of the municipality and, after aggregation, at the state level (Table 6). At the national level, we estimate that 29.7 million Brazilians (15.6%) are infected with *A. lumbricoides*. The corresponding 95% CI ranges between 26.9 and 32.4 million people. For hookworm, we estimate the number of infected people at 4.7 million (2.5%), whereas the 95% CI range between 4.0 and 5.8 million people. For *T. trichiura*, we estimate that 19.2 million Brazilians are infected (10.1%) with a 95% CI ranging between 16.6 and 21.6 million people.

Discussion

To our knowledge, we present the first model-based, spatially explicit risk maps of soil-transmitted helminth infections for Brazil at high spatial resolution (5 x 5 km).

Our estimates include the number of infected people at the state level (aggregation of municipality-level data) as well as for the entire country. At the national level, we estimate overall prevalences for *A. lumbricoides*, *T. trichiura* and hookworm of 15.6%, 10.1% and 2.5%, respectively, which translates to 4.7 million infections with hookworm, 19.2 million infections with *T. trichiura* and 29.7 infections with *A. lumbricoides*. Our estimates were derived from advanced Bayesian geostatistical models, including two validation approaches. Model construction took into account the relation between the parasite prevalence data with remotely sensed climatic and environmental data and socioeconomic variables obtained from readily available databases. From the initial set of 45 covariates, between 12 and 16 covariates remained in the final models, depending on the soil-transmitted helminth species investigated.
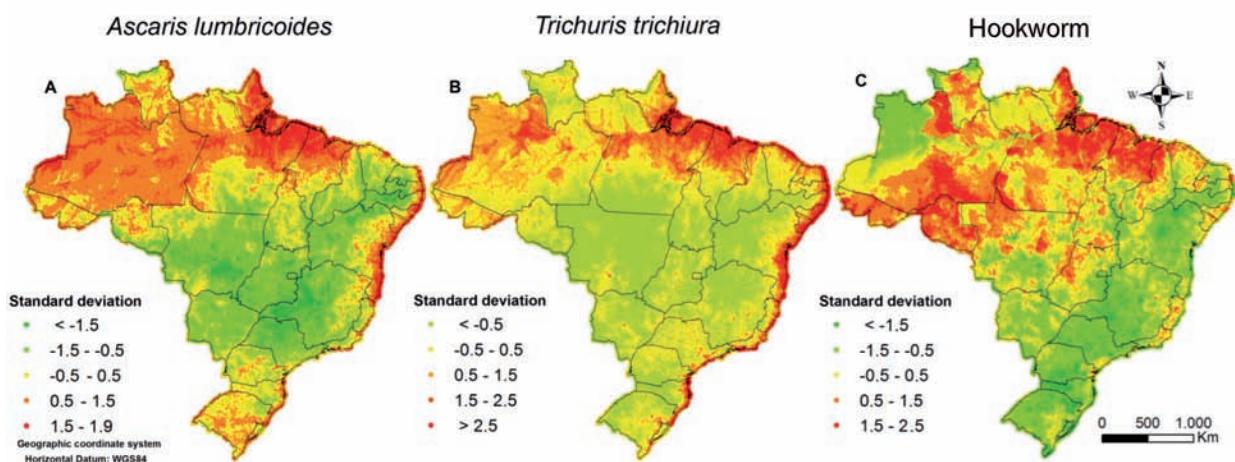

Fig. 3. Maps of the error of prediction.

Table 6. Geostatistical model-based estimates of the number of people infected with *A. lumbricoides*, *T. trichiura* and hookworm in each one of the 26 states and the Federal District in Brazil.

