Bayesian geostatistical and mathematical models to assess the geographical distribution of neglected tropical diseases

INAUGURALDISSERTATION

zur

Erlangung der Würde eines Doktors der Philosophie

vorgelegt der

Philosophisch-Naturwissenschaftlichen Fakultät

der Universität Basel

von

Yingsi Lai

aus China

Basel, 2016

Genehmigt von der Philosophisch-Naturwissenschaftlichen Fakultät auf Antrag von Prof. Dr. Jürg Utzinger, PD Dr. Penelope Vounatsou, and Asst. Prof. Anna-Sofie Stensgaard.

Basel, den 19. April 2016

Prof. Dr. Jörg Schibler

Dekan

Summary

Neglected tropical diseases (NTDs) are a group of communicable diseases affecting more than one billion of the world's poorest population. Soil-transmitted helminth infections, schistosomiasis, and foodborne trematodiasis are among the most important NTDs. Soil-transmitted helminth infections are caused by a group of parasite nematode worms (i.e., *Ascaris lumbricoides*, *Trichuris trichiura*, and hookworm) through contact with parasite eggs or larvae which thrive in warm and moist soil. They are widely endemic in the tropics and sub-tropics and ranked on the top among all NTDs burden, contributing to the global disease burden with 5.2 million disability-adjusted life years (DALYs). Schistosomiasis is caused by trematode parasites of the genus *Schistosoma*. It is the second highest in terms of NTD burden and responsible for around 3.3 million DALYs worldwide. More than 90% of schistosomiasis cases occur in Africa. Clonorchiasis is one of the most important foodborne trematodiasis and it is caused by infection with the Chinese liver fluke, *Clonorchis sinensis*. China accounts for around 85% of the global infected people and most cases occur in the southern and the northeastern parts of the country. For all the three diseases, preventive chemotherapy is advocated by WHO as a key strategy for morbidity control. Furthermore, integrated approaches are highly recommended to achieve sustainable control and elimination. Such approaches may include preventive chemotherapy in combination with improvement of water, sanitation, and hygiene, as well as better information, education, and communication.

To implement control strategies cost-effectively, high-resolution maps depicting the geographical distribution of disease risk are important. These maps provide useful information for spatial targeting of control measures and for long-term monitoring and surveillance. Geostatistical modeling is the most rigorous inferential approach for high-resolution risk mapping of NTDs. It is a data-driven approach, which relates georeferenced disease data (usually point-referenced) with potential predictors (e.g., environmental and socioeconomic factors) that are considered important for disease transmission. Location-specific random effects can explain geographical variation in the data, assuming that neighboring areas have similar infection status due to common disease exposures they receive. Geostatistical models are highly parameterized, however Bayesian model formulations provide a flexible inferential framework and powerful computational tools such as Markov chain Monte Carlo (MCMC) simulation or approximations (e.g., integrated nested Laplace approximation (INLA)) are applied for model fit.

A good coverage and a fine amount of disease data are necessary to capture the spatial heterogeneity of the infection risk. Due to lack of large surveys covering the whole study region, this PhD thesis is based on historical survey data that are compiled via bibliometric searches. Publications however are either report the survey data as point-referenced (with geographical information at the survey location) or as areal, aggregated over several locations

within an administrative level (e.g., county or district). The areal data can provide useful information especially when the spatial coverage of point-referenced data is low. Geostatistical model for jointly analysing point-level and areal survey data are not available. Furthermore, historical data are generated from studies with different designs between locations, including different population age-groups. Geostatistical models that align survey data across locations to a common age group do not exist in the field of NTDs. Ignoring the age-heterogeneity of the data can lead to biased estimation because models cannot distinguish whether risk differences between locations is due to differences in age or to exposures. Mathematical models can be used to age-align the surveys, but there is no model formulation allowing changes of the shape of the age-prevalence curve over space as a result of the varying endemicity.

The overall goal of the thesis is to develop Bayesian geostatistical and mathematical models for analysing georeferenced NTD survey data and to provide tools and knowledge for disease control and prevention.

In Chapter 2 surveys pertaining to soil-transmitted helminth infections in People's Republic of China (P.R. China) were compiled. Bayesian geostatistical models were developed and used to estimate the disease risk throughout the country at high spatial resolution. Advanced Bayesian variable selection methods were employed to identify the most important predictors. Results indicate that the prevalence of soil-transmitted helminth infections in P.R. China considerably decreased from 2005 onwards. Yet, some 144 million people were estimated to be infected in 2010. High prevalence (>20%) was predicted in large areas of Guizhou and the southern part of Hubei and Sichuan provinces for *Ascaris lumbricoides* infection, in large areas of Hainan, the eastern part of Sichuan, and the southern part of Yunnan provinces for hookworm infection, as well as in a few small areas of south P.R. China for *Trichuris trichiura* infection.

In Chapter 3 a systematic review was carried out to identify prevalence surveys to soil-transmitted helminth infections in South Asia. Bayesian geostatistical models were applied to identify important environmental and socioeconomic predictors, and to estimate infection risk at high spatial resolution across the study region. Results show that 397 million of South Asia population was infected with at least one species of soil-transmitted helminths in 2015. *A. lumbricoides* was the most common infection species. Moderate to high prevalence (>20%) of any soil-transmitted helminth infection was predicted in the northeastern part and some northern areas of the study region as well as the southern coastal-line areas of India. The annual treatment needs for the school-aged population requiring preventive chemotherapy was estimated at 187 million doses. The study highlights the need for up-to-date surveys to accurately evaluate the disease burden in the region.

In Chapter 4 georeferenced survey data of *C. sinensis* infection were obtained via a systematic review and additional data were provided by the National Institute of Parasitic

Diseases, Chinese Center for Diseases Control and Prevention. Bayesian geostatistical models were applied to quantify the relation between infection risk and important predictors, and to predict the risk of infection across P.R. China at high spatial resolution. The results show an increasing risk of *C. sinensis* infection over time, particularly from 2005 onwards, which urges the Chinese government to pay more attention on the public health importance of the diseases. Highly endemic areas (>20%) were concentrated in southern and northeastern parts of the country. The provinces with the highest risk of infection and the largest number of infected people were Guangdong, Guangxi and Heilongjiang.

In Chapter 5 a systematic review was conducted to identify relevant surveys pertaining to prevalence of *Schistosoma* infection in sub-Saharan Africa. Bayesian geostatistical meta-analysis and rigorous variable selection were used to obtain up-to-date risk estimates of schistosomiasis at high spatial resolution, based on environmental and socioeconomic predictors. The literature search identified *Schistosoma haematobium* and *Schistosoma mansoni* surveys at 9,318 and 9,140 unique locations, respectively. Results show a decreased infection risk from 2000 onwards, yet suggesting that 163 million Africans were infected in 2012. Mozambique had the highest prevalence of *Schistosoma* infection among 44 countries of sub-Saharan Africa. Annualised treatment needs with praziquantel were estimated at 123 million doses for school-aged children and 247 million for the entire population.

In Chapter 6 a Bayesian geostatistical modeling approach was developed to analyse jointly areal and point-referenced survey data. We assumed that the point-referenced data arise from a binomial distribution and that the aggregated area data follow a Poisson binomial distribution which was approximated by a two parameter shifted binomial distribution. Results from extensive simulations shows that our proposed model has better predictive ability and improves parameter estimation compared to models that treat area data as points, located at the centroid of the areas. We applied the new model to obtain high spatial resolution estimates of the infection risk of clonorchiasis in an endemic region of P.R. China.

In Chapter 7 we integrated geostatistical and mathematical transmission models of schistosomiasis within a single model formulation to analyse age-heterogenous *S. mansoni* data from Côte d'Ivoire. A series of age-specific risk maps of *S. mansoni* infection in Côte d'Ivoire were produced at high geographical resolution, which allow us to identify the most important age groups of the population to treat at a given place. We predicted that the infection risk reached the peak at younger ages in high risk areas and at older ages in low risk areas. Moreover, a more rapid decline rate of infection risk was observed at older ages in high risk areas compared to that in moderate and low risk ones.

In summary, this PhD thesis contributes to the fields of spatial statistics and of epidemiology of NTDs with (i) statistical methodology for modeling spatially-structured disease data, having heterogeneous geographical support (i.e., georeferenced at point or area level) across the study region and they are collected over different age groups between

locations, (ii) applications on soil transmitted helminth infections, schistosomiasis, and clonorchiasis in sub-Saharan Africa, South Asia, and P.R. China, to obtain spatially explicit estimates of disease risk, number of infected people, and annual treatment needs for preventive chemotherapy at different administrative levels, and (iii) large amount of geo-referenced data on NTD surveys conducted at over 10,750 unique locations that are available via the open access Global Neglected Tropical Diseases Database (GNTD). The innovative statistical methodology for analysing historical survey data, heterogeneous in space can be readily applied to other disease survey data. The up-to-date, model-based, high-resolution risk maps and estimates of treatment needs provide useful tools and information for guiding disease control and interventions.

Acknowledgements

Contents

List of Figures

List of Tables

List of Abbreviations

BCI: Bayesian credible interval

BSAT: base saturation as percentage of ECEsoil

BULK: bulk density

CFRAG: percentage of coarse fragments

China CDC: Chinese Center for Diseases Control and Prevention

CLPC: percentage of clay

CNKI: China National Knowledge Internet

CSC: China Scholarship Council

DALYs: disability-adjusted life years

DHS: Demographic and Health Surveys

DIC: deviance information criterion

DR Congo: Democratic Republic of the Congo

DRAIN: FAO soil drainage class

ELISA: enzyme-linked immunosorbent assay

GAHI: Global Atlas of Helminth Infections

GDP: gross domestic product

GMRF: Gaussian Markov random field

GNTD: Global Neglected Tropical Diseases Database

GYPS: gypsum content

HII: human influence index

IEC: information, education, and communication

IMR: infant mortality rates

INLA: integrated nested Laplace approximation

IPD: the National Institute of Parasitic Diseases

JMP: Joint Monitoring Programme

LAMP: loop-mediated isothermal amplification

Lao PDR: Lao People's Democratic Republic

LSMS: Living Standards Measurement Study

LST: land surface temperature

MAE: mean absolute error

MCMC: Markov chain Monte Carlo

ME: mean error

MICS: Multiple Cluster Indicator Surveys

NaOH: sodium hydroxide

NDVI: normalized difference vegetation index

NMIG: normal mixture of inverse Gammas

NTDs: Neglected tropical diseases

P.R. China: People's Republic of China

PCT: Preventive Chemotherapy and Transmission Control

peNMIG: normal mixture of inverse Gammas with parameter expansion

PHAQ: pH measured in water

POC-CCA: point-of-care circulating cathodic antigen

PSCL: FAO texture class

RAMPS: reparameterized and marginalized posterior sampling

SD: standard deviation

SDTO: percentage of sand

SEDAC: Socioeconomic Data and Applications Center

SPDE: stochastic partial differential equations

STPC: percentage of silt

SWBD: Shuttle Radar Topography Mission Water Body Data

Swiss TPH: Swiss Tropical and Public Health Institute

TAWC: available water capacity

TOTC: organic carbon content

TOTN: total nitrogen

WASH: water, sanitation, and hygiene

WHO: World Health Organization

WHS: World Health Surveys

Chapter 1   Introduction

1.1 Neglected tropical diseases

Neglected tropical diseases (NTDs) are a diverse group of communicable diseases that were historically overlooked but among the most common chronic infections in the world's poorest population (Hotez et al. 2007;Mackey et al. 2014). The 17 core NTDs that WHO identified include: dengue and chikungunya and rabies, which are viral infections; Buruli ulcer, endemic treponematoses, leprosy, and trachoma, which are bacterial infections; Chagas disease, human African trypanosomiasis, and leishmaniases, which are protozoan infections; and cysticercosis/taeniasis, dracunculiasis, echinococcosis, foodborne trematodiasis, lymphatic filariasis, onchocerciasis, schistosomiasis, and soil-transmitted helminth infections, which are helminthial or metazoan infections. They are endemic in 149 countries and territories and affect more than one billion people (WHO 2010b). With a few exceptions, NTDs are mostly resulted in low mortality but high morbidity conditions (Hotez 2013). It was estimated that all NTDs together accounting for 26 million disability-adjusted life years (DALYs) in 2010, which is comparable to that of the "big three diseases" (i.e., HIV/AIDS, tuberculosis, and malaria, corresponding to 81 million, 49 million, and 83 million DALYs, respectively) (Hotez 2015;Murray et al. 2012).

Control of NTDs is considered as "low-hanging fruit", as low-cost and highly cost-effective intervention approaches are available for many of these diseases, and in addition, controlling of these diseases has simultaneous and sustainable effects on poverty reduction (Hotez et al. 2009;Molyneux 2010). However, many of individuals still have far less access to the resources (Mackey et al. 2014). As some of NTDs are co-endemic and share similar control strategies, tackling of these diseases through co-implementation, for example, conducting integrating drug distribution programs, can be even more effective and affordable (Brady, Hooper, & Ottesen 2006;Laxminarayan et al. 2006).

Preventive chemotherapy is identified as a key strategy for tackling, often jointly, a number of NTDs, according to the NTD roadmap published by WHO in 2012 (WHO 2012a). The road map also set specific targets for eradication, elimination, and intensified control of different NTDs (WHO 2012a). Five public-health strategies to overcoming NTDs were highlighted by the WHO report in 2013, which include: (1) preventive chemotherapy, (2) innovative and intensified disease-management, (3) vector control and pesticide management, (4) safe drinking-water and basic sanitation/hygiene services, and (5) education and veterinary public-health services (WHO 2013). These strategies can be more effective when combined and delivered locally (WHO 2013).

In order to cost-effective implementation of control strategies, high-resolution maps depicting the geographical distribution of disease risk are important to identify areas with highest risk, but they are not yet available for many NTDs. This PhD thesis focuses on risk estimates of three important NTDs: soil-transmitted helminth infections, schistosomiasis, and

clonorchiasis (one of the most important foodborne trematodiasis), and provides appropriate methodologies for the estimation.

1.1.1 Soil-transmitted helminth infections

Soil-transmitted helminth infections are caused by a group of parasite nematode worms through contact with parasite eggs or larvae that thrive in warm and moist soil of the world's tropical and subtropical countries (Bethony et al. 2006).

1.1.1.1 Parasites and life cycles

Three main kinds of worms that infect people are the roundworm (*Ascaris lumbricoides*), the whipworm (*Trichuris trichiura*), and the hookworms (*Necator americanus* and *Ancylostoma duodenale*), which vary greatly in shape and size. Adult worms living in the intestine produce thousands of eggs every day, which leave the body with faeces and contaminate the soil and water systems in areas with no latrine systems or of poor sanitation (Figure 1.1-1.3). In soil, eggs develop into infective stages. People get infected with *A. lumbricoides* and *T. trichiura* by ingesting the infective eggs via intake of food/water or putting into the mouths unwashed hands, which are contaminated with eggs. On the other hand, people are infected with hookworm by contacting the infective larvae with skin. Inside the human body, the three worms travel to their final locations in different ways: after ingestion, trichuris eggs develop into larvae and travel directly to the colon, where they further develop into adult worms; ascaris larvae penetrate the intestinal mucosa, migrate through liver and lungs, re-enter the gastrointestinal tract, and develop into adult worms; after skin penetration, hookworm larvae enter the afferent circulation, pass through the lungs, migrate into gastrointestinal tract, and turn to egg-laying adults (Bethony et al. 2006). In addition, *A. duodenale* larvae are also infective through oral digestion (Loukas & Prociv 2001). Soil-transmitted helminths can live in human intestine for several years. As they do not reproduce within the host, re-infection occurs only when people re-contact the infective stages of worms in environment.

1.1.1.2 Clinical conditions

When larvae migrate through skin and viscera, acute manifestations, from temporary skin itch to severe pneumonia, can occur (Bethony et al. 2006). With regards to intestinal parasitism, only infections with relatively high intensity can lead to evident symptoms, which include intestinal manifestations (e.g., diarrhea and abdominal pain), general malaise and weakness, malnutrition, and impaired physical growth (WHO 2006). Particularly, very heavy infection by *A. lumbricoides* can lead to severe consequences such as intussusception, volvulus, complete obstruction, bowel infarction, intestinal perforation, and peritonitis (Das 2014;Viliamizar et al. 1996). Heavy *T. trichiura* infection can result in serious manifestations (e.g., chronic dysentery and rectal prolapse) (Bundy & Cooper 1989). The major hookworm-related pathology is caused by intestinal blood loss, as the worms use their cutting apparatus

to attach the intestinal mucosa and submucosa, leading to mechanical and chemical rupture of capillaries and arterioles (Hotez et al. 2004). Iron-deficiency anaemia and hypoalbumnemia occur when blood loss exceeds the intake and reserves of iron and protein of the host (Stoltzfus et al. 1997). Children and women of child-bearing age are at particular risk of hookworm infection, as they have reduced iron reserves (Bethony et al. 2006). Sever iron-deficiency anaemia during pregnancy caused by hookworm disease can result in adverse outcomes for both mother and infant (Bundy, Chan, & Savioli 1995).
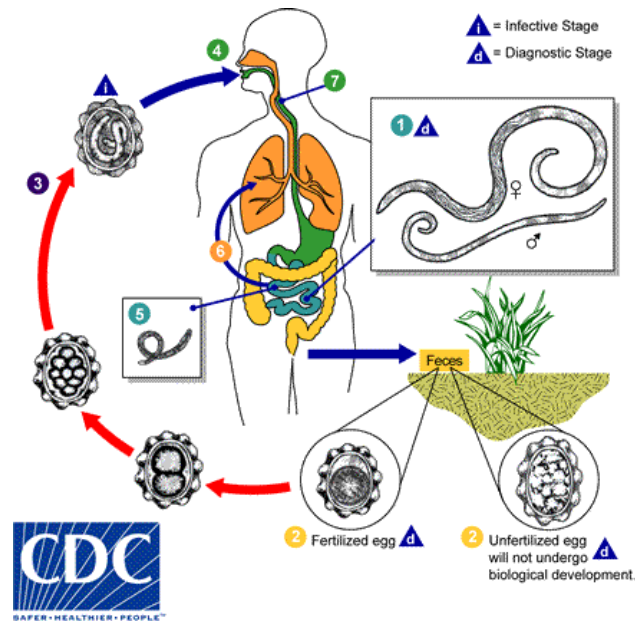
Figure 1.1: Life cycle of *Ascaris lumbricoides* (source: CDC)
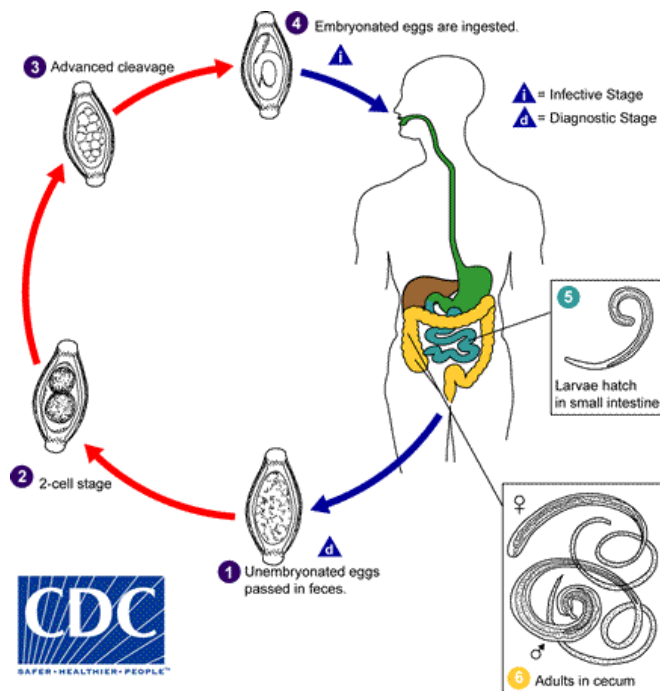
Figure 1.2: Life cycle of *Trichuris trichiura* (source: CDC)
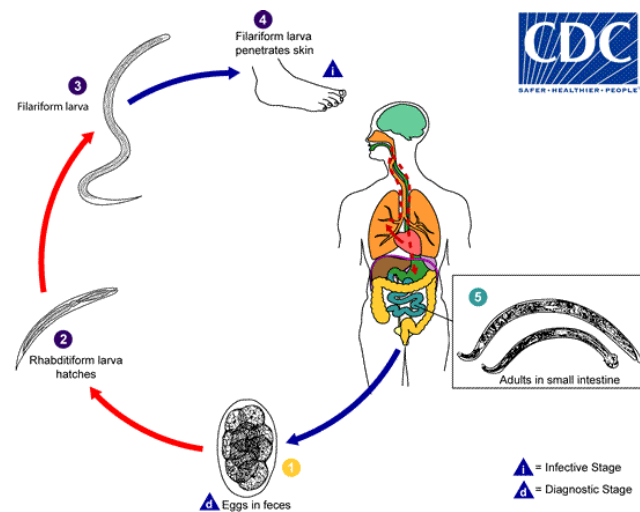
Figure 1.3: Life cycle of hookworm (source: CDC)

1.1.1.3 Epidemiology

Soil-transmitted helminth infections are widely endemic in tropics and sub-tropics, particularly in areas of East Asia and Pacific Islands (included China), sub-Saharan Africa, South Asia (included India), Latin America and Caribbean (de Silva et al. 2003) (Figure 1.4). It was estimated that in 2010 the global numbers of people infected with *A. lumbricoides*, *T. trichuris*, hookworm, and any soil-transmitted helminth species were 819 million, 465 million, 439 million, and 1.45 billion, respectively (Pullan et al. 2014). The overall DALYs caused by soil-transmitted helminth infections in 2010 were estimated to 5.2 million, ranked the top among all NTDs (Murray et al. 2012). Regarding to each disease, *A. lumbricoides*, *T. trichuris*, and hookworm infections took into account 1.3 million, 0.6 million, and 3.2 million DALYs, resplectively (Murray et al. 2012). Populations affected by these diseases are often live in poverty, with less access to clean water and sanitation infrastructures, inadequate hygiene practices, and low education (Hotez et al. 2007;Steinmann et al. 2010).

1.1.1.4 Risk factors

Environmental and climatic conditions are important for transmission of soil-transmitted helminth infections. Several factors (e.g., land surface temperature, soil moisture, soil types, vegetation, land cover, rainfall, and altitude) influence the development and survival of the worms' free-living infective stages (Brooker et al. 2003;Hohmann et al. 2001;Tchuem Tchuenté 2011). For example, warm temperature and adequate moisture in soil can speed up the development of eggs/larvae (Brooker, Clements, & Bundy 2006). On the other hand, socioeconomic factors (e.g., drinking water sources, sanitation, personal hygiene, education, poverty, and clinical features) have equal importance for transmission of the diseases (Escobedo, Canete, & Nunez 2008;Hohmann et al. 2001;Knopp et al. 2010;Norhayati,

Oothuman, & Fatmah 1998;Pinheiro et al. 2011). With poor sanitation, faeces from infected people may easily contaminate the soil; with bad hygiene practice, people can simply get infected due to, for example unwashed hands or unclean vegetables.
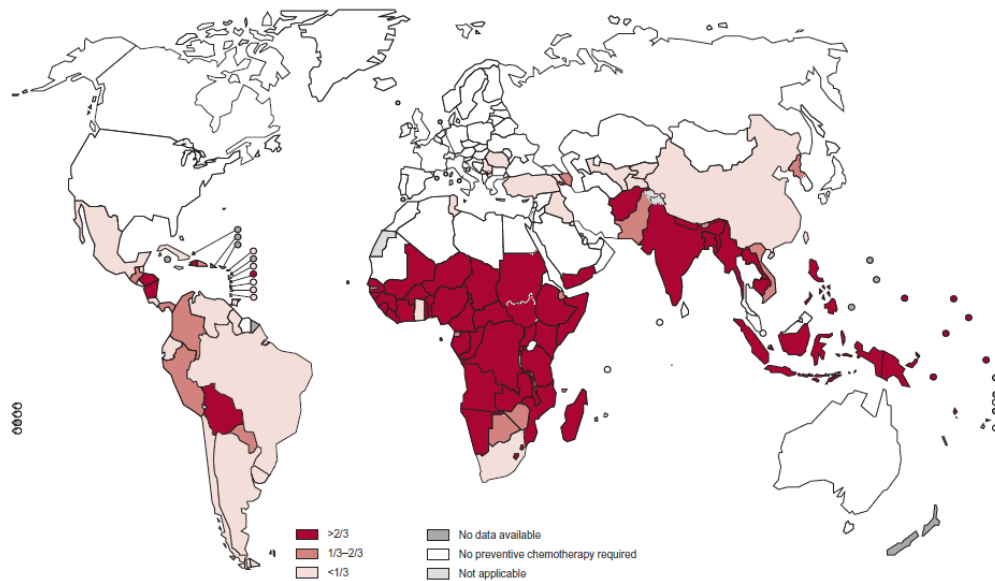
Figure 1.4: Endemic regions of soil-transmitted helminth infections and proportion of children (aged 1-14 years) in each endemic country requiring preventive chemotherapy for the diseases, 2011 (sources:WHO 2013).

1.1.1.5 Diagnosis and treatment

WHO recommended Kato-Katz technique as the standard method for evaluating prevalence and intensity of soil-transmitted helminth infections in endemic areas, due to its relative simplicity, speed, and low cost (Montressor et al. 1998). On average, 41.7 mg of stool on a microscopic slide is examined for the detection and quantification of helminth eggs in a single Kato-Katz thick smear (Speich et al. 2014b). However, Kato-Katz method may result in low sensitivity if only one single smear is examined, particularly in low transmission settings (Booth et al. 2003). By multiple stool sampling, the sensitivity can be increased (Knopp et al. 2008). Other commonly used techniques include direct smear microscopy, formol-ether concentration (FEC), McMaster, FLOTAC, and Mini-FLOTAC, which also rely on visual examination of a small sample of stool to determine the presence and number of soil-transmitted helminth eggs (Nikolay, Brooker, & Pullan 2014). A single FLOTAC shows higher sensitivity than multiple Kato-Katz thick smears in detecting low-intensity infections, thus is considered as an alternative for anthelmintic drug efficacy studies and for monitoring and evaluation of deworming programs (Knopp et al. 2009;Knopp et al. 2011). Generally, if sources are permitting, combination of different methods are suggested for a more reliable evaluation of the prevalence and intensity of infections. In addition, ultrasonography and endoscopy are useful in clinical practices for diagnosis of intestinal complications due to ascariasis (Umetsu et al. 2014).

Benzimidazole drugs albendazole and mebendazole are commonly used for treatment of soil-transmitted helminth infections, with an aim to removal of worms from the gastrointestinal tract (Bethony et al. 2006). As board-spectrum anthelminthic drugs, they have different efficacies on the three types of infections: a single dose of both drugs is effective for *A. lumbricoides* infection; single-dose of albendazole shows high efficacy for hookworm, on contrary to that of mebendazole, which appears a low cure rate; a single dose of the two drugs is not satisfactory for the treatment of *T. trichiura* (Keiser & Utzinger 2008;Keiser & Utzinger 2010). In this way, several doses of mebendazole are needed for *T. trichiura* and hookworm infections (Keiser & Utzinger 2010). On the other hand, mebendazole is poorly absorbed so the therapeutic activity is mainly on the adult worms in gastrointestinal tract, while the absorption of albendazole is better especially with fat in the diet, thus it is also used for treatment of disorders caused by larvae migration through tissues (Dayan 2003). Mild and transient side effects can occur, including diarrhoea, nausea, abdominal discomfort, headache, and fatigue (Ray 2015). Although there is no confident evidence showing that albendazole and mebendazole are embryotoxic and teratogenic in human, concerns arise with the use of them in very young children and pregnancy women, as benzimidazole drugs have been shown embryotoxic and teratogenic in some animal species (Acs et al. 2005;Horton 1997). In preventive chemotherapy, WHO allows the use of albendazole and mebendazole in pregnant women of the second and third trimesters, as well as in lactating women, by considering that the benefit of treatment outweighs the risk (WHO 2002b;WHO 2006).

Levamisole, pyrantel pamoate, and ivermectin are also used for treatment against soil-transmitted helminth infections. However, a single oral dose of the above drugs has less efficacious for *T. trichiura* infection (Keiser & Utzinger 2010). Oxantel pamoate is a pyrimidine derivate developed from pyrantel, with excellent activity against *T. trichiura* but only low efficacy against *A. lumbricoides* and hookworm (Keiser et al. 2013;Moser et al. 2016). Therefore, it is necessary to combine oxantel pamoate with a partner drug (e.g., albendazole) in order to have a broad treatment of all three types of soil-transmitted helminth infections (Speich et al. 2014a;Speich et al. 2015). Although there is no conclusive evidence for drug resistance among human soil-transmitted helminth infections, monitoring drug efficacy in control programmes is necessary in order to maximize the ability to detect any drug resistance cases (Vercruysse et al. 2011).

1.1.1.6 Control and prevention

Periodic large-scale preventive chemotherapy is advocated by WHO in infection risk areas to control morbidity (WHO 2006). Frequent anthelmintic drug administrations are necessary to maximize the benefit of preventive chemotherapy, as reinfections of soil-transmitted helminths can happen rapidly after treatment (Jia et al. 2012). Albendazole or mebendazole (at a single dosage of 400mg or 500mg, respectively) are recommended for treatment of all school-aged children twice each year in high-risk (prevalence≥50%) areas and once each

year in moderate-risk (prevalence≥20% and <50%) areas (WHO 2006). Control strategies only focusing on school-based deworming may be inadequate, thus extension of treatments should be also considered to other populations (e.g., preschool-aged children, women of childbearing age, and high occupational exposure adults) or to the whole community (Anderson et al. 2013;Karagiannis-Voules et al. 2015a;Lo et al. 2015).

As re-infections occur rapidly, long-term solutions should be applied. Studies reveal that water, sanitation, and hygiene (WASH) interventions, especially the improvement of sanitation, appear to significantly reduce odds of infection (Strunz et al. 2014;Ziegelbauer et al. 2012). Furthermore, health education shows a positive impact on control of the diseases (Al-Delaimy et al. 2014;Gyorkos et al. 2013). In general, in order to achieve a durable reduction or elimination of soil-transmitted helminth infections, integrated control approaches are required, for example preventive chemotherapy with improvements of WASH and better information, education, and communication (IEC) (Jia et al. 2012;Strunz et al. 2014;Ziegelbauer et al. 2012).

1.1.2 Schistosomiasis

Schistosomiasis, also known as bilharizia, is a disease caused by trematode parasites of the genus *Schistosoma* (Gryseels et al. 2006).

1.1.2.1 Parasites and life cycles

There are three main species of schistosomes infecting human beings, that is *Schistosoma haematobium* presenting in Africa and the Middle East, *Schistosoma mansoni* in Africa, the Middle East, and South America, and *Schistosoma japonica* in Asia, primarily the Philippines and China (Colley et al. 2014). Besides, there are the other three species of only local importance with restricted distributions: *Schistosoma mekongi*, found along the Mekong River and its tributaries in Cambodia and Lao People's Democratic Republic (Lao PDR), and *Schistosoma guineensis* and *Schistosoma intercalatum* in parts of West and central Africa (Chu et al. 2012;Muth et al. 2010;Tchuem Tchuenté et al. 2003b). Adult schistosomes have a cylindrical body with two terminal suckers, a complex tegument, a blind digestive tract, and reproductive organs, usually 7-20mm in length (Gryseels et al. 2006). They have separate sexes but male and female worms live much of the time as embraced couple (Gryseels et al. 2006).

The eggs produced by adult worms are excreted in faeces or urine of infected human and shed into environment (Figure 1.5). In freshwater, eggs are hatch into miracidia, which infect the intermediate host, freshwater snails. Inside the snails, they further develop into multicellular sporocysts through asexual replication and eventually the cercariae. Mature cercariae are released into water, penetrate the skin of human host and turn into schistosomulae, which further migrate through blood circulation and reach perivesicular (for *S. haematobium*) or mesenteric (for other species) destination (Gryseels et al. 2006). It takes

about 5-7 weeks for schistosomulae becoming adults and producing eggs (Colley et al. 2014). There are a specific range of suitable snail hosts for each species, for example *S. haematobium*, *S. mansoni*, and *S. japonica* are transmitted by *Bulinus*, *Biomphalaria*, and *Oncomelania* snails, respectively.
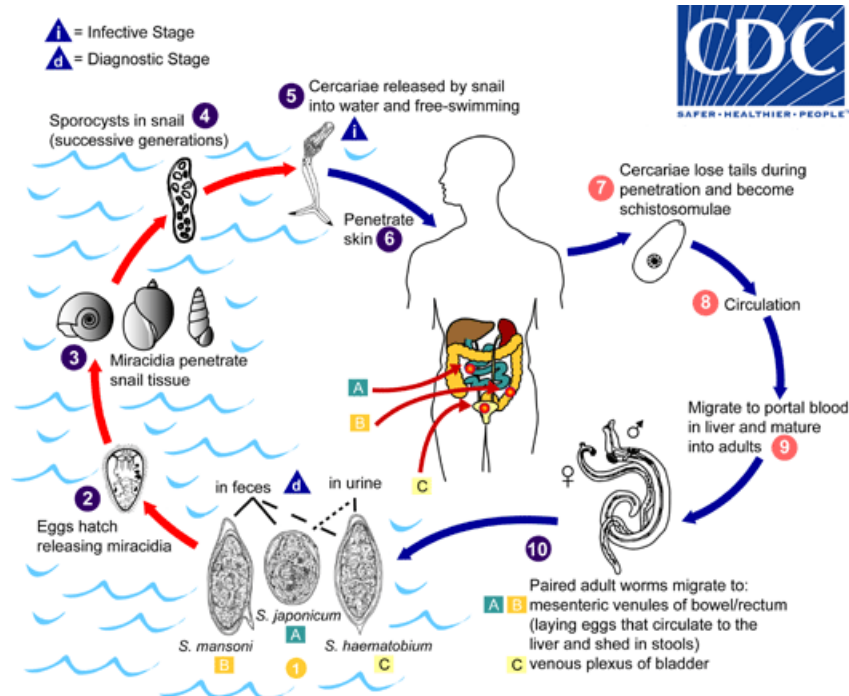
Figure 1.5: Life cycle of schistosomes (source: CDC)

1.1.2.2 Clinical conditions

Schistosome infections usually cause intestinal and hepatosplenic schistosomiasis, except for the infection of *S. haematobium*, which often leads to urogenital schistosomiasis. The earliest symptoms of infection include temporary or sometimes consist of skin rash or prurit, which are induced by percutaneous penetration of cercariae, often unrecognized in endemic areas (Appleton 1984). Generally, the morbidity of schistosomiasis is predominantly caused by host's immune response to schistosome eggs (Burke et al. 2009). Acute schistosomiasis, known as Katayama syndrome, appears between weeks to months after non-immune individuals exposed to first schistosome infection or heavy reinfection, typical clinical presentations of which include nocturnal fever, cough, myalgia, headache, and abdominal tenderness (Ross et al. 2007).

In general, chronic schistosomiasis is the most common form of the disease and often results in chronic anaemia, undernutrition, and children's growth stunting (King & Dangerfield-Cha 2008). Chronic intestinal schistosomiasis frequently presents as non-specific abdominal pain, diarrhea, dyspepsia, tenesmus, and anal pain (Elbaz & Esmat 2013). Some people can further develop to hepatosplenic disease, with clinical features such as splenomegaly, portal hypertension, oesophageal varices, haematemesis, melaena, and ascites

(De Cock 1986). One the other hand, urogenital schistosomiasis often appears with a classic clinical presentation as haematuria, often with urinary frequency, burning micturition, and suprapublic discomfort (Colley et al. 2014). If not properly treated, long-standing urinary complications may result in serious sequelae such as chronic bladder ulcers, leucoplakia, vesical granuloma, contracted bladder, bladder neck contracture, and stricture ureters, which may further lead to lethal consequences from renal failure or bladder cancer (Khalaf, Shokeir, & Shalaby 2012). Female genital schistosomiasis not only affects women's reproductive health, but also make them at a higher risk of HIV acquisition (Kjetland, Leutscher, & Ndhlovu 2012).

1.1.2.3 Epidemiology

It was estimated in 2008 that approximately 240 million people were infected with schistosomiasis in 76 endemic countries or territories of Africa, the Americas, the Eastern Mediterranean, and eastern Asia (WHO 2010a) (Figure 1.6). More than 90% of all cases occur in Africa (Stothard et al. 2009). In addition, due to lack of morbidity control, most of severe cases are found in Africa, even though more pathogenic type of schistosomiasis appears in Asia (Bruun & Aagaard-Hansen 2008). A global burden caused by schistosomiasis in 2010 was estimated to 3.3 million DALYs, ranked the second among all NTDs (Murray et al. 2012).
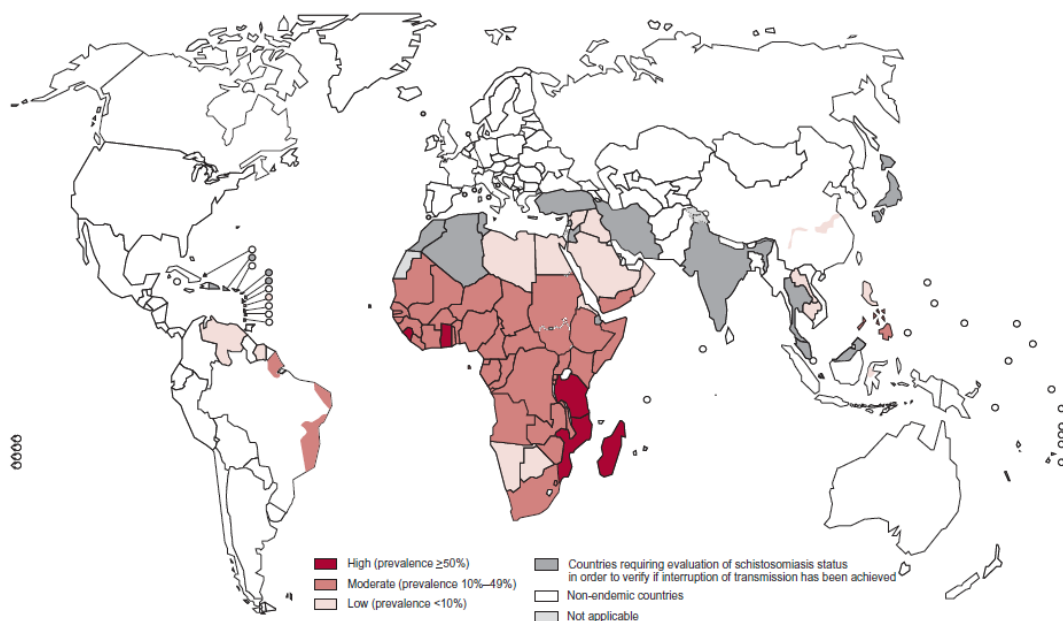
Figure 1.6: Endemic regions of schistosomiasis, 2011 (source: WHO 2013)

Schistosomiasis is an age-related disease, a typical pattern of which shows the prevalence and infection intensity increase in young children, reach a peak during school age to early adulthood, then decline and become stable at a certain age level (Woolhouse 1991). This pattern may be attributed to age-related water contact activities and/or development of

acquired resistance (immunity) (Warren 1973;Yang 2003). However, a few researchers reported that high prevalence levels may continue during adult life or a second peak may exist in older age groups in some endemic populations (Enk et al. 2008;Mutapi, Gryseels, & Roddam 2003;Raso et al. 2007).

1.1.2.4 Risk factors

Schistosomiasis is a water-associated disease that exposure to contaminated water is a determined risk factor for transmission. Environmental factors, such as temperature, precipitation, vegetation, and land cover, play an important role on transmission by either influencing the intermediate host snail population and the parasite development outside human host, or affecting human activities related to water contact. For example, very low temperature limits the snail distribution and parasite maturation outside the human host, while temperature too high may limit the fecundity and survival of snails (Appleton 1977;McCreesh & Booth 2013). High precipitation can either increase the risk of transmission by increasing the contact rate of human beings to contaminated water or decrease it by creating for example fast-flowing water unsuitable for cercaria or snail survival (McCreesh & Booth 2013). Ecological transformations, such as construction of dams and changes of irrigation schemes, can influence the distribution of snail species and thus become potential risk factors (Steinmann et al. 2006).

Socioeconomic factors (e.g., education, occupation, and wealth/poverty) influencing the behavior of people, are important for schistosomiasis transmission (Gazzinelli et al. 2006;Huang & Manderson 2005;Ximenes et al. 2003). Particularly, improvement of WASH can become a protection factor: water from safe supplies is schistosome-free and hence reduces the exposure to contaminated water; improvement of sanitation declines the risk of egg contamination with excreta to fresh water bodies; soap use related to better hygiene may protect people from infection during human water contact (Grimes et al. 2015;Utzinger et al. 2003).

1.1.2.5 Diagnosis and treatment

The gold standard diagnosis of active schistosomiasis is the detection of eggs in excreta (i.e., urine for *S. haematobium* and stool for other species) via microscopic examination, but the sensitivity is low due to large inter- and intra-specimen variations (Lamberton et al. 2014). Schistosome eggs are easy to identify on microscopy according to their characteristic shapes and sizes (Gray et al. 2011). Kato-Katz thick smear stool examination is recommended by WHO for intestinal schistosomiasis, as it is simple, rapid and low cost (Teesdale & Amin 1976). The sensitivity can be increased by increasing number of stool specimens and slides per sample (Raso et al. 2007). Direct thick smear and formalin based techniques for sedimentation and concentration are sometimes used for detection of intestinal schistosomiasis in endemic areas (Gray et al. 2011). Urine sedimentation, centrifugation and filtration are applied to microscopically detection of *S. haematobium* eggs. In endemic areas,

microhaematuria on reagent strips or self-reported blood in urine are rapid ways for indicating potential infections, but with low specificity (Gray et al. 2011). On the other hand, PCR techniques based on the detection of *Schistosoma* species DNA in faeces, serum, plasma and urine show high sensitivity and specificity, thus may become potential alternatives for diagnosis of the disease (Enk, Silva, & Rodrigues 2012;Gomes et al. 2010;Sandoval et al. 2006;Wichmann et al. 2009). Serological assays detecting antibodies against schistosomal antigens show high sensitive, but are unable to distinguish active or past infections (Colley et al. 2014). Point-of-care circulating cathodic antigen (POC-CCA) dipstick/cassette test overcomes this difficulty and is recommended for the rapid identification of *S. mansoni* infections in large-scale epidemiological surveys (Coulibaly et al. 2013a;Foo et al. 2015;Mwinzi et al. 2015). Additional new diagnostic techniques, such as a modified version of the miracidium hatching test for diagnosis of any schistosome species, or the Mini-FLOTAC for the detection of *S. mansoni* or *S. japonicum*, need further evaluation (Knopp et al. 2013).

Praziquantel is the current drug of choice for schistosomiasis, which is effective for all *Schistosoma* species and considered safe for treatment of children and pregnant women (WHO 2002b). A standard dose of 40 mg/kg is recommended for treatment of *S. haematobium* and *S. mansoni* infections, while 60mg/kg in split doses is recommended for treatment of *S. japonicum* and *S. mekongi* infections, or treatment in populations with high initial egg counts (Colley et al. 2014;Gryseels et al. 2006). Side effects are mild and transient, commonly including abdominal pain, headache, nausea, dizziness and fever (Jaoko, Muchemi, & Oguya 1996). Since praziquantel has little effect on eggs and immature schistosome worms, repeated treatment for several weeks can be more effective in treating the initially resistant immature forms after they have matured into drug-susceptible adult worms (King et al. 2011). Even though there is no clear evidence for existing of praziquantel resistance, the threat of emerging resistance remains, as resistance can be induced in animals under laboratory conditions and a reduced susceptibility of the drug in *S. mansoni* has been found in some endemic foci (Wang, Wang, & Liang 2012). Artemisinin derivatives (e.g., artemether and artesunate), active against *S. japonicum*, *S. mansoni*, and *S. haematobium*, mainly target the developmental stages of the parasites (Liu et al. 2011). Therefore, artemisinin derivatives in combination with praziquantel can increase the cure rates in schistosomiasis treatment (del Villar et al. 2012). However, investigations for dosing, formulation, and drug interactions are needed before standardizing the combination treatment (Colley et al. 2014). In addition, such treatment is not recommended in malaria endemic regions to avoid the potential induction of artemisinin resistance in malaria parasites (Gryseels et al. 2006).

1.1.2.6 Control and prevention

Large-scale preventive chemotherapy with praziquantel is advocated by WHO for morbidity control of schistosomiasis (WHO 2002a;WHO 2006). In high-risk (prevalence≥

50%) areas, preventive chemotherapy is recommended to all school-aged children once a year, while in moderate-risk (prevalence≥10% and <50%) and low-risk (prevalence<10%) areas, all school-aged children are recommended to be treated once every two years and twice during their primary schooling age, respectively. In high-risk and moderate-risk areas, adults considered to be at risk are also recommended for preventive chemotherapy (WHO 2006). In addition, inclusion of preschool-aged children to the administration of praziquantel is suggested by WHO in ongoing public-health interventions (WHO 2011b).

Besides preventive chemotherapy, behavioural modification, which reduces both the exposure of people to contaminated water and the contamination of snail habitat by human excreta with schistsome eggs, may be a possible approach for control of the disease (Colley et al. 2014). However, people's behavior is difficult to change unless in conjunction with other interventions, such as health education and improvements of WASH (Grimes et al. 2015;Lansdown et al. 2002). Increasing access to safe water and adequate sanitation are suggested as important and sustainable measures to reduce the risk of schistosome infection (Grimes et al. 2014). On the other hand, snail control is considered as an alternative measure for interruption of transmission (Lardans & Dissous 1998). The use of molluscicide niclosamide is the primary method for chemical snail control, because of its very low toxicity for humans and livestock and its ability to kill snails, their eggs, and cercariae at low concentrations (King & Bertsch 2015). However, as long-term continued use of mollusciding needs high labor cost and may have a negative impact on aquatic life, mollusciding should be restricted to areas that aim for schistosomiasis elimination (King & Bertsch 2015;Knopp et al. 2012;Oliveira & Paumgartten 2000).

To achieve sustainable control and widespread elimination of schistosomiasis, integrated control strategies should be applied, for example preventive chemotherapy in combination with behavioural modification, health education, improvements of WASH, and snail control.

1.1.3 Clonorchiasis

Clonorchiasis is one of the most important foodborne trematodiasis (Fürst, Keiser, & Utzinger 2012). It is caused by infection with the Chinese liver fluke, *Clonorchis sinensis* (Lun et al. 2005).

1.1.3.1 Parasite and life cycle

The adult fluke *C. sinensis* is a leaf-shaped slender digenetic trematode, 15–20 mm long and 3–4 mm wide (Hong & Fang 2012). Eggs laid by hermaphroditic adult worms reach the intestine with bile fluids and are emitted with the faeces into the water (Qian et al. 2016) (Figure 1.7). The first intermediate hosts, freshwater snails, ingest the eggs, which are further hatch into miracidiae. Inside the nails, miracidiae subsequently develop to sporocysts, rediae and cercariae, through asexual reproduction (Lun et al. 2005). The free-swimming cercariae leave the snails and adhere to the second intermediate host, freshwater fish or shrimp, in

which cercariae develop into mature metacercariae (Hong & Fang 2012). The definite hosts, human beings or other piscivorous mammals, get infected by eating raw or insufficient cooked infected fish. Metacercariae reach the human small intestine and further navigate to the liver, where they develop into adult flukes and reach the stage of sexual reproduction (Rim 1986). The egg productivity of an adult worm in human is estimated at around 4000 per day (Kim et al. 2011). Usually after four weeks of infection, eggs can be detected in faeces (Hsü & Wang 1938).
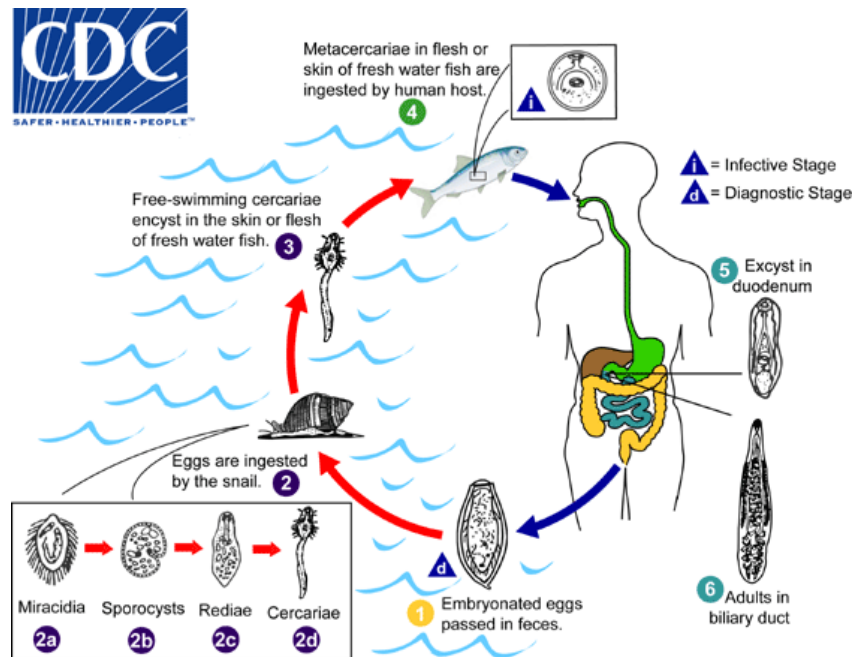
Figure 1.7: Life cycle of *Clonorchis sinensis* (source: CDC)

1.1.3.2 Clinical conditions

The clinical manifestations of clonorchiasis tend to relate to worm burden but are variable and unspecific (Kim et al. 2011;Lun et al. 2005;Rim 1986). People with small infection intensity have few or mild symptoms (e.g., abdominal discomfort, diarrhea, and/or malaise), while people with moderate to high infection intensity present more pronounced symptoms (e.g., fever, chills, anorexia, weight loss, colic, fatigue, and/or abdominal distension) (Lun et al. 2005). Typical physical signs of clonorchiasis include jaundice, hepatomegaly, and liver tenderness. Chronic infection usually results in complications in liver and biliary systems (e.g., cholelithiasis, cholangitis, and cholecystitis) (Qian et al. 2016). Furthermore, *C. sinensis* is classified as a definite carcinogen, as infection can increase the risk of cholangiocarcinoma, according to different studies (Bouvard et al. 2009;Fürst et al. 2012;Qian et al. 2012;Shin et al. 2010).

1.1.3.3 Epidemiology
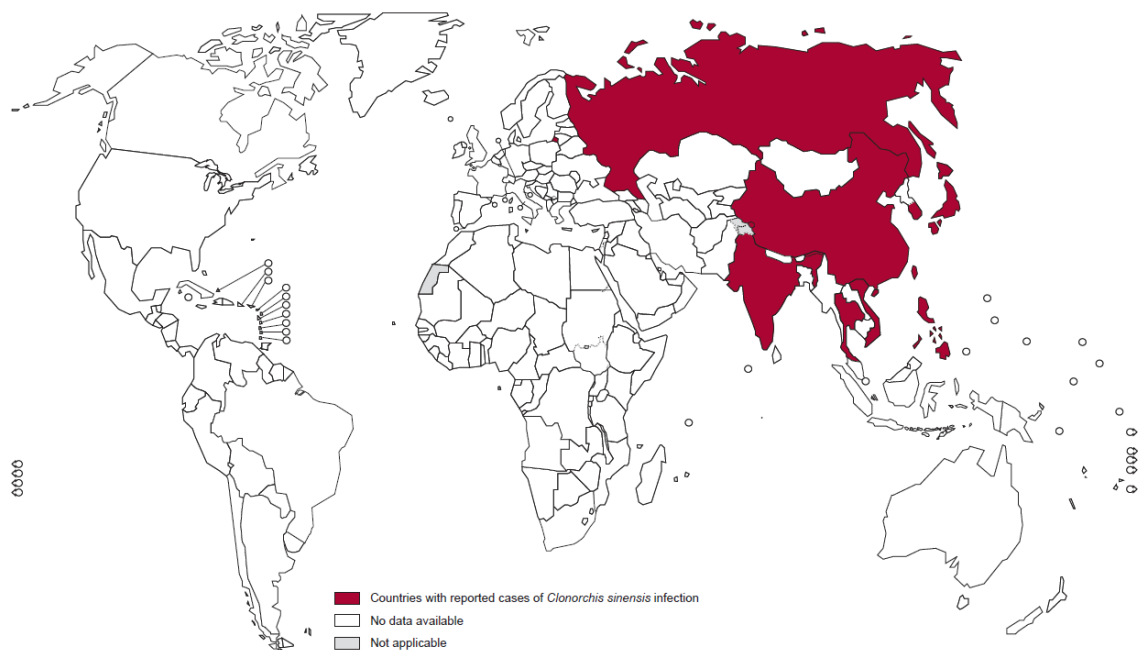
Figure 1.8: Endemic regions of clonorchiasis (source: WHO 2013)

It was estimated conservatively that around 15 million people were infected with *C. sinensis* in 2004, predominantly in countries of Asia, particularly in China, South Korea, northern Vietnam and parts of Russia (Fürst et al. 2012;Qian et al. 2012;Qian et al. 2016;Qian, Chen, & Yan 2013) (Figure 1.8). China accounts for around 85% of the global infected people, corresponding to 12.5 million people infected (Qian et al. 2012). Two major endemic regions were identified for human clonorchiasis in China, namely the provinces of Guangdong and Guangxi in the south and the provinces of Heilongjiang and Jilin in the north-east (Lun et al. 2005;Qian et al. 2012;Qian et al. 2016). In South Korea, *C. sinensis* infection is the major intestinal parasitic infection, with an estimation of 1.2 million people infected, according to a nationwide survey in 2004 (Kim et al. 2009). High endemic areas were reported along the four major rivers (Nakdong-fang, Seomjin-gang, Zoungsan-gang, and Guem-gang) in the southern part of the country (Cho et al. 2008). Around one million people in Vietnam (mainly the northern part) and 3000 people in the far east of Russia were reported to be infected with *C. sinensis* by a WHO report in 1995, however, there is no updated country-level reports for the two countries since then (Chau et al. 2001;Kino et al. 1998;WHO 1995).

Clonorchiasis was estimated to attribute to a disease burden of 275 thousand DALYs in 2005 (Fürst et al. 2012). However, the burden was considered to be largely underestimated due to the exclusion of light to moderate infections in the calculation (Qian et al. 2016). In general, males show higher prevalence than women and the prevalence increases with age (Fang et al. 2008;Qian et al. 2012).

1.1.3.4 Risk factors

Environmental and climatic factors affect the endemicity of *C. sinensis* infection, mainly thought influencing the distribution of the intermediate hosts. For example, temperature and climatic change have an impact on the activities, survival and reproduction rate of the intermediate hosts, thus are regarded as potential risk factors (Li et al. 1983;Petney et al. 2013). Factors such as precipitation, land cover/usage, and aquaculture that affect the presence, quality, and current of fresh water bodies (reservoirs for intermediate hosts), can also be potential risk factors (Keiser & Utzinger 2005). Areas adjacent to water bodies were reported to correlate with high infection risk of *C. sinensis*, however, such situation may be changing due to improvement of trade and transportation channels (Keiser & Utzinger 2005;Sripa et al. 2010).

One the other hand, socioeconomic factors and consumption of raw freshwater fish are important in understanding the epidemiology of clonorchiasis (Phan et al. 2011). Consumption of raw fish dishes is a traditional rooted culture practice in some areas of China, while in other areas it is considered delicious or highly nutritious by some people (Qian et al. 2013a;Tang et al. 1963a;Zheng 2009). In addition, lack of self-protection awareness of food hygiene influencing people's behavior of raw-fish-consumption, can be an important risk factor (Han et al. 2013).

1.1.3.5 Diagnosis and treatment

The gold standard of diagnosis for *C. sinensis* infection is the detection of eggs in stool (Qian et al. 2016). Kato-Katz method is the most widely used technique with the advantages of simplicity, low cost and the ability to quantify the infection intensity, but the sensitivity is low (Hong et al. 2003;Qian et al. 2016). Direct stool smear and formalin-ether concentration technique are sometimes used but also with low sensitivity (Hong et al. 2003;Qian et al. 2013b). Multiple Kato-Katz thick smears are recommended to increase the accuracy of diagnosis (Qian et al. 2013b). On the other hand, immunodiagnostic techniques are employed as supplementary methods, among which serodiagnosis by the enzyme-linked immunosorbent assay (ELISA) is the most commonly used one (Qu, Chen, & Zeng 1980). However, the main limitations of ELISA is its cross-reactivity and inability to differentiate between past and active infection (Chen, Hu, & Shen 1988). Other immunodiagnostic techniques such as complement fixation, agglutination, and immunoelectrophoresis are seldom used in epidemiological studies (Qian et al. 2016). Molecular biological methods such as PCR-based/coupled technologies and loop-mediated isothermal amplification (LAMP) technique have been developed showing high performance and accuracy, however, they are inconvenient for large-scale epidemiological surveys due to the need for laboratory facilities, trained personnel, and financial supports (Han et al. 2012;Huang et al. 2012). In addition, imaging diagnosis is a complementary method for clonorchiasis in clinical practices (Choi & Hong 2007).

Praziquantel is the only recommended medicine by WHO for treatment of clonorchiasis, which is safe, well tolerated and effective (WHO 2013;Yangco et al. 1987). Mild and transient adverse effects are sometimes reported, such as dizziness, sleepiness, headache, and diarrhea (Hong & Fang 2012). In rare occasions, sever adverse events (e.g., anaphylactic reaction) may occur (Lee, Lim, & Hong 2011;Shen et al. 2007). Broad-spectrum anthelmintic drugs, such as albendazole and tribendimidine, also show good efficacy in clinical trials when against *C. sinensis* infection or its co-infection with other helminths (Liu et al. 1991;Xu et al. 2014).

1.1.3.6 Control and prevention

The recommended treatment guidelines of clonorchiasis by WHO for preventive chemotherapy advocate praziquantel administration for all residents every year in high endemic areas (prevalence ≥20%) and for all residents every two years or individuals regularly eating raw fish every year in moderate endemic areas (prevalence <20%) (WHO 2013). In order to maintain control sustainability, a comprehensive control strategy, such as preventive chemotherapy in combination with IEC and environmental modification, should be considered (Oh et al. 2014;Zhang, Huang F.Y., & Geng Y.J. 2009). Through IEC, residents may conscientiously reduce or stop the consumption of raw fish and pay more attention to food hygiene (Wu et al. 2012). In addition, by improving sanitation facilities around fish ponds, the chance of intermediate hosts becoming infected with *C. sinensis* will decrease (Wu et al. 2012;Zhang et al. 2009).

1.2 Geographical distribution of disease risk

High-resolution maps depicting the geographical distribution of disease risk assist disease control by delivering control interventions at areas of highest risk, monitoring and evaluating effectiveness of control programmes. Geostatistical modeling is the most rigorous inferential approach for predicting the disease risk at areas without observed data by relating survey data with potential predictors (e.g., environmental and socioeconomic factors). In the absence of single surveys covering large areas, survey data are compiled from bibliometric searches. Mathematical models can be combined with geostatistical models to address data heterogeneities between locations. Furthermore, the number of individuals infected and the number of people at risk can be estimated by overlapping the high-resolution risk maps with gridded population surfaces. Population-adjusted prevalence and treatment needs for preventive chemotherapy can be further calculated at different administrative levels.

1.2.1. Data

1.2.1.1 Disease data

Goereferenced disease data (e.g., prevalence, or number of people infected from a sample of people examined) are fundamental for the data-driven geostatistical modeling. To capture the heterogeneity of infection risk over the surface, it is necessary to have a good coverage and a fine amount of observed data across the study region. Disease data can be obtained from systematic review of related literature, or through other sources (e.g., thesis, working papers, research reports, or personal communication). Recently, two large databases are available for neglected tropical diseases, that is the Global Neglected Tropical Diseases Database (GNTD, www.gntd.org) and the Global Atlas of Helminth Infections (GAHI, www.thiswormyworld.org). The GNTD database is a georeferenced, open-access global database, which is constantly updated and can be utilized by researchers, scientists, disease control managers, and policy makers (Hürlimann et al. 2011). Data obtained and extracted into the GNTD database follows a standard protocol, according to which detailed information for each reference is recorded (e.g., sources, survey description, location information, and parasitological data) (Hürlimann et al. 2011). Data compiled from bibliometric searches are often heterogeneous in the age groups of the population surveyed across locations and in the diagnostic methods used between the studies collecting these data. Furthermore publications may either report the actual geographical location of the survey or may provide aggregated survey data within areas. The above data characteristics complicate their analysis.

1.2.1.2 Environmental, climatic and socioeconomic data

Environmental, climatic and socioeconomic factors play an important role in transmission of NTDs, thus they are employed as predictors in geostatistical modeling. The spatial distribution of these data is usually obtained from different readily accessible data sources. For example, environmental proxies such as land surface temperature (LST), normalized difference vegetation index (NDVI), and land cover are accessible from the remote sensing source MODIS/Terra (Savtchenko et al. 2004). Geographical climatic zones can be obtained from a digital map of Köppen-Geiger climate classification (Kottek et al. 2006). Socioeconomic proxies such as human influence index (HII), gross domestic product (GDP), and infant mortality rates (IMR) can be accessed from Socioeconomic Data and Applications Center (SEDAC, http://sedac.ciesin.columbia.edu). WASH indicators (e.g., proportion of households with improved sanitation, proportion of households with improved drinking water sources, and proportion of households practicing open defecation) can be obtained from data compiled from household surveys conducted by Demographic and Health Surveys (DHS), Multiple Cluster Indicator Surveys (MICS), World Health Surveys (WHS) and Living Standards Measurement Study (LSMS).

1.2.2 Bayesian geostatistical modeling

Three distinct types of spatial data exist: geostatistical or point-referenced data (data observed at a set of locations), areal data (aggregated data available over a set of regions with common boarders), and point pattern data (sets of locations in a region with particular event occurring) (Banerjee, Carlin, & Gelfand 2014). This PhD thesis is mainly focus on point-referenced observed data. Geostatistical modeling is the most rigorous inferential approach to predict the infection risk of NTDs based on point-referenced data. Environmental, climatic, and socioeconomic factors are often employed into the models as explanatory variables (predictors). In addition, as neighbouring areas have similar infection status due to common disease exposures, location-specific parameters (random effects) are introduced to explain the residual spatial variation after accounting for all known explanatory variables (Diggle, Tawn, & Moyeed 1998). These random effects are assumed to be latent observations of underlying spatial process that follow a zero-mean multivariate Gaussian distribution. If the spatial association only depends upon distance between locations (isotropy), the covariance matrix can be constructed by parametric functions of distance such as exponential or Matérn functions, suggesting a decrease of spatial correlation with an increase of distance (Matérn 1960).

The Bayesian approach provides a coherent framework for geostatistical modeling of both the observed data and the unknown parameters such as the coefficients of the predictors and the spatial process parameters (Banerjee et al. 2014). Inferences are based on posterior distributions that are not available analytically in most realistic problems (Gelman et al. 2013). Markov chain Monte Carlo (MCMC) methods, such as the Metropolis-Hastings algorithm and the Gibbs sampler, are powerful computational tools in Bayesian inference that approximate the posterior distributions from samples simulated iteratively from a particular Markov chain, whose stationary distribution is the posterior distribution of interest (Gelfand & Smith 1990;Geman & Geman 1984;Hastings 1970;Metropolis et al. 1953).

Implementation of MCMC algorithms for geostatistical modeling requires repeated inversions of the covariance matrix, whose dimension is large for large spatial datasets, leading to heavy computations. The approaches to overcome the above so-called "big n problem" can be grouped into two main categories: those that approximate the exact likelihood and those that reduce the dimension of the problem (Banerjee et al. 2014). Stochastic partial differential equations (SPDE) and integrated nested Laplace approximation (INLA) approximate the likelihood by constructing a Gaussian Markov random field (GMRF) representation of the latent spatial process and estimate the posterior distribution by a Laplace approximation (Lindgren, Rue, & Lindstrom 2011;Rue, Martino, & Chopin 2009). On the other hand, Gaussian predictive processes project the spatial process to a lower dimensional subspace (Banerjee et al. 2008;Banerjee et al. 2010;Finley et al. 2009).

In recent years, Bayesian geostatistical modeling approaches have been widely applied to obtain high-resolution infection risk estimates for many NTDs at regional, country and continental scales. Examples include mapping leishmaniasis incidence in Brazil (Karagiannis-Voules et al. 2013), and risk of lymphatic filariasis in Uganda (Stensgaard et al. 2011a), of schistosomiasis in East and West Africa (Schur et al. 2011a;Schur et al. 2011b), of soil-transmitted helminthiasis in sub-Saharan Africa and South America (Chammartin et al. 2013c;Karagiannis-Voules et al. 2015a), and of trachoma in Southern Sudan (Clements et al. 2010b).

1.2.3 Mathematical modeling

Mathematical models play an important role on understanding the transmission dynamics of NTDs, and thus support control and elimination programmes (Basáñez & Anderson 2015). Compared to data-driven statistical models, mathematical models are sets of equations or computing rules that represent a biological system by a quantitative description of the process that drives the dynamical changes (Nouvellet, Cucunubá, & Gourbière 2015). The models can be deterministic or stochastic (Basáñez et al. 2012). Maximum likelihood methods and Bayesian inference algorithms can be employed for parameter estimation (Anderson, Truscott, & Hollingsworth 2014;Gambhir, Singh, & Michael 2015;Raso et al. 2007;Truscott, Turner, & Anderson 2015). In recent years, mathematical models have been widely applied for many neglected tropical diseases, either to study the factors influencing the persistence of the diseases (e.g., dengue (Medeiros et al. 2011) and lymphatic filariasis (Gambhir et al. 2015)), or to evaluate different strategies or intervention scenarios for control of the disease (e.g., leprosy (Blok et al. 2015), human African trypanosomiasis (Rock et al. 2015), Chagas disease (Nouvellet et al. 2015), soil-transmitted helminth infections (Anderson et al. 2014;Anderson et al. 2013;Levecke et al. 2015;Truscott, Hollingsworth, & Anderson 2014;Truscott et al. 2015), lymphatic filariasis (Kastner et al. 2015;Stolk, Stone, & de Vlas 2015), and schistosomiasis (French et al. 2010;French et al. 2015;Lamberton et al. 2015;Zhang, Feng, & Milner 2007)). In addition, cost-benefit and cost-effectiveness models are used for economic and financial evaluation of NTDs (Lee, Bartsch, & Gorham 2015).

