Soil-transmitted helminthiasis in mainland China from 2016 to 2020: a population-based study

Hui-Hui Zhu,[a] Ji-Lei Huang,[a] Chang-Hai Zhou,[a] Ting-Jun Zhu,[a] Jin-Xin Zheng,[b] Mi-Zhen Zhang,[a] Men-Bao Qian,[a,b,*] Ying-Dan Chen,[a] and Shi-Zhu Li[a,b,**]

[a]National Institute of Parasitic Diseases, Chinese Center for Disease Control and Prevention (Chinese Center for Tropical Diseases Research); NHC Key Laboratory of Parasite and Vector Biology; WHO Collaborating Center for Tropical Diseases; National Center for International Research on Tropical Diseases, Shanghai 200025, China
[b]School of Global Health, Chinese Center for Tropical Diseases Research-Shanghai Jiao Tong University School of Medicine, Shanghai, China

The Lancet Regional Health - Western Pacific 2023;■: 100766

Published Online XXX
https://doi.org/10.1016/j.lanwpc.2023.100766

Summary

Background Soil-transmitted helminthiasis is epidemic in China and many other countries of the world, and has caused substantial burdens to human health. We conducted successive national monitoring in China from 2016 to 2020 to analyze the prevalence, changing trends, and factors influencing soil-transmitted helminthiasis, which provided a reference for future control strategies.

Methods Soil-transmitted helminth monitoring was carried out in 31 provinces (autonomous regions or municipalities, herein after referred to as "provinces") throughout China. Each province determined the number and location of monitoring sites (counties), and a unified sampling method was employed. At least 1,000 subjects were investigated in each monitoring county. Stool samples were collected and the modified Kato-Katz thick smear method was employed for stool examination. Infection data and the details of factors influencing soil-transmitted helminthiasis from 2016 to 2020 were collected from national monitoring sites. Additional influencing factors such as environment, climate and human activities were obtained from authoritative websites. Prevalence of soil-transmitted helminths was presented by species, province, sex, and age group. ArcGIS software was used to conduct spatial autocorrelation and hotspot analysis on the infection data. A Poisson distribution model and SaTScan software were used to analyze the infection data with retrospective spatiotemporal scan statistics. A database was built by matching village-level infection rate data with influencing factors. Subsequently, machine learning methods, including a Linear Regression (LR), a Random Forest (RF), a Gradient Boosted Machine (GBM), and an Extreme gradient boosting (XGBOOST) model was applied to construct a model to analyze the main influencing factors of soil-transmitted helminthiasis.

Findings The infection rates of soil-transmitted helminths at national monitoring sites from 2016 to 2020 were 2.46% (6,456/262,380), 1.78% (5,293/297,078), 1.29% (4,200/326,207), 1.40% (5,959/424,766), and 0.84% (3,485/415,672), respectively. The infection rate of soil-transmitted helminths in 2020 decreased by 65.85% compared to that in 2016. From 2016 to 2020, the infection rate of soil-transmitted helminthiasis was relatively high in southern and southwestern China, including Hainan, Yunnan, Sichuan, Guizhou, and Chongqing. In general, the infection rate was higher in females than in males, with the highest rate in the population aged 60 years and above, and the lowest in children aged 0–6 years. Global autocorrelation and hotspot analyses revealed spatial aggregation in both the national and local distribution of soil-transmitted helminthiasis in China from 2016 to 2020. The hotspots were concentrated in southwestern China. The spatiotemporal scanning analysis revealed aggregation years from 2016 to 2017 located in southwestern China, including Yunnan, Sichuan, Chongqing, Guizhou and Guangxi. The RF model was the best fit model for the infection rate of soil-transmitted helminths in China. The top six influencing factors of this disease in the model were landform, barefoot farming, isothermality, temperature seasonality, year, and the coverage of sanitary toilets.

*Corresponding author. National Institute of Parasitic Diseases, Chinese Center for Disease Control and Prevention (Chinese Center for Tropical Diseases Research); NHC Key Laboratory of Parasite and Vector Biology; WHO Collaborating Center for Tropical Diseases; National Center for International Research on Tropical Diseases, Shanghai 200025, China.
**Corresponding author. National Institute of Parasitic Diseases, Chinese Center for Disease Control and Prevention (Chinese Center for Tropical Diseases Research); NHC Key Laboratory of Parasite and Vector Biology; WHO Collaborating Center for Tropical Diseases; National Center for International Research on Tropical Diseases, Shanghai 200025, China.
E-mail addresses: qianmb@nipd.chinacdc.cn (M.-B. Qian), lisz@chinacdc.cn (S.-Z. Li).

Interpretation The overall infection rate of soil-transmitted helminths in China showed a decreasing trend from 2016–2020 due to the implementation of control measures and the economic boom in China. However, there are still areas with high infection rates and the distribution of such areas exhibit spatiotemporal aggregation. As a strategic next step, control measures should be adjusted to local conditions based on the main influencing factors and the prevalence of different sites to aid in the control and elimination of soil-transmitted helminthiasis.

Funding This research was funded by the National Key Research and Development Program of China (Grant Nos. 2021YFC2300800 and 2021YFC2300804) and the National Natural Science Foundation of China (Grant No. 32161143036).

Keywords: Soil-transmitted helminthiasis; Infection rate; Influencing factors; China

Research in context

Evidence before this study

We searched PubMed, Web of Science and CNKI (China National Knowledge Infrastructure) for studies on soil-transmitted helminthiasis in China, published in English and Chinese, up to October 11, 2022. We used a combination of search terms, including "soil-transmitted helminthiasis or soil-transmitted helminth", "China", "spatial aggregation", and " influencing factors". We found that previously the prevalence of soil-transmitted helminth in China was obtained using three national surveys in 1988–1990, 2001–2004, and 2014–2015, with prevalence of 53.58%, 19.56%, and 4.49%, respectively. The most recent data were acquired by national monitoring from 2016 to 2019. No spatial aggregation research based on the national data of soil-transmitted helminthiasis was found. The influencing factors of soil-transmitted helminthiasis have been studied widely, and factors such as safe drinking water, sanitary toilets, hygiene habits, economic status, and sanitation were determined to be important.

Added value of this study

To the best of our knowledge, this study provided a first longitudinal comparison on the prevalence of soil-transmitted helminthiasis in China from 2016 to 2020. It also carried out the first spatial aggregation analysis, and defined the in-depth relationships between the influencing factors and infection rates of soil-transmitted helminths based on national data.

Implications of all the available evidence

The overall infection rate of soil-transmitted helminths in China showed a decreasing trend from 2016 to 2020, and spatial aggregation existed in both the national and local distributions during that period. The top six factors influencing the disease were landform, barefoot farming, isothermality, temperature seasonality, year, and coverage of sanitary toilets. The in-depth relationship between factors and the infection rates of soil-transmitted helminths was illustrated. All such data provided important information for guiding the prevention and control of soil-transmitted helminthiasis in China.