| State | Abbreviation | Population | Predicted infected population | | | | | |
| | | | *A. lumbricoides* | | *T. trichiura* | | Hookworm | |
			Estimate	95% CI	Estimate	95% CI	Estimate	95% CI
Acre	AC	732,793	137,472	116,148; 157,991	137,985	125,381; 151,542	23,890	14,290; 38,985
Alagoas	AL	3,120,922	640,726	546,162; 727,487	309,908	260,285; 360,779	237,191	206,606; 276,826
Amapá	AP	668,689	201,945	186,164; 214,917	107,793	84,590; 131,331	28,955	25,210; 33,101
Amazonas	AM	3,480,937	987,890	859,792; 1,125,039	444,168	373,505; 554,514	124,966	80,062; 191,800
Bahia	BA	14,021,432	1,448,414	1,364286; 1,549,369	412,231	335,113; 503,370	273,418	231,354; 315,482
Ceará	CE	8,448,055	830,444	685,138; 967,303	827,065	689,362; 954,631	274,562	223,874; 326,940
Distrito Federal	DF	2,562,963	84,835	57,411; 121,485	26,399	13,584; 47,672	44,340	24,349; 73,045
Espírito Santo	ES	3,512,672	364,265	322,815; 419,062	318,951	244,131; 393,068	98,355	72,713; 133,833
Goiás	GO	6,004,045	258,174	208,941; 315,213	460,511	387,862; 548,170	55,838	43,830; 69,647
Maranhão	MA	6,569,683	1,655,561	1,509,057; 1,807,320	1,321,821	1,181,230; 1,413,139	392,868	352,136; 429,658
Mato Grosso	MT	3,033,991	191,749	158,071; 227,246	285,196	236,348; 341,628	14,260	11,833; 18,204
Mato Grosso do Sul	MS	2,449,341	146,716	115,609; 180,027	96,260	82,788; 112,180	23,024	17,391; 32,087
Minas Gerais	MG	19,595,309	1,062,066	979,766; 1,156,124	482,045	389,947; 585,900	207,711	172,439; 252,780
Pará	PA	7,588,078	1,760,435	1,569,974; 1,916,749	1,232,304	1,075,231; 1,380,272	289,106	246,613; 334,635
Paraíba	PB	3,766,834	315,285	271,213; 363,500	258,405	215,463; 310,764	132,593	111,122; 155,947
Paraná	PR	10,439,601	1,169,236	963,576; 1,375,940	223,408	178,518; 282,914	144,067	103,353; 185,825
Pernambuco	PE	8,796,032	605,168	535,679; 680,813	243,651	205,828; 281,474	211,105	177,680; 244,530
Piauí	PI	3,119,015	322,195	296,619; 348,394	392,684	335,918; 420,444	58,326	47,410; 69,866
Rio de Janeiro	RJ	15,993,583	1,754,497	1,532,186; 1,986,404	1,004,398	846,061; 1,175,529	782,087	655,737; 905,237
Rio Grande do Norte	RN	3,168,133	279,747	232,541; 333,288	188,188	164,110; 216,384	145,735	121,340; 171,080
Rio Grande do Sul	RS	10,695,532	1,801,128	1,464,219; 2,114,507	257,763	192,520; 365,788	151,877	80,217; 272,737
Rondônia	RO	1,560,501	200,213	176,961; 222,528	357,355	291,814; 429,294	22,784	18,883; 28,090
Roraima	RR	451,227	78,018	69,986; 85,869	50,312	38,535; 62,586	7,671	5,957; 10,108
Santa Catarina	SC	6,249,682	803,710	613,094; 973,076	104,370	79,996; 140,618	115,620	81,871; 156,868
São Paulo	SP	41,252,160	2,037,857	1,773,843; 2,297,746	643,534	519,778; 800,292	878,672	734,289; 1,068,431
Sergipe	SE	2,068,031	429,944	384,034; 469,444	157,378	130,493; 186,537	166,684	139,593; 195,223
Tocantins	TO	1,383,453	144,295	121,001; 163,663	121,329	106,941; 148,445	16,464	12,590; 21,997
Total		190,732,694	29,697,081	26,874,237; 32,405,485	19,168,636	16,631,891; 21,648,161	4,749,245	3,967,241; 5,760,128