Mathematical models can be also used to describe useful epidemiology quantities such as the age-related pattern of the disease risk. Models available for estimating age-prevalence curve for example of schistosomiasis include Hairston's two-stage catalytic models (Hairston 1965), Holford and Hardy's immigration-death model (Holford & Hardy 1976), Chan *et al*'s fully age-structured models (Chan et al. 1995), and Yang *et al*'s semi-stochastic acquired immunity model (Yang 2003;Yang, Coutinho, & Massad 1997). These models can be combined with geostatistical models to standardise age-heterogeneous survey data and obtain age-specific risk estimates, however such formulations are not yet available for NTD data.

1.3 Goal and objectives

1.3.1 Goal

The overall goal of the thesis is to develop Bayesian geostatistical and mathematical models for analysing geo-referenced NTD survey data and to provide tools and knowledge for disease control and prevention.

1.3.2 Specific objectives

The following specific objectives were linked to this goal:

(i) produce population-adjusted risk maps for soil-transmitted helminth and for *C. sinensis* infections at high spatial resolution in P.R. China and estimate the number of infected people by province (Chapter 2 and 4);

(ii) provide geo-referenced estimates of soil-transmitted helminth infection risk and the number of infected by country across South Asia (Chapter 3);

(iii) predict the distribution of schistosomiasis infection risk in sub-Saharan Africa, evaluate temporal trends and provide spatially explicit estimates of people infected and of treatment needs for preventive chemotherapy by country (Chapter 5);

(iv) develop Bayesian geostatistical models for analyzing jointly point-referenced and areal NTD survey data (Chapter 6);

(v) integrate Bayesian geostatistical and mathematical transmission models of schistosomiasis within a single model formulation and obtain age-specific estimates of the disease risk at high geographical resolution from age-heterogeneous surveys (Chapter 7).

Chapter 2　Bayesian geostatistical modelling of soil-transmitted helminth survey data in the People's Republic of China

Ying-Si Lai[1,2], Xiao-Nong Zhou[3], Jürg Utzinger[1,2] and Penelope Vounatsou[1,2]

[1]Department of Epidemiology and Public Health, Swiss Tropical and Public Health Institute, P.O. Box, CH-4002 Basel, Switzerland

[2]University of Basel, Petersplatz 1, CH-4003 Basel, Switzerland

[3]National Institute of Parasitic Diseases, Chinese Center for Disease Control and Prevention; WHO Collaborating Centre for Malaria, Schistosomiasis and Filariasis; Key Laboratory of Parasite and Vector Biology, Ministry of Health, Shanghai 200025, People's Republic of China

Abstract

Background: Soil-transmitted helminth infections affect tens of millions of individuals in the People's Republic of China (P.R. China). There is a need for high-resolution estimates of at-risk areas and number of people infected to enhance spatial targeting of control interventions. However, such information is not yet available for P.R. China.

Methods: A geo-referenced database compiling surveys pertaining to soil-transmitted helminthiasis, carried out from 2000 onwards in P.R. China, was established. Bayesian geostatistical models relating the observed survey data with potential climatic, environmental and socioeconomic predictors were developed and used to predict at-risk areas at high spatial resolution. Predictors were extracted from remote sensing and other readily accessible open-source databases. Advanced Bayesian variable selection methods were employed to develop a parsimonious model.

Results: Our results indicate that the prevalence of soil-transmitted helminth infections in P.R. China considerably decreased from 2005 onwards. Yet, some 144 million people were estimated to be infected in 2010. High prevalence (>20%) of the roundworm *Ascaris lumbricoides* infection was predicted for large areas of Guizhou province, the southern part of Hubei and Sichuan provinces, while the northern part and the south-eastern coastal-line areas of P.R. China had low prevalence (<5%). High infection prevalence (>20%) with hookworm was found in Hainan, the eastern part of Sichuan and the southern part of Yunnan provinces. High infection prevalence (>20%) with the whipworm *Trichuris trichiura* was found in a few small areas of south P.R. China. Very low prevalence (<0.1%) of hookworm and whipworm infections were predicted for the northern parts of P.R. China.

Conclusions: We present the first model-based estimates for soil-transmitted helminth infections throughout P.R. China at high spatial resolution. Our prediction maps provide useful information for the spatial targeting of soil-transmitted helminthiasis control interventions and for long-term monitoring and surveillance in the frame of enhanced efforts to control and eliminate the public health burden of these parasitic worm infections.

Keywords: Soil-transmitted helminths, *Ascaris lumbricoides*, *Trichuris trichiura*, Hookworm, Bayesian geostatistics, People's Republic of China

2.1 Background

Soil-transmitted helminths are a group of parasitic nematode worms causing human infection through contact with parasite eggs (*Ascaris lumbricoides* and *Trichuris trichiura*) or larvae (hookworm) that thrive in the warm and moist soil of the world's tropical and subtropical countries (Bethony et al. 2006). More than 5 billion people are at risk of soil-transmitted helminthiasis (Pullan & Brooker 2012). Estimates published in 2003 suggest that 1,221 million people were infected with *A. lumbricoides*, 795 million with *T. trichiura* and 740 million with hookworms (de Silva et al. 2003). The greatest number of soil-transmitted helminth infections at that time occurred in the Americas, the People's Republic of China (P.R. China), East Asia and sub-Saharan Africa (Hotez et al. 2006). Socioeconomic development and large-scale control efforts have lowered the number of people infected with soil-transmitted helminths in many parts of the world (Bethony et al. 2006). For the year 2010, the global burden due to soil-transmitted helminthiasis has been estimated at 5.2 million disability-adjusted life years (Murray et al. 2012).

In P.R. China, there have been two national surveys for parasitic diseases, including soil-transmitted helminthiasis. Both surveys used the Kato-Katz technique as the diagnostic approach, based on a single Kato-Katz thick smear obtained from one stool sample per individual. The first national survey was conducted from 1988 to 1992 and the second in 2001-2004. In the first survey, there were a total of 2,848 study sites with approximately 500 people examined per site. The survey indicated overall prevalences of 47.0%, 18.8% and 17.2% for *A. lumbricoides*, *T. trichiura* and hookworm infections, respectively, corresponding to 531 million, 212 million and 194 million infected people, respectively (Xu et al. 1995). The second survey involved 687 study sites and there were 356,629 individuals examined overall. Analyses of the data revealed considerably lower prevalences for soil-transmitted helminth infections than in the first survey; *A. lumbricoides*, hookworm and *T. trichiura* prevalences were 12.7%, 6.1% and 4.6%, respectively (Coordinating Office of the National Survey on the Important Human Parasitic Diseases 2005). However, interventions were less likely to reach marginalized communities in the poorest areas (Zheng et al. 2009) and the diseases re-emerged whenever control measures were discontinued (Li et al. 2010;Wang et al. 2012). To overcome the challenge of parasite infections in P.R. China, in 2005, the Chinese Ministry of Health issued the "National Control Program on Important Parasitic Diseases from 2006 to 2015" with its target to reduce the prevalence of helminth infections by 70% by the year 2015 (Zheng et al. 2009). The key strategy for control was large-scale administration of anthelminthic drugs in high prevalence areas, especially targeting school-aged children and people living in rural areas (Li et al. 2010;Zhou, Bergquist, & Tanner 2013).

Maps depicting the geographical distribution of the disease risk can aid control programmes to deliver cost-effective interventions and assist in monitoring and evaluation. The Coordinating Office of the National Survey on the Important Human Parasitic Diseases (2005) in P.R. China obtained prevalence maps by averaging the data of the second national

survey within each province. To our knowledge, high-resolution, model-based maps using available national survey data are not available to date in P.R. China. Model-based geostatistics predict the disease prevalence at places without observed data by quantifying the relation between the disease risk at observed locations with potential predictors such as socioeconomic, environmental, climatic and ecological information, the latter often obtained via remote sensing. Model-based geostatistics have been used before to map and predict the geographical distribution of soil-transmitted helminth infections in Africa (Pullan et al. 2011;Raso et al. 2006), Asia and Latin America (Chammartin et al. 2013c;Pullan et al. 2008;Scholte et al. 2013). Model-based geostatistics typically employ regression analysis with random effects at the locations of the observed data. The random effects are assumed to be latent observations from a zero-mean Gaussian process, which models spatial correlation to the data via a spatially structured covariance. Bayesian formulations enable model fit via Markov chain Monte Carlo (MCMC) simulation algorithms (Diggle et al. 1998;Gelfand et al. 1990) or other computational algorithms (e.g. integrated nested Laplace approximations (INLA) (Rue et al. 2009)). INLA is a computational approach for Bayesian inference and is an alternative to MCMC to overcome computational burden for obtaining the approximated posterior marginal distribution for the latent variables, as well as for the hyperparameters (Cameletti et al. 2013).

In this study, we aimed to: (i) identify the most important climatic, environmental and socioeconomic determinants of soil-transmitted helminth infections; and (ii) develop model-based Bayesian geostatistics to assess the geographical distribution and number of people infected with soil-transmitted helminths in P.R. China.

2.2 Methods

2.2.1 Ethical considerations

The work presented here is based on soil-transmitted helminth survey data derived from the second national survey and additional studies identified through an extensive review of the literature. All data in our study was extracted from published sources and they are aggregated over villages, towns or counties; therefore, do not contain information that is identifiable at individual or household level. Hence, there are no specific ethical considerations.

2.2.2 Disease data

Geo-referenced data on soil-transmitted helminth infections from the second national survey conducted in P.R. China from 2001 to 2004 were provided by the National Institute of Parasitic Diseases, Chinese Center for Diseases Control and Prevention (IPD, China CDC; Shanghai, P.R. China). Moreover, an extensive literature search was undertaken in PubMed

and China National Knowledge Internet (CNKI) from January 1, 2000 until April 25, 2013 to identify studies reporting village, town and county-level prevalence data of soil-transmitted helminth infections in P.R. China. Data were excluded if (i) they were from hospital surveys, post-intervention surveys, drug efficacy studies and clinical trials; (ii) reports on disease infection among travellers, military personnel, expatriates, mobile populations and other displaced or migrating populations; (iii) the geographical coordinates could not be identified; and (iv) the diagnostic technique was not reported (Hürlimann et al. 2011). Data were entered into the Global Neglected Tropical Diseases (GNTD) database, which is a geo-referenced, open-access source (Hürlimann et al. 2011). Geographical coordinates for the survey locations were obtained via Google maps, a free web mapping service application and technology system. As we focus on recent data pertaining to soil-transmitted helminth infections in P.R. China, we only considered surveys carried out from 2000 onwards.

2.2.3 Climatic, demographic and environmental data

Climatic, demographic and environmental data were downloaded from different readily accessible remote sensing data sources, as shown in Table 2.1. Land surface temperature (LST) and normalized difference vegetation index (NDVI) were calculated to annual averages and land cover data was summarised to the most frequent category over the period of 2001-2004. Moreover, land cover data were re-grouped into six categories based on between-class similarities: (i) forest; (ii) shrubland and savanna; (iii) grassland; (iv) cropland; (v) urban; and (vi) wet areas. Monthly precipitation values were averaged to obtain a long-term average for the period 1950-2000. Four climatic zones were considered: (i) equatorial; (ii) arid; (iii) warm; and (iv) snow/polar. The following 13 soil types, which may be related to the viability of parasites or microorganisms living in the soil, were used: (i) percentage of coarse fragments (CFRAG, % >2 mm); (ii) percentage of sand (SDTO, mass %); (iii) percentage of silt (STPC, mass %); (iv) percentage of clay (CLPC, mass %); (v) bulk density (BULK, km/dm3); (vi) available water capacity (TAWC, cm/m); (vii) base saturation as percentage of ECEsoil (BSAT); (viii) pH measured in water (PHAQ); (ix) gypsum content (GYPS, g/kg); (x) organic carbon content (TOTC, g/kg); (xi) total nitrogen (TOTN, g/kg); (xii) FAO texture class (PSCL); and (xiii) FAO soil drainage class (DRAIN). Human influence index (HII) was included in the analysis to capture direct human influence on ecosystems (Sanderson et al. 2002). Urban/rural extent was considered as a binary indicator. Gross domestic product (GDP) per capita was used as a proxy of people's socioeconomic status. We obtained GDP per capita for each county from the P.R. China Yearbook full-text database in 2008.

Moderate Resolution Imaging Spectroradiometer (MODIS) Reprojection Tool version 4.1 (EROS; Sioux Falls, USA) was applied to process MODIS/Terra data. All remotely sensed data were aligned over a prediction grid of 5×5 km spatial resolution using Visual Fortran version 6.0 (Digital Equipment Corporation; Maynard, USA). Data at the survey locations were also extracted in Visual Fortran. As the outcome of interest (i.e. infection prevalence

with a specific soil-transmitted helminth species) is not available at the resolution of the covariates for surveys aggregated over counties, we linked the centroid of those counties with the average value of each covariate within the counties. Distances to the nearest water bodies were calculated using ArcGIS version 9.3 (ERSI; Redlands, USA). For county-level surveys, the distances of all the 5×5 km pixel centroids to their nearest water bodies within the county were extracted and averaged. The arithmetic mean was used as a summary measure of continuous data, while the most frequent category was used to summarise categorical variables.

Table 2.1: Remote sensing data sources[a].

Source	Data type	Data period	Temporal resolution	Spatial resolution
MODIS/Terra[b]	LST[j]	2001-2012	8 days	1 km
MODIS/Terra[b]	NDVI[k]	2001-2012	16 days	1 km
MODIS/Terra[b]	Land cover	2001-2004	Yearly	1 km
WorldClim[c]	Elevation	2000	-	1 km
WorldClim[c]	Precipitation	1950-2000	Monthly	1 km
SWBD[d]	Water bodies	2000	-	30 m
Köppen-Geiger[e]	Climate zones	1976-2000	-	50 km
ISRIC[f]	Soil types	-	-	8 km
Atlas of the Biosphere[g]	Soil-moisture	1950-1999	-	50 km
SEDAC[h]	Population data	2000; 2010	-	5 km
SEDAC[h]	HII[l]	1995-2004	-	1 km
SEDAC[h]	Urban extents	1990-2000	-	1 km
China Yearbook[i]	GDP per capita	2008	-	County-level

[a] Land cover data accessed on June 1, 2011 and other data accessed on April 1, 2013.
[b] Moderate Resolution Imaging Spectroradiometer (MODIS)/Terra , available at: https://lpdaac.usgs.gov/
[c] Available at: http://www.worldclim.org/current.
[d] Shuttle Radar Topography Mission Water Body Data (SWBD), available at:
http://gis.ess.washington.edu/data/vector/worldshore/index.html.
[e] World maps of Köppen-Geiger climate classification, available at: http://koeppen-geiger.vu-wien.ac.at/shifts.htm.
[f] International Soil Reference and Information Center, available at: http://www.isric.org/data/isric-wise-derived-soil-properties-5-5-arc-minutes-global-grid-version-12.
[g] Available at: http://www.sage.wisc.edu/atlas/maps.php?datasetid=23&includerelatedlinks=1&dataset=23.
[h] Socioeconomic data and applications center, available at: http://sedac.ciesin.org/.
[i] China yearbook full-text database, available at: http://acad.cnki.net/Kns55/brief/result.aspx?dbPrefix=CYFD.
[j] Land surface temperature (LST) day and night.
[k] Normalized difference vegetation index.
[l] Human influence index.

2.2.4 Statistical analysis

The survey year was grouped into two categories: before 2005 and from 2005 onwards. Land cover, climatic zones, soil texture and soil drainage were included into the model as categorical covariates. Continuous variables were standardised to mean 0 and standard deviation 1 using the command "std()" in Stata version 10 (Stata Corp. LP; College Station, USA). Pearson's correlation was calculated between continuous variables. One of the two

variables, which had correlation coefficient greater than 0.8, was dropped to avoid collinearity (Dormann et al. 2013). Preliminary analysis indicated that for this dataset, three categories were sufficient to encapsulate for non-linearity of continuous variables, therefore we constructed 3-level categorical variables based on their distribution. Subsequent variable selection incorporated within the geostatistical model selected the most probable functional form (linear vs. categorical). Bivariate and multivariate logistic regressions were carried out in Stata version 10.

Bayesian geostatistical logistic regression models with location-specific random effects were fitted to obtain spatially explicit soil-transmitted helminth infection estimates. Let Y_i, n_i and p_i be the number of positive individuals, the number of those examined and the probability of infection at location i ($i = 1, 2, \ldots, L$), respectively. We assume that Y_i arises from a binominal distribution $Y_i \sim Bn(p_i, n_i)$, where:

$\text{logit}(p_i) = \beta_0 + \sum_{k=1} \beta_k \times X_i^{(k)} + \varepsilon_i + \phi_i$. β_k is the regression coefficient of the kth covariate $X_i^{(k)}$, ε_i is a location-specific random effect and ϕ_i is an exchangeable non-spatial random effect. To estimate the parameters, we formulate our model in a Bayesian framework. We assumed $\varepsilon = (\varepsilon_1, \ldots, \varepsilon_L)$ followed a zero-mean multivariate normal distribution, $\varepsilon \sim MVN(0, \Sigma)$, where Matérn covariance function $\Sigma_{ij} = \sigma_{sp}^2 (\kappa d_{ij})^\upsilon K_\upsilon(\kappa d_{ij}) / (\Gamma(\upsilon) 2^{\upsilon-1})$. d_{ij} is the Euclidean distance between locations i and j, κ is a scaling parameter, υ is a smoothing parameter fixed to 1 and K_υ denotes the modified Bessel function of second kind and order υ. The spatial range $\rho = \sqrt{8}/\kappa$, is the distance at which spatial correlation becomes negligible (<0.1) (Karagiannis-Voules et al. 2013). We assumed that ϕ_i follows a zero-mean normal distribution $\phi_i \sim N(0, \sigma_{nonsp}^2)$. A normal prior distribution was assigned to the regression coefficients, that is $\beta_0, \beta_k \sim N(0, 1000)$ and loggamma priors were adopted for the precision parameters, $\tau_{sp} = 1/\sigma_{sp}^2$ and $\tau_{nonsp} = 1/\sigma_{nonsp}^2$ on the log scale, that is $\log(\tau_{sp}) \sim \log gamma(1, 0.00005)$ and $\log(\tau_{nonsp}) \sim \log gamma(1, 0.00005)$. Furthermore, we assumed the following prior distribution for range parameter $\log(\rho) \sim \log gamma(1, 0.01)$

The most widely used computational approach for Bayesian geostatistical model fit is MCMC simulation. However, large spatial covariance matrix calculations can increase computational time and possibly introduce numerical errors. Hence, we fitted the geostatistical model using the stochastic partial differential equations (SPDE)/INLA (Lindgren et al. 2011;Rue et al. 2009) approach, readily implemented in the INLA R-package (available at: http://www.r-inla.org). Briefly, the spatial process assuming a Matérn covariance matrix Σ can be represented as a Gaussian Markov random field (GMRF) with mean zero and a symmetric positive definite precision matrix Q (defined as the inverse Σ)

(Cameletti et al. 2013). The SPDE approach constructs a GMRF representation of the Matérn field on a triangulation (a set of non-intersecting triangles where any two triangles meet in at most a common edge or corner) partitioning the domain of the study region (Lindgren et al. 2011). Subsequently, the INLA algorithm is used to estimate the posterior marginal (or joint) distribution of the latent Gaussian process and hyperparameters by Laplace approximation (Rue et al. 2009).

Bayesian variable selection, using normal mixture of inverse Gammas with parameter expansion (peNMIG) spike-and-slab priors (Scheipl, Fahrmeir, & Kneib 2012) was applied on the model with independent random effect for each location to identify the best set of predictors (i.e., climatic, environmental, and socioeconomic). In particular, we assumed a normal distribution for the regression coefficients with a hyperparameter for the variance σ_B^2 to be a mixture of inverse Gamma distributions, that is $\beta_k \sim N(0, \sigma_B^2)$, where $\sigma_B^2 \sim I_k IG(a_\sigma, b_\sigma) + (1 - I_k)\upsilon_0 IG(a_\sigma, b_\sigma)$ and a_σ, b_σ are fixed parameters. υ_0 is some small positive constant (Chammartin et al. 2013b) and the indicator I_k has a Bernoulli prior distribution $I_k \sim bern(\pi_k)$, where $\pi_k \sim beta(a_\pi, b_\pi)$. We set $(a_\sigma, b_\sigma) = (5, 25)$, $(a_\pi, b_\pi) = (1, 1)$ and $\upsilon_0 = 0.00025$. The above prior of mixed inverse Gamma distributions is called a mixed spike and slab prior for β_k, as one component of the mixture $\upsilon_0 IG(a_\sigma, b_\sigma)$ (when $I_k = 0$) is a narrow spike around zero that strongly shrinks β_k to zero, while the other component $IG(a_\sigma, b_\sigma)$ (when $I_k = 1$) is a wide slab that moves β_k away from zero. The posterior distribution of I_k determines which component of the mixture is predominant contributing to the inclusion or exclusion of β_k. For categorical variables, we applied a peNMIG prior developed by Scheipl *et al*. (Scheipl et al. 2012), which allows to include or exclude blocks of coefficients by improving "shrinkage" properties. Let β_{kh} be the regression coefficient for the hth category of the kth predictor, then $\beta_{kh} = \alpha_k \xi_{hk}$, where α_k is assigned a NMIG prior described above and $\xi_{hk} \sim N(m_{hk}, 1)$. Here $m_{hk} = o_{hk} - (1 - o_{hk})$ and $o_{hk} \sim bern(0.5)$, allow to shrink $|\xi_{hk}|$ towards 1. Hence, α_k models the overall contribution of the kth predictor and ξ_{hk} estimates the effects of each element β_{kh} of the predictor (Chammartin et al. 2013b). In addition, we introduced another indicator I_d for selection of either a categorical or a linear form of a continuous variable. Let β_{kd1} and β_{kd2} indicate coefficients of the categorical and linear form of kth predictor, respectively, then $\beta_k = I_d \beta_{kd1} + (1 - I_d)\beta_{kd2}$, where $I_d \sim Be(0.5)$. MCMC simulation was employed to estimate the model parameters for variable selection in OpenBUGS version 3.0.2 (Imperial College and Medical Research Council; London, UK) (Lunn et al. 2009). Convergence was assessed by the Gelman and Rubin diagnostic (Gelman & Rubin 1992), using the coda library in R (Plummer et al. 2006). In Bayesian variable

selection, all models arising from any combination of covariates are fitted and the posterior probability for each model to be the true one is calculated. The predictors corresponding to the highest joint posterior probability of indicators $(I_1, I_2, ... I_k, ..., I_K)$ were subsequently used as the best set of predictors to fit the final geostatistical model.

A 5 × 5 km grid was overlaid to the P.R. China map, resulting in 363,377 pixels. Predictions for each soil-transmitted helminth species were obtained via INLA at the centroids of the grid's pixels. An overall soil-transmitted helminth prevalence was calculated assuming independence in the risk between any two species, that is, $p_S = p_A + p_T + p_h - p_A \times p_T - p_A \times p_h - p_T \times p_h + p_A \times p_T \times p_h$, where p_S, p_A, p_T, and p_h indicate the predicted prevalence of overall soil-transmitted helminth, *A. lumbricoides*, *T. trichiura*, and hookworm, respectively, for each pixel. The number of infected individuals at pixel level was estimated by multiplying the median of the corresponding posterior predictive distribution of the infection prevalence with the population density.

2.2.5 Model validation

Our model was fitted on a subset of the data, including approximately 80% of survey locations. Validation was performed on the remaining 20% by estimating the mean predictive error (ME) between the observed π_i and predicted prevalence $\hat{\pi}_i$ at location i, where $ME = 1/N * \sum_{i=1}(\pi_i - \hat{\pi}_i)$ and N is the total number of test locations. In addition, we calculated Bayesian credible intervals (BCI) of various probability and the percentage of observations included in these intervals.

2.3 Results

2.3.1 Data summaries

The final dataset included 1,187 surveys for hookworm infection carried out at 1,067 unique locations; 1,157 surveys for *A. lumbricoides* infection at 1,052 unique locations; and 1,138 surveys for *T. trichiura* infection at 1,028 unique locations. The overall prevalence was 9.8%, 6.6% and 4.1% for *A. lumbricoides*, hookworm and *T. trichiura* infection, respectively. Details about the number of surveys by location type, study year, diagnostic method and infection prevalence are shown in Table 2.2. The geographical distribution of locations and observed prevalence for each soil-transmitted helminth species are shown in Figure 2.1. Maps of the spatial distribution of environmental/climatic, soil types and socioeconomic covariates used in Bayesian variable selection are provided in Additional file.
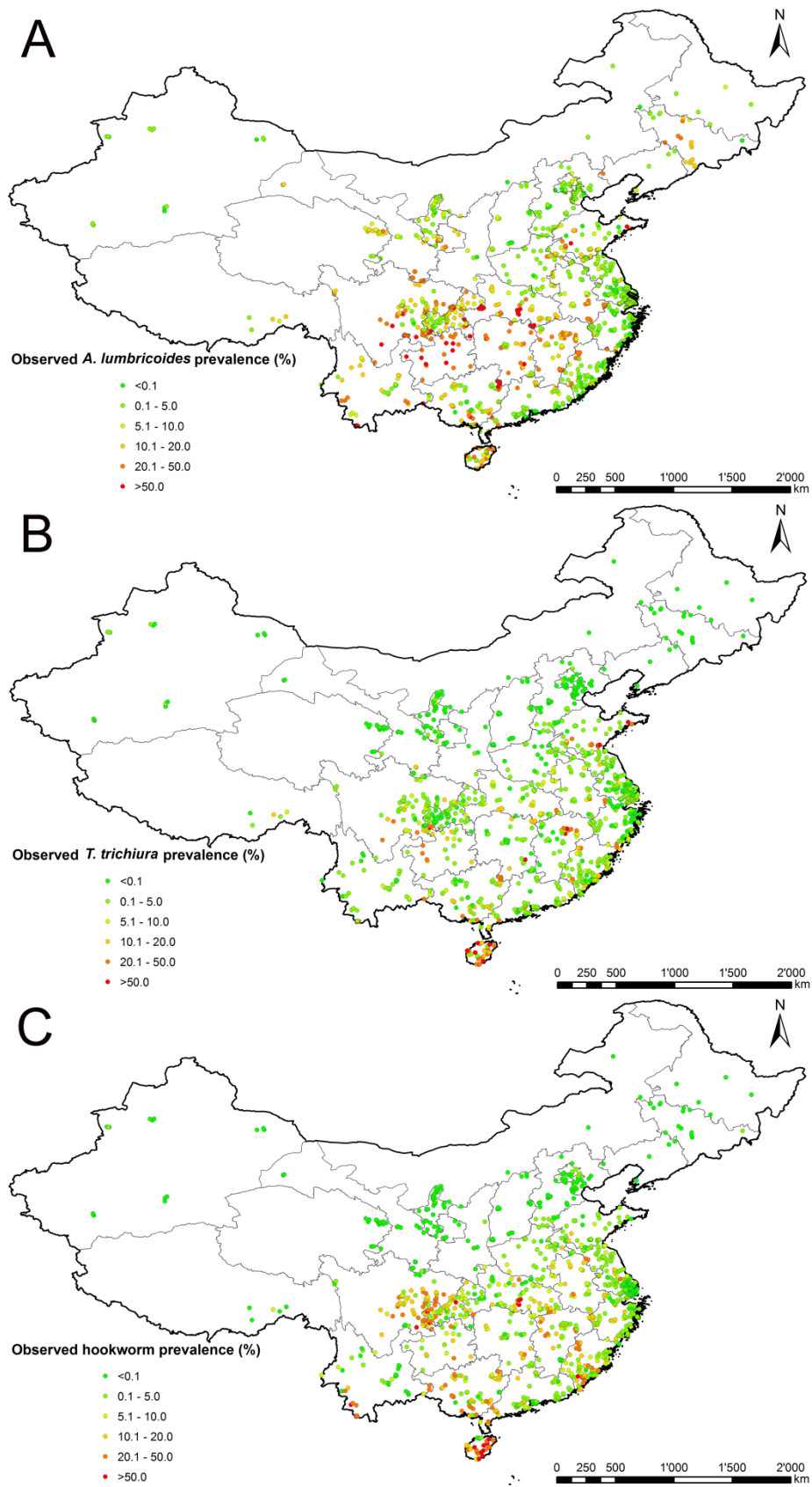
Figure 2.1: Survey locations and observed prevalence across P.R. China. The maps show the survey locations and observed prevalence for (A) *A. lumbricoides*, (B) *T. trichiura* and (C) hookworm.

Table 2.2: Overview of the number of soil-transmitted helminth surveys.

		A. lumbricoides	*T. trichiura*	**Hookworm**
		Number of surveys (percentage)		
Location types	Villages/towns	842 (72.8)	822 (72.2)	838 (70.6)
	Counties	315 (27.2)	316 (27.8)	349 (29.4)
Study year finished	2000-2004	739 (63.9)	737 (64.8)	775 (65.3)
	2005-2010	418 (36.1)	401 (35.2)	412 (34.7)
Test methods	Kato-Katz	1124 (97.2)	1112 (97.7)	1151 (97.0)
	Stool sedimentation	3 (0.26)	3 (0.26)	3 (0.25)
	Flotation method[a]	16 (1.4)	16 (1.4)	19 (1.6)
	Ether-concentration[b]	1 (0.09)	1 (0.09)	1 (0.08)
	Other diagnostic method	13 (1.1)	6 (0.53)	13 (1.1)
Observed prevalence (%)	<0.1	126 (10.9)	374 (32.9)	364 (30.7)
	0.1-5.0	513 (44.3)	523 (46.0)	396 (33.4)
	5.1-10.0	152 (13.1)	95 (8.4)	149 (12.6)
	10.1-20.0	158 (13.7)	71 (6.2)	134 (11.3)
	20.1-50.0	155 (13.4)	58 (5.1)	118 (10.0)
	>50.0	53 (4.6)	17 (1.5)	26 (2.2)
Total		1,157 (100)	1,138 (100)	1,187 (100)

[a]Stool flotation method or McMaster salt flotation method.
[b]Formalin ethyl acetate concentration method.

2.3.2 Spatial statistical modelling and variable selections

The models with the highest posterior probabilities selected the following covariates: GDP per capita, elevation, NDVI, LST at day, LST at night, precipitation, pH measured in water, and climatic zones for *T. trichiura*; GDP per capita, elevation, NDVI, LST at day, LST at night, precipitation, bulk density, gypsum content, organic carbon content, climatic zone and land cover for hookworm; and GDP per capita, elevation, NDVI, LST at day and climatic zone for *A. lumbricoides*. The corresponding posterior probabilities of the respective models were 33.2%, 23.6% and 21.4% for *T. trichiura*, hookworm and *A. lumbricoides*, respectively.

The parameter estimates that arose from the Bayesian geostatistical logistic regression fit are shown in Tables 2.3, 2.4 and 2.5. The infection risk of all three soil-transmitted helminth species decreased considerably from 2005 onwards. We found significant positive association between NDVI and the prevalence of *A. lumbricoides*. A negative association was found between GDP per capita, arid or snow/polar climatic zones and the prevalence of *A. lumbricoides*. High precipitation and LST at night are favourable conditions for the presence of hookworm, while high NDVI, LST at day, urban or wet land covers and arid or snow/polar climatic zones are less favourable. Elevation, LST at night, NDVI larger than 0.45 and equatorial climatic zone were associated with a higher odds of *T. trichiura* infection, while LST at day, arid or snow climatic zones were associated with a lower odds of *T. trichiura* infection.

2.3.3 Model validation results

Model validation indicated that the Bayesian geostatistical logistic regression models were able to correctly estimate within a 95% BCI 84.2%, 81.5% and 79.3% for *T. trichiura*, hookworm and *A. lumbricoides*, respectively. A plot of coverage for the full range of credible intervals is presented in Additional file. The MEs for hookworm, *A. lumbricoides* and *T. trichiura* were 0.56%, 1.7%, and 2.0% respectively, suggesting that our model may slightly under-estimate the risk of each of the soil-transmitted helminth species.

Table 2.3: Posterior summaries (median and 95% BCI) of the geostatistical model parameters for *A. lumbricoides*.

	Estimate[†]
Year	0.34 (0.32; 0.36)[*]
GDP per capita (yuan)	
≤12,000	1.00
12,000-24,000	0.89 (0.68; 1.18)
>24,000	0.59 (0.41; 0.86)[*]
Elevation (m)	
≤55	1.00
55-400	1.21 (0.89; 1.63)
>400	1.54 (0.96; 2.46)
NDVI	
≤0.45	1.00
0.45-0.55	2.41 (2.05; 2.84)[*]
>0.55	1.30 (1.03; 1.64)[*]
LST at day (°C)	
≤21	1.00
21-23	1.07 (0.83; 1.37)
>23	1.08 (0.76; 1.54)
Climatic zones	
Warm	1.00
Equatorial	1.73 (0.42; 7.08)
Arid	0.41 (0.18; 0.98)[*]
Snow	0.41 (0.20; 0.84)[*]
Range (km)	243.1 (182.8; 321.4)
Spatial variance (σ^2_{sp})	2.64 (1.97; 3.51)
Non-spatial variance (σ^2_{nonsp})	0.91 (0.77; 1.08)

[†]Regression coefficients are provided as odds ratios.
[*]Significant correlation based on 95% Bayesian credible interval (BCI).

2.3.4 Predictive risk maps of soil-transmitted helminth infections

Figures 2.2, 2.3 and 2.4 present species-specific predictive risk maps of soil-transmitted helminth infections for the period 2005 onwards. High prevalence of *A. lumbricoides* (>20%) was predicted in large areas of Guizhou province and the southern part of Sichuan and Hubei provinces. Moderate to high prevalence (5-20%) were predicted for large areas of Hunan,

Yunnan, Jiangxi, some southern areas of Gansu and Anhui provinces and Chongqing city. For the northern part of P.R. China and the south-eastern coastal-line areas, low prevalences were predicted (<5%). The high prediction uncertainty shown in Figure 2.2B is correlated with high prevalence areas. High infection prevalence (>20%) with *T. trichiura* was predicted for a few small areas of the southern part of P.R. China. Moderate-to-high prevalence (5-20%) was predicted for large areas of Hainan province. High hookworm infection prevalence (>20%) was predicted for Hainan, eastern parts of Sichuan and southern parts of Yunnan provinces. Low prevalence (0.1-5%) of *T. trichiura* and hookworm infections were predicted for most areas of the southern part of P.R. China, while close to zero prevalence areas were predicted for the northern part.

Table 2.4: Posterior summaries (median and 95% BCI) of the geostatistical model parameters for *T. trichiura*.

	Estimate[†]
Year	0.26 (0.24; 0.28)[*]
GDP per capita	1.02 (0.83; 1.25)
Elevation	1.80 (1.37; 2.37)[*]
NDVI	
≤0.45	1.00
0.45-0.55	2.64 (1.99; 3.52)[*]
>0.55	1.59 (1.10; 2.32)[*]
LST at day	0.62 (0.48; 0.81)[*]
LST at night	3.61 (2.08; 6.32)[*]
Precipitation	1.23 (0.79; 1.91)
pH measured in water	
≤5.95	1.00
5.95-7.00	1.39 (1.00; 1.95)
>7.00	1.49 (0.96; 2.30)
Climatic zones	
Warm	1.00
Equatorial	6.40 (1.25; 31.49)[*]
Arid	0.10 (0.03; 0.36)[*]
Snow	0.07 (0.02; 0.22)[*]
Range (km)	138.8 (104.3; 179.1)
Spatial variance (σ^2_{sp})	4.19 (3.22; 5.08)
Non-spatial variance (σ^2_{nonsp})	1.09 (0.88; 1.37)

[†]Regression coefficients are provided as odds ratios.
[*]Significant correlation based on 95% Bayesian credible interval (BCI).

2.3.5 Estimates of number of people infected

Figure 2.5 shows the combined soil-transmitted helminth prevalence and the number of infected individuals from 2005 onwards. Table 2.6 summarises the population-adjusted predicted prevalence and the number of infected individuals, stratified by province. The overall population-adjusted predicted prevalence of *A. lumbricoides*, hookworm and *T. trichiura* infections were, respectively, 6.8%, 3.7% and 1.8%, corresponding to 85.4, 46.6

and 22.1 million infected individuals. The overall population-adjusted predicted prevalence for combined soil-transmitted helminth infections was 11.4%.

Table 2.5: Posterior summaries (median and 95% BCI) of the geostatistical model parameters for hookworm.

	Estimate[†]
Year	0.27 (0.25; 0.29)[*]
GDP per capita (yuan)	
≤12,000	1.00
12,000-24,000	1.28 (0.89; 1.85)
>24,000	0.89 (0.53; 1.50)
Elevation (m)	
≤55	1.00
55-400	1.34 (0.91; 1.98)
>400	1.32 (0.73; 2.38)
NDVI	
≤0.45	1.00
0.45-0.55	0.44 (0.36; 0.52)[*]
>0.55	0.36 (0.27; 0.47)[*]
LST at day	0.32 (0.23; 0.45)[*]
LST at night	7.35 (3.88; 14.12)[*]
Precipitation	3.17 (1.89; 5.48)[*]
Bulk density (km/dm^3)	
≤1.29	1.00
1.29-1.36	0.82 (0.52; 1.30)
>1.36	0.66 (0.37; 1.17)
Gypsum content (g/kg)	
≤0	1.00
0-1	1.20 (0.88; 1.63)
>1	1.17 (0.73; 1.87)
Organic carbon content (g/kg)	
≤11	1.00
11-12.5	0.73 (0.44; 1.20)
>12.5	0.81 (0.46; 1.43)
Climatic zones	
Warm	1.00
Equatorial	1.87 (0.34; 10.13)
Arid	0.17 (0.03; 0.83)[*]
Snow	0.05 (0.01; 0.21)[*]
Land cover	
Croplands	1.00
Forests	0.83 (0.57; 1.22)
Shrublands and savannas	1.07 (0.67; 1.70)
Grasslands	0.63 (0.13; 2.58)
Urban	0.35 (0.22; 0.58)[*]
Wet areas	0.15 (0.07; 0.32)[*]
Range (km)	186.1 (126.8; 296.5)
Spatial variance (σ^2_{sp})	5.07 (3.72; 6.63)
Non-spatial variance (σ^2_{nonsp})	0.88 (0.69; 1.22)

[†]Regression coefficients are provided as odds ratios.
[*]Significant correlation based on 95% Bayesian credible interval (BCI).

Table 2.6: Population-adjusted predicted prevalence (%) and number of individuals ($\times 10^6$) infected with soil-transmitted helminths by province[†].

Province	Population	A. lumbricoides		T. trichiura	
		Prevalence	No. of people infected	Prevalence	No. of people infected
Anhui	54.9	4.3 (2.9; 6.8)	2.4 (1.6; 3.8)	1.4 (0.77; 2.6)	0.78 (0.42; 1.4)
Beijing	17.0	0.68 (0.40; 1.3)	0.12 (0.07; 0.22)	0.05 (0.02; 0.23)	0.01 (0.00; 0.04)
Chongqing	26.7	10.2 (7.9; 12.4)	2.7 (2.1; 3.3)	1.2 (0.78; 1.7)	0.31 (0.21; 0.47)
Fujian	32.8	1.9 (1.5; 2.7)	0.63 (0.48; 0.89)	2.0 (1.4; 3.2)	0.67 (0.44; 1.0)
Gansu	25.6	6.6 (3.8; 11.5)	1.7 (0.97; 2.9)	0.30 (0.10; 1.3)	0.08 (0.02; 0.34)
Guangdong	91.1	3.0 (2.1; 4.4)	2.7 (1.9; 4.0)	2.3 (1.4; 3.5)	2.1 (1.3; 3.2)
Guangxi	37.6	6.9 (5.1; 9.4)	2.6 (1.9; 3.5)	3.6 (2.3; 5.7)	1.4 (0.86; 2.1)
Guizhou	31.4	27.9[*] (19.5; 37.6)	8.7 (6.1; 11.8)	5.2 (2.8; 9.6)	1.6 (0.87; 3.0)
Hainan	6.7	7.5 (5.0; 10.5)	0.50 (0.33; 0.70)	18.3[*] (11.9; 25.9)	1.2 (0.80; 1.7)
Hebei	75.5	1.3 (0.77; 2.1)	0.95 (0.58; 1.6)	0.09 (0.04; 0.20)	0.07 (0.03; 0.15)
Heilongjiang	42.3	2.1 (0.99; 4.7)	0.90 (0.42; 2.0)	0.02 (0.01; 0.07)	0.01 (0.00; 0.03)
Henan	84.3	2.2 (1.5; 3.2)	1.8 (1.2; 2.7)	0.62 (0.34; 1.2)	0.52 (0.29; 1.1)
Hubei	58.2	18.3 (14.4; 22.7)	10.6 (8.4; 13.2)	3.2 (1.7; 6.7)	1.9 (0.98; 3.9)
Hunan	55.1	17.7 (12.5; 24.9)	9.7 (6.9; 13.7)	1.8 (0.9; 3.6)	0.99 (0.50; 2.0)
Jiangsu	74.3	1.2 (0.88; 1.8)	0.91 (0.65; 1.3)	0.72 (0.45; 1.5)	0.54 (0.34; 1.1)
Jiangxi	36.3	11.3 (8.1; 15.7)	4.1 (2.9; 5.7)	3.3 (2.0; 5.8)	1.2 (0.73; 2.1)
Jilin	29.2	7.2 (4.3; 11.8)	2.1 (1.2; 3.5)	0.02 (0.00; 0.09)	0.01 (0.00; 0.03)
Liaoning	43.1	3.5 (1.4; 9.1)	1.5 (0.61; 3.9)	0.02 (0.00; 0.08)	0.01 (0.00; 0.03)
Nei Mongol	29.7	2.2 (1.0; 5.2)	0.65 (0.30; 1.6)	0.01[#] (0.01; 0.06)	0.00[#] (0.00; 0.02)
Ningxia Hui	6.3	3.8 (2.5; 5.4)	0.24 (0.16; 0.34)	0.07 (0.03; 0.26)	0.00[#] (0.00; 0.02)
Qinghai	5.0	5.7 (3.4; 9.5)	0.28 (0.17; 0.47)	0.05 (0.01; 0.20)	0.00[#] (0.00; 0.01)
Shaanxi	34.2	5.8 (2.6; 12.5)	2.0 (0.89; 4.3)	0.84 (0.30; 2.3)	0.29 (0.10; 0.78)
Shandong	93.4	5.2 (3.7; 7.3)	4.9 (3.4; 6.9)	2.1 (1.4; 3.3)	2.0 (1.3; 3.1)
Shanghai	15.0	0.32[#] (0.21; 0.54)	0.05[#] (0.03; 0.08)	0.46 (0.27; 0.81)	0.07 (0.04; 0.12)
Shanxi	35.5	1.7 (0.88; 3.9)	0.59 (0.31; 1.4)	0.14 (0.04; 0.41)	0.05 (0.01; 0.15)
Sichuan	94.6	14.8 (11.5; 19.3)	14.0[*] (10.9; 18.2)	3.9 (2.4; 6.9)	3.7[*] (2.2; 6.5)
Tianjin	9.8	0.66 (0.32; 1.3)	0.06 (0.03; 0.13)	0.01[#] (0.00; 0.05)	0.00[#] (0.00; 0.00)
Xinjiang Uygur	25.0	2.4 (1.1; 7.1)	0.60 (0.26; 1.8)	0.05 (0.02; 0.15)	0.01 (0.00; 0.04)
Tibet	2.7	3.3 (1.5; 7.7)	0.09 (0.04; 0.21)	0.47 (0.15; 1.4)	0.01 (0.00; 0.04)
Yunnan	39.5	13.6 (9.0; 19.4)	5.4 (3.5; 7.7)	3.5 (2.0; 6.5)	1.4 (0.77; 2.6)
Zhejiang	45.4	0.80 (0.58; 1.2)	0.36 (0.26; 0.54)	0.64 (0.38; 1.1)	0.29 (0.17; 0.50)
Total	1,257.9	6.8 (6.2; 7.5)	85.4 (77.8; 94.0)	1.8 (1.5; 2.1)	22.1 (18.7; 26.2)

(Table 2.6 continues in next page)

For *A. lumbricoides*, the predicted prevalence ranged from 0.32% (Shanghai) to 27.9% (Guizhou province). Shanghai had the smallest (0.05 million) and Sichuan province the largest number (14.8 million) of infected individuals. For *T. trichiura*, the predicted prevalence ranged from 0.01% (Tianjin) to 18.3% (Hainan province). The smallest number of infected individuals were found in Nei Mongol, Ningxia Hui, Qinghai provinces and Tianjin (<0.01 million) whereas the largest number, 3.7 million, was predicted for Sichuan province. For hookworm, Ningxia Hui and Qinghai province had the lowest predicted prevalence (<0.01%), while Hainan province had the highest (22.1%). The provinces of Gansu, Nei Mongol, Ningxia Hui, Qinghai, Xinjiang Uygur and Tibet, and the cities of Beijing, Shanghai and Tianjin each had less than 10,000 individuals infected with hookworm. Sichuan province had the largest predicted number of hookworm infections (14.3 million).

The predicted combined soil-transmitted helminth prevalence ranged from 0.70% (Tianjin) to 40.8% (Hainan province). The number of individuals infected with soil-transmitted helminths ranged from 0.07 million (Tianjin) to 29.0 million (Sichuan province). Overall, slightly more than one out of ten people in P.R. China is infected with soil-transmitted helminths, corresponding to more than 140 million infections in the year 2010.

(continued from previous page Table 2.6)

Province	Population	Hookworm		Any soil-transmitted helminth	
		Prevalence	No. of people infected	Prevalence	No. of people infected
Anhui	54.9	4.6 (2.6; 8.1)	2.5 (1.5; 4.4)	10.1 (7.6; 14.2)	5.5 (4.2; 7.8)
Beijing	17.0	0.02 (0.01; 0.06)	0.00[#] (0.00; 0.01)	0.77 (0.47; 1.4)	0.13 (0.08; 0.24)
Chongqing	26.7	10.3 (7.6; 13.5)	2.8 (2.0; 3.6)	20.6 (17.1; 24.0)	5.5 (4.6; 6.4)
Fujian	32.8	8.6 (6.4; 11.6)	2.8 (2.1; 3.8)	12.3 (10.1; 15.3)	4.0 (3.3; 5.0)
Gansu	25.6	0.02 (0.00; 0.16)	0.00[#] (0.00; 0.04)	7.0 (4.1; 11.9)	1.8 (1.0; 3.0)
Guangdong	91.1	4.3 (2.6; 7.1)	3.9 (2.4; 6.5)	9.0 (7.1; 12.4)	8.2 (6.4; 11.3)
Guangxi	37.6	7.8 (5.4; 11.8)	2.9 (2.0; 4.4)	17.4 (14.0; 21.8)	6.5 (5.3; 8.2)
Guizhou	31.4	4.1 (2.1; 7.7)	1.3 (0.65; 2.4)	34.6 (25.9; 43.3)	10.9 (8.1; 13.6)
Hainan	6.7	22.1[*] (16.0; 29.7)	1.5 (1.1; 2.0)	40.8[*] (33.6; 48.6)	2.7 (2.2; 3.3)
Hebei	75.5	0.12 (0.06; 0.31)	0.09 (0.04; 0.23)	1.5 (0.97; 2.4)	1.1 (0.73; 1.8)
Heilongjiang	42.3	0.04 (0.01; 0.26)	0.02 (0.00; 0.11)	2.2 (1.1; 4.8)	0.93 (0.44; 2.0)
Henan	84.3	1.5 (0.83; 2.7)	1.2 (0.70; 2.3)	4.3 (3.2; 5.9)	3.6 (2.7; 4.9)
Hubei	58.2	5.9 (4.0; 8.9)	3.4 (2.3; 5.2)	24.9 (20.7; 30.1)	14.5 (12.0; 17.5)
Hunan	55.1	3.5 (2.0; 6.6)	1.9 (1.1; 3.7)	22.1 (16.6; 29.5)	12.2 (9.2; 16.3)
Jiangsu	74.3	2.2 (1.5; 3.5)	1.6 (1.1; 2.6)	4.1 (3.2; 5.7)	3.1 (2.4; 4.3)
Jiangxi	36.3	5.0 (3.1; 7.7)	1.8 (1.1; 2.8)	18.6 (14.8; 23.6)	6.8 (5.4; 8.6)
Jilin	29.2	0.04 (0.01; 0.28)	0.01 (0.00; 0.08)	7.3 (4.3; 11.9)	2.1 (1.3; 3.5)
Liaoning	43.1	0.03 (0.00; 0.21)	0.01 (0.00; 0.09)	3.5 (1.5; 9.2)	1.5 (0.64; 5.0)
Nei Mongol	29.7	0.01 (0.00; 0.04)	0.00[#] (0.00; 0.01)	2.2 (1.0; 5.2)	0.66 (0.31; 1.6)
Ningxia Hui	6.3	0.00[#] (0.00; 0.03)	0.00[#] (0.00; 0.00)	3.9 (2.6; 5.5)	0.25 (0.16; 0.34)
Qinghai	5.0	0.00[#] (0.00; 0.03)	0.00[#] (0.00; 0.00)	5.8 (3.5; 9.6)	0.29 (0.17; 0.48)
Shaanxi	34.2	0.39 (0.09; 1.9)	0.14 (0.03; 0.66)	7.0 (3.7; 13.4)	2.4 (1.3; 4.6)
Shandong	93.4	0.78 (0.44; 1.6)	0.73 (0.41; 1.5)	7.9 (6.2; 10.4)	7.4 (5.8; 9.7)
Shanghai	15.0	0.02 (0.01; 0.05)	0.00[#] (0.00; 0.01)	0.82 (0.58; 1.2)	0.12 (0.09; 0.18)
Shanxi	35.5	0.07 (0.02; 0.27)	0.02 (0.01; 0.10)	1.9 (1.1; 4.1)	0.68 (0.37; 1.4)
Sichuan	94.6	15.1 (10.9; 21.4)	14.3[*] (10.3; 20.3)	30.6 (26.0; 36.2)	29.0[*] (24.6; 34.2)
Tianjin	9.8	0.03 (0.01; 0.13)	0.00[#] (0.00; 0.01)	0.70[#] (0.36; 1.3)	0.07[#] (0.04; 0.13)
Xinjiang Uygur	25.0	0.01 (0.00; 0.09)	0.00[#] (0.00; 0.02)	2.5 (1.1; 7.2)	0.62 (0.28; 1.8)
Tibet	2.7	0.03 (0.01; 0.17)	0.00[#] (0.00; 0.00)	3.9 (1.9; 8.1)	0.10 (0.05; 0.22)
Yunnan	39.5	2.7 (1.6; 5.0)	1.1 (0.62; 2.0)	19.0 (13.89; 25.1)	7.5 (5.5; 9.9)
Zhejiang	45.4	3.0 (1.9; 4.9)	1.4 (0.88; 2.2)	4.4 (3.3; 6.4)	2.0 (1.5; 2.9)
Total	1,257.9	3.7 (3.2; 4.3)	46.6 (40.7; 53.7)	11.4 (10.8; 12.2)	143.8 (135.9; 153.8)

[†]Estimates based on Gridded population of 2010; calculations based on the median and 95% Bayesian credible interval (BIC) of the posterior distribution of the predicted risk from 2005 onwards; [*]highest prevalence/ largest number of infected individuals among provinces; [#]lowest prevalence/ smallest number of infected individuals among provinces.

2.4 Discussion

To our knowledge, we present the first model-based, nation-wide predictive infection risk maps of soil-transmitted helminths for P.R. China. Previous epidemiological studies (Coordinating Office of the National Survey on the Important Human Parasitic Diseases 2005) were mainly descriptive, reporting prevalence estimates at specific locations or visualized at province level using interpolated risk surface maps. We carried out an extensive literature search and collected published georeferenced soil-transmitted helminth prevalence data across P.R. China, alongside the ones from the second national survey that had been completed in 2004. Bayesian geostatistical models were utilised to identify climatic/environmental and socioeconomic factors that were significantly associated with infection risk, and hence, the number of infected individuals could be calculated at high spatial resolution. We derived species-specific risk maps. Additionally, we produced a risk map with any soil-transmitted
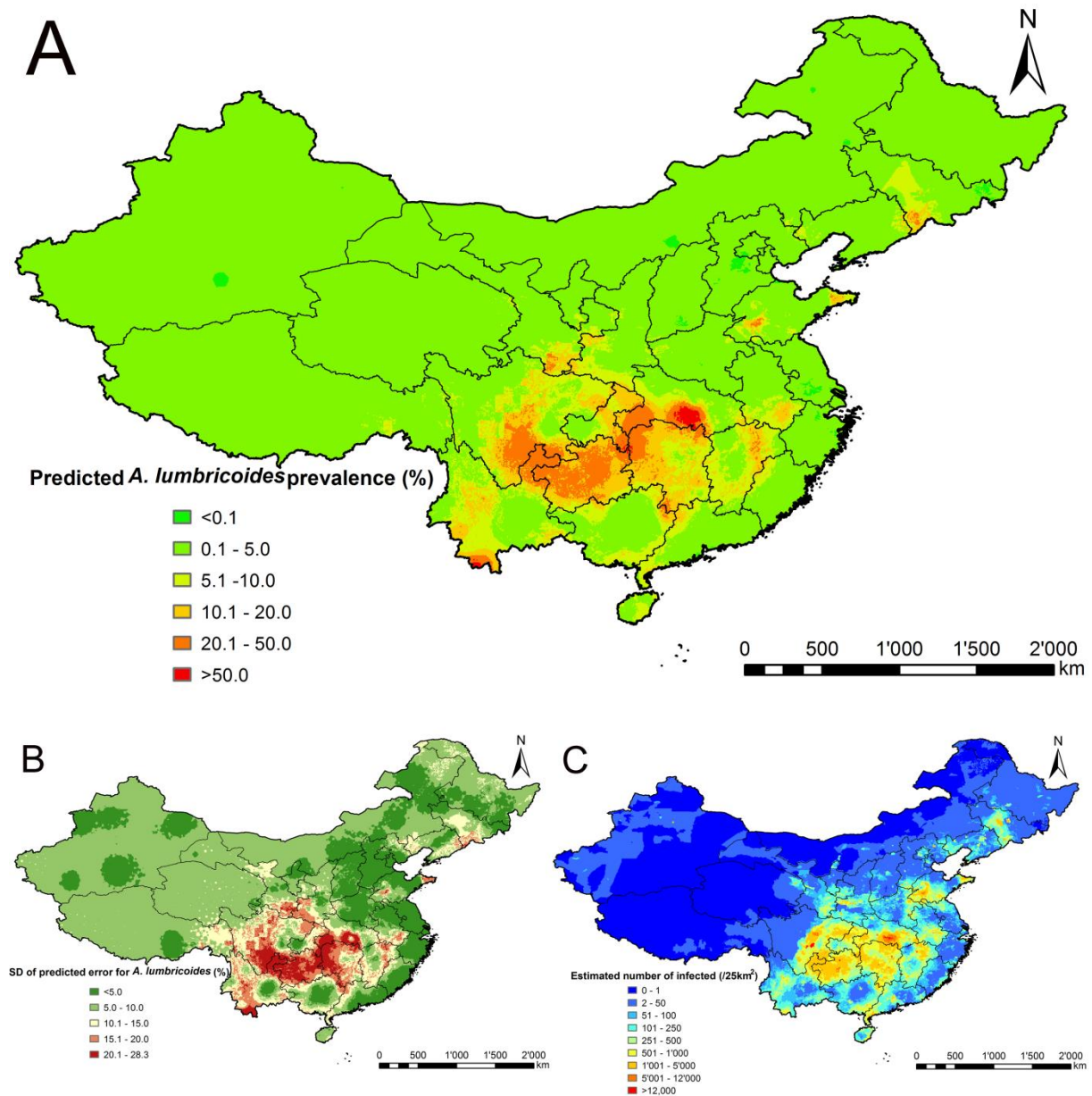
Figure 2.2: The geographical distribution of *A. lumbricoides* infection risk in P.R. China. The maps show the situation from 2005 onwards based on the median and standard deviation of the posterior predictive distribution. Estimates of (A) infection prevalence, (B) prediction uncertainty and (C) number of infected individuals.

helminth infection, which is particularly important for the control of soil-transmitted helminthiasis, as the same drugs (mainly albendazole and mebendazole) are used against all three species (Keiser & Utzinger 2008;WHO 2002a).

Model validation suggested good predictive ability of our final models. In particular, 84.2%, 81.5% and 79.3% of survey locations were correctly predicted within a 95% BCI for *T. trichiura*, hookworm and *A. lumbricoides*, respectively. The combined soil-transmitted helminth prevalence (11.4%) is supported by the current surveillance data reported to China
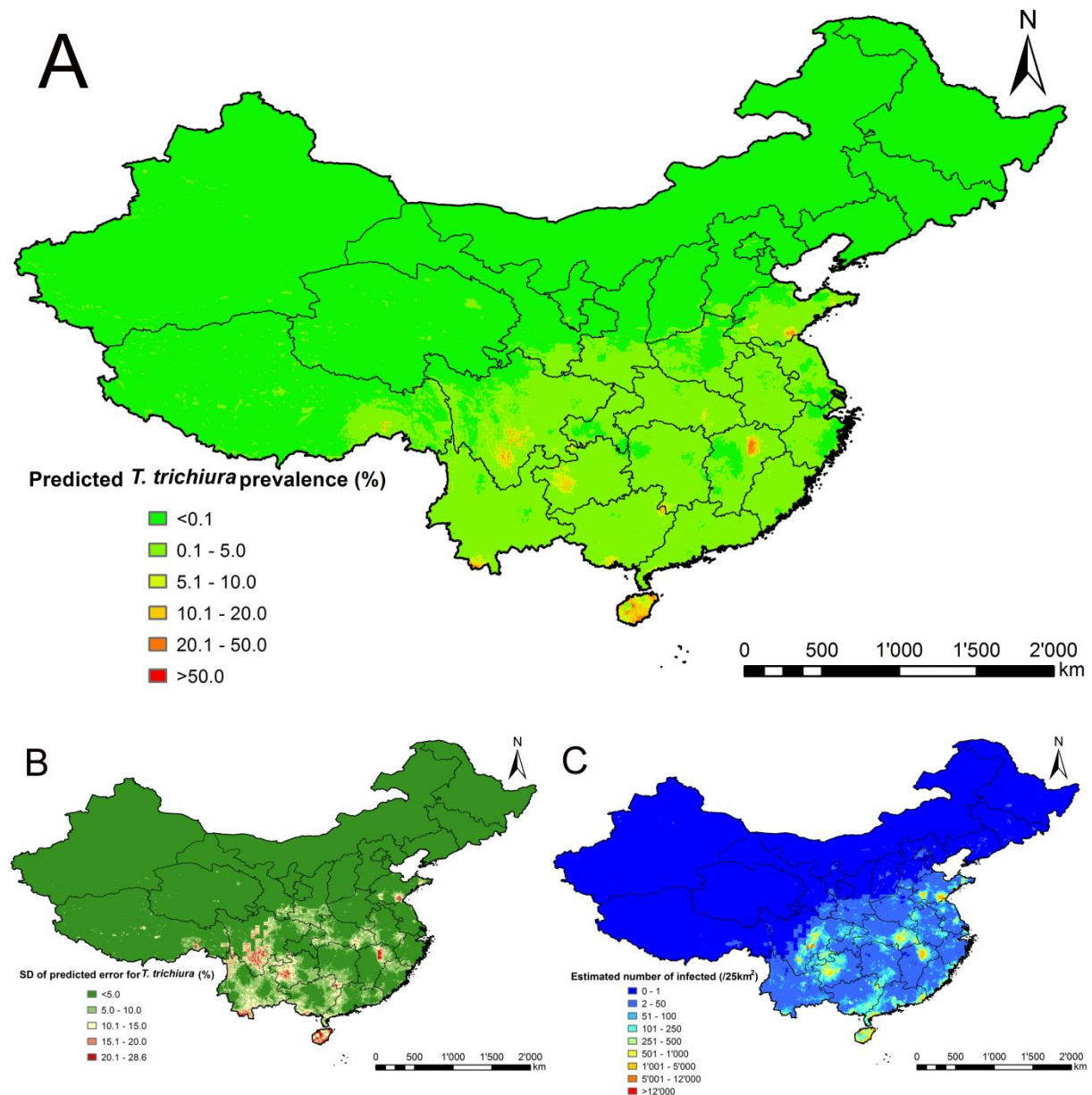
Figure 2.3: The geographical distribution of *T. trichiura* infection risk in P.R. China. The maps show the situation from 2005 onwards based on the median and standard deviation of the posterior predictive distribution. Estimates of (A) infection prevalence, (B) prediction uncertainty and (C) number of infected individuals.

CDC that shows infection rates in many areas of P.R. China around 10%. We found that all ME were above zero, hence the predictive prevalence slightly under-estimated the true prevalence of each of the three soil-transmitted helminth species. The combined soil-transmitted helminth prevalence estimates assume that the infection of each species is independent of each other. However, previous research reported significant associations, particularly between *A. lumbricoides* and *T. trichiura* (Booth & Bundy 1992;Tchuem Tchuenté et al. 2003a). Hence, our assumption may over-estimate the true prevalence of soil-
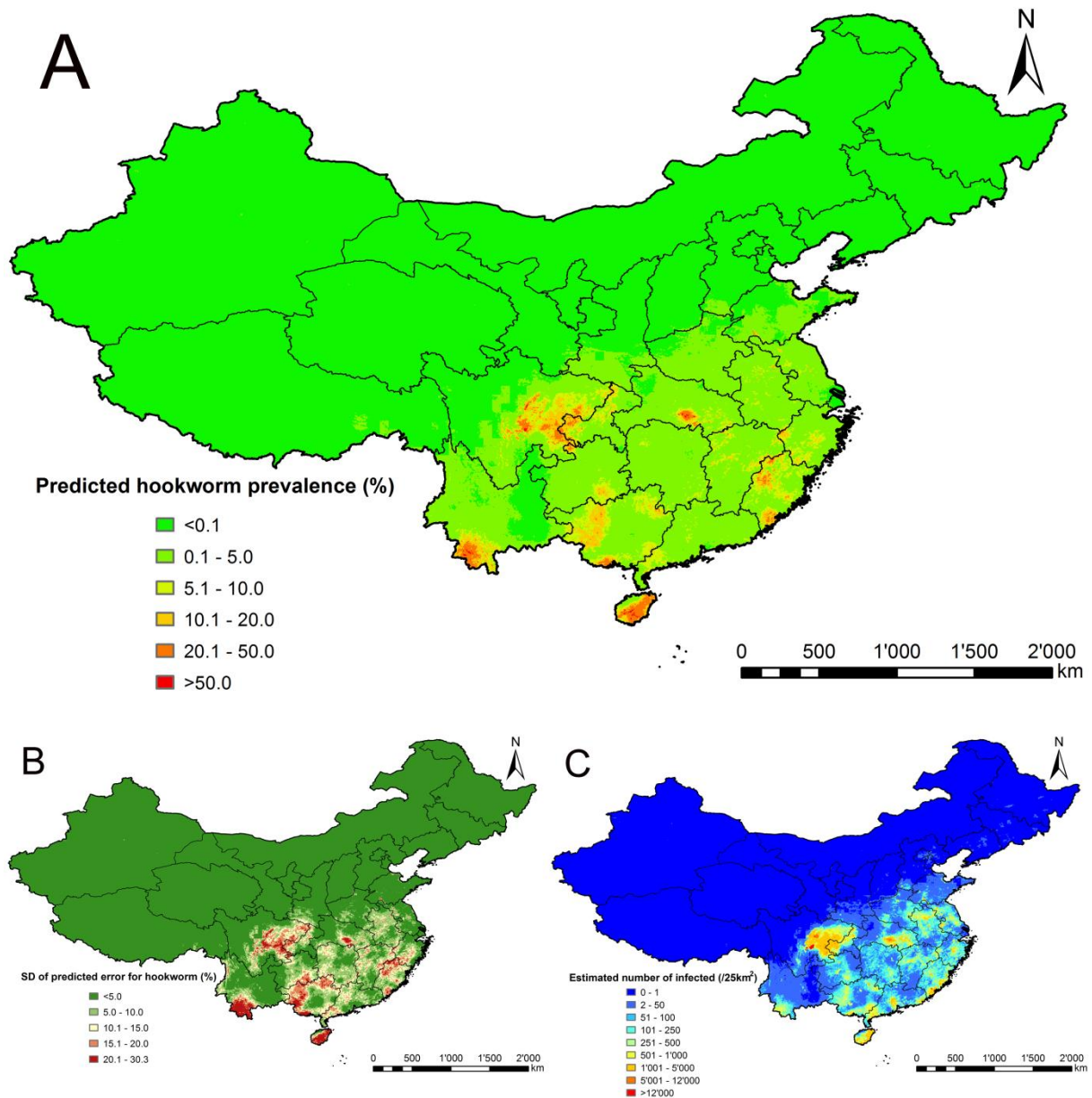
Figure 2.4: The geographical distribution of hookworm infection risk in P.R. China. The maps show the situation from 2005 onwards based on the median and standard deviation of the posterior predictive distribution. Estimates of (A) infection prevalence, (B) prediction uncertainty and (C) number of infected individuals.

transmitted helminths. Unfortunately we do not have co-infection data from P.R. China, and thus we are unable to calculate a correction factor.