Introduction

Soil-transmitted helminths mainly include the hookworms, *Ascaris lumbricoides* and *Trichuris trichiura*.[1,2] They live in the intestine of (human) hosts and their eggs are passed in the faeces. Eggs initially undergo a period of maturation for 2–3 weeks, and then become infectious eggs/larvae. Soil-transmitted helminths have a direct life cycle, which requires no intermediate hosts or vectors. Thus, eggs/larvae could be transmitted through direct contact or during food preparation because of poor hygiene.[3] Soil-transmitted helminths usually cause asymptomatic or chronic infections, acquiring nutrition from their hosts.[4] Because children and pregnant women suffer more from malnutrition, they are more susceptible to soil-transmitted helminthiasis.[5] Severe infection can damage the gut, leading to malnutrition, developmental disorders, anemia, poor immunity in children,[6,7] and poor delivery prognosis in pregnant women.[8] About 2.5 billion people in developing countries worldwide are infected with at least one species of soil-transmitted helminth, among which 10% of the population is infected with two or more species.[9] In 2015, approximately 360 million people in East Asia were infected with soil-transmitted helminths, accounting for a quarter of the population.[10]

Soil-transmitted helminthiasis was once highly prevalent in China.[1,11] A national survey on the distribution of human parasites (the first national survey) was carried out in China in 1988–1992, in which the infection rate of soil-transmitted helminths was 53.58% and the estimated number of infected individuals was 536 million.[11] After the first national survey, large-scale population-based deworming was implemented, and the infection rate of soil-transmitted helminths significantly decreased.[1] In 2001–2004, a national survey on the status of human key parasitic diseases in China (the second national survey) was conducted, and the prevalence of soil-transmitted helminths dropped to 19.56%, with an estimated number of infected individuals of 129 million.[12] After the second national survey, comprehensive demonstration plots for STHs were established and control strategies of "health education as leadership, control of infection source as priority" were carried out.[1] With prevention and control efforts, and the improvement of living standards, the infection rate of soil-transmitted helminthiasis continues to decrease. The 2015 National Survey on the Current Status of Major Human Parasitic Diseases in China (the third national survey) showed that the national soil-transmitted helminth infection rate was 4.49%, and 29.12 million people were estimated to be infected.[13]

Between 2006 and 2015, a total of 22 counties from 22 provinces were monitored for soil-transmitted helminths each year.[14] In 2016, a nationwide soil-transmitted helminthiasis monitoring system was established with the support of the government, and the monitoring coverage has expanded greatly each year.[14] The monitoring system aims to understand the endemic status and factors influencing soil-transmitted helminthiasis to provide reference for the formation of control strategies and the evaluation of control effects. The monitoring system was implemented by annual field surveys and nationwide data reports. Each year since 2016, national monitoring has been carried out in 31 provinces following the Operation Manual,[15] and the data are reported annually to the Chinese CDC. After 5 years of monitoring and disease control, the epidemic status of soil-transmitted helminthiasis have changed.[16–20] Along with an improved ecosystem and a rapidly-developing economy, the factors influencing transmission have also changed.

Previous studies on the factors influencing soil-transmitted helminthiasis were mostly based on individual level data. Cross-sectional surveys and logistic regression models found that factors such as safe drinking water, sanitary toilets, hygiene habits, economic status, and sanitation were the main influencing factors, and these findings has played a positive role in guiding prevention and control efforts.[21–26] To further understand the prevalence of soil-transmitted helminthiasis, and the factors influencing this disease, this study analyzed Chinese national monitoring data from 2016 to 2020.

Methods

Data sources

National infection data of soil-transmitted helminths

The main data source for this study was field data from the national monitoring system for soil-transmitted helminths, dated 2016 to 2020. The data were collected between June and October in each year. Soil-transmitted helminth monitoring was carried out in 31 provinces across China.[15] Considering the representativeness of each location and the feasibility of the work, each province determined the number and location of monitoring sites (counties), such that 10%–15% of counties, including those with high, medium, and low endemic status, were covered each year. After the determination of monitoring sites, a standardized sampling method was applied. Each monitoring site was divided into five areas according to geographical location, including east, west, south, north, and central. One administrative village in one town was randomly selected from each area and approximately 200 permanent residents were selected from each administrative village each year. A total of 1,000 individuals were surveyed at each monitoring site. Stool samples were collected and examined using the modified Kato-Katz thick smear method (two smears for a single sample) to detect the eggs of soil-transmitted helminths.[27,28]

Additional variables

Five environmental variables, including elevation, slope, Normalized Difference Vegetation Index (NDVI), land cover, and landform were included in the analyses. All variables were obtained from the SRTM (The NASA Shuttle Radar Topographic Mission, http://srtm.csi.cgiar.org/) website, except for the landform data, which were obtained from the survey results of national monitoring.

There were 19 climate variables (named Bio1-Bio19) obtained from the WorldClim website (http://www.worldclim.org). There were also seven humanistic variables, among which the human influence index (HII) and the human footprint (HFP) data were downloaded from the SDAC (Socioeconomic Data and Applications Center, https://sedac.ciesin.columbia.edu) website. The spatial resolution of the data obtained from the websites was 5 km × 5 km. The other five variables, including the main industries, annual per capita net income of farmers, types of safe drinking water, the coverage of sanitary toilets, and the frequency of barefoot farming for each village were obtained from the survey data of the national monitoring system.

Spatial data

The Chinese administrative map used in this study was obtained from the official website of the Ministry of Civil Affairs of the People's Republic of China (http://xzqh.mca.gov.cn/map). The figure number is GS (2022) 1873.

Data analysis

Infection rate analysis

SAS 9.2 software (SAS Institute Inc., Cary, NC, USA) was used to analyze the national monitoring data of soil-transmitted helminthiasis from 2016 to 2020. The total infection rate, as well as the infection rate by species, province, sex, and age group, were obtained. Comparisons between infection rates were performed using a χ^2 test, with $\alpha = 0.05$ as the test level.

Spatiotemporal aggregation analysis

ArcGIS software[29] was used to correlate the infection rates of each province with their spatial location, to build a spatial distribution database for spatiotemporal aggregation analysis.

Spatial autocorrelation analysis was performed using Moran's *I* statistic as the evaluation index to determine whether soil-transmitted helminthiasis exhibited spatial aggregation. To identify the clustering locations with high values (hot spots) and low values (cold spots) of statistically significant soil-transmitted helminth infection rates, hot spot analysis (Getis-Ord *Gi**) was performed using the hotspot analysis tool in ArcGIS. The location of high value clustering areas of soil-transmitted helminthiasis was subsequently determined.

Retrospective spatiotemporal scan statistics were analyzed using SaTScan with a Poisson distribution model.[30–32] To calculate the log-likelihood ratio (LLR) and the corresponding *P* value and relative risk (RR) for different scanning windows, the maximum radius of the scan analysis was set at 50% of the population at risk, the time scale was set as 50% of the study period, the time unit was years, and the number of Monte Carlo simulations was 999. When $P < 0.05$, the larger the LLR, the larger the risk ratio of soil-transmitted helminth infections in the scanning window, and the more likely that the window represents the aggregation area.