Using a simple linear regression approach, Colston and Saboyá (2013) estimated that the overall prevalence of soil-transmitted helminth infection among school-aged children in Brazil is 29.3%. Brazil-specific data obtained from a Bayesian geostatistical meta-analysis of historic soil-transmitted helminth survey data at the continental level revealed similar prevalence estimates for *A. lumbricoides* (14.3%) and *T. trichiura* (10.1%) (Chammartin et al., 2013b). However, for hookworm, a five-fold difference in the estimated prevalence was found: 2.5% in the current study compared to the 12.3% obtained by Chammartin et al. (2013b). Differences in the spatial and temporal coverage of the data might explain this observation. For example, in terms of spatial coverage, risk profiling carried out by Chammartin et al. (2013b) focussed on entire Latin America, whereas the current analysis pertains to Brazil only. In regards of temporality, the hookworm data employed by Chammartin et al. (2013b) covered a 10-year period (2000-2009), as opposed to a 5-year period (2005-2009) utilised in the current analysis.

The geographical distribution of *A. lumbricoides* and *T. trichiura* infection indicated high prevalence along the Atlantic forest areas, which are humid and warm and thus provide a suitable environment for transmission. Our geostatistical models were able to identify the most important climatic factors for parasite development, and these corroborate with prior laboratory investigations (Beer, 1973; Crompton and Pawlowski, 1985). The similar spatial patterns of *A. lumbricoides* and *T. trichiura* predicted by our models are supported by the similar life cycles of these two helminth species outside the human host (Bethony et al., 2006; Knopp et al., 2012). Of note, the overall prevalence of *A. lumbricoides* is somewhat higher than that of *T. trichiura*, which might be explained by different morphologies of the parasite eggs (*A. lumbricoides* eggs show higher resilience to climate extremes in the natural environment compared *T. trichiura* eggs), higher egg production rates of *A. lumbricoides* compared to *T. trichiura*, and somewhat faster reinfection of *A. lumbricoides* compared to *T. trichiura* following anthelminthic drug administration (Brooker and Bundy, 2009; Jia et al., 2012). Hookworm has a distinctively different life cycle. In fact, hookworm eggs hatch and human infection occurs when the third-stage larvae (L_3) actively penetrate the skin. Hence, hookworm larvae are vulnerable to desiccation and need humid and warm environment for survival (Bethony et al., 2006). The geographical distribution of the hookworm risk predicted by our model indicates a high prevalence on the Amazonia Legal region.

Our spatial predictions for each of the three soil-transmitted helminth species studied showed strong correlations with socioeconomic and environmental covariates (precipitation and temperature). Our results therefore suggest that these variables are important drivers for the spatial distribution of *A. lumbricoides*, *T. trichiura* and hookworm, which should be considered when control strategies are planned. Importantly, our models predict the risk of soil-transmitted helminth infections at high spatial resolutions (i.e. 5 x 5 km) aggregated at various levels from the municipality to the national. Our fine-scale estimates (e.g. municipality level) can be utilised, alongside guidelines put forward by the World Health Organization (WHO) for preventive chemotherapy (WHO, 2006), to calculate the cost of preventive chemotherapy at the administrative unit where interventions will be implemented. Colston and Saboyá (2013), who estimated the prevalence of soil-transmitted helminth infection in the school-aged population of Brazil at 29.3% and using WHO treatment guidelines, came up with an initial cost of US$ 17.5 million for administering anthelminthic drugs to school-aged children.

It must be admitted that our prediction maps may not capture disease focality with high accuracy since our geospatial analysis is based on survey data aggregated at the municipality level. The surveys were carried out at higher resolution (i.e. villages and towns), but disaggregated information of the exact survey locations were not available. Our models therefore assumed that there is no variation in our predictors within the municipalities although this might not be true, especially not with respect to the socioeconomic factors. In addition, parts of Brazil lack point prevalence data. For example, there are no prevalence data available for soil-transmitted helminth for large areas covering the Amazon as well as the Central and Southern regions. Interestingly, we estimate a lower risk in those areas compared to Chammartin et al. (2013b), who used data from neighbouring countries to derive inference in such areas. We recommend that future parasitological surveys be conducted in municipalities where no or only limited data are currently available to improve model accuracy. Finally, advances have been made with multivariate spatial models for risk profiling of multiple health outcomes (for a recent example of a joint spatial model for hookworm-*Schistosoma mansoni* co-infection, see Schur et al. (2011b)). It would be interesting to develop and validate joint spatial models for *A. lumbricoides* and *T. trichiura*, as these two helminth species share similar transmission pathways (Booth and Bundy, 1992; Bethony et al., 2006; Knopp et al., 2012).