Our results indicate that several environmental and climatic predictors are significantly associated with soil-transmitted helminth infections. For example, LST at night was significantly associated with *T. trichiura* and hookworm, suggesting that temperature is an important driver of transmission. Similar results have been reported by other researchers
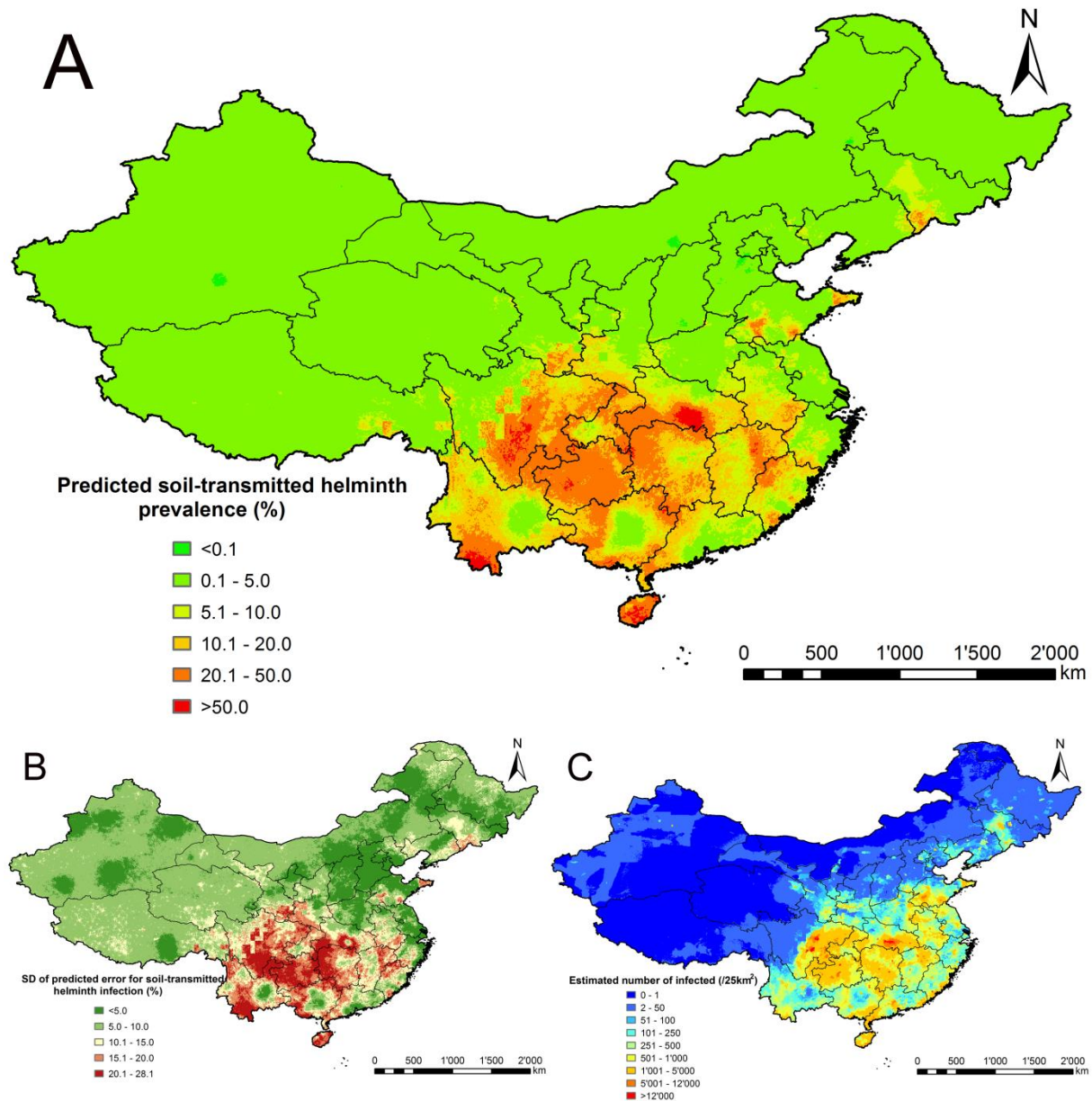
Figure 2.5: The geographical distribution of soil-transmitted helminth infection risk in P.R. China. The maps show the situation from 2005 onwards based on the median and standard deviation of the posterior predictive distribution. Estimates of (A) infection prevalence, (B) prediction uncertainty and (C) number of infected individuals.

(Pullan & Brooker 2012;Tchuem Tchuenté 2011). Our results suggest that the risk of infection with any of the soil-tansmitted helminth species is higher in equatorial or warm zones, compared to the arid and snow/polar zones. This is consistent with earlier findings that extremely arid environments limit the transmission of soil-transmitted helminths (Pullan & Brooker 2012), while equatorial or warm zones provide temperatures and soil moisture that are particularly suitable for larval development (Tchuem Tchuenté 2011). However, we found a positive association between elevation and *T. trichiura* infection risk, which contradicts earlier reports (Flores et al. 2001;Gunawardena et al. 2011). The reason may be the altitude

effect, i.e., the negative correlation between altitude and economy in P.R. China (Zhai S & Sun A 2012). The low socioeconomic development in high altitude or mountainous areas might result in limited access to healthcare services (Schratz et al. 2010;Yap et al. 2013).

On the other hand, it is reported that socioeconomic factors are closely related with the behaviour of people, which in turn impacts the transmission of soil-transmitted helminths (Brooker et al. 2006). Indeed, wealth, inadequate sewage discharge, drinking of unsafe water, lack of sanitary infrastructure, personal hygiene habits, recent travel history, low education and demographic factors are strongly associated with soil-transmitted helminth infections (Escobedo et al. 2008;Hohmann et al. 2001;Knopp et al. 2010;Norhayati et al. 1998;Pinheiro et al. 2011). Our results show that GDP per capita has a negative effect on *A. lumbricoides* infection risk. Other socioeconomic proxies such as sanitation level, number of hospital beds and percentage of people with access to tap water might be more readily able to explain the spatial distribution of infection risk.

Model-based estimates adjusted for population density indicate that the highest prevalence of *A. lumbricoides* occurred in Guizhou province. *T. trichiura* and hookworm were most prevalent in Hainan province. Although the overall soil-transmitted helminth infection risk decreased over the past several years, Hainan province had the highest risk in 2010, followed by Guizhou and Sichuan provinces. These results are consistent with the reported data of the second national survey on important parasitic diseases (Coordinating Office of the National Survey on the Important Human Parasitic Diseases 2005), and hence more effective control strategies are needed in these provinces.

The targets set out by the Chinese Ministry of Health in the "National Control Program on Important Parasitic Diseases from 2006 to 2015" are to reduce the prevalence of soil-transmitted helminth infections by 40% until 2010 and up to 70% until 2015 (Zheng et al. 2009). The government aims to reach these targets by a series of control strategies, including anthelminthic treatment, improvement of sanitation, and better information, education and communication (IEC) campaigns (Bergquist & Whittaker 2012). Preventive chemotherapy is recommended for populations older than 3 years in areas where the prevalence of soil-transmitted helminth infection exceeds 50%, while targeted drug treatment is recommended for children and rural population in areas where infection prevalences range between 10% and 50% (Ministry of Health 2006). Our models indicate that the first step of the target, i.e., reduction of prevalence by 40% until 2010, has been achieved. Indeed, the prevalence of *T. trichiura*, hookworm and *A. lumbricoides* dropped from 4.6%, 6.1% and 12.7% in the second national survey between 2001 and 2004 (Coordinating Office of the National Survey on the Important Human Parasitic Diseases 2005) to 1.8%, 3.7% and 6.8% in 2010, which corresponds to respective reductions of 60.9%, 39.3% and 46.5%. The combined soil-transmitted helminth prevalence dropped from 19.6% to 11.4% in 2010, a reduction of 41.8%. These results also suggest that, compared to *T. trichiura* and *A. lumbricoides*, more effective strategies need to be tailored for hookworm infections.

The data of our study stem largely from community-based surveys. However, the information extracted from the literature is not disaggregated by age, and hence we were not able to obtain age-adjusted predictive risk maps. In addition, more than 96% of observed surveys used the Kato-Katz technique (Katz, Chaves, & Pellegrino 1972;Speich et al. 2010). We assumed that the diagnostic sensitivity was similar across survey locations. However, the sensitivity depends on the intensity of infection, and hence varies in space (Booth et al. 2003). The above data limitations are known in geostatistical meta-analyses of historical data (Chammartin et al. 2013b) and we are currently developing methods to address them.

Acknowledgements

2.5 Additional file

2.5.1 Spatial distribution of environmental/climatic, soil types and socioeconomic factors across P.R. China
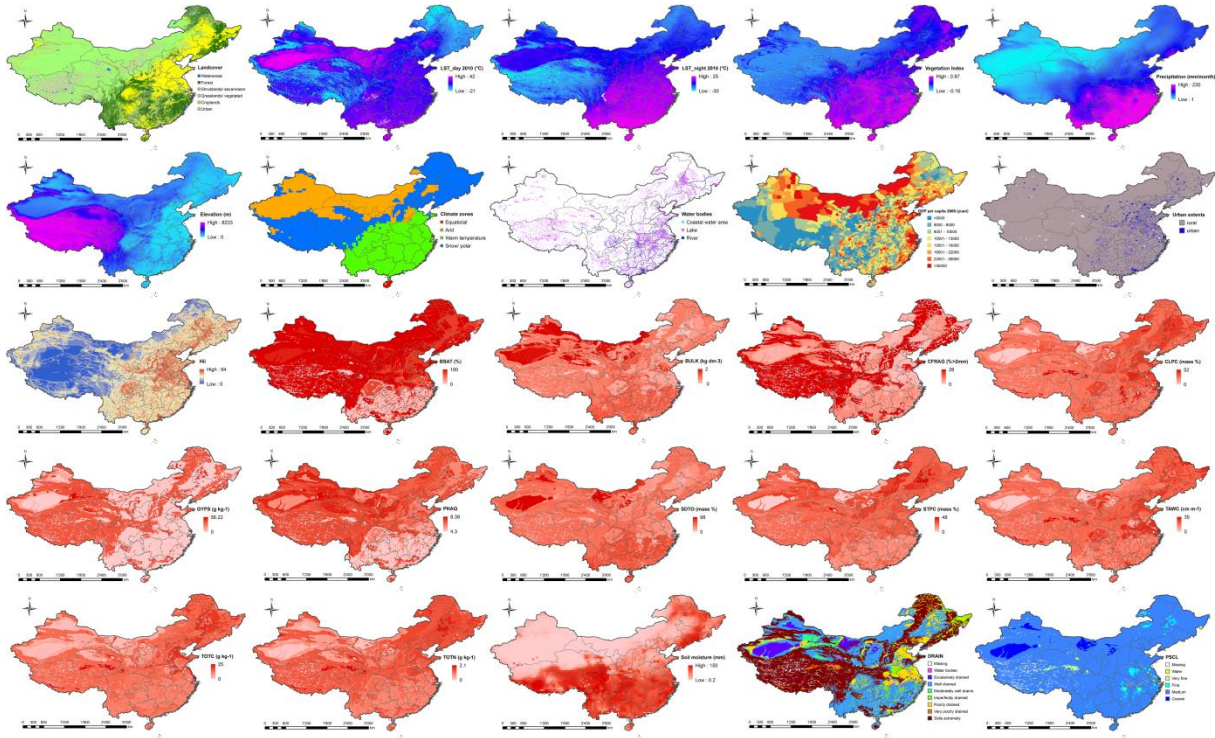
2.5.2 Model validation results
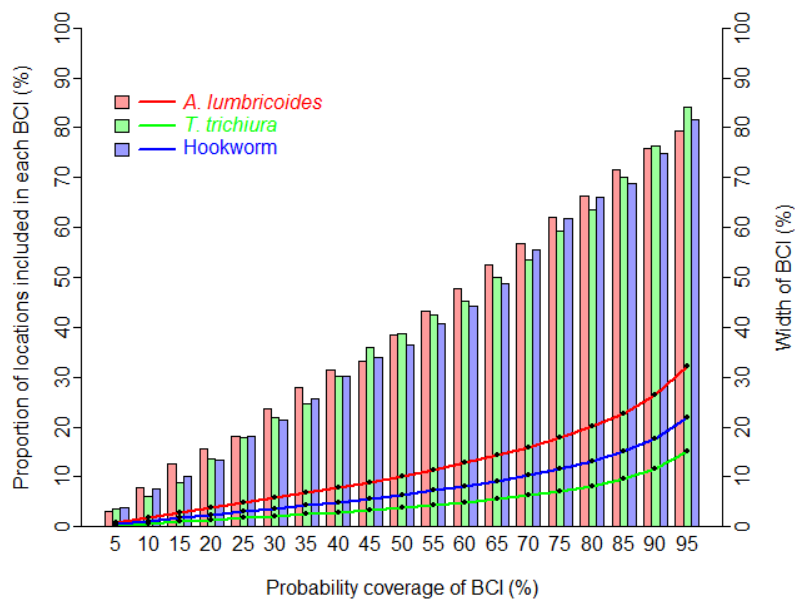
Percentage of survey locations with observed prevalence included within the Bayesian credible interval (BCI) of various probability coverage cut-offs (bar plots) calculated from the posterior predicted distribution. Solid lines indicate the corresponding width of BCI.

Chapter 3   Risk profiling of soil-transmitted helminth infection and the estimated number of infected people in South Asia

Ying-Si Lai[1,2], Patricia Biedermann[1,2], Akina Shrestha[1,2], Frédérique Chammartin[1,2], Natacha à Porta[3], Antonio Montresor[3], Nerges F. Mistry[4], Jürg Utzinger[1,2], Penelope Vounatsou[1,2]

[1]Swiss Tropical and Public Health Institute, Basel, Switzerland

[2]University of Basel, Basel, Switzerland

[3]Department of Control of Neglected Tropical Diseases, World Health Organization, Geneva, Switzerland

[4] Foundation for Medical Research, Mumbai, Maharashtra, India

Abstract

Background: In South Asia, millions of people are infected with soil-transmitted helminthes (*Ascaris lumbricoides*, *Trichuris trichiura* and hookworm). However, high-resolution risk profiles and the estimated number of people infected have yet to be determined. In turn, such information will assist control programs to identify priority areas for resources allocation.

Methodologies: We pursued a systematic review to identify prevalence surveys to soil-transmitted helminth infections in South Asia. PubMed and ISI Web of Science were searched from inception to 15 January 2016, without restriction of language, study design, and survey date. We utilized Bayesian geostatistical models to identify environmental and socioeconomic predictors, and to estimate infection risk at high spatial resolution across South Asia.

Principal Findings: A total of 501, 465, and 384 georeferenced surveys were identified for *A. lumbricoides*, hookworm and *T. trichiura*, respectively. We estimate that 397 million people (95% Bayesian credible interval (BCI) 365 to 436 million), approximatedly one-quarter of the South Asia population, was infected with at least one species of soil-transmitted helminths in 2015. *A. lumbricoides*, was the most predominant species. Moderate to high prevalence (>20%) of any soil-transmitted helminth infection was predicted in the northeastern part and some northern areas of the study region as well as the southern coastal-line areas of India. The annual treatment needs for the school-aged population requiring preventive chemotherapy was estimated at 187 million doses (95% BCI 167-211 million).

Conclusions/Significance: Our risk maps provide an overview of the geographic distribution of soil-transmitted helminth infections in South Asia and highlight the need for up-to-date surveys to accurately evaluate the disease burden in the region.

3.1 Introduction

Soil-transmitted helminths (i.e., *Ascaris lumbricoides*, *Trichuris trichiura*, and hookworm) are widespread particularly in resource-constrained settings and marginalized populations (Bethony et al. 2006). Indeed, soil-transmitted helminth infections are among the most prevalent of the neglected tropical diseases (NTDs), and they rank among the top three according to global prevalence and population at risk of all NTDs (Hotez et al. 2007). In 2010, it was estimated that 819 million people were infected with *A. lumbricoides*, 465 million with *T. trichuris*, and 439 million with hookworm (Pullan et al. 2014), accounting for a global burden of 5.2 million disability-adjusted life years (DALYs), respectively (Murray et al. 2012). The regions with the highest prevalence of soil-transmitted helminth infections are East Asia and Pacific Islands (including China), sub-Saharan Africa, South Asia (including India), and Latin America and Caribbean (Bethony et al. 2006;de Silva et al. 2003).

According to the World Bank, South Asia consists of six mainland countries, namely Afghanistan, Bangladesh, Bhutan, India, Nepal, and Pakistan, and two island countries Maldives and Sri Lanka (Lobo et al. 2011). Four of these countries (i.e., Bangladesh, India, Nepal, and Pakistan), account for 97% of the population in South Asia. Even though regional economic growth in South Asia was projected to gradually increase according a World Bank report in 2014 (World Bank 2014), there are still a large number of people living in poverty. Indeed, in 2010 and 2011, about 950 million people in Bangladesh, India, Nepal, and Pakistan lived on less than US$2 per day (World Bank 2016). Moreover, South Asia still has the highest rates and largest numbers of malnourished children, which is improving only very slowly (World Bank 2005).

It was estimated that in 2010, there were 298 million, 140 million, and 101million people in South Asia infected with *A. lumbricoides*, hookworm and *T. trichuris*, respectively, thus accounting for more than one-quarter of the world's soil-transmitted helminth infections (Pullan et al. 2014). In 2001, the World Health Assembly set the global target of regular deworming of at least 75% of school-aged children at risk of soil-transmitted helminthiasis by 2010 (WHO 2011a). Periodic large-scale preventive chemotherapy is recommended by the World Health Organization (WHO) when prevalence in school-aged children exceeds a pre-defined threshold (WHO 2006). Interestingly, a school-based national survey in Sri Lanka showed that the country had a prevalence of soil-transmitted helminth infections in 2003 below the WHO's threshold warranting preventive chemotherapy (Pathmeswaran et al. 2005). Data from the WHO Preventive Chemotherapy and Transmission Control (PCT) databank showed that before 2010, only Bhutan achieved the target of preventive chemotherapy with coverage of at least 75% of school-aged children at risk (WHO 2016). Bangladesh reached this target for the first time in 2012 and also in subsequent years, while Nepal reached this 75% coverage level only in 2012/2013. For Pakistan, no data are available for drug coverage of school-aged children from 2010 onwards. Information is lacking on infection risk of soil-transmitted helminthes and coverage of preventive chemotherapy in the Maldives.

High-resolution, model-based risk maps depicting the geographic distribution of soil-transmitted helminthiasis can assist disease control programs by helping governments and policy makers deliver and monitor preventive chemotherapy and other interventions. Large-scale risk estimates of soil-transmitted helminth infections have been generated for China, Latin America, and sub-Saharan Africa (Chammartin et al. 2013c;Karagiannis-Voules et al. 2015a;Lai et al. 2013). However, risk maps for soil-transmitted helminthiasis are currently lacking for South Asia. Bayesian geostatistical modeling is a powerful approach to produce risk maps for NTDs, by relating disease survey data to potential risk factors, thus predicting infection risk in areas without observed data (Clements et al. 2010a;Lai et al. 2015;Pullan et al. 2011).

In this paper we present the first comprehensive risk estimates of soil-transmitted helminthiasis in four countries of mainland South Asia (i.e., Bangladesh, India, Nepal, and Pakistan). Despite considerable efforts, we only obtained very little information on geo-referenced soil-transmitted helminth infection survey data in Afghanistan and Bhutan, and hence, these countries were not included in our Bayesian geostatistical modeling (Allen et al. 2004;Lobo et al. 2011).

3.2 Methods

3.2.1 Ethics statement

The work presented here is facilitated by soil-transmitted helminthiasis survey data primarily derived from the peer-reviewed literature. All data in our study were aggregated at the unit of villages, towns, or districts, and do not contain information that is identifiable at individual or household level. Hence, there were no specific ethics issues that warranted attention.

3.2.2 Soil-transmitted helminth infection data

A systematic review was undertaken following the PRISMA guidelines (Moher et al. 2009). We searched PubMed and ISI Web of Science from inception to January 15, 2016 for relevant publications that reported data of infection prevalence with any of the three common soil-transmitted helminths in Bangladesh, India, Nepal, and Pakistan. The following search terms were utilized: helminth* (OR ascari*, OR trichur*, OR hookworm*, OR necator, OR ankylostom*, OR ancylostom*, OR geohelminth*, OR nematode*) AND South Asia (OR Bangladesh, OR India, OR Nepal, OR Pakistan). The grey literatures (e.g., reports from the Ministry of Health or research groups, PhD thesis, or unpublished research through personal communication) were also considered. As we tried to identify all potentially relevant studies, we set no restriction for language of publication, date of survey, or study design in our search strategy. Further criteria were applied to exclude the ones that were not fit for our analysis.

With regard to inclusion, exclusion, and extraction of survey data, we followed the protocol put forth by Chammartin and colleagues (Chammartin et al. 2013c). In brief, we excluded case report, *in-vitro* studies, non-human studies, and surveys that did not report soil-transmitted helminth infection prevalence data. We also excluded case-control studies, clinical trials, or drug efficacy or intervention studies (except for baseline data or control groups), or locations with preventive chemotherapy occurred within one year (if such information was mentioned in the corresponding literatures), or studies done in specific groups that might not be representative (e.g., travelers, military personnel, expatriates, nomads, or displaced or migrating population).

Data were georeferenced and entered into the open-access Global Neglected Tropical Diseases (GNTD) database (Hürlimann et al. 2011). Our final analysis included data derived from surveys conducted from 1950 onwards, either school- or community-based, aggregated at village or town level, or on administrative divisions of level two or three (district level).

3.2.3 Climatic, demographic, environmental, and socioeconomic data

Climatic, demographic, and environmental data were obtained from readily accessible data sources, as shown in Table 3.1. Land surface temperature (LST) and normalized difference vegetation index (NDVI) were averaged over the period of 2000-2015, while land cover was summarized by the most frequent category over the period of 2001-2012. According to similar classes, land cover data were further re-grouped into seven categories: (i) grasslands; (ii) forests; (iii) scrublands and savannas; (iv) croplands; (v) urban; (vi) wet areas; and (vii) barren areas.

Socioeconomic data such as human influence index (HII), urban extents, and infant mortality rates (IMR) were downloaded from the Socioeconomic Data and Applications Center (Table 3.1). Geo-referenced water, sanitation, and hygiene (WASH) data for Bangladesh, Nepal, and Pakistan were extracted from the most recent Demographic and Health Surveys (DHS). For India, WASH information were obtained from the Census of India 2011, which were aggregated at administrative division of level three, stratified by rural and urban areas. The following indicators were extracted: proportion of households practicing open defecation, proportion of households with improved sanitation, and proportion of households with improved drinking water sources. An overview of WASH sources and data summaries of the relevant indicators are given in Table 3.2.

Visual Fortran version 6.0 (Digital Equipment Corporation; Maynard, United States of America) was employed to extract the environmental and socioeconomic data at survey locations. We linked the survey locations with missing data to the values at the nearest pixels. Surveys aggregated over districts were linked with the average values of the covariates within the districts and georeferenced using the corresponding centroids.

Table 3.1: Remote sensing data sources[a].

Source	Data type	Data period	Temporal resolution	Spatial resolution
MODIS/Terra[b]	LST[k]	2000-2015	8 days	1 km
MODIS/Terra[b]	NDVI[l]	2000-2015	16 days	1 km
MODIS/Terra[b]	Land cover	2001-2012	Yearly	500 m
WorldClim[c]	Elevation	2000	-	1 km
WorldClim[c]	Bioclimatic variables	1950-2000	-	1 km
SWBD[d]	Water bodies	2000	-	30 m
Köppen-Geiger[e]	Climate zones	1976-2000	-	50 km
ISRIC[f]	pH in water	-	-	10 km
Atlas of the Biosphere[g]	Soil moisture	1950-1999	-	50 km
WorldPop[h]	Grid population	2015	-	1 km
SEDAC[i]	HII[m]	1995-2004	-	1 km
SEDAC[i]	Urban extents	1990-2000	-	1 km
SEDAC[i]	IMR[n]	2000	-	4 km
GADM[j]	Geographic administrative boundaries	2012	-	-

[a]Data accessed on 01 January 2016.
[b]Moderate Resolution Imaging Spectroradiometer (MODIS)/Terra , available at: http://modis.gsfc.nasa.gov/.
[c]Available at: http://www.worldclim.org/current.
[d]Shuttle Radar Topography Mission Water Body Data (SWBD), available at:
http://gis.ess.washington.edu/data/vector/worldshore/index.html.
[e]World Maps of Köppen-Geiger climate classification, available at: http://koeppen-geiger.vu-wien.ac.at/shifts.htm.
[f]International Soil Reference and Information Center, available at: http://www.isric.org/data/isric-wise-derived-soil-properties-5-5-arc-minutes-global-grid-version-12.
[g]Available at: http://www.sage.wisc.edu/atlas/maps.php?datasetid=23&includerelatedlinks=1&dataset=23.
[h]The WorldPop project, available at: http://www.worldpop.org.uk/.
[i]Socioeconomic data and applications center, available at: http://sedac.ciesin.org/.
[j]Global Administrative Areas database, available at: http://www.gadm.org/.
[k]Land surface temperature (LST) day and night.
[l]Normalized difference vegetation index.
[m]Human influence index.
[n]Infant Mortality Rates.

3.2.4 Statistical analysis

Survey years were grouped into three periods (before 1980, 1980 to 1999, and from 2000 onwards) to study temporal trends. Continuous variables were standardized to mean zero and standard deviation (SD) one. Based on exploratory analysis, we converted continuous variables into categorical variables based on plotting of disease prevalence with each continuous variable to capture the non-linear relations. Pearson's correlation was used to check for continuous variables with a high correlation coefficient (>0.8) to avoid colinearity.

Bayesian variable selection was applied to identify the best set of predictors using a stochastic search approach (Scheipl et al. 2012). For each continuous covariate, a binary indicator was included in the model to indicate the exclusion/inclusion probability of the corresponding covariate. The priors for the coefficients of the covariates were constructed by

Table 3.2: Overview of WASH sources and data summaries of the relevant indicators by country.

Country		Bangladesh	India	Nepal	Pakistan
Sources		DHS[a]	Census of India[b]	DHS[a]	DHS[a]
Data periord		1999-2011	2011	2001-2011	2006
Number of locations		1661	2172	800	957
Type of locations		Point	Aggregated at administrative division of level three	Point	Point
Mean proportion (%)	Urban				
		Sanitation[c]	56.82	69.67	43.17	75.50
		Water[d]	99.44	76.70	92.86	95.10
		Defecation[e]	2.26	25.24	19.91	4.74
	Rural	Sanitation[c]	40.34	31.30	24.06	34.91
		Water[d]	97.77	63.75	80.40	86.25
		Defecation[e]	11.69	63.42	58.34	42.70

[a]Demographic and Health Surveys (DHS), available at: http://dhsprogram.com/.
[b]Census of India 2011, available at: http://censusindia.gov.in/.
[c]Proportion of households with improved sanitation.
[d]Proportion of households with improved drinking water sources.
[e]Proportion of households practicing open defecation.

a narrow spike (i.e., a normal distribution with variance close to zero to shrink the coefficient to zero) and a wide slab (i.e., a normal distribution that supports a non-zero coefficient). Inverse gamma prior distributions were employed for the variance parameters. We selected the covariates with inclusion probabilities (mean posterior distribution of indicators) greater than 0.5 for the final geostatistical analysis. Moreover, an adapted version of the above priors was used for categorical variables to include or exclude all categories of the variables simultaneously (Chammartin et al. 2013b). An additional indicator was introduced for each continuous variable to select either its linear or non-linear form, as details are provided elsewhere (Lai et al. 2013). The following 23 variables were considered for Bayesian variable selection: mean diurnal range, isothermality, temperature annual range, annual precipitation, precipitation of driest month, precipitation seasonality, precipitation of warmest quarter, precipitation of coldest quarter, elevation, HII, IMR, LST in the daytime, soil moisture, soil pH, NDVI, distance to the nearest freshwater bodies, proportion households with improved sanitation, proportion of households with improved water sources, proportion of households practicing open defecation, survey type (school- or community-based), urban extents, land cover, and climatic zones.

For each soil-transmitted helminth species, Bayesian geostatistical logistic regression models with spatial structure random effects were developed to obtain the spatially explicit estimates of infection risk (Chammartin et al. 2013a). Similar models were fitted on WASH indicators for Bangladesh, Nepal, and Pakistan using urban/rural as a covariate, as survey

locations of these data were not aligned in space with infection prevalence data. Geostatistical model predictions estimated the WASH indicators at the disease survey locations. Markov chain Monte Carlo (MCMC) simulation was applied to estimate the model parameters in Winbugs version 1.4 (Imperial College London and Medical Research Council; London, United Kingdom) (Lunn et al. 2009). Two chains were run and convergence was assessed by the Brooks-Gelman-Rubin diagnostic (Brooks & Gelman 1998).

The model was fitted on a random subset of 80% of the survey locations, and it was validated on the remaining 20% by calculations of the mean predictive error and the percentages of observations included in Bayesian credible intervals (BCI) of various probability coverages of the predictive distributions (Lai et al. 2015). A 5×5 km grid was overlaid to the study region, resulting in 222,555 pixels. Prediction of infection risk for each soil-transmitted helminth species was done at the centroids of the grid's pixels using Bayesian kriging (Diggle et al. 1998). We assumed independence of either species of soil-transmitted helminth and estimated the prevalence of infected by any species using the formula $p_S = p_A + p_T + p_h - p_A \times p_T - p_A \times p_h - p_T \times p_h + p_A \times p_T \times p_h$ where p_S, p_A, p_T and p_h indicate the predicted prevalence of any soil-transmitted helminth, *A. lumbricoides*, *T. trichiura*, and hookworm infections, respectively. Population-adjusted prevalence for each country was estimated by overlaying the pixel-based infection risk on gridded population to obtain the number of infected individuals at each pixel, which was then summed up within country and divided by the country population. The numbers of anthelmintic doses for preventive chemotherapy and the numbers of population requiring were estimated at pixel-level according to WHO control guidelines (WHO 2006) and summarized by country. In detail, we calculated the annualized pixel-level numbers of anthelmintic doses for school-aged children and for pre-school-aged children as zero at pixels with estimated prevalence <20%, as the corresponding population at pixels with estimated prevalence ≥20% and <50%, and as the double corresponding population at pixels with estimated prevalence ≥50%. The pixel-level numbers of school-aged children and pre-schooled-aged children requiring for preventive chemotherapy were calculated as zero at pixels with estimated prevalence <20%, and as the corresponding population at pixels with estimated prevalence ≥20%.

3.3 Results

3.3.1 Data summaries

We identified 3,394 records by systematically reviewing the peer-reviewed literature and an additional 11 records from the grey literature and personal communication. After excluding records according to our study protocol, 207 records remained, resulting in 501 surveys for *A. lumbricoides* at 430 unique locations, 384 surveys for *T. trichiura* at 332 unique locations, and 465 surveys for hookworm at 404 unique locations (Figure 3.1). Table 3.3 shows an overview of the soil-transmitted helminth surveys included in the final
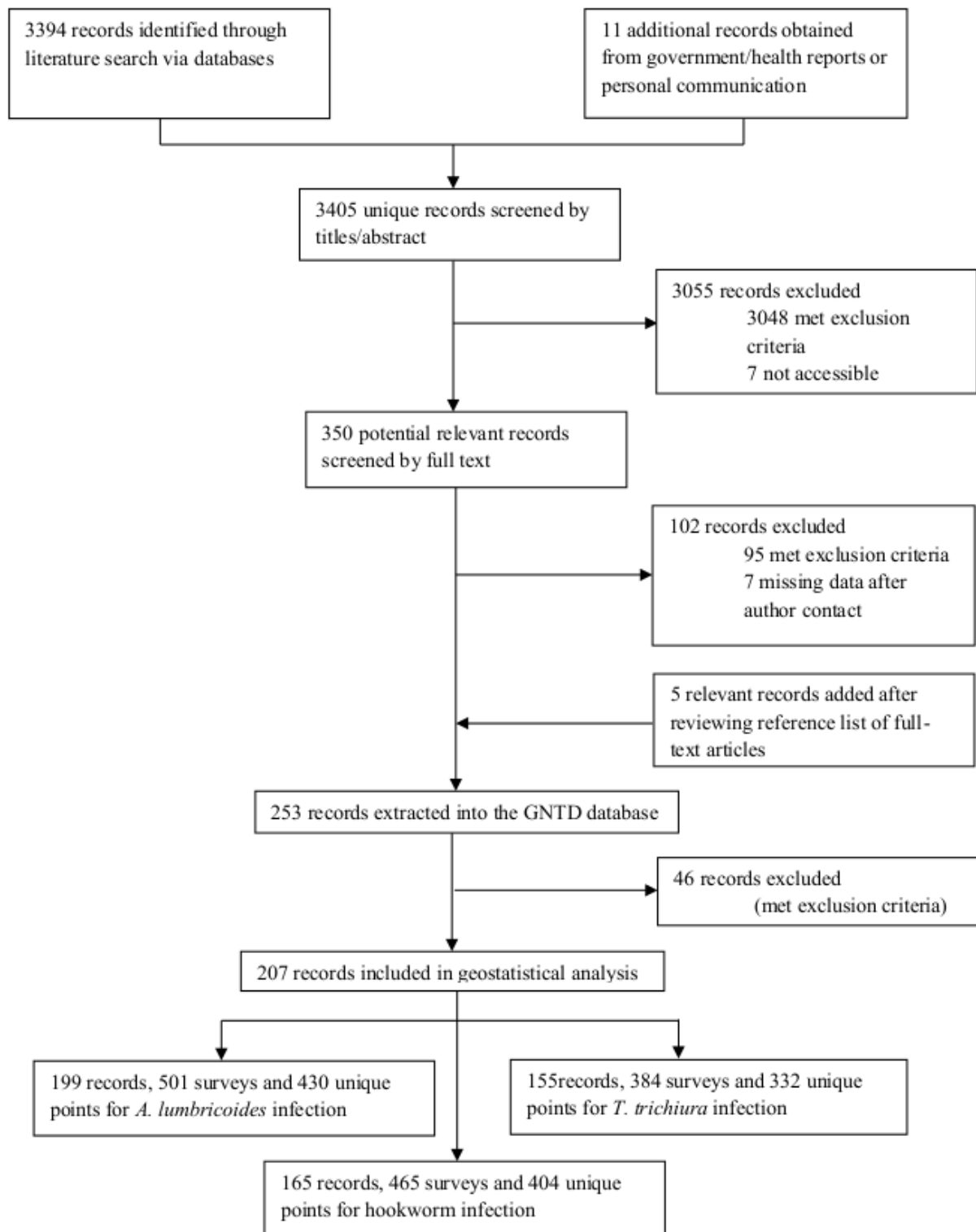
Figure 3.1: Data selection flow chart.

analysis, stratified by country. Figure 3.2 displays the geographic distribution of locations and observed prevalence for each soil-transmitted helminth species. There were only few surveys in the southwestern part of Pakistan and the central part of India.
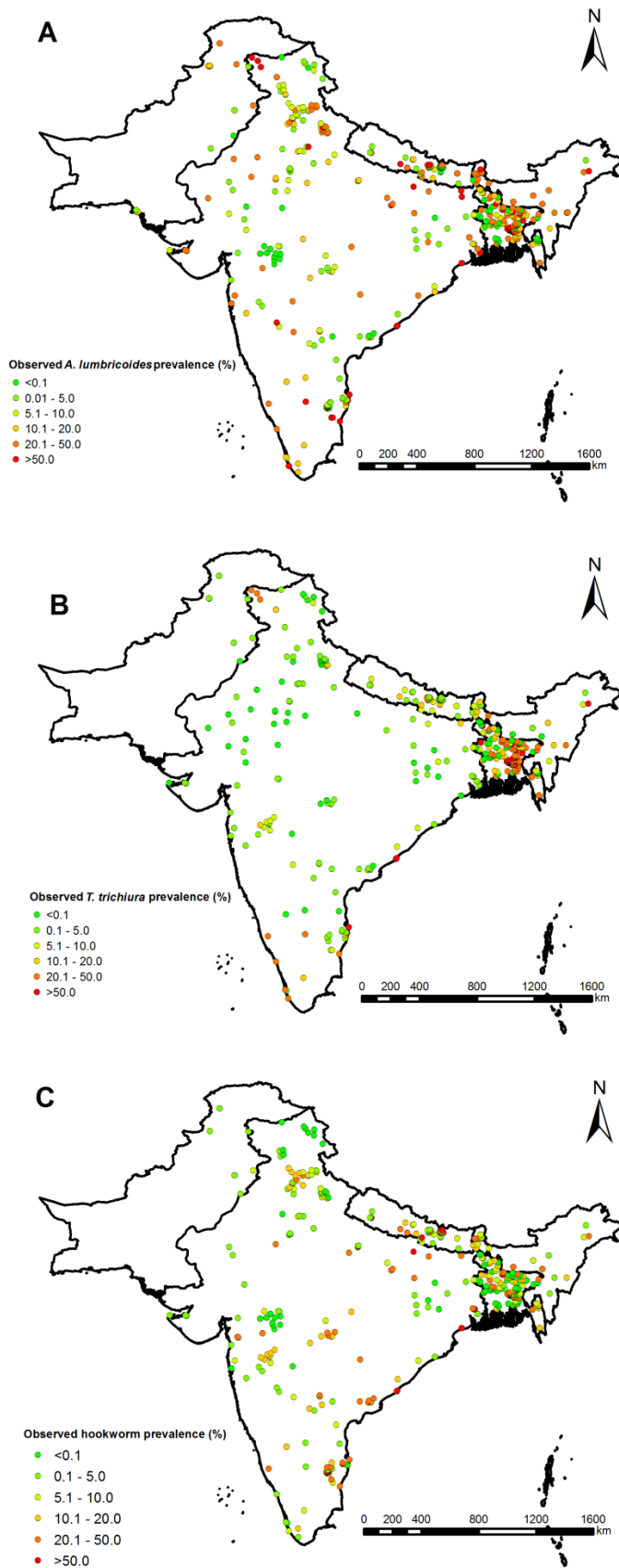
Figure 3.2: Survey locations and observed prevalence over the study region for (A) *A. lumbricoides*, (B) *T. trichiur*a and (C) hookworm.

Table 3.3: Overview of soil-transmitted helminth surveys.

Country		Bangladesh	India	Nepal	Pakistan	Total
		A. lumbricoides				
Relevant papers		24	128	33	14	199
Total surveys/locations		131/117	293/255	57/40	20/18	501/430
Survey type (surveys/locations)	School	8/5	91/83	34/26	10/9	143/123
	Community	123/112	202/172	23/14	10/9	358/307
Location type (surveys/locations)	Point	107/104	155/134	35/25	9/8	306/271
	District	24/13	138/121	22/15	11/10	195/159
Period		1957-2012	1963-2013	1995-2015	1976-2012	1957-2015
Year of Survey (surveys/locations)	<1980	7/5	43/38	0/0	6/6	56/49
	1980-2000	18/9	82/70	12/8	7/7	119/94
	>=2000	106/105	168/155	45/35	7/6	326/301
Diagnostic method (%)[a]	KK	2.24	38.17	40.00	0.00	27.46
	Dir	4.48	5.99	26.67	23.08	8.63
	Flot	75.37	18.30	0.00	23.08	31.33
	Concen	8.21	22.40	20.00	23.08	18.44
	NS	9.70	15.14	13.33	30.77	14.14
Raw prevalence (%)		56.60	16.70	18.60	10.30	19.40
		T. trichiura				
Relevant papers		21	95	30	9	155
Total surveys/locations		127/116	193/168	53/38	11/10	384/332
Survey type (surveys/locations)	School	7/4	88/82	31/24	4/4	130/114
	Community	120/112	105/86	22/14	7/6	254/218
Location type (surveys/locations)	Point	107/105	102/87	34/25	3/3	246/220
	District	20/11	91/81	19/13	8/7	138/112
Period		1957-2012	1963-2013	1995-2015	1976-2012	1957-2015
Year of Survey (surveys/locations)	<1980	7/5	31/26	0/0	2/2	40/33
	1980-2000	14/8	53/42	10/8	4/4	81/62
	>=2000	106/105	109/104	43/33	5/4	263/246
Diagnostic method (%)[a]	KK	2.34	43.72	43.64	0.00	28.77
	Dir	3.13	4.65	21.82	33.33	7.34
	Flot	78.91	9.3	0.00	13.33	31.15
	Concen	7.81	25.12	20.00	26.67	18.73
	NS	7.81	17.21	14.55	26.67	14.01
Raw prevalence (%)		45.40	6.20	16.00	2.60	10.30
		Hookworm				
Relevant papers		20	104	32	9	165
Total surveys/locations		128/116	272/238	54/40	11/10	465/404
Survey type (surveys/locations)	School	7/5	74/66	34/26	4/4	119/101
	Community	121/111	198/172	20/14	7/6	346/303
Location type (surveys/locations)	Point	107/105	169/147	32/25	2/2	310/279
	District	21/11	103/91	22/15	9/8	155/125
Period		1957-2012	1962-2013	1995-2015	1978-2012	1957-2015
Year of Survey (surveys/locations)	<1980	6/5	52/43	0/0	1/1	59/49
	1980-2000	18/9	78/67	10/8	5/5	111/89
	>=2000	104/104	142/133	44/35	5/4	295/276
Diagnostic method (%)[a]	KK	1.53	36.70	40.35	0.00	26.57
	Dir	4.58	4.38	24.56	29.41	7.37
	Flot	77.86	21.89	0.00	5.88	34.38
	Concen	7.63	22.90	21.05	23.53	18.50
	NS	8.40	14.14	14.04	41.18	13.19
Raw prevalence (%)		13.40	17.50	17.10	3.70	16.50

[a]KK: Kato-Katz; Dir: direct smear; Flot: stool flotation; Concen: stool concentration; NS: not stated.

3.3.2 Variable selection and geostatistical modeling

The selected variables from Bayesian variable selection are listed in Table 3.4. Maps of spatial distributions of the selected variables and the WASH indicators are shown in Figures 3.3 and 3.4. In the final geostatistical logistic regression models, the infection risk

decreased from 2000 onwards for hookworm, while the infection risk first increased in 1980-1999 and then decreased from 2000 onwards for *A. lumbricoides* and *T. trichiura* (Table 3.4). A negative association was identified for the prevalence of *A. lumbricoides* with LST in the daytime, whereas a positive association was found with HII. The model showed a lower prevalence of *A. lumbricoides* in school-aged children compared to that in the community population. Negative associations were identified for the *T. trichiura* infection risk with LST in the daytime and precipitation seasonality. A positive association was found for hookworm infection risk with proportion of households practicing open defecation, whereas a negative association was found with average NDVI.
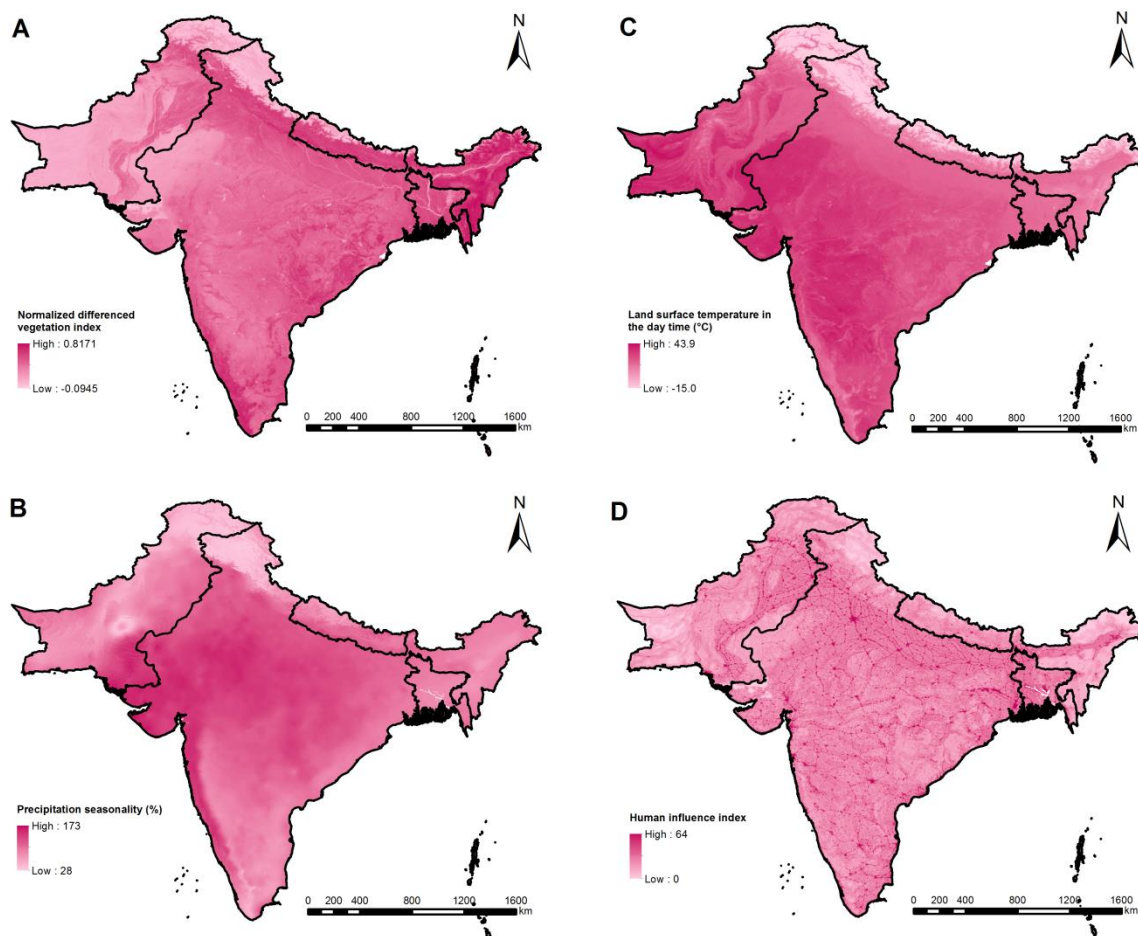
Figure 3.3: Spatial distributions of the selected variables: (A) normalized differnced vegetation index, (B) precipitation seasonality, (C) Land surface temperature in the day time and (D) human influence index.

3.3.3 Model Validation

Model validation indicated that the geostatistical logistic regression models were able to correctly estimate (within the 95% BCI) 83.1%, 77.0% and 72.0% of locations for *T. trichiura*,
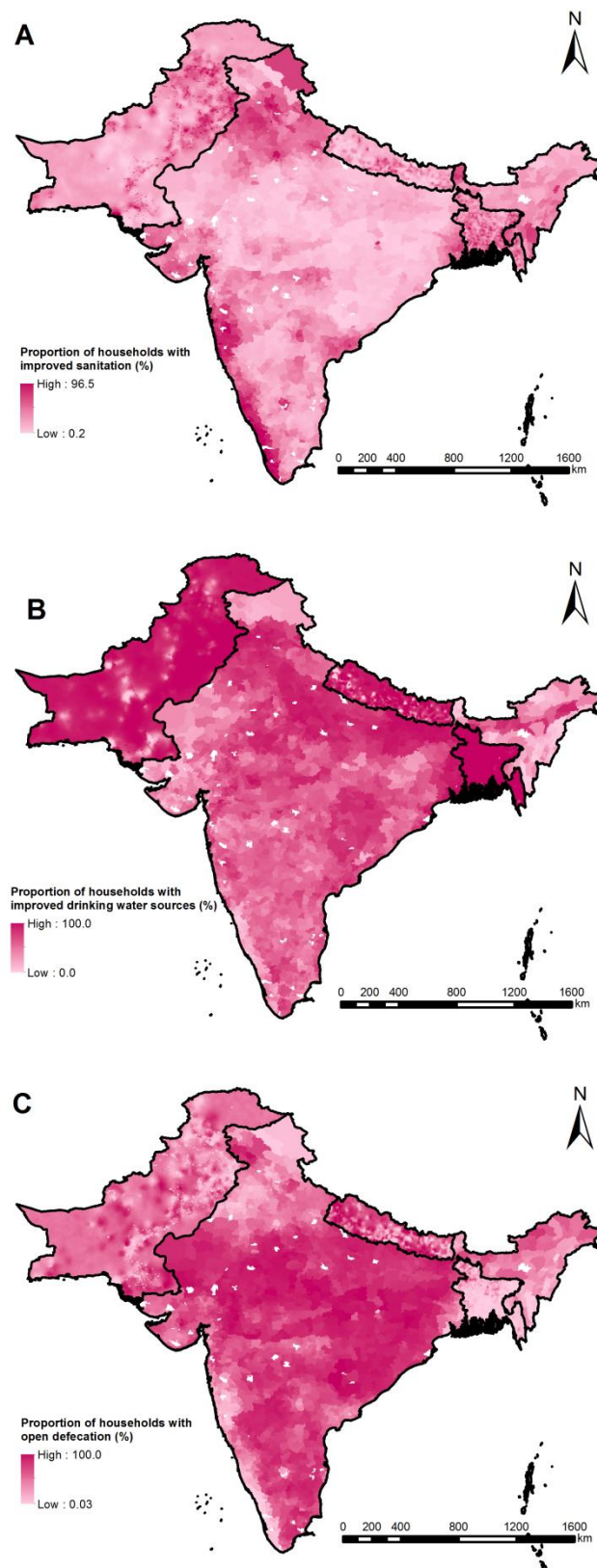
Figure 3.4: Spatial distribution of the WASH indicators: (A) proportion households with improved sanitation, (B) proportion of households with improved water sources and (C) proportion of households practicing open defecation.

Table 3.4: Posterior summaries (median and 95% Bayesian credible interval) of the geostatistical model parameters.

A. lumbricoides	Estimate
Period (<1980)[a]	
1980-1999	0.58 (0.46; 0.69)[b]
≥2000	-0.10 (-0.20; 0.01)
Survey type (school-based)[a]	
Community-based	0.33 (0.24; 0.43)[b]
Land surface temperature in the day time (25-30°C)[a]	
≤8	-0.04 (-3.09; 2.90)
8-20	0.53 (-0.21; 1.21)
20-25	-0.05 (-0.70; 0.53)
30-35	-0.90 (-1.38; -0.52)[b]
>35	-1.14 (-1.73; -0.47)[b]
Human influence index (≤22)[a]	
22-32	0.32 (-0.06; 0.79)
>32	1.30 (0.65; 2.00)[b]
Range (km)	115.53 (64.66; 193.98)
Spatial variance (σ^2_{sp})	1.87 (1.30; 2.61)
Non-spatial variance (σ^2_{nonsp})	1.17 (0.81; 1.63)
T. trichiura	Estimate
Period (<1980) [a]	
1980-1999	1.40 (1.25; 1.55)[b]
≥2000	0.37 (0.21; 0.52)[b]
Precipitation seasonality (bio15, 90-110%)[a]	
≤70	-0.46 (-2.15; 0.73)
70-90	0.50 (-0.27; 1.32)
110-130	-1.31 (-1.90; -0.68)[b]
>130	-1.71 (-2.74; -0.59)[b]
Land surface temperature in the day time (≤26.5°C)[a]	
26.5-31	-0.09 (-0.72; 0.44)
>31	-1.34 (-2.15; -0.62)[b]
Range (km)	133.87 (57.18; 359.72)
Spatial variance (σ^2_{sp})	2.20 (1.06; 3.48)
Non-spatial variance (σ^2_{nonsp})	1.12 (0.71; 1.88)
Hookworm	Estimate
Period (<1980)[a]	
1980-1999	-0.67 (-0.85; -0.48)[b]
≥2000	-0.60 (-0.78; -0.40)[b]
Normalized differenced vegetation index (≤0.40)[a]	
0.40-0.53	-0.53 (-0.93; -0.04)[b]
>0.53	0.21 (-0.03; 0.65)
Open defecation (≤15%)[a]	
15-60	0.42 (0.01; 0.71)[b]
>60	0.14 (-0.28; 0.69)
Range (km)	196.97 (104.62; 326.47)
Spatial variance (σ^2_{sp})	2.35 (1.52; 3.53)
Non-spatial variance (σ^2_{nonsp})	1.00 (0.56; 1.33)

[a]In brackets, baseline values are reported; [b]important effect based on 95% Bayesian credible interval (BCI).

A. lumbricoides and hookworm, respectively. The mean errors for *T. trichiura*, *A. lumbricoides* and hookworm were 3.9%, 4.7%, and 5.0% respectively, suggesting our models may under-estimate the infection risk of the three species.

3.3.4 Predictive risk maps

Figures 3.5A-C and Figures 3.5D-F present the species-specific predictive risk maps and the corresponding prediction uncertainty, respectively. A predictive infection risk map of any soil-transmitted helminth species and the associated prediction error are shown in Figures 3.6A and 3.6B. Moderate to high prevalence (>20%) of *A. lumbricoides* was mainly predicted in eastern areas of Bangladesh and some northern areas of Pakistan and India. Low prevalence (<5%) was predicted in areas of southern Pakistan and central India. Most of the study region had low prevalence (<5%) of *T. trichiura* infection, while the eastern areas of Bangladesh are characterized by moderate to high prevalence (>20%). Moderate to high hookworm prevalence (>20%) was predicted in some areas of southern and northern India.

3.3.5 Estimates of population-adjusted prevalence and number of people infected

Table 3.5 summarizes the population-adjusted predicted prevalence and estimated number of individuals infected with soil-transmitted helminths, stratified by country. Figure 3.6C shows the estimated number of individuals infected with any soil-transmitted helminth in South Asia. In the whole study region, the overall population-adjusted predicted prevalence of *A. lumbricoides*, *T. trichiura*, and hookworm were 15.0% (95% BCI: 12.8-17.4%), 4.9% (4.1-6.1%), and 8.8% (7.2-10.4%), respectively, corresponding to 245 million (95% BCI: 210-285 million), 81 million (67-100 million), and 145 million (117-171 million) infected individuals. The overall population-adjusted predicted prevalence of infected with any soil-transmitted helminth species was 24.2% (22.3-26.6%), which is equivalent to 397 million (365-436 million) infected individuals. The annual treatment needs for school-aged children requiring preventive chemotherapy with albendazole or mebendazole was estimated at 187 million (167-211 million) doses.

Bangladesh showed the highest population-adjusted predicted prevalence of *A. lumbricoides* (23.2%; 19.8-27.2%), *T. trichiura* (19.1%; 15.9-22.9%), and any soil-transmitted helminth species (39.4%; 35.9-43.4%). Nepal had the highest predicted prevalence of hookworm infection (12.4%; 9.2-16.2%) and the second highest of any soil-transmitted helminth infection in the region. India had the largest numbers of individuals estimated to be infected with *A. lumbricoides* (176 million; 148-211 million), *T. trichiura* (40 million; 31-53 million), hookworm (114 million; 91-136 million), and any soil-transmitted helminth (285 million; 257-319 million).

3.4 Discussion

We pursued a systematic review to collect available georeferenced data pertaining to prevalence of soil-transmitted helminth infections in South Asia, used Bayesian variable selection to identified important predictors, and developed Bayesian geostatistical logistic regression models for spatially explicit estimates of infection risk. To our knowledge, we present the first model-based, high-resolution infection risk estimates of the three soil-
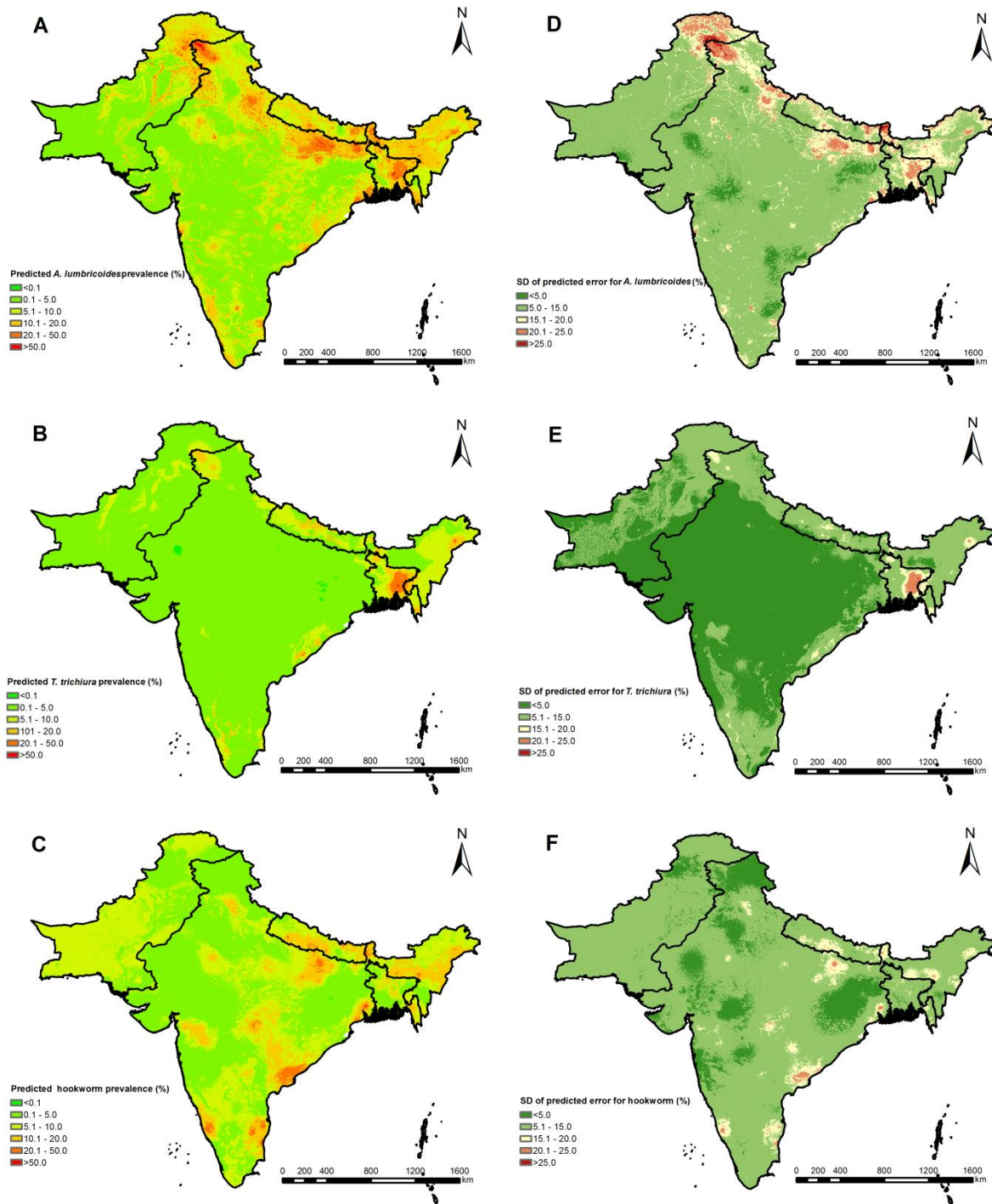
Figure 3.5: Species-specific model-based predictive risk maps from 2000 onwards. Predictive prevalence based on the median of the posterior predictive distribution of infection risk for (A) *A. lumbricoides*, (B) *T. trichiura* and (C) hookworm. Prediction uncertainty based on the standard deviation of the posterior predictive distribution of infection risk for (D) *A. lumbricoides*, (E) *T. trichiura* and (F) hookworm.
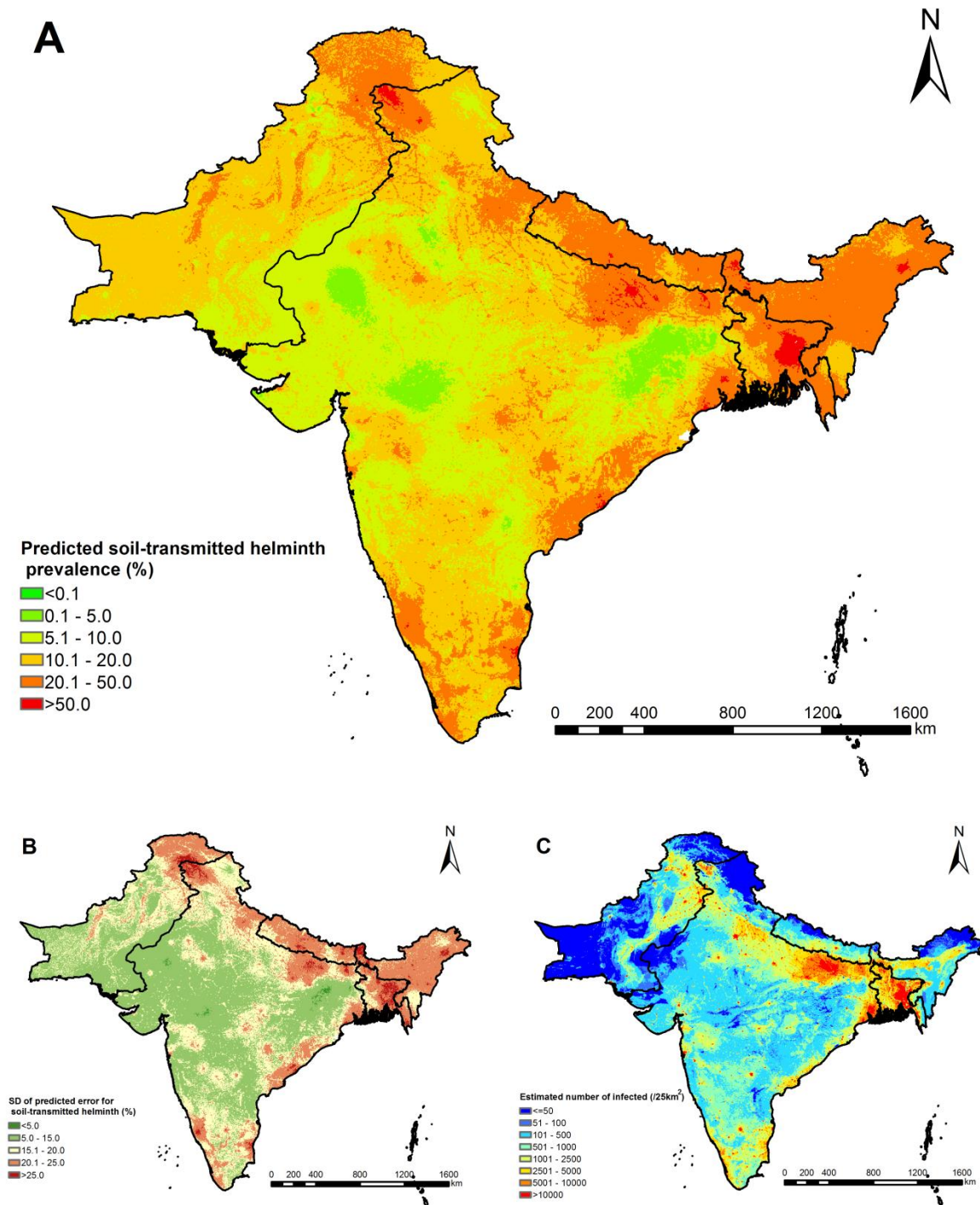
Figure 3.6: Model-based predictive risk map of any soil-transmitted helminth species from 2000 onwards. (A) Predictive prevalence based on the median of the posterior predictive distribution of infection risk. (B) Prediction uncertainty based on the standard deviation of the posterior predictive distribution of infection risk. (C) Number of infected based on the predictive prevalence and gridded population of 2015.

Table 3.5: Population-adjusted predicted prevalence (%) and number of individuals ($\times 10^6$) infected by soil-transmitted helminths by country[a].

Countries		Bangladesh	India	Nepal	Pakistan	Total
Population		157.35	1258.49	32.67	188.82	1637.33
Population of school-aged children		31.35	243.54	7.46	41.45	323.8
Population of pre-school-aged children		11.99	95.00	2.57	19.72	129.28
A. lumbricoides	Prevalence	23.22 (19.75; 27.19)	13.98 (11.79; 16.73)	17.34 (13.41; 21.88)	13.75 (10.68; 17.28)	14.97 (12.84; 17.42)
	No. of entire population infected	36.53 (31.08; 42.78)	175.92 (148.34; 210.59)	5.67 (4.38; 7.15)	25.97 (20.16; 32.64)	245.19 (210.21; 285.25)
	No. of school-aged children infected	5.94 (4.95; 7.12)	27.01 (22.59; 32.84)	1.02 (0.79; 1.34)	4.49 (3.39; 5.86)	38.49 (32.90; 45.9)
T. trichiura	Prevalence	19.06 (15.89; 22.91)	3.20 (2.47; 4.23)	7.50 (5.40; 10.64)	4.14 (2.58; 5.90)	4.94 (4.09; 6.09)
	No. of entire population infected	29.99 (25.00; 36.04)	40.29 (31.04; 53.23)	2.45 (1.76; 3.48)	7.82 (4.87; 11.14)	80.86 (66.91; 99.65)
	No. of school-aged children infected	5.97 (4.98; 7.18)	7.80 (6.01; 10.30)	0.56 (0.40; 0.79)	1.72 (1.07; 2.45)	16.10 (13.33; 19.84)
Hookworm	Prevalence	9.48 (7.22; 12.03)	9.06 (7.20; 10.83)	12.38 (9.15; 16.22)	6.03 (4.01; 9.07)	8.84 (7.17; 10.44)
	No. of entire population infected	14.92 (11.37; 18.92)	113.97 (90.66; 136.32)	4.04 (2.99; 5.30)	11.38 (7.57; 17.13)	144.68 (117.45; 170.96)
	No. of school-aged children infected	2.97 (2.26; 3.77)	22.06 (17.55; 26.38)	0.92 (0.68; 1.21)	2.50 (1.66; 3.76)	28.50 (23.19; 33.71)
Any soil-transmitted helminth	Prevalence	39.39 (35.88; 43.38)	22.66 (20.43; 25.34)	30.83 (26.62; 35.10)	21.22 (18.40; 24.78)	24.23 (22.29; 26.63)
	No. of entire population infected	61.98 (56.46; 68.26)	285.15 (257.05; 318.94)	10.07 (8.70; 11.47)	40.07 (34.75; 46.78)	396.73 (364.88; 436.02)
	No. of school-aged children infected	11.48 (10.41; 12.71)	49.32 (44.29; 55.65)	2.10 (1.81; 2.42)	7.71 (6.63; 9.12)	70.52 (64.96; 77.76)
School-aged children requiring preventive chemotherapy ($\times 10^6$)	Model-based estimate	21.63 (19.89; 23.4)	100.23 (88.3; 113.04)	4.41 (3.86; 4.92)	16.18 (13.53; 19.27)	142.46 (129.69; 156.36)
	WHO estimate[b]	31.71	157.27	7.53	21.32	217.83
Pre-school-aged children requiring preventive chemotherapy ($\times 10^6$)	Model-based estimate	8.27 (7.61; 8.95)	39.1 (34.44; 44.09)	1.52 (1.33; 1.70)	7.7 (6.43; 9.17)	56.6 (51.52; 62.09)
	WHO estimate[b]	14.90	63.12	2.88	9.12	90.02
Number of anthelmintic doses for school-aged children ($\times 10^6$)		31.95 (28.74; 35.39)	129.65 (112.27; 149.63)	5.88 (4.89; 6.76)	20.08 (16.24; 24.69)	187.29 (167.23; 210.56)
Number of anthelmintic doses for pre-school-aged children ($\times 10^6$)		12.22 (10.99; 13.53)	50.58 (43.79; 58.37)	2.03 (1.69; 2.33)	9.55 (7.72; 11.74)	74.26 (66.38; 83.39)

[a]Estimates are based on Gridded population of 2015; calculations are using on the median and 95% BIC of the posterior predictive distribution of the infection risk from 2000 onwards; [b]Obtained from WHO, PCT databank (http://www.who.int/neglected_diseases/preventive_chemotherapy/sth/en/) for the year 2014 (WHO 2016).

transmitted helminth species as well as a risk map of any soil-transmitted helminth infection in South Asia. The latter map is particularly important in terms of disease control as preventive chemotherapy with albendazole or mebendazole is based on the overall prevalence of any soil-transmitted helminth, usually estimated for school-aged population (Keiser & Utzinger 2008;WHO 2002a).

