

RESEARCH ARTICLE

Risk map development for soil-transmitted helminth infections in Argentina

Eliana M. Alvarez Di Fino^{1,2}, Jorge Rubio¹, Marcelo C. Abril², Ximena Porcasi¹, María V. Periago^{2,3*}

1 Mario Gulich Institute for Higher Space Studies, National University of Cordoba, National Commission of Space Activities (UNC_CONAE), Córdoba, Argentina, **2** Fundación Mundo Sano, Buenos Aires, Argentina, **3** Consejo Nacional de Investigaciones Científica y Técnicas (CONICET), Buenos Aires, Argentina

* vperiago@mundosano.org



Abstract

Background

Soil-transmitted helminths (STHs) comprise a group of helminth parasites that are included in the list of Neglected Tropical Diseases and require a passage through the soil to become infective. Several studies have detected that infection with STHs are associated with certain socioeconomic, environmental and soil characteristics. In Argentina, the presence of these parasites has been detected through a few point studies conducted in localities from 11 of the 23 provinces that comprise the country.

Methods/Principal findings

The most important characteristics previously associated with the presence of STHs were identified and ranked through the use of an expert survey and the Analytical Hierarchy Process (AHP) in order to construct a risk map of STHs specific for Argentina. Prevalence data from previous studies was used to validate the generated risk map. The map shows that half of Argentina, from the Central provinces to the North, contains localities with the characteristics necessary for the development of these parasites.

Conclusions/Significance

The predicted map should serve as a useful tool for guiding the identification of survey areas for the generation of baseline data, detecting hotspots of infection, planning and prioritizing areas for control interventions, and eventually performing post-implementation surveillance activities.

OPEN ACCESS

Citation: Alvarez Di Fino EM, Rubio J, Abril MC, Porcasi X, Periago MV (2020) Risk map development for soil-transmitted helminth infections in Argentina. *PLoS Negl Trop Dis* 14(2): e0008000. <https://doi.org/10.1371/journal.pntd.0008000>

Editor: Martin Walker, Royal Veterinary College, UNITED KINGDOM

Received: July 25, 2019

Accepted: December 17, 2019

Published: February 10, 2020

Peer Review History: PLOS recognizes the benefits of transparency in the peer review process; therefore, we enable the publication of all of the content of peer review and author responses alongside final, published articles. The editorial history of this article is available here: <https://doi.org/10.1371/journal.pntd.0008000>

Copyright: © 2020 Alvarez Di Fino et al. This is an open access article distributed under the terms of the [Creative Commons Attribution License](https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited.

Data Availability Statement: All relevant data are within the manuscript and its Supporting Information files.

Author summary

Given the lack of information on the prevalence of soil-transmitted helminths (STH) in Argentina, the aim of this study was to generate a national risk map to guide the implementation of surveys to determine baseline data. For this purpose, factors most associated

Funding: This study was funded by Fundación Mundo Sano. Mundo Sano did not play any role in the study design, data collection and analysis, decision to publish or preparation of the manuscript.

Competing interests: The authors have declared that no competing interests exist.

with the presence of STHs from previous studies conducted in other countries were ranked by experts in the field through the use of an Analytical Hierarchy Process. These characteristics were grouped into categories: socioeconomic, environmental and soil characteristics. The resulting map was validated against point prevalence studies conducted in the country and it showed that the top half of the country, from the Central provinces to the North, have localities with conditions appropriate for the development of these parasites.

Introduction

Neglected Tropical Diseases (NTDs) are endemic in tropical and subtropical areas of the world and include 20 diseases [1]. Among them are a group of parasites called soil-transmitted helminths (STHs) that are exclusively human (no animal reservoir) and which require a passage through the soil to become infective. The species that compose this group may be divided into those that infect through the fecal-oral route, *Ascaris lumbricoides*, *Trichuris trichiura*, and those that infect via penetration of the skin, the hookworms (*Ancylostoma duodenale* and *Necator americanus*) and *Strongyloides stercoralis*. This last species is not included in the list due to certain unique characteristics that make it hard to detect, quantify and treat [2] using World Health Organization (WHO) guidelines. For control of the morbidity caused by infection with these parasites, WHO guidelines suggest preventive chemotherapy (PC) with albendazole or mebendazole in high-risk groups: pre-school, school-aged children and women of childbearing age. Deworming should be performed once a year if the prevalence is between 20 and 50%, twice a year if it's more than 50% and it should be individualized in those communities with less than 20% prevalence [3].

Studies conducted in different parts of the world have detected that infection with STH is associated to certain characteristics that may be grouped basically into socioeconomic and environmental. For example, these parasites are more prevalent in populations that lack access to safe water, improved sanitation and hygiene (WASH), those that live in precarious houses with overcrowding, and those without waste management [4–12]. Some research groups have tried to create risk maps by identifying associations with environmental variables such as temperature, precipitation, humidity, NDVI, elevation, land cover in different countries or regions of the world [11,13–17]. Associations with soil characteristics such as composition, bulk density, organic carbon content, moisture and acidity have also been used in these maps [18–23].

Based on these past experiences, in the current study, the most important socioeconomic, environmental and soil factors previously associated with the presence of STH were identified. A survey was created in order to obtain the input from different experts in the field to rank them in order of importance from least to most important. The answers from the different participants of the survey were used to create a matrix using the Analytical Hierarchy Process (AHP) originally developed by Saaty [24] in order to construct a risk map for STHs specific for Argentina.

In Argentina, the presence of all five species of STHs have been detected through point studies conducted in a few cities spread throughout 11 of the 23 provinces of the country and the prevalences found in these studies vary from very low to very high depending on the province, the locality within the province and the coprological technique used for analysis of the fecal samples (more or less sensitive). The prevalence data obtained from these studies was used to validate the generated risk map and a machine model approach using a decision tree (DT) algorithm was used to determine the predicted prevalence for each city to see if it

matched with the prevalence obtained from the studies. Therefore, the aim of this study is to identify high risk areas in order to be able to prioritize surveillance of STHs and subsequently be able to plan effective control interventions in those places where the presence of these parasites is verified.

Methods

Study area

Argentina is located in the southern cone of South America (minimum and maximum longitude: -73.580 and -53.590; minimum and maximum latitude: -55.050 and -21.780), covering a total area of 3,761,274 km². In this study, we analyzed the risk for STH infection in 3526 cities of Argentina that correspond to points of census cities provided by the Argentinian Institute of Statistics and Censuses (INDEC); since not all cities count with the variables used to create the risk map.

Identification of parameters

A literature review was performed in order to identify the most important factors associated to the prevalence of STH infections and the main ones are summarized in Table 1 [13,20–22,25,26]. These were grouped by categories into socioeconomic (SEC), environmental (EC) and soil characteristics (SC).

An AHP was used as a multiple criteria decision making tool to determine the importance of these factors for the development of STH infections. This methodology introduced by Saaty [24] allows breaking down a complex problem into components and arranging them in hierarchic levels [27]. A pairwise comparison judgment matrix that included the 19 factors identified through the literature review was used to survey experts in the field of STHs and members of the STH Coalition (<http://www.childrenwithoutworms.org/sth-coalition>). The participants were asked to evaluate the relationship between infection with STH and the factors from each of the categories from Table 1 rating them in importance from 1 to 9. The ranking system was as follows: 1 designated equal importance; 3 moderate importance of one vs the other, 5 essential or strong importance, 7 very strong importance, 9 extreme importance and even numbers were designated as intermediate values between two adjacent judgements [28]. From the eigenvalues of this matrix, the coefficient for each group of characteristics was obtained in order to generate three different maps.

The risk associated to SEC, EC and SC was determined by a linear combination of the factors mentioned in Table 1, using the following general equation:

$$Risk = Coef_{F_1} \times F_1 + Coef_{F_2} \times F_2 + \dots + Coef_{F_N} \times F_N$$

Table 1. List of the most important factors associated with soil-transmitted helminth infections grouped into categories.