Analysis of influencing factors

Village-level soil-transmitted helminth infection rates were calculated according to the monitoring database and then matched with the map layer along with environmental, climate, and humanistic variables to form the database for infection rates and influencing factors of soil-transmitted helminthiasis in 2016–2020. The database included 7,941 village-level data entries and 35 variables. The details of the variables are shown in Table 1.

R software (4.2.1) was then used to analyze the factors influencing soil-transmitted helminthiasis, according to the established database. The Caret (Classification and Regression Training) package in R was used for model construction and calculation. The recursive feature elimination (RFE) method was used to screen the model variables.[33] Considering the model functions and the research demands, Linear Regression (LM), Random Forest (RF), Gradient Boosted Machine (GBM), and Extreme Gradient Boosting (XGBOOST) models were selected as the machine learning classification algorithms. First, the database for this study was subdivided into a 70% training dataset and a 30% testing dataset. The training set was then divided into five cross-validation subsets, of which four subsets were used to build the model and the remaining subset was used to evaluate the strengths and weaknesses of the model. This continuous learning process was iterated approximately 1000 times to obtain the model parameters. The model built on the training set was then evaluated for its performance. The testing set performance was predicted as well. Finally, the optimal model was selected to calculate the importance of each variable and the marginal effect of the model. The overall evaluation of the model was made using three indicators: mean absolute error (MAE), root mean squared error (RMSE) and the coefficient of determination (R-square, R^2).

Ethical statement

This article was based on monitoring data from the National Institute of Parasitic Diseases (NIPD) at the Chinese Center for Disease Control and Prevention (China CDC). The study obtained approval from the Institutional Ethical Review Committee of NIPD (document No. 2021006). No personal information was disclosed. Oral informed consent for publication was obtained from all participants. For participants under 18 years of age, oral informed consent was obtained from their guardian.

Role of the funding source

The funding sources for this study did not influence nor participate in study design, data collection, data analysis, data interpretation, or drafting of the manuscript. The corresponding author had full access to all of the data and final responsibility for the decision to submit the study for publication.

Results

Prevalence of soil-transmitted helminthiasis in China from 2016 to 2020

Overview

The infection rates of soil-transmitted helminths at national monitoring sites from 2016 to 2020 were 2.46% (6,456/262,380), 1.78% (5,293/297,078), 1.29% (4,200/326,207), 1.40% (5,959/424,766), and 0.84% (3,485/415,672), respectively. Thus, the overall infection rate decreased over the period (Fig. 1; Supplementary Table S1). The infection rate was decreased by 65.85% in 2020 compared to that in 2016. The infection rate of hookworm was the highest during those five years, followed by *A. lumbricoides*, and *T. trichiura*. Over 5 years, the infection rates of hookworm, *A. lumbricoides* and *T. trichiura* decreased by 61.94%, 77.65%, and 64.44%, respectively.

Infection rates by region

From 2016 to 2020, the infection rates of soil-transmitted helminths in China were high at

Variables	Median	Interquartile range (IQR)	Details
Bio1	17.5	12.3–20	Annual mean temperature (°C)
Bio2	10	7.9–11.6	Mean diurnal range (max–min temp) (°C) (%)
Bio3	27	24–30	Isothermality
Bio4	8861	7390–10757	Temperature seasonality
Bio5	33.6	31.3–34.9	Max temperature of the warmest month (°C)
Bio6	−0.7	−10.7–5.1	Min temperature of the coldest month (°C)
Bio7	34.2	29–41.6	Temperature annual range (°C)
Bio8	25.9	24–28.6	Mean temperature of the wettest quarter (°C)
Bio9	6	−1.8–10.8	Mean temperature of the driest quarter (°C)
Bio10	28.4	25.3–29.9	Mean temperature of the warmest quarter (°C)
Bio11	5.6	−2.2–9.7	Mean temperature of the coldest quarter (°C)
Bio12	888	619–1297	Annual precipitation (mm)
Bio13	200	155–244	Precipitation of the wettest month (mm)
Bio14	9	3–21	Precipitation of the driest month (mm)
Bio15	79	60–100	Precipitation seasonality (mm)
Bio16	501	377–599	Precipitation of the wettest quarter (mm)
Bio17	33	13–92	Precipitation of the driest quarter (mm)
Bio18	473	370–556	Precipitation of the warmest quarter (mm)
Bio19	34	13–93	Precipitation of the coldest quarter (mm)
Income	9500	5500–14705	Annual income of whole family in each village
Fraction of sanitary toilets	63.24	1.761–99.558	Fraction of sanitary toilets in each village
Slope	1.12	0.2513–4.7662	
NDVI	0.36	0.2228–0.5432	Normalized Difference Vegetation Index
Soil moisture	88.96	37.9885–115.4698	
Landcover	12	9–12	
HII	26	18–30	Human influence index
HFP	40	28–46	Human footprint
POPdensity	252.76	92.61–596.17	
GDP	914.7	333.8–2561.4	Gross Domestic Product in each village
Elevation	198	57–736	
Year	Categorial variable	2016, 2017, 2018, 2019, 2020	Year of investigation
Landform	Categorial variable	Plain, Basin, Hill, Mountain, Plateau	Landform of each village
Drinking.water	Categorial variable	Tap water, Well water, Pothole water, Rivers and lakes, Others	Main source of drinking water in each village
Industry	Categorial variable	Farming, Animal raising Fishery, Trade, Forestry	Main industry of each village
Barefoot farming	Categorial variable	Yes, Sometimes, No	How often do people farming barefooted in each village

Table 1: **Independent variables used in machine learning.**

monitoring sites in Hainan, Yunnan, Sichuan, Guizhou, and Chongqing (Fig. 2). In 2016, Yunnan had the highest infection rate (14.13%), followed by that in Guizhou (9.04%), Sichuan (8.94%) and Hainan (4.87%). In 2020, Hainan province exhibited the highest infection rate (6.34%), followed by Yunnan (5.80%), Sichuan (3.66%), and Guizhou (1.81%) (Fig. 2; Supplementary Table S2).

Compared to 2016, the infection rate in Hainan province in 2020 increased by 30.07%, while that in Yunnan, Sichuan, Guizhou and Chongqing decreased by 59.97%, 59.02%, 79.96%, and 71.09%, respectively. No soil-transmitted helminth infections were observed in Anhui Province in 2016. Similarly, no soil-transmitted helminth infections were found in Shanxi, Shanghai, or Heilongjiang in 2017, Beijing, Shanghai or Heilongjiang in 2018, Beijing, Shanghai, Hebei or Inner Mongolia in 2019, or Shanxi, Beijing, Shanghai, Hebei, Heilongjiang or Inner Mongolia in 2020.

Infection rates by sex

Soil-transmitted helminth infection rates decreased in males and females from 2016 to 2020 (Fig. 3; Supplementary Table S3). However, infection rates were higher among females than males (Fig. 3) ($P < 0.001$ in all years except 2019). The infection rates in males and females decreased by 65.76% and 66.30% over the five years, respectively.