Our estimates suggest that in 2015, approximately 15.0% (95% BCI: 12.8-17.4%), 4.9% (4.1-6.1%), and 8.8% (7.2-10.4%) of the population in South Asia are infected with *A. lumbricoides*, *T. trichiura*, and hookworm infection, respectively, corresponding to 245 million (210-285 million), 81 million (67-100 million), and 145 million (117-171 million) people for the three species, respectively. We estimated lower numbers of infection for *A. lumbricoides* and *T. trichuris* while similar number of infection for hookworm, compared to the previous estimates in 2010 by Pullan and colleagues (Pullan et al. 2014). Of note, the later estimates were obtained by direct empirical approaches based on aggregated prevalence data at administrative level two or higher (Pullan et al. 2014), while our estimates are based on a rigorous Bayesian geostatistical model. We estimate that the number of school-aged children requiring preventive chemotherapy is 142 million (130-156 million) doses, which is lower than the 218 million doses estimated by WHO in 2014 (WHO 2016).

Our final models had reasonable predictive ability, as revealed by model suggesting that they were able to correctly predict 83.1%, 77.0%, and 72.0% of locations for *T. trichiura*, *A. lumbricoides*, and hookworm, respectively. However, as mean errors for all three species were larger than zero, our models may under-estimate the true prevalence of each species. We estimated a prevalence of any soil-transmitted helminth infection by assuming independence of the three species, which might over-estimate the reported prevalence as some researchers suggested a positive association between *A. lumbricoides* and *T. trichiura* (Booth & Bundy 1992;Tchuem Tchuenté et al. 2003a). However, it is difficult to adjust the calculation by adding a correction factor due to lack of co-infection data in South Asia (de Silva & Hall 2010). Our compiled survey data must be treated with caution, as sampling efforts and diagnostic approaches were not uniform. For example, more than 25% of the survey employed the widely-used Kato-Katz technique, while almost 15% had missing diagnostic information. As it was difficult to assess the quality of the diagnostic approach in a given survey and the number of surveys with detailed diagnostic information was rather low, we analyzed the data regardless of the diagnostic method and assumed common sensitivity and specificity across all surveys, which obviously might bias predictions (Booth et al. 2003;Nikolay et al. 2014).

We identified several climatic and environmental factors that were associated with soil-transmitted helminth infection, such as LST in the day time, precipitation seasonality, and NDVI. This finding is consistent with other reports emphasizing that environmental conditions play an important role in transmission (Appleton & Gouws 1996;Brooker et al. 2003;Tchuem Tchuenté 2011). A similar relationship was found between LST in the daytime

and *T. trichiura* infection risk in China (Lai et al. 2013). Socioeconomic factors impact the transmission of soil-transmitted helminths, mainly via influencing the behavior of people (Brooker et al. 2006). We found that HII was positively related to the infection risk of *A. lumbricoides*, indicating that direct human influence on ecosystems may have an effect on helminth transmission. Improvements of WASH are considered as interventions for sustainable control of soil-transmitted helminthiasis (Campbell et al. 2014). A systematic review and meta-analysis compiling results from individual-level studies showed a significant relation between WASH and soil-transmitted helminth infection risk (Strunz et al. 2014). Our results show that higher proportion of households practicing open defecation had a positive effect on hookworm infection risk, which is consistent with previous results (Freeman et al. 2015). However, the Bayesian variable selection did not identify important WASH indicators for either *A. lumbricoides* and *T. trichiura*. The effect of WASH can differ between genders, or sub-groups with exposure-related behavior patterns, thus our data that aggregated within villages or areas may be difficult to detect those variations (Freeman et al. 2013;Karagiannis-Voules et al. 2015b;Lai et al. 2015). In addition, a slight bias in prediction of the WASH indicators might exist, as each country implemented their own survey with different methodologies and in different years.

We included all publicly available point-specific survey data pertaining to soil-transmitted helminth infections in South Asia, as obtained through a systematic review of PubMed and ISI Web of Science. However, a considerable large amount of data which could not be accessed; indeed, approximately 40% of our survey data were aggregated at district level, and were not available at survey locations even after contacting the authors. To avoid data scarcity, we treated the data as point-specific data by setting the centroids of districts at the survey locations. This approach may lead to bias in the estimates of spatial parameters because we ignored within-district variation. We encourage researchers to share data disaggregated at the survey locations, to support secondary analyses for estimates of disease burden at high geographic resolution. Our study identified the areas with sparse data and thus it can help in the planning of future surveys. Furthermore, national surveys after large-scale deworming are important for monitoring and assessing control interventions and for avoiding overtreatment of population if the treatment estimates are relied on historic data. On the other hand, even though we excluded data from intervention studies or locations with preventive chemotherapy occurred within one year, if such information was mentioned in the corresponding literatures, we cannot obtain detailed geographic information of large preventive chemotherapy programmes in the whole study region. In addition, it is noted that India has implemented mass drug administration for lymphatic filariasis with almost 100% geographical coverage, and Bangladesh and Nepal are also with high rates of coverage (Lobo et al. 2011). Therefore we assumed that the effect of preventive chemotherapy for lymphatic filariasis is similar for the study region.

We estimated low-to-moderate (<50%) prevalence of hookworm infection in the northeastern part of Maharashtra State in India. Pullan and Brooker (Pullan & Brooker 2012) put forth very low risk of hookworm in these areas (prevalence <0.1%), however, their estimates are not supported by observed survey data in several villages of Nagpur district, which show prevalence of hookworm higher than 15% (Kulkarni et al. 1978). On the other hand, our models might over-estimate the risk of soil-transmitted helminth infection in the very high mountainous areas of the northern part of the study region, where the prediction uncertainty was high. Due to lack of survey data in these areas, further surveys are needed in order to have more precise estimates. Never the less, the predictions of the northern very high mountainous areas do not influence much the population-adjusted predicted prevalence as the population density and the estimated number of infected people in those areas are quite low (Figure 3.5C). As few survey data were available in the southwestern part of Pakistan, risk estimates in this region should be interpreted cautiously.

Our results reveal that any soil-transmitted helminth infection prevalence was higher than 20% in all the four countries subjected to detailed Bayesian-based geostatistical risk profiling, thus more efforts are needed to focus on control and intervention activities inside these countries. Furthermore, with the exception of Bangladesh, more effective strategies should be tailored for *A. lumbricoides* and hookworm infections compared to *T. trichiura* infection, the predicted prevalence of which was low (<5%) in most areas of the study region. We found that the infection risk in community population was higher than that of school-aged children for *A. lumbricoides*, and negligible difference between school-aged population and the entire community in the other two species. These findings support suggestions of other researchers that control strategies focusing on school-based deworming needs to be reassessed and extend treatments to other populations (e.g., preschool-aged children, women of childbearing age, and high occupational exposure adults) or to the whole community should be considered (Anderson et al. 2013;Karagiannis-Voules et al. 2015a;Lo et al. 2015).

We do not provide estimates for Afghanistan, Bhutan, and the island countries of Maldives and Sri Lanka. In fact, only very sparse georeferenced data were revealed by our system review for Afghanistan, Bhutan, and Maldives, and thus, it is difficult to infer reliable estimates. Even though surveys on soil-transmitted helminthiasis were carried out in Bhutan in 1985, 1986, 1989, and 2003, data with precise survey locations were not available (Allen et al. 2004). To our knowledge, Bhutan has had a school deworming program in place since 19888, but detailed reports on school deworming are not available (Allen et al. 2004). The survey conducted in 2003 observed an overall prevalence of 16.5% for soil-transmitted helminth infection in five schools of the Western region, which suggested a continuation of deworming in the country (Allen et al. 2004). On the other hand, we did not include Sri Lanka for further analysis because data disaggregated at village/school level were not publicly available after 2000. Sri Lanka launched a major 10-year deworming program between 1994 and 2005 and it is considered as a country where preventive chemotherapy on soil-transmitted

helminth infections is not necessary any longer according to the observed low prevalence from a national survey conducted in 2003 (Pathmeswaran et al. 2005). However, a school-based cross-sectional survey conducted in 2009 reported that the prevalence bounced back after cessation of preventive chemotherapy to above 20% in four districts of plantation sector (Kandy, Kegalle, Nuwara Eliya and Ratnapuram), suggesting that effective sustainable control activities should be undertaken in the this sector in order to maintain a low prevalence (Gunawardena et al. 2011).

In conclusion, we present the first model-based, high-resolution risk estimates of soil-transmitted helminth infections in four countries of South Asia, using data obtained from a systematic review and applying Bayesian geostatistical modeling for prediction based on environmental and socioeconomic predictors. The risk maps provide an overview of the geographic distribution of the diseases and highlight the need for up-to-date surveys to accurately evaluate the disease burden in the region.

Acknowledgements

Chapter 4   Bayesian geostatistical analysis and risk mapping of clonorchiasis in the People's Repubic of China

Ying-Si Lai[1,2], Xiao-Nong Zhou[3,4], Zhi-Heng Pan[5], Jürg Utzinger[1,2], Penelope Vounatsou[1,2]

[1]Swiss Tropical and Public Health Institute, Basel, Switzerland

[2]University of Basel, Basel, Switzerland

[3]National Institute of Parasitic Diseases, Chinese Center for Disease Control and Prevention, Shanghai, People's Republic of China

[4]WHO Collaborating Centre for Tropical Diseases, Key Laboratory of Parasite and Vector Biology, Ministry of Health, Shanghai, People's Republic of China

[5]Tianjin Modern Vocational Technology College, Tianjin, People's Republic of China

Abstract

Background: Clonorchiasis, one of the most important food-borne trematodiasis, affects millions of people in the People's Republic of China (P.R. China). Spatially explicit risk estimates of *Clonorchis sinensis* infection are needed in order to target control interventions.

Methods: Georeferenced survey data pertaining to infection prevalence of *C. sinensis* in P.R. China from 2000 onwards were obtained via a systematic review. Additional data were provided by the National Institute of Parasitic Diseases, Chinese Center for Diseases Control and Prevention. Bayesian geostatistical models were applied to quantify the relation between infection risk and important predictors, and to predict the risk of infection across P.R. China at high spatial resolution.

Principal Findings: We observed that the risk of *C. sinensis* infection increased over time, particularly from 2005 onwards. We estimate that around 14.8 million (95% Bayesian credible interval 13.8-15.8 million) people in P.R. China were infected with *C. sinensis* in 2010. Highly endemic areas (≥ 20%) were concentrated in southern and northeastern parts of the country. The provinces with the highest risk of infection and the largest number of infected people were Guangdong, Guangxi and Heilongjiang.

Conclusions/Significance: Our results provide spatially relevant information for guiding clonorchiasis control intervention in P.R. China. The trend toward higher risk of *C. sinensis* infection in the recent past urges the Chinese government to pay more attention on the public health importance of clonorchiasis and to target interventions to high-risk areas.

4.1 Introduction

Clonorchiasis is an important food-borne trematodiasis in Asia, caused by chronic infection with *Clonochis sinensis* (Lun et al. 2005;Qian et al. 2016). Symptoms of clonochiasis are related to worm burden; ranging from no or mild non-specific symptoms to liver and biliary disorders (Kim et al. 2011;Rim 1986). *C. sinensis* is classified as a carcinogen (Bouvard et al. 2009), as infection increases the risk of cholangiocarcinoma (Qian et al. 2012). Conservative estimates suggest that around 15 million people were infected with *C. sinensis* in 2004, over 85% of whom were concentrated in the People's Republic of China (P.R. China) (Coordinating Office of the National Survey on the Important Human Parasitic Diseases 2005;Fang et al. 2008;Qian et al. 2012). It has also been estimated that, in 2005, clonorchiasis caused a disease burden of 275,000 disability-adjusted life years (DALYs), though light and moderate infections were excluded from the calculation (Fürst et al. 2012).

Therefore, two national surveys have been conducted for clonorchiasis in P.R. China; the first national survey done in 1988-1992 and the second national survey in 2001-2004. Of note, the two surveys used an insensitive diagnostic approach with only one stool sample subjected to a single Kato-Katz thick smear. The first survey covered 30 provinces/autonomous regions/municipalities (P/A/M) with around 1.5 million people screened and found an overall prevalence of 0.37% (Yu et al. 1994). Data from the second survey, which took place in 31 P/A/M and screened around 350,000 people, showed an overall prevalence of 0.58% (Coordinating Office of the National Survey on the Important Human Parasitic Diseases 2005). Another dataset in the second national survey is a survey pertaining to clonorchiasis conducted in 27 endemic P/A/M using triplicate Kato-Katz thick smears from single stool sample. The overall prevalence was 2.4%, corresponding to 12.5 million infected people (Fang et al. 2008). Two main endemic settings were identified namely the provinces of Guangdong and Guangxi in the south and the provinces of Heilongjiang and Jilin in the northeast (Lun et al. 2005;Qian et al. 2012;Qian et al. 2016). In the latter setting, the prevalence was especially high in Korean (minority) communities. In general, males showed higher infection prevalence than females and the prevalence increases with age (Fang et al. 2008;Qian et al. 2012).

The life cycle of *C. sinensis* involves specific snails as first intermediate hosts, freshwater fish or shrimp as the second intermediate host, and humans or other piscivorous mammals as definitive hosts, who become infected through consumption of raw or insufficiently cooked infected fish (Keiser & Utzinger 2009;Lun et al. 2005;Qian et al. 2016;Sripa et al. 2010). Behavioral, environmental, and socioeconomic factors that influence the transmission of *C. sinensis* or the distribution of the intermediate hosts affect the endemicity of clonorchiasis. For example, temperature, rainfall, land cover/usage, and climate change that affect the activities and survival of intermediate hosts, are considered as potential risk factors (Keiser & Utzinger 2005;Petney et al. 2013). Socioeconomic factors and consumption of raw freshwater fish are particular important in understanding the epidemiology of clonorchiasis (Phan et al.

2011). Consumption of raw fish dishes is a deeply rooted cultural practice in some areas of P.R. China, while in other areas it has become popular in recent years, partially explained by food that is considered delicious or highly nutritious (Lun et al. 2005;Qian et al. 2013a;Qian et al. 2016;Tang et al. 1963b).

Treatment with praziquantel is one of the most important measures for the management of clonorchiasis, focusing on infected individuals or entire at-risk groups through preventive chemotherapy (Choi et al. 2010;WHO 2013). Furthermore, information, education, and communication (IEC), combined with preventive chemotherapy, is suggested for maintaining control sustainability (Oh et al. 2014). Elimination of raw or insufficient cooked fish or shrimp is an effective way for prevention of infection, but this strategy is difficult to be implemented due to deeply rooted traditions and misconceptions of people (Lun et al. 2005). Environmental modification is an additional way of controlling clonorchiasis by removing unimproved lavatories built adjacent to fish ponds in endemic areas, thus preventing water contamination by feces (Lun et al. 2005;Zhang et al. 2009).

Maps displaying the risk where a specific disease occurs are useful to guide prevention and control interventions. To our knowledge, only a province-level prevalence map of *C. sinensis* infection is available for P.R. China, while high-resolution, model-based risk estimates based on up-to-date survey data are currently lacking (Lun et al. 2005). Bayesian geostatistical modeling is a rigorous inferential approach to put forth risk maps. The utility of this approach has been demonstrated for a host of neglected tropical diseases, such as leishmaniasis, lymphatic filariasis, schistosomiasis, soil-transmitted helminthiasis, and trachoma (Chammartin et al. 2013c;Clements et al. 2010c;Karagiannis-Voules et al. 2013;Karagiannis-Voules et al. 2015a;Lai et al. 2013;Lai et al. 2015;Stensgaard et al. 2011b). The approach relies on the qualification of the relation between disease risk at observed locations and potential risk factors (e.g., environmental and socioeconomic factors), thus predicting infection risk in areas without observed data (Lai et al. 2015). Random effects are usually introduced to the regression equation to capture the spatial correlation between locations via a spatially structured Gaussian process (Lai et al. 2013).

Here, we compiled available survey data on clonorchiasis in P.R. China, identified important climatic, environmental and socioeconomic determinants, and developed Bayesian geostatistical models to estimate the risk of *C. sinensis* infection at high spatial resolution throughout the country.

4.2 Methods

4.2.1 Ethics statement

This work is based on clonorchiasis survey data extracted from the peer-reviewed literature and national surveys. All data were aggregated and do not contain any information at individual-level or household level. Hence there are no specific ethical issues that warranted attention.

4.2.2 Disease data

A systematic review was undertaken in PubMed, ISI Web of Science, China National Knowledge Internet (CNKI), and Wanfang Data from 2000 until 10 January 2016 to identify studies reporting school, village, town, and county-level prevalence data of clonorchiasis in P.R. China. The search terms were "clonorchi*" (OR "liver fluke*") AND "China" for Pubmed and ISI Web of Science, and "huazhigaoxichong" (OR "ganxichong") for CNKI and Wanfang. Government reports and other grey literature (e.g., MSc and PhD thesis, working reports from research groups) were also considered. There were no restrictions on language or study design. The dataset at county-level of clonorchiasis conducted in 27 endemic P/A/M in the second national survey were provided by the National Institute of Parasitic Diseases, Chinese Center for Diseases Control and Prevention (NIPD, China CDC; Shanghai, P.R. China).

Titles and abstracts of articles were screened to identify potentially relevant publications. Full text articles were obtained from seemingly relevant pieces that were screened for *C. sinensis* infection prevalence data. Data were excluded if they stemmed from hospital-based surveys, case-control studies, clinical trials, drug efficacy studies, or intervention studies (except for baseline or control group data). Studies on clearly defined populations (e.g., travellers, military personnel, expatriates, nomads, or displaced or migrating populations) that are not representative of the general population were also excluded. We further excluded data based on direct smear or serum diagnostics due to the known low sensitivity or the unability to differentiate between past and active infection, respectively. All valuable data were georeferenced and entered into the open-access Global Neglected Tropical Diseases (GNTDs) database (Hürlimann et al. 2011).

4.2.3 Environmental, socioeconomic, and demographic Data

Environmental, socioeconomic, and demographic data were obtained from different accessible remote sensing data sources (Table 4.1). Land cover data were re-grouped to the following five categories: (i) forests, (ii) scrublands and grass, (iii) croplands, (iv) urban, and (v) wet areas. They were summarized by the most frequent category for each pixel over the period 2001-2004. Land surface temperature (LST) and normalized difference vegetation index (NDVI) were averaged annually. Climate zone data were grouped to four categories: (i) equatorial, (ii) arid, (iii) warm temperate, and (iv) snow with polar. We used human influence

index (HII), urban extents, and gross domestic product (GDP) per capita as socioeconomic proxies. The latter was obtained from the P.R. China yearbook full-text database at county-level for the year 2008 and georeferenced for the purpose of our study. Details about data processing are provided in Lai *et al.* (Lai et al. 2013). We georeferenced surveys reporting aggregated data at county level by the county centroid and linked them to the average values of our covariates within the specific county.

Table 4.1: Remote sensing data sources[a].

Source	Data type	Data period	Temporal resolution	Spatial resolution
MODIS/Terra[b]	LST[j]	2001-2015	8 days	1 km
MODIS/Terra[b]	NDVI[k]	2001-2015	16 days	1 km
MODIS/Terra[b]	Land cover	2001-2004	Yearly	1 km
WorldClim[c]	Elevation	2000	-	1 km
WorldClim[c]	Precipitation	1950-2000	Monthly	1 km
SWBD[d]	Water bodies	2000	-	30 m
Köppen-Geiger[e]	Climate zones	1976-2000	-	50 km
ISRIC[f]	Soil pH	-	-	10 km
Atlas of the Biosphere[g]	Soil-moisture	1950-1999	-	50 km
SEDAC[h]	Population data	2010	-	5 km
SEDAC[h]	HII[l]	1995-2004	-	1 km
SEDAC[h]	Urban extents	1990-2000	-	1 km
China Yearbook[i]	GDP per capita	2008	-	County-level

[a]Land cover data accessed on 01 June 2011 and other data accessed on 01 January 2016.
[b]Moderate Resolution Imaging Spectroradiometer (MODIS)/Terra , available at: https://lpdaac.usgs.gov/
[c]Available at: http://www.worldclim.org/current.
[d]Shuttle Radar Topography Mission Water Body Data (SWBD), available at:
http://gis.ess.washington.edu/data/vector/worldshore/index.html.
[e]World maps of Köppen-Geiger climate classification, available at: http://koeppen-geiger.vu-wien.ac.at/shifts.htm.
[f]International Soil Reference and Information Center, available at: http://www.isric.org/data/isric-wise-derived-soil-properties-5-5-arc-minutes-global-grid-version-12.
[g]Available at: http://www.sage.wisc.edu/atlas/maps.php?datasetid=23&includerelatedlinks=1&dataset=23.
[h]Socioeconomic data and applications center, available at: http://sedac.ciesin.org/.
[i]China yearbook full-text database, available at: http://acad.cnki.net/Kns55/brief/result.aspx?dbPrefix=CYFD.
[j]Land surface temperature (LST) day and night.
[k]Normalized difference vegetation index.
[l]Human influence index.

4.2.4 Statistical analysis

We grouped survey years into two categories (before 2005 and from 2005 onwards). We standardized continuous variables to mean zero and standard deviation one (SD=1). We calculated Pearson's correlation between continuous variables and dropped one variable

among pairs with correlation coefficient greater than 0.8 to avoid collinearity. Furthermore, continuous variables were converted to categorical ones based on their tertiles.

We carried out Bayesian variable selection to identify the most important predictors using spike-and-slab prior distributions for the regression coefficients. Methodological details have been described elsewhere (Lai et al. 2013). Through the variable selection, we also identified the best functional form (i.e., linear or categories) for continuous variables.

We developed Bayesian geostatistical logistic regression models with location-specific random effects to obtain spatially explicit *C. sinensis* risk estimates. We assumed that the number of positive individuals Y_i arises from a binominal distribution $Y_i \sim Bn(p_i, n_i)$, where n_i and p_i are the number of individuals examined and the probability of infection at location i ($i =1, 2, \ldots, L$), respectively. We modelled the covariates on the logit scale, that is $\mathrm{logit}(\mathrm{p}_i) = \beta_0 + \sum_{k=1} \beta_k \times X_i^{(k)} + \varepsilon_i$, where β_k is the regression coefficient of the kth covariate $X_i^{(k)}$. We assumed that location-specific random effects $\vec{\varepsilon} = (\varepsilon_1, \ldots, \varepsilon_L)^T$ followed a multivariate normal distribution $\vec{\varepsilon} \sim MVN(0, \Sigma)$, with exponential correlation function $\Sigma_{ij} = \sigma_{sp}^2 \exp(-\rho d_{ij})$, where d_{ij} is the Euclidean distance between locations, and ρ is the parameter corresponding to the correlation decay. We estimated the spatial range as the minimum distance with spatial correlation less than 0.1 by $-\log(0.1)/\rho$. We formulated the model in a Bayesian framework and applied Markov Chain Monte Carlo (MCMC) simulation to estimate the model parameters in Winbugs version 1.4 (Imperial College London and Medical Research Council; London, United Kingdom) (Lunn et al. 2009). We assessed the convergence using the Brooks-Gelman-Rubin diagnostic (Brooks & Gelman 1998).

We fitted the model on a random subset of 80% survey locations and used the remaining 20% for model validation. Mean error and the percentage of observations covered by 95% Bayesian credible interval (BCI) of posterior predicted prevalence were calculated to access the model performance. Bayesian kriging was employed to predict the *C. sinensis* infection risk at the centroids of pixels from a 5 × 5 km grid over P.R. China (Diggle et al. 1998). Population-adjusted prevalence for each province was calculated by summing the pixel-level predicted number of infected individuals (estimated by the pixel-level predicted infection risk overlaid on gridded population) within each province and dividing by the population of the province.

4.3 Results

4.3.2 Data summaries

A data selection flow chart for the systematic review is presented in Figure 4.1. We identified 7,575 records through literature search and obtained one additional report provided by NIPD,

China CDC (Shanghai, P.R. China). According to our inclusion and exclusion criteria, we obtained 143 records for the final analysis, resulting in 691 surveys for *C. sinensis* at 633 unique locations published from 2000 onwards. A summary of our survey data, stratified by province, is provided in Table 4.2. The geographic distribution of locations and observed *C. sinensis* prevalence are shown in Figure 4.2. We obtained data from most provinces, except Iner Mongolia, Ningxia, Qinghai, and Tibet. We collected more than 50 surveys in Guangdong, Guangxi, Hunan, and Jiangsu provinces. Over 45% of surveys were conducted from 2005 onwards. Around 90% of surveys used the Kato-Katz technique for diagnosis, while 0.14% surveys had no information on the diagnostic technique employed. The overall raw prevalence, calculated as the total number of infected divided by the total number of people examined from all observed surveys, was 9.7%.
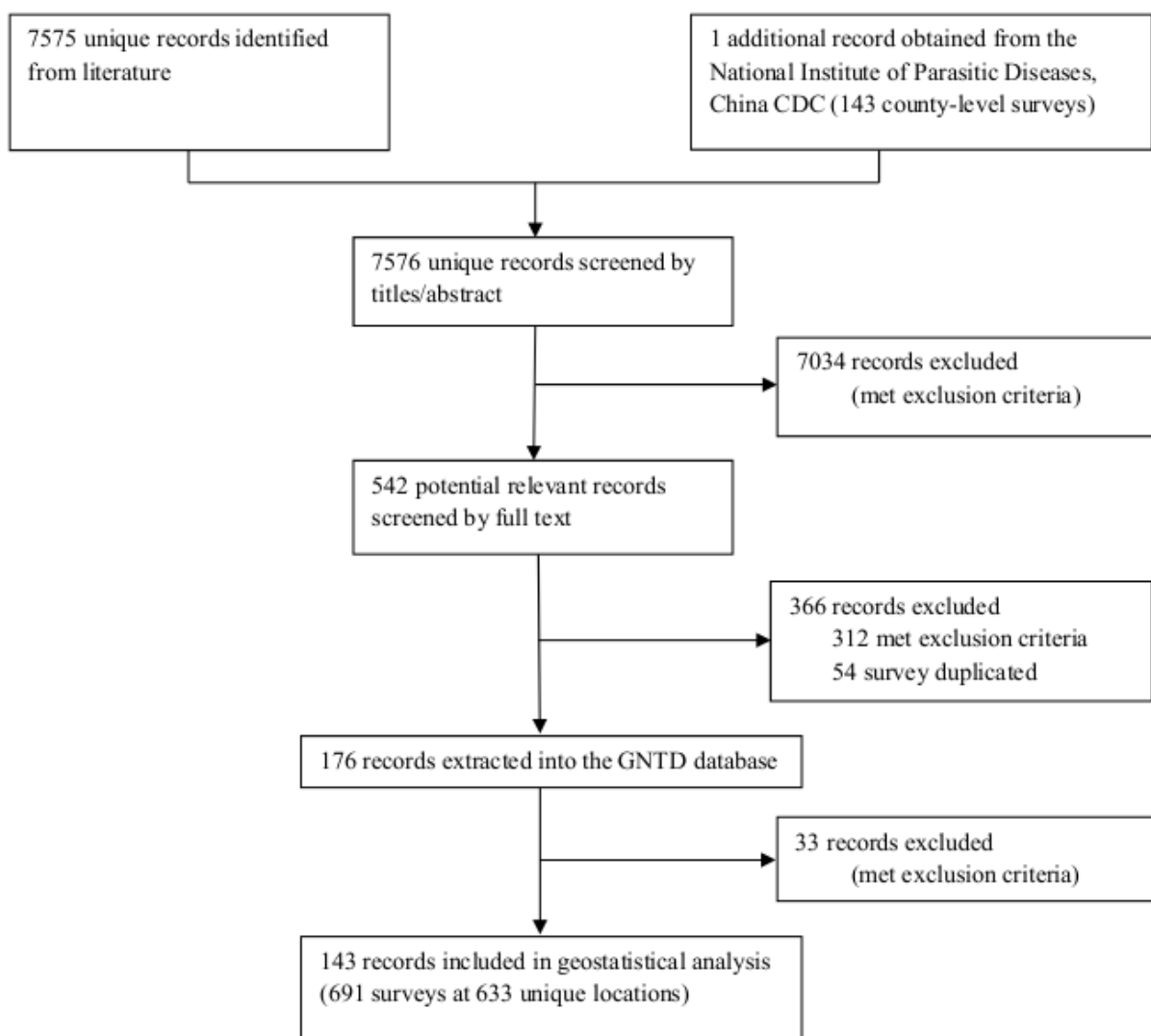
Figure 4.1: Data selection flow chart.

Table 4.2: Overview of clonorchiasis survey data in China.

Province	Relevant papers	Total[a]	Raw prevalence (%)	Location type[a] Point	County	Period	Year of Survey[a] 2000-2004	>2005	Diagnostic method (%)[b] KK	Digest	Conc	Sendi	NS/ missing
Anhui	3	12/12	0.65	1/1	11/11	2001-2014	11/11	1/1	100.00	0.00	0.00	0.00	0.00
Beijing	1	3/3	0.00	0/0	3/3	2004-2004	3/3	0/0	100.00	0.00	0.00	0.00	0.00
Chongqing	4	17/17	0.37	6/6	11/11	2002-2009	15/15	2/2	94.12	0.00	0.00	0.00	0.00
Fujian	5	18/18	0.27	17/17	1/1	2002-2011	16/16	2/2	100	0.00	0.00	0.00	0.00
Gansu	1	3/3	0.00	0/0	3/3	2004-2004	3/3	0/0	100	0.00	0.00	0.00	0.00
Guangdong	53	211/186	14.64	164/145	47/41	2000-2015	46/35	165/151	94.31	0.47	0.00	0.00	0.47
Guangxi	17	60/57	17.27	37/34	23/23	2000-2014	24/23	36/34	88.33	0.00	0.00	3.33	0.00
Guizhou	2	4/4	0.04	1/1	3/3	2004-2013	3/3	1/1	100.00	0.00	0.00	0.00	0.00
Hainan	4	10/10	0.30	7/7	3/3	2002-2004	10/10	0/0	100.00	0.00	0.00	0.00	0.00
Hebei	1	3/3	0.02	0/0	3/3	2004-2004	3/3	0/0	100.00	0.00	0.00	0.00	0.00
Heilongjiang	10	34/32	36.24	11/11	23/21	2001-2012	30/29	4/3	67.65	0.00	2.94	29.41	0.00
Henan	3	21/21	0.10	3/3	18/18	2000-2011	20/20	1/1	100.00	0.00	0.00	0.00	0.00
Hubei	4	37/37	1.98	3/3	34/34	2000-2004	37/37	0/0	16.22	0.00	83.78	0.00	0.00
Hunan	8	65/60	22.78	52/47	13/13	2002-2012	25/25	40/35	100.00	0.00	0.00	0.00	0.00
Jiangsu	14	53/39	0.60	28/25	25/14	2000-2014	26/21	27/18	100.00	0.00	0.00	0.00	0.00
Jiangxi	3	9/9	0.08	6/6	3/3	2002-2004	9/9	0/0	100.00	0.00	0.00	0.00	0.00
Jilin	7	25/23	14.75	15/13	10/10	2002-2012	12/11	13/12	100.00	0.00	0.00	0.00	0.00
Liaoning	3	13/13	0.77	4/4	9/9	2004-2007	9/9	4/4	100.00	0.00	0.00	0.00	0.00
Iner Mongolia	0	0/0	-	0/0	0/0	-	0/0	0/0	-	-	-	-	-
Ningxia	0	0/0	-	0/0	0/0	-	0/0	0/0	-	-	-	-	-
Qinghai	0	0/0	-	0/0	0/0	-	0/0	0/0	-	-	-	-	-
Shaanxi	1	3/3	0.00	0/0	3/3	2004-2004	3/3	0/0	100.00	0.00	0.00	0.00	0.00
Shando-ng	10	36/34	0.06	13/13	23/21	2000-2012	22/20	14/14	88.89	0.00	2.78	0.00	0.00
Shanghai	1	3/3	0.00	0/0	3/3	2004-2004	3/3	0/0	100.00	0.00	0.00	0.00	0.00
Shanxi	1	3/3	0.00	0/0	3/3	2004-2004	3/3	0/0	100.00	0.00	0.00	0.00	0.00
Sichuan	4	24/22	0.11	2/2	22/20	2003-2012	18/16	6/6	100.00	0.00	0.00	0.00	0.00
Tianjin	2	8/5	0.18	0/0	8/5	2004-2004	8/5	0/0	100.00	0.00	0.00	0.00	0.00
Xinjiang Uygur	1	4/4	0.03	0/0	4/4	2004-2004	4/4	0/0	100.00	0.00	0.00	0.00	0.00
Tibet	0	0/0	-	0/0	0/0	-	0/0	0/0	-	-	-	-	-
Yunnan	2	9/9	0.00	6/6	3/3	2004-2004	9/9	0/0	100.00	0.00	0.00	0.00	0.00
Zhejiang	1	3/3	0.00	0/0	3/3	2004-2004	3/3	0/0	100.00	0.00	0.00	0.00	0.00
Total	143	691/633	9.69	376/344	315/289	2000-2015	375/363	316/288	90.45	0.14	4.78	1.74	0.14

[a]Presented as surveys/points.

[b]KK: Kato-Katz; Digest: sodium hydroxide (NaOH) digestion, Conc: stool concentration; Sendi: stool sendimentation; NS/missing: not stated or missing.

4.3.2 Variable selection, geostatistical modeling, and model validation

We considered a total of 13 variables (i.e., land cover, urban extents, precipitation, GDP per capita, HII, soil moisture, elevation, LST in the daytime, LST at night, NDVI, distance to the nearest open water bodies, PH in water, and climate zones) for Bayesian variable selection. Elevation, NDVI, distance to the nearest open water bodies, and land cover were selected for the final Bayesian geostatistical logistic regression model.

The parameter estimates arising from the geostatistical model fit are shown in Table 4.3. The infection risk of *C. sinensis* increased from 2005 onwards. Elevation had a negative effect on infection risk. People living at distance between 2.5 and 7.0 km from the nearest open water bodies had a lower risk compared to those living in close proximity (<2.5 km). The risk of *C. sinensis* infection was lower in areas covered by forest, shrub, and grass compared to crop. Furthermore, NDVI was positively correlated with the risk of *C. sinensis* infection.

Table 4.3: Posterior summaries (median and 95% Bayesian credible interval) of the geostatistical model parameters for clonorchiasis.

Variable	Estimate
Year	
2000-2004	1.00
>=2005	0.43 (0.38; 0.48)[a]
Elevation	-1.34 (-1.97; -0.83)[a]
NDVI	0.61 (0.41; 0.81)[a]
Distance to the nearest open water bodies (km)	
<=2.5	1.00
2.5 - 7.0	-0.46 (-0.72; -0.22)[a]
>7.0	0.05 (-0.38; 0.46)
Land cover	
Crop	1.00
Forest	-0.91 (-1.37; -0.42)[a]
Shrub and grass	-0.57 (-1.07; -0.16)[a]
Urban	0.06 (-0.36; 0.48)
Wet	0.24 (-0.39; 0.98)
Spatial range (km)	259.80 (199.70; 373.01)
σ_{sp}^2	13.64 (9.90; 20.51)

[a]Significant correlation based on 95% Bayesian credible interval.

Model validation indicated that the Bayesian geostatistical logistic regression models were able to correctly estimate (within a 95% BCI) 71.7% of locations for *C. sinensis*. The mean error was -0.07%, suggesting that our model may slightly over-estimate the infection risk of *C. sinensis*.

4.3.3 Predictive risk maps and estimates of number of people infected

Figure 4.2A shows the model-based predicted risk map of *C. sinensis* for P.R. China. High prevalence (≥20%) was estimated in some areas of southern and northeast parts of Guangdong province, southwest and north parts of Guangxi province, southwest part of Hunan province, west bordering region of Heilongjiang and Jilin provinces, and eastern part of Heilongjiang province. Most regions of northwestern China and eastern costal-line areas had zero to very low prevalence (<0.01%). The prediction uncertainty is shown in Fig 4.2C.

Table 4.4 reports the population-adjusted predicted prevalence and the number of individuals infected with *C. sinensis* in P.R. China, stratified by province, based on gridded population of 2010. The overall population-adjusted predicted prevalence of clonorchiasis was 1.18% (95% BCI: 1.10%-1.25%) in 2010, corresponding to 14.8 million (95% BCI: 13.8-15.8 million) infected individuals. The three provinces with the highest infection risk were Heilongjiang (7.21%, 95% BCI: 5.95-8.84%), Guangdong (6.96%, 95% BCI: 6.62-7.27%) and Guangxi (5.52%, 95% BCI: 4.97-6.06%) provinces. Provinces with very low risk estimates (median predicted prevalence < 0.01%) are Gansu, Ningxia, Qinghai, Shanghai, Shanxi, Tibet, and Yunnan. Guangdong, Heilongjiang, and Guangxi were the top three
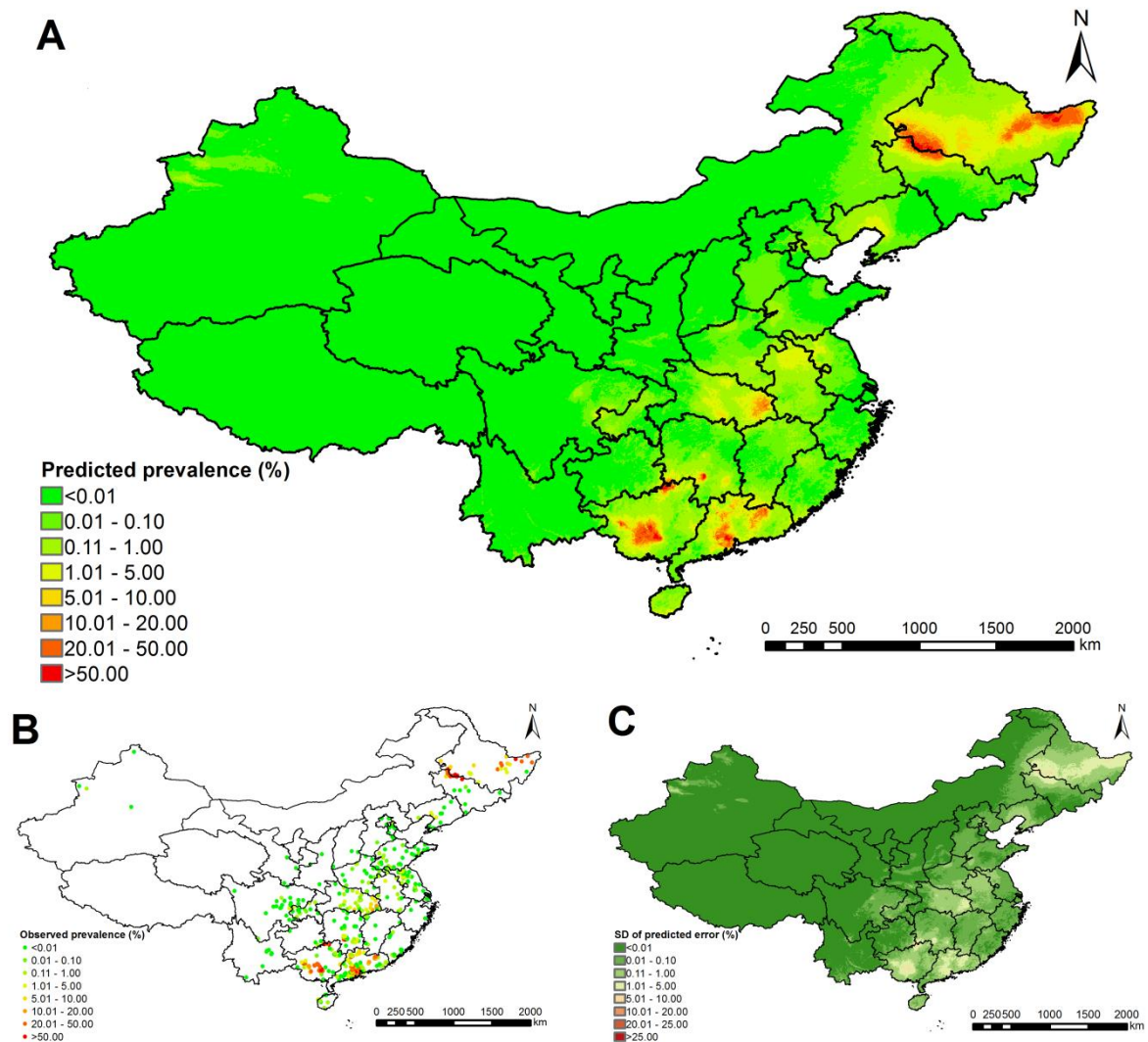
Figure 4.2: Model-based prediction risk maps of *C. sinensis* infection from 2005 onwards. (A) Predictive prevalence based on the median of the posterior predictive distribution of infection risk. (B) Survey locations and observed prevalence over P.R. China. (C) Prediction uncertainty based on the standard deviation of the posterior predictive distribution of infection risk.

provinces with the highest number of people infected: 6.34 million (95% BCI: 6.03-6.62 million), 3.05 million (2.52-3.74 million) and 2.08 million (1.87-2.28 million), respectively.

4.4 Discussion

To our knowledge, we present the first model-based, high-resolution estimates of *C. sinensis* infection risk in P.R. China. Risk maps were produced through Bayesian geostatistical modeling of clonorchiasis survey data from 2000 onwards, readily adjusting for environmental/climatic predictors. Our methodology is based on a rigorous approach for spatially explicit estimation of neglected tropical diseases (Karagiannis-Voules et al. 2015a).

Table 4.4: Population-adjusted predicted prevalence and estimated number of infected individuals per province of in 2010[a].

Provinces	Population ($\times 10^6$)	Prevalence (95% BCI[b]) (%)	No. infected (95% BCI[b]) ($\times 10^3$)
Anhui	54.89	0.66 (0.51; 0.89)	363.67 (279.09; 489.63)
Beijing	16.99	0.02 (0.00; 0.11)	2.64 (0.20; 17.98)
Chongqing	26.72	0.13 (0.09; 0.20)	35.96 (23.52; 52.15)
Fujian	32.8	0.09 (0.05; 0.17)	29.33 (17.08; 56.51)
Gansu	25.6	0.00 (0.00; 0.00)	0.10 (0.01; 0.88)
Guangdong	91.06	6.96 (6.62; 7.27)	6341.38 (6030.46; 6622.18)
Guangxi	37.62	5.52 (4.97; 6.06)	2077.94 (1867.98; 2280.70)
Guizhou	31.37	0.05 (0.02; 0.08)	15.33 (7.58; 26.67)
Hainan	6.68	0.46 (0.26; 0.71)	30.42 (17.05; 47.62)
Hebei	75.52	0.04 (0.02; 0.09)	32.20 (13.77; 67.43)
Heilongjiang	42.28	7.21 (5.95; 8.84)	3050.57 (2515.22; 3737.51)
Henan	84.3	0.11 (0.07; 0.16)	91.32 (60.46; 138.85)
Hubei	58.21	2.26 (1.90; 2.66)	1313.52 (1103.91; 1548.28)
Hunan	55.14	0.61 (0.51; 0.75)	333.64 (283.77; 410.98)
Jiangsu	74.3	0.19 (0.15; 0.24)	144.31 (113.62; 177.60)
Jiangxi	36.26	0.15 (0.09; 0.25)	54.72 (31.76; 91.40)
Jilin	29.19	2.07 (1.76; 2.43)	605.43 (514.64; 707.94)
Liaoning	43.09	0.32 (0.24; 0.45)	139.05 (104.56; 194.56)
Iner Mongolia	29.73	0.08 (0.05; 0.12)	24.09 (14.87; 37.06)
Ningxia	6.27	0.00 (0.00; 0.00)	0.03 (0.01; 0.18)
Qinghai	4.96	0.00 (0.00; 0.00)	0.00 (0.00; 0.01)
Shaanxi	34.26	0.01 (0.00; 0.03)	2.32 (0.26; 11.53)
Shandong	93.37	0.04 (0.03; 0.07)	41.06 (24.86; 65.24)
Shanghai	14.95	0.00 (0.00; 0.04)	0.36 (0.01; 5.34)
Shanxi	35.45	0.00 (0.00; 0.02)	1.38 (0.22; 5.39)
Sichuan	94.63	0.04 (0.02; 0.06)	34.48 (20.55; 57.33)
Tianjin	9.76	0.03 (0.02; 0.06)	2.97 (1.58; 5.41)
Xinjiang Uygur	24.97	0.01 (0.00; 0.06)	3.58 (1.10; 14.62)
Tibet	2.68	0.00 (0.00; 0.00)	0.00 (0.00; 0.00)
Yunnan	39.49	0.00 (0.00; 0.01)	1.31 (0.30; 4.24)
Zhejiang	45.35	0.02 (0.00; 0.08)	8.31 (1.04; 36.06)
Total	1257.89	1.18 (1.10; 1.25)	14844.08 (13796.77; 15767.23)

[a]Estimates based on gridded population of 2010; calculations based on the median and 95% Bayesian credible interval of the posterior distribution of the predictive risk from 2005 onwards.
[b]Bayesian credible interval.

Surveys pertaining to prevalence of *C. sinensis* in P.R. China were obtained through a systematic review in both Chinese and worldwide scientific databases to obtain published work from 2000 onwards. Additional data were provided by the NIPD, China CDC.

We estimated that 14.8 million (95% BCI: 13.8-15.8 million; 1.18%) people in P.R. China were infected with *C.* sinensis in 2010, which is slightly higher than the previous estimates of 12.5 million people for the year 2004, based on empirical analysis of data from a large survey

of clonorchiasis conducted from 2002-2004 in 27 endemic P/A/M. The mean error for the model validation was slightly smaller than zero, suggesting that our model might somewhat over-estimate the true prevalence of clonorchiasis. The overall raw prevalence of the observed data was 9.7%. This can be an over-estimation of the overall prevalence as many surveys from literatures are likely to be conducted in places with relatively high infection risk (preferential sampling). Our population-adjusted, model-based estimates is much lower (1.18%, 95% BCI: 1.10-1.25%) and it should reflect the actual situation because it takes into account the distribution of the population and of the disease risk across the country. Our model may overestimate the overall infection risk for Heilongjiang province due to the lack of observed data in the north part of the province.

We found an increase of infection risk of *C. sinensis* from 2005 onwards, which may be due to several reasons, such as higher consumption of raw fish, lack of self-protection awareness of food hygiene, low health education, and rapid growth of aquaculture (Han et al. 2013;Keiser & Utzinger 2005). Consumption of raw freshwater fish is related with *C. sinensis* infection risk, however (June et al. 2013;Phan et al. 2011), such information is unavailable for P.R. China.

Elevation is one of the most important predictors in our model. Different elevation levels correspond to different environmental/climatic conditions that can influence the distribution of intermediate host snails. Our results show a positive effect of NDVI with the prevalence of *C. sinensis*. We found that distance to nearest water bodies was significantly related to infection risk. Traditionally, areas adjacent to water bodies were reported to have a higher prevalence of *C. sinensis*, however, due to improvement of trade and transportation channels, this situation may be changing, which may explain our result showing a non-linear relationship between distance to nearest water bodies and infection risk (Keiser & Utzinger 2005;Qian et al. 2016). Furthermore, our analysis supports earlier observations, suggesting an association between land cover and infection risk (Keiser & Utzinger 2005;Petney et al. 2013).

Interestingly, the risk of infection with other neglected tropical diseases, such as soil-transmitted helminthiasis and schistosomiasis, has declined in P.R. China over the past 10-15 years due to socioeconomic development and long-term efforts on control and prevention (Yang et al. 2014). However, clonorchiasis, the major food-borne trematodiasis in P.R. China, shows an increasing temporal trend, which urges the Chinese government to pay more attention to this disease. Our risk estimates indicate several high endemic areas in P.R. China, where control strategies should be focused, such as the western areas of Pearl River delta region and the northeastern part of Heyuan city in Guangdong province, Nanning city, the northern part of Liuzhou city, and the northwestern part of Guilin city in Guangxi province, the western part of Qiyang city in Hunan province, Jiamusi city, the southern part of Daqing city, and the eastern part of Hegang city in Heilongjiang Province, and the northeastern part of Baicheng city and the northern part of Songyuan city in Jilin province. In China the city corresponds to an administrative division of level two.

The recommended treatment guidelines of clonorchiasis by WHO advocate praziquantel administration for all residents every year in high endemic areas (prevalence ≥20%) and for all residents every two years or individuals regularly eating raw fish every year in moderate endemic areas (prevalence <20%) (WHO 2013). As re-infection or super-infection is common in heavy endemic areas, repeated preventive chemotherapy is necessary to interrupt transmission (Choi et al. 2010). On the other hand, to maintain control sustainability, a comprehensive control strategy must be implemented, including IEC, preventive chemotherapy, and improvement of environment sanitation (Oh et al. 2014;Zhang et al. 2009). Through IEC, residents may conscientiously reduce or stop consumption of raw fish. Furthermore, by removing unimproved lavatories around fish ponds, the likelihood of fish becoming infected with cercariae declines (Wu et al. 2012). A successful example of comprehensive control strategies is Shangdong province, where clonorchiasis was endemic, but after rigorous implementation of comprehensive control programs for more than 10 years, the disease has been well controlled (Liu et al. 2010).

The Chinese Ministry of Health set a goal to halve the prevalence for clonorchiasis (compared to that observed in the second national survey in 2001-2004) in highly endemic areas by 2015 using integrated control measures (Ministry of Health 2006). In practice, control measures are carried out in endemic villages or counties with available survey data. However, large-scale control activities are lacking in most endemic provinces, as control plans are difficult to make when the epidemiology is only known at provincial level (Qian et al. 2013). Our high-resolution infection risk estimates provide important information for targeted control.

Approximately half of our survey data were aggregated at county-level. To avoid data sparsity, we included these data in the analysis using the centroids of the counties as survey locations. As the actual observed survey locations were not known, we assumed a uniform distribution of the infection risk within these counties. To assess the effect of this assumption on our estimates, we simulated data over a number of hypothetical survey locations within the counties and compared predictions based on approaches using the county aggregated data together with the data at individual georeferenced survey locations and using the data at individual georeferenced survey locations only (excluded the county aggregated data). The former approach gave substantially better disease risk prediction compared to the later one.

In conclusion, we present the first model-based, high-resolution risk estimates of *C. sinensis* infection risk in P.R. China, and identified areas of high priority for control. Our findings show an increased trend of infection risk from 2005 onwards, suggesting that the government should put more efforts on control activities of clonorchiasis in P.R. China.

Acknowledgements

Chapter 5   The spatial distribution of schistosomiasis and treatment needs in sub-Saharan Africa: a systematic review and geostatistical analysis

Yingsi Lai[1,2], Patricia Biedermann[1,2], Uwem F Ekpo[3], Amadou Garba[4], Els Mathieu[5], Nicholas Midzi[6], Pauline Mwinzi[7], Eliézer K N'Goran[8,9], Giovanna Raso[1,2], Rufin K Assaré[1,2], Moussa Sacko[10], Nadine Schur[1,2], Idrissa Talla[11], Louis-Albert Tchuem Tchuenté[12], Seydou Touré[13], Mirko S Winkler[1,2], Jürg Utzinger[1,2], Penelope Vounatsou[1,2]

[1]Department of Epidemiology and Public Health, Swiss Tropical and Public Health Institute, Basel, Switzerland

[2]University of Basel, Basel, Switzerland

[3]Department of Biological Sciences, Federal University of Agriculture, Abeokuta, Nigeria

[4]Réseau International Schistosomose, Environnement, Amenagement et Lutte, Niamey, Niger

[5]National Center of Infectious Diseases, Centers for Disease Control and Prevention, Atlanta, United Sates of America

[6]National Institute of Health Research, Causeway Harare, Zimbabwe

[7]Centre for Global Health Research, Kenya Medical Research Institute, Kisumu, Kenya

[8]Unité de Formation et de Recherche Biosciences, Université Félix Houphouët-Boigny, Abidjan, Côte d'Ivoire

[9]Centre Suisse de Recherches Scientifiques en Côte d'Ivoire, Abidjan, Côte d'Ivoire

[10]Institut National de Recherche en Santé Publique, Bamako, Mali

[11]Direction de la Lutte contre la Maladie, Ministère de la Santé, Dakar, Senegal

[12]Laboratory of Parasitology and Ecology, University of Yaoundé, and Center for Schistosomiasis and Parasitology, Yaoundé, Cameroon

[13]Programme National de Lutte Contre la Schistosomiase, Ministère de la Santé, Ouagadougou, Burkina Faso

Summary

Background: Schistosomiasis affects more than 200 million individuals, mostly in sub-Saharan Africa, but empirical estimates of the disease burden in this region are unavailable. We used geostatistical modelling to produce high-resolution risk estimates of infection with *Schistosoma spp* and of the number of doses of praziquantel treatment needed to prevent morbidity at different administrative levels in 44 countries.

Methods: We did a systematic review to identify surveys including schistosomiasis prevalence data in sub-Saharan Africa via PubMed, ISI Web of Science, and African Journals Online, from inception to May 2, 2014, with no restriction of language, survey date, or study design. We used Bayesian geostatistical meta-analysis and rigorous variable selection to predict infection risk over a grid of 1,155,818 pixels at 5×5 km, on the basis of environmental and socioeconomic predictors and to calculate the number of doses of praziquantel needed for prevention of morbidity.

Findings: The literature search identified *Schistosoma haematobium* and *Schistosoma mansoni* surveys done in, respectively, 9,318 and 9,140 unique locations. Infection risk decreased from 2000 onwards, yet estimates suggest that 163 million (95% Bayesian credible interval (BCI) 155 million to 172 million; 18.5%, 17.6–19.5) of the sub-Saharan African population was infected in 2012. Mozambique had the highest prevalence of schistosomiasis in school-aged children (52.8%, 95% BCI 48.7–57.8). Low-risk countries (prevalence among school-aged children lower than 10%) included Burundi, Equatorial Guinea, Eritrea, and Rwanda. The numbers of doses of praziquantel needed per year were estimated to be 123 million (95% BCI 121 million to 125 million) for school-aged children and 247 million (239 million to 256 million) for the entire population.

Interpretation: Our results will inform policy makers about the number of treatments needed at different levels and will guide the spatial targeting of schistosomiasis control interventions.

5.1 Introduction

Schistosomiasis is a chronic disease that affects around 240 million people worldwide, leading to a burden of 3.3 million disability-adjusted life-years (DALYs) (Murray et al. 2012;WHO 2010a). More than 90% of all cases occur in Africa (Stothard et al. 2009). The two main schistosome species that infect human beings in Africa are *Schistosoma mansoni*, which causes intestinal and hepatic schistosomiasis, and *Schistosoma haematobium*, which causes urogenital schistosomiasis (Colley et al. 2014). *Schistosoma intercalatum* and *Schistosoma guineensis* have limited distribution and are rarely reported (Chu et al. 2012;Tchuem Tchuenté et al. 2003b).

Successful schistosomiasis control programmes have been implemented in north Africa (i.e., Egypt, Libya, Morocco, and Tunisia) (Amarir et al. 2011;Rollinson et al. 2013;WHO 2009;WHO/EMRO 2007), but in many countries of sub-Saharan Africa schistosomiasis is still far from being under control, let alone eliminated. Although schistosomiasis-endemic countries are encouraged to treat at least 75% of school-aged children at risk of morbidity (Savioli et al. 2009), less than 14% of people needing preventive chemotherapy were treated in 2012 (WHO 2014). Control programmes emphasising preventive chemotherapy supported by the Schistosomiasis Control Initiative, have been set up in several countries, including Burkina Faso, Mali, Niger, Tanzania, Uganda, and Zambia, where prevalence and intensity of infection with *Schistosoma spp* have been reduced substantially (Fenwick et al. 2009;Garba et al. 2009). Other countries (e.g., Angola, Benin, Cameroon, Central African Republic, Madagascar, and Senegal) have launched large-scale control programmes, facilitated by donated praziquantel (WHO 2014), but, for various reasons, including ineffective coverage of drug distribution and rapid reinfection owing to interruption of preventive chemotherapy, the reduction in disease risk has been lower than expected (Gray et al. 2010). Whenever resources allow, integrated control activities, incorporating preventive therapy with praziquantel, transmission reduction through environmental modification, health education and promotion, and improved water and sanitation, are being recommended (Gray et al. 2010;Utzinger et al. 2003;WHO 2010b).

Risk maps depicting the geographical distribution of schistosomiasis assist disease control by helping to focus control interventions in the areas of highest risk. Bayesian geostatistical modeling enables prediction of disease risk in areas without observed data by relating survey data to potential predictors (Chammartin et al. 2013c). Such models have been used for small-scale and large-scale schistosomiasis risk profiling (Clements et al. 2006;Clements et al. 2008a;Clements et al. 2010a;Clements et al. 2008b;Ekpo et al. 2013;Hodges et al. 2012;Koroma et al. 2010;Raso et al. 2005;Soares Magalhães et al. 2011). Schur and colleagues (Schur et al. 2011a;Schur et al. 2011b) presented prevalence maps for *Schistosoma spp* infection in east and west Africa in 2011. Their estimates, however, were based mainly on data recorded before large-scale preventive chemotherapy control programmes had started and, therefore, do not reflect the current situation. Additionally, high-resolution estimates of

the disease in central and southern Africa are unavailable. We aimed to systematically review survey data and produce high-resolution risk estimates of *Schistosoma spp* infection and to estimate the number of doses of praziquantel that would be needed for prevention of morbidity at country and district levels for the whole sub-Saharan Africa region.

5.2 Methods

5.2.1 Search strategy and selection criteria

We did a systematic review following the PRISMA guidelines (Moher et al. 2009). We searched for relevant publications pertaining to prevalence of *Schistosoma spp* infection in sub-Saharan Africa, in PubMed, ISI Web of Science, and African Journals Online, from inception to May 2, 2014. We applied the search string "schisto* (OR mansoni, OR bilhar*, OR haema*) AND sub-Saharan Africa (OR Angola, OR Benin, OR Botswana, OR Burkina Faso, OR Burundi, OR Cameroon, OR Central African Republic, OR Chad, OR Congo*, OR Côte d'Ivoire, OR Cote d'Ivoire, OR Ivory Coast, OR Djibouti, OR Eritrea, OR Ethiopia, OR Gabon, OR Gambia, OR Ghana, OR Guinea*, OR Kenya, OR Lesotho, OR Liberia, OR Madagascar, OR Malawi, OR Mali, OR Mauritania, OR Mozambique, OR Namibia, OR Niger, OR Nigeria, OR Rwanda, OR Senegal, OR Sierra Leone, OR Somalia, OR South Africa, OR Sudan, OR Swaziland, OR Tanzania, OR Togo, OR Tunisia, OR Uganda, OR Zambia, OR Zimbabwe)". Government reports and other grey literature (eg, PhD theses, working papers from research groups, or unpublished research reports obtained through personal communication) were also considered. We set no parameters for language, survey date, or study design.

We initially screened titles and abstracts to identify potentially relevant articles. We excluded case reports, in-vitro studies, non-human studies, or those that did not report on schistosomiasis. Quality control for each country was done by rechecking 15% of randomly selected papers deemed irrelevant. If any misclassifications were identified, the selection for the whole country was rechecked. Full-text reports for potentially relevant papers were obtained and screened. At this stage we additionally excluded studies without prevalence data, those done in specific groups of patients (e.g., with specified diseases) or clearly defined population groups (i.e., travellers, military personnel, expatriates, nomads, and displaced or migrating populations) not representative of the general population, studies that used either indirect diagnostic techniques (because such tests cannot distinguish between active and cleared infection) or direct stool smear (because of low diagnostic sensitivity), reports of case-control studies, clinical trials, drug efficacy, intervention studies (except for baseline data or control groups), studies that reported on species other than *S. haematobium* and *S. mansoni*, and surveys done before 1950, that were not community based or school based, or were done in places where population deworming had been done within 1 year, or study findings

reported aggregated within regions (i.e., administrative division of level one). We reviewed the reference lists of full-text articles for further possible data sources. Duplicates were removed. If important information was missing (e.g., survey year, location names or coordinates, numbers of individuals assessed and positive, etc) or if surveys were aggregated, we contacted the authors for clarification.

5.2.2 Ethics committee approval

This study was based on surveys of *Schistosoma spp* infection derived from the peer-reviewed literature and other sources that included statements of ethics approval in the original reports. Because the data are aggregated and do not contain identifiable individual-level or household-level information, no specific ethics approval was needed for this study.

5.2.3 Data extraction

From the screened references we obtained information, such as number of people assessed, number of positive cases (disease prevalence), age group, diagnostic approach, year of survey and name of study setting or coordinates of survey location. Data were georeferenced and entered into the Global Neglected Tropical Diseases database (Hürlimann et al. 2011) for 44 countries: Angola, Benin, Botswana, Burkina Faso, Burundi, Cameroon, Central African Republic, Chad, Congo, Côte d'Ivoire, Democratic Republic of the Congo, Djibouti, Equatorial Guinea, Eritrea, Ethiopia, Gabon, Ghana, Guinea, Guinea-Bissau, Kenya, Lesotho, Liberia, Madagascar, Malawi, Mali, Mauritania, Mozambique, Namibia, Niger, Nigeria, Rwanda, Senegal, Sierra Leone, Somalia, South Africa, South Sudan, Sudan, Swaziland, Tanzania, The Gambia, Togo, Uganda, Zambia, and Zimbabwe. All information and coordinates of locations were checked and approved for each record by personnel working at the Global Neglected Tropical Diseases database (http://www.gntd.org) and who had not been involved in data extraction for this study.

5.2.4 Environmental, socioeconomic, and population data

We obtained environmental and demographic data from different readily accessible remote sensing sources (Appendix). We used socioeconomic predictors, such as the Human Influence Index, urban extents, gross domestic product, and infant mortality rates, which were obtained from the Socioeconomic Data and Applications Center (http://sedac.ciesin.columbia.edu). Additionally, we included the proportion of households with improved sanitation and the proportion of households with improved drinking water sources, that were estimated from predictions based on geostatistical models fitted with data collected during household surveys done by Demographic and Health Surveys, Multiple Cluster Indicator Surveys, World Health Surveys, and the Living Standards Measurement Study (Appendix). The 2012 population was projected from that of 2010 by taking into account the country-specific annual population growth rate obtained from the United Nations World Population Prospects (United Nations 2016).

5.2.5 Statistical analysis

We grouped survey years into three categories—before 1980, 1980–99, and from 2000 onwards—to assess temporal trends. Survey type (school-based or community-based) was included in geostatistical model as a covariate. Plots of disease prevalence with each continuous variable indicated non-linear relations that could best be captured by converting continuous variables to categorical variables based on tertiles. Mean prevalence for each country was calculated as the mean of prevalence values reported in observed surveys identified within the country.

We did a meta-analysis by developing Bayesian geostatistical logistic regression models that related survey data on *Schistosoma spp* infection with potential environmental and socioeconomic predictors. Spatially structured random effects were introduced into the models. Spatial correlation was assumed to follow a Matérn function—i.e., decreasing correlation with increasing distance between two locations. To overcome computational burden due to large spatial correlation matrices, we fitted the models with the integrated nested Laplace approximations (INLA) package in R (version 3.0.1) (Lindgren et al. 2011;Rue et al. 2009). Further details of our model fitting approach are provided in the Appendix. Separate geostatistical models were fitted for the mainland of sub-Saharan Africa and for Madagascar because correlation with distance does not have the same meaning between locations separated by land and by sea.

We used Bayesian variable selection to identify the best set of predictors (Scheipl et al. 2012). Furthermore, we grouped the continuous variables with Pearson's correlation coefficient values greater than 0.8 and selected a maximum of one predictor from each of the highly correlated variable groups. We also identified the best functional form (i.e., continuous or categorical) for each continuous covariate (Appendix). Prediction of risk of infection with *S. haematobium* or *S. mansoni* was done with the INLA package over a grid of 1,155,818 pixels across sub-Saharan Africa at 5×5 km spatial resolution. We estimated population-adjusted prevalence for school-aged children (age 5–14 years) and communities (no age restriction) by overlaying the pixel-based disease risk predictions on the population to obtain the number of infected individuals, and recalculated the prevalence for different administrative levels. We assumed independence between the two *Schistosoma spp*. To predict the combined prevalence, we used the formula $p_s = p_h + p_m - p_h \times p_m$, where p_s is the predicted prevalence of schistosomiasis, p_h the predicted prevalence of *S. haematobium*, and p_m the predicted prevalence of *S. mansoni*. To assess the effect of co-infection, we assumed that the probability of being infected with one species (A) given that the other species (A) is present, increased by $a\%$, that is $P(A|B) = (1 + a\%)P(A)$, where P is probability. Thus, with use of probability laws, $P(A \cup B) = P(A) + P(B) - P(B)P(A|B)$, we estimated the probabilities of being infected by either species for $a = 25$, $a = 50$, and $a = 75$.