Socioeconomic characteristics (SEC)	Environmental characteristics (EC)	Soil characteristics (SC)
<ul style="list-style-type: none"> • F1 = % Population with sanitation at home • F2 = % Urban population • F3 = % Population with access to safe drinking water • F4 = % Population with computer • F5 = % Population with overcrowding • F6 = Unemployment rate 	<ul style="list-style-type: none"> • F1 = Normalized Difference Vegetation Index (NDVI) • F2 = Maximum temperature • F3 = Minimum temperature • F4 = Altitude • F5 = Precipitation driest month • F6 = Annual precipitation • F7 = Land Cover type • F8 = Precipitation wettest month • F9 = Mean annual temperature 	<ul style="list-style-type: none"> • F1 = Soil acidity • F2 = Bulk density • F3 = Organic carbon content • F4 = Gypsum content

<https://doi.org/10.1371/journal.pntd.0008000.t001>

Where Risk represents total risk associated to each characteristic, determined by the addition of each factor (F) multiplied by its coefficient (Coef). For SEC the risk was calculated using F1 to F6, as seen in Table 1, for EC using F1 to F9 and for SC using F1 to F4.

Data collection

To characterize the socioeconomic situation of each city, data were obtained from the last national census performed in 2010 (data processed with Rdatam +Sp) [29,30]. The census was performed through a questionnaire and personal interview in all the localities of the country regardless of its population size [29]. This value represents the average of each city; therefore, in big cities, there may be areas of a locality that have values well below average and areas of a locality that have values well above average, and on the contrary small villages could be aggregated to the nearest city for social data. Therefore, not all census data are exactly georeferenced and the resulting map is not continuous.

Environmental and soil data were obtained from WorldClim version 2 [30] at 30 arc seconds (~1000 m) for a large time series (1970 to 2000). Coverage of data from WorldClim is available until -50°S latitude and this is why the maps generated through this software appear to be cut at that point. In addition, the state of vegetation was included through MODIS TERRA Normalized Difference Vegetation Index (NDVI) products (MOD13 A2 version 5) distributed by NASA's Land Processed Distributed Active Archive Center (LP DAAC) were obtained for a ten-year period (2004 to 2014) and the mean of the warmest month was calculated and used as a general vegetation approach (January NDVI mean). Land cover type was obtained from the Latin American SERENA project [31], containing 23 main categories of land cover (needle leaf and broadleaf forest, barren land, grassland, wetlands, crops and urban areas, among others). To evaluate the cover diversity around each city, a diversity approach was calculated using a 9x9 pixel window as the number of different covers observed within that window [32]. Finally, altitude was obtained from the Space Shuttle Radar Topography Mission (SRTM), available from the NASA-Earth data repository (<https://search.earthdata.nasa.gov/>), with a spatial resolution of 3 arc seconds (~90 m) and vertical overall accuracy of 12 m.

Data for the soil characteristics were obtained from a Soil Atlas [33] for Argentina from the National Institute of Agricultural Technology (INTA) at a scale of 1:500,000. This Atlas presents regions of the different soil orders, 12 in total [34], which have different characteristics of acidity, density, organic carbon content and gypsum. These characteristics are categorized as described in Table 2, therefore in Table 3 the characteristics for each order of soil used in this study is described. Only eight of the 12 orders of soil are present in Argentina, therefore the table only lists those that correspond.

GIS analysis

QGIS software was used to develop the risk maps for each characteristic [35]. Risk information is represented through vector layers for the three groups of characteristics in each city of Argentina. A summary of the methodology is presented in Fig 1.

Table 2. Categorization of soil characteristics based on standard measurement and the measurements normalized and used in the current study.

Soil Characteristic	Standard measurement	Normalized for this study
Acidity	>7, 7, <7	1, 0, -1
Bulk density	High, medium, low	1, 0.5, 0
Organic carbon content	Very high, high, medium, low, very low	1, 0.66, 0.5, 0.33, 0
Gypsum content	Yes, no	-1, 0

<https://doi.org/10.1371/journal.pntd.0008000.t002>

Table 3. Categorization of each soil characteristic found in each of the eight types of soil orders found in the territory that comprises Argentina.

Soil Order	Characteristics of each order of soil found in Argentina (Acidity; Bulk density; Organic carbon content; Gypsum content)
Alfisols	-1; 0; 0; 0
Aridisols	1; 0; 0; -1
Entisols	0; 0; 0.5; 0
Inceptisols	0; 0; -1; 0
Mollisols	1; 0.5; 0.66; 0
Oxisols	1; 0; 1; 0
Ultisols	1; 0.5; 0.33; 0
Vertisols	0; 0; 0.33; 0

<https://doi.org/10.1371/journal.pntd.0008000.t003>

Risk model representation

As a first approach, a Red Green Blue (RGB) color risk map was developed to describe the predominant characteristic that increases the risk of STHs. Therefore, the risk associated to SEC is represented in red, to EC it is green and to SC it is blue. The combination obtained from the mix of these colors for each city represents the overall risk associated to STHs so that the preponderance of each characteristic may be observed since although all of these characteristics are important for transmission, we don't know the importance of one over the other.

To evaluate the effectiveness and coherence of the risk models obtained, the maps were compared with the risk from 20 cities with known prevalence for STH, based on different prevalence studies conducted in 11 of the 23 provinces that make up the country. The prevalence studies included (Table 4) are those performed in areas without MDA interventions and conducted at the community or school level, i.e. not in hospital settings. Prevalence data were ranked in four categories based on WHO guidelines [36], with a prevalence of $\leq 10\%$ considered as low, a prevalence between 10% and 20% considered as moderate, a prevalence $\geq 20\%$ and $< 50\%$ considered as high and finally a prevalence $\geq 50\%$ considered as very high.

Then, a machine learning technique, concretely a decision tree (DT) classification model, which tries to discriminate classes of objects according to their characteristics in successive steps, was used to obtain a final STH risk. This technique was applied to discriminate low, middle, high and very high prevalence cities. There are different algorithms that make decisions about which variable contributes most significantly in the final classification [37]. Here, a C50 algorithm was trained with the 20 cities mentioned above, using the three types of characteristics as attributes to discriminate prevalence categories (C5.0 R statistical package) [38,39]. As a measure of model accuracy, the observed error was compared the expected distribution for an ordinal variable.

Results

Risk associated to SEC, EC and SC

Twenty-two experts on STH participated in the survey developed for the AHP in order to determine their importance in relation to the presence of STH infection. The importance of the contribution of each characteristic within each category was obtained with respect to their relationship with STH infection and the remaining variables. That is the relative frequency of response for each value. The characteristics are ordered from least to most important. For SEC, those turned out to be percent population with computer, unemployment rate, percent urban population, percent population with overcrowding, percent population with access to safe drinking water, and percent population with sanitation at home. For EC they were NDVI,