Infection rates by age groups

From 2016 to 2020, the infection rate was the highest among people aged 60 years and above, followed by the age group of 45–59 year olds. Children aged 0–6 years had the lowest infection rates (Fig. 4; Supplementary

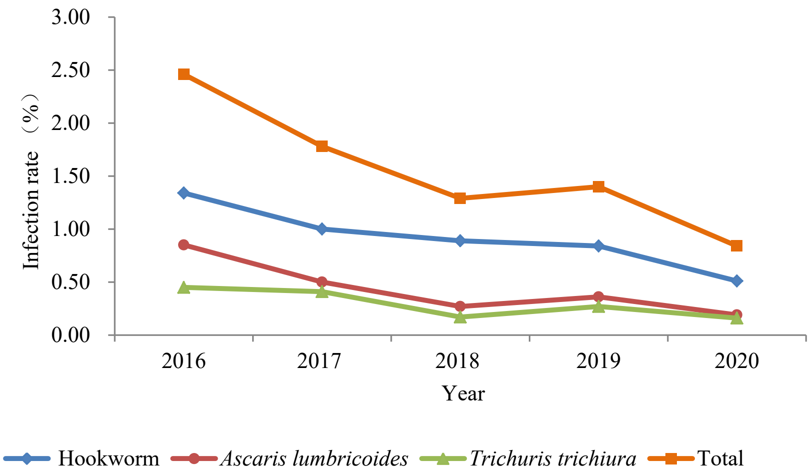

Fig. 1: Prevalence of soil-transmitted helminths from 2016 to 2020 in national monitoring.

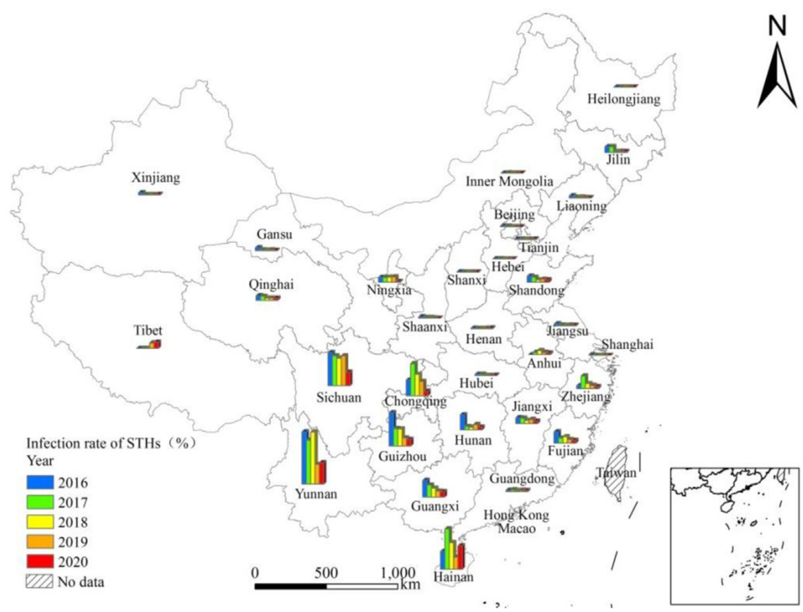

Fig. 2: Region distribution of soil-transmitted helminths from 2016 to 2020 in national monitoring.

Table S1). The infection rates decreased in all age groups over the 5 years. Over the 5 years, the infection rates of 0–6-year-olds, 7–14-year-olds, 15–44-year-olds, 45–59-year-olds, and over 60 years old decreased by 73.57%, 70.95%, 64.28%, 66.80%, and 66.13%, respectively.

Analysis of spatiotemporal aggregation of soil-transmitted helminthiasis in China

Global spatial autocorrelation analysis

The results of global autocorrelation analyses showed that the Moran's I value of the distribution of soil-transmitted helminthiasis in China from 2016 to 2020 was greater than 0, with a Z value between 0.21 and 0.34, and a P value less than 0.01 (Table 2). These findings suggested the existence of spatial clustering for the distribution of soil-transmitted helminthiasis in China over the 5 years.

The results of hotspot analysis showed that local spatial aggregation existed in the distribution of soil-transmitted helminths in China from 2016 to 2020. In 2016, the hotspots were distributed in the four provinces of Sichuan, Guizhou, Yunnan, and Guangxi. In 2017–2018, the hotspots were distributed in the five provinces of Sichuan, Guizhou, Yunnan, Chongqing and Hainan. The hotspots in 2019 were distributed in Sichuan, Guizhou, Yunnan, and Chongqing, while in

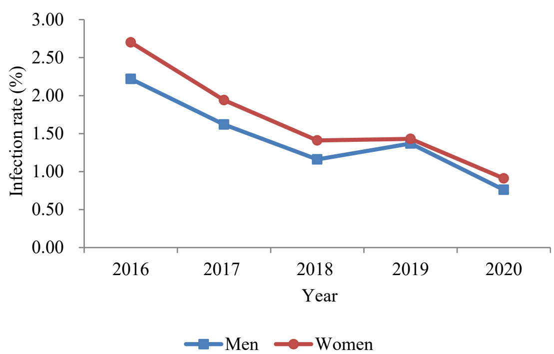

Fig. 3: Gender distribution of soil-transmitted helminths from 2016 to 2020 in national monitoring.

2020, they were distributed in Sichuan, Yunnan, and Hainan. The cold spots in 2016–2017 and 2019 were distributed in Beijing, Hebei and Tianjin. No cold spots were observed in 2018 and 2020 (Fig. 5).

Spatiotemporal aggregation analysis

The results of the retrospective spatiotemporal scanning analysis revealed clustering in the distribution of soil-transmitted helminthiasis in China during 2016–2017, suggesting that there was spatiotemporal aggregation for the distribution. The aggregation was located in southwestern China, which included Yunnan, Sichuan, Chongqing, Guizhou, and Guangxi. The aggregation center was located at 24.9753° longitude and 101.4870° latitude, the aggregation radius was 845.33 km, the number of observations was 8242, the expected value was 1464.52, and the RR (relative risk) was 7.89 [LLR (log-likelihood ratio) = 8541.40, $P < 0.01$] (Fig. 6).

Analysis of the factors influencing soil-transmitted helminthiasis

Variable screening and model comparison

The RFE results showed that the RMSE value of the model gradually decreased with an increase in the number of variables in the model. That is, the accuracy of the model gradually increased and was the highest when all 35 variables were included.

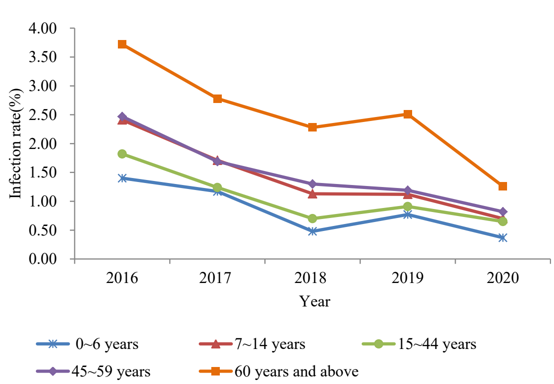

Fig. 4: Age group distribution of soil-transmitted helminths from 2016 to 2020 in national monitoring.