Conclusions

Our predictive risk maps for the three common soil-transmitted helminth infections across Brazil should be useful tools for prioritising control interventions and to set-up effective surveillance-response mechanisms. Moreover, important data gaps have been identified and these must be filled to further improve predictive risk estimates. However, already at this stage, as can be judged from the maps shown here, useful information for spatial targeting of control interventions is provided. Moreover, the estimates presented here might serve as a benchmark for monitoring and evaluation of future efforts towards disease elimination.

Acknowledgements

References

Anderson RM, May RM, 1985. Helminth infections of humans: mathematical models, population dynamics, and control. Adv Parasitol 24, 1-101.

Ault SK, Nicholls RS, Saboya MI, 2012. The Pan American Health Organization's role and perspectives on the mapping and modeling of the neglected tropical diseases in Latin America and the Caribbean: an overview. Geospat Health 6, S7-S9.

Beer RJS, 1973. Morphological descriptions of the egg and larval stages of *Trichuris suis*, Schrank 1788. Parasitology 67, 263-268.

Bethony J, Brooker S, Albonico M, Geiger SM, Loukas A, Diemert D, Hotez PJ, 2006. Soil-transmitted helminth infections: ascariasis, trichuriasis, and hookworm. Lancet 367, 1521-1532.

Booth M, Bundy DAP, 1992. Comparative prevalences of *Ascaris lumbricoides*, *Trichuris trichiura* and hookworm infections and the prospects for combined control. Parasitology 105, 151-157.

Brooker S, Bundy DAP, 2009. Soil-transmitted helminths (geohelminths). In: Manson's Tropical Diseases (21st ed.). Cook GC, Zumla AI (eds). Saunders Elsevier, 1515-1548 pp.

Brooker S, Jardim-Botelho A, Quinnell RJ, Geiger SM, Caldas IR, Fleming F, Hotez PJ, Correa-Oliveira R, Rodrigues LC,

Bethony JM, 2007. Age-related changes in hookworm infection, anaemia and iron deficiency in an area of high *Necator americanus* hookworm transmission in south-eastern Brazil. Trans R Soc Trop Med Hyg 101, 146-154.

Carvalho OS, Scholte RGC, Guimarães RJPS, Freitas CC, Drummond SC, Amaral RS, Dutra LV, Oliveira G, Massara CL, Enk MJ, 2010. The Estrada Real project and endemic diseases: the case of schistosomiasis, geoprocessing and tourism. Mem Inst Oswaldo Cruz 105, 532-536.

Chammartin F, Scholte RGC, Guimarães LH, Tanner M, Utzinger J, Vounatsou P, 2013b. Soil-transmitted helminth infection in South America: a systematic review and geostatistical meta-analysis. Lancet Infect Dis 13, 507-518.

Chammartin F, Scholte RGC, Malone J, Bavia ME, Nieto P, Utzinger J, Vounatsou P, 2013a. Modelling the geographical distribution of soil-transmitted helminth infections in Bolivia. Parasit Vectors 6, 152.

Clements ACA, Moyeed R, Brooker S, 2006. Bayesian geostatistical prediction of the intensity of infection with *Schistosoma mansoni* in East Africa. Parasitology 133, 711-719.

Colston J, Saboyá M, 2013. Soil-transmitted helminthiasis in Latin America and the Caribbean: modelling the determinants, prevalence, population at risk and costs of control at subnational level. Geospat Health 7, 321-340.

Crompton DWT, Pawlowski ZS, 1985. Life history and development of *Ascaris lumbricoides* and the persistence of humam ascariasis. In: Ascariasis and its Public Health Significance. Taylor and Francis, London, 9-23 pp.

de Silva NR, Brooker S, Hotez PJ, Montresor A, Engels D, Savioli L, 2003. Soil-transmitted helminth infections: updating the global picture. Trends Parasitol 19, 547-551.