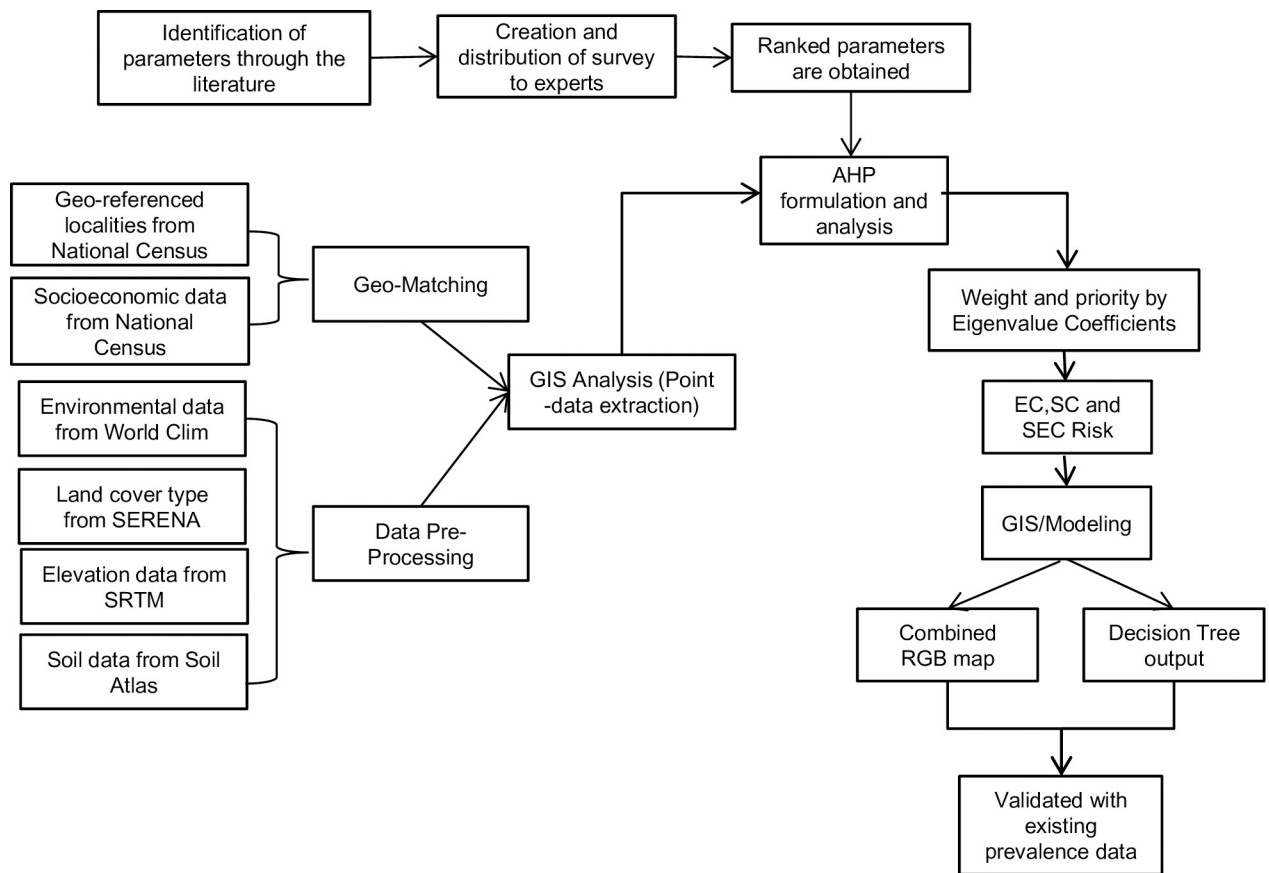


Fig 1. Summary of the methodology used to obtain and validate the predicted map.

<https://doi.org/10.1371/journal.pntd.0008000.g001>

altitude, land cover, precipitation driest month, precipitation wettest month, maximum temperature, annual precipitation, minimum temperature and mean annual temperature. And finally for SC, they were gypsum content, soil acidity, bulk density and organic carbon content.

From the order obtained, a decision matrix was elaborated, from which the relative importance of each variable was analyzed and therefore, a matrix of eigenvalues was obtained with its corresponding coefficients. Fig 2 shows the weight and the final coefficients for indicators of SEC, EC and SC, reflecting their order of importance with respect to their association with the presence of STH infection.

Map of risk areas

Separate risk maps were developed for each group of characteristics individually using the coefficients calculated. The risk map for socioeconomic characteristics (Fig 3) shows that the socioeconomic risk is present in many localities of the country, regardless of geographical location or province.

The environmental characteristics map (Fig 4) shows highest risk in the Northeast Argentine region with a gradient of risk that goes from North to South and the lowest risk observed in the Patagonia Region. The Northeast of Argentina is characterized by a subtropical climate, with abundant vegetation (jungle in Misiones province, tall forests in Chaco and Formosa province, low bushes forests in south Corrientes and north Entre Rios province) and large rivers. Therefore, the map is compatible with the land cover areas and characteristics of the country.

Table 4. Risk values for socioeconomic (SEC), environmental (EC) and soil (SC) characteristics for each city in Argentina with prevalence data for soil-transmitted helminths, observed prevalence categories and predicted categories of risk using the decision tree (DT) algorithm.

City	Province	EC risk	SC risk	SEC risk	Prevalence	DT Prediction
Puerto Iguazú[41]	Misiones	0.8	0.5	0.3	Very high	Low
Aristóbulo del Valle[42]	Misiones	0.9	1.0	0.3	Very high	Very high
Tartagal[43]	Salta	0.9	1.0	0.2	Very high	Very high
Misión Km 6[44]	Salta	0.9	1.0	0.6	Very high	Very high
San Ramón de la Nueva Orán[45]	Salta	0.9	1.0	0.2	Very high	Very high
Bahía Blanca[46]	Buenos Aires	0.5	1.0	0.2	Moderate	Moderate
Berisso[47]	Buenos Aires	0.6	1.0	0.2	Low	Moderate
General Mansilla[48]	Buenos Aires	0.7	0.0	0.2	Low	Low
Coronel Brandsen[49]	Buenos Aires	0.7	1.0	0.3	Moderate	Moderate
Santa Fe[50]	Santa Fe	0.7	1.0	0.3	Very high	Low
Rosario[51]	Santa Fe	0.7	1.0	0.3	Low	Low
Famaillá [52]	Tucumán	0.8	1.0	0.1	High	Low
Villa Burreyacu[53]	Tucumán	0.8	1.0	0.2	Moderate	Low
San Cayetano[54]	Corrientes	0.8	1.0	0.4	High	Low
Santa Ana[55]	Corrientes	0.8	0.6	0.3	Moderate	Low
Añatuya[40]	Santiago del Estero	0.8	1.0	0.3	Low	Low
Pampa del Indio[56]	Chaco	0.9	0.9	0.3	Low	Moderate
Neuquén[57]	Neuquén	0.5	0.9	0.1	Low	Low
Córdoba[58]	Córdoba	0.6	0.9	0.3	Moderate	Low
Mendoza[59]	Mendoza	0.7	0.9	0.0	Low	Low

<https://doi.org/10.1371/journal.pntd.0008000.t004>

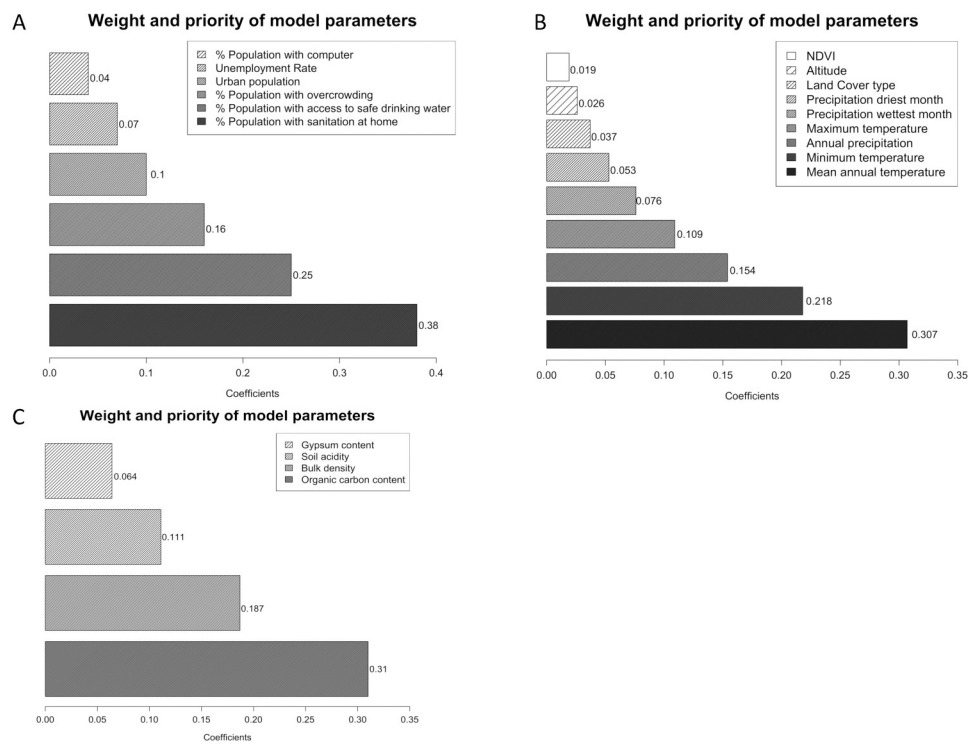


Fig 2. Analytical Hierarchy Process coefficients for each indicator of socioeconomic (A), environmental (B) and soil characteristics (C). These values were obtained through the use of a decision matrix from which the relative importance of each variable was analyzed.