All variables were included to build the machine learning model, with a total of four models (LR, RF, GBM and XGBOOST) being built. By comparing the R^2, RMSE and MAE values of the testing sets of the four models, we found that the RF model had the largest R^2 and the smallest RMSE and MAE values (Table 3). Thus, the RF model was considered the optimal model for fitting the infection rate data of soil-transmitted helminths in China.

Importance and marginal effects of variables

All variables were included in the RF machine learning model, and the importance of each was compared by drawing its RMSE loss diagram. The six most important variables were landform, barefoot farming, isothermality (Bio3), temperature seasonality (Bio4), year, and coverage of sanitary toilets (Fig. 7).

Based on the RF model, the partial dependence plots (PDPs) of important variables and infection rates were plotted, and the RR values of the effects of different types of categorical variables on infection rates were calculated. The PDPs showed that the predicted infection rates of plateaus and mountains in terrain variables were significantly higher than those of the other three terrains (Fig. 8-A). Compared to plains, mountains and plateaus were risk factors for soil-transmitted helminth infections, and their RR values were 1.61 and 1.18, respectively (both $P < 0.05$). For the barefoot farming variable, the predicted infection rates of laborers who went barefoot, sometimes barefoot, and never barefoot decreased gradually (Fig. 8-B). Compared with never going barefoot, sometimes and often going barefoot were risk factors for soil-transmitted helminth infection, with RR values of 1.53 and 2.55, respectively. For the year variable, the predicted infection rate for 2020 was significantly lower than the other years, and the infection rates in 2016–2019 were significantly different from those in 2020 ($P < 0.0001$) (Fig. 8-E).

The Bio3 variable (isothermality) is the percentage of the average annual temperature in the annual temperature range. When Bio3 was below 25%, the infection rate of soil-transmitted helminths decreased slowly with an increase in Bio3. When the Bio3 value was between 25% and 29%, there was no obvious change in the infection rate with changes in Bio3. However, when the Bio3 value was between 30% and 52%, the infection rate of soil-transmitted helminths increased rapidly with an increase in Bio3 (Fig. 8-C).

Bio4 represents the temperature seasonality standard deviation. At Bio4 < 10,000, the infection rate of soil-transmitted helminths decreased with an increase in Bio4. When Bio4 > 10,000, the infection rate did not change significantly with an increase in Bio4, but rather, stabilized at a low level (less than 0.5%) (Fig. 8-D).

Year	Moran 's *I*	Expected Moran's *I* value	Variance	Z value	P value
2016	0.343543	−0.034483	0.004885	5.408657	0.000000
2017	0.248602	−0.034483	0.005361	3.866203	0.000111
2018	0.268392	−0.035714	0.00491	4.340001	0.000014
2019	0.306901	−0.033333	0.007128	4.029911	0.000056
2020	0.212669	−0.033333	0.007298	2.879539	0.003983

Table 2: **Global spatial autocorrelation analysis of soil-transmitted helminths in China from 2016 to 2020.**

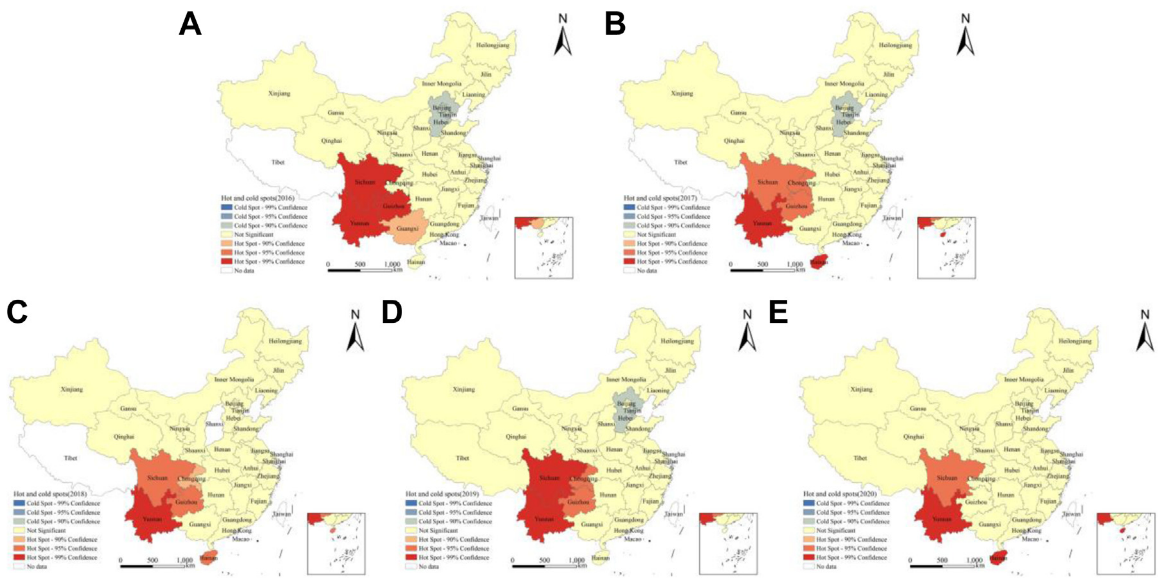

Fig. 5: Hotspot analysis of distribution of soil-transmitted helminth in China in years of 2016 (A), 2017 (B), 2018 (C), 2019 (D), and 2020 (E).

Fig. 6: Space-time scanning analysis of distribution of soil-transmitted helminths in China from 2016 to 2020.

Model	MRSE	MAE	R^2
RF	3.52	1.36	0.56
XGBOOST	3.72	1.49	0.48
LR	4.46	2.11	0.27
GBM	4.10	1.64	0.38

Table 3: **Parameters of model performance in testing set of machine learning.**

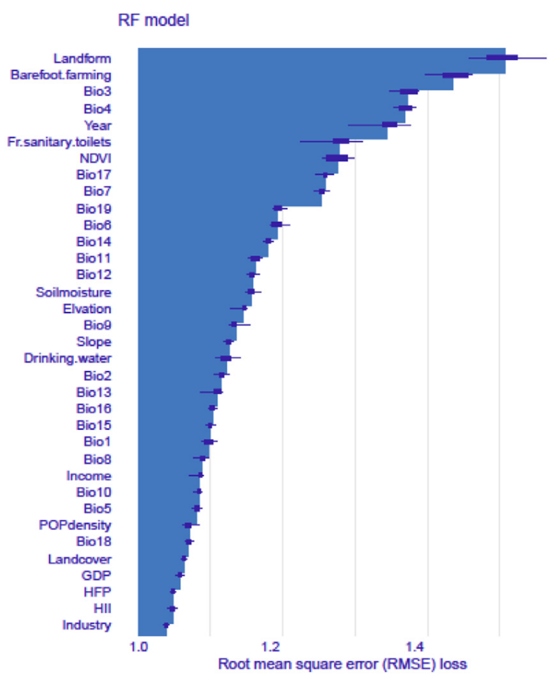

Fig. 7: RMSE loss of variables in RF model.