George EI, McCulloch RE, 1993. Variable selection via Gibbs sampling. J Am Stat Assoc 88, 881-889.

Guimarães RJPS, Freitas CC, Dutra LV, Scholte RCG, Martins-Bedé FT, Fonseca FR, Amaral RS, Drummond SC, Felgueiras CA, Oliveira GC, Carvalho OS, 2010. A geoprocessing approach for schistosomiasis studying and control in the state of Minas Gerais - Brazil. Mem Inst Oswaldo Cruz 105, 524-531.

Hijmans RJ, Cameron SE, Parra JL, Jones PG, Jarvis A, 2005. Very high resolution interpolated climate surfaces for global land areas. Int J Climatol 25, 1965-1978.

Hotez PJ, Brindley PJ, Bethony JM, King CH, Pearce EJ, Jacobson J, 2008. Helminth infections: the great neglected tropical diseases. J Clin Invest 118, 1311-1321.

Hotez PJ, Molyneux DH, Fenwick A, Kumaresan J, Ehrlich Sachs S, Sachs JD, Savioli L, 2007. Control of neglected tropical diseases. N Engl J Med 357, 1018-1027.

Jardim-Botelho A, Brooker S, Geiger SM, Fleming F, Souza Lopes AC, Diemert DJ, Corrêa-Oliveira R, Bethony JM, 2008. Age patterns in undernutrition and helminth infection in a rural area of Brazil: associations with ascariasis and hook-

worm. Trop Med Int Health 13, 458-467.

Jia TW, Melville S, Utzinger J, King CH, Zhou XN, 2012. Soil-transmitted helminth reinfection after drug treatment: a systematic review and meta-analysis. PLoS Negl Trop Dis 6, e1621.

Katz N, Chaves A, Pellegrino J, 1972. A simple device for quantitative stool thick-smear technique in schistosomiasis mansoni. Rev Inst Med Trop São Paulo 14, 397-400.

Knopp S, Steinmann P, Keiser J, Utzinger J, 2012. Nematode infections: soil-transmitted helminths and *Trichinella*. Infect Dis Clin North Am 26, 341-358.

Lustigman S, Prichard RK, Gazzinelli A, Grant WN, Boatin BA, McCarthy JS, Basáñez MG, 2012. A research agenda for helminth diseases of humans: the problem of helminthiases. PLoS Negl Trop Dis 6, e1582.

Nascimento SA, Moitinho MLR, 2005. *Blastocystis hominis* and other intestinal parasites in a community of Pitanga city, Paraná state, Brazil. Rev Inst Med Trop 47, 213-217.

PAHO, 2009. Elimination of neglected diseases and other poverty-related infections. Washington DC, Pan American Health Organization (Technical Report CD49-09).

Pedrazzani ES, Mello DA, Pripas S, Fucci M, Barbosa CAA, Santoro MC, 1988. Intestinal helminthiasis. II. Prevalence and correlation with income, family size, anemia and nutritional status. Rev Saude Publica 22, 384-389 (in Portugese).

Pullan RL, Brooker SJ, 2012. The global limits and population at risk of soil-transmitted helminth infections in 2010. Parasit Vectors 5, 81.

Raso G, Vounatsou P, Singer BH, N'Goran EK, Tanner M, Utzinger J, 2006. An integrated approach for risk profiling and spatial prediction of *Schistosoma mansoni*-hookworm coinfection. Proc Natl Acad Sci USA 103, 6934-6939.

Sabin Vaccine Institute, Inter-American Development Bank, Pan American Health Organization, 2011. A call to action: addressing soil-transmitted helminths in Latin America and the Caribbean. Technical Report.

Schneider MC, Aguilera XP, Barbosa da Silva Junior J, Ault SK, Najera P, Martinez J, Requejo R, Nicholls RS, Yadon Z, Silva JC, Leanes LF, Periago MR, 2011. Elimination of neglected diseases in Latin America and the Caribbean: a mapping of selected diseases. PLoS Negl Trop Dis 5, e964.