<https://doi.org/10.1371/journal.pntd.0008000.g002>

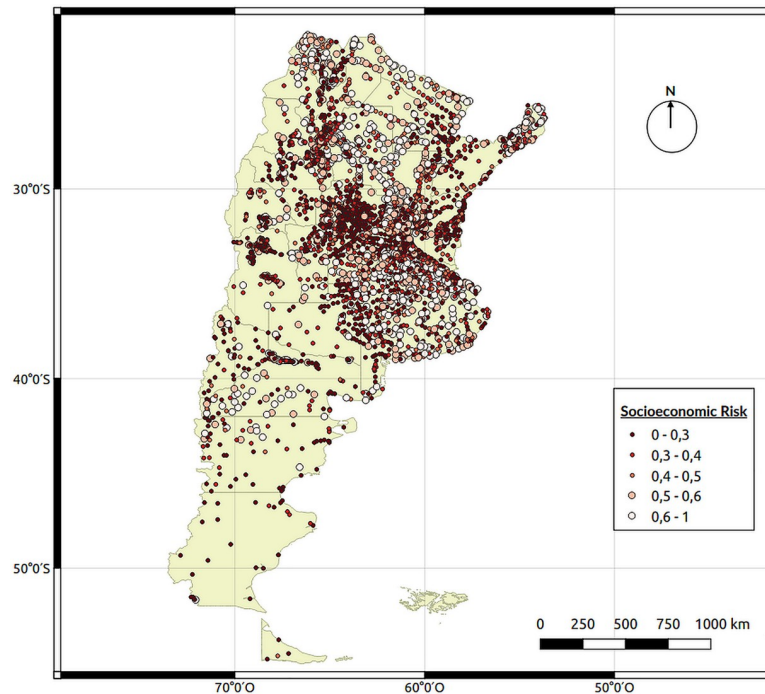


Fig 3. Socioeconomic (SEC) risk map. Each point on the map represents a georeferenced city with the average socioeconomic data obtained from the last national census performed in 2010. Map created using QGIS Geographic Information System. Open Source Geospatial Foundation Project. <http://qgis.osgeo.org>.

<https://doi.org/10.1371/journal.pntd.0008000.g003>

Fig 5 shows the risk map for soil characteristics and although the risk is concentrated on the northern half of Argentina, there are some localities with high risk in the southern provinces of Neuquén, Rio Negro and Chubut. This is because of the four soil characteristics included; the one of most relative importance is the content of organic carbon (0.31). The areas shown as high risk are those areas found alongside rivers that cross the Patagonian region, where the soil is much more fertile.

The final RGB map (Fig 6), which takes into account the risk from the individual socioeconomic, environmental and soil characteristics, shows which of the characteristics are more predominant in each locality based on a color scale. Since all of these characteristics are important for the transmission of STHs and all of them need to be present we can observe that the southern area of the country does not have the environmental characteristics adequate for the development of these parasites (reflecting red-violet tones). In the central area of the country, there are some localities between the provinces of Buenos Aires, la Pampa, Córdoba, San Luis and Santa Fe where all three characteristics are present (reflecting pale pink tones). The northern area of the country shows a predominance of appropriate environmental characteristics (reflecting greener tones) with several localities with the presence of all three of them. Fig 7 shows a closer look from two Northern provinces (Formosa and Chaco) to show the examples of localities with a combination of all three characteristics.

Modeling

The cities analyzed and their prevalence are shown in Table 4, as well as the values of the three risk characteristics modeled for each city, and the prevalence obtained and the predicted results using the DT methodology. The DT obtained by the C50 algorithm identified EC to be

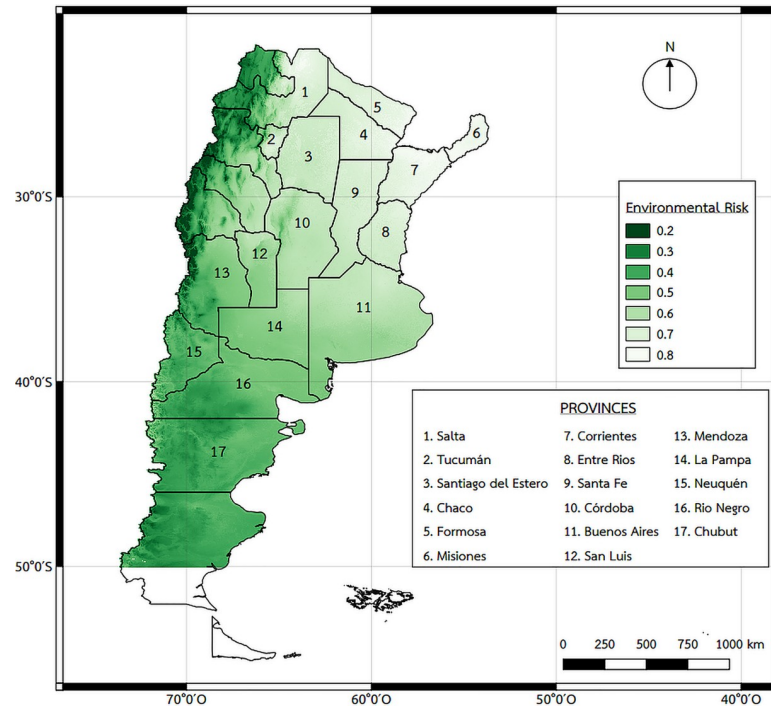


Fig 4. Environmental characteristics (EC) risk map. This data was obtained using different sources such as WorldClim which is available until -50°S latitude, MODIS TERRA NDVI products obtained for a ten-year period (2004 to 2014), the mean of the warmest month calculated through a general vegetation approach (January NDVI mean), the SERENA project for land cover type data and altitude was obtained from the Space Shuttle Radar Topography Mission. Map created using QGIS Geographic Information System. Open Source Geospatial Foundation Project. <http://qgis.osgeo.org>.

<https://doi.org/10.1371/journal.pntd.0008000.g004>

the most important characteristic to classify the 20 cities into the four ordinal categories of risk (expressed as prevalence, Fig 8). The percent of each characteristic used to obtain the final decision tree were as follows: 100% for environmental, 80% for soil and 50% for socio-economic. The predicted results are shown in Table 4 as “DT Prediction”. Using the DT methodology, the expected random error of the ordinal data (four categories) were significantly higher compared with the DT algorithm error (1.25 versus 0.75, $p = 0.0004$).

Through the use of this algorithm, a total of 247 localities from 5 northern provinces were classified as very high risk: 114 localities from Misiones, 54 localities from Formosa, 34 localities from Salta, 23 localities from Chaco and 22 localities from Corrientes. These are all small localities with an average number of 1,143 inhabitants. The total population at very high risk of infection by STHs would be approximately 1,464,235.

The prevalence obtained in each study through the analysis of fecal samples usually matches the one predicted or varies by one order (i.e. moderate instead of low or vice versa). Nonetheless, in four of the cities included in this study the variation between real and predicted was greater. These four studies, from the cities of Puerto Iguazu, Santa Fe, Famailla and San Cayetano all have high risk with respect to the environment and low risk with respect to the socio-economic conditions since the INDEC takes the average of each of the cities. The discrepancy here then could be explained by the real SEC of the populations included in the study, which is really much lower than the one reflected by the index since these studies were conducted in marginal areas of the city with no access to running water or improved sanitation.

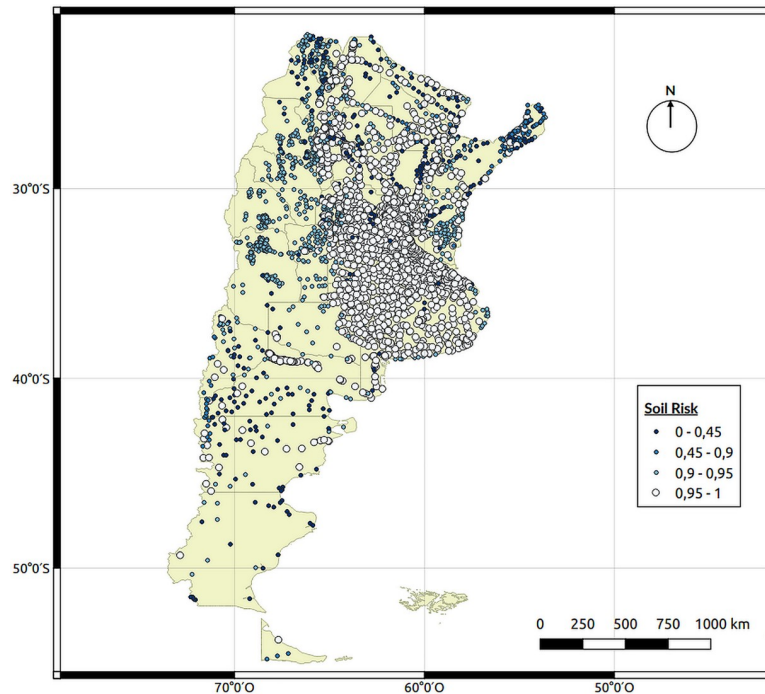


Fig 5. Soil characteristics (SC) risk map. Each point represents a city and the different characteristics of the soil were obtained from a Soil Atlas of Argentina from the National Institute of Agricultural Technology. Map created using QGIS Geographic Information System. Open Source Geospatial Foundation Project. <http://qgis.osgeo.org>.