A random subset of 80% of survey locations was used to create the training set for model fit, and the remaining 20% of locations were used as the test set for validation. The mean error and proportion of observations included in Bayesian credible intervals (BCI) for various probability coverages of predictions at the test locations were calculated to assess model performance. We used the calculation $ME = 1/N \sum_{i=1}(\pi_i - \hat{\pi}_i)$, where π indicates the observed prevalence, $\hat{\pi}$ the median of the posterior distribution of the predictive prevalence at test location i, and N the total number of test locations. Thus, positive values of mean error suggest that the model under-estimates the observed prevalence.

We estimated the numbers of doses of praziquantel needed for prevention of morbidity at pixel level, according to the approach of Schur and colleagues (Schur, Vounatsou, & Utzinger 2012) and WHO schistosomiasis control guidelines (WHO 2002a;WHO 2006). The numbers of school-aged children living in areas with low, moderate, and high prevalence were calculated per pixel and summarised by country.

5.2.6 Role of the funding sources

The funders of the study had no role in study design, data collection, data analysis, data interpretation, or writing of the report. Y-SL and PV (corresponding author) had full access to all the data in the study and had final responsibility for the decision to submit for publication.

5.3 Results

Of 9,319 references identified in published articles and 78 from other sources, 1,274 were suitable for inclusion in the meta-analysis (Figure 5.1). The 78 references from other sources contributed 29.4% of surveys used for *S. haematobium* and 29.1% for *S. mansoni*. Data were available for *S. haematobium* from all countries except Djibouti and for both *Schistosoma spp* from all countries except from Lesotho. We identified none or very few surveys done after 2000 in Botswana, Central African Republic, Chad, Congo, Equatorial Guinea, Guinea-Bissau, Liberia, Madagascar, Namibia, Somalia, and The Gambia (Appendix). Diagnostic methods used and incomplete information that could affect assessment of the observed prevalence, are provided in the Appendix. 8,052 (72.8%) of the surveys of *S. mansoni* used the WHO-recommended Kato-Katz technique for diagnosis; 973 (8.8%) did not specify the diagnostic approach taken. 4,428 (37.6%) and 3,486 (29.6%) of surveys of *S. haematobium* used, respectively, urine filtration and reagent strips for diagnosis; 1,401 (11.9%) surveys did not specify the diagnostic technique. The mean prevalence values calculated from survey data were 24.7% for *S. haematobium* and 14.8% for *S. mansoni*.

We considered 35 variables for Bayesian modeling (Appendix). In the final geostatistical logistic regression model, the infection risks for the two *Schistosoma spp* were lower in the entire community than that in school-aged children from 1980 onwards, and decreased substantially from 2000 onwards (Table 5.1). The prevalence of *S. haematobium* in the
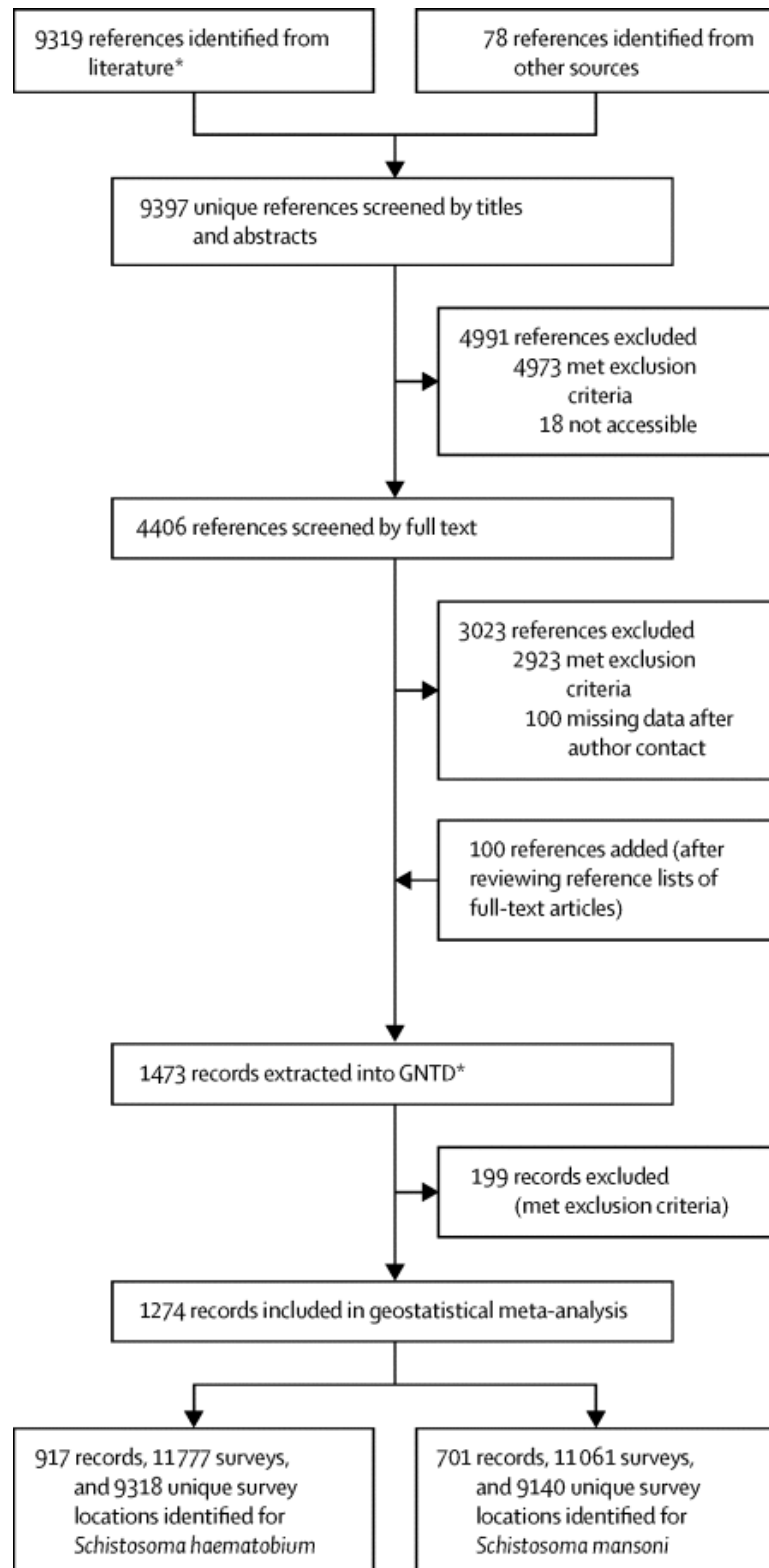
Figure 5.1: Data search and selection. GNTD=Global Neglected Tropical Diseases Database. *After duplicates removed.

school-aged population was lower in the period from 2000 onwards than in the two preceding periods. Prevalence of *S. mansoni* in school-aged children increased in the 1980s and 1990s, after which it declined. Distance to the nearest body of fresh water had a negative effect on risk of infection with either species (Table 5.1). Positive associations were identified between

the prevalence of *S. haematobium* and land surface temperature in the daytime and at night and normalised difference vegetation index, whereas a negative association was found with precipitation during the coldest quarter of the year (Table 5.1). For *S. mansoni* prevalence was lower in areas with barren land and those with large variation in precipitation. Positive associations were found with average rainfall during the driest month, areas with wet land cover, mean temperature of the driest quarter, soil moisture, and areas with high infant mortality, whereas negative associations were found with land surface temperature in the daytime and pH measured in water (Table 5.1). The parameter estimates for Madagascar are shown in the Appendix.

Table 5.1: Posterior summaries (median and 95% BCI) of the geostatistical model parameters[†].

S. haematobium	Estimate
Period (<1980)[#]	
1980-1999	-0.14 (-0.20; -0.08)[*]
≥2000	-0.60 (-0.66; -0.54)[*]
Survey type (school-based)[#]	
Community-based	-0.49 (-0.54; -0.43)[*]
Period×survey type (<1980×school-based)[#]	
1980-1999×community	-0.12 (-0.19; -0.06)[*]
≥2000×community	0.15 (0.08; 0.23)[*]
Mean diurnal temperature range	0.17 (-0.08; 0.41)
Annual precipitation (≤650 mm)[#]	
650-1150	0.02 (-0.33; 0.37)
>1150	-0.07 (-0.50; 0.36)
Precipitation of warmest quarter	0.14 (-0.02; 0.31)
Precipitation of driest quarter (≤0 mm)[#]	
0-30	0.15 (-0.09; 0.39)
>30	-0.35 (-0.75; 0.05)
Precipitation of coldest quarter (≤3 mm)[#]	
3-200	-0.22 (-0.49; 0.04)
>200	-0.67 (-1.04; -0.30)[*]
Land surface temperature at night	0.38 (0.25; 0.51)[*]
Land surface temperature at day	0.21 (0.06; 0.36)[*]
Normalized differenced vegetation index	0.20 (0.08; 0.32)[*]
Gross domestic product (≤1 million US$ per 0.25° grid cell)[#]	
1-3	0.10 (-0.06; 0.25)
>3	-0.12 (-0.35; 0.11)
Infant mortality rate (≤830 per 10,000 live births)[#]	
≤830	0.00
830-1150	-0.12 (-0.40; 0.16)
>1150	0.11 (-0.27; 0.48)
Soil moisture (≤0.5 mm)[#]	
0.5-30	0.15 (-0.16; 0.46)
>30	-0.17 (-0.61; 0.27)
pH measured in water (≤6)[#]	
6-7	0.14 (-0.08; 0.35)
>7	-0.28 (-0.80; 0.24)
Distance to water bodies (≤6 km)[#]	
6-25	-0.20 (-0.31; -0.08)[*]
>25	-0.33 (-0.51; -0.15)[*]
Improved sanitation (≤7.5%)[#]	
7.5-25	0.11 (-0.01; 0.24)
>25	0.02 (-0.15; 0.18)
Climatic zone (Equatorial)[#]	
Arid	-0.22 (-0.59; 0.15)
Warm	0.15 (-0.32; 0.62)

(Continues in next page)

(Continued from previous page)

S. haematobium	Estimate
Land cover (grass)[#]	
Forest	-0.16 (-0.37; 0.06)
Shrub	-0.01 (-0.20; 0.18)
Crop	-0.04 (-0.23; 0.16)
Urban	-0.21 (-0.45; 0.04)
Wet	0.16 (-0.15; 0.47)
Barren	-0.37 (-0.82; 0.07)
Range (km)	142.23 (120.67; 161.44)
Spatial variance (σ^2_{sp})	4.24 (3.64; 4.84)
Non-spatial variance (σ^2_{nonsp})	1.85 (1.75; 1.94)
S. mansoni	Estimate
Period (<1980)[#]	
1980-1999	0.76 (0.68; 0.83)[*]
≥2000	0.21 (0.13; 0.29)[*]
Survey type (school-based)	
Community-based	0.00 (-0.07; 0.06)
Period×survey type (<1980×school-based)[#]	
1980-1999×community	-0.75 (-0.83; -0.68)[*]
≥2000×community	-0.99 (-1.08; -0.90)[*]
Mean temperature of driest quarter	0.53 (0.29; 0.76)[*]
Temperature annual range	0.75 (0.35; 1.14)[*]
Precipitation of driest month (≤0 mm)[#]	
0-12	0.93 (0.47; 1.39)[*]
>12	0.53 (-0.05; 1.12)
Precipitation of coldest quarter (≤15 mm)[#]	
15-285	0.04 (-0.38; 0.47)
>285	-0.23 (-0.70; 0.23)
Precipitation seasonality	-0.60 (-1.04; -0.17)[*]
Precipitation of warmest quarter	-0.10 (-0.30; 0.10)
pH measured in water	-0.32 (-0.61; -0.03)[*]
Distance to water bodies	-0.32 (-0.46; -0.18)[*]
Elevation (≤300 m)	
300-900	0.23 (-0.06; 0.53)
>900	-0.11 (-0.61; 0.38)
Infant mortality rate (≤800 per 10,000 live births)[#]	
800-1120	0.61 (0.26; 0.95)[*]
>1120	0.63 (0.16; 1.10)[*]
Land surface temperature at day (≤28 ℃)[#]	
28-34	-0.28 (-0.48; -0.08)[*]
>34	-0.33 (-0.65; -0.02)[*]
Soil moisture (≤0.5 mm)[#]	
0.5-50	0.78 (0.26; 1.29)[*]
>50	0.76 (0.05; 1.47)[*]
Improved water sources (≤6 %)[#]	
6-25	0.06 (-0.12; 0.23)
>25	0.18 (-0.03; 0.38)
Land cover (grass)[#]	
Forest	-0.11 (-0.37; 0.15)
Shrub	-0.03 (-0.34; 0.27)
Crop	0.11 (-0.10; 0.32)
Urban	-0.20 (-0.52; 0.12)
Wet	0.50 (0.25; 0.75)[*]
Barren	-1.44 (-2.54; -0.38)[*]
Range (km)	130.22 (114.21; 147.03)
Spatial variance (σ^2_{sp})	7.70 (6.64; 8.85)
Non-spatial variance (σ^2_{nonsp})	1.79 (1.68; 1.89)

[†]Results based on surveys in mainland sub-Saharan Africa;
[*]important effect based on 95% Bayesian credible interval (BCI);
[#]in brackets, baseline values are reported.

Model validation suggested that the Bayesian geostatistical logistic regression models were able to correctly estimate 76.8% and 71.9% of locations within the 95% BCI for

S. mansoni and *S. haematobium*, respectively (Appendix). The mean error for *S. mansoni* was 2.1% and for *S. haematobium* was 2.7%, which suggest that our models might have slightly underestimated the risk for both species.

The model-based maps of predicted prevalence are shown in Figures 5.2–5.4. In 2012, the population-adjusted prevalence values for *S. haematobium* and *S. mansoni* in school-aged children in sub-Saharan Africa were 17.4% (95% BCI 16.4–18.5) and 8.0% (7.4–8.6), (Table 5.2), and 14.6% (13.7–15.5) and 4.6% (4.2–5.1), respectively, in the entire population (Appendix). 24.0% (95% BCI 23.0–25.1) school-aged children and 18.5% (17.6–19.5) of the entire population were infected with either *Schistosoma species*. Mozambique had the highest predicted prevalence of schistosomiasis among school-aged children (Table 5.2). The lowest prevalence values (less than 10%) in school-aged children were found in Burundi, Equatorial Guinea, Eritrea, Lesotho, and Rwanda. Additionally, the population-adjusted prevalence for 1980–99 was 29.6% (95% BCI 28.5–30.7) in school-aged children and 22.0% (21.1–23.0) in the entire population, which suggests notable reductions over time.

In our estimation of the effect of co-infection, we calculated that the probability of being infected by either species in school-aged children was 23.6% (95% BCI 22.7–24.7), 23.3% (22.4–24.4), and 22.9% (22.0–24.0), for $a = 25$, $a = 50$, and $a = 75$, respectively. The independent assumption, therefore, has very little effect on the overall infection risk. At the pixel level, we were able to classify the numbers of school-aged children living in low (less than 10%), moderate (10–50%), and high (more than 50%) prevalence areas, and the number of doses of praziquantel needed for prevention of morbidity (Table 5.3, Appendix). In sub-Saharan Africa, 44.1 million (95% BCI 41.4 million to 47.3 million) school-aged children were living in areas with the highest risk of schistosomiasis. The numbers of doses of praziquentel needed for prevention of morbidity in sub-Saharan Africa were estimated to be 122.8 million for school-aged children and 247.5 million for the entire population, which are close to WHO estimates (Table 5.3, Appendix).

5.4 Discussion

We were able to calculate model-based, high-resolution schistosomiasis risk estimates for sub-Saharan Africa, overall and by *Schistosoma spp*, by doing a systematic review, compiling georeferenced survey data, and using advanced Bayesian geostatistical modeling based on multiple environmental and socioeconomic predictors. This method and our findings will be important for prioritising and targeting interventions for morbidity control and, ultimately, elimination (WHO 2002a;WHO 2010b).

Table 5.2: Population-adjusted prevalence and number of school-aged children infected with *Schistosoma spp* in 2012*.

Country	Population	S. haematobium		S. mansoni		Schistosomiasis	
		Prevalence	Number infected	Prevalence	Number infected	Prevalence	Number infected
Angola	6441	24.8 (15.1–39.3)	1595 (972–2531)	4.2 (1.9–15.0)	273 (123–964)	28.5 (18.4–43.7)	1835 (1187–2815)
Benin	2431	32.3 (24.1–40.7)	785 (587–990)	5.0 (2.4–10.7)	122 (59–260)	35.9 (26.9–43.9)	872 (653–1067)
Botswana	437	11.6 (3.9–25.5)	51 (17–111)	5.7 (2.8–12.5)	25 (12–55)	17.3 (8.1–32.0)	75 (35–140)
Burkina Faso	4727	26.4 (23.1–30.4)	1248 (1091–1435)	1.7 (1.1–2.8)	80 (53–132)	27.9 (24.4–31.8)	1318 (1153–1503)
Burundi	2650	0.1 (0.0–0.6)	3 (1–16)	6.6 (4.5–10.2)	174 (121–271)	6.7 (4.7–10.3)	178 (124–273)
Cameroon	5952	8.8 (7.3–10.4)	522 (436–618)	5.9 (4.5–7.9)	350 (271–468)	13.8 (12.0–16.1)	820 (715–958)
Central African Republic	1115	9.7 (6.6–14.0)	109 (73–156)	14.4 (6.2–30.4)	161 (70–339)	22.6 (14.7–38.8)	252 (164–432)
Chad	3637	26.4 (17.6–36.3)	959 (641–1319)	7.3 (3.7–15.1)	264 (136–550)	31.8 (22.7–42.1)	1158 (826–1530)
Congo	856	12.3 (7.7–19.9)	105 (66–170)	6.8 (2.0–17.5)	59 (17–150)	18.8 (12.4–28.5)	161 (106–244)
Côte d'Ivoire	5032	9.0 (7.2–11.2)	451 (361–565)	9.1 (7.5–11.2)	456 (377–564)	17.4 (15.0–20.2)	875 (756–1015)
DR Congo	17758	13.4 (9.6–19.0)	2376 (1697–3381)	20.5 (16.2–25.7)	3647 (2885–4571)	31.4 (26.7–37.8)	5584 (4739–6719)
Djibouti	269	8.6 (1.0–60.3)	23 (3–162)	18.6 (2.1–74.9)	50 (6–201)	33.2 (6.2–81.6)	89 (17–219)
Equatorial Guinea	179	1.4 (0.3–5.8)	3 (1–10)	2.6 (0.3–14.0)	5 (1–25)	4.5 (1.2–15.7)	8 (2–28)
Eritrea	1608	2.2 (0.8–5.0)	35 (13–81)	6.8 (3.9–10.7)	109 (63–172)	8.8 (5.6–13.4)	142 (90–215)
Ethiopia	26497	8.3 (4.6–14.1)	2206 (1227–3724)	8.9 (7.4–10.8)	2358 (1970–2869)	16.5 (12.7–21.6)	4383 (3371–5732)
Gabon	368	15.3 (6.3–36.6)	56 (23–135)	4.0 (0.7–15.0)	15 (3–55)	20.2 (9.3–43.3)	74 (34–159)
Ghana	5634	22.3 (19.4–26.5)	1256 (1095–1491)	1.3 (0.7–2.3)	71 (37–130)	23.3 (20.5–27.3)	1312 (1152–1537)
Guinea	2877	12.3 (8.3–17.1)	354 (239–491)	14.6 (9.9–20.5)	421 (286–590)	24.4 (19.0–30.7)	703 (545–884)
Guinea-Bissau	369	23.7 (14.7–38.2)	88 (54–141)	1.5 (0.2–11.6)	6 (1–43)	25.7 (15.7–40.0)	95 (58–148)
Kenya	11110	10.2 (7.0–14.6)	1131 (773–1624)	5.9 (4.8–7.5)	661 (531–831)	15.5 (12.3–19.8)	1721 (1367–2200)
Lesotho	523	4.4 (0.6–23.8)	23 (3–125)	2.3 (0.1–19.3)	12 (0–101)	8.3 (1.5–32.0)	43 (8–167)
Liberia	1004	16.7 (9.7–26.5)	167 (98–266)	9.0 (5.4–16.0)	90 (55–160)	24.5 (16.4–35.9)	246 (165–360)
Madagascar	5841	6.6 (5.5–14.3)	387 (323–833)	8.1 (6.9–9.5)	474 (403–558)	14.6 (12.8–21.4)	851 (749–1252)
Malawi	4240	24.1 (19.9–28.8)	1021 (845–1221)	7.3 (4.9–10.5)	308 (208–446)	29.8 (25.5–35.4)	1265 (1081–1500)
Mali	4084	29.4 (26.6–32.5)	1201 (1084–1328)	7.8 (6.2–9.5)	318 (252–390)	34.2 (31.3–37.1)	1399 (1279–1517)
Mauritania	969	24.6 (19.0–34.0)	238 (184–330)	1.8 (0.6–6.7)	18 (6–65)	26.5 (20.2–35.9)	256 (196–348)
Mozambique	6722	47.1 (42.4–51.9)	3167 (2853–3490)	10.5 (7.5–14.7)	708 (506–991)	52.8 (48.7–57.8)	3552 (3272–3885)
Namibia	584	7.6 (4.4–15.4)	44 (26–90)	2.9 (1.3–8.2)	17 (8–48)	10.5 (6.4–18.4)	61 (37–107)
Niger	5109	18.9 (16.2–21.7)	968 (828–1108)	0.4 (0.2–0.8)	20 (12–43)	19.3 (16.5–22.1)	984 (842–1127)
Nigeria	43088	22.0 (19.9–24.6)	9470 (8566–10592)	4.4 (3.2–6.0)	1879 (1395–2587)	25.2 (23.0–27.8)	10846 (9918–11998)
Rwanda	3301	0.1 (0.0–0.2)	2 (1–6)	3.8 (2.2–5.8)	125 (74–192)	3.9 (2.3–5.9)	127 (76–194)
Senegal	3490	18.3 (16.2–21.3)	639 (566–743)	2.2 (1.5–4.1)	78 (52–143)	20.3 (17.9–23.5)	708 (626–820)
Sierra Leone	1708	21.6 (15.9–28.4)	369 (271–486)	18.4 (13.7–24.7)	314 (234–422)	35.0 (28.9–42.1)	597 (493–719)
Somalia	2657	22.8 (16.9–30.0)	607 (450–797)	2.3 (0.7–6.7)	62 (19–177)	24.9 (18.8–31.8)	661 (499–846)
South Africa	9689	16.1 (12.2–20.7)	1561 (1181–2006)	8.6 (5.3–14.0)	837 (511–1353)	22.6 (17.9–28.9)	2186 (1734–2798)
South Sudan	2813	18.5 (10.9–29.4)	521 (308–826)	14.4 (9.2–22.5)	405 (260–633)	30.6 (22.3–41.7)	860 (627–1174)
Sudan	9456	20.7 (16.1–24.9)	1955 (1520–2358)	12.1 (9.3–16.0)	1141 (876–1511)	31.0 (25.7–35.2)	2935 (2435–3332)
Swaziland	326	18.4 (9.3–35.3)	60 (30–115)	11.2 (2.8–39.6)	37 (9–129)	29.1 (15.1–52.7)	95 (49–172)
Tanzania	12953	19.9 (17.0–23.3)	2572 (2207–3017)	7.1 (5.2–9.7)	923 (668–1263)	25.9 (22.8–29.7)	3349 (2955–3848)
The Gambia	565	16.8 (10.4–32.3)	95 (59–183)	7.5 (1.9–23..8)	42 (11–135)	24.5 (15.1–43.0)	138 (85–243)
Togo	1457	20.6 (18.8–22.6)	300 (274–330)	2.2 (1.8–2.8)	32 (26–41)	22.4 (20.7–24.5)	326 (301–357)
Uganda	10223	4.3 (1.7–11.1)	436 (175–1135)	10.3 (8.8–12.1)	1057 (901–1238)	14.2 (11.3–20.5)	1455 (1157–2096)
Zambia	4059	23.1 (19.2–28.0)	936 (781–1138)	5.0 (2.9–8.9)	204 (119–360)	27.2 (22.5–32.2)	1105 (912–1308)
Zimbabwe	3816	25.2 (22.4–28.2)	960 (856–1075)	7.6 (6.4–9.1)	290 (246–347)	30.5 (27.9–33.5)	1163 (1066–1280)
Total†	238626	17.4 (16.4–18.5)	41418 (39153–44107)	8.0 (7.4–8.6)	19029 (17671–20612)	24.0 (23.0–25.1)	57236 (54881–59931)

Numbers are thousands. DR Congo=Democratic Republic of the Congo. *Estimates are based on gridded population estimates of children aged 5–14 years in 2012; calculations are based on the median and 95% Bayesian credible interval of the posterior predictive distribution of the risk from 2000 onwards. †District level estimates for each country can be obtained from the authors upon request.
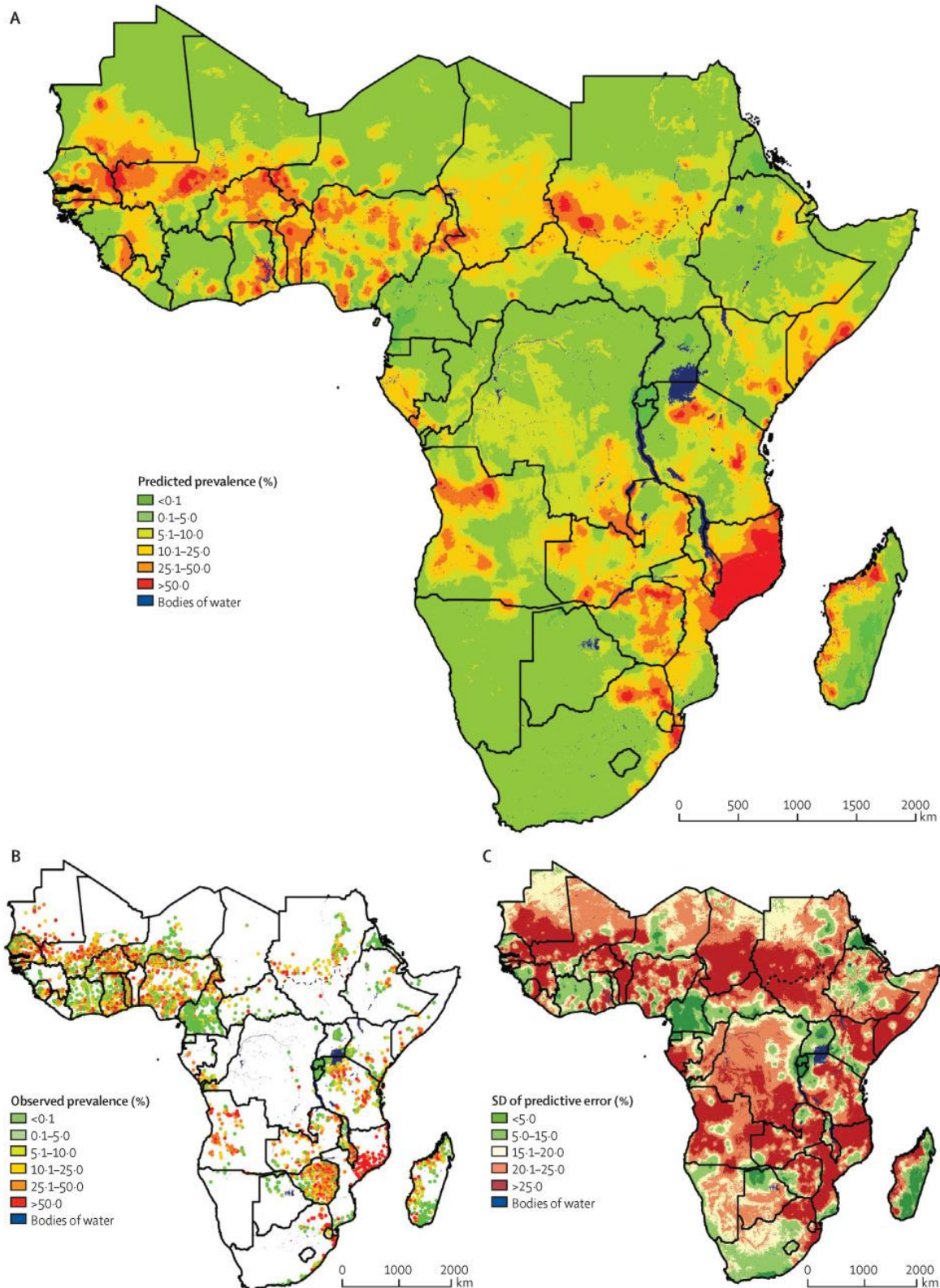
Figure 5.2: Prevalence of *Schistosoma haematobium* infection in school-aged children in sub-Saharan Africa, from 2000 onwards (A) Predictive prevalence based on the median of the posterior predictive distribution. (B) Observed prevalence over the whole study period. (C) Prediction uncertainty based on the SD of the posterior predictive distribution.
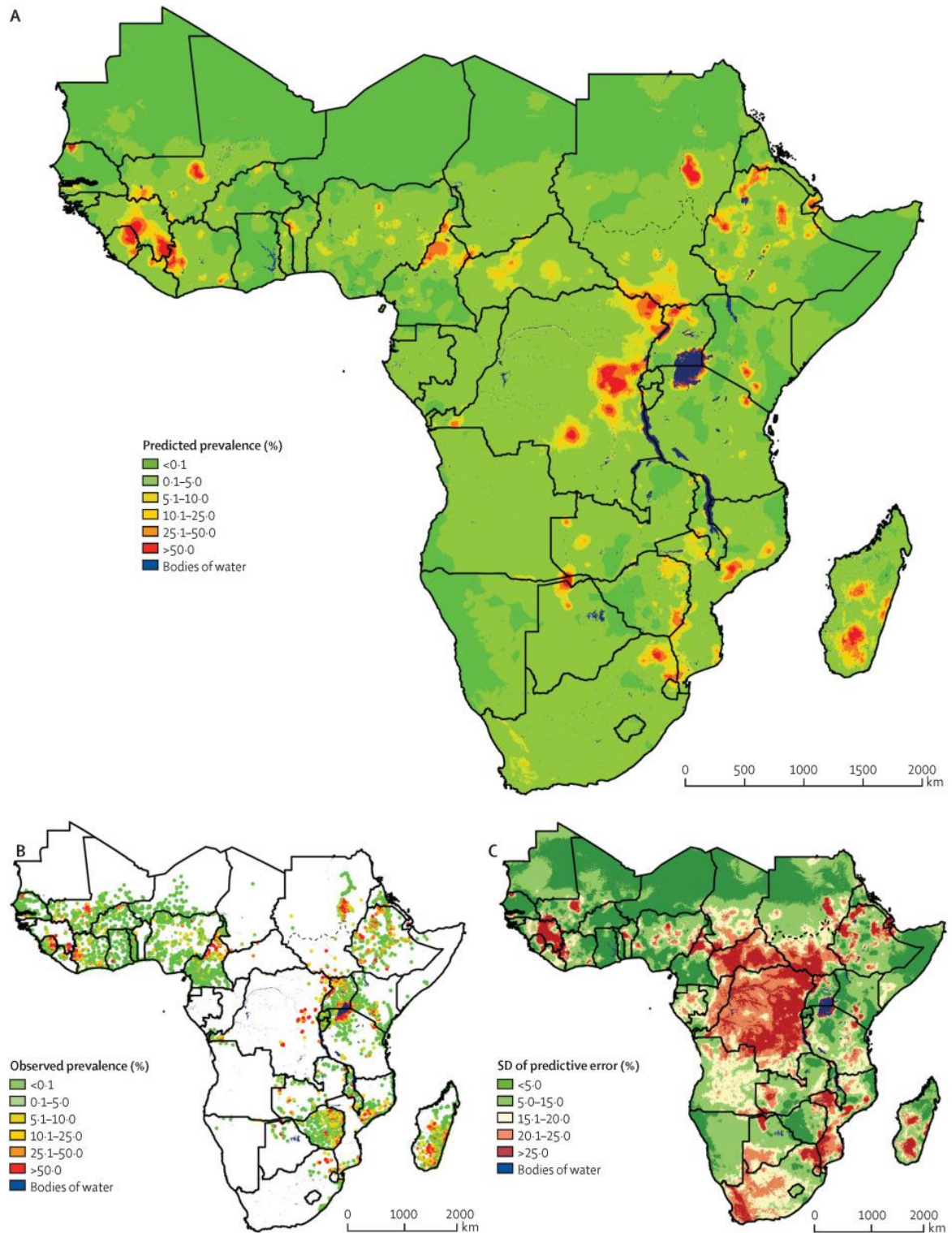
Figure 5.3: Prevalence of *Schistosoma mansoni* infection in school-aged children in sub-Saharan Africa, from 2000 onwards (A) Predictive prevalence based on the median of the posterior predictive distribution. (B) Observed prevalence over the whole study period. (C) Prediction uncertainty based on the SD of the posterior predictive distribution.
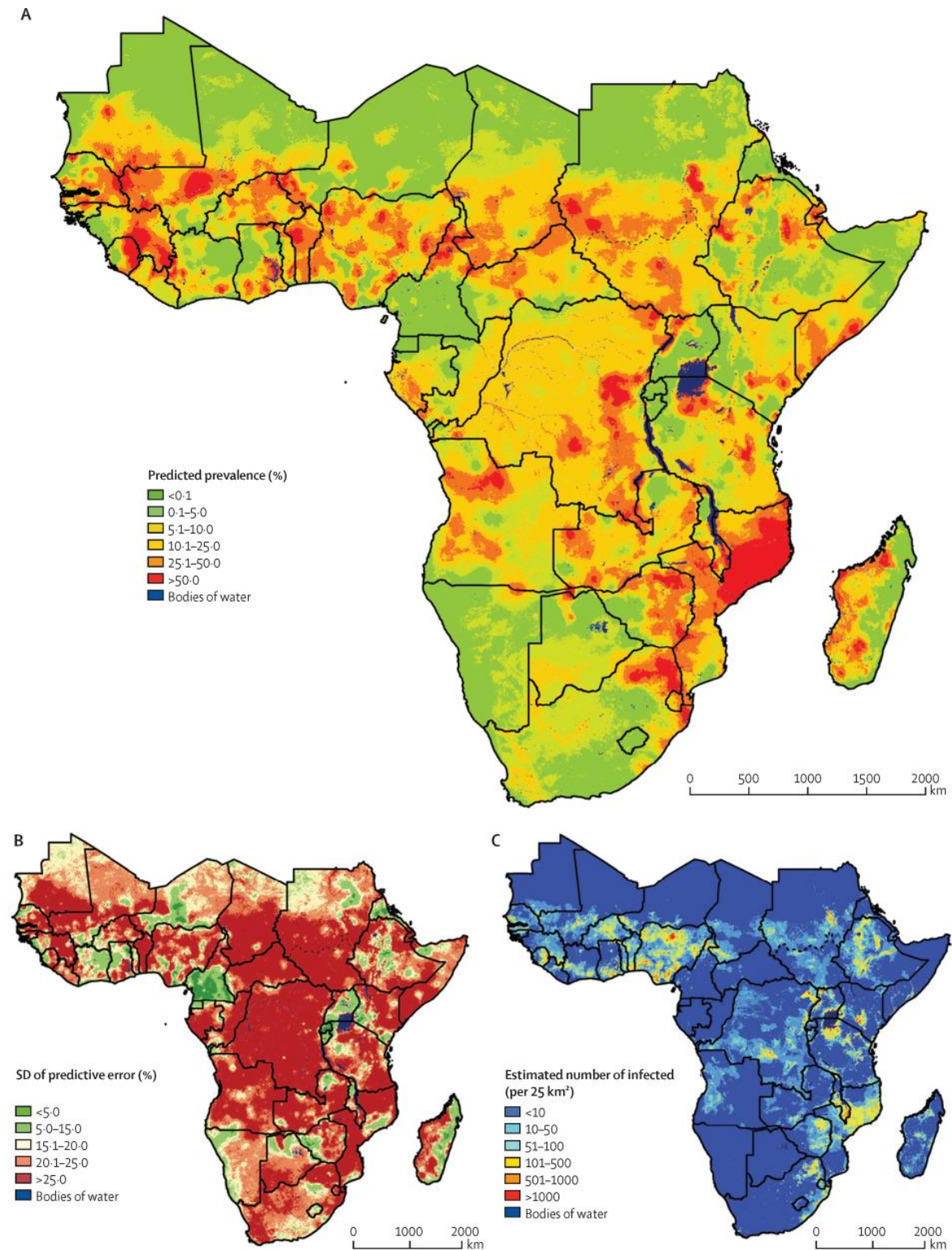
Figure 5.4: Prevalence of *Schistosoma spp* infection in school-aged children in sub-Saharan Africa, from 2000 onwards (A) Predictive prevalence based on the median of the posterior predictive distribution. (B) Prediction uncertainty based on the SD of the posterior predictive distribution. (C) Number of infected school-aged children, based on the predictive prevalence and gridded population estimates of school-aged children in 2012.

Table 5.3: Estimated number of school-aged children at risk of schistosomiasis and number of praziquantel doses needed for prevention of morbidity in school-aged children and the entire population in 2012*.

Country	School-aged children at low risk†	School-aged children at moderate risk‡	School-aged children at high risk§	Doses needed for school-aged children		Doses needed for entire population	
				Model-based estimate	WHO estimate**	Model-based estimate	WHO estimate**
Angola	2214 (1490–3577)	2591 (1775–3205)	1421 (707–2797)	3557 (3005–4346)	2755	7260 (5197–10276)	4984
Benin	559 (342–893)	1124 (952–1303)	736 (448–1020)	1491 (1293–1665)	1281	3479 (2678–4208)	2398
Botswana	253 (156–348)	128 (69–184)	49 (17–127)	201 (170–254)	143	410 (257–729)	167
Burkina Faso	1538 (1262–1837)	2226 (2006–2436)	963 (744–1214)	2588 (2446–2750)	1749	5310 (4691–6046)	2196
Burundi	2176 (1907–2343)	417 (264–631)	62 (25–132)	994 (952–1068)	395	1311 (1146–1605)	907
Cameroon	4063 (3759–4299)	1303 (1084–1584)	572 (427–780)	2589 (2486–2714)	3555	4337 (3908–4873)	9923
Central African Republic	555 (343–727)	358 (258–484)	189 (99–408)	560 (489–700)	521	1173 (866–1767)	875
Chad	1218 (822–1675)	1421 (1195–1626)	979 (577–1469)	2104 (1821–2406)	1754	4288 (3299–5361)	3470
Congo	422 (249–582)	335 (202–481)	87 (38–205)	402 (355–478)	287	755 (564–1085)	326
Côte d'Ivoire	2740 (2334–3088)	1802 (1510–2090)	494 (384–652)	2308 (2207–2434)	2396	4135 (3734–4678)	3880
DR Congo	6337 (4926–7592)	6688 (5887–7509)	4747 (3748–6073)	10200 (9515–11054)	9895	22019 (19254–25517)	18027
Djibouti	48 (3–233)	103 (17–201)	65 (4–244)	159 (99–255)	NA	451 (135–1016)	NA
Equatorial Guinea	158 (114–176)	19 (3–50)	2 (0–21)	64 (60–80)	27	78 (62–152)	52
Eritrea	1260 (1112–1385)	262 (180–362)	84 (37–153)	635 (591–694)	344	934 (754–1184)	514
Ethiopia	15612 (12948–17822)	8118 (6613–9590)	2732 (1804–4097)	12003 (11191–13092)	11677	20302 (17334–24507)	22092
Gabon	191 (88–289)	118 (58–192)	49 (14–162)	177 (144–249)	163	365 (216–720)	321
Ghana	2441 (2050–2820)	2265 (1982–2539)	918 (717–1189)	2868 (2728–3054)	3110	6164 (5527–7026)	6632
Guinea	1415 (1108–1660)	879 (707–1070)	577 (386–784)	1489 (1358–1633)	1219	3097 (2562–3702)	2062
Guinea-Bissau	139 (67–221)	155 (107–200)	67 (34–136)	195 (165–240)	127	403 (280–605)	180
Kenya	6861 (5896–7729)	3138 (2524–3836)	1068 (756–1587)	4947 (4654–5347)	5657	8630 (7375–10578)	11762
Lesotho	401 (173–510)	105 (12–233)	14 (0–142)	202 (177–302)	NA	292 (184–778)	NA
Liberia	388 (223–585)	435 (300–545)	167 (90–306)	520 (454–617)	403	1057 (793–1469)	1040
Madagascar	3759 (3352–3990)	1524 (1326–1831)	530 (431–966)	2559 (2477–2831)	2750	4244 (3926–5564)	6454
Malawi	1356 (1017–1650)	1892 (1665–2078)	987 (771–1302)	2388 (2237–2589)	2922	5067 (4468–5870)	6782
Mali	1223 (1054–1386)	1638 (1460–1825)	1228 (1039–1411)	2452 (2343–2560)	2340	5569 (5088–6017)	5669
Mauritania	421 (246–514)	337 (286–465)	206 (140–313)	519 (470–596)	297	1096 (892–1509)	662
Mozambique	1081 (771–1402)	2075 (1801–2367)	3564 (3159–3988)	4955 (4727–5203)	5074	12916 (11988–13889)	13456
Namibia	435 (348–489)	112 (68–178)	36 (13–86)	238 (218–275)	190	380 (292–617)	453
Niger	2793 (2528–3040)	1607 (1386–1875)	702 (502–912)	2437 (2315–2575)	2876	4216 (3751–4716)	5733
Nigeria	17936 (16300–19459)	16947 (15731–18163)	8154 (6970–9661)	22624 (21827–23603)	22537	45491 (42357–49466)	60622
Rwanda	2965 (2815–3125)	299 (155–428)	30 (2–99)	1172 (1134–1228)	445	1330 (1207–1524)	757
Senegal	1823 (1499–1995)	1160 (1010–1470)	499 (411–616)	1693 (1626–1784)	1866	3065 (2808–3449)	4180
Sierra Leone	456 (313–666)	726 (545–884)	507 (358–710)	1031 (938–1144)	583	2513 (2089–3061)	1440
Somalia	1282 (971–1551)	820 (634–1044)	548 (365–779)	1390 (1259–1549)	285	2733 (2253–3337)	517
South Africa	5027 (3987–5913)	2885 (2364–3520)	1757 (1280–2404)	4883 (4514–5359)	2458	12225 (10056–15242)	5249
South Sudan	998 (637–1369)	1080 (898–1237)	719 (441–1098)	1596 (1411–1849)	NA	3600 (2827–4653)	NA
Sudan	3270 (2722–3981)	3700 (3274–4069)	2477 (1872–2968)	5422 (4999–5759)	2238	12407 (1722–13824)	5815
Swaziland	93 (21–190)	156 (94–196)	67 (21–176)	181 (143–246)	155	423 (253–734)	307
Tanzania	5309 (4511–6049)	5079 (4541–5551)	2562 (2094–3126)	6870 (6528–7272)	5985	14077 (12645–15778)	10135
The Gambia	178 (53–353)	292 (158–415)	75 (31–243)	290 (245–396)	168	554 (382–1007)	184
Togo	638 (563–701)	600 (534–675)	218 (175–259)	731 (705–758)	888	1451 (1341–1562)	1750
Uganda	6597 (5275–7334)	2766 (2134–3646)	856 (625–1469)	4442 (4207–4973)	3963	6911 (6099–8735)	8625
Zambia	1423 (1133–1787)	1805 (1587–2025)	818 (572–1097)	2200 (2028–2380)	2244	4386 (3736–5069)	4626
Zimbabwe	1126 (936–1322)	1779 (1618–1942)	902 (787–1088)	2173 (2091–2280)	1495	4947 (4576–5467)	3060
Total	111180 (106851–115307)	83472 (80721–86037)	44124 (41413–47251)	122811 (120883–125019)	109217	247477 (239365–256262)	238284

Numbers are thousands. DR Congo=Democratic Republic of the Congo. NA=not applicable. *Pixel-level risk estimates are used in the calculations; calculations are based on the median and 95% Bayesian credible interval of the posterior predictive distribution of the risk from 2000 onwards. †Prevalence <10%. ‡Prevalence 10–50%. §Prevalence >50%. **WHO estimate were obtained from reference (WHO 2015)

Our analysis suggests that, in 2012, around 163 million people in sub-Saharan Africa were infected with one of two *Schistosoma spp*, 57 million (35%) of whom were school-aged children. Our estimated number of doses needed each year for the entire population (247.5 million) was close to the 238 million estimated by WHO. However, at the national level, for several countries, the estimates differed substantially. For example, we estimated that lower numbers of doses would be needed for Cameroon, Kenya, and Malawi, but higher numbers for Somalia, South Africa, and Sudan than the numbers suggested by WHO. By contrast with the WHO estimates, our results are based on risk profiles that take into account within-country variation and temporal trends (Schur et al. 2012) and on population estimates at the pixel level projected in 2012 from the 2010 WorldPop data (United Nations 2016). WHO collected population data from national offices of statistics, national plans of action for neglected tropical disease control, and from UN population estimates (WHO 2012b). Our analysis allowed us to estimate treatment needs at different administrative levels, which provides important information to guide the distribution of praziquantel within countries. The amount of praziquantel donated by WHO is projected to increase every year, from 27 million doses in 2012, to 44 million in 2013, 75 million in 2014, 100 million in 2015, and up to 250 million in 2016. At its peak, the donation will be sufficient to treat all school-aged children at risk of schistosomiasis. Of note, our mean survey-derived prevalence by country might differ from our population adjusted estimates (owing to the predictive distribution of all pixels across the country) because observed data from non-national surveys might not be representative of the country.

Besides preventive chemotherapy, other measures to fight schistosomiasis should be used, such as snail control and behavioural modification. The use of molluscicides, mainly niclosamide, is the primary method for chemical snail control. Although niclosamide is used in low concentrations and is non-toxic to people, it might have a negative effect on aquatic life (Colley et al. 2014). Its use should, therefore, be restricted to areas most likely to achieve schistosomiasis elimination (Knopp et al. 2012). Behavioural modification in conjunction with improvements in water sources and sanitation is another possible approach, although provision of safe water for washing, bathing, and recreation, while effective, is expensive (Colley et al. 2014;Grimes et al. 2014).

Our Bayesian geostatistical models identified several environmental predictors that are related to increased risk of *Schistosoma spp* infection, such as bodies of water (exposure to contaminated water is an important risk factor), temperature, and precipitation for both species, and land cover for *S. mansoni*. Negative disease risks were found for substantial distance from the nearest body of water and barren land cover. Temperature and precipitation affected transmission in more complex ways, by affecting the intermediate host snail population and human activities related to water contact. Low temperature limits snail distribution and cercaria maturation (McCreesh & Booth 2013), and high temperature might negatively affect fecundity and survival of adult snails (Appleton 1977). High rainfall might

increase the risk of transmission, but might also decrease it, for example by creating fast-flowing water unsuitable for cercaria or snail survival (McCreesh & Booth 2013).

Various socioeconomic factors (e.g., wealth or poverty, education, and occupation) affect the behaviour of people (Gazzinelli et al. 2006;Huang & Manderson 2005;Ximenes et al. 2003) and, therefore, their exposure to disease. Improved water sources and sanitation are particularly important in schistosomiasis prevention of morbidity and in sustaining the benefits of chemotherapy (Aagaard-Hansen, Mwanga, & Bruun 2009;Asaolu & Ofoezie 2003;Steinmann et al. 2006;Utzinger et al. 2003). We found that areas with high infant mortality, which might reflect low socioeconomic status, generally had high risk of *S. mansoni* infection. We did not, however, find significant correlations between disease risk and improved water sources or sanitation. Socioeconomic factors are more likely to be associated with small-scale variation in disease risk, in contrast to climatic factors, which lead to large-scale variations. Our survey data, however, are aggregated at cluster level and variation within clusters, where socioeconomic or behavioural factors are most likely to have effects, cannot be estimated.

WHO produced a global distribution map of schistosomiasis and country-specific disease risk estimates for 2009 (WHO 2011c) and Schur and colleagues presented model-based *Schistosoma spp* prevalence estimates for east and west Africa in 2010 (Schur et al. 2011a;Schur et al. 2011b). Our analysis includes notably more data from later surveys and our findings suggest that the disease risk is lower than was estimated previously. For example, for countries in east and west Africa we identified 5,940 and 4,961 surveys for *S. mansoni* and *S. haematobium*, respectively, that were done after 2000, compared with 941 and 857 surveys included in the analysis of Schur and colleagues (Schur et al. 2011a;Schur et al. 2011b). Additionally, we included survey data obtained after 2000 from countries that had not been represented in previous analyses, such as Benin, Burundi, Eritrea, Guinea, Rwanda, Sierra Leone, and Togo.

Clements and colleagues produced model-based geostatistical maps of *S. haematobium* risk in Burkina Faso, Mali, Niger, and northwest Tanzania (Clements et al. 2006;Clements et al. 2008a), and of *S. mansoni* for the Great Lakes region of east Africa (Clements et al. 2010a). Our findings support their results. Our estimates also confirm previously model-based maps for Sierra Leone and Ghana (Hodges et al. 2012;Koroma et al. 2010;Soares Magalhães et al. 2011). Compared with *S. haematobium* risk maps modeled for Nigeria by Ekpo and colleagues (Ekpo et al. 2013), although we obtained similar patterns overall in Nigeria, we estimated higher risks in some southern areas and lower risks around the Lake Chad basin. These differences could be due to the larger number of surveys on which our analysis was based. Indeed, we identified surveys done in 1,210 unique locations for Nigeria, in 927 of which the surveys were done after 2000, compared with 368 unique locations in the study of

Ekpo and colleagues (Ekpo et al. 2013). Survey data from countries bordering Nigeria might have provided additional information in our analysis.

Although we tried to collect all available survey data on *Schistosoma spp* infection in sub-Saharan Africa, there are some large surveys with location-specific data to which we were unable to get access, such as the national schistosomiasis survey done in Mozambique in 2005–07 (Augusto et al. 2009). Most of the survey data we obtained for Mozambique were collected before 1980, but our estimated risk maps have similar patterns to those in the distribution maps of the 2005–07 national survey (Augusto et al. 2009), with high *S. haematobium* prevalence in the northern part and low *S. mansoni* prevalence in most of the country. We identified no published data for the northwest of South Sudan on *S. haematobium* prevalence. The Global Atlas of Helminth Infection website (http://www.thiswormyworld.org) reports unpublished survey data from the areas with low observed prevalence, but these are not available in an open-access format and, hence, we could not include them in our analysis. Our predictions for this region indicate a low to moderate risk of *S. haematobium* infection (less than 25%).

For countries or regions with scarce or old survey data, our estimates should be interpreted cautiously. For example, few surveys are available for western Democratic Republic of the Congo, which might explain the high uncertainty in the estimates of that region. The infection risk around the Lake Chad basin of Nigeria could be underestimated because very few surveys have been done since 2000 due to insurgency and insecurity. We could find no location-specific surveys for Lesotho or any *S. haematobium* surveys for Djibouti. Additionally, we identified no or very few surveys done after 2000 in Botswana, Central African Republic, Chad, Congo, Equatorial Guinea, Gabon, Guinea-Bissau, Madagascar, Namibia, and Somalia. National surveys in those countries would be helpful to understand the current disease situation more clearly. We do not have enough data to estimate when the prevalence decline started after 2000 and, therefore, we did not use more refined time-frames to estimate the temporal trends. This limitation might have led to overestimation of the infection risk in some areas.

We extracted data from sources that clearly indicated study designs and conduct of surveys and adhered to exclusion criteria to remove data that were potential sources of bias. For example, we excluded studies done in populations or subgroups not representative of the general population. Data are, however, obtained from studies with different designs, different diagnostic methods, and covering different age groups. Additionally, the distribution of the surveys varies across time periods and the survey locations might over-represent endemic areas. Statistical models can address some differences in data characteristics (Diggle, Menezes, & Su 2010;Wang et al. 2008), but there is a need to do national surveys repeatedly for monitoring and assessment of control interventions. In many countries in Africa, malaria control programmes do national surveys to assess interventions, but many neglected tropical disease programmes lack the resources to collect representative data and might unnecessarily

overtreat populations by relying on historical data. An important risk of overtreatment is resistance leading to reduced drug efficacy (Wang et al. 2012). Our climatic, environmental, and socioeconomic predictors were extracted from well known sources and widely used databases (Appendix) and, therefore, we are confident that the data quality is good.

To estimate schistosomiasis prevalence we assumed independence of *S. haematobium* and *S. mansoni* infection risk. This approach is supported by Chammartin and colleagues (Chammartin et al. 2014b), who analysed data from a 2012 national survey in Côte d'Ivoire and reported that species co-infection is less likely to occur other than simply by chance. By contrast, Meurs and colleagues (Meurs et al. 2012) reported high co-infection in some villages of Senegal. Although in areas with high species co-infection risk we might have overestimated the schistosomiasis prevalence, our calculations of co-infection risk for a range of values quantifying species dependence showed that the independence assumption has very little effect on the overall infection risk.

As assessment of the quality of different diagnostic approaches and procedures is difficult, and to avoid discarding surveys with incomplete information about the diagnostic method, we assumed similar diagnostic sensitivity and specificity across surveys, which might have led to bias and could explain some of the regional variation in predicted prevalence (Booth et al. 2003). Moreover, most of survey data were aggregated across different age groups. Nevertheless, we distinguished between community-based and school-based surveys and, therefore, were able to calculate risk estimates for the entire population and school-aged children, respectively. Our geostatistical model assumed a stationary spatial process across sub-Saharan Africa, but predictors might affect the disease risk differently in regions with different ecological, socioeconomic, and health system profiles. Treatment coverage data are available from WHO yearly from 2006 onwards, aggregated at the country level. Unfortunately the survey data are not high-resolution temporally and do not allow estimation of the annual schistosomiasis risk that can be related to the treatment coverage we have already discussed.

Geostatistical modeling enabled us to estimate the geographical distribution of *Schistosoma spp* infection risk across sub-Saharan Africa, by prevalence, number of infected individuals, and doses needed for prevention of morbidity. Our overall estimate for number of doses of praziquantel needed per year is similar to that reported by WHO, but at the national level these estimates differed substantially for several countries. Nevertheless, our estimates provide information that will help national control programmes and international donors to support schistosomiasis control activities in Africa.

Acknowledgements

5.5 Appendix

5.5.1 Remote sensing data sources

Source	Data type	Data period	Temporal resolution	Spatial resolution
MODIS/Terra[b]	LST[l]	2000-2012	8 days	1 km
MODIS/Terra[b]	NDVI[m]	2000-2012	16 days	1 km
MODIS/Terra[b]	Land cover	2001-2004	Yearly	1 km
WorldClim[c]	Elevation	2000	-	1 km
WorldClim[c]	Bioclimatic variables	1950-2000	-	1 km
SWBD[d]	Water bodies	2000	-	30 m
Köppen-Geiger[e]	Climate zones	1976-2000	-	50 km
ISRIC[f]	pH in water	-	-	10 km
Atlas of the Biosphere[g]	Soil moisture	1950-1999	-	50 km
WorldPop[h]	Age and sex specific grid population	2010	-	1 km
SEDAC[i]	HII[n]	1995-2004	-	1 km
SEDAC[i]	Urban extents	1990-2000	-	1 km
SEDAC[i]	GDP[o]	1990, 2025	-	25 km
SEDAC[i]	IMR[p]	2000	-	4 km
DHS, MICS, WHS, and LSMS[j]	Sanitation	1991-2013	-	Household surveys
DHS, MICS, WHS, and LSMS[j]	Drinking-water sources	1991-2013	-	Household surveys
GADM[k]	Geographical administrative boundaries	2012	-	-

[a] Land cover data accessed in June 2011 and other data asccessed in November 2013.

[b] Moderate Resolution Imaging Spectroradiometer (MODIS)/Terra, available at: http://modis.gsfc.nasa.gov/.

[c] Available at: http://www.worldclim.org/current.

[d] Shuttle Radar Topography Mission Water Body Data (SWBD), available at:
http://gis.ess.washington.edu/data/vector/worldshore/index.html.

[e] World Maps of Köppen-Geiger climate classification, available at: http://koeppen-geiger.vu-wien.ac.at/shifts.htm.

[f] International Soil Reference and Information Center, available at: http://www.isric.org/data/isric-wise-derived-soil-properties-5-5-arc-minutes-global-grid-version-12.

[g] Available at: http://www.sage.wisc.edu/atlas/maps.php?datasetid=23&includerelatedlinks=1&dataset=23.

[h] The WorldPop project, available at: http://www.worldpop.org.uk/.

[i] Socioeconomic data and applications center, available at: http://sedac.ciesin.org/.

[j] Demographic and Health Surveys (DHS), available at: http://dhsprogram.com/; Multiple Cluster Indicator Surveys (MICS), available at: http://www.childinfo.org/mics_available.html; World Health Surveys (WHS), available at: http://www.who.int/healthinfo/survey/en/; and Living Standards Measurement Study (LSMS), available at: http://iresearch.worldbank.org/lsms/lsmssurveyFinder.htm.

[k] Global Administrative Areas database, available at: http://www.gadm.org/.

[l] Land surface temperature (LST) day and night.

[m] Normalized difference vegetation index.

[n] Human influence index.

[o] Gross domestic product.

[p] Infant mortality rates.

5.5.2 Processing of environmental and socioeconomic data

Land surface temperature (LST) and normalized difference vegetation index (NDVI) were averaged over the period of 2000-2012. Land cover was summarised by the most frequent category within the pixel that the point belonged over the period of 2001-2004. We combined similar land cover classes and re-grouped them into seven categories: (i) grasslands; (ii) forests; (iii) scrublands and savannas; (iv) croplands; (v) urban; (vi) wet areas; and (vii) barren areas. As data for infant mortality rates (IMR) of South Sudan is missing from the socioeconomic data and applications center (SEDAC), we replaced it with the average IMR of the country in 2000 obtained from the World Bank (http://data.worldbank.org/). We compiled all available household data from Demographic and Health Surveys (DHS), Multiple Cluster Indicator Surveys (MICS), World Health Surveys (WHS), and Living Standards Measurement Study (LSMS), and extracted geo-referenced water and sanitation (WASH) indicators, such as proportion of households with improved drinking-water sources and sanitation adhering to definitions utilised by the WHO/UNICEF Joint Monitoring Programme (JMP) (http://www.wssinfo.org/definitions-methods/watsan-categories/). Overall, we obtained WASH indicators at around 38,000 locations across Africa covering the period of 1991-2013. Bayesian geostatistical logistic regression models considering urban/rural as a covariate were fitted on the WASH indicators and model-based predictions were obtained on a 5×5 km grid over Africa. Environmental and socioeconomic data at the survey locations were extracted using Visual Fortran version 6.0 (Digital Equipment Corporation; Maynard, USA). As the outcome of interest (i.e. infection prevalence) is not available at the resolution of the covariates for surveys aggregated over administration divisions of level two or three, we linked the centroid of those divisions with the average value of each covariate within the divisions.

5.5.3 Geostatistical model fitting

Bayesian geostatistical logistic regression models were fitted to obtain spatially explicit *Schistosoma* infection estimates by introducing location-specific random effects. Let Y_i, n_i and p_i be the number of positive individuals, the number of examined individuals and the probability of infection at location i ($i = 1, 2,…, L$), respectively. Y_i was assumed to be generated by a binominal distribution $Y_i \sim Bn(p_i, n_i)$. In particular, $\text{logit}(p_i) = X_i^T \beta + \varepsilon_i + \phi_i$, where X_i and β are the vector of covariates and coefficients, respectively. ε_i and ϕ_i indicate location-specific spatial and exchangeable random effects, respectively. We assumed $\vec{\varepsilon} = (\varepsilon_1,...,\varepsilon_L)^T$ is arising from a zero-mean multivariate normal distribution $\vec{\varepsilon} \sim MVN(0, \Sigma)$ with a Matérn covariance matrix $\Sigma_{ij} = \sigma_{sp}^2 (\kappa d_{ij})^\upsilon K_\upsilon (\kappa d_{ij}) / (\Gamma(\upsilon) 2^{\upsilon-1})$. d_{ij} is the Euclidean distance between locations i and j, κ is a scaling parameter, υ is a smoothing parameter fixed to 1, and K_υ denotes the modified Bessel function of second kind and order υ. The

spatial range $\rho = \sqrt{8}/\kappa$ is regarded as the distance that spatial correlation becomes negligible (<0.1). We assumed ϕ_i is arising from a zero-mean normal distribution $\phi_i \sim N(0, \sigma_{nonsp}^2)$. A Bayesian inferential framework was adopted to estimate the parameters as well as hyperparameters. For regression coefficients, non-informative normal distribution priors $\beta_0, \beta_k \sim N(0,1000)$ are used. The priors used for hyper parameters τ_{sp} and κ were $\log(\tau_{sp}) \sim \log normal(0,100)$ and $\log(\kappa) \sim \log normal(0,100)$ respectively, where $\sigma_{sp}^2 = 1/(4\pi\kappa^2\tau_{sp}^2)$ and $\rho = \sqrt{8}/\kappa$. To assess model sensitivity to the prior specification, we tried alternative prior distributions for τ_{sp} and κ such as $\log(\tau_{sp}) \sim \log normal(0,10)$ and $\log(\kappa) \sim \log normal(0,10)$. The model estimates were almost the same. Model fitting was undertaken in INLA by using the homonymous R-package (available at www.r-inla.org).

5.5.4 Bayesian variable selection

We proposed normal mixture of inverse gammas priors for coefficients of continuous covariates, which is constructed by two components: a narrow spike that shrinks coefficients to zero and a wide slab that moves coefficients away from zero. The covariates are selected if the wide slab component is predominant in the posterior distribution (i.e., posterior median of slab component is greater than 0.5). For categorical variables, we used an adapted version of the above prior to allow simultaneously inclusion or exclusion of all coefficients related to the variable. Furthermore, the variable selection procedure enabled to select a maximum of one predictor from each of the highly correlated variable groups and to identify the best functional form (i.e., continuous or categorical) of each continuous covariate. The following paragraphs show the details.

Highly correlated predictors were grouped together into G groups. Variable selection was carried out using a Bayesian logistic regression model with exchangeable random effects, that is, where T_{2i} and T_{3i} are indicator variables corresponding to the second and third categories of survey year, respectively, and ϕ_i is an exchangeable random effect as described above. We introduced $I_{j^{(g)}}$ as the indicator for the jth predictor in group g. For groups with more than one predictor, we assumed that $(I_{1^{(g)}}, ..., I_{J^{(g)}}, I_{J^{(g)}+1})$ follows a multinomial distribution with $J^{(g)} + 1$ categories $(I_{1^{(g)}}, ..., I_{J^{(g)}}, I_{J^{(g)}+1}) \sim multi(\pi_{1^{(g)}}, ..., \pi_{J^{(g)}}, \pi_{J^{(g)}+1}, 1)$, where $J^{(g)}$ is the total number of predictors, $I_{J^{(g)}+1}$ indicates the absence of any predictor in group g. We assume that the probabilities related to the indicators of the predictors follow a Dirichlet distribution, that is $(\pi_{1^{(g)}}, ..., \pi_{J^{(g)}}, \pi_{J^{(g)}+1}) \sim dirch(1,...,1,1)$. Thus, for each group, a maximum of one predictor can be selected. For groups with just one predictor, we assumed $I_{j^{(g)}} \sim bern(\pi_{j^{(g)}})$, where $\pi_{j^{(g)}} \sim beta(1,1)$.

We proposed a normal mixture of inverse Gammas (NMIG) prior for $\beta_{j^{(g)}}$, that is

$\beta_{j^{(g)}} \sim N(0, \sigma^2_{j^{(g)}})$, where $\sigma^2_{j^{(g)}} \sim I_{j^{(g)}} IG(a_{\sigma_B}, b_{\sigma_B}) + (1 - I_{j^{(g)}}) \upsilon_0 IG(a_{\sigma_B}, b_{\sigma_B})$ and $(a_{\sigma_B}, b_{\sigma_B}) = (5, 25)$

are fixed parameters of non-informative inverse-gamma distribution and $\upsilon_0 = 0.00025$ is a

small positive constant. The above prior for $\beta_{j^{(g)}}$ is called a mixed spike and slab prior as one

component of the mixture $\upsilon_0 IG(a_{\sigma_B}, b_{\sigma_B})$ (when $I_{j^{(g)}} = 0$) is a narrow spike around zero that

strongly shrinks $\beta_{j^{(g)}}$ to zero, while the other component $IG(a_{\sigma_B}, b_{\sigma_B})$ (when $I_{j^{(g)}} = 1$) is a

wide slab that moves $\beta_{j^{(g)}}$ away from zero. For coefficients of categorical variables, normal

mixture of inverse Gammas with parameter expansion (peNMIG) priors were applied, which

allows simultaneous including or excluding a block of coefficients by improving "shrinkage"

properties. Let $\beta_{j^{(g)}h}$ be the regression coefficient for the hth category of the jth predictor in

group g, $\beta_{j^{(g)}h} = \alpha_{j^{(g)}} \xi_{j^{(g)}h}$ where $\alpha_{j^{(g)}}$ follows a NMIG prior described above and

$\xi_{j^{(g)}h} \sim N(m_{j^{(g)}h}, 1)$ with prior mean $m_{j^{(g)}h}$ either 1 or -1 in equal probability, allowing to shrink

$|\xi_{j^{(g)}h}|$ towards 1. In this way, $\alpha_{j^{(g)}}$ models the overall contribution of the jth predictor in

group g, and $\xi_{j^{(g)}h}$ estimates the effects of each category for the predictors (Chammartin et al.