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

Scholte RGC, Freitas CC, Dutra LV, Guimarães RJ, Drummond SC, Oliveira G, Carvalho OS, 2012b. Utilizing environmental, socioeconomic data and GIS techniques to estimate the risk for ascariasis and trichuriasis in Minas Gerais, Brazil. Acta Trop 121, 112-117.

Schur N, Gosoniu L, Raso G, Utzinger J, Vounatsou P, 2011b. Modelling the geographical distribution of co-infection risk from single-disease surveys. Stat Med 30, 1761-1776.

Schur N, Hürlimann E, Garba A, Traoré MS, Ndir O, Ratard RC, Tchuem Tchuenté LA, Kristensen TK, Utzinger J, Vounatsou P, 2011a. Geostatistical model-based estimates of schistosomiasis prevalence among individuals aged ≤20 years in West Africa. PLoS Negl Trop Dis 5, e1194.

Speich B, Knopp S, Mohammed KA, Khamis IS, Rinaldi L, Cringoli G, Rollinson D, Utzinger J, 2010. Comparative cost assessment of the Kato-Katz and FLOTAC techniques for soil-transmitted helminth diagnosis in epidemiological surveys. Parasit Vectors 3, 71.

Steinmann P, Zhou XN, Matthys B, Li YL, Li HJ, Chen SR, Yang Z, Fan W, Jia TW, Vounatsou P, Utzinger J, 2007. Spatial risk profiling of *Schistosoma japonicum* in Eryuan county, Yunnan province, China. Geospat Health 2, 59-73.

Tchuem Tchuenté LA, 2011. Control of soil-transmitted helminths in sub-Saharan Africa: diagnosis, drug efficacy concerns and challenges. Acta Trop 120, S4-S11.

Utzinger J, Raso G, Brooker S, de Savigny D, Tanner M, Ørnbjerg N, Singer BH, N'Goran EK, 2009. Schistosomiasis and neglected tropical diseases: towards integrated and sustainable control and a word of caution. Parasitology 136, 1859-1874.

WHO, 2002. Prevention and control of schistosomiasis and soil-transmitted helminthiasis: report of a WHO expert committee. WHO Tech Rep Ser 912, 1-57.

WHO, 2006. Preventive chemotherapy in human helminthiasis: coordinated use of anthelminthic drugs in control interventions: a manual for health professionals and programme managers. Geneva: World Health Organization, 1-62.

Ziegelbauer K, Speich B, Mäusezahl D, Bos R, Keiser J, Utzinger J, 2012. Effect of sanitation on soil-transmitted helminth infection: systematic review and meta-analysis. PLoS Med 9, e1001162.

Appendix

Geostatistical model

Let $Y_{i,l}$ be the number of observed infections with disease l (l = 1, *A. lumbricoides*; l = 2, *T. trichiura*; and l = 3, hookworm infection) at location s_i (i = 1, ..., n) of the study region $A \subset R^2$ and let N_i be the total number of tested individuals for each helminth infection. We assume that $Y_{i,l}$ are binomially distributed, that is $Y_{i,l} \sim Bin(p_{i,l}, N_i)$ where $p_{i,l}$ is the probability that an infection with helminth species l is detected at site s_i. We model the influence of the associated set of m environmental and socioeconomic covariates $\mathbf{X}_{i,l}$ = (1, $X_{i1,l}$, ..., $X_{im,l})^T$ of the infection on the logit, such as logit($p_{i,l}$) = $\mathbf{X}_{i,l}^T \beta_l + v_{i,l} + w_{i,l}$, with the vector of regression coefficients β_l = ($\beta_{0,l}$, $\beta_{1,l}$, ..., $\beta_{m,l})^T$ location-specific random effect parameters $v_{i,l}$ and exchangeable random effect parameters $w_{i,l}$. Unobserved spatial variation is introduced on $v_{i,l}$ by assuming that $\mathbf{v}_l$ = ($v_{1,l}$, ..., $v_{n,l})^T$ follows a latent stationary Gaussian process over the study region with $\mathbf{v}_l \sim MVN(\mathbf{0}, \Sigma_l)^T$. The elements of the variance-covariance matrix Σ_l are related to an isotropic exponential correlation function defined by $\Sigma_{ij,l} = \sigma_l^2 \exp(-\rho_l d_{ij})$ with spatial variance σ_l^2, spatial decay ρ_l and the Euclidean distance d_{ij} between any pair of locations s_i and s_j. The distance at which spatial correlation between locations drops below 5% is known as spatial range and can be calculated by $3/\rho$ if exponential correlation functions are employed. We further assume that $w_{i,l} \sim N(0, \tau_l^2)$ with variance τ_l^2 that can be interpreted as nugget of the spatial process.