Finally, for the availability of sanitary toilets variable, the higher the coverage of sanitary toilets, the lower the infection rate of soil-transmitted helminths (Fig. 8-F).

Discussion

Soil-transmitted helminths have been prevalent in China for a long time, and seriously endanger the health of the Chinese population.[1] In 2016, a monitoring system was preliminarily established, and by 2020, the monitoring system for soil-transmitted helminthiasis covered 31 provinces and more than 400 counties.[16–20] The monitoring results revealed that the infection rates of soil-transmitted helminths dropped from 2.46% in 2016 to 0.84% in 2020. Additionally, the overall infection rate of soil-transmitted helminth and the species-specific infection rates all decreased by more than 60%, and the number of provinces where soil-transmitted helminths were not detected increased each year from 2016 to 2020.

The prevalence of soil-transmitted helminthiasis in China showed distinctive characteristics. First, regionalization was obvious. The highly prevalent areas were mainly distributed in the southwestern and southern regions of China, such as Hainan, Yunnan, Sichuan, Guizhou, and Chongqing. Those regions have higher temperatures and humidity than other regions in China, which makes them more suitable for helminth survival. The underdeveloped economy, as well as crop and vegetable planting in some regions were also factors contributing to the spread of diseases such as hookworms.[14] Second, the distribution by population was distinct. The infection rate of soil-transmitted helminthiasis was higher in females than in males, and was higher in the age group over 60 years, followed by that of the 45–59 years and 7–14 years age groups. Culturally, middle-aged and elderly individuals mainly participate in agricultural activities in rural areas, while school-aged children 7–14 years tended to have more outdoor activities that increase exposure.[16–20] The unhealthy diet and poor hygiene habits of these three age groups increased their risk of infection. Consequently, the control work of the “Fourteenth Five-Year Plan” focuses on school-age children, women, and the elderly. Finally, hookworm has become the dominant species of soil-transmitted helminths in China, and should be the focus of future prevention and control measures.

Spatiotemporal clustering analysis is an important method in spatial epidemiology, as it can not only detect spatiotemporal variations in disease risk and the geographic location of high risk areas, but also provides evidence of the causes of disease.[34] Few studies have been carried out on the temporal and spatial aggregation of soil-transmitted helminthiasis in China. Chen et al.[35] used spatiotemporal clustering to analyze the distribution of soil-transmitted helminthiasis in Jiangxi province from 2016 to 2019, and showed that spatiotemporal aggregation existed there. However, no studies have reported on the nationwide spatiotemporal distribution of soil-transmitted helminths. The results of global spatial autocorrelation analyses in this study showed that spatial aggregation existed for the overall distribution of soil-transmitted helminths in China. Hotspot analysis also revealed that the hot spots were concentrated in the southwestern and southern areas of the country, while the cold spots were concentrated in northeastern parts of China.

Retrospective spatiotemporal scanning analysis showed that the spatial distribution of helminth infections was aggregated in the southwest during 2016–2017. This was consistent with the high infection rates in 2016–2017. Thus, future prevention and control work must focus on those regions and pilot locations must be established in high infection areas. The spread of soil-transmitted helminths was relatively low in northeastern China because of the climate there, however, prevention and control efforts cannot be relaxed in

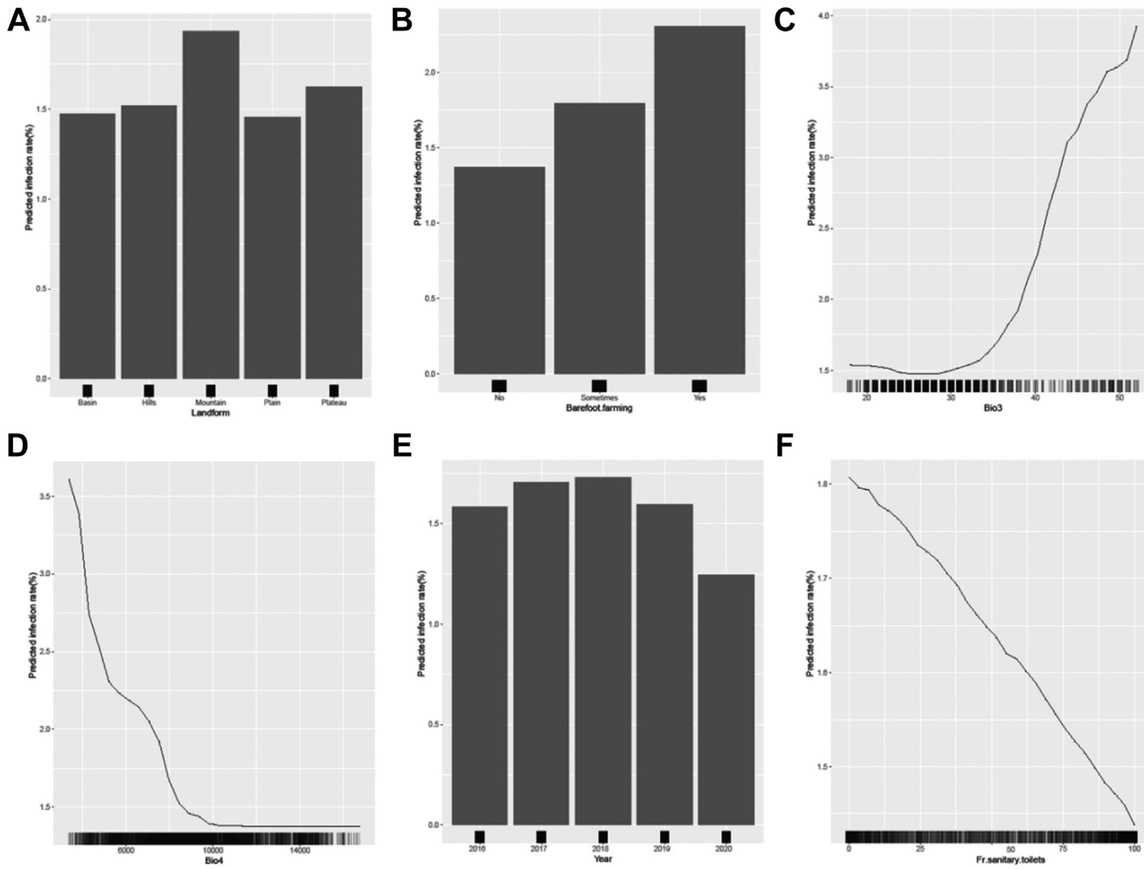

Fig. 8: Partial dependence plots of the top 6 influencing factors: landform (A), barefoot farming (B), Bio3 (C), Bio4 (D), year (E), and fraction of sanitary toilets (F) based on RF model.

that location, and routine monitoring still needs to be carried out in a timely and effective manner. The spatiotemporal clustering analysis provided scientific evidence to confirm the spatiotemporal aggregation of soil-transmitted helminths in China, which is important to guide prevention and control efforts.