2013b). To select either categorical or linear form of a continuous variable, we introduced

another indicator I_d. Let $\vec{\beta}_{j^{(g)}d1}$ and $\beta_{j^{(g)}d2}$ indicate coefficients of the categorical and linear

form of the jth predictor in group g, respectively. Then for continuous predictor $X^{(g)}_{ji}$, we

have $\beta_{j^{(g)}} X^{(g)}_{ji} = I_d \vec{\beta}_{j^{(g)}d1} \vec{Z}^{(g)}_{ji} + (1 - I_d) \beta_{j^{(g)}d2} R^{(g)}_{ji}$, where $\vec{Z}^{(g)}_{ji}$ and $R^{(g)}_{ji}$ are categorical and linear

form of the predictor, respectively, and the prior $I_d \sim bern(0.5)$. Fitting was undertaken

through Markov chain Monte Carlo (MCMC) simulation in OpenBUGS version 3.0.2

(Imperial College and Medical Research Council; London, UK). We assessed convergence by

the Gelman and Rubin diagnostic(Gelman & Rubin 1992), using the coda library in R

(Plummer et al. 2006).

5.5.5 Overview of survey data in sub-Saharan Africa

Country	Relevant papers	*Schistosoma haematobium*							
		Total	Mean prevalence (%)	Survey type		Period	Year of Survey		
		Survey/ location		Survey/location			Survey/location		
				School	Community		<1980	1980-2000	>=2000
Angola	13	120/94	34.2	22/22	98/72	1952-2012	118/93	0/0	2/1
Benin	8	39/38	38.8	30/29	9/9	1952-2012	2/2	11/11	26/25
Botswana	2	24/22	0.9	9/8	15/14	1965-1983	15/15	9/8	0/0
Buikina Faso	22	482/286	31.8	293/190	189/96	1950-2008	182/102	115/108	185/94
Burundi	1	22/22	0.0	22/22	0/0	2007-2007	0/0	0/0	22/22
Cameroon	33	898/784	14.6	765/698	133/86	1950-2011	33/23	505/467	360/345
Central African Repbulic	3	32/31	12.2	0/0	32/31	1950-1971	32/31	0/0	0/0
Chad	4	18/18	19.8	12/12	6/6	1950-1993	6/6	12/12	0/0
Congo	12	97/68	25.3	70/50	27/18	1962-1992	27/22	70/55	0/0
Côte d'Ivoire	24	360/255	16.8	260/162	100/93	1970-2013	39/35	183/109	138/123
DR Congo	11	133/99	20.8	103/81	30/18	1950-2010	38/23	86/67	9/9
Djibouti	0	-	-	0	0	-	0	0	0
Equatorial Guinea	3	3/3	0.0	0/0	3/3	1987-1988	0/0	3/3	0/0
Eritrea	2	48/46	0.0	46/45	2/1	1952-2002	8/7	0/0	40/40
Ethiopia	21	114/71	17.7	62/39	52/32	1952-2011	35/31	71/40	8/8
Gabon	9	19/10	29.8	9/4	10/6	1966-2013	6/4	7/4	6/4
Ghana	28	303/271	26.9	234/221	69/50	1950-2010	72/57	55/53	176/173
Guinea	5	69/57	15.2	58/48	11/9	1995-2011	0/0	33/29	36/28
Guinea-Bissau	2	21/20	46.5	0/0	21/20	1950-1983	20/19	1/1	0/0
Kenya	64	649/511	19.5	535/437	114/74	1970-2013	78/70	179/136	392/338
Lesotho	0	-	-	0	0	-	0	0	0
Liberia	8	126/80	21.6	12/3	114/77	1968-1987	17/15	109/72	0/0
Madagascar	13	270/241	17.4	43/27	227/214	1952-2002	223/209	42/34	5/4
Malawi	16	123/114	37.5	103/98	20/16	1952-2012	6/6	70/66	47/45
Mali	34	978/600	32.0	54/50	924/550	1950-2008	22/19	916/554	40/37
Mauritania	11	118/86	35.3	82/60	36/26	1960-2011	63/51	29/26	26/20
Mozambique	11	144/142	67.8	11/11	133/131	1953-2008	131/131	4/4	9/7
Namibia	4	38/32	13.0	10/7	28/25	1965-1987	23/22	15/13	0/0
Niger	57	789/561	34.0	574/420	215/141	1950-2010	17/16	577/402	195/172
Nigeria	188	1488/1210	21.8	1074/914	414/296	1952-2012	75/46	379/269	1034/927
Rwanda	1	30/30	0	30/30	0/0	2008-2008	0/0	0/0	30/30
Senegal	53	660/544	25.7	354/315	306/229	1958-2009	35/32	257/222	368/325
Sierra Leone	8	64/58	27.1	11/10	53/48	1950-2010	10/10	41/36	13/12
Somalia	9	68/59	46.1	9/5	59/54	1952-1987	42/37	26/24	0/0
South Africa	29	193/151	39.3	179/143	14/8	1955-2010	117/104	61/34	15/13
Sudan & South Sudan	26	520/294	9.4	221/134	299/160	1958-2013	9/7	185/170	326/119
Swaziland	3	6/6	8.1	5/5	1/1	1967-2010	1/1	0/0	5/5
Tanzania	113	642/535	29.8	513/444	129/91	1950-2013	130/106	300/253	212/196
The Gambia	6	88/45	35.2	0/0	88/45	1953-1983	85/42	3/3	0/0
Togo	10	1197/1125	20.4	1137/1084	60/41	1952-2009	42/40	61/59	1094/1052
Uganda	6	94/62	3.2	76/48	18/14	1957-2009	16/13	2/2	76/50
Zambia	32	222/203	26.6	194/182	28/21	1969-2011	43/38	71/68	108/104
Zimbabwe	49	468/434	32.1	412/394	56/40	1953-2011	18/15	219/213	231/212
Total	917	11777/9318	24.7	7634/6452	4143/2866	1950-2013	1836/1500	4707/3627	5234/4540

(Continues in next page)

(Continued from previous page)

| Country | Relevant papers | Total Survey/ location | Mean prevalence (%) | Survey type Survey/location | | Period | Year of Survey Survey/location | | |
				School	Community		<1980	1980-2000	>=2000
Angola	2	18/18	0.8	0/0	18/18	1956-1965	18/18	0/0	0/0
Benin	7	46/37	10.6	26/25	20/12	1987-2012	0/0	10/10	36/27
Botswana	5	35/27	29.3	14/9	21/18	1965-2001	17/16	17/12	1/1
Buikina Faso	11	224/121	5.7	192/99	32/22	1973-2008	31/21	5/5	188/95
Burundi	10	109/59	15.5	66/34	43/25	1952-2007	1/1	86/37	22/22
Cameroon	22	825/757	6.9	762/706	63/51	1962-2011	13/11	467/453	345/332
Central African Repbulic	2	2/2	30.3	0/0	2/2	1950-1971	2/2	0/0	0/0
Chad	3	7/6	12.1	0/0	7/6	1967-2006	6/5	0/0	1/1
Congo	1	2/2	0.0	0/0	2/2	1965-1965	2/2	0/0	0/0
Côte d'Ivoire	37	646/521	30.6	547/432	99/89	1972-2013	21/20	133/115	492/424
DR Congo	25	202/159	32.7	112/91	90/68	1950-2011	40/29	147/117	15/15
Djibouti	1	1/1	74.3	0/0	1/1	1998-1998	0/0	1/1	0/0
Equatorial Guinea	2	3/2	6.2	0/0	3/2	1987-1989	0/0	3/2	0/0
Eritrea	4	50/47	8.4	47/45	3/2	1952-2002	9/8	1/1	40/40
Ethiopia	92	659/511	20.5	302/239	357/272	1952-2013	177/144	384/335	98/85
Gabon	1	1/1	0.0	0/0	1/1	1992-1992	0/0	1/1	0/0
Ghana	8	97/96	3.4	85/85	12/11	1956-2008	10/9	9/9	78/78
Guinea	5	67/57	35.0	57/48	10/9	1995-2011	0/0	31/29	36/28
Guinea-Bissau	1	1/1	0.7	0/0	1/1	1983-1983	0/0	1/1	0/0
Kenya	62	1440/1088	19.1	1385/1051	55/37	1967-2013	60/55	160/108	1220/957
Lesotho	0	-	-	0	0	-	0	0	0
Liberia	6	101/73	16.8	10/4	91/69	1978-1987	12/12	89/67	0/0
Madagascar	23	550/473	15.9	165/114	385/359	1951-2012	378/358	171/126	1/1
Malawi	8	54/54	11.5	44/44	10/10	1977-2012	3/3	21/21	30/30
Mali	24	923/587	15.1	77/73	846/514	1974-2006	16/16	838/517	69/66
Mauritania	4	35/23	10.2	35/23	0/0	1994-2011	0/0	20/17	15/9
Mozambique	7	135/132	8.6	1/1	134/131	1953-2004	131/131	0/0	4/1
Namibia	4	38/32	34.6	10/7	28/25	1965-1987	23/22	15/13	0/0
Niger	11	231/204	2.5	182/164	49/40	1982-2010	0/0	69/63	162/150
Nigeria	56	852/784	2.2	731/685	121/99	1959-2011	21/18	81/73	750/700
Rwanda	3	33/33	3.8	32/32	1/1	1951-2008	1/1	0/0	32/32
Senegal	34	189/158	20.5	33/29	156/129	1961-2012	5/5	129/109	55/50
Sierra Leone	10	112/106	17.1	68/65	44/41	1984-2011	0/0	30/26	82/81
Somalia	3	10/9	0	4/3	6/6	1952-1986	8/7	2/2	0/0
South Africa	21	114/89	23.6	99/81	15/8	1955-2005	103/78	10/10	1/1
Sudan & South Sudan	30	348/291	23.8	149/133	199/158	1971-2010	47/34	178/166	123/96
Swaziland	1	1/1	15.2	1/1	0/0	1967-1967	1/1	0/0	0/0
Tanzania	66	317/273	19.5	220/200	97/73	1954-2012	80/62	87/73	150/150
The Gambia	2	5/5	28.8	0/0	5/5	1956-1992	4/4	1/1	0/0
Togo	9	1195/1121	3.5	1138/1082	57/39	1976-2009	36/34	65/61	1094/1052
Uganda	54	815/645	23.6	631/513	184/132	1957-2012	22/19	71/54	722/607
Zambia	22	138/126	9.8	119/114	19/12	1969-2011	16/15	25/23	97/92
Zimbabwe	31	430/408	10.9	389/376	41/32	1964-2010	17/14	196/194	217/204
Total	701	11061/9140	14.8	7733/6608	3328/2532	1950-2013	1331/1175	3554/2852	6176/5427

5.5.6 Location type, diagnostic methods, and incomplete information for schistosomiasis survey data in sub-Saharan Africa

Country	*Schistosoma haematobium*								
	Location type[##]		Diagnostic method* (%)					Missing information***	
	Point	District[#]	Sedi	Filtr	Cen	RS	NS	Study year	Number of examined
Angola	106/86	14/8	24.2	0.7	75.2	0.0	0.0	21 (17.5%)	42 (35.0%)
Benin	36/35	3/3	0.0	74.4	14.0	0.0	11.6	13 (33.3%)	0 (0.0%)
Botswana	24/22	0/0	57.7	0.0	34.6	0.0	7.7	0 (0.0%)	0 (0.0%)
Buikina Faso	481/286	1/0	0.6	16.7	25.1	19.1	38.5	14 (2.9%)	31 (6.4%)
Burundi	22/22	0/0	0.0	100.0	0.0	0.0	0.0	0 (0.0%)	0 (0.0%)
Cameroon	890/776	8/8	1.5	46.0	5.7	43.5	3.2	3 (0.3%)	54 (6.0%)
Central African Repbulic	7/6	25/25	0.0	0.0	75.8	3.0	21.2	2 (6.3%)	0 (0.0%)
Chad	18/18	0/0	2.5	10.0	42.5	2.5	42.5	0 (0.0%)	0 (0.0%)
Congo	95/66	2/2	0.0	26.1	25.2	48.7	0.0	2 (2.1%)	1 (1.0%)
Côte d'Ivoire	360/255	0/0	0.0	29.6	8.5	59.5	2.5	2 (0.6%)	4 (1.1%)
DR Congo	133/99	0/0	5.2	0.0	24.9	67.1	2.9	4 (3.0%)	5 (3.8%)
Djibouti	0	0	0.0	0.0	0.0	0.0	0.0	0	0
Equatorial Guinea	3/3	0/0	0.0	100.0	0.0	0.0	0.0	1 (33.3%)	0 (0.0%)
Eritrea	48/46	0/0	0.0	83.3	16.7	0.0	0.0	0 (0.0%)	0 (0.0%)
Ethiopia	107/64	7/7	5.7	19.7	34.4	31.2	9.0	9 (7.9%)	15 (13.2%)
Gabon	18/10	1/0	0.0	100.0	0.0	0.0	0.0	5 (26.3%)	3 (15.8%)
Ghana	277/248	26/23	6.3	56.6	27.5	6.0	3.6	41 (13.5%)	17 (5.6%)
Guinea	69/57	0/0	15.7	33.7	42.7	1.1	6.7	0 (0.0%)	0 (0.0%)
Guinea-Bissau	21/20	0/0	56.1	0.0	41.5	0.0	2.4	0 (0.0%)	0 (0.0%)
Kenya	643/508	6/3	0.6	38.5	30.1	10.5	20.4	39 (6.0%)	36 (5.5%)
Lesotho	0	0	0.0	0.0	0.0	0.0	0.0	0	0
Liberia	125/79	1/1	0.0	54.5	23.9	0.0	21.6	0 (0.0%)	50 (39.7%)
Madagascar	270/241	0/0	0.7	17.4	0.0	1.4	80.6	14 (5.2%)	8 (3.0%)
Malawi	116/108	7/6	0.0	34.1	22.0	43.4	0.6	5 (4.1%)	7 (5.7%)
Mali	976/599	2/1	0.2	93.4	0.3	0.4	5.7	142 (14.5%)	121 (12.4%)
Mauritania	118/86	0/0	0.0	43.4	10.3	36.0	10.3	11 (9.3%)	0 (0.0%)
Mozambique	105/103	39/39	83.5	15.8	0.6	0.0	0.0	1 (0.7%)	0 (0.0%)
Namibia	38/32	0/0	60.5	0.0	0.0	0.0	39.5	0 (0.0%)	9 (23.7%)
Niger	782/554	7/7	0.6	79.3	0.0	11.1	9.0	18 (2.3%)	45 (5.7%)
Nigeria	1432/1163	56/47	11.8	10.8	16.0	51.6	9.9	225 (15.1%)	128 (8.6%)
Rwanda	0/0	30/30	0.0	96.8	0.0	0.0	3.2	0 (0.0%)	0 (0.0%)
Senegal	648/532	12/12	0.4	43.8	15.8	36.6	3.5	107 (16.2%)	42 (6.4%)
Sierra Leone	63/57	1/1	14.9	1.4	33.8	1.4	48.7	8 (12.5%)	35 (54.7%)
Somalia	68/59	0/0	68.8	2.6	27.3	0.0	1.3	2 (2.9%)	25 (36.8%)
South Africa	178/139	15/12	18.0	21.9	11.8	22.8	25.4	19 (9.8%)	81 (42.0%)
Sudan & South Sudan	515/290	5/4	20.0	9.1	18.4	0.2	52.4	30 (5.8%)	76 (14.6%)
Swaziland	6/6	0/0	0.0	0.0	71.4	0.0	28.6	0 (0.0%)	0 (0.0%)
Tanzania	617/518	25/17	0.6	55.2	23.3	16.2	4.9	43 (6.7%)	22 (3.4%)
The Gambia	88/45	0/0	0.0	14.6	77.1	1.0	7.3	2 (2.3%)	1 (1.1%)
Togo	1175/1105	22/20	0.0	7.9	0.4	91.0	0.7	6 (0.5%)	5 (0.4%)
Uganda	94/62	0/0	2.0	89.8	3.1	0.0	5.1	3 (3.2%)	2 (2.1%)
Zambia	210/193	12/10	7.1	47.5	25.1	18.6	1.7	43 (19.4%)	6 (2.7%)
Zimbabwe	451/424	17/10	30.1	48.5	8.7	10.0	2.6	53 (11.3%)	166 (35.5%)
Total	11433/9022	344/296	6.8	37.6	14.1	29.6	11.9	888 (7.5%)	1037 (8.8%)

(Continues in next page)

(Continued from previous page)

Country	Location type[##]		Diagnostic method** (%)							Missing information***	
	Point	District[#]	KK	Sedi	Filtr	Cen	Con	Othe	NS	Study year	Number of examined
Angola	15/15	3/3	5.0	0.0	0.0	0.0	95.0	0.0	0.0	0 (0.0%)	10 (55.6%)
Benin	46/37	0/0	95.7	0.0	0.0	0.0	4.3	0.0	0.0	10 (21.7%)	4 (8.7%)
Botswana	35/27	0/0	5.6	38.9	0.0	0.0	41.7	2.8	11.1	0 (0.0%)	4 (11.4%)
Buikina Faso	224/121	0/0	17.8	0.0	0.0	0.0	6.6	0.0	75.6	13 (5.8%)	13 (5.8%)
Burundi	109/59	0/0	98.7	0.0	0.0	0.0	0.7	0.7	0.0	6 (5.5%)	49 (45.0%)
Cameroon	821/755	4/2	43.3	1.6	0.0	0.0	1.1	51.3	2.9	1 (0.1%)	32 (3.9%)
Central African Repbulic	2/2	0/0	0.0	0.0	0.0	0.0	6.3	0.0	93.8	0 (0.0%)	0 (0.0%)
Chad	7/6	0/0	25.0	62.5	0.0	0.0	0.0	0.0	12.5	0 (0.0%)	0 (0.0%)
Congo	2/2	0/0	0.0	0.0	0.0	0.0	100.0	0.0	0.0	0 (0.0%)	0 (0.0%)
Côte d'Ivoire	644/521	2/0	97.2	0.0	0.0	0.2	2.3	0.2	0.2	3 (0.5%)	16 (2.5%)
DR Congo	202/159	0/0	75.3	0.8	2.6	0.0	6.0	0.4	15.0	16 (7.9%)	16 (7.9%)
Djibouti	1/1	0/0	0.0	0.0	0.0	0.0	0.0	100.0	0.0	0 (0.0%)	0 (0.0%)
Equatorial Guinea	3/2	0/0	100.0	0.0	0.0	0.0	0.0	0.0	0.0	2 (66.7%)	0 (0.0%)
Eritrea	50/47	0/0	80.0	18.0	0.0	0.0	2.0	0.0	0.0	1 (2.0%)	1 (2.0%)
Ethiopia	649/502	10/9	53.2	8.5	0.0	0.9	34.9	0.6	1.9	55 (8.3%)	116 (17.6%)
Gabon	1/1	0/0	100.0	0.0	0.0	0.0	0.0	0.0	0.0	0 (0.0%)	0 (0.0%)
Ghana	96/95	1/1	79.1	18.1	0.0	1.0	0.0	1.9	0.0	10 (10.3%)	0 (0.0%)
Guinea	67/57	0/0	98.8	0.0	0.0	0.0	0.0	0.0	1.2	0 (0.0%)	0 (0.0%)
Guinea-Bissau	1/1	0/0	0.0	0.0	0.0	0.0	100.0	0.0	0.0	0 (0.0%)	0 (0.0%)
Kenya	1432/1082	8/6	92.6	2.8	0.0	0.0	3.1	0.1	1.4	33 (2.3%)	19 (1.3%)
Lesotho	0	0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0	0
Liberia	100/72	1/1	1.9	24.1	0.0	0.0	68.5	0.0	5.6	0 (0.0%)	28 (27.7%)
Madagascar	550/473	0/0	13.4	0.3	0.0	0.0	29.6	0.6	56.0	9 (1.6%)	10 (1.8%)
Malawi	52/52	2/2	78.2	2.3	0.0	0.0	0.0	18.4	1.2	3 (5.6%)	1 (1.9%)
Mali	884/549	39/38	89.0	0.0	0.0	0.0	0.0	0.3	10.7	96 (10.4%)	84 (9.1%)
Mauritania	35/23	0/0	71.4	0.0	0.0	0.0	0.0	26.5	2.0	0 (0.0%)	0 (0.0%)
Mozambique	96/93	39/39	10.1	10.8	2.0	0.0	77.0	0.0	0.0	0 (0.0%)	0 (0.0%)
Namibia	38/32	0/0	0.0	0.0	26.3	0.0	73.7	0.0	0.0	0 (0.0%)	9 (23.7%)
Niger	231/204	0/0	75.8	0.0	0.0	0.0	17.5	1.2	5.6	1 (0.4%)	30 (13.0%)
Nigeria	840/772	12/12	81.8	4.9	0.0	0.3	6.2	1.5	5.3	55 (6.5%)	36 (4.2%)
Rwanda	1/1	32/32	97.0	0.0	0.0	0.0	0.0	0.0	3.0	0 (0.0%)	3 (9.1%)
Senegal	189/158	0/0	83.2	0.0	0.0	0.0	0.5	1.4	15.0	4 (2.1%)	1 (0.5%)
Sierra Leone	95/89	17/17	82.2	0.0	0.0	2.5	0.9	0.9	13.6	13 (11.6%)	20 (17.9%)
Somalia	10/9	0/0	0.0	50.0	0.0	0.0	30.0	20.0	0.0	0 (0.0%)	0 (0.0%)
South Africa	105/81	9/8	2.9	0.0	11.0	14.7	25.7	27.9	17.7	16 (14.0%)	27 (23.7%)
Sudan & South Sudan	343/287	5/4	64.1	0.5	0.0	0.0	0.0	0.0	35.3	34 (9.8%)	65 (18.7%)
Swaziland	1/1	0/0	0.0	0.0	0.0	0.0	0.0	0.0	100.0	0 (0.0%)	0 (0.0%)
Tanzania	308/265	9/8	68.7	10.0	0.3	1.3	7.4	4.2	8.2	26 (8.2%)	3 (0.9%)
The Gambia	5/5	0/0	20.0	80.0	0.0	0.0	0.0	0.0	0.0	0 (0.0%)	0 (0.0%)
Togo	1174/1102	21/19	99.1	0.3	0.0	0.0	0.4	0.1	0.1	1 (0.1%)	0 (0.0%)
Uganda	809/643	6/2	94.6	0.7	0.0	0.7	1.1	1.7	1.2	58 (7.1%)	17 (2.1%)
Zambia	136/124	2/2	64.2	9.1	0.0	0.0	15.2	5.5	6.1	14 (10.1%)	3 (2.2%)
Zimbabwe	422/403	8/5	52.6	36.2	0.0	7.7	0.7	0.9	1.8	25 (5.8%)	164 (38.1%)
Total	10831/8930	230/210	72.8	4.2	0.3	0.7	8.3	5.0	8.8	505 (4.6%)	781 (7.1%)

*Sedi: urine sedimentation; Filtr: urine filtration; Cen: urine centrifugation; RS: reagent strip; NS: not stated;**KK: Kato-Katz; Sedi: stool sedimentation; Filtr: stool filtration; Cen: stool centrifugation; Con: stool concentration; Othe: including FLOTAC, Faust's method and stool flotation; NS: not stated;***listed as number of surveys (percentage); [##]listed as number of survey/number of unique location; [#]administrative divisions of level two or three.

5.5.7 Variable selection using peNMIG spike-and-slab priors[#]

Variables	*Schistosoma haematobium*	*Schistosoma mansoni*
Group1[†]		
Annual mean temperature	-	-
Max temperature of warmest month	-	-
Min temperature of coldest month	-	-
Mean temperature of wettest quarter	-	-
Mean temperature of driest quarter	-	Selected
Mean temperature of warmest quarter	-	-
Mean temperature of coldest quarter	-	-
LST at night	Selected	-
Group 2[†]		
Mean diurnal temperature range	Selected	-
Isothermality	-	-
Temperature seasonality	-	-
Temperature annual range	-	Selected
Group 3[†]		
Annual precipitation	Selected	-
Precipitation of wettest month	-	-
Precipitation of wettest quarter	-	-
Group 4[†]		
Precipitation of driest month	-	Selected
Precipitation of driest quarter	Selected	-
Variables moderately correlated	-	-
Precipitation seasonality	-	Selected
Precipitation of warmest quarter	Selected	Selected
Precipitation of coldest quarter	Selected	Selected
LST in the day time	Selected	Selected
NDVI	Selected	-
Land cover	Selected	Selected
Elevation	-	Selected
Water distance	Selected	Selected
Climatic zones	Selected	-
pH measured in water	Selected	Selected
Soil moisture	Selected	Selected
Human influence index (HII)	-	-
Urban extents	-	-
Gross domestic product (GDP)	Selected	-
Infant mortality rates (IMR)	Selected	Selected
Proportion of improved sanitation	Selected	-
Proportion of improved drinking-water sources	-	Selected
Survey type	Selected	Selected

[#]Results are based on surveys in mainland sub-Saharan Africa; [†]a maxiumum of one variable can be selected in each highly correlated group.

5.5.8 Posterior summaries of the geostatistical model parameters for Madagascar

Schistosoma haematobium	Estimate (95% BCI)
Survey type (school-based)†	
Community-based	-0.56 (-0.68; -0.44)‡
Isothermality	0.22 (-0.42; 0.85)
Annual precipitation (≤1000 mm)†	
1001-1500	1.62 (0.04; 3.13)‡
>1500	1.39 (-0.39; 3.10)
Temperature seasonality	0.34 (-0.64; 1.24)
Max temperature of warmest month	0.40 (-0.20; 0.98)
Precipitation seasonality	1.53 (0.92; 2.17)‡
Climatic zone (Equatorial)†	
Arid	0.75 (-0.58; 2.02)
Warm	-1.88 (-3.48; -0.31)‡
Land cover (Grass)†	
Forest	1.50 (0.40; 2.62)‡
Shrub	-0.05 (-1.28; 1.19)
Urban	-6.99 (-33.16; 15.71)
Wet	1.54 (-0.85; 3.95)
Range (km)	89.94 (35.61; 201.56)
Spatial variance	2.16 (1.02; 4.35)
Non-spatial variance	1.86 (1.15; 2.94)
Schistosoma mansoni	Estimate (95% BCI)
Survey type (school-based)†	
Community-based	0.28 (0.17; 0.39)‡
Mean Temperature of Coldest Quarter	-0.22 (-0.84; 0.40)
Precipitation of coldest quarter (≤28 mm)†	
29-100	0.09 (-0.85; 0.68)
>100	0.53 (-0.69; 1.74)
Land surface temperature in the day time	0.49 (-0.03; 1.02)
Land surface temperature at night (≤15.0 ℃)†	
15.1-17.3	0.27 (-0.44; 0.97)
>17.3	-0.39 (-1.34; 0.57)
pH measured in water (≤5.15)†	
5.16-5.35	0.95 (0.13; 1.78)‡
>5.35	1.45 (0.27; 2.63)‡
Normalized differenced vegetation index (≤0.43)†	
0.44-0.60	-0.66 (-1.22; -0.11)‡
>0.60	-0.69 (-1.68; 0.30)
Climatic zone (Equatorial)†	
Arid	-0.81 (-2.53; 0.88)
Warm	0.66 (-0.20; 1.52)
Range (km)	143.11 (119.31; 172.59)
Spatial variance	3.20 (2.00; 5.20)
Non-spatial variance	2.13 (1.72; 2.63)

BCI=Bayesian credible interval. †baseline values are reported in brackets. ‡Important effect based on 95% BCI.

5.5.9 Proportion of locations included in the Bayesian credible interval (BCI) of various probability coverage cut-offs (Bar plots) calculated from the posterior predicted distribution, and the corresponding width of BCI (Solid lines)
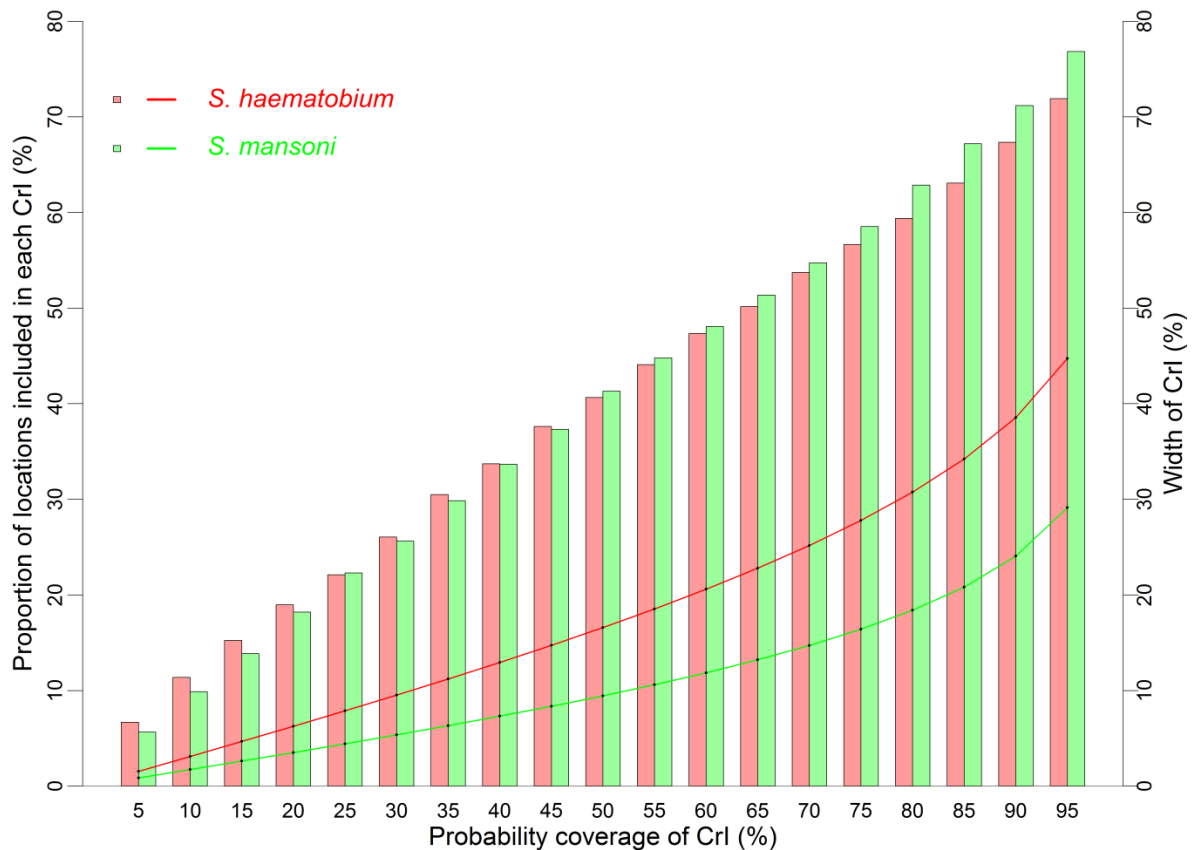
5.5.10 Population-adjusted prevalence (%) and number of individuals from the entire population ($\times 10^3$) infected with *Schistosoma* in 2012[*]

Country	Population	*S. haematobium*		*S. mansoni*		Schistosomiasis	
		Prevalence	Number infected	Prevalence	Number infected	Prevalence	Number infected
Angola	20163	21.1 (13.6; 33.6)	4262 (2747; 6769)	2.4 (1.0; 7.4)	487 (201; 1493)	23.5 (14.9; 36.4)	4739 (2995; 7331)
Benin	8772	28.3 (20.6; 37.9)	2478 (1807; 3324)	2.5 (1.0; 5.5)	220 (88; 483)	30.0 (22.4; 39.4)	2630 (1969; 3453)
Botswana	2016	9.4 (3.1; 24.6)	189 (62; 495)	2.8 (1.2; 6.9)	57 (24; 139)	12.4 (5.9; 27.7)	251 (118; 558)
Burkina Faso	17347	22.6 (19.4; 26.7)	3912 (3361; 4634)	0.8 (0.5; 1.4)	136 (82; 248)	23.3 (20.0; 27.5)	4042 (3473; 4764)
Burundi	8822	0.1 (0.0; 0.4)	7 (1; 39)	3.3 (2.3; 5.0)	290 (200; 445)	3.4 (2.3; 5.2)	302 (207; 457)
Cameroon	20335	7.3 (6.2; 8.8)	1484 (1259; 1779)	3.0 (2.3; 4.2)	615 (460; 848)	9.9 (8.6; 11.8)	2016 (1743; 2393)
Central African Republic	4434	7.7 (5.0; 11.4)	341 (222; 504)	8.8 (3.8; 21.6)	388 (166; 956)	15.8 (10.2; 28.5)	701 (454; 1265)
Chad	11752	22.0 (14.6; 31.7)	2585 (1711; 3726)	4.1 (1.5; 9.5)	479 (178; 1119)	25.4 (17.9; 34.8)	2989 (2103; 4094)
Congo	3830	9.0 (5.3; 16.8)	344 (202; 645)	3.4 (1.0; 10.8)	129 (39; 415)	12.6 (7.5; 21.5)	483 (287; 824)
Côte d'Ivoire	19967	6.7 (5.4; 8.5)	1337 (1068; 1690)	5.0 (4.0; 6.6)	994 (808; 1313)	11.5 (9.9; 13.6)	2297 (1971; 2713)

(Continues in next page)

(Continued from previous page)

Country	Population	*S. haematobium*		*S. mansoni*		Schistosomiasis	
		Prevalence	Number infected	Prevalence	Number infected	Prevalence	Number infected
DR Congo	68273	10.9 (7.9; 15.5)	7467 (5367; 10563)	12.8 (9.4; 17.0)	8768 (6422; 11574)	22.7 (18.6; 28.0)	15530 (12688; 19130)
Djibouti	951	7.6 (1.0; 47.8)	72 (9; 455)	10.7 (1.3; 60.7)	101 (13; 578)	24.0 (5.1; 66.2)	228 (48; 630)
Equatorial Guinea	721	1.0 (0.2; 3.8)	7 (1; 27)	1.1 (0.1; 10.3)	8 (1; 74)	2.5 (0.6; 11.7)	18 (5; 85)
Eritrea	5526	1.6 (0.6; 4.0)	87 (32; 223)	3.4 (1.7; 6.2)	189 (96; 340)	5.0 (3.1; 9.0)	275 (172; 497)
Ethiopia	86962	6.7 (3.6; 11.3)	5835 (3101; 9827)	4.7 (3.8; 6.1)	4077 (3305; 5273)	11.1 (7.9; 15.5)	9626 (6905; 13505)
Gabon	1542	12.1 (4.6; 35.6)	186 (71; 549)	2.0 (0.4; 8.3)	30 (6; 127)	14.3 (6.0; 37.4)	220 (92; 576)
Ghana	24375	18.1 (15.5; 21.3)	4415 (3777; 5203)	0.5 (0.2; 1.0)	120 (58; 252)	18.5 (15.9; 21.8)	4515 (3875; 5308)
Guinea	10079	9.7 (6.6; 13.7)	981 (664; 1384)	9.0 (5.9; 13.4)	907 (595; 1350)	17.5 (13.4; 23.1)	1765 (1355; 2328)
Guinea-Bissau	1523	18.0 (9.4; 39.8)	274 (143; 607)	0.6 (0.1; 7.0)	10 (1; 107)	19.0 (10.3; 41.1)	290 (157; 625)
Kenya	42146	8.2 (5.5; 12.3)	3439 (2334; 5173)	3.2 (2.5; 4.2)	1355 (1056; 1790)	11.3 (8.5; 15.1)	4772 (3585; 6346)
Lesotho	2179	3.4 (0.3; 20.7)	74 (6; 452)	1.3 (0.0; 16.1)	28 (1; 351)	6.0 (0.8; 27.1)	131 (16; 590)
Liberia	3841	13.3 (7.7; 27.8)	511 (297; 1066)	4.5 (2.5; 9.1)	174 (95; 350)	17.5 (11.3; 31.8)	671 (432; 1220)
Madagascar	21150	5.3 (4.1; 12.6)	1127 (868; 2667)	9.2 (7.8; 11.1)	1946 (1660; 2350)	14.4 (12.5; 20.9)	3040 (2648; 4427)
Malawi	15490	21.0 (17.1; 25.6)	3246 (2645; 3962)	3.7 (2.4; 5.9)	574 (365; 918)	24.1 (19.9; 29.1)	3729 (3083; 4511)
Mali	16116	26.0 (23.1; 29.4)	4198 (3724; 4731)	4.3 (3.4; 5.6)	692 (550; 909)	28.9 (26.0; 32.1)	4653 (4184; 5170)
Mauritania	3627	20.1 (15.4; 30.6)	730 (557; 1110)	0.9 (0.3; 3.5)	32 (12; 126)	21.0 (16.1; 31.4)	762 (583; 1137)
Mozambique	23667	43.4 (39.0; 48.1)	10283 (9228; 11395)	5.9 (4.0; 8.6)	1399 (958; 2037)	46.9 (42.5; 51.5)	11108 (10069; 12179)
Namibia	2403	6.3 (3.2; 14.2)	152 (78; 342)	1.5 (0.7; 4.8)	35 (17; 115)	8.0 (4.4; 15.9)	191 (106; 383)
Niger	16427	15.0 (12.8; 17.5)	2463 (2102; 2868)	0.2 (0.1; 0.4)	27 (15; 68)	15.2 (12.9; 17.6)	2491 (2119; 2890)
Nigeria	165415	18.0 (16.0; 20.5)	29733 (26488; 33858)	2.1 (1.5; 3.1)	3511 (2508; 5165)	19.6 (17.6; 22.1)	32421 (29058; 36554)
Rwanda	11014	0.0 (0.0; 0.1)	4 (1; 15)	1.7 (1.0; 3.1)	191 (109; 339)	1.8 (1.0; 3.1)	196 (113; 342)
Senegal	12681	14.1 (12.1; 16.7)	1788 (1538; 2122)	1.1 (0.6; 2.2)	134 (82; 282)	15.1 (13.1; 17.7)	1913 (1661; 2249)
Sierra Leone	5714	19.6 (14.1; 25.6)	1122 (805; 1463)	11.0 (8.2; 15.0)	630 (468; 860)	27.5 (22.3; 33.9)	1574 (1272; 1934)
Somalia	9478	21.2 (15.4; 28.8)	2007 (1459; 2734)	1.3 (0.4; 4.0)	126 (34; 380)	22.4 (16.5; 29.4)	2120 (1562; 2790)
South Africa	50110	12.4 (9.1; 17.3)	6202 (4583; 8658)	4.7 (2.6; 8.1)	2339 (1320; 4043)	16.4 (12.6; 20.8)	8204 (6331; 10412)
South Sudan	10567	15.4 (8.5; 24.6)	1629 (899; 2599)	8.8 (4.9; 13.9)	926 (519; 1464)	23.4 (15.3; 31.8)	2468 (1618; 3356)
Sudan	34188	16.9 (13.3; 21.3)	5787 (4557; 7294)	7.5 (5.3; 10.3)	2553 (1829; 3517)	23.7 (19.7; 28.1)	8110 (6737; 9610)
Swaziland	1216	15.5 (7.5; 31.9)	189 (91; 388)	6.4 (1.2; 27.3)	77 (14; 332)	22.6 (11.1; 41.8)	275 (136; 509)
Tanzania	46549	16.7 (14.2; 20.3)	7787 (6599; 9427)	4.4 (3.1; 6.6)	2040 (1433; 3073)	20.6 (17.8; 24.1)	9608 (8280; 11221)
The Gambia	1754	14.4 (8.5; 28.7)	252 (150; 503)	3.2 (0.9; 12.4)	57 (15; 218)	18.0 (11.1; 32.8)	316 (194; 575)
Togo	5863	15.8 (14.3; 17.6)	924 (837; 1031)	0.9 (0.7; 1.2)	52 (41; 72)	16.5 (15.0; 18.3)	970 (881; 1075)
Uganda	34841	3.7 (1.1; 9.6)	1278 (389; 3344)	5.7 (4.8; 6.8)	1986 (1670; 2381)	9.1 (6.5; 14.8)	3186 (2258; 5148)
Zambia	13916	19.8 (15.8; 24.8)	2756 (2199; 3452)	2.6 (1.4; 5.3)	356 (197; 741)	22.1 (17.8; 27.1)	3081 (2475; 3773)
Zimbabwe	13200	21.2 (19.1; 24.2)	2799 (2517; 3188)	3.8 (3.2; 4.7)	505 (420; 621)	24.1 (22.0; 27.0)	3175 (2906; 3563)
Total†	880941	14.6 (13.7; 15.5)	128190 (120475; 136466)	4.6 (4.2; 5.1)	40763 (37236; 45100)	18.5 (17.6; 19.5)	162988 (155254; 172211)

*Estimates are based on gridded population estimates in 2012; calculations are based on the median and 95% Bayesian credible interval of the posterior predictive distribution of the risk from 2000 onwards.

5.5.11 Bar plots with 95% Bayesian credible intervals for (A) population-adjusted prevalence, (B) number of infected individuals, and (C) praziquantel treatment needs in 2012 by each country of sub-Saharan Africa for school-aged children (5-14 years)

5.5.12 Bar plots with 95% Bayesian credible intervals for (A) population-adjusted prevalence, (B) number of infected individuals, and (C) praziquantel treatment needs in 2012 by each country of sub-Saharan Africa for entire population.

Chapter 6   Geostatistical-meta analyses of point-referenced and areal neglected tropical disease survey data

Ying-Si Lai[1,2] and Penelope Vounatsou[1,2]

[1]Department of Epidemiology and Public Health, Swiss Tropical and Public Health Institute, P.O. Box, CH-4002 Basel, Switzerland

[2]University of Basel, Petersplatz 1, CH-4003 Basel, Switzerland

Summary

Efforts for control and elimination of neglected diseases require maps of the geographical distribution of disease risk and spatially explicit estimates of the treatment needs. Geostatistical meta-analyses are often carried out on survey data compiled from bibliometric searches. These data often consists of combined point-referenced and areal aggregated surveys. In this paper, we developed a Bayesian geostatistical joint modeling approach that can analyse together areal and point-referenced survey data. We assumed that the point-referenced data arise from a binomial distribution and that the aggregated area data follow a Poisson binomial distribution which is approximated by a two parameter shifted binomial distribution. Results from simulated data shows that our proposed formulation has better predictive ability and provides more precise estimates of the model parameters compared to models that discard the area data or include them as points at the centroids of the areas. We have applied the new models to map the clonorchiasis risk in an endemic region in P.R. China.

Keywords: Poisson binomial, Bayesian geostatistical joint modeling, areal data, point-referenced data

6.1 Introduction

Neglected tropical diseases (NTDs) are a diverse group of communicable diseases prevalent widely in the world's poorest countries or regions of tropics and sub-tropics and affect more than one billion people (Hotez et al. 2007;Mackey et al. 2014;WHO 2010b). They had received little attention in the past, but over the last years there are a lot of efforts from national governments in endemic countries, global health initiatives, funding agencies, and philanthropists to control and potentially elimination of them (WHO 2013). Control efforts of the disease require maps of the geographical distribution of the infection risk and estimates of the number of infected people to help control programs to identify the high risk areas and assess the number of treatment needs. Geostatistical modeling applied on disease survey data is the most rigorous and commonly used approach for risk mapping of NTDs (Lai et al. 2015). These data are often extracted from published sources (peer reviewed publications and relevant reports) via bibliometric searches because single surveys covering large areas are not available. Publications may provide geographical information of the survey location or report survey data aggregated over an administrative level such as county or district. Therefore the analysis data consist of both geographical data types, point-referenced and areal data. Geostatistical analyses either ignore the areal data or treat them as point data, geo-referenced at the centroid of the area (Lai et al. 2013;Pullan et al. 2011). However, the areal data can provide useful disease information, especially when the spatial coverage of point-referenced data is low. Analyses treating the areal data as point referenced at the centroid of the area may improve model predictive ability but bias the estimates of the model parameters, particularly the spatial variation, since a uniform distribution of disease risk is considered within the area.

In the statistical literature the problem of combining areal and point data is known as the change of support problem (Jodar et al. 2015). Kriging is one of the approaches that has been proposed to address change of support problem for obtaining point estimates from both point and areal data arising from Gaussian, Poisson or binomial distributions (Goovaerts 2008a;Goovaerts 2008b;Velasco-Forero et al. 2009;Webster et al. 1994). In particular, Goovaerts employed binomial kriging to map breast-cancer incidence rates over a region of Michigan by geostatistical interpolation of observed individual-level and areal-level disease data (Goovaerts 2010). However, these methods are mainly based on interpolation of observed data that cannot take into account predictors (i.e., environmental and socioeconomic factors) which can play an important role on disease transmission and therefore improve predictions. Gelfand et al provided a unifying approach including fully Bayesian kriging for change of support problems, where areal data can be sensibly viewed as averages over point data and risk predictors can be incorporated into the geostatistical models (Gelfand, Zhu, & Carlin 2001). Smith and Cowles developed a Bayesian geostatistical model that combined point-level data with areal data by assuming that the latter was derived from an average of all point measurements within the area and formed a joint distribution framework of spatial random effects by defining an adjusted covariance function (Smith & Cowles 2007). They

further adopted a reparameterized and marginalized posterior sampling (RAMPS) algorithm to enhance the efficiency of MCMC sampling for model fit (Cowles, Yan, & Smith 2009;Smith, Yan, & Cowels 2008). Such models can be fitted for responses arising from Gaussian distributions and not survey risk data that are typically binomially distributed.

The goal of this research is to develop Bayesian geostatistical models that can jointly analyse both point-referenced and areal survey data that originally arise from binomial distributions and underlying the same spatial process. We view the total number of screened individuals within the area as a sum of independent Bernoulli variables that each one arises from a certain location and has a location-specific Bernoulli probability. We further assume that the total number of infected within the area follows a Poisson binomial distribution (Wang 1993). The probability density function of the Poisson binomial distribution is rather complicated and several approaches have been used to approximate it (Butler & Stephens 1993;Fernandez & Williams 2010;Pekoz et al. 2009;Pekoz et al. 2010). Approximations developed by Butler and Stephens, or Fernandez and Williams are relatively complex and difficult applicable in practical settings (Butler & Stephens 1993;Fernandez & Williams 2010), while approximations such as Gaussian, Poisson and binomial distributions are simple but can lead to large approximation errors (Pekoz et al. 2009). Instead, we employ the shifted binomial approximation with two parameters to match the first two moments of the non-identical location probabilities which has relatively low approximation error and can be easily applied (Pekoz et al. 2009). We applied the proposed models on simulated and real survey data of human clonorchiasis from P.R. China. In Section 6.2, we describe the disease survey data. In Section 6.3, we present the Bayesian geostatistical joint modeling framework. Simulation results are given in Section 6.4 and the results of the real application are shown in Section 6.5.

6.2 Data

The survey data which motivated this work are related to food-borne trematodiasis which is among the 17 core NTDs (Hotez et al. 2007). Clonorchiasis, caused by infection with Clonochis sinensis in human, is one of the most important food-borne trematodiasis in Asia (Qian et al. 2016). China accounts for over 85% of worldwide infected cases and the highest endemic regions include the neighboring provinces of Guangdong and Guangxi (Qian et al. 2012). We obtained clonorchiasis data from the open-access Global Neglected Tropical Diseases (GNTD) database (Hürlimann et al. 2011), which includes all available survey data that can be extracted from literature searches. Our data cover surveys carried out during the period of 2000-2015 in Guangdong and Guangxi provinces and includes 111 surveys at 100 unique villages (i.e., point locations) and 70 surveys with aggregated data within 49 counties (i.e., areas). Figure 6.1 shows the distribution of survey locations and areas and display the

observed prevalence data. The coverage of point-referenced data over the study region is low, as large areas of western and central parts have no point-survey data (Figure 6.1A).

In addition, over the study region, elevation data were obtained from the WorldClim (http://www.worldclim.org/current) at 1×1 km spatial resolution and population density data of the year 2010 were downloaded from the Socioeconomic Data and Applications Center (http://sedac.ciesin.org/) at 5×5 km.
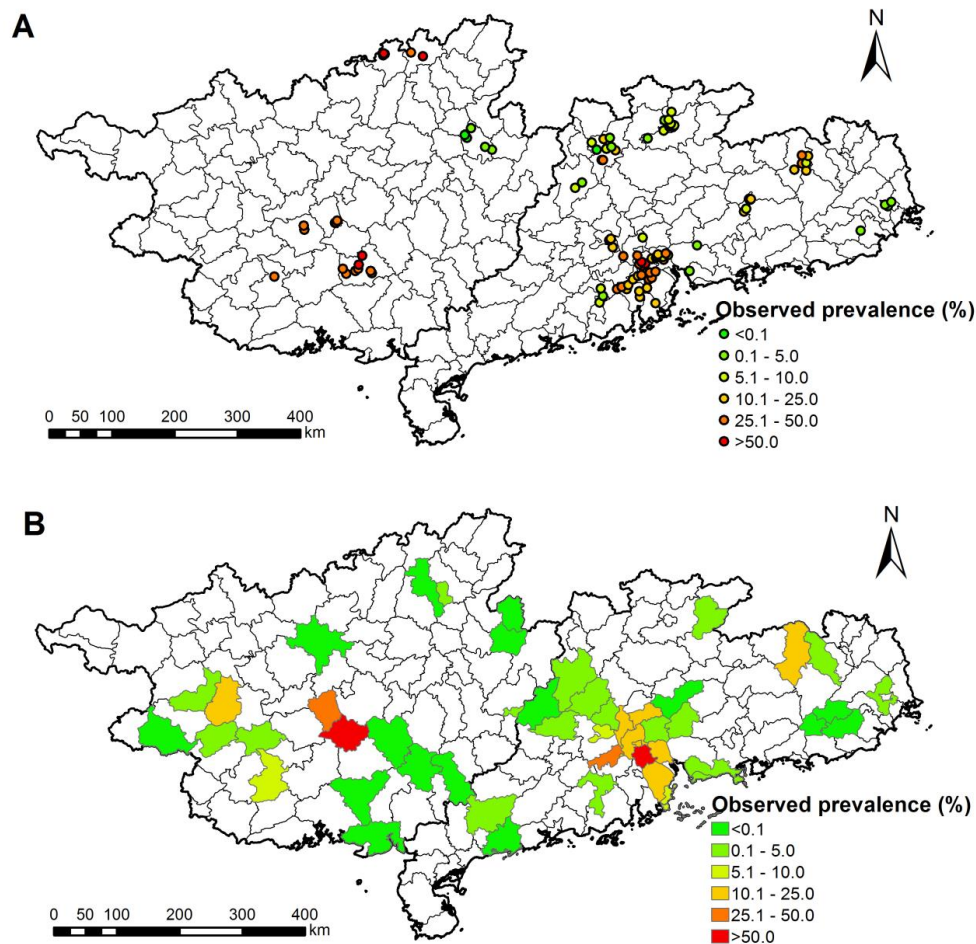
Figure 6.1: Geographical distribution of observed clonorchiasis survey data in Guangdong and Goangxi Provinces in P.R. China. (A) Point-referenced data and (B) Aggregated data at county-level (areal data).

6.3 Bayesian geostatistical modeling

6.3.1 Model specification

We assumed a number of latent points within the areas representing the survey locations. Surveys are more likely to be conducted in places with high population density, therefore we sampled the latent points according to their population density. We overlay a regular grid over the study region at 5×5 km spatial resolution. Let m_k indicates the number of latent points that sampled from the regular grid based on the population density in area k. We define p_i to

be the probability of infection at location i, where i is either location of point-referenced data or point sampled from areas with survey data (i.e. $i \in L \cup A$, where L and A contain only of the locations of the observed point-referenced data and of the sampled points within the areas with survey data, respectively).

Let y_i and n_i be the number of positive and the number of screened individuals, respectively, at location i, from the point-referenced data. We assume that y_i follows a binomial distribution, that is, $y_i \sim Bin(p_i, n_i)$ for $i \in L$. We define Y_k, and N_k to be the number of positive and the number of screened individuals, respectively within area k. We assume that Y_k follows a Poisson binomial distribution. Within each survey area, the numbers of screened people at the sampled points are considered to be the same. We apply a shifted binomial distribution with two parameters to approximate Poisson binomial distribution. According to formulations provided by Peköz (Pekoz et al. 2009), $Y_k \sim Bin(P'_k, N'_k)$, where $P'_k = \sum_{i \in A_k} p_i / m_k$ and $N'_k = [(\sum_{i \in A_k} p_i \times \sum_{i \in A_k} p_i / \sum_{i \in A_k} p_i^2) \times N_k / m_k]$. Here A_k contains only the points sampled from the regular grid in survey area k and [ ] indicates the function of round.

Following a standard Bayesian geostatistical model formulation, we model predictors on the logit scale of the infection probability p_i, that is, $\text{logit}(p_i) = \beta_0 + \sum_{s=1} \beta_s X_i^{(s)} + w_i$, where β_0 is the intercept, $X_i^{(s)}$ is the predictor s, β_s is the corresponding regression coefficient, and w_i is the location-specific random effect, respectively. We assume $\boldsymbol{w} \sim MVN(0, \Sigma)$ with an exponential covariance function $\Sigma_{ij} = \sigma^2 \exp(-\rho d_{ij})$, where d_{ij} is the Euclidean distance between locations i and j, and ρ corresponds to the rate of correlation decay.

6.3.2 Model implementation

Non-informative normal prior distributions were adopted for the regression coefficients (i.e., $\beta_0, \beta_s \sim N(0,100)$) and gamma priors were assigned to the precision parameter (i.e., $\tau^2 \sim G(0.01, 0.01)$, where $\tau^2 = 1/\sigma^2$) as well as the correlation decay parameter (i.e., $\rho \sim G(0.01, 0.01)$). The models were fitted using Markov chain Monte Carlo (MCMC) simulation in Winbugs version 1.4 (Imperial College London and Medical Research Council; London, UK) (Lunn et al. 2009). Two chains were run and convergence was assessed using the Brooks-Gelman-Rubin diagnostic (Brooks & Gelman 1998).

6.4 Simulation study

We carried out a simulation study to assess the parameter estimation and predictive ability of the proposed models compared to models which ignore or treat the areal data as points, referenced at the centroids of the areas. Furthermore, we assessed the sensitivity of the proposed model on the number of sampled points selected within the areas.

6.4.1 Simulation data

We generated 30 datasets following closely the structure of the survey data in our clonorchiasis application, which consists of surveys at 100 unique locations and 49 areas. In each area, we sampled 12 points (according to the population density of the study region) that represent the unobserved locations of the surveys. We generated a Gaussian process process $w = (w_{1,...,}w_S)^T$ over the S=688 locations (i.e., 100+12×49), setting $\sigma^2 = 2$ and $\rho = 5$ and considered a single covariate X that corresponds to the elevation predictor of our actual application. We assumed equal number of screened individuals at each location ($n_i = 500$, $i = 1, ..., S$) and simulated the number of infected ones from a binomial distribution $y_i \sim Bin(p_i, n_i)$, where $p_i = \text{logit}^{-1}(\beta_0 + \beta_1 X_i + w_i)$, setting $\beta_0 = -2$ and $\beta_1 = -2$. The number of infected and the total number of screened individuals at each area k were calculated as explained in Section 6.3, that is $Y_k = \sum_{i \in A_k} y_i$ and $N_k = \sum_{i \in A_k} n_i$. Simulations were carried out in R version 3.2.2 (R Foundation for Statistical Computing; Vienna, Austria).

6.4.2 Model validation

We fitted 6 models (A-F) that all use the point-referenced data at the 100 locations. Models A-C are standard geostatistical models. Model A includes the 12 points sampled within each area. Model B treats the areas as points, located at the corresponding centroid. Model C discards the area data and fits only the point-referenced ones. Models D-F are the proposed models considering $m_k = 2$, 5 and 8, respectively. We randomly selected 20% of the point-referenced and of sampled within areas points for model validation. The remaining data were used for model fit. The mean absolute error (MAE) and predictive log score were employed for model validation (Gneiting & Raftery 2007). Lower mean absolute error and higher log score suggest a model with better predictive ability. For each model parameter, we calculated the MAE and the proportion of datasets having the true parameter value within the 95% Bayesian credible interval (BCI) of the posterior distribution of the parameter (termed as proportion of inclusion). The proportion of datasets having all the true values of parameter within the 95% BCIs of their posterior distributions as well as the average width of 95% BCIs were calculated to assess the overall model performance in estimating the parameters.

6.4.3 Results

The mean absolute error and log score show similar patterns regarding to predictive ability of the six models (Figure 6.2). For visualising better the differences between the models in Figure 6.2A and 6.2C, we connected datasets of the same model with lines. As expected, the full model (Model A) has the best predictive performance (highest log score and lowest MAE). In contrast, Model C which fits only the point-referenced data, has the worst predictive ability (lowest log score and highest MAE). Our proposed joint models (D-F), show that the higher the number of points m_k sampled within the areas, the higher the model's predictive ability. However, the differences in the validation measures between
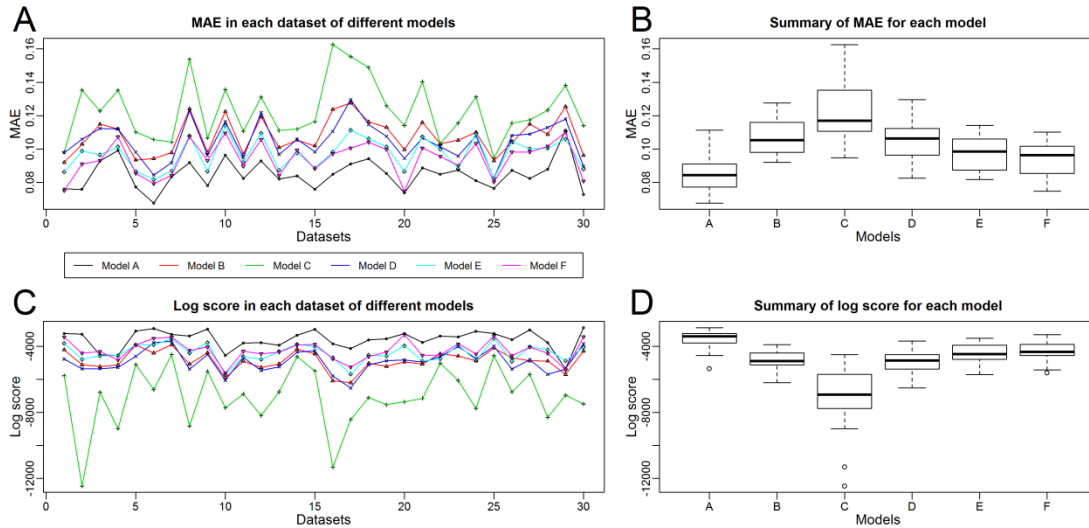
Figure 6.2: Predictive performance of models fitted on the 30 simulated datasets. (A) and (C) present the mean absolute error and log score for each dataset, respectively. (B) and (D) depict the distribution of the mean absolute error and log score over all datasets by each model.

Model E ($m_k = 5$) and F ($m_k = 8$) are small. Model B which treats areas as points located at their centroids shows similar predictive ability with the joint Model D that has $m_k = 2$.

Table 6.1 presents the evaluation results of parameter estimation from the six models. The regression coefficient β_1 and the spatial correlation decay parameter ρ were estimated best by Model A (highest proportion of inclusion, smallest mean width of the 95% BCI and smallest MAE), followed by Models F and E. Model C shows high proportion of inclusion, but the mean width of 95% BCI and MAE are both very large, suggesting that estimation uncertainty is high. The spatial precision parameter τ^2 ($=1/\sigma^2$) was estimated well by the Models A, F and E. Model B presents the lowest proportion of inclusion and highest MAE.

6.5 Application

Following the results of the simulation study, we fitted our proposed joint formulation on our clonorchiasis survey data, considering 8 points sampled within each area. Figure 6.3A and B show the predictive risk map of clonorchiasis and the corresponding prediction uncertainty. The model suggests that the high infection risk areas are concentrated in the western part of Pearl River Delta of Guangdong, in central Guangxi as well as in some small areas of northern Guangxi. Table 6.2 presents the posterior summaries of the model parameters. Results show that elevation has an important negative effect on the disease risk.

Table 6.1: Evaluation of the parameters estimates of the six models fitted on the simulation datasets.

	β_0	β_1	ρ	τ^2 $(1/\sigma^2)$	**All**
Proportion of datasets with 95% Bayesian credible intervals of estimated parameters including the true values of the parameters					
Model A	90	93.33	90	96.67	73.33
Model B	46.67	63.33	70	36.67	6.67
Model C	86.67	86.67	90	100	70
Model D	63.33	40	80	73.33	20
Model E	80	70	83.33	100	43.33
Model F	83.33	86.67	83.33	100	60
Mean width of 95% Bayesian credible intervals of estimated parameters					
Model A	0.6	0.56	2.48	0.22	
Model B	0.55	1.08	7.01	0.46	
Model C	1.07	2.1	9.11	0.49	
Model D	0.5	0.99	6.13	0.37	
Model E	0.58	1.08	3.99	0.32	
Model F	0.6	1.19	3.35	0.31	
Mean absolute error*					
Model A	0.2	0.16	0.82	0.07	
Model B	0.31	0.5	2.67	0.27	
Model C	0.36	0.63	2.73	0.14	
Model D	0.22	0.59	2.16	0.14	
Model E	0.22	0.47	1.42	0.09	
Model F	0.23	0.4	1.18	0.08	

*Calculated as the mean (over all simulated datasets) of the absolute difference between the true value of parameters and samples drawn from the posterior distribution of the corresponding parameter.

Table 6.2: Posterior summaries of the parameters obtained by fitting the join model on the clonorciasis survey data.

Parameter	Median (95% BCI*)
β_0	-6.27 (-6.69, -5.88)
β_1 (i.e., Elevation)	-1.11 (-2.40, -0.06)
ρ	2.66 (1.82, 3.65)
τ^2 $(1/\sigma^2)$	0.057 (0.036, 0.087)

*BCI indicates a Bayesian credible interval

6.6 Discussion

In this paper, we developed a Bayesian geostatistical joint modeling approach that can analyse together areal and point-referenced survey data. We assumed that the point-referenced data
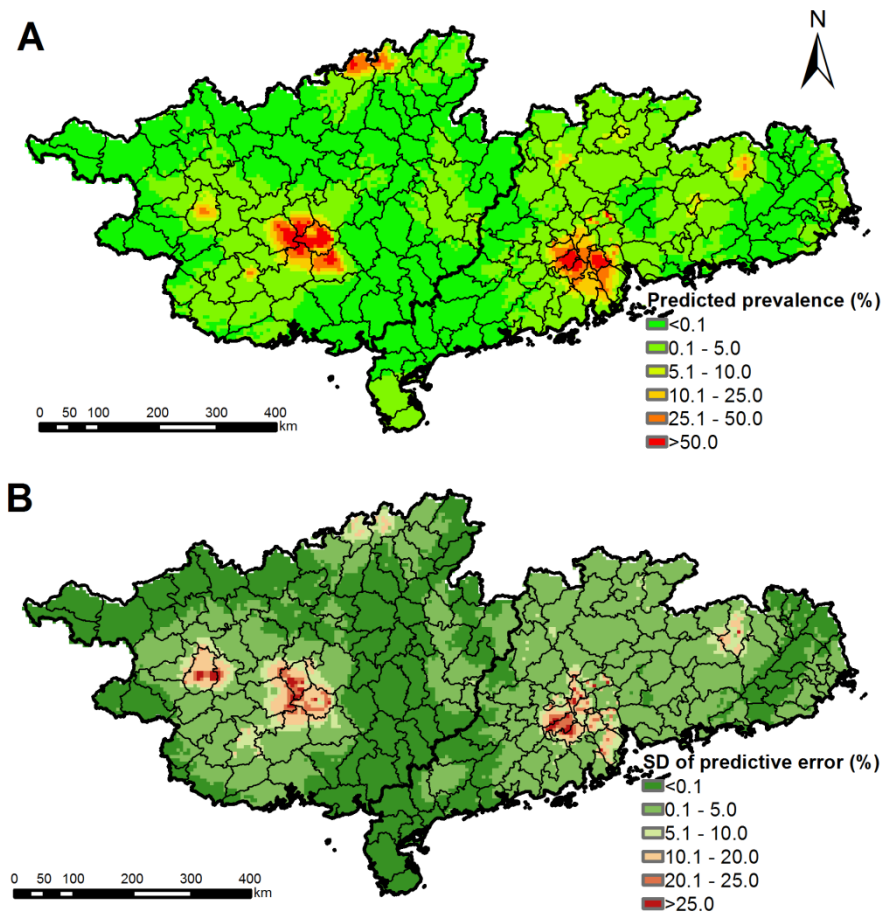
Figure 6.3: Infection risk estimates based on the chlonorciasis survey data. Median (A) and standard deviation (B) of the posterior predictive distribution of clonorchiasis risk.

arise from a binomial distribution and that the aggregated area data follow a Poisson binomial distribution that was approximated by a two parameter shifted binomial distribution. Results from simulated data shows that our proposed formulation has better predictive ability and provides more precise estimates of the model parameters compared to models that discard the area data or include them as points at the centroids of the areas. We have applied the new models to map the clonorchiasis risk in an endemic region in P.R. China, however our approach can be readily implemented to disease mapping of any survey data that have different spatial supports (i.e., points and areas).

Our geostatistical joint model introduces latent points within the areas and it tries to capture the variation of the risk within the areas using the disease risk-predictors relation that is estimated from the point-referenced data. As we have seen from the simulations, the number of latent points influences the performance of the model. However, the performance is not linear to the number of points because the improvement of performance between 5 and 8 latent points is smaller than that between 2 and 5 points. When using small number of latent points (i.e., 2), the predictive ability of the model is similar to the one treating areal as single point data, but the precision of the spatial parameter estimates in the join model is better.

Surveys often are conducted at locations with more residents, therefore it is sensible to select the locations of the latent points in the areas according to the population density.

The computational burden during the model fit and prediction increases with the number of latent points especially because the time required to inverse the correlation matrix of the Gaussian process increases in cubic order with the size of the matrix. Predictive process approximations of the Gaussian process (Banerjee et al. 2008;Banerjee et al. 2010;Finley et al. 2009) or approximate likelihood approaches for large geostatistical data can be applied to speed computations (Banerjee et al. 2014).