Prior specifications and implementation details

Bayesian model formulation requires the specification of prior distributions of all model parameters. In this study, we adopted non-informative Normal and Gamma prior distributions with mean 0 and variance of 100 for the regression coefficients β_l and for the variance parameters σ_l^2 and τ_l^2, respectively. Furthermore, uniform prior distributions with minimum bounds of 30 m and maximum bounds of 3,000 km were chosen for ρ_l (the range of distances between the survey locations varies from 15 m to almost 3,000 km, with a median of about 800 km).

Models were fitted in WinBUGS 1.4.3 (Imperial College London and Medical Research Council; London, UK) using MCMC simulation. Models were run for two chains with a thinning of 10 and a burn-in of 1,000 iterations. Starting values for the chains were based on non-spatial model estimates and semi-variogram estimates for the spatial model parameters. Convergence was assessed by inspection of ergodic averages of selected model parameters. After convergence, samples of 500 iterations were extracted for each parameter and each soil-transmitted helminth species.

Samples from the predictive posterior distribution at each prediction locations were obtained via Bayesian kriging. Kriging was performed in Fortran 95 (Digital Equipment Corporation; Maynard, MA, USA) based on code written by the authors and the samples obtained from model fit. The results were summarised based on the median and the SD of the prediction.

Variable selection

Gibbs variable selection was employed on non-spatial logistic regression models to select parsimonious sets of the environmental and socioeconomic covariates for each soil-transmitted helminth infection (George and McCulloch, 1993). A vector of indicator variables $\mathbf{I}$ = ($I_{0,l}$, $I_{1,l}$, ..., $I_{m,l})^T$ was linked to the vector of regression coefficients in the models, such as $\mathbf{X}_{i,l}^T \beta_l$, I_l to specify between inclusion and exclusion of the covariates where I_k = 1 if the k-th covariate is included and zero otherwise. Priors for the joint distribution of $\beta_{k,l}$ and $I_{k,l}$ are defined by $P(\beta_{k,l}|I_{k,l}) = P(\beta_{k,l}|I_{k,l})P(I_{k,l})$ in the Gibbs variable selection approach. Furthermore, $P(\beta_k | I_k)$ is given by $I_{k,l} N(0, \tau_{I=1,l}^2) + (1 - I_{k,l}) N(0, \tau_{I=0,l}^2)$. In this study, we adopted non-informative Bernoulli prior distributions with probability of 0.5 for $P(I_{k,l})$ and set the inclusion and exclusion variances to $\tau_{I=1}^2$ = 100 and $\tau_{I=0}^2$ = 0.01, respectively.

Model validation

Model validation was performed to obtain estimates of the accuracy of predictive ability of the geostatistical models for each infection. The MAE provides information about the absolute distances between the observed $p_{i,l}$ and median predicted outcome $\hat{p}_{i,l}$ at the i-th test locations (i = 1, ..., k), such as

$$\text{MAE} = \frac{1}{k} \sum_{i=1}^{k} |p_{i,l} - \hat{p}_{i,l}|.$$

The outcome of the BCI approach is the proportion of test locations correctly predicted within the 95% BCIs of the posterior predictive distribution (restricted by the lower percentiles $c_{i,l}^l$ and upper percentiles $c_{i,l}^u$),

$$\text{MAE} = \frac{1}{k} \sum_{i=1}^{k} \min \left(\text{I}(c_{i,l}^l < p_{i,l}), \text{I}(c_{i,l}^u < p_{i,l}) \right).$$