Analyses of the factors influencing soil-transmitted helminth infections showed that among the variables in the landform model, plateaus and mountains were risk factors, compared to basins. Those findings were consistent with the results of the second national survey.[12,36] The economic development rates of plateau and mountain areas were lower than those of plain areas, which contributed to the high prevalence of soil-transmitted helminths yet was not beneficial to its' prevention and control.[36] For example, Yunnan, Sichuan, and Guizhou are provinces with high infection rates, and all are densely mountainous areas with low economic development rates.

Barefoot farming is also a risk factor for soil-transmitted helminth infections, and a higher frequency of barefoot labor is associated with a greater risk of infection. Those findings are also consistent with the findings of previous studies.[37–40] Hookworm is the dominant species of soil-transmitted helminths, and barefoot labor is a known risk factor for hookworm infection. Therefore, changing the culture of barefoot labor is an important strategy for soil-transmitted helminth control.

This study also showed that a higher coverage of sanitary toilets was associated with a lower infection rate of soil-transmitted helminths. Previous soil-transmitted helminth control efforts have also considered toilet renovation as a necessary step in achieving good results.[41–43] For future control efforts, toilet renovations in poor rural areas should remain an important consideration.

Previous studies on the relationships between the distribution of soil-transmitted helminths and climatic factors concluded that temperature, humidity and rainfall were important factors influencing the geographical distribution of *A. lumbricoides*, *T. trichiura*, and hookworms. However, those studies did not conduct an in-depth analysis of the relationship between climatic factors and soil-transmitted helminth infection rates.[44–47]

In this study, a machine learning model was constructed, upon which the impact of climate factors on the infection rates of soil-transmitted helminths was analyzed. The variables Bio3 (Isothermality) and Bio4 (Temperature seasonality) both reflected the degree of regional temperature fluctuations, where larger Bio3 and smaller Bio4 values represented smaller temperature fluctuations. The results of this study showed that, within a certain range, a larger Bio3 and smaller Bio4 were associated with higher helminth infection rates, thus, suggesting that soil-transmitted helminths have a certain tolerance to temperature changes. The development of soil-transmitted helminth eggs/larvae requires suitable temperatures and humidity levels in vitro environment. In areas where the temperature changes dramatically or rapidly and surpasses the eggs/larvae's carrying capacity, the survival of soil-transmitted helminths will be adversely affected, manifesting as a reduction in population infection rates.

This study had several limitations. In the machine learning models, important variables such as the coverage of sanitary toilets and barefoot farming were obtained from the national monitoring survey sites, and thus, the national raster data were not obtained. Therefore, the nationwide infection rates of soil-transmitted helminths could not be predicted and only the important influencing factors for infections were shown. Also, additional human variables such as dietary habits should be considered in the models. For future work, we suggest the collection of additional high quality data to improve the results. Moreover, the Kato-Katz method used for national monitoring could not distinguish similar eggs, such as the eggs from *A. lumbricoides* and *Ascaris suum*. Thus, the use of molecular methods such as PCR should be considered to improve the accuracy of future examinations.

Conclusion

The prevention and control efforts of soil-transmitted helminthiasis in China have achieved great results, which have been confirmed by three national surveys of parasitic diseases since 1988.[11–13] However, the continued national prevalence and high local prevalence of disease indicated that much work is still required for the prevention and control of soil-transmitted helminthiasis. We should continue to improve the national monitoring system, which not only provides data for the prevention and control of soil-transmitted helminthiasis, but serves as an intervention itself.[48] With the great regional differences in prevalence throughout China, educational prevention and control sites should be established in highly prevalent areas to reduce infection rates, and the elimination of soil-transmitted helminthiasis in areas with low infection rates should be prioritized.

Contributors

SZL and HHZ conceived and designed this study; JLH, CHZ, TZJ, MZZ, MBQ and YDC participated in the field survey, HHZ and JLH collected data, HHZ, JLH, and JXZ analyzed the data; HHZ drafted the manuscript; SZL and MBQ made critical revisions of the manuscript. All authors reviewed the manuscript.

Data sharing statement

Following approval by the National Institute of Parasitic Diseases, Chinese Center for Disease Control and Prevention (Shanghai, China), the datasets underlying the results of this article will be made available to investigators. Please email the corresponding author for more information.

Editor note

The Lancet Group takes a neutral position with respect to territorial claims in published maps and institutional affiliations.

Declaration of interests

None.

Acknowledgments

We thank all of the disease control agencies at provincial, municipal, and county levels engaged in the national monitoring for their hard work.

Appendix A. Supplementary data

Supplementary data related to this article can be found at https://doi.org/10.1016/j.lanwpc.2023.100766.

References

1 Tang LH, Xu LQ, Chen YD. *Parasitic disease control and research in China*. Beijing: Beijing Science and Technology Press; 2012.
2 Bethony J, Brooker S, Albonico M, et al. Soil-transmitted helminth infections: ascariasis, trichuriasis, and hookworm. *Lancet*. 2006;367(9521):1521–1532.
3 Truscott JE, Turner HC, Farrell SH, Anderson RM. Soil-transmitted helminths: mathematical models of transmission, the impact of mass drug administration and transmission elimination criteria. *Adv Parasitol*. 2016;94:133–198.
4 Salazar-castanon VH, Legorreta-herrera M, Rodriguez-sosa M. Helminth parasites alter protection against Plasmodium infection. *BioMed Res Int*. 2014;2014:913696.
5 Brooker S, Kabatereine NB, Smith JL, et al. An updated atlas of human helminth infections: the example of East Africa. *Int J Health Geogr*. 2009;8:42.
6 Hotez PJ, Kamath A. Neglected tropical diseases in sub-saharan Africa: review of their prevalence, distribution, and disease burden. *PLoS Negl Trop Dis*. 2009;3(8):e412.
7 Brooker S, Hotez PJ, Bundy DA. Hookworm-related anaemia among pregnant women: a systematic review. *PLoS Negl Trop Dis*. 2008;2(9):e291.
8 Hotez P. Hookworm and poverty. *Ann N Y Acad Sci*. 2008;1136:38–44.
9 Ranjan S, Passi SJ, Singh SN. Prevalence and risk factors associated with the presence of Soil-Transmitted Helminths in children studying in Municipal Corporation of Delhi Schools of Delhi, India. *J Parasit Dis*. 2015;39(3):377–384.
10 Lai YS, Biedermann P, Shrestha A, et al. Risk profiling of soil-transmitted helminth infection and estimated number of infected people in South Asia: a systematic review and Bayesian geostatistical Analysis. *PLoS Negl Trop Dis*. 2019;13(8):e0007580.
11 Xu LQ, Chen YD, Sun FH, et al. A national survey on current status of the important parasitic diseases in human population. *Chin J Parasitol Parasit Dis*. 2005;23:332–340.
12 Wang LD. *Report on the National survey of current status of major human parasitic diseases in China*. Beijing: People's Medical Publishing House; 2008.
13 Zhou XN. *Report on the National survey of important human parasitic diseases in China*22015;2. Beijing: People's Medical Publishing House; 2015.