In conclusion, our Bayesian geostatistical joint modeling approach can be applied for analysing together areal and point-referenced survey data. This approach improves both the predictive ability and model parameter estimates compared to geostatistical regression models fitting only point-referenced data or treating areal data as single points. Considering larger number of latent points within the areas increases the model performance but also leads to high computational burden.

Acknowledgements

Chapter 7 Bayesian geostatistical modeling of age-heterogeneous *Schistosoma masoni* survey data in Côte d'Ivoire

Ying-Si Lai[1,2], Giovanna Raso[1,2], Jürg Utzinger[1,2] and Penelope Vounatsou[1,2]

[1]Department of Epidemiology and Public Health, Swiss Tropical and Public Health Institute, P.O. Box, CH-4002 Basel, Switzerland

[2]University of Basel, Petersplatz 1, CH-4003 Basel, Switzerland

Abstract

Background: Geostatistical model-based estimates of schistosomiasis risk are often based on survey data extracted through bibliometric searches due to the lack of single disease surveys covering large areas. These data are aggregated over age ranges that differ from one location to another. Infection risk is age-related however existing geostatistical models ignore the age-heterogeneity of the data leading to potentially biased estimates.

Methods: We integrated geostatistical and mathematical transmission models of schistosomiasis within a single model formulation and obtained age-specific estimates of the disease risk at high geographical resolution using *Schistosoma masoni* data from Côte d'Ivoire. Models took into account age-structured, water contact patterns or other characteristics of acquired immunity.

Results: A series of age-specific risk maps of *S. mansoni* infection in Côte d'Ivoire were produced. We predicted that the infection risk peaks at younger ages in high risk areas and at older ages in low risk areas. Furthermore, a more rapid decline rate of infection risk was observed at older ages in high risk areas compared to that in moderate and low risk areas.

Conclusions: We provide models that allow estimation of age-specific infection risk and of age-prevalence curves at high geographical resolution using compilations of age-heterogeneous survey data. The models can identify the most important age groups of the population to treat at a given place and evaluate interventions at population level from imperfect data.

Keywords: Bayesian geostatistics, *Schistosoma mansoni*, age-heterogeneous surveys, age-prevalence curve, immigration-death model, acquired immunity

7.1 Background

Schistosomiasis, one of the most prevalent neglected tropical diseases, affects more than 200 million people and causes a global burden of 3.3 million disability-adjusted life years (DALYs) (Murray et al. 2012;WHO 2010a). *Schistosoma masoni* is one of the two main species affecting African people (Colley et al. 2014). In Côte d'Ivoire, very high prevalence of *S. mansoni* was found in the western part while many other areas of the country remain at low risk (Lai et al. 2015;Yapi et al. 2014). The prevalence of schistosomiasis has an age-distinctive profile (French et al. 2010). A well-known pattern is that the prevalence rises in young children, reaches a peak during school age to early adulthood, then declines and becomes stable at a certain age level (Woolhouse 1991). This pattern may be explained by the age-related water contact activities and/or the development of the acquired resistance (immunity) (Warren 1973;Yang 2003).

Mathematical modeling can be used to estimate the age-prevalence curve of the disease based on observed prevalence data across a range of age groups. Hairston developed two-stage catalytic models to analyse age-prevalence data of schistosomiasis by extending the catalytic models introduced by Muench (Hairston 1965;Muench 1959). These models assume constant rates at which parasites are acquired or lost, respectively. Holford and Hardy employed an age-dependent immigration-death model to *Schistosoma* infection assuming that the worm immigration rate decreases monotonically with age according to Makeham's function (Holford & Hardy 1976). Chan *et al* adapted mathematical models with fully age-structured partial differential equations to model the transmission of human schistosomiasis (Chan et al. 1995). These models have been extended by other researchers to capture intervention effects (French et al. 2010;French et al. 2015;Zhang et al. 2007). Yang *et al* also proposed a semi-stochastic model to analyse the effect of acquired immunity on the relationship between prevalence and age, by introducing a fixed period of time after which human hosts build an immune response from the first infection (Yang 2003;Yang et al. 1997). Maximum likelihood was usually applied to estimate the model's parameters, however this approach has several shortcomings, such as not all the parameters can be estimated simultaneously and no confidence intervals can be calculated for the predicted age-specific prevalence. Raso *et al* developed a Bayesian formulation for Holford and Hardy's immigration-death model, which can draw inference for all parameters via Markov chain Monte Carlo (MCMC) simulation as well as obtain credible intervals of predicted age-specific prevalence of *S. mansoni* (Raso et al. 2007).

Over the last years, international efforts for control and elimination of schistosomiasis have been intensified. Georeferenced disease risk estimates and maps aid control programs by indicating appropriate interventions within a region based on its endemicity and WHO recommendations (Lai et al. 2015). Lai *et al* obtained high spatial resolution maps of schistosomiasis across sub-Saharan Africa using Bayesian geostatistical models and historical schistosomiasis survey data extracted from the open access global neglected tropical diseases

(GNTD) database (Hürlimann et al. 2011;Lai et al. 2015). Geostatistical models are widely used to predict disease risk at areas without observed data by relating survey data to potential predictors (e.g., environmental, climatic, and/or socioeconomic information) taking into account geographical dependence (Chammartin et al. 2013c). Historical data are compilations of age-heterogeneous survey data identified from peer reviewed literature and published reports and they are used in geostatistical modeling due to the lack of single disease surveys covering large areas. The age-heterogeneity can bias the estimation of the relation between the disease risk and its predictors. Mathematical models can be used to age-align the surveys, however there is no model formulation which allows changes of the shape of the age prevalence curve over space as a result of the varying endemicity.

In this work, we extended the Holford and Hardy's immigration-death model to allow geographical dependence of the age-prevalence curve and developed Bayesian geostatistical models that incorporate the immigration-death model to age-standardise the surveys data and obtain age-specific risk estimation. We implemented these models on age-heterogeneous *S. mansoni* survey data across Côte d'Ivoire. For the same purpose, we also proposed a Bayesian geostatistical acquired immunity model by combining Bayesian geostatistical model with an adaptation of Yang *et al*'s semi-stochastic model that takes into accounting the effect of acquired immunity on transmission. The performance of the models in fitting the data and in their predictive ability was assessed.

7.1 Methods

7.2.1 Data sources and data process

Cross-sectional age-specific *S. mansoni* infection surveys were extracted from the GNTD database (Hürlimann et al. 2011) for Côte d'Ivoire. A summary of the data sources derived from GNTD database as well as the diagnostic techniques are listed in Table 7.1. For each survey, we group individuals into age classes according to their ages: when individuals were younger than 20 years old, we considered each year of age to a separate class, while when individuals were equal or older than 20 years, we grouped every five years of age into a different class. We represented the 5-year age group by their mean age, weighted by the proportion of examined individuals in each year of age. We assumed that the disease prevalence becomes constant in older ages and grouped all individuals above 65 years in one category to overcome data scarcity for older individuals. We allocated survey data from individuals above 65 years in the last age class of 65. The survey year was treated as a binary indicator with a cut of at 2005.

Environmental and socioeconomic data obtained from remote sensing sources and other sources are shown in Table 7.2. In addition, gridded population of 2010 was obtained from WorldPop (http://www.worldpop.org.uk).

Table 7.1: Summary of *S. mansoni* survey data sources.

First author & publication date	Places of surveys	Year of surveys	Number of villages/schools	Age range	Diagnostics (Kato-Katz)	
					Samples per stool	Stool Specimens
Yapi et al, 2005	The savannah zone in the north & the forest zone in the west	1997-1999	45 villages	2-87	1	1
Raso et al, 2005	Mountainous region of Man	2001-2002	55 schools	1-20	1	1
Raso et al, 2004	The village of Zouatta II	2002	1 village	0-91	1	3
Rohner et al, 2010	15-20 km south of Toumodi	2006-2007	5 villages	5-17	2	1
Coulibaly et al, 2012	Azaguié	2010	4 villages (11 schools)	8-12	3	3
Coulibaly et al, 2013	Villages of Azaguié Makouguié and Azaguié M'Bromé	2011	2 villages	0-82	2	2
Becker et al, 2011	Léléblé	2009	1 village	0-75	2	1
Fürst et al, 2012	Taabo	2010	13 villages	18-87	2	1
Yapi et al, 2014	Across the country	2011-2012	92 schools	5-16	2	1

7.2.2 Geostatistical model

We assumed that Y_{ij} the number of positive individuals at location i and age class j, arises from a binomial distribution $Y_{ij} \sim Bin(P_{ij}, N_{ij})$, where N_{ij} and P_{ij} are the number of examined individuals and the probability of infection, respectively. We modelled the P_{ij} on the logit scale $logit(P_{ij}) = logit(k_{ij}) + w_j$, where k_{ij} is the infection risk explained by observed climatic, environmental, and socioeconomic factors that influence the distribution of worms within an individual of age class j, and w_j is the location-specific random effect introducing spatial correlation due to unknown spatially structure factors. We assumed a Gaussian process for w: $w \sim MVN(0, \Sigma_w)$, with a covariance function $\Sigma_{wiv} = \sigma_w^2 \exp(-\rho_w d_{iv})$, where d_{iv} is the Euclidean distance between locations i and v, and ρ_w corresponds to the rate of correlation decay of w.

7.2.3 Immigration-death model

According to Holford and Hardy's immigration-dealth model (Holford & Hardy 1976), $k_{ij} = (1 - \exp(-M_{ij}))^2$, under the assumption that the number of worms of either sex in an individual of age class j at location i follows a Poisson distribution with parameter M_{ij},

Table 7.2: Environmental and socioeconomic data sources[a].

Variable	Source	Data period	Temporal resolution	Spatial resolution
LST[b] on day time	MODIS/Terra[e]	2000-2012	8 days	1 km
LST at night	MODIS/Terra	2000-2012	8 days	1 km
NDVI[c]	MODIS/Terra	2000-2012	16 days	1 km
Land cover	MODIS/Terra	2001-2004	Yearly	1 km
Elevation	WorldClim[f]	2000	-	1 km
Precipitation	WorldClim	1950-2000	-	1 km
Precipitation Seasonality	WorldClim	1950-2000	-	1 km
Distance to water bodies	Calculation from locations to the nearst water bodies (SWBD[g])	2000	-	30 m
Soil moisture	Atlas of the Biosphere[h]	1950-1999	-	50 km
HII[d]	SEDAC[i]	1995-2004	-	1 km
Proportion of improved sanitation	Bayesian kriging of DHS, MICS, WHS, and LSMS[j]	1991-2012	-	5 km
Proportion of improved drinking-water sources	Bayesian kriging of DHS, MICS, WHS, and LSMS	1991-2012	-	5 km

[a]Land cover data accessed in June 2011 and other data asccessed in December 2015.
[b]Land surface temperature.
[c]Normalized difference vegetation index.
[d]Human influence index.
[e]Moderate Resolution Imaging Spectroradiometer (MODIS)/Terra, available at: http://modis.gsfc.nasa.gov/.
[f]Available at: http://www.worldclim.org/current.
[g]Shuttle Radar Topography Mission Water Body Data (SWBD), available at:
http://gis.ess.washington.edu/data/vector/worldshore/index.html.
[h]Available at: http://nelson.wisc.edu/sage/data-and-models/atlas/data.php?incdataset=Soil%20Moisture.
[i]Socioeconomic data and applications center, available at: http://sedac.ciesin.org/.
[j]Demographic and Health Surveys (DHS), available at: http://dhsprogram.com/; Multiple Cluster Indicator Surveys (MICS), available at: http://www.childinfo.org/mics_available.html; World Health Surveys (WHS), available at: http://www.who.int/healthinfo/survey/en/; and Living Standards Measurement Study (LSMS), available at: http://iresearch.worldbank.org/lsms/lsmssurveyFinder.htm.

where M_{ij} is determined by both the worm immigration rate $\lambda_i(\tau)$ at location i and age τ (in years) and the worm death rate δ, i.e. $M_{ij} = \int_0^{a_{ij}} \lambda_i(\tau)\exp(-\delta a_{ij})\exp(\delta\tau)d\tau$. Here a_{ij} is the weighted mean age of age class j at location i. δ is assumed constant across all locations and age classes. We considered a Makeham's function to model the age-dependent immigration rate, $\lambda_i(\tau)$ at location i, that is $\lambda_i(\tau) = A\tau^{r_1}\exp(-b_i\tau^{r_2}) + c_i$, where A, b_i, c_i are larger than zero, and r_1 and r_2 are two fixed parameters that affect the shape of the function. Holford and Hardy adopted $r_1 = 0$ and $r_2 = 1$ following the assumption that the water contact activities of an individual decreases with age (Holford & Hardy 1976). c_i is regarded as the baseline immigration rate. We assumed a common A across all locations and b_i was considered to location-specific, modelled on the log scale by a Gaussian process $\log(b) \sim MVN(m_b, \Sigma_b)$,

where $\Sigma_{biv} = \sigma_b^2 \exp(-\rho_b d_{iv})$, ρ_b corresponds to the rate of correlation decay of $\log(b)$. Environmental, climatic, and/or socioeconomic factors may have an important effect on *Schistosoma* transmission, therefore, we assumed that the baseline immigration rate c_i is determined by the factors mentioned above, that is $c_i = \exp(\beta_0 + \beta_t T_i + \sum_{s=1} \beta_s X_i^{(s)})$ (Model 1), where T_i is the category of survey year, $X_i^{(s)}$ is the sth covariate, β_0 is the intercept, and β_t and β_s are the corresponding coefficients, respectively. Besides Model 1, we explored an alternative formulations assuming that the average of immigration is determined by environmental, climatic, and/or socioeconomic factors, that is $\frac{1}{65} \int_0^{65} \lambda_i(\tau) d\tau = \exp(\beta_0 + \beta_t T_i + \sum_{s=1} \beta_s X_i^{(s)})$, giving $c_i = \exp(\beta_0 + \beta_t T_i + \sum_{s=1} \beta_s X_i^{(s)}) - \int_0^{65} A\tau^{r1} \exp(-b_i \tau^{r2}) d\tau$ (Model 2). Here we took an upper age limit of 65 years old due to sparse observed data for older individuals.

7.2.4 Acquired immunity model

Evidence indicates a protective resistance to reinfection of schistosomiasis (Colley et al. 2014), although it can take several years to develop (Fitzsimmons et al. 2012), and the acquired immunity is partial protective (Warren 1973). Yang *et al* modelled the acquired immunity by a semi-stochastic model assuming that human host builds up a partially effective and everlasting immune response after elapsing a fixed period of years from the first infection (Yang 2003). It is a semi-stochastic model as the distribution of worms among human population is treated stochastically while the demographic structure of human population is treated deterministic (Yang et al. 1997).

We employed the Yang *et al*'s model to assess its ability to fit the age-prevalence curve compared to immigration-death model and defined λ_i^0 and λ_i' the transmission rates among non-immune and immune individuals, at location i, respectively (Yang 2003). The model assumes that the transmission rates depend only on immunity and are constant across ages. We assumed that λ_i^0 is location-dependent and determined by environmental, climatic, and/or socioeconomic factors, that is $\lambda_i^0 = \exp(\beta_0 + \beta_t T_i + \sum_{s=1} \beta_s X_i^{(s)})$. As the acquired immunity is partially protective, we considered that the transmission rate among immune individuals is given by $\lambda_i' = \theta_i \lambda_i^0$, where θ_i is location specific and $0 < \theta_i < 1$, modelled on the logit scale by a Gaussian process $logit(\theta_i) \sim MVN(m_\theta, \Sigma_\theta)$, where $\Sigma_{\theta iv} = \sigma_\theta^2 \exp(-d_{iv} \rho_\theta)$. Following the formulation of Yang *et al*'s model, k_{ij} can be written as follows:

$$k_{ij} = \begin{cases} 1 - \exp(-\lambda_i^0(1 - \exp(-\delta a_{ij}))/\delta), & \text{for } a_{ij} \leq L \\ 1 - \exp(-\lambda_i^0(a_{ij} - L + (1 - \exp(-\delta L))/\delta)) - \int_0^{a_{ij}} f_1(\tau) f_2(\tau) d\tau, & \text{for } a_{ij} > L \end{cases},$$

(Model 3), where $f_1(\tau) = \lambda_i^0(1 - \exp(-\delta(a_{ij} - \tau)))\exp(-\lambda_i^0 \tau)$, and $f_2(\tau) = \exp(-\lambda_i^0/\delta(\exp(-\delta(a_{ij} - \tau - L)) - \exp(-\delta(a_{ij} - \tau)) + \lambda_i'/\lambda_i^0(1 - \exp(-\delta(a_{ij} - \tau - L)))))$. L corresponds to the fixed period of years from the first infection until the building up of the acquired immunity.

7.2.5 Practical implementation

We formulate a Bayesian framework for the above models to estimate the parameters as well as hyperparameters. To define the best value for r_1 and r_2 in Model 1, we considered a range of fixed values, $r_1 = 0,1,1.5,2$ and $r_2 = 0.5,1,2$ and chose the combination that fitted the model best. The deviance information criterion (DIC) showed that the optimal combination was $r_1 = 0$ and $r_2 = 1$ (Spiegelhalter et al. 2002). The prior distributions for model's parameters were adopted as follows: $\beta_0, \beta_t, \beta_s \sim N(0,1)$, $\sigma_w^2, \sigma_b^2, \sigma_\theta^2 \sim IG(2.01,0.99)$, $\rho_w, \rho_b, \rho_\theta \sim G(0.01,0.01)$, $m_b, m_\theta \sim N(0,1)$, $\delta \sim N(0,1)I(0,)$, $L \sim N(0,0.01)I(0,15)$. For Model 1, $A \sim N(0,1)$, while for Model 2, $A \sim unif(0, \min(F_i))$, where $F_i = 65\exp(\beta_0 + \beta_t T_i + \sum_{s=1} \beta_s X_i^{(s)})b_i/(1 - \exp(-65b_i))$, to make sure $c_i > 0$. MCMC simulation was employed to estimate the model parameters in Openbugs version 3.0.2 (Imperial College London and Medical Research Council; London, United Kingdom) (Lunn et al. 2009). Two chain samplers were run and convergence was assessed by Brooks-Gelman-Rubin diagnostic (Brooks & Gelman 1998). Bayesian variable selection was used to select the most important covariates present in Table 7.2 (Lai et al. 2013;Scheipl et al. 2012).

7.2.6 Validation and prediction

We randomly selected a subset of approximately 90% of locations for model fitting (training set) and subsequently assessed the model performance on the remaining 10% (test set). We selected the best model of the three according to the DIC and log score (Gneiting & Raftery 2007;Spiegelhalter et al. 2002) (i.e. the model with the lowest DIC and the highest log score measure). Mean error and the percentage of observations included in Bayesian credible intervals (BCI) of various probability coverages of predictions at the test locations were also calculated to evaluate the performance of the best model. Bayesian kriging was done to predict the *S. mansoni* infection risk for ages ranging from 1 to 65 years at the centroids of the pixels of a 5 × 5 km grid. We fitted the training set and produced the predictive age-prevalence curves for the test locations. We calculated the median and 95% BCI of the posterior distribution of the population-adjusted mean predictive distribution of the prevalence of all pixels for each age. We also grouped the pixels into low, moderate and high risk groups according to the peak of the median predictive distribution of the prevalence in each pixel across all ages. Low, moderate, and high risk groups were defined as peak prevalence less than 10%, between 10% and 50%, and higher than 50%, respectively, and the age of peak prevalence in each risk group was calculated.

7.3 Results

7.3.1 Data description

The age-heterogeneous data derived mainly from 9 references (Table 7.1), that included 218 surveys at 215 locations. 46% of surveys were carried out between 1997 and 2004 and the rest were from 2005 onwards. The diagnostic technique was Kato-Katz for all surveys, but with different numbers of stool specimens and samples per stool. Figure 7.1A shows the locations and overall observed prevalence at each location. 74.4% of surveys were carried out among young individuals (children and adolescents) with age bellow 20 years, 5.1% of surveys included adults (equal or older than 20 years), while the remaining 20.5% of the surveys included both children and adults (Figure 7.1B). Elevation and normalized difference vegetation index (NDVI) were selected during the variable selection process.
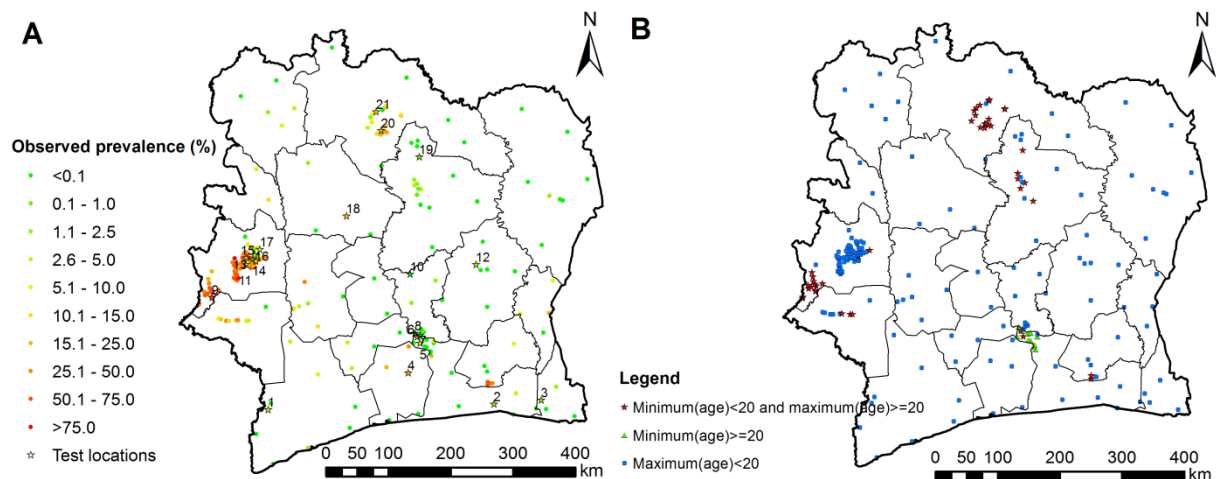
Figure 7.1: Survey locations of observed survey data of *S. mansoni* across Côte d'Ivoire. (A) Overall observed prevalence and locations with indexes for test set; and (B) locations of surveys carried out among both children and adults (with minimum age younger than 20 and maximum age equal to or older than 20), among only adults (with minimum age equal to or older than 20) and among young individuals (with maximum age younger than 20).

7.3.2 Model selection and parameter summaries

Table 7.3 lists the posterior summaries of the three Bayesian geostatistical models. Large spatial variations of infection risk (σ_w^2) and small spatial decay (ρ_w) parameters were estimated from all the three models. Model 2 had the lowest DIC and the largest log score (i.e. best model). The best model estimated a decrease of the average transmission rate from 2005 onwards. Elevation had a negative effect while NDVI had a positive effect on transmission. The model may under-estimate the infection risk as the mean error is larger than zero. It was also able to correctly predict the age-specific infection risk (within the 95% BCI) at the 74.2% of the test locations.

Table 7.3: Posterior summaries of the parameters of the three models.

Parameters		Model 1	Model 2	Model 3
		Posterior estimate (median & 95% Bayesian credible interval)		
Year	<2005	0	0	0
	≥2005	-0.44 (-0.79; -0.29)	-0.19 (-0.29; -0.09)	-0.39 (-0.7; -0.1)
Elevation		-0.13 (-0.20; -0.06)	-0.09 (-0.13; -0.03)	-0.57 (-0.72; -0.48)
NDVI		-0.07 (-0.15; 0.05)	0.07 (0.01; 0.11)	0.1 (0.04; 0.16)
ρ_w		0.72 (0.69; 0.85)	0.71 (0.69; 0.80)	0.73 (0.69; 0.89)
σ_w^2		19.46 (14.09; 27.89)	16.70 (12.11; 22.53)	24.59 (19.26; 32.34)
δ		0.15 (0.13; 0.17)	0.16 (0.14; 0.19)	0.05 (0.03; 0.06)
m_b		-1.63 (-1.99; -1.30)	-1.81 (-2.70; -1.13)	-
ρ_b		4.71 (1.55; 13.21)	86.00 (17.78; 307.00)	-
σ_b^2		4.05 (2.04; 11.83)	4.64 (2.02; 10.37)	-
A		0.74 (0.70; 0.85)	0.72 (0.63; 0.80)	-
m_θ		-	-	0.02 (-0.23; 0.71)
ρ_θ		-	-	12.07 (3.37; 76.96)
σ_θ^2		-	-	0.43 (0.23; 1.29)
L		-	-	7.66 (6.66; 8.39)
Model validation				
DIC		3990	3706	4053
Log score		-308.68	-293.78	-300.75

7.3.3 Age-specific risk prediction

Figure 7.2 presents a series of predictive age-specific risk maps of *S. mansoni* infection from age 2 to 65 in Côte d'Ivoire. All the maps show that the western part and some small areas in southern and eastern parts have higher risk of *S. mansoni* infection compared to other parts of the country. In high risk areas, the infection risk increased significantly with age until around 15 to 17 years old and then decreased subsequently and became stable at older ages (after 45). Other areas with lower infection risk also show a similar trend (i.e., an initial increase and then decrease after a certain age), but the changes were minor. The prediction uncertainties of age-specific infection risk (Figure 7.3) show an increase with age. High uncertainties were mainly in the western and southern parts of the country.

The geographical distribution of the 21 test survey locations is shown in Figure 7.1A. The predictive age-prevalence curves and the corresponding observed data are depicted in
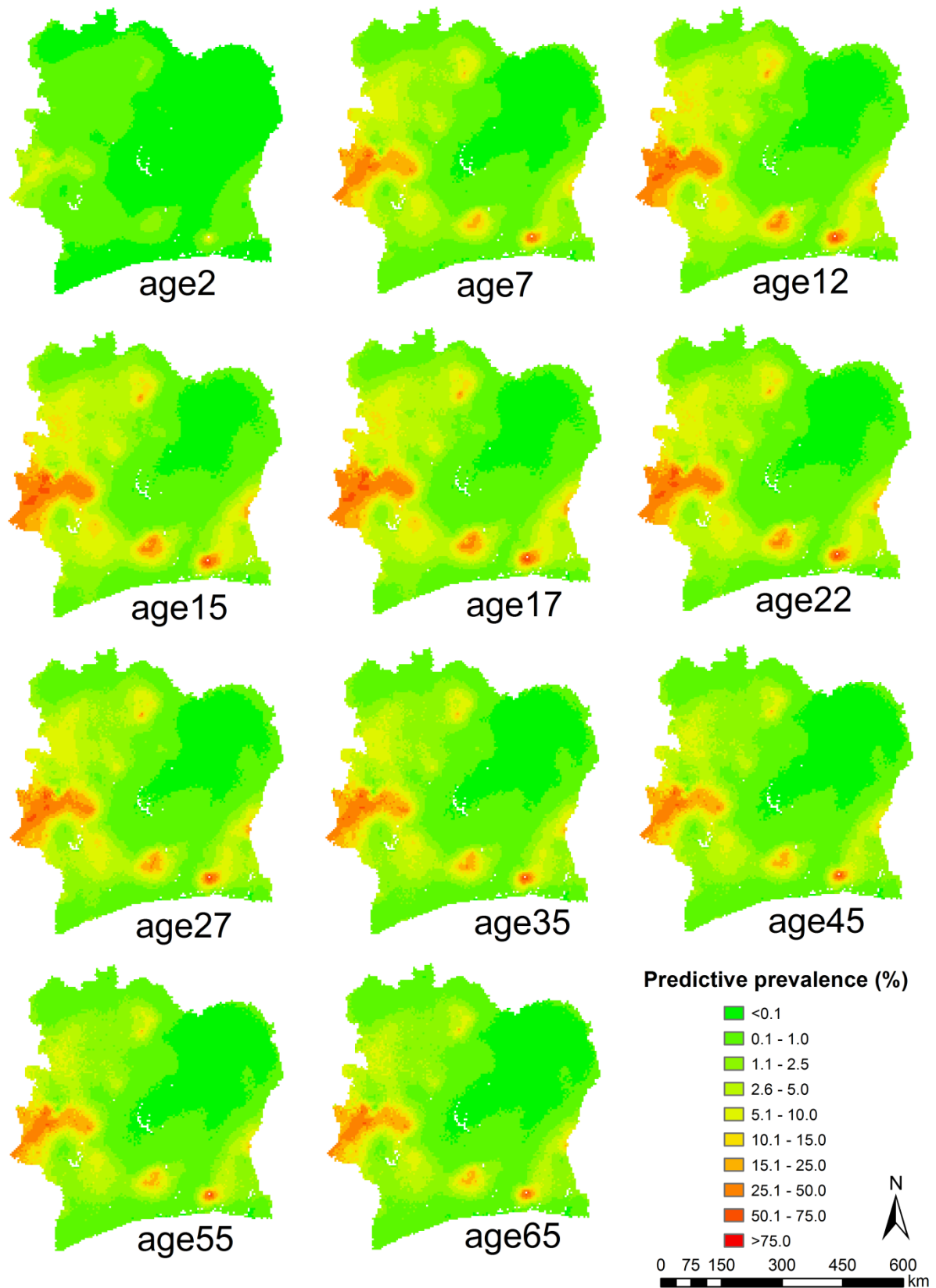
Figure 7.2: The geographical distribution of age-specific *S. mansoni* prevalence in Côte d'Ivoire. The maps show the risk estimates from 2005 onwards based on the median of the posterior predictive distribution of infection risk.
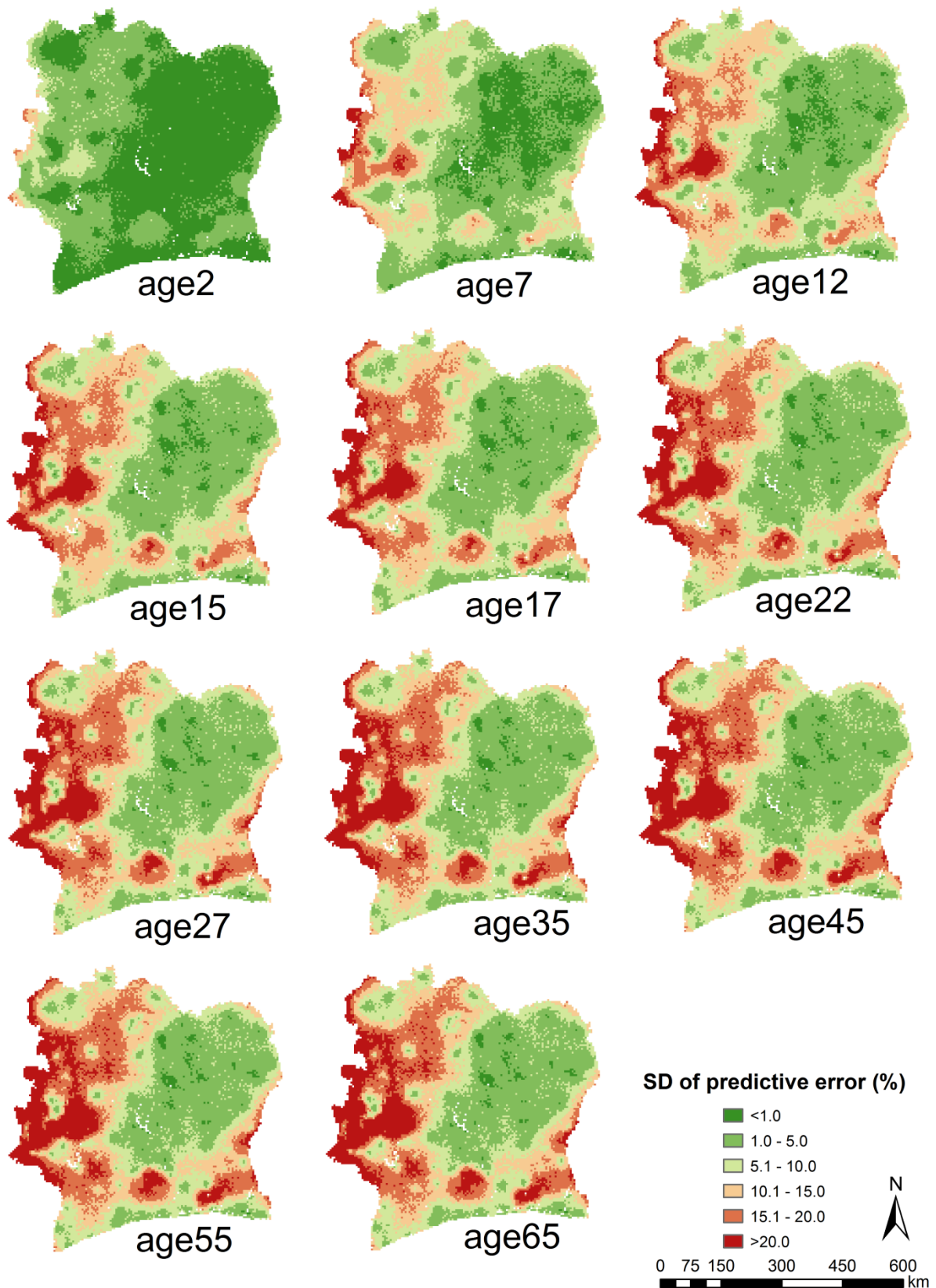
Figure 7.3: The geographical distribution of age-specific prediction uncertainty of *S. mansoni* infection risk in Côte d'Ivoire. The maps are based on the standard deviation of the posterior predictive distribution of infection risk from 2005 onwards.

Figure 7.4 for each one of the 21 test locations. Thirteen out of the 21 locations had observed data only in school-aged children, while 4 out of the 21 locations (i.e., Location 5, 19, 20 and 21) had observed data within full age range. At Location 5, very low prevalence was observed in young children while low to moderate prevalence was observed in older ages. The median of the predictive distribution of the age-prevalence curve shows a slow increase until twenties, then a slow decrease and maintained stable at older ages. At Location 19, most of the observed prevalence was very low and the median of the predictive age-prevalence curve
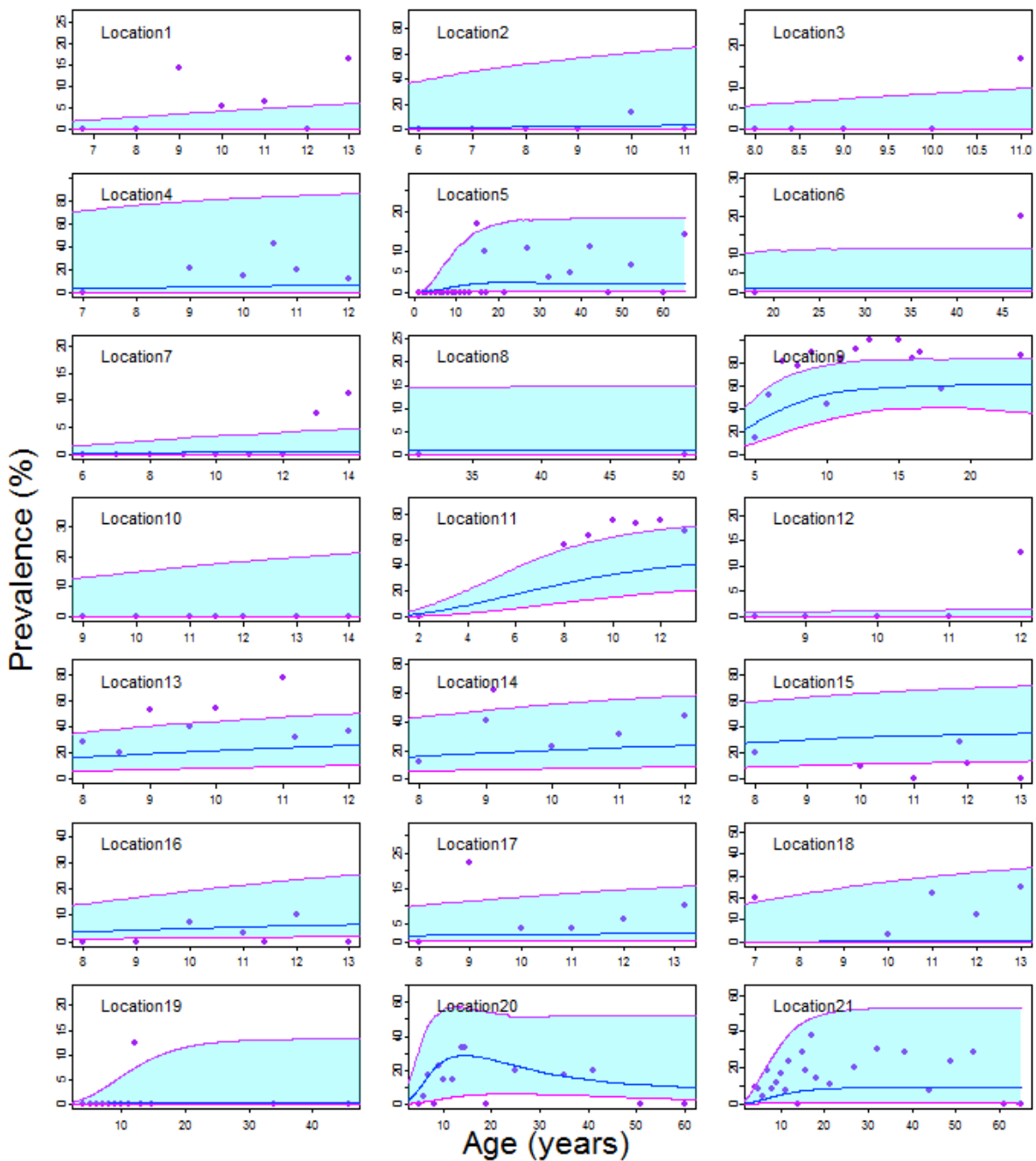
Figure 7.4: The predictive age-prevalence curves of *S. mansoni* at test locations. Purple dots, blue lines and light blue areas indicate observed data, median and 95% Bayesian credible interval of posterior predictive prevalence, respectively.

shows a very low and stable prevalence with age. At Location 20, the observed prevalence first increased with age, reached a peak at around 15 and then decreased, where the predictive age-prevalence curve shows a very similar pattern. At Location 21, the observed prevalence increased with age until around 20, then it remained between 10% and 40%, and dropped down to around zero after the age of 60 years. The median of the predictive distribution of the age-prevalence curve shows an increase until the age of 20 and it is maintained stable with prevalence around 10% afterwards. For all predictive age-prevalence curves, the widths of the 95% BCIs increase with age.

Figure 7.5A shows the age-specific, population-adjusted mean predictive prevalence curve of *S. mansoni* over Côte d'Ivoire. The mean prevalence increased with age, reached the peak at around 20 years and decreased slowly afterwards. In high risk areas, the mean prevalence reached the peak at around 16 to 17, and for moderate risk areas, the peak moved to around 18 to 19, years. In low risk areas, the peak was reached after 20 years and the prevalence became quite stable with just a very slow decrease afterwards (Figure 7.5B).
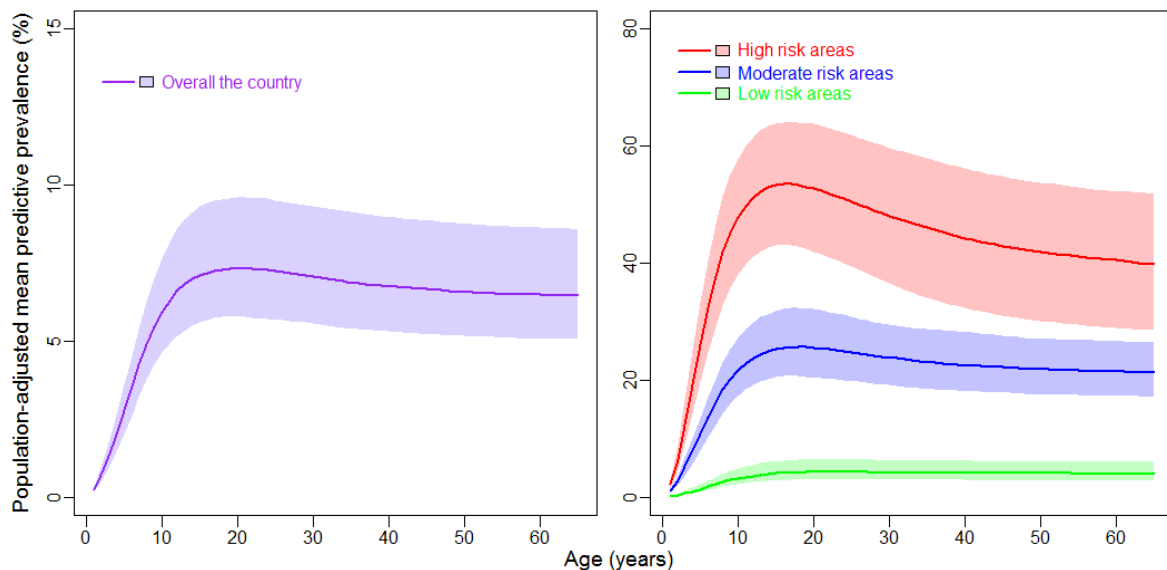
Figure 7.5: Age-population-adjusted mean predictive prevalence curves of *S. mansoni* in Côte d'Ivoir. (A) For the whole country and (B) for high, moderate and low risk areas. Solid lines and light-color areas indicate the median and 95% Bayesian credible interval of posterior distribution of population-adjusted mean predictive prevalence, respectively.

7.4 Discussion

In this study, we integrated geostatistical and mathematical transmission models of schistosomiasis within a single model formulation and obtained the first age-specific estimates of the disease risk at high geographical resolution from age-heterogeneous surveys. Previous researchers studying age-prevalence patterns of schistosomiasis have mainly focussed on fitting transmission models for a single survey and the results were difficult to

extend to other areas as the conditions of transmission were different (Chan et al. 1995;Hairston 1965;Holford & Hardy 1976;Muench 1959;Raso et al. 2007;Yang 2003). We assumed that the transmission of *S. mansoni* is determined by potential factors (e.g., environmental, climatic, and/or socioeconomic) and embedded modified transmission models into Bayesian gestatistical model to address age-heterogeneity. Three models were developed using *S. mansoni* survey data from Côte d'Ivoire based on age-structured, water contact patterns or characteristics of acquired immunity, and the best one was selected according to the model performance. We assumed that the water contact or acquired immunity patterns were similar in geographically close locations and introduced the spatial specific parameters, thus were able to predict the age-prevalence curves at locations without any survey data or without survey data of full age range.

We predicted that the infection risk peaks at younger ages in high risk areas and at older ages in low risk areas (Figure 7.5B). This is consistent with the "peak shift" pattern, i.e., the peak age of infection shifts towards younger ages in areas of high transmission (Woolhouse 1998). Furthermore, Figure 7.5B shows a more rapid decline rate of infection risk at older ages in high risk areas compared to that in moderate and low risk areas, which confirms the observations of other researchers (Butterworth et al. 1988). WHO recommends that mass drug administration for schistosomiasis should be implemented to all school-aged children in endemic areas, with treatment frequency according to the prevalence levels in school surveys; while access to praziquantel for passive case treatment is recommended for community-based interventions (WHO 2002a). In moderate and low risk areas, as the infection risk declines quite slowly after the peak age and then remains stable, mass drug administration that focusses only at school-aged children may not have a sufficient impact on the community-wide parasite transmission (Anderson et al. 2013;Lelo et al. 2014), thus hard to achieve transmission elimination.

We estimated a decrease of transmission from 2005 onwards, which may be attributed to the control programmes of schistosomiasis in Côte d'Ivoire in recent years (Rollinson et al. 2013). We found a negative effect of elevation on transmission, that is consistent with previous studies showing a low infection risk in high elevation areas in Côte d'Ivoire (Assare et al. 2015;Beck-Worner et al. 2007;Raso et al. 2005). We assumed a linear relation between the climatic predictors (elevation and NDVI) and the infection risk. However this assumption may not hold, as environmental factors influence transmission in a more complex way by affecting the snail habitat, the human activities, and cercaria maturation (Liang et al. 2007). Furthermore, many other factors may have a more important and direct effect on transmission, the distribution of which we cannot obtain, such as the infected snail density. These may partly explain the large spatial variation of the predictive infection risk.

The majority (74.4%) of the surveys were carried out in young individuals while only one fifth of the surveys included population of full age range. The small number of older age individuals may also contribute to the high prediction uncertainties in older ages and explain

partly the large spatial variation of the predictive infection risk. A systematic review of surveys pertaining schistosomiasis prevalence data in sub-Saharan Africa (Lai et al. 2015) shows that the age distribution in the observed survey data from Côte d'Ivoire reflects that observed in many countries of the sub-continent. We suggest therefore more community-based surveys to be carried out in the future covering a full range of ages in order to better understand and predict more precisely the age-prevalence patterns across different areas in the countries. Moreover, evidence suggest that deworming that only focusses on school-aged children may not be sufficient for *S. mansoni* elimination (Lelo et al. 2014;Lo et al. 2015).

Our best model (Model 2) considered a location-specific, age-dependent worm immigration rate, the average of which was assumed to be determined by environmental, climatic, and/or socioeconomic predictors. As the function for the age-dependent worm immigration rate is based on the assumption that the water contact activities decrease with age (Holford & Hardy 1976), this model emphasized the effect of water contact on the age-prevalence pattern of *S. mansoni*. Many researchers emphasized that development of acquired immunity can be a better interpretation for the "peak shift" pattern or the rapid decline rate of infection in high risk areas (Butterworth et al. 1988;Woolhouse 1998). In our study, however, the Model 3 that addressed the important effect of acquired immunity on age-prevalence pattern did not improve the fit of the data. Development of a Bayesian geostatistical model embedded both water contact and acquired immunity patterns may take into account both effects on age-prevalence pattern. However, such model may be too complex and computationally expensive.

Our models assumed a unique peak of prevalence in age-prevalence pattern of *S. mansoni*. However, a few researchers reported a second peak of prevalence or intensity may exist in older age groups, probably due to an increase of water contact activities resulting from occupational change or a decay of acquire immunity (Mutapi et al. 2003;Raso et al. 2007). The second peak might be captured by modifying the age-dependent immigration rate function in Model 1 and Model 2, or by addressing the lasting of acquired immunity for a certain duration instead of lifelong (Chan et al. 1996). Model 1 and Model 2 assumed that the distribution of the number of worms of either sex within an individual follows a Poisson distribution and that the exposure to infection is homogeneous for all individuals in the same age class (Holford & Hardy 1976). However, heterogeneity in exposure exists for schistosomes and leads to an overdispersed distribution in host population, that is, the majority of hosts are uninfected or lightly infected while a minority of them have heavy infections (Anderson & May 1985;French et al. 2010). The distribution of worms in such situations is usually assumed to be negative binomial determined by an aggregation parameter and mean worm burden) (Chan et al. 1995). The aggregation parameter can differ across different levels of infection intensity and across different locations (French et al. 2010). Without knowing both the observed infection intensity and prevalence, it is difficult to estimate the aggregation parameter, thus we use Poisson distribution instead in our models.

Our *S. mansoni* data extracted from different surveys that used the Kato-Katz diagnostic technique. However, the numbers of stool specimen and sample per stool differ among surveys. Even though it is known that the sensitivity of Kato-Katz is affected both by the number of specimens and the level of infection (Carneiro et al. 2012;Lamberton et al. 2014), it is difficult to take into account the diagnostic sensitivity into the modeling because many of the surveys report aggregated data and do not provide details on the number of sampling efforts.

7.5 Conclusion

We developed Bayesian geostatistical models to analyse age-heterogeneous *S. mansoni* prevalence data in Côte d'Ivoire. Age-specific infection risk was estimated and age-prevalence curves were predicted across the country. These estimates are important for planning control programmes in areas of different endemicity, targeting different age groups. The models can be applied to other *Schistosoma* species, such as *S. haematobium* and *S. japonicum*.

Acknowledgements

Chapter 8　Discussion

The work of this PhD thesis contributes to the fields of spatial statistics and of epidemiology of NTDs with (i) statistical methodology for modeling spatially-structured disease data, having heterogeneous geographical support (i.e., georeferenced at point or area level) across the study region and they are collected over different age groups between locations, (ii) applications on soil transmitted helminth infections, schistosomiasis, and clonorchiasis in sub-Saharan Africa, South Asia, and P.R. China, to obtain spatially explicit estimates of disease risk, number of infected people, and annual treatment needs for preventive chemotherapy at different administrative levels, and (iii) large amount of geo-referenced data on NTD surveys conducted at over 10,750 unique locations that are available via the open access GNTD database. Six manuscripts were produced and included in the thesis as chapters, where the detailed methodologies, results and discussions are presented. This section highlights the main contributions of this research and the significance of the findings, discusses the limitations of the methods, and proposes extensions of the work.

8.1 Significance

8.1.1 Spatial statistics: methodology for survey data heterogeneous in space

Estimates of disease risk at high spatial resolution over large geographical areas are often rely on historical data that are extracted from publications due to the lack of single surveys covering the whole study regions. These data are either reported at the survey location (point-referenced) or they are aggregated over several locations within an administrative level such as county or district (areal data). Spatial analyses often discard the areal data or treat them as point referenced at the centroid of the area. However the areal data can provide useful information especially when the spatial coverage of point-referenced data is low. Furthermore, treating areal data as point-referenced may improve model prediction ability but bias the estimates of the model parameters, particularly the spatial variation, since a uniform distribution of disease risk is considered within the area. Smith and Cowles (2007) developed a Bayesian geostatistical model that combined point-reference data with areal data by assuming that the latter was derived from the average of all point measurements within the area and formed a joint distribution framework of spatial random effects by defining an adjusted covariance function (Smith & Cowles 2007). However, such model only fits responses arising from a Gaussian distribution and not survey data that are typically binomially distributed.

In Chapter 6 we developed a Bayesian geostatistical joint modeling approach that analyses together the areal and point-referenced data arising from a binomial distribution. As the actual survey locations cannot be identified for the areal data, we assumed a number of latent points within the areas that represent the survey locations of the area. The responses at these points were assumed to arise from a binomial distribution and share the same spatial process as the

available point-referenced data in the study. In addition, we assumed that the sum of the binomial responses at the latent points follow a Poisson binomial distribution which was approximated by a shifted binomial distribution with two parameters (Pekoz et al. 2009;Wang 1993). Our modeling approach shows a better prediction performance and improved parameter estimation on simulated data compared to that of previous approaches. We have applied the methodology to estimate the clonorchiasis risk in an endemic region in P.R. China, however our approach can be readily applied to analyse any other neglected tropical disease survey data.

Historical disease survey data are often aggregated over heterogeneous age groups between survey locations. Geostatistical analyses of NTDs ignore age-heterogeneity and treat data as if they come from a common age group. This approach can lead to biased estimation because models cannot distinguish whether different risk between locations is due to differences in age or to exposures. Mathematical models can be used to age-align the surveys, but there is no model formulation allowing changes of the shape of the age-prevalence curve over space as a result of the varying endemicity.

In Chapter 7 we integrated geostatistical and mathematical transmission models within a single model formulation to analyse age-heterogenous survey data of *S. mansoni* in Côte d'Ivoire. Firstly, Holford and Hardy's immigration-death model and Yang *et al*'s semi-stochastic model were extended to allow geographical dependence of age-prevalence curve. The former model estimates the age-prevalence curve based on water contact patterns, while the later one is based on characteristics of acquired immunity (Holford & Hardy 1976;Yang et al. 1997). These models were further incorporated into Bayesian geostatistical modeling to age-standardise the surveys data. The assessment of model performance in fitting the data and in prediction showed that the best model was the one adopted a location-specific age-dependent worm immigration rate derived from Holford and Hardy's immigration-death model, where the average of the immigration rate was assumed to be determined by the risk predictors such as climatic and environmental factors.

8.1.2 Epidemiology: implications for disease control

This thesis contributes to the field of spatial epidemiology of NTDs by introducing innovative statistical methods, tools and knowledge for disease control, monitoring and evaluation.

In particular, we conducted the first, to our knowledge, geostatistical analyses combining point-referenced and areal survey data to make use of all possible available data and improve model-based disease risk mapping (Chapter 6). Furthermore we have introduced mathematical transmission models within geostatistical modeling of neglected diseases and for first time we produced age-specific risk maps from historical schistosomiasis data and showed how the age of the peak prevalence changes in space according to the disease endemicity (Chapter 7).

Results of the thesis were translated to high-resolution risk maps, number of infected people and treatment requirements for preventive chemotherapy of different NTDs across different regions. These tools and knowledge were either not available or were based on simplified calculations that ignored the variation of the diseases in space. The risk maps are important for guiding disease control and interventions, by informing policy-makers about the priority areas to target control measures and allocate resources. In addition, the maps, providing the baseline estimates of the disease risk, are useful for accessing the effectiveness of control interventions in the future. We obtained both, species-specific and overall risk maps of soil-transmitted helminth and of schistosome infections. The latter are particularly important for morbidity control as similar drugs are used against different species (Keiser & Utzinger 2008;WHO 2002a).

In Chapter 2, our results show that the prevalence of soil-transmitted helminth infections in P.R. China considerably decreased from 2005 onwards, yet, some 144 million people were estimated to be infected in 2010. The Chinese Ministry of Health set the target to reduce the prevalence of soil-transmitted helminth infections by 40% until 2010 and up to 70% until 2015 (Zheng et al. 2009). The government aims to reach these targets by a series of control strategies, including preventive chemotherapy, improvement of sanitation, and information, education and communication (IEC) campaigns (Bergquist & Whittaker 2012). Our models indicate that the first step of the target, i.e. an overall reduction of prevalence by 40% until 2010, has been achieved. On the other hand, we identified provinces such as Hainan, Guizhou, and Sichuan with very high disease prevalence, requiring more effective control strategies.

In Chapter 3 our risk maps depict the geographical distribution of soil-transmitted helminth infections in South Asia, highlighting the need for up-to-date surveys to accurately evaluate the disease burden for this region. The model-based results revealed that in 2015, all four countries in the study region (i.e., Bangladesh, India, Nepal, and Pakistan) had soil-transmitted helminth infection prevalence higher than 20%. Furthermore, our findings show that the infection risk of community-based surveys was higher than that of school-based surveys for *A. lumbricoides*. Negligible differences were found between school-aged population and the entire community for the other two species. These results support suggestions of other researchers that control strategies focusing on school-based deworming needs to be reassessed and treatments should be extended to other populations (e.g., preschool-aged children, women of childbearing age, and high occupational exposure adults) or to the whole community (Anderson et al. 2013;Karagiannis-Voules et al. 2015a;Lo et al. 2015).

In Chapter 4 our geostatistical analyses suggest the areas mainly in the southern and northeastern parts of P.R. China that control interventions for *C. sinensis* infection should be concentrated. Around 14.8 million people in the country were infected with *C. sinensis* in

2010 and an increased trend of the infection risk over time was estimated, urging the Chinese government to pay more attention to the public health importance of this disease.

Chapter 5 provides up-to-date, spatially explicit estimates of schistosomiasis risk and of the number of infected people across sub-Saharan Africa. Infection risk decreased from 2000 onwards, yet estimates suggest that around 163 million people in sub-Saharan Africa were infected with at least one of *Schistosoma* species in 2012, 57 million of whom were school-aged children. Mozambique had the highest prevalence of *Schistosoma* infection in school-aged children (52.8%), while low-risk countries (prevalence among school-aged children <10%) included Burundi, Equatorial Guinea, Eritrea and Rwanda. Annualised treatment needs with praziquantel were estimated at 123 million doses for school-aged children and 247 million for the entire population, which are similar to the ones reported by WHO. However, we obtained different estimates for a number of countries. The amount of praziquantel in the WHO donation is planned to increase every year: from 27 million in 2012 to 44 million in 2013, 75 million in 2014, 100 million in 2015, and up to 250 million in 2016 and subsequent years. Our results indicate that when the donation reaches the full scale, the praziquantel will be sufficient to treat all the population at risk of schistosomiasis in sub-Saharan Africa. Furthermore, our analysis allowed us to estimate treatment needs at different administrative levels, providing important information for the control programmes to guide the distribution of praziquantel within countries.

8.1.3 Contribution towards a global database of NTDs

Burden estimation of NTDs at high spatial resolution requires availability of georeferenced survey data. The open-access GTND database (Hürlimann et al. 2011) that was initiated by the EU-CONTRAST project (2006-2010) compiles and geo-references all available survey data on NTDs across the world. This thesis has extended the database with data on soil-transmitted helminth infections and food-borne trematodiasis covering the regions of South Asia and P.R. China. In addition, we updated the schistosomiasis data in Africa with the most recent surveys. These data were identified through systematic reviews and searches of published and "grey" (e.g., theses, working papers from research groups, or unpublished research reports through personal communication) literature, extracted, geo-located, and entered into the database. We have contributed to the GNTD database with a total of survey data collected at over 10,750 unique locations In particularly, we extracted survey data at 8,414 unique locations for schistosomiasis in sub-Saharan Africa, 1,610 unique locations for soil-transmitted helminth infections in South Asia and P.R. China, and 726 unique locations for food-borne trematodiasis in P.R. China. These data are available for utilization by other researchers, scientists, disease control managers, and policy makers. On the other hand, the georeferenced surveys, showing the geographical distribution and coverage of the available disease data, provided important information for planning of future surveys.

8.2 Limitations

The statistical analyses of this work rely on historical data compiled from studies that may differ in the design, diagnostic methods, and age groups of the population covered. As it was difficult to assess the quality of different diagnostic techniques/procedures in most of the surveys and in order to avoid discarding surveys with incomplete information about the diagnostic method, we analyzed the data regardless of the diagnostic method and assumed common sensitivity and specificity across all surveys. However, the sensitivity depends on the intensity of infection and the diagnostic approaches, and hence varies in space (Booth et al. 2003;Nikolay et al. 2014). The age-heterogeneity of the survey data across locations motivated the methodological work of Chapter 7, however due to time constraints the methodology could not be implemented to geostatistical analyses other than those involving the schistosomiasis data from Côte d'Ivoire.

Survey locations from historical data may over-represent endemic areas, as many surveys are likely to be conducted in places with relatively high infection risk (preferential sampling). Ignoring preferential sampling could lead to over-estimation of the infection risk in low risk areas (Diggle et al. 2010). On the other hand, we introduced temporal trend as a time covariate into the models, assuming that the risk varies uniformly across time, thus ignored the variation of the spatial process over time. Spatio-temporal models can be applied for joint analyses of space and time components for risk profiling (Chammartin et al. 2014a). However, the distribution of surveys from historical data varies across different time periods, which makes it difficult for such models to produce reliable spatio-temporal estimates.

We estimated prevalence of any soil-transmitted helminth infection and schistosomiasis by assuming independence of the respective species. Previous researchers suggested a positive association between *A. lumbricoides* and *T. trichiura*, hence, our assumption may over-estimate the true prevalence of soil-transmitted helminthes (Booth & Bundy 1992;Tchuem Tchuenté et al. 2003a). However, it is difficult to adjust the calculation by adding a correction factor due to lack of co-infection data in the study regions (de Silva & Hall 2010). Our independence assumption of *Schistosoma* species is supported by a study which analysed data from a national survey in Côte d'Ivoire in 2012 and reported that species co-infection is less likely to occur than simply by chance (Chammartin et al. 2014b). In contrast, another study reported high co-infection of *S. haematobium* and *S. mansoni* in some villages of Senegal (Meurs et al. 2012). Although we might over-estimate the schistosomiasis prevalence in areas with high species co-infection risk, our results showed that the independence assumption has very little impact on the overall infection risk in sub-Saharan Africa, when we calculated the coinfection risk for a range of values quantifying species dependence (Chapter 5).

We assumed that the regression coefficients are constant across the study region. However, the relationships between the predictors and the disease risk might vary in space, as a result of a geographical variation in the effect (e.g., across eco-regions) or some unmeasured factors (e.g., intervention levels and health-system performance) that can differ in space (Karagiannis-Voules et al. 2015a). A model with spatially varying coefficients can take into account the varying effects (Gelfand et al. 2003;Giardina et al. 2014). Furthermore we considered a stationary, and isotropic spatial process across the study regions, assuming that spatial association depends only upon distance and not on location as well as direction in space (Ecker & Gelfand 2003). We have not used models in the thesis that have spatially varying effects or address non-stationarity because these models are computationally very intensive especially when modeling data over large areas (Banerjee et al. 2004;Schmidt & O'Hagan 2003;Vounatsou et al. 2009).

8.3 Extension of the work

Our geostatistical analyses and models can be extended to other NTDs and regions, with priority the estimation of the soil transmitted helminthiasis risk in Southeast Asia, Central America, and Caribbean, and of the schistosomiasis risk in endemic regions of Northern South America, Carribean, the Eastern Mediterranean, and Eastern and Southeast Asia. These estimates together with the already accomplished risk maps by Chammartin *et al* (2013c), Karagiannis-Voules *et al* (2015a), Scholte *et al* (2014), and our work (Lai *et al* 2013, 2015 and Chapter 2, 3 and 5) will complete the model-based risk mapping of soil-transmitted helminth infections and schistosomiasis at global scale. In addition, the high-resolution risk maps we produced can be further used for estimation of disease burden maps.

As some NTDs share similar control strategies, tackling of these diseases through co-implementation (e.g., conducting integrating drug distribution, improving WASH, and enhancing IEC) in highly co-endemic areas can be very effective and affordable (Brady et al. 2006;Laxminarayan et al. 2006). Co-endemic risk maps of diseases will support planning for co-implementation, thus, are important for disease control and prevention. The most readily extension is to estimate the geographical distribution of co-endemic risk of soil-transmitted helminth infections and schistosomiasis in sub-Saharan Africa, as risk estimates of both diseases are already available (i.e., Chapter 5 and Karagiannis-Voules *et al* 2015a).

Our Bayesian geostatistical and mathematical modeling approach for analysing age-heterogeneous survey data of schistosomiasis can be further developed to obtain age-specific risk maps of other NTDs (e.g, soil-transmitted helminth infections), by defining appropriate transmission models of the corresponding disease. The predictive ability of the models may be improved if intensity data are incorporated into the analyses, however, such models may not be easy to use for large-scale risk mapping if intensity data are not available.

Chapter 9   Conclusion

Over the last years a lot of attention has been raised for control, elimination, and eradication of NTDs, which were historically overlooked although they affect more than one billion of people living in the world's poorest areas. Accessing the geographical distribution of these diseases at high spatial resolution is important for disease control, by targeting control interventions at areas of highest risk and evaluating effectiveness of control programmes. This PhD thesis focuses on three important NTDs, namely soil-transmitted helminth infections, schistosomiasis, and clonorchiasis. Systematic reviews were carried out to collect available disease survey data at different regions, geo-reference and enter them into the open-access GNTD database. Data-driven Bayesian geostatistical models were applied to estimate the disease risk based on a suite of important environmental and socioeconomic predictors. Advanced models were developed to address inherent data characteristics, i.e., data with age heterogeneity and data with heterogeneous geographical support. Up-to-date, model-based, high-resolution risk maps, the number of infected people, and the annual treatment needs for preventive chemotherapy were estimated for soil-transmitted helminth infections in P.R. China and South Asia, clonorchiasis in P.R. China, and schistosomiasis in sub-Saharan Africa. Our work contributes to the fields of spatial statistics and of epidemiology of NTDs with both innovative statistical methodology for the spatial analysis of heterogeneous survey data across space and with tools and knowledge for disease control, monitoring and evaluation.

Bibliography

Aagaard-Hansen, J., Mwanga, J. R., & Bruun, B., 2009. Social science perspectives on schistosomiasis control in Africa: past trends and future directions. *Parasitology* 136(13), pp. 1747-1758.

Acs, N., Banhidy, F., Puho, E., & Czeizel, A. E., 2005. Population-based case-control study of mebendazole in pregnant women for birth outcomes. *Congenital Anomalies* 45(3), pp. 85-88.

Al-Delaimy, A. K., Al-Mekhlafi, H. M., Lim, Y. A. L., Nasr, N. A., Sady, H., Atroosh, W. M., & Mahmud, R., 2014. Developing and evaluating health education learning package (HELP) to control soil-transmitted helminth infections among Orang Asli children in Malaysia. *Parasites & Vectors* 7, p. 416.

Allen, H., Sithey, G., Padmasiri, E. A., & Montresor, A., 2004. Epidemiology of soil-transmitted helminths in the western region of Bhutan. *The Southeast Asian Journal of Tropical Medicine and Public Health* 35(4), pp. 777-779.

Amarir, F., El, M. B., Fellah, H., Sebti, F., Mohammed, L., Handali, S., Wilkins, P., El Idrissi, A. L., Sadak, A., & Rhajaoui, M., 2011. National serologic survey of Haematobium schistosomiasis in Morocco: evidence for elimination. *The American Journal of Tropical Medicine and Hygiene* 84(1), pp. 15-19.

Anderson, R., Truscott, J., & Hollingsworth, T. D., 2014. The coverage and frequency of mass drug administration required to eliminate persistent transmission of soil-transmitted helminths. *Philosophical Transactions of the Royal Society B-Biological Sciences* 369(1645), p. 20130435.

Anderson, R. M. & May, R. M., 1985. Helminth infections of humans: mathematical models, population dynamics, and control. *Advances in Parasitology* 24, pp. 1-101.

Anderson, R. M., Truscott, J. E., Pullan, R. L., Brooker, S. J., & Hollingsworth, T. D., 2013. How effective is school-based deworming for the community-wide control of soil-transmitted helminths? *Plos Neglected Tropical Diseases* 7(2), p. e2027.

Appleton, C. C., 1977. The influence of above-optimal constant temperatures on South African Biomphalaria pfeifferi (Krauss) (Mollusca: Planorbidae). *Transactions of the Royal Society of Tropical Medicine and Hygiene* 71(2), pp. 140-143.

Appleton, C. C., 1984. Schistosome dermatitis--an unrecognized problem in South Africa? *South African Medical Journal* 65(12), pp. 467-469.

Appleton, C. C. & Gouws, E., 1996. The distribution of common intestinal nematodes along an altitudinal transect in KwaZulu-Natal, South Africa. *Annals of Tropical Medicine and Parasitology* 90(2), pp. 181-188.