<https://doi.org/10.1371/journal.pntd.0008000.g005>

Discussion

The lack of prevalence baseline data for STHs in Argentina makes it difficult to determine the need for a control program. A few point prevalence studies conducted in localities from 11 of the 23 provinces that make up the country have shown the presence of all five species of STHs with varying prevalences, ranging from 0.8% to 88.6% [60]. Given the high prevalence found in localities from the Northwest and Northeast of the country and the consequences these infections have on children; it is important to determine the extent of the problem in order to be able to design a tailor-made control program that is implemented either at the national or regional scale. Therefore, the aim of this study was to identify risk areas for the entire country using variables previously associated with the presence of STHs in several studies from other countries, including neighboring Bolivia [13] and Brazil [25].

This is the first risk map for STHs generated specifically for Argentina; it is also one of the few maps generated for this group of parasites that is done at the country scale using socioeconomic, environmental and soil characteristics. The limitation is usually in the availability of census data with socioeconomic characteristics and fortunately for Argentina these were readily available from the last national census. The Analytical Hierarchy Process (AHP) is a common approach used in decision making of complex systems [61–64]. In this study, it has allowed identification of hierarchical relations between separate components (variables identified by a literature review) of a complex problem such as STH transmission. The weight (rank) of the variables within the categories included (socioeconomic, environmental and soil characteristics) were defined by 22 national and international experts in the field of STHs. This result based on expert knowledge may be applied to any geographic region where the SEC, EC and

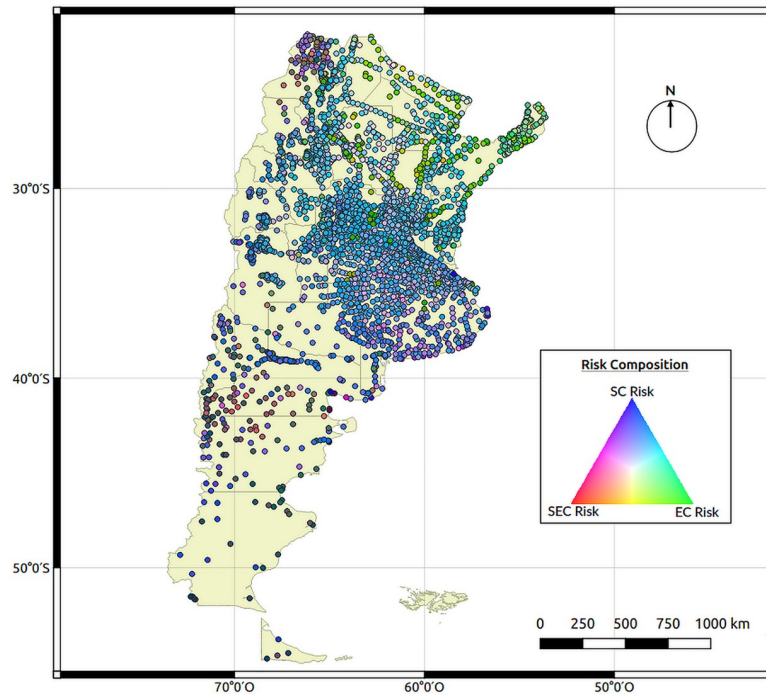


Fig 6. Final RGB map developed to describe the predominant characteristic that increases the risk of soil transmitted helminths, the risk associated to socioeconomic characteristics is red, to environmental characteristics is green and to soil characteristics is blue. Each point represents the combination obtained from the mix of these colors for each city. Map created using QGIS Geographic Information System. Open Source Geospatial Foundation Project. <http://qgis.osgeo.org>.

<https://doi.org/10.1371/journal.pntd.0008000.g006>

SC characteristics are available, attempting to standardize a wide range of risk factors used in several previous studies.

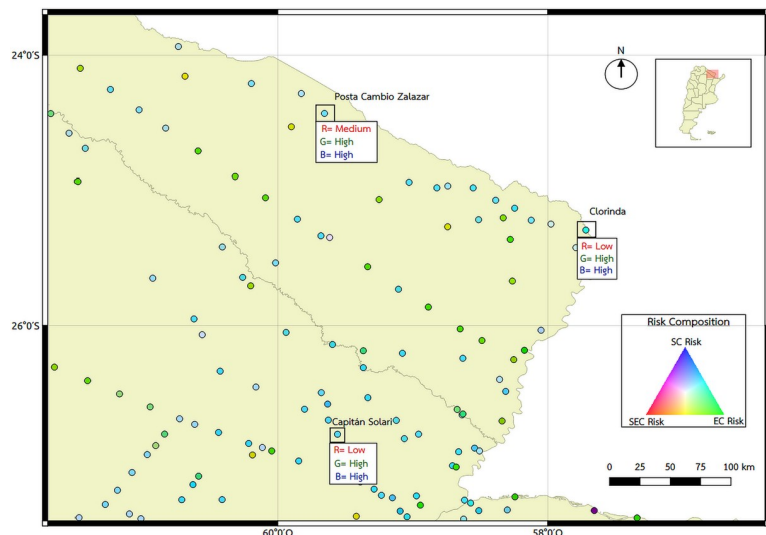


Fig 7. Description of the values of SEC, EC and SC risk of Northern provinces, two localities from Formosa (Posta Cambio Zalazar and Clorinda) and one from Chaco (Capitán Solari). Map created using QGIS Geographic Information System. Open Source Geospatial Foundation Project. <http://qgis.osgeo.org>.

<https://doi.org/10.1371/journal.pntd.0008000.g007>

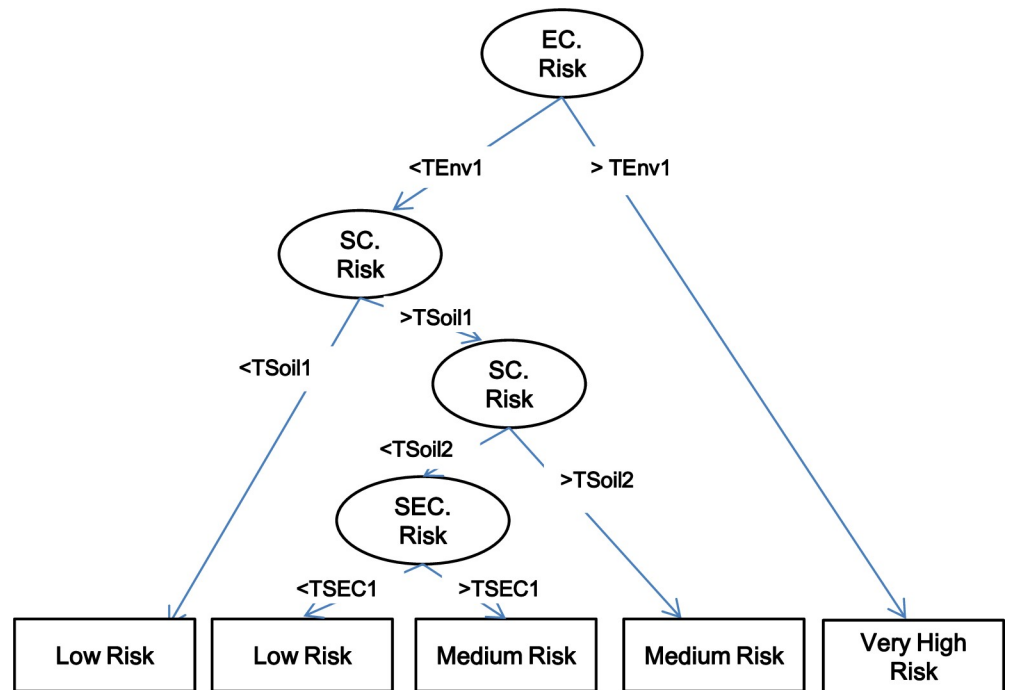


Fig 8. Hierarchical Decision Tree algorithm obtained with the prevalence training data set from the 20 studies published (Table 4). Thresholds (T) with respect to the prevalence categories of risk identified in each node (characteristic) which corresponds to the characteristics included in the study, environment (TEnv), soil (TSoil) and socioeconomic (TSec). Squares are the final classification. Given this is a machine learning technique, the values would change responding to the different training sets offered (observed prevalence data).