14 Zhu HH, Zhou CH, Zhang MZ, et al. Engagement of the National Institute of Parasitic Diseases in control of soil-transmitted helminthiasis in China. *Adv Parasitol*. 2020;110:217–244.
15 China CDC. *Operation manual of the National surveillance on STH*. Beijing: China CDC; 2011.
16 Zhu HH, Huang JL, Zhu TJ, et al. National surveillance on soil-transmitted helminthiasis in the People's Republic of China. *Acta Trop*. 2020;205:105351.
17 Zhu HH, Huang JL, Zhu TJ, et al. National surveillance of soil-transmitted helminth infections in 2017 [J]. *Chin J Parasitol Parasit Dis*. 2019;37(1):12–17.
18 Zhu HH, Huang JL, Zhu TJ, et al. Soil-transmitted helminthiasis — China, 2018 [J]. *China CDC Weekly*. 2020;2(3):34–38.
19 Zhu HH, Huang JL, Chen YD, et al. Analysis on endemic status of soil-transmitted nematode infection in China in 2019. *Chin J Parasitol Parasit Dis*. 2021;39(5):666–673.
20 Gao F. *Surveillance report on infectious diseases of China in 2020*. Beijing: Chinese center for disease control and prevention; 2021.
21 JYH Aw, Clarke NE, Mayfield HJ, Lau CL, Richardson A, Nery VS. Novel statistical approaches to identify risk factors for soil-transmitted helminth infection in Timor-Leste. *Int J Parasitol*. 2021;51(9):729–739.
22 Nisha M, Aiman M, Asyhira N, et al. Risk factors associated with soil transmitted helminth (STH) infection in two indigenous communities in Malaysia. *Trop Biomed*. 2020;37(2):379–388.
23 Gupta A, Acharya AS, Rasania SK, Ray TK, Jain SK. Prevalence and risk factors of soil-transmitted helminth infections in school age children (6-14 years) - a cross-sectional study in an urban resettlement colony of Delhi. *Indian J Public Health*. 2020;64(4):333–338.
24 Deka S, Barua D, Bahurupi Y, Kalita D. Assessment of the prevalence of soil-transmitted helminth infections and associated risk factors among school-aged children in a flood-affected area of Northeast India. *Am J Trop Med Hyg*. 2021;105(2):480–489.
25 Okoyo C, Campbell SJ, Williams K, Simiyu E, Owaga C, Mwandawiro C. Prevalence, intensity and associated risk factors of soil-transmitted helminth and schistosome infections in Kenya: impact assessment after five rounds of mass drug administration in Kenya. *PLoS Negl Trop Dis*. 2020;14(10):e0008604.
26 Vaz NS, Clarke NE, Richardson A, et al. Risk factors for infection with soil-transmitted helminths during an integrated community level water, sanitation, and hygiene and deworming intervention in Timor-Leste. *Int J Parasitol*. 2019;49(5):389–396.
27 A comparison of the sensitivity and fecal egg counts of the McMaster egg counting and Kato-Katz thick smear methods for soil-transmitted helminths. *PLoS Negl Trop Dis*. 2011;5(6):e1201.
28 Yu SH, Xu LQ. *Colour atlas of human parasitology*. Beijing: China Science and Technology Press; 1992.
29 Robinson TP. Spatial statistics and geographical information systems in epidemiology and public health. *Adv Parasitol*. 2000;47:81–128.
30 Robertson C, Nelson TA. Review of software for space-time disease surveillance. *Int J Health Geogr*. 2010;9:16.
31 Masrur A, Yu M, Luo W, Dewan A. Space-time patterns, change, and propagation of COVID-19 risk relative to the intervention scenarios in Bangladesh. *Int J Environ Res Publ Health*. 2020;17(16):5911.
32 Kulldorff M, Athas WF, Feurer EJ, Miller BA, Key CR. Evaluating cluster alarms: a space-time scan statistic and brain cancer in Los Alamos, New Mexico. *Am J Public Health*. 1998;88(9):1377–1380.
33 Li L, Ching WK, Liu ZP. Robust biomarker screening from gene expression data by stable machine learning-recursive feature elimination methods. *Comput Biol Chem*. 2022;100:107747.
34 Xue JB, Xia S, Zhang X, et al. Pattern analysis of tempo-spatial distribution of schistosomiasis in marshland epidemic areas in stage of transmission control. *Chin J Schisto Control*. 2016;28(6):624–629+643.
35 Chen Z, Ge J, Wen Q, Qiu TT, Zeng XJ. Spatial-temporal clustering analysis of soil-transmitted helminth transmission risk in Jiangxi Province, 2016-2019. *China Trop Med*. 2022;22(9):1–7.
36 Tian H, Luo J, Zhong B, Huang Y, Xie H, Liu L. Challenges in the control of soil-transmitted helminthiasis in Sichuan, Western China. *Acta Trop*. 2019;199:105132.
37 Brandon-mong GJ, Abdullah NA, Shukor N, et al. Soil-Transmitted Helminths in Malaysia landscape: an aborigines study. *Trop Biomed*. 2017;34(2):363–374.
38 Rahimi BA, Mahboobi BA, Wafa MH, Sahrai MS, Stanikzai MH, Taylor WR. Prevalence and associated risk factors of soil-transmitted helminth infections in Kandahar, Afghanistan. *BMC Infect Dis*. 2022;22(1):361.
39 Jiraanankul V, Aphijirawat W, Mungthin M, et al. Incidence and risk factors of hookworm infection in a rural community of central Thailand. *Am J Trop Med Hyg*. 2011;84(4):594–598.
40 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.
41 Hesham al-mekhlafi M, Surin J, Atiya AS, Ariffin WA, Mohammed mahdy AK, Che abdullah H. Pattern and predictors of soil-transmitted helminth reinfection among aboriginal schoolchildren in rural Peninsular Malaysia. *Acta Trop*. 2008;107(2):200–204.
42 Freeman MC, Chard AN, Nikolay B, Garn JV, Okoyo C, Kihara J. Associations between school- and household-level water, sanitation and hygiene conditions and soil-transmitted helminth infection among Kenyan school children. *Parasit Vectors*. 2015;8:412.
43 Sangalang SO, Medina SAJ, Ottong ZJ, et al. Protocol for a trial assessing the impacts of school-based WaSH interventions on children's health literacy, handwashing, and nutrition status in low- and middle-income countries. *Int J Environ Res Publ Health*. 2020;18(1):226.
44 Xu LQ, Yu SH, Xu SH. *Distribution and pathogenic impact of human parasites in China*. Beijing: People's Medical Publishing House; 1999.
45 Zhou XN. *Soil-transmitted parasitic diseases*. Beijing: People's Publishing House; 2011.
46 Pullan RL, Brooker SJ. The global limits and population at risk of soil-transmitted helminth infections in 2010. *Parasit Vectors*. 2012;5:81.
47 Fleitas PE, Kehl SD, Lopez W, et al. Mapping the global distribution of Strongyloides stercoralis and hookworms by ecological niche modeling. *Parasit Vectors*. 2022;15(1):197.
48 Qian MB, Li SZ, Zhou XN. Epidemiology and control of soil-transmitted nematodiasis in China. *J Trop Dis Parasitol*. 2022;20(3):126–129.