Asaolu, S. O. & Ofoezie, I. E., 2003. The role of health education and sanitation in the control of helminth infections. *Acta Trop* 86(2-3), pp. 283-294.

Assare, R. K., Lai, Y. S., Yapi, A., Tian-Bi, Y. N., Ouattara, M., Yao, P. K., Knopp, S., Vounatsou, P., Utzinger, J., & N'Goran, E. K., 2015. The spatial distribution of *Schistosoma mansoni* infection in four regions of western Cote d'Ivoire. *Geospatial Health* 10(1), p. 345.

Augusto, G., Nala, R., Casmo, V., Sabonete, A., Mapaco, L., & Monteiro, J., 2009. Geographic distribution and prevalence of schistosomiasis and soil-transmitted helminths among schoolchildren in Mozambique. *The American Journal of Tropical Medicine and Hygiene* 81(5), pp. 799-803.

Banerjee, S., Carlin, B. P., & Gelfand, A. E. 2014. *Hierarchical Modeling and Analysis for Spatial Data, Second Edition.* Chapman and Hall/CRC.

Banerjee, S., Finley, A. O., Waldmann, P., & Ericsson, T., 2010. Hierarchical spatial process models for multiple traits in large genetic trials. *Journal of the American Statistical Association* 105(490), pp. 506-521.

Banerjee, S., Gelfand, A. E., Finley, A. O., & Sang, H. Y., 2008. Gaussian predictive process models for large spatial data sets. *Journal of the Royal Statistical Society Series B-Statistical Methodology* 70(4), pp. 825-848.

Banerjee, S., Gelfand, A. E., Knight, J. R., & Sirmans, C. F., 2004. Spatial modeling of house prices using normalized distance-weighted sums of stationary processes. *Journal of Business & Economic Statistics* 22(2), pp. 206-213.

Basáñez, M. G. & Anderson, R. M., 2015. Mathematical models for neglected tropical diseases: essential tools for control and elimination, part A. *Advances in Parasitology* 87, p. xiii-xvii.

Basáñez, M. G., McCarthy, J. S., French, M. D., Yang, G. J., Walker, M., Gambhir, M., Prichard, R. K., & Churcher, T. S., 2012. A research agenda for helminth diseases of humans: modelling for control and elimination. *Plos Neglected Tropical Diseases* 6(4), p. e1548.

Beck-Worner, C., Raso, G., Vounatsou, P., N'Goran, E. K., Rigo, G., Parlow, E., & Utzinger, J., 2007. Bayesian spatial risk prediction of Schistosoma mansoni infection in western Cote d'Ivoire using a remotely-sensed digital elevation model. *American Journal of Tropical Medicine and Hygiene* 76(5), pp. 956-963.

Becker, S. L., Sieto, B., Silue, K. D., Adjossan, L., Kone, S., Hatz, C., Kern, W. V., N'Goran, E. K., & Utzinger, J., 2011. Diagnosis, clinical features, and self-reported morbidity of *Strongyloides stercoralis* and hookworm infection in a Co-endemic setting. *Plos Neglected Tropical Diseases* 5(8), p. e1292.

Bergquist, R. & Whittaker, M., 2012. Control of neglected tropical diseases in Asia Pacific: implications for health information priorities. *Infectious Diseases of Poverty* 1(1), p. 3.

Bethony, J., Brooker, S., Albonico, M., Geiger, S. M., Loukas, A., Diemert, D., & Hotez, P. J., 2006. Soil-transmitted helminth infections: ascariasis, trichuriasis, and hookworm. *The Lancet* 367(9521), pp. 1521-1532.

Blok, D. J., de Vlas, S. J., Fischer, E. A., & Richardus, J. H., 2015. Mathematical modelling of leprosy and its control. *Advances in Parasitology* 87, pp. 33-51.

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

Booth, M., Vounatsou, P., N'Goran, E. K., Tanner, M., & Utzinger, J., 2003. The influence of sampling effort and the performance of the Kato-Katz technique in diagnosing *Schistosoma mansoni* and hookworm co-infections in rural Côte d'Ivoire. *Parasitology* 127, pp. 525-531.

Bouvard, V., Baan, R., Straif, K., Grosse, Y., Secretan, B., El Ghissassi, F., Benbrahim-Tallaa, L., Guha, N., Freeman, C., Galichet, L., & Cogliano, V., 2009. A review of human carcinogens-Part B: biological agents. *The Lancet Oncology* 10(4), pp. 321-322.

Brady, M. A., Hooper, P. J., & Ottesen, E. A., 2006. Projected benefits from integrating NTD programs in sub-Saharan Africa. *Trends in Parasitology* 22(7), pp. 285-291.

Brooker, S., Clements, A. C. A., & Bundy, D. A. P., 2006. Global epidemiology, ecology and control of soil-transmitted helminth infections. *Advances in Parasitology* 62, pp. 221-261.

Brooker, S., Singhasivanon, P., Waikagul, J., Supavej, S., Kojima, S., Takeuchi, T., Luong, T. V., & Looareesuwan, S., 2003. Mapping soil-transmitted helminths in Southeast Asia and implications for parasite control. *Southeast Asian Journal of Tropical Medicine and Public Health* 34(1), pp. 24-36.

Brooks, S. P. & Gelman, A., 1998. General methods for monitoring convergence of iterative simulations. *Journal of Computational and Graphical Statistics* 7(4), pp. 434-455.

Bruun, B. & Aagaard-Hansen, J. 2008. *The social context of schistosomiasis and its control: an introduction and annotated bibliography.* Geneva, Switzerland: World Health Organization.

Bundy, D. A. P., Chan, M. S., & Savioli, L., 1995. Hookworm infection in pregnancy. *Transactions of the Royal Society of Tropical Medicine and Hygiene* 89(5), pp. 521-522.

Bundy, D. A. P. & Cooper, E. S., 1989. Trichuris and trichuriasis in humans. *Advances in Parasitology* 28, pp. 107-173.

Burke, M. L., Jones, M. K., Gobert, G. N., Li, Y. S., Ellis, M. K., & Mcmanus, D. P., 2009. Immunopathogenesis of human schistosomiasis. *Parasite Immunology* 31(4), pp. 163-176.

Butler, K. & Stephens, M. 1993. *The distribution of a sum of binomial random variables.* Stanford University, Stanford, California, USA, April 1993, prepared for the Office of Naval Research., 467.

Butterworth, A. E., Fulford, A. J. C., Dunne, D. W., Ouma, J. H., & Sturrock, R. F., 1988. Longitudinal studies on human schistosomiasis. *Philosophical Transactions of the Royal Society of London Series B-Biological Sciences* 321(1207), pp. 495-511.

Cameletti, M., Lindgren, F., Simpson, D., & Rue, H., 2013. Spatio-temporal modeling of particulate matter concentration through the SPDE approach. *Asta-Advances in Statistical Analysis* 97(2), pp. 109-131.

Campbell, S. J., Savage, G. B., Gray, D. J., Atkinson, J. A., Soares Magalhaes, R. J., Nery, S. V., McCarthy, J. S., Velleman, Y., Wicken, J. H., Traub, R. J., Williams, G. M., Andrews, R. M., & Clements, A. C., 2014. Water, sanitation, and hygiene (WASH): a critical

component for sustainable soil-transmitted helminth and schistosomiasis control. *Plos Neglected Tropical Diseases* 8(4), p. e2651.

Carneiro, T. R., Pinheiro, M. C. C., de Oliveira, S. M., Hanemann, A. L. D., Queiroz, J. A. N., & Bezerra, F. S. M., 2012. Increased detection of schistosomiasis with Kato-Katz and SWAP-IgG-ELISA in a Northeastern Brazil low-intensity transmission area. *Revista da Sociedade Brasileira de Medicina Tropical* 45(4), pp. 510-513.

Chammartin, F., Guimaraes, L. H., Scholte, R. G. C., Bavia, M. E., Utzinger, J., & Vounatsou, P., 2014a. Spatio-temporal distribution of soil-transmitted helminth infections in Brazil. *Parasites & Vectors* 7, p. 440.

Chammartin, F., Houngbedji, C. A., Hurlimann, E., Yapi, R. B., Silue, K. D., Soro, G., Kouame, F. N., EK, N. G., Utzinger, J., Raso, G., & Vounatsou, P., 2014b. Bayesian risk mapping and model-based estimation of *Schistosoma haematobium-Schistosoma mansoni* co-distribution in Cote d'Ivoire. *Plos Neglected Tropical Diseases* 8(12), p. e3407.

Chammartin, F., Scholte, R. G. C., Malone, J. B., Bavia, M. E., Nieto, P., Utzinger, J., & Vounatsou, P., 2013a. Modelling the geographical distribution of soil-transmitted helminth infections in Bolivia. *Parasites & Vectors* 6, p. 152.

Chammartin, F., Hürlimann, E., Raso, G., N'Goran, E. K., Utzinger, J., & Vounatsou, P., 2013b. Statistical methodological issues in mapping historical schistosomiasis survey data. *Acta Tropica* 128(2), pp. 345-352.

Chammartin, F., Scholte, R. G. C. C., Guimaraes, L. H., Tanner, M., Utzinger, J., & Vounatsou, P., 2013c. Soil-transmitted helminth infection in South America: a systematic review and geostatistical meta-analysis. *The Lancet Infectious Diseases* 13(6), pp. 507-518.

Chan, M. S., Anderson, R. M., Medley, G. F., & Bundy, D. A. P., 1996. Dynamic aspects of morbidity and acquired immunity in schistosomiasis control. *Acta Tropica* 62(2), pp. 105-117.

Chan, M. S., Guyatt, H. L., Bundy, D. A. P., Booth, M., Fulford, A. J. C., & Medley, G. F., 1995. The development of an age structured model for schistosomiasis transmission dynamics and control and its validation for Schistosoma mansoni. *Epidemiology and Infection* 115(2), pp. 325-344.

Chau, L. V., De, N. V., Son, D. T., Hien, N. T. T., Cong, L. D., & Waikagyl, J. 2001. *Epidemiology of clonorchiasis in Northern Vietnam.* IWMI Books, Reports, International Water Management Institute, H029043.

Chen, X. B., Hu, S. F., & Shen, J. L., 1988. Inspection of antibody level in treated patients with clonorchiasis. *Bengbu Yi Xue Yuan Xue Bao (in Chinese)* 13, pp. 215-217.

Cho, S. H., Lee, K. Y., Lee, B. C., Cho, P. Y., Cheun, H. I., Hong, S. T., Sohn, W. M., & Kim, T. S., 2008. Prevalence of clonorchiasis in southern endemic areas of Korea in 2006. *The Korean Journal of Parasitology* 46(3), pp. 133-137.

Choi, D. & Hong, S. T., 2007. Imaging diagnosis of clonorchiasis. *The Korean Journal of Parasitology* 45(2), pp. 77-85.

Choi, M. H., Park, S. K., Li, Z., Ji, Z., Yu, G., Feng, Z., Xu, L., Cho, S. Y., Rim, H. J., Lee, S. H., & Hong, S. T., 2010. Effect of control strategies on prevalence, incidence and re-

infection of clonorchiasis in endemic areas of China. *Plos Neglected Tropical Diseases* 4(2), p. e601.

Chu, T., Liao, C., Huang, Y., Chang, Y., Costa, A., Ji, D., Nara, T., Tsubouchi, A., Chang, P. W., Chiu, W., & Fan, C., 2012. Prevalence of *Schistosoma intercalatum* and *S. haematobium* Infection among Primary Schoolchildren in Capital Areas of Democratic Republic Of Sao Tome and Principe, West Africa. *Iranian Journal of Parasitology* 7(1), pp. 67-72.

Clements, A. C., Deville, M. A., Ndayishimiye, O., Brooker, S., & Fenwick, A., 2010a. Spatial co-distribution of neglected tropical diseases in the east African great lakes region: revisiting the justification for integrated control. *Tropical Medicine & International Health* 15(2), pp. 198-207.

Clements, A. C., Garba, A., Sacko, M., Toure, S., Dembele, R., Landoure, A., Bosque-Oliva, E., Gabrielli, A. F., & Fenwick, A., 2008a. Mapping the probability of schistosomiasis and associated uncertainty, West Africa. *Emerging Infectious Diseases* 14(10), pp. 1629-1632.

Clements, A. C., Lwambo, N. J., Blair, L., Nyandindi, U., Kaatano, G., Kinung'hi, S., Webster, J. P., Fenwick, A., & Brooker, S., 2006. Bayesian spatial analysis and disease mapping: tools to enhance planning and implementation of a schistosomiasis control programme in Tanzania. *Tropical Medicine & International Health* 11(4), pp. 490-503.

Clements, A. C. A., Brooker, S., Nyandindi, U., Fenwick, A., & Blair, L., 2008b. Bayesian spatial analysis of a national urinary schistosomiasis questionnaire to assist geographic targeting of schistosomiasis control in Tanzania, East Africa. *International Journal for Parasitology* 38(3-4), pp. 401-415.

Clements, A. C. A., Kur, L. W., Gatpan, G., Ngondi, J. M., Emerson, P. M., Lado, M., Sabasio, A., & Kolaczinski, J. H., 2010b. Targeting Trachoma Control through Risk Mapping: The Example of Southern Sudan. *Plos Neglected Tropical Diseases* 4(8), p. e799.

Clements, A. C. A., Kur, L. W., Gatpan, G., Ngondi, J. M., Emerson, P. M., Lado, M., Sabasio, A., & Kolaczinski, J. H., 2010c. Targeting trachoma control through risk mapping: the example of Southern Sudan. *Plos Neglected Tropical Diseases* 4(8), p. e799.

Colley, D. G., Bustinduy, A. L., Secor, W. E., & King, C. H., 2014. Human schistosomiasis. *The Lancet* 383(9936), pp. 2253-2264.

Coordinating Office of the National Survey on the Important Human Parasitic Diseases, 2005. A national survey on current status of the important parasitic diseases in human population. *Chinese journal of parasitology and parasitic diseases (in Chinese)* 23(5 Suppl), pp. 332-340.

Coulibaly, J. T., Furst, T., Silue, K. D., Knopp, S., Hauri, D., Ouattara, M., Utzinger, J., & N'Goran, E. K., 2012. Intestinal parasitic infections in schoolchildren in different settings of Cote d'Ivoire: effect of diagnostic approach and implications for control. *Parasites & Vectors* 5, p. 135.

Coulibaly, J. T., N'Gbesso, Y. K., Knopp, S., N'Guessan, N. A., Silue, K. D., van Dam, G. J., N'Goran, E. K., & Utzinger, J., 2013a. Accuracy of urine circulating cathodic antigen test for the diagnosis of *Schistosoma mansoni* in preschool-aged children before and after treatment. *Plos Neglected Tropical Diseases* 7(3), p. e2109.

Coulibaly, J. T., N'Gbesso, Y. K., N'Guessan, N. A., Winkler, M. S., Utzinger, J., & N'Goran, E. K., 2013b. Epidemiology of schistosomiasis in two high-risk communities of south Cote d'Ivoire with particular emphasis on pre-school-aged children. *The American Journal of Tropical Medicine and Hygiene* 89(1), pp. 32-41.

Cowles, M. K., Yan, J., & Smith, B., 2009. Reparameterized and marginalized posterior and predictive sampling for complex Bayesian geostatistical models. *Journal of Computational and Graphical Statistics* 18(2), pp. 262-282.

Das, A. K., 2014. Hepatic and biliary ascariasis. *Journal of Global Infectious Diseases* 6(2), pp. 65-72.

Dayan, A. D., 2003. Albendazole, mebendazole and praziquantel. Review of non-clinical toxicity and pharmacokinetics. *Acta Tropica* 86(2-3), pp. 141-159.

De Cock, K. M., 1986. Hepatosplenic schistosomiasis: a clinical review. *Gut* 27(6), pp. 734-745.

de Silva, N. & Hall, A., 2010. Using the prevalence of individual species of intestinal nematode worms to estimate the combined prevalence of any species. *Plos Neglected Tropical Diseases* 4(4), p. e655.

de Silva, N. R., Brooker, S., Hotez, P. J., Montresor, A., Engels, D., & Savioli, L., 2003. Soil-transmitted helminth infections: updating the global picture. *Trends in Parasitology* 19(12), pp. 547-551.

del Villar, L. P., Burguillo, F. J., Lopez-Aban, J., & Muro, A., 2012. Systematic review and meta-analysis of artemisinin based therapies for the treatment and prevention of schistosomiasis. *PLoS One* 7(9).

Diggle, P. J., Menezes, R., & Su, T. L., 2010. Geostatistical inference under preferential sampling. *Journal of the Royal Statistical Society Series C-Applied Statistics* 59, pp. 191-232.

Diggle, P. J., Tawn, J. A., & Moyeed, R. A., 1998. Model-based geostatistics. *Journal of the Royal Statistical Society Series C-Applied Statistics* 47, pp. 299-326.

Dormann, C. F., Elith, J., Bacher, S., Buchmann, C., Carl, G., Carre, G., Marquez, J. R. G., Gruber, B., Lafourcade, B., Leitao, P. J., Munkemuller, T., McClean, C., Osborne, P. E., Reineking, B., Schroder, B., Skidmore, A. K., Zurell, D., & Lautenbach, S., 2013. Collinearity: a review of methods to deal with it and a simulation study evaluating their performance. *Ecography* 36(1), pp. 27-46.

Ecker, M. D. & Gelfand, A. E., 2003. Spatial modeling and prediction under stationary non-geometric range anisotropy. *Environmental and Ecological Statistics* 10(2), pp. 165-178.

Ekpo, U. F., Hurlimann, E., Schur, N., Oluwole, A. S., Abe, E. M., Mafe, M. A., Nebe, O. J., Isiyaku, S., Olamiju, F., Kadiri, M., Poopola, T. O., Braide, E. I., Saka, Y., Mafiana, C. F., Kristensen, T. K., Utzinger, J., & Vounatsou, P., 2013. Mapping and prediction of schistosomiasis in Nigeria using compiled survey data and Bayesian geospatial modelling. *Geospatial Health* 7(2), pp. 355-366.

Elbaz, T. & Esmat, G., 2013. Hepatic and intestinal schistosomiasis: review. *Journal of Advanced Research* 4(5), pp. 445-452.

Enk, M. J., Lima, A. C. L., Drummond, S. C., Schall, V. T., & Coelho, P. M. Z., 2008. The effect of the number of stool samples on the observed prevalence and the infection intensity with *Schistosoma mansoni* among a population in an area of low transmission. *Acta Tropica* 108(2-3), pp. 222-228.

Enk, M. J., Silva, G. O. E., & Rodrigues, N. B., 2012. Diagnostic accuracy and applicability of a PCR system for the detection of *Schistosoma mansoni* DNA in human urine samples from an endemic area. *PLoS One* 7(6), p. e38947.

Escobedo, A. A., Canete, R., & Nunez, F. A., 2008. Prevalence, risk factors and clinical features associated with intestinal parasitic infections in children from San Juan y Martínez, Pinar del Río, Cuba. *West Indian Medical Journal* 57(4), pp. 377-382.

Fang, Y. Y., Chen, Y. D., Li, X. M., Wu, J., Zhang, Q. M., & Ruan, C. W., 2008. Current prevalence of *Clonorchis sinensis* infection in endemic areas of China. *Zhongguo Ji Sheng Chong Xue Yu Ji Sheng Chong Bing Za Zhi (in Chinese)* 26(2), pp. 99-103, 109.

Fenwick, A., Webster, J. P., Bosque-Oliva, E., Blair, L., Fleming, F. M., Zhang, Y., Garba, A., Stothard, J. R., Gabrielli, A. F., Clements, A. C. A., Kabatereine, N. B., Toure, S., Dembele, R., Nyandindi, U., Mwansa, J., & Koukounari, A., 2009. The Schistosomiasis Control Initiative (SCI): rationale, development and implementation from 2002-2008. *Parasitology* 136(13), pp. 1719-1730.

Fernandez, M. & Williams, S., 2010. Closed-form expression for the Poisson-binomial probability density function. *Ieee Transactions on Aerospace and Electronic Systems* 46(2), pp. 803-817.

Finley, A. O., Sang, H. Y., Banerjee, S., & Gelfand, A. E., 2009. Improving the performance of predictive process modeling for large datasets. *Computational Statistics & Data Analysis* 53(8), pp. 2873-2884.

Fitzsimmons, C. M., Jones, F. M., Pinot de, M. A., Protasio, A. V., Khalife, J., Dickinson, H. A., Tukahebwa, E. M., & Dunne, D. W., 2012. Progressive cross-reactivity in IgE responses: an explanation for the slow development of human immunity to schistosomiasis? *Infection and Immunity* 80(12), pp. 4264-4270.

Flores, A., Esteban, J. G., Angles, R., & Mas-Coma, S., 2001. Soil-transmitted helminth infections at very high altitude in Bolivia. *Transactions of the Royal Society of Tropical Medicine and Hygiene* 95(3), pp. 272-277.

Foo, K. T., Blackstock, A. J., Ochola, E. A., Matete, D. O., Mwinzi, P. N., Montgomery, S. P., Karanja, D. M., & Secor, W. E., 2015. Evaluation of point-of-contact circulating cathodic antigen assays for the detection of *Schistosoma mansoni* infection in low-, moderate-, and high-prevalence schools in western Kenya. *The American Journal of Tropical Medicine and Hygiene* 92(6), pp. 1227-1232.

Freeman, M. C., Chard, A. N., Nikolay, B., Garn, J. V., Okoyo, C., Kihara, J., Njenga, S. M., Pullan, R. L., Brooker, S. J., & Mwandawiro, C. S., 2015. Associations between school- and household-level water, sanitation and hygiene conditions and soil-transmitted helminth infection among Kenyan school children. *Parasites & Vectors* 8, p. 412.

Freeman, M. C., Clasen, T., Brooker, S. J., Akoko, D. O., & Rheingans, R., 2013. The impact of a school-based hygiene, water quality and sanitation intervention on soil-transmitted

helminth reinfection: a cluster-randomized trial. *The American Journal of Tropical Medicine and Hygiene* 89(5), pp. 875-883.

French, M. D., Churcher, T. S., Gambhir, M., Fenwick, A., Webster, J. P., Kabatereine, N. B., & Basanez, M. G., 2010. Observed reductions in *Schistosoma mansoni* transmission from large-scale administration of praziquantel in Uganda: a mathematical modelling study. *Plos Neglected Tropical Diseases* 4(11), p. e897.

French, M. D., Churcher, T. S., Webster, J. P., Fleming, F. M., Fenwick, A., Kabatereine, N. B., Sacko, M., Garba, A., Toure, S., Nyandindi, U., Mwansa, J., Blair, L., Bosque-Oliva, E., & Basanez, M. G., 2015. Estimation of changes in the force of infection for intestinal and urogenital schistosomiasis in countries with schistosomiasis control initiative-assisted programmes. *Parasites & Vectors* 8, p. 558.

Fürst, T., Keiser, J., & Utzinger, J., 2012. Global burden of human food-borne trematodiasis: a systematic review and meta-analysis. *The Lancet Infectious Diseases* 12(3), pp. 210-221.

Fürst, T., Silue, K. D., Ouattara, M., N'Goran, D. N., Adiossan, L. G., N'Guessan, Y., Zouzou, F., Kone, S., N'Goran, E. K., & Utzinger, J., 2012. Schistosomiasis, soil-transmitted helminthiasis, and sociodemographic factors influence quality of life of adults in Côte d'Ivoire. *Plos Neglected Tropical Diseases* 6(10), p. e1855.

Gambhir, M., Singh, B. K., & Michael, E., 2015. The Allee effect and elimination of neglected tropical diseases: a mathematical modelling study. *Adv.Parasitol.* 87, pp. 1-31.

Garba, A., Toure, S., Dembele, R., Boisier, P., Tohon, Z., Bosque-Oliva, E., Koukounari, A., & Fenwick, A., 2009. Present and future schistosomiasis control activities with support from the Schistosomiasis Control Initiative in West Africa. *Parasitology* 136(13), pp. 1731-1737.

Gazzinelli, A., Velasquez-Melendez, G., Crawford, S. B., LoVerde, P. T., Correa-Oliveira, R., & Kloos, H., 2006. Socioeconomic determinants of schistosomiasis in a poor rural area in Brazil. *Acta Tropica* 99(2-3), pp. 260-271.

Gelfand, A. E., Hills, S. E., Racinepoon, A., & Smith, A. F. M., 1990. Illustration of Bayesian-inference in normal data models using Gibbs sampling. *Journal of the American Statistical Association* 85(412), pp. 972-985.

Gelfand, A. E., Kim, H. J., Sirmans, C. F., & Banerjee, S., 2003. Spatial modeling with spatially varying coefficient processes. *Journal of the American Statistical Association* 98(462), pp. 387-396.

Gelfand, A. E. & Smith, A. F. M., 1990. Sampling-based approaches to calculating marginal densities. *Journal of the American Statistical Association* 85(410), pp. 398-409.

Gelfand, A. E., Zhu, L., & Carlin, B. P., 2001. On the change of support problem for spatio-temporal data. *Biostatistics.* 2(1), pp. 31-45.

Gelman, A., Carlin, J. C., Stern, H. S., Dunson, D. B., Vehtari, A., & Rubin, D. B. 2013. *Bayesian Data Analysis, Third Edition.* Chapman and Hall/CRC.

Gelman, A. & Rubin, D. B., 1992. Inference from iterative simulation using multiple sequences. *Statistical Science* 7(4), pp. 457-511.

Geman, S. & Geman, D., 1984. Stochastic relaxation, Gibbs distributions, and the Bayesian restoration of images. *Ieee Transactions on Pattern Analysis and Machine Intelligence* 6(6), pp. 721-741.

Giardina, F., Kasasa, S., Sie, A., Utzinger, J., Tanner, M., & Vounatsou, P., 2014. Effects of vector-control interventions on changes in risk of malaria parasitaemia in sub-Saharan Africa: a spatial and temporal analysis. *The Lancet Global Health* 2(10), p. E601-E615.

Gneiting, T. & Raftery, A. E., 2007. Strictly proper scoring rules, prediction, and estimation. *Journal of the American Statistical Association* 102(477), pp. 359-378.

Gomes, L. I., Marques, L. H. D., Enk, M. J., de Oliveira, M. C., Coelho, P. M. Z., & Rabello, A., 2010. Development and evaluation of a sensitive PCR-ELISA system for detection of schistosoma infection in feces. *Plos Neglected Tropical Diseases* 4(4), p. e664.

Goovaerts, P., 2008a. Accounting for rate instability and spatial patterns in the boundary analysis of cancer mortality maps. *Environmental and Ecological Statistics* 15(4), pp. 421-446.

Goovaerts, P., 2008b. Kriging and semivariogram deconvolution in the presence of irregular geographical units. *Mathematical Geosciences* 40(1), pp. 101-128.

Goovaerts, P., 2010. Combining areal and point data in geostatistical interpolation: applications to soil science and medical geography. *Mathematical Geosciences* 42(5), pp. 535-554.

Gray, D. J., Mcmanus, D. P., Li, Y., Williams, G. M., Bergquist, R., & Ross, A. G., 2010. Schistosomiasis elimination: lessons from the past guide the future. *The Lancet Infectious Diseases* 10(10), pp. 733-736.

Gray, D. J., Ross, A. G., Li, Y. S., & Mcmanus, D. P., 2011. Diagnosis and management of schistosomiasis. *British Medical Journal* 342, p. d2651.

Grimes, J. E., Croll, D., Harrison, W. E., Utzinger, J., Freeman, M. C., & Templeton, M. R., 2015. The roles of water, sanitation and hygiene in reducing schistosomiasis: a review. *Parasites & Vectors* 8, p. 156.

Grimes, J. E. T., Croll, D., Harrison, W. E., Utzinger, J., Freeman, M. C., & Templeton, M. R., 2014. The relationship between water, sanitation and schistosomiasis: a systematic review and meta-analysis. *Plos Neglected Tropical Diseases* 8(12), p. e3296.

Gryseels, B., Polman, K., Clerinx, J., & Kestens, L., 2006. Human schistosomiasis. *The Lancet* 368(9541), pp. 1106-1118.

Gunawardena, K., Kumarendran, B., Ebenezer, R., Gunasingha, M. S., Pathmeswaran, A., & de Silva, N., 2011. Soil-transmitted helminth infections among plantation sector schoolchildren in Sri Lanka: prevalence after ten years of preventive chemotherapy. *Plos Neglected Tropical Diseases* 5(9), p. e1341.

Gyorkos, T. W., Maheu-Giroux, M., Blouin, B., & Casapia, M., 2013. The impact of a health education intervention on soil-transmitted helminth infections in grade 5 schoolchildren of the Peruvian Amazon: a cluster-randomized controlled trial. *Tropical Medicine & International Health* 18, p. 61.

Hairston, N. G., 1965. An analysis of age-prevalence data by catalytic models - a contribution to study of bilharziasis. *Bulletin of the World Health Organization* 33(2), pp. 163-175.

Han, S., Zhang, X., Chen, R., Wen, J., Li, Y., Shu, J., Ling, H., & Zhang, F., 2013. Trends in prevalence of clonorchiasis among patients in Heilongjiang province, Northeast China (2009-2012): implications for monitoring and control. *PLoS One* 8(11), p. e80173.

Han, S., Zhang, X. L., Wen, J. S., Li, Y. H., Shu, J., Ling, H., & Zhang, F. M., 2012. A combination of the Kato-Katz methods and ELISA to improve the diagnosis of clonorchiasis in an endemic area, China. *PLoS One* 7(10), p. e46977.

Hastings, W. K., 1970. Monte-Carlo sampling methods using Markov chains and their applications. *Biometrika* 57(1), pp. 97-109.

Hodges, M. H., Soares Magalhaes, R. J., Paye, J., Koroma, J. B., Sonnie, M., Clements, A., & Zhang, Y., 2012. Combined spatial prediction of schistosomiasis and soil-transmitted helminthiasis in Sierra Leone: a tool for integrated disease control. *Plos Neglected Tropical Diseases* 6(6), p. e1694.

Hohmann, H., Panzer, S., Phimpachan, C., Southivong, C., & Schelp, F. P., 2001. Relationship of intestinal parasites to the environment and to behavioral factors in children in the Bolikhamxay province of Lao PDR. *Southeast Asian Journal of Tropical Medicine and Public Health* 32(1), pp. 4-13.

Holford, T. R. & Hardy, R. J., 1976. Stochastic-model for analysis of age-specific prevalence curves in schistosomiasis. *Journal of Chronic Diseases* 29(7), pp. 445-458.

Hong, S. T., Choi, M. H., Kim, C. H., Chung, B. S., & Ji, Z., 2003. The Kato-Katz method is reliable for diagnosis of *Clonorchis sinensis* infection. *Diagnostic Microbiology and Infectious Disease* 47(1), pp. 345-347.

Hong, S. T. & Fang, Y., 2012. Clonorchis sinensis and clonorchiasis, an update. *Parasitology International* 61(1), pp. 17-24.

Horton, R. J., 1997. Albendazole in treatment of human cystic echinococcosis: 12 years of experience. *Acta Tropica* 64(1-2), pp. 79-93.

Hotez, P. J., 2013. NTDs V.2.0: "blue marble health"--neglected tropical disease control and elimination in a shifting health policy landscape. *Plos Neglected Tropical Diseases* 7(11), p. e2570.

Hotez, P. J., 2015. Blue marble health and "the big three diseases": HIV/AIDS, tuberculosis, and malaria. *Microbes and Infection* 17(8), pp. 539-541.

Hotez, P. J., Brooker, S., Bethony, J. M., Bottazzi, M. E., Loukas, A., & Xiao, S. H., 2004. Hookworm infection. *New England Journal of Medicine* 351(8), pp. 799-807.

Hotez, P. J., Bundy, D. A. P., Beegle, K., Brooker, S., Drake, L., de, S. N., Montresor, A., Engels, D., Jukes, M., Chitsulo, L., Chow, J., Laxminarayan, R., Michaud, C., Bethony, J., Correa-Oliveira, R., Shuhua, X., Fenwick, A., & Savioli, L. 2006. Helminth infections: soil-transmitted helminth infections and schistosomiasis. In: Jamison, D. T., Breman, J. G. , & Measham, A. R., eds. *Disease Control Priorities in Developing Countries. 2nd edition.* Washington (DC): World Bank.

Hotez, P. J., Fenwick, A., Savioli, L., & Molyneux, D. H., 2009. Rescuing the bottom billion through control of neglected tropical diseases. *The Lancet* 373(9674), pp. 1570-1575.

Hotez, P. J., Molyneux, D. H., Fenwick, A., Kumaresan, J., Sachs, S. E., Sachs, J. D., & Savioli, L., 2007. Current concepts - Control of neglected tropical diseases. *New England Journal of Medicine* 357(10), pp. 1018-1027.

Hsü, H. F. & Wang, L. S., 1938. Studies on certain problems of *Clonorchis sinensis*. IV. Notes on the resistance of cysts in fish flesh, the migration route, and the morphology of the young worm in the final host. *Chinese Medical Journal (Engl)* Suppl 2, pp. 385-400.

Huang, S. Y., Zhao, G. H., Fu, B. Q., Xu, M. J., Wang, C. R., Wu, S. M., Zou, F. C., & Zhu, X. Q., 2012. Genomics and molecular genetics of *Clonorchis sinensis*: Current status and perspectives. *Parasitology International* 61(1), pp. 71-76.

Huang, Y. X. & Manderson, L., 2005. The social and economic context and determinants of schistosomiasis japonica. *Acta Tropica* 96(2-3), pp. 223-231.

Hürlimann, E., Schur, N., Boutsika, K., Stensgaard, A. S., Laserna de Himpsl, M., Ziegelbauer, K., Laizer, N., Camenzind, L., Di, P. A., Ekpo, U. F., Simoonga, C., Mushinge, G., Saarnak, C. F., Utzinger, J., Kristensen, T. K., & Vounatsou, P., 2011. Toward an open-access global database for mapping, control, and surveillance of neglected tropical diseases. *Plos Neglected Tropical Diseases* 5(12), p. e1404.

Jaoko, W. G., Muchemi, G., & Oguya, F. O., 1996. Praziquantel side effects during treatment of *Schistosoma mansoni* infected pupils in Kibwezi, Kenya. *East African Medical Journal* 73(8), pp. 499-501.

Jia, T. W., Melville, S., Utzinger, J., King, C. H., & Zhou, X. N., 2012. Soil-transmitted helminth reinfection after drug treatment: a systematic review and meta-analysis. *Plos Neglected Tropical Diseases* 6(5), p. e1621.

Jodar, J., Sapriza, G., Herrera, C., Lamban, L. J., & Medina, A., 2015. Combining point and regular lattice data in geostatistical interpolation. *Journal of Geographical Systems* 17(3), pp. 275-296.

June, K. J., Cho, S. H., Lee, W. J., Kim, C., & Park, K. S., 2013. Prevalence and risk factors of clonorchiasis among the populations served by primary healthcare posts along five major rivers in South Korea. *Osong Public Health and Research Perspectives* 4(1), pp. 21-26.

Karagiannis-Voules, D. A., Biedermann, P., Ekpo, U. F., Garba, A., Langer, E., Mathieu, E., Midzi, N., Mwinzi, P., Polderman, A. M., Raso, G., Sacko, M., Talla, I., Tchuente, L. A., Toure, S., Winkler, M. S., Utzinger, J., & Vounatsou, P., 2015a. Spatial and temporal distribution of soil-transmitted helminth infection in sub-Saharan Africa: a systematic review and geostatistical meta-analysis. *The Lancet Infectious Diseases* 15(1), pp. 74-84.

Karagiannis-Voules, D. A., Odermatt, P., Biedermann, P., Khieu, V., Schar, F., Muth, S., Utzinger, J., & Vounatsou, P., 2015b. Geostatistical modelling of soil-transmitted helminth infection in Cambodia: do socioeconomic factors improve predictions? *Acta Tropica* 141(Pt B), pp. 204-212.

Karagiannis-Voules, D. A., Scholte, R. G., Guimaraes, L. H., Utzinger, J., & Vounatsou, P., 2013. Bayesian geostatistical modeling of leishmaniasis incidence in Brazil. *Plos Neglected Tropical Diseases* 7(5), p. e2213.

Kastner, R. J., Stone, C. M., Steinmann, P., Tanner, M., & Tediosi, F., 2015. What is needed to eradicate lymphatic filariasis? A model-based assessment on the impact of scaling up mass drug administration programs. *Plos Neglected Tropical Diseases* 9(10), p. e0004147.

Katz, N., Chaves, A., & Pellegrino, J., 1972. A simple device for quantitative stool thick-smear technique in *Schistosomiasis mansoni*. *Revista do Instituto de Medicina Tropical de São Paulo* 14(6), pp. 397-400.

Keiser, J., Tritten, L., Silbereisen, A., Speich, B., Adelfio, R., & Vargas, M., 2013. Activity of oxantel pamoate monotherapy and combination chemotherapy against Trichuris muris and hookworms: revival of an old drug. *Plos Neglected Tropical Diseases* 7(3), p. e2119.

Keiser, J. & Utzinger, J., 2005. Emerging foodborne trematodiasis. *Emerging Infectious Diseases* 11(10), pp. 1507-1514.

Keiser, J. & Utzinger, J., 2008. Efficacy of current drugs against soil-transmitted helminth infections - Systematic review and meta-analysis. *Jama-Journal of the American Medical Association* 299(16), pp. 1937-1948.

Keiser, J. & Utzinger, J., 2009. Food-borne trematodiases. *Clinical Microbiology Reviews* 22(3), pp. 466-483.

Keiser, J. & Utzinger, J., 2010. The drugs we have and the drugs we need against major helminth infections. *Advances in Parasitology* 73, pp. 197-230.

Khalaf, I., Shokeir, A., & Shalaby, M., 2012. Urologic complications of genitourinary schistosomiasis. *World Journal of Urology* 30(1), pp. 31-38.

Kim, J. H., Choi, M. H., Bae, Y. M., Oh, J. K., Lim, M. K., & Hong, S. T., 2011. Correlation between discharged worms and fecal egg counts in human clonorchiasis. *Plos Neglected Tropical Diseases* 5(10), p. e1339.

Kim, T. S., Cho, S. H., Huh, S., Kong, Y., Sohn, W. M., Hwang, S. S., Chai, J. Y., Lee, S. H., Park, Y. K., Oh, D. K., & Lee, J. K., 2009. A nationwide survey on the prevalence of intestinal parasitic infections in the Republic of Korea, 2004. *The Korean Journal of Parasitology* 47(1), pp. 37-47.

King, C. H. & Bertsch, D., 2015. Historical perspective: snail control to prevent schistosomiasis. *Plos Neglected Tropical Diseases* 9(4), p. e0003657.

King, C. H. & Dangerfield-Cha, M., 2008. The unacknowledged impact of chronic schistosomiasis. *Chronic Illness* 4(1), pp. 65-79.

King, C. H., Olbrych, S. K., Soon, M., Singer, M. E., Carter, J., & Colley, D. G., 2011. Utility of repeated praziquantel dosing in the treatment of schistosomiasis in high-risk communities in Africa: a systematic review. *Plos Neglected Tropical Diseases* 5(9), p. e1321.

Kino, H., Inaba, H., Van, D. N., Van, C. L., Son, D. T., Hao, H. T., Toan, N. D., Cong, L. D., & Sano, M., 1998. Epidemiology of clonorchiasis in Ninh Binh Province, Vietnam. *The Southeast Asian Journal of Tropical Medicine and Public Health* 29(2), pp. 250-254.

Kjetland, E. F., Leutscher, P. D. C., & Ndhlovu, P. D., 2012. A review of female genital schistosomiasis. *Trends in Parasitology* 28(2), pp. 58-65.

Knopp, S., Becker, S. L., Ingram, K. J., Keiser, J., & Utzinger, J., 2013. Diagnosis and treatment of schistosomiasis in children in the era of intensified control. *Expert Review of Anti-Infective Therapy* 11(11), pp. 1237-1258.

Knopp, S., Glinz, D., Rinaldi, L., Mohammed, K. A., N'Goran, E. K., Stothard, J. R., Marti, H., Cringoli, G., Rollinson, D., & Utzinger, J., 2009. FLOTAC: A promising technique for detecting helminth eggs in human faeces. *Transactions of the Royal Society of Tropical Medicine and Hygiene* 103(12), pp. 1190-1194.

Knopp, S., Mgeni, A. F., Khamis, I. S., Steinmann, P., Stothard, J. R., Rollinson, D., Marti, H., & Utzinger, J., 2008. Diagnosis of soil-transmitted helminths in the era of preventive chemotherapy: effect of multiple stool sampling and use of different diagnostic techniques. *Plos Neglected Tropical Diseases* 2(11), p. e331.

Knopp, S., Mohammed, K. A., Ali, S. M., Khamis, I. S., Ame, S. M., Albonico, M., Gouvras, A., Fenwick, A., Savioli, L., Colley, D. G., Utzinger, J., Person, B., & Rollinson, D., 2012. Study and implementation of urogenital schistosomiasis elimination in Zanzibar (Unguja and Pemba islands) using an integrated multidisciplinary approach. *Bmc Public Health* 12, p. 930.

Knopp, S., Mohammed, K. A., Stothard, J. R., Khamis, I. S., Rollinson, D., Marti, H., & Utzinger, J., 2010. Patterns and risk factors of helminthiasis and anemia in a rural and a peri-urban community in Zanzibar, in the context of helminth control programs. *Plos Neglected Tropical Diseases* 4(5), p. e681.

Knopp, S., Speich, B., Hattendorf, J., Rinaldi, L., Mohammed, K. A., Khamis, I. S., Mohammed, A. S., Albonico, M., Rollinson, D., Marti, H., Cringoli, G., & Utzinger, J., 2011. Diagnostic accuracy of Kato-Katz and FLOTAC for assessing anthelmintic drug efficacy. *Plos Neglected Tropical Diseases* 5(4), p. e1036.

Koroma, J. B., Peterson, J., Gbakima, A. A., Nylander, F. E., Sahr, F., Magalhaes, R. J. S., Zhang, Y. B., & Hodges, M. H., 2010. Geographical distribution of intestinal schistosomiasis and soil-transmitted helminthiasis and preventive chemotherapy strategies in Sierra Leone. *Plos Neglected Tropical Diseases* 4(11), p. e891.

Kottek, M., Grieser, J., Beck, C., Rudolf, B., & Rubel, F., 2006. World map of the Koppen-Geiger climate classification updated. *Meteorologische Zeitschrift* 15(3), pp. 259-263.

Kulkarni, S. W., Panicker, P. V., Gadkari, A. S., & Handa, B. K., 1978. Prevalence and patterns of parasitic infections in rural areas around Nagpur. *Indian Journal of Medical Research* 68, pp. 583-591.

Lai, Y. S., Biedermann, P., Ekpo, U. F., Garba, A., Mathieu, E., Midzi, N., Mwinzi, P., N'Goran, E. K., Raso, G., Assare, R. K., Sacko, M., Schur, N., Talla, I., Tchuente, L. A., Toure, S., Winkler, M. S., Utzinger, J., & Vounatsou, P., 2015. Spatial distribution of schistosomiasis and treatment needs in sub-Saharan Africa: a systematic review and geostatistical analysis. *The Lancet Infectious Diseases* 15(8), pp. 927-940.

Lai, Y. S., Zhou, X. N., Utzinger, J., & Vounatsou, P., 2013. Bayesian geostatistical modelling of soil-transmitted helminth survey data in the People's Republic of China. *Parasites & Vectors* 6, p. 359.

Lamberton, P. H., Crellen, T., Cotton, J. A., & Webster, J. P., 2015. Modelling the effects of mass drug administration on the molecular epidemiology of schistosomes. *Advances in Parasitology* 87, pp. 293-327.

Lamberton, P. H. L., Kabatereine, N. B., Oguttu, D. W., Fenwick, A., & Webster, J. P., 2014. Sensitivity and specificity of multiple Kato-Katz thick smears and a circulating cathodic antigen test for *Schistosoma mansoni* diagnosis pre- and post-repeated-praziquantel treatment. *Plos Neglected Tropical Diseases* 8(9), p. e3139.

Lansdown, R., Ledward, A., Hall, A., Issae, W., Yona, E., Matulu, J., Mweta, M., Kihamia, C., Nyandindi, U., & Bundy, D., 2002. Schistosomiasis, helminth infection and health education in Tanzania: achieving behaviour change in primary schools. *Health Education Research* 17(4), pp. 425-433.

Lardans, V. & Dissous, C., 1998. Snail control strategies for reduction of schistosomiasis transmission. *Parasitology Today* 14(10), pp. 413-417.

Laxminarayan, R., Mills, A. J., Breman, J. G., Measham, A. R., Alleyne, G., Claeson, M., Jha, P., Musgrove, P., Chow, J., Shahid-Salles, S., & Jamison, D. T., 2006. Advancement of global health: key messages from the Disease Control Priorities Project. *The Lancet* 367(9517), pp. 1193-1208.

Lee, B. Y., Bartsch, S. M., & Gorham, K. M., 2015. Economic and financial evaluation of neglected tropical diseases. *Advances in Parasitology* 87, pp. 329-417.

Lee, J. M., Lim, H. S., & Hong, S. T., 2011. Hypersensitive reaction to praziquantel in a clonorchiasis patient. *Korean Journal of Parasitology* 49(3), pp. 273-275.

Lelo, A. E., Mburu, D. N., Magoma, G. N., Mungai, B. N., Kihara, J. H., Mwangi, I. N., Maina, G. M., Kinuthia, J. M., Mutuku, M. W., Loker, E. S., Mkoji, G. M., & Steinauer, M. L., 2014. No apparent reduction in schistosome burden or genetic diversity following four years of school-based mass drug administration in mwea, central kenya, a heavy transmission area. *Plos Neglected Tropical Diseases* 8(10), p. e3221.

Levecke, B., Anderson, R. M., Berkvens, D., Charlier, J., Devleesschauwer, B., Speybroeck, N., Vercruysse, J., & Van, A. S., 2015. Mathematical inference on helminth egg counts in stool and its applications in mass drug administration programmes to control soil-transmitted helminthiasis in public health. *Adv.Parasitol.* 87(Advances in Parasitology), pp. 193-247.

Li, B. Z., Wang, C. X., Li, D. H., Liu, T. C., Deng, L. J., & Liu, Y. Z., 1983. Ecologica study on the Parafossarulus striatulus, the first intermediate host of Clonorchis sinensis in Liaoning province, China. *Dong Wu Xue Za Zhi (in Chinese)* 18(1), pp. 3-6.

Li, T., He, S. Y., Zhao, H., Zhao, G. H., & Zhu, X. Q., 2010. Major trends in human parasitic diseases in China. *Trends in Parasitology* 26(5), pp. 264-270.

Liang, S., Seto, E. Y., Remais, J. V., Zhong, B., Yang, C., Hubbard, A., Davis, G. M., Gu, X., Qiu, D., & Spear, R. C., 2007. Environmental effects on parasitic disease transmission exemplified by schistosomiasis in western China. *Proceedings of the National Academy of Sciences* 104(17), pp. 7110-7115.

Lindgren, F., Rue, H., & Lindstrom, J., 2011. An explicit link between Gaussian fields and Gaussian Markov ranadom fields: the stochastic partial differential equation approach.

Journal of the Royal Statistical Society Series B-Statistical Methodology 73(4), pp. 423-498.

Liu, B., Li, H. J., Gui, J. J., & Wei, J. B., 2010. Overview of the status of intestinal parasitic diseases in Shandong. *Journal of Pathogen Biology* 5(6), pp. 471-472.

Liu, R., Dong, H. F., Guo, Y., Zhao, Q. P., & Jiang, M. S., 2011. Efficacy of praziquantel and artemisinin derivatives for the treatment and prevention of human schistosomiasis: a systematic review and meta-analysis. *Parasites & Vectors* 4, p. 201.

Liu, Y. H., Wang, X. G., Gao, P., & Qian, M. X., 1991. Experimental and clinical-trial of albendazole in the treatment of *Clonorchiasis sinensis*. *Chinese Medical Journal* 104(1), pp. 27-31.

Lo, N. C., Bogoch, I. I., Blackburn, B. G., Raso, G., N'Goran, E. K., Coulibaly, J. T., Becker, S. L., Abrams, H. B., Utzinger, J., & Andrews, J. R., 2015. Comparison of community-wide, integrated mass drug administration strategies for schistosomiasis and soil-transmitted helminthiasis: a cost-effectiveness modelling study. *The Lancet Global Health* 3(10), p. e629-e638.

Lobo, D. A., Velayudhan, R., Chatterjee, P., Kohli, H., & Hotez, P. J., 2011. The neglected tropical diseases of India and South Asia: review of heir prevalence, distribution, and control or elimination. *Plos Neglected Tropical Diseases* 5(10), p. e1222.

Loukas, A. & Prociv, P., 2001. Immune responses in hookworm infections. *Clinical Microbiology Reviews* 14(4), pp. 689-703.

Lun, Z. R., Gasser, R. B., Lai, D. H., Li, A. X., Zhu, X. Q., Yu, X. B., & Fang, Y. Y., 2005. Clonorchiasis: a key foodborne zoonosis in China. *The Lancet Infectious Diseases* 5(1), pp. 31-41.

Lunn, D., Spiegelhalter, D., Thomas, A., & Best, N., 2009. The BUGS project: Evolution, critique and future directions. *Statistics in Medicine* 28(25), pp. 3049-3067.

Mackey, T. K., Liang, B. A., Cuomo, R., Hafen, R., Brouwer, K. C., & Lee, D. E., 2014. Emerging and reemerging neglected tropical diseases: a review of key characteristics, risk factors, and the policy and innovation environment. *Clinical Microbiology Reviews* 27(4), pp. 949-979.

Matérn, B., 1960. Spatial variation. *Meddelanden fran Statens Skogsforskningsinstitut* 49(5), pp. 1-144.

McCreesh, N. & Booth, M., 2013. Challenges in predicting the effects of climate change on *Schistosoma mansoni* and *Schistosoma haematobium* transmission potential. *Trends in Parasitology* 29(11), pp. 548-555.

Medeiros, L. C. D., Castilho, C. A. R., Braga, C., de Souza, W. V., Regis, L., & Monteiro, A. M. V., 2011. Modeling the dynamic transmission of dengue fever: investigating disease persistence. *Plos Neglected Tropical Diseases* 5(1), p. e942.

Metropolis, N., Rosenbluth, A. W., Rosenbluth, M. N., Teller, A. H., & Teller, E., 1953. Equations of state calculations by fast computing machines. *Journal of Chemical Physics* 21, pp. 1087-1091.

Meurs, L., Mbow, M., Vereecken, K., Menten, J., Mboup, S., & Polman, K., 2012. Epidemiology of mixed *Schistosoma mansoni* and *Schistosoma haematobium* infections in northern Senegal. *International Journal for Parasitology* 42(3), pp. 305-311.

Ministry of Health, 2006. Notice of the Ministry of Public Health Concerning Publishing "National Control Program on Important Parasitic Diseases in 2006-2015". *Gazette of the Ministry of Health of People's Republic of China* 33, pp. 41-44.

Moher, D., Liberati, A., Tetzlaff, J., & Altman, D. G., 2009. Preferred reporting items for systematic reviews and meta-analyses: the PRISMA Statement. *Plos Medicine* 6(7), p. e1000097.

Molyneux, D. H., 2010. Neglected tropical diseases--beyond the tipping point? *The Lancet* 375(9708), pp. 3-4.

Montressor, A., Crompton, D. W. T., Hall, A., Bundy, D. A. P., & Savioli, L. 1998. *Guidelines for the evaluation of soil-transmitted helminthiasis and schistosomiasis at community level: a guide for managers of control programmes (WHO/CTD/SIP/98.1).* Geneva: World Health Organization.

Moser, W., Ali, S. M., Ame, S. M., Speich, B., Puchkov, M., Huwyler, J., Albonico, M., Hattendorf, J., & Keiser, J., 2016. Efficacy and safety of oxantel pamoate in school-aged children infected with *Trichuris trichiura* on Pemba Island, Tanzania: a parallel, randomised, controlled, dose-ranging study. *The Lancet Infectious Diseases* 16(1), pp. 53-60.

Muench, H. 1959. *Catalytic model in epidemiology.* Harvard University Press, Cambridge, MA, USA.

Murray, C. J. L., Vos, T., Lozano, R., Naghavi, M., Flaxman, A. D., Michaud, C., Ezzati, M., Shibuya, K., Salomon, J. A., Abdalla, S., et al, 2012. Disability-adjusted life years (DALYs) for 291 diseases and injuries in 21 regions, 1990-2010: a systematic analysis for the Global Burden of Disease Study 2010. *The Lancet* 380(9859), pp. 2197-2223.

Mutapi, F., Gryseels, B., & Roddam, A., 2003. On the calculation of intestinal schistosome infection intensity. *Acta Tropica* 87(2), pp. 225-233.

Muth, S., Sayasone, S., Odermatt-Biays, S., Phompida, S., Duong, S., & Odermatt, P., 2010. *Schistosoma mekongi* in Cambodia and Lao People's Democratic Republic. *Advances in Parasitology, Vol 72* 72, pp. 179-203.

Mwinzi, P. N., Kittur, N., Ochola, E., Cooper, P. J., Campbell, C. H., Jr., King, C. H., & Colley, D. G., 2015. Additional evaluation of the point-of-contact circulating cathodic antigen assay for *Schistosoma mansoni* infection. *Frontiers in Public Health* 3, p. 48.

Nikolay, B., Brooker, S. J., & Pullan, R. L., 2014. Sensitivity of diagnostic tests for human soil-transmitted helminth infections: a meta-analysis in the absence of a true gold standard. *International Journal for Parasitology* 44(11), pp. 765-774.

Norhayati, M., Oothuman, P., & Fatmah, M. S., 1998. Some risk factors of Ascaris and Trichuris infection in Malaysian aborigine (Orang Asli) children. *Medical Journal of Malaysia* 53(4), pp. 401-407.

Nouvellet, P., Cucunubá, Z. M., & Gourbière, S., 2015. Ecology, evolution and control of Chagas disease: a century of neglected modelling and a promising future. *Advances in Parasitology* 87, pp. 135-191.

Oh, J. K., Lim, M. K., Yun, E. H., Cho, H., Park, E. Y., Choi, M. H., Shin, H. R., & Hong, S. T., 2014. Control of clonorchiasis in Korea: effectiveness of health education for community leaders and individuals in an endemic area. *Tropical Medicine & International Health* 19(9), pp. 1096-1104.

Oliveira, E. C. & Paumgartten, F. J. R., 2000. Toxicity of Euphorbia milli latex and niclosamide to snails and nontarget aquatic species. *Ecotoxicology and Environmental Safety* 46(3), pp. 342-350.

Pathmeswaran, A., Jayatissa, R., Samarasinghe, S., Fernando, A., de Silva, R. P., Thattil, R. O., & de Silva, N. R., 2005. Health status of primary schoolchildren in Sri Lanka. *Ceylon Medical Journal* 50(2), pp. 46-50.

Pekoz, E. A., Rollin, A., Cekanavicius, V., & Shwartz, M., 2009. A Three-Parameter Binomial Approximation. *Journal of Applied Probability* 46(4), pp. 1073-1085.

Pekoz, E. A., Shwartz, M., Christiansen, C. L., & Berlowitz, D., 2010. Approximate models for aggregate data when individual-level data sets are very large or unavailable. *Statistics in Medicine* 29(21), pp. 2180-2193.

Petney, T. N., Andrews, R. H., Saijuntha, W., Wenz-Mucke, A., & Sithithaworn, P., 2013. The zoonotic, fish-borne liver flukes Clonorchis sinensis, Opisthorchis felineus and Opisthorchis viverrini. *International Journal for Parasitology* 43(12-13), pp. 1031-1046.

Phan, V. T., Ersboll, A. K., Do, D. T., & Dalsgaard, A., 2011. Raw-fish-eating behavior and fishborne zoonotic trematode infection in people of northern Vietnam. *Foodborne Pathogens and Disease* 8(2), pp. 255-260.

Pinheiro, I. D., de Castro, M. F., Mitterofhe, A., Pires, F. A. C., Abramo, C., Ribeiro, L. C., Tibirica, S. H. C., & Coimbra, E. S., 2011. Prevalence and risk factors for giardiasis and soil-transmitted helminthiasis in three municipalities of Southeastern Minas Gerais State, Brazil risk factors for giardiasis and soil-transmitted helminthiasis. *Parasitology Research* 108(5), pp. 1123-1130.

Plummer, M., Best, N., Cowles, K., & Vines, K., 2006. CODA: convergence diagnosis and output analysis for MCMC. *R News* 6(1), pp. 7-11.

Pullan, R. L., Bethony, J. M., Geiger, S. M., Cundill, B., Correa-Oliveira, R., Quinnell, R. J., & Brooker, S., 2008. Human helminth co-infection: analysis of spatial patterns and risk factors in a Brazilian community. *Plos Neglected Tropical Diseases* 2(12), p. e352.

Pullan, R. L. & Brooker, S. J., 2012. The global limits and population at risk of soil-transmitted helminth infections in 2010. *Parasites & Vectors* 5, p. 81.

Pullan, R. L., Gething, P. W., Smith, J. L., Mwandawiro, C. S., Sturrock, H. J. W., Gitonga, C. W., Hay, S. I., & Brooker, S., 2011. Spatial modelling of soil-transmitted helminth infections in Kenya: a disease control planning tool. *Plos Neglected Tropical Diseases* 5(2), p. e958.

Pullan, R. L., Smith, J. L., Jasrasaria, R., & Brooker, S. J., 2014. Global numbers of infection and disease burden of soil transmitted helminth infections in 2010. *Parasites & Vectors* 7, p. 37.

Qian, M. B., Chen, Y. D., Fang, Y. Y., Tan, T., Zhu, T. J., Zhou, C. H., Wang, G. F., Xu, L. Q., & Zhou, X. N., 2013a. Epidemiological profile of *Clonorchis sinensis* infection in one community, Guangdong, People's Republic of China. *Parasites & Vectors* 6, p. 194.

Qian, M. B., Chen, Y. D., Liang, S., Yang, G. J., & Zhou, X. N., 2012. The global epidemiology of clonorchiasis and its relation with cholangiocarcinoma. *Infectious Diseases of Poverty* 1(1), p. 4.

Qian, M. B., Chen, Y. D., & Yan, F., 2013. Time to tackle clonorchiasis in China. *Infectious Diseases of Poverty* 2(1), p. 4.

Qian, M. B., Utzinger, J., Keiser, J., & Zhou, X. N., 2016. Clonorchiasis. *The Lancet* 387(10020), pp. 800-810.

Qian, M. B., Yap, P., Yang, Y. C., Liang, H., Jiang, Z. H., Li, W., Utzinger, J., Zhou, X. N., & Keiser, J., 2013b. Accuracy of the Kato-Katz method and formalin-ether concentration technique for the diagnosis of *Clonorchis sinensis*, and implication for assessing drug efficacy. *Parasites & Vectors* 6(1), p. 314.

Qu, Z. Q., Chen, R. X., & Zeng, M. A., 1980. Preliminary study on the diagnosis of clonorchiasis by ELISA. *Sichuan Yi Xue (in Chinese)* 1, pp. 238-243.

Raso, G., Luginbuhl, A., Adjoua, C. A., Tian-Bi, N. T., Silue, K. D., Matthys, B., Vounatsou, P., Wang, Y., Dumas, M. E., Holmes, E., Singer, B. H., Tanner, M., N'Goran, E. K., & Utzinger, J., 2004. Multiple parasite infections and their relationship to self-reported morbidity in a community of rural Côte d'Ivoire. *International Journal of Epidemiology* 33(5), pp. 1092-1102.

Raso, G., Matthys, B., N'Goran, E. K., Tanner, M., Vounatsou, P., & Utzinger, J., 2005. Spatial risk prediction and mapping of *Schistosoma mansoni* infections among schoolchildren living in western Côte d'Ivoire. *Parasitology* 131(Pt 1), pp. 97-108.

Raso, G., Vounatsou, P., Gosoniu, L., Tanner, M., N'Goran, E. K., & Utzinger, J., 2006. Risk factors and spatial patterns of hookworm infection among schoolchildren in a rural area of western Côte d'Ivoire. *International Journal for Parasitology* 36(2), pp. 201-210.

Raso, G., Vounatsou, P., Mcmanus, D. P., N'Goran, E. K., & Utzinger, J., 2007. A Bayesian approach to estimate the age-specific prevalence of *Schistosoma mansoni* and implications for schistosomiasis control. *International Journal for Parasitology* 37(13), pp. 1491-1500.

Ray, S. D., 2015. Side effects of drugs annual: a worldwide yearly survey of new data in adverse drug reactions. 37, pp. 2-651. (Book Series)

Rim, H. J., 1986. The current pathobiology and chemotherapy of clonorchiasis. *The Korean Journal of Parasitology* 24 Suppl, pp. 1-141.

Rock, K. S., Stone, C. M., Hastings, I. M., Keeling, M. J., Torr, S. J., & Chitnis, N., 2015. Mathematical models of human african trypanosomiasis epidemiology. *Advances in Parasitology* 87, pp. 53-133.

Rohner, F., Zimmermann, M. B., Amon, R. J., Vounatsou, P., Tschannen, A. B., N'Goran, E. K., Nindjin, C., Cacou, M. C., Te-Bonle, M. D., Aka, H., Sess, D. E., Utzinger, J., &

Hurrell, R. F., 2010. In a randomized controlled trial of iron fortification, anthelmintic treatment, and intermittent preventive treatment of malaria for anemia control in Ivorian children, only anthelmintic treatment shows modest benefit. *Journal of Nutrition* 140(3), pp. 635-641.

Rollinson, D., Knopp, S., Levitz, S., Stothard, J. R., Tchuem Tchuente, L. A., Garba, A., Mohammed, K. A., Schur, N., Person, B., Colley, D. G., & Utzinger, J., 2013. Time to set the agenda for schistosomiasis elimination. *Acta Tropica* 128(2), pp. 423-440.

Ross, A. G., Vickers, D., Olds, G. R., Shah, S. M., & Mcmanus, D. P., 2007. Katayama syndrome. *The Lancet Infectious Diseases* 7(3), pp. 218-224.

Rue, H., Martino, S., & Chopin, N., 2009. Approximate Bayesian inference for latent Gaussian models by using integrated nested Laplace approximations. *Journal of the Royal Statistical Society Series B-Statistical Methodology* 71(2), pp. 319-392.

Sanderson, E. W., Jaiteh, M., Levy, M. A., Redford, K. H., Wannebo, A. V., & Woolmer, G., 2002. The human footprint and the last of the wild. *Bioscience* 52(10), pp. 891-904.

Sandoval, N., Siles-Lucas, M., Perez-Arellano, J. L., Carranza, C., Puente, S., Lopez-Aban, J., & Muro, A., 2006. A new PCR-based approach for the specific amplification of DNA from different *Schistosoma* species applicable to human urine samples. *Parasitology* 133(Pt 5), pp. 581-587.

Savioli, L., Gabrielli, A. F., Montresor, A., Chitsulo, L., & Engels, D., 2009. Schistosomiasis control in Africa: 8 years after World Health Assembly Resolution 54.19. *Parasitology* 136(13), pp. 1677-1681.

Savtchenko, A., Ouzounov, D., Ahmad, S., Acker, J., Leptoukh, G., Koziana, J., & Nickless, D., 2004. Terra and aqua MODIS products available from NASA GES DAAC. *Trace Constituents in the Troposphere and Lower Stratosphere* 34(4), pp. 710-714.

Scheipl, F., Fahrmeir, L., & Kneib, T., 2012. Spike-and-slab priors for function selection in structured additive regression models. *Journal of the American Statistical Association* 107(500), pp. 1518-1532.

Schmidt, A. M. & O'Hagan, A., 2003. Bayesian inference for non-stationary spatial covariance structure via spatial deformations. *Journal of the Royal Statistical Society Series B-Statistical Methodology* 65(3), pp. 743-758.

Scholte, R. G. C., Gosoniu, L., Malone, J. B., Chammartin, F., Utzinger, J., & Vounatsou, P., 2014. Predictive risk mapping of schistosomiasis in Brazil using Bayesian geostatistical models. *Acta Tropica* 132, pp. 57-63.

Scholte, R. G. C., Schur, N., Bavia, M. E., Carvalho, E. M., Utzinger, J., & Vounatsou, P., 2013. Spatial analysis and risk mapping of Soil Transmitted Helminths in Brazil, using Bayesian geostatiscal models. *Geospatial Health* 8(1), pp. 97-110.

Schratz, A., Pineda, M. F., Reforma, L. G., Fox, N. M., Le, A. T., Tommaso Cavalli-Sforza, L., Henderson, M. K., Mendoza, R., Utzinger, J., Ehrenberg, J. P., & Tee, A. S., 2010. Neglected diseases and ethnic minorities in the Western Pacific Region exploring the links. *Advances in Parasitology* 72, pp. 79-107.

Schur, N., Hürlimann, E., Garba, A., Traoré, M. S., Ndir, O., Ratard, R. C., Tchuente, L. A. T., Kristensen, T. K., Utzinger, J., & Vounatsou, P., 2011a. Geostatistical model-based

estimates of Schistosomiasis prevalence among individuals aged <= 20 years in West Africa. *Plos Neglected Tropical Diseases* 5(6), p. e1194.

Schur, N., Hürlimann, E., Stensgaard, A. S., Chimfwembe, K., Mushinge, G., Simoonga, C., Kabatereine, N. B., Kristensen, T. K., Utzinger, J., & Vounatsou, P., 2011b. Spatially explicit *Schistosoma* infection risk in eastern Africa using Bayesian geostatistical modelling. *Acta Trop.*

Schur, N., Vounatsou, P., & Utzinger, J., 2012. Determining treatment needs at different spatial scales using geostatistical model-based risk estimates of schistosomiasis. *Plos Neglected Tropical Diseases* 6(9), p. e1773.

Shen, C. H., Choi, M. H., Bae, Y. M., Yu, G., Wang, S. H., & Hong, S. T., 2007. A case of anaphylactic reaction to praziquantel treatment. *American Journal of Tropical Medicine and Hygiene* 76(3), pp. 603-605.

Shin, H. R., Oh, J. K., Lim, M. K., Shin