<https://doi.org/10.1371/journal.pntd.0008000.g008>

In Argentina, the resulting individual maps show those areas of the country with the appropriate characteristics for the transmission of STHs. Given certain biological characteristics of these parasites: 1) no animal reservoir, transmission from human to human; 2) life cycle with the development of infective stages in the environment; 3) transmission by the fecal-oral route (*A. lumbricoides* and *T. trichiura*) and by larval penetration (hookworms and *S. stercoralis*), it is important to consider these three categories of characteristics in a combined map in order to more accurately assess their presence. In order to obtain a final STH risk assigning weight or coefficients to the three different characteristics used (EC, SC and SEC) in an objective manner, a Decision Tree (DT) algorithm was used. These DT classifiers are included into machine learning techniques where the number and quality of training data sets are crucial. Therefore, increasing the number of sites (cities) will produce more accurate predictions of final STH risk. Nonetheless, the error measure obtained is about 50% lower than those expected by random assignment of the four risk prevalence categories (low, moderate, high, very high)

The combined map generated was validated using prevalence data from the different point studies conducted and in general it was able to correctly predict the prevalence of STHs in 15 of the 20 studies included. This evidences certain limitation with have been previously addressed in other studies, such as the parasitology method used for diagnosis [26], variations of socioeconomic conditions within localities and the unavailability to measure certain soil characteristics using remote sensing (i.e. salt content) [13]. Moreover, uncertainty measurements need to be considered. In the case of the three components used in this study, we can assume that they would be the sum of the error of each individual one. Therefore, the percent

relative error for the social [29,65] and environmental characteristics is around 10% [30,31,66]. On the other hand, the error for the soil is larger since the scale available in the maps is an approximation based on soil order and there is a lack of information on the precise composition of the soil with respect to the variables of interest (gypsum content, soil acidity, bulk density and organic carbon content) [34].

The resulting RGB map was generated so that the contribution of each category of characteristics used can be inferred, some areas have only two of the three categories, i.e. SEC and EC, or EC and SC, and those areas that show a combination of all three would be the ones to prioritize for the conduction of baseline surveys to determine the presence of STHs. According to this map, the top half of the country, from Central Argentina to the North, has localities with the characteristics necessary for the development of these parasites. Using the DT methodology, a total of 1,464,235 people would be at very high risk of infection by STHs. With the current availability of geospatial and environmental data and the tendency of an open data policy, the work scheme presented herein for the development of an STH risk map, is highly reproducible in other areas.

Conclusion

The predictive map generated takes into consideration the combination between socioeconomic, environmental and soil characteristics that are more closely associated with the risk of soil-transmitted helminths as classified by the use of a DT methodology and therefore should serve as a useful tool for guiding the identification of survey areas for the generation of baseline data, detecting hotspots of infection, planning and prioritizing areas for control interventions, and eventually performing post-implementation surveillance activities.

Acknowledgments

We would like to thank Marcelo Scavuzzo and Sofía Lanfri from the Mario Gulich Institute for Higher Space Studies, Teófilo Tabanera Space Center, Córdoba, Argentina for their valuable guidance/input in the early stages of this study. We would also to thank the 22 experts from the STH community, including STH Coalition members, who took the time to answer the survey for this study.

Author Contributions

Conceptualization: Marcelo C. Abril, Ximena Porcasi, María V. Periago.

Data curation: Eliana M. Alvarez Di Fino, Ximena Porcasi, María V. Periago.

Formal analysis: Eliana M. Alvarez Di Fino, María V. Periago.

Funding acquisition: Marcelo C. Abril.

Methodology: Ximena Porcasi, María V. Periago.

Project administration: María V. Periago.

Resources: Jorge Rubio.

Software: Jorge Rubio.

Supervision: Marcelo C. Abril.

Writing – original draft: Eliana M. Alvarez Di Fino, Jorge Rubio, María V. Periago.

Writing – review & editing: Eliana M. Alvarez Di Fino, Marcelo C. Abril, Ximena Porcasi, María V. Periago.

References

1. WHO. Investing to overcome the global impact of neglected tropical diseases: third WHO report on neglected tropical diseases. Geneva, Switzerland, 2015.
2. Krolewiecki AJ, Lammie P, Jacobson J, Gabrielli AF, Levecke B, et al. A public health response against *Strongyloides stercoralis*: time to look at soil-transmitted helminthiasis in full. *PLoS Negl Trop Dis*. 2013; 7(5):e2165. <https://doi.org/10.1371/journal.pntd.0002165> PMID: 23675541
3. WHO. Guideline: preventive chemotherapy to control soil-transmitted helminth infections in at-risk population groups. Geneva, Switzerland. 2017.
4. Campbell SJ, Nery SV, Wardell R, D'Este CA, Gray DJ, et al. Water, Sanitation and Hygiene (WASH) and environmental risk factors for soil-transmitted helminth intensity of infection in Timor-Leste, using real time PCR. *PLoS Negl Trop Dis*. 2017; 11(3):e0005393. <https://doi.org/10.1371/journal.pntd.0005393> PMID: 28346536
5. Campbell SJ, Savage GB, Gray DJ, Atkinson JA, Soares Magalhaes RJ, et al. Water, Sanitation, and Hygiene (WASH): a critical component for sustainable soil-transmitted helminth and schistosomiasis control. *PLoS Negl Trop Dis*. 2014; 8(4): e2651. <https://doi.org/10.1371/journal.pntd.0002651> PMID: 24722335
6. Clasen T, Boisson S, Routray P, Cumming O, Jenkins M, et al. The effect of improved rural sanitation on diarrhoea and helminth infection: design of a cluster-randomized trial in Orissa, India. *Emerg Themes Epidemiol*. 2012; 9(1):7. <https://doi.org/10.1186/1742-7622-9-7> PMID: 23148587
7. O'Reilly CE, Freeman MC, Ravani M, Migele J, Mwaki A, et al. The impact of a school-based safe water and hygiene programme on knowledge and practices of students and their parents: Nyanza Province, western Kenya, 2006. *Epidemiol Infect*. 2008; 136(1):80–91. <https://doi.org/10.1017/S0950268807008060> PMID: 17306051
8. Strunz EC, Addiss DG, Stocks ME, Ogden S, Utzinger J, et al. Water, sanitation, hygiene, and soil-transmitted helminth infection: a systematic review and meta-analysis. *PLoS Med*. 2014; 11(3): e1001620. <https://doi.org/10.1371/journal.pmed.1001620> PMID: 24667810
9. Gamboa MI, Giambelluca LA, Navone GT. Spatial distribution of intestinal parasites in the City of La Plata, Argentina. *Medicina (B Aires)*. 2014; 74(5):363–370. PMID: 25347897
10. Gamboa MI, Navone GT, Orden AB, Torres MF, Castro LE, et al. Socio-environmental conditions, intestinal parasitic infections and nutritional status in children from a suburban neighborhood of La Plata, Argentina. *Acta Trop*. 2011; 118(3):184–189. <https://doi.org/10.1016/j.actatropica.2009.06.015> PMID: 19577532
11. Pullan RL, Bethony JM, Geiger SM, Cundill B, Correa-Oliveira R, et al. Human helminth co-infection: analysis of spatial patterns and risk factors in a Brazilian community. *PLoS Negl Trop Dis*. 2008; 2(12): e352. <https://doi.org/10.1371/journal.pntd.0000352> PMID: 19104658
12. Pullan RL, Kabatereine NB, Quinnell RJ, Brooker S. Spatial and genetic epidemiology of hookworm in a rural community in Uganda. *PLoS Negl Trop Dis*. 2010; 4(6):e713. <https://doi.org/10.1371/journal.pntd.0000713> PMID: 20559556
13. Chammartin F, Scholte RG, Malone JB, Bavia ME, Nieto P, et al. Modelling the geographical distribution of soil-transmitted helminth infections in Bolivia. *Parasit Vectors*. 2013; 6:152. <https://doi.org/10.1186/1756-3305-6-152> PMID: 23705798
14. Chammartin F, Guimaraes LH, Scholte RG, Bavia ME, Utzinger J, et al. Spatio-temporal distribution of soil-transmitted helminth infections in Brazil. *Parasit Vectors*. 2014; 7:440. <https://doi.org/10.1186/1756-3305-7-440> PMID: 25230810
15. Soares Magalhaes RJ, Langa A, Pedro JM, Sousa-Figueiredo JC, Clements AC, et al. Role of malnutrition and parasite infections in the spatial variation in children's anaemia risk in northern Angola. *Geospat Health*. 2013; 7(2):341–354. <https://doi.org/10.4081/gh.2013.91> PMID: 23733295
16. Karagiannis-Voules DA, Biedermann P, Ekpo UF, Garba A, Langer E, et al. Spatial and temporal distribution of soil-transmitted helminth infection in sub-Saharan Africa: a systematic review and geostatistical meta-analysis. *Lancet Infect Dis*. 2015; 15(1):74–84. [https://doi.org/10.1016/S1473-3099\(14\)71004-7](https://doi.org/10.1016/S1473-3099(14)71004-7) PMID: 25486852
17. Owada K, Lau CL, Leonardo L, Clements ACA, Yakob L, et al. Spatial distribution and populations at risk of *A. lumbricoides* and *T. trichiura* co-infections and infection intensity classes: an ecological study. *Parasit Vectors*. 2018; 11(1):535. <https://doi.org/10.1186/s13071-018-3107-y> PMID: 30285906
18. Knopp S, Mohammed KA, Simba Khamis I, Mgeni AF, Stothard JR, et al. Spatial distribution of soil-transmitted helminths, including *Strongyloides stercoralis*, among children in Zanzibar. *Geospat Health*. 2008; 3(1):47–56. <https://doi.org/10.4081/gh.2008.231> PMID: 19021108

19. Chammartin F, Scholte RG, Guimaraes LH, Tanner M, Utzinger J, et al. Soil-transmitted helminth infection in South America: a systematic review and geostatistical meta-analysis. *Lancet Infect Dis*. 2013; 13(6):507–518. [https://doi.org/10.1016/S1473-3099\(13\)70071-9](https://doi.org/10.1016/S1473-3099(13)70071-9) PMID: 23562238
20. Lai YS, Zhou XN, Utzinger J, Vounatsou P. Bayesian geostatistical modelling of soil-transmitted helminth survey data in the People's Republic of China. *Parasit Vectors*. 2013; 6:359. <https://doi.org/10.1186/1756-3305-6-359> PMID: 24350825
21. Khieu V, Schar F, Forrer A, Hattendorf J, Marti H, et al. High prevalence and spatial distribution of *Strongyloides stercoralis* in rural Cambodia. *PLoS Negl Trop Dis*. 2014; 8(6):e2854. <https://doi.org/10.1371/journal.pntd.0002854> PMID: 24921627
22. Oluwole AS, Ekpo UF, Karagiannis-Voules DA, Abe EM, Olamiju FO, et al. Bayesian geostatistical model-based estimates of soil-transmitted helminth infection in Nigeria, including annual deworming requirements. *PLoS Negl Trop Dis*. 2015; 9(4):e0003740. <https://doi.org/10.1371/journal.pntd.0003740> PMID: 25909633
23. Chaiyos J, Suwannatrai K, Thinkhamrop K, Pratumchart K, Sereewong C, et al. MaxEnt modeling of soil-transmitted helminth infection distributions in Thailand. *Parasitol Res*. 2018; 117(11):3507–3517. <https://doi.org/10.1007/s00436-018-6048-7> PMID: 30120589
24. Saaty TL. *The Analytic Hierarchy Process*. New York: McGraw-Hill; 1980.
25. Scholte RG, Schur N, Bavia ME, Carvalho EM, Chammartin F, et al. Spatial analysis and risk mapping of soil-transmitted helminth infections in Brazil, using Bayesian geostatistical models. *Geospat Health*. 2013; 8(1):97–110. <https://doi.org/10.4081/gh.2013.58> PMID: 24258887
26. Ngui R, Shafie A, Chua KH, Mistam MS, Al-Mekhlafi HM, et al. Mapping and modelling the geographical distribution of soil-transmitted helminthiasis in Peninsular Malaysia: implications for control approaches. *Geospat Health*. 2014; 8(2):365–376. <https://doi.org/10.4081/gh.2014.26> PMID: 24893014
27. Vaidya OS, Kumar S. Analytic hierarchy process: An overview of applications. *Eur J Oper Res*. 2006; 169(1):1–29. <https://doi.org/10.1016/j.ejor.2004.04.028>
28. Saaty RW. The analytic hierarchy process—what it is and how it is used. *Math Modelling*. 1987; 9(3–5):161–176. [https://doi.org/10.1016/0270-0255\(87\)90473-8](https://doi.org/10.1016/0270-0255(87)90473-8)
29. Argentina. Instituto Nacional de Estadística y Censos Censo nacional de población, hogares y viviendas 2010: Censo del Bicentenario: resultados definitivos, Seria B n°2. 1st ed. Buenos Aires: Instituto Nacional de Estadística y Censos—INDEC; 2012. 378 pp. ISBN 978-950-896-421-2
30. Fisk SE, Hijmans RJ. *WorldClim 2: new 1-km spatial resolution climate surfaces for global land areas*. *Int J Climatol*. 2017; 37(12):4302–4315. <https://doi.org/10.1002/joc.5086>
31. Blanco PD, Colditz RR, Saldaña GL, Hardtke LA, Llamas RM, et al. A land cover map of Latin America and the Caribbean in the framework of the SERENA project. *Remote Sens Environ*. 2013; 132:13–31. <https://doi.org/10.1016/j.rse.2012.12.025>
32. Porcasi X, Andreo V, Ferral A, Guimarey P, Santini MS, et al. Regional scale environmental variables complementing a Risk Model of Chagas Disease vectorial transmission. San Miguel de Tucumán, Argentina; 2018. p. 1–15.
33. INTA. *Atlas de Suelos de la República Argentina*. Escala 1:500.000 y 1:1.000.000. Buenos Aires: CIRN. Instituto de Evaluación de Tierras. Buenos Aires; 1990. Tomo I, 731 pp. Tomo II, 677 pp.
34. Staff SS. *Soil Taxonomy. A Basis System of Soil Classification for Making and Interpreting Soil Surveys*. Agriculture Handbook No. 436. Second edition. Washington, DC: United States Department of Agriculture. Natural Resources Conservation Center; 1999. 886 pp.
35. QGIS Development Team. *QGIS Geographic Information System*. Open Source Geospatial Foundation Project; 2018. <http://qgis.osgeo.org>.
36. Montresor A, Crompton D.W.T., Hall A, Bundy D.A.P., Savioli L. Guidelines for the evaluation of soil-transmitted helminthiasis and schistosomiasis at community level. A guide for managers of control programmes. Geneva: World Health Organization; 1998. 1–45 p.
37. Lakshmi BN, Raghunandhan GH. Proceeding of the National Conference on Innovations in Emerging Technology. Kongu Engineering College, Perundurai, Erode, Tamilnadu, India. 17&18 February 2011. Institute of Electrical and Electronics Engineers (IEEE); 2011;27–32.
38. Quinlan R. *C4.5: Programs for Machine Learning*. Morgan Kaufmann Publishers; 1993. <http://www.rulequest.com/see5-unix.html>.
39. R Core Team, R: A Language and Environment for Statistical Computing, R Foundation for Statistical Computing, Vienna. 2019 [Online]. Available: <https://www.R-project.org/>
40. Periago MV, Garcia R, Astudillo OG, Cabrera M, Abril MC. Prevalence of intestinal parasites and the absence of soil-transmitted helminths in Anatuya, Santiago del Estero, Argentina. *Parasit Vectors*. 2018; 11: 638. <https://doi.org/10.1186/s13071-018-3232-7> PMID: 30547815

41. Candela-Senti E, Goizueta C, Cabrera M, Muñoz-Antoli C, Periago MV. High prevalence and intensity of hookworm in an indigenous community from Puerto Iguazu, Misiones, Argentina; New Orleans, LA, United States. *American Society of Tropical Medicine and Hygiene*. 2018; 617.
42. Gamboa MI, Kozubsky LE, Costas ME, Garraza M, Cardozo MI, et al. [Associations between geohelminths and socioenvironmental conditions among different human populations in Argentina]. *Rev Panam Salud Publica*. 2009; 26(1):1–8. <https://doi.org/10.1590/s1020-49892009000700001> PMID: 19814875
43. Taranto NJ, Cajal SP, De Marzi MC, Fernandez MM, Frank FM, et al. Clinical status and parasitic infection in a Wichí Aboriginal community in Salta, Argentina. *Trans R Soc Trop Med Hyg*. 2003; 97(5):554–558. [https://doi.org/10.1016/s0035-9203\(03\)80026-3](https://doi.org/10.1016/s0035-9203(03)80026-3) PMID: 15307425
44. Menghi CI, Iuvaro FR, Dellacasa MA, Gatta CL. [Survey of intestinal parasites among an aboriginal community in Salta]. *Medicina (B Aires)*. 2007; 67(6 Pt 2): 705–708. PMID: 18422062
45. Krolewiecki AJ, Ramanathan R, Fink V, McAuliffe I, Cajal SP, et al. Improved diagnosis of *Strongyloides stercoralis* using recombinant antigen-based serologies in a community-wide study in northern Argentina. *Clin Vaccine Immunol*. 2010; 17(1):1624–1630. <https://doi.org/10.1128/CVI.00259-10> PMID: 20739501
46. Costamagna SR, García S, Visciarelli E, Casas N. Epidemiología de las parasitosis en Bahía Blanca (Provincia de Buenos Aires) Argentina—1994–1999. *Parasitol Latinoam*. 2002; 57:103–110. <http://dx.doi.org/10.4067/S0717-77122002000300004>
47. Molina N, Pezzani B, Ciarmela M, Orden A, Rosa D, et al. Intestinal parasites and genotypes of *Giardia intestinalis* in school children from Berisso, Argentina. *J Infect Dev Ctries*. 2011; 5(5):527–534. <https://doi.org/10.3855/jidc.1660> PMID: 21795821
48. Basualdo JA, Córdoba MA, De Luca MM, Ciarmela ML, Pezzani BC, et al. Intestinal parasitoses and environmental factors in a rural population of Argentina, 2002–2003. *Rev Inst Med Trop Sao Paulo*. 2007; 49(4):251–255. <https://doi.org/10.1590/s0036-46652007000400011> PMID: 17823756
49. Zonta ML, Navone GT, Oyhenart EE. Parasitosis intestinales en niños de edad preescolar y escolar: situación actual en poblaciones urbanas, periurbanas y rurales en Brandsen, Buenos Aires, Argentina. *Parasitol Latinoam*. 2007 62(1–2):54–60. <http://dx.doi.org/10.4067/S0717-77122007000100009>.
50. Beltramino D, Lurá MC, Carrera E. El tratamiento antihelmíntico selectivo frente al tratamiento masivo. Experiencia en dos comunidades hiperendémicas. *Rev Panam Salud Pública*. 2003 13(1):10–18. <https://doi.org/10.1590/s1020-49892003000100002> pmid: 12744797
51. Indelman P, Echenique C, Bertorini G, Racca L, Gomez C, et al. Parasitosis intestinales en una población pediátrica de la ciudad de Rosario, Santa Fe, Argentina. *Acta Biochim Clin Latinoam*. 2011; 45(2):329–334.
52. Dib J, Oquilla J, Lazarte SG, Gonzalez SN. Parasitic prevalence in a suburban school of famailla, tucuman, Argentina. *ISRN Microbiol*; 2012:560376. <https://doi.org/10.5402/2012/560376> PMID: 23724325
53. Silvia GL, Christian A, Juana O, Silvia NG. Alarming levels of infection by enteric parasites in children of Burreyacu, Province of Tucuman, Argentina. *Intl J Trop Med*. 2006; 1(1):40–43.
54. Borda CE, Rea MJ, Rosa JR, Maidana C. Intestinal parasitism in San Cayetano, Corrientes, Argentina. *Bull Pan Am Health Organ*. 1996; 30(3):227–233. PMID: 8897723
55. Milano AM, Oscherov EB, Palladino AC, Bar AR. [Children enteroparasitosis in north east Argentine urban area]. *Medicina (B Aires)*. 2007; 67(3):238–242. PMID: 17628910
56. Periago MV, Delgado C, Wasilewsky S, Cabrera M. Intestinal polyparasitism in Pampa del Indio, Chaco Province, Argentina. Baltimore, MD, United States. *American Society of Tropical Medicine and Hygiene*. 2017; 566.
57. Soriano SV, Manacorda AM, Pierangeli NB, Navarro MC, Grenovero MS, et al. Parasitosis intestinales y su relación con factores socioeconómicos y condiciones de habitat en niños de Neuquén, Patagonia, Argentina. In: (FLAP) Federacion Latinoamericana de Parasitología, editor. Bahía, Brasil. 2005;154–165.
58. Guignard S, Arienti H, Freyre L, Lujan H, Rubinstein H. Prevalence of enteroparasites in a residence for children in the Cordoba Province, Argentina. *Eur J Epidemiol*. 2000; 16(3):287–293. <https://doi.org/10.1023/a:1007651714790> PMID: 10870946
59. Bontti S. Enteroparasitosis en el Departamento de Lavalle, Mendoza. *Medicina (B Aires)*. 1999; 59 (Supl III):40–41.
60. Socías ME, Fernandez A, Gil JF, Krolewiecki AJ. [Soil transmitted helminthiasis in Argentina. A systematic review]. *Medicina (B Aires)*. 2014; 74(1):29–36. PMID: 24561837
61. Avila Mogollon RM. El AHP (Proceso Analítico Jerárquico) y su aplicación para determinar los usos de las tierras. El Caso de Brasil (Proyecto GCP/RLA/126). Informe Técnico N°2. The Food and Agriculture

- Organization of the United Nations (FAO). 2000; 65 pp. http://www.fao.org/tempref/GI/Reserved/FTP_FaoRlc/old/proyecto/139jpn/document/3dctos/sirtplan/infotec/2ahptx.pdf
62. Nejati J, Bueno-Marl R, Collantes F, Hanafi-Bojd AA, Vatandoost H, et al. Potential risk areas of *Aedes albopictus* in South-Eastern Iran: A vector of Dengue Fever, Zika, and Chikungunya. *Front Microbiol.* 2017; 8:1660. <https://doi.org/10.3389/fmicb.2017.01660> PMID: 28928720
 63. Chen Z, Chen T, Qu Z, Yang S, Ji X, et al. Use of evidential reasoning and AHP to assess regional industrial safety. *PLoS One.* 2018 13(5):e0197125. <https://doi.org/10.1371/journal.pone.0197125> PMID: 29795593
 64. Cabrera-Barona P, Ghorbanzadeh O. Comparing Classic and Interval Analytical Hierarchy Process Methodologies for Measuring Area-Level Deprivation to Analysis Health Inequalities. *Int J Environ Res Public Health.* 2018; 15(1):140. <https://doi.org/10.3390/ijerph15010140>
 65. Instituto Nacional de Estadística y Censos—I.N.D.E.C. Tablas de errores de muestreo. Bases trimestrales de microdatos de la Encuesta Permanente de Hogares (EPH). Tercer trimestre de 2014 en adelante. 1st ed. Ciudad Autónoma de Buenos Aires: Instituto Nacional de Estadística y Censos—INDEC. 2015; 46 pp.
 66. Hijmans RJ, Cameron SE, Para JL, Jones PG, Jarvis A. Very high resolution interpolated climate surfaces for global land areas. *Int J Climatol.* 2005; 25(15):1965–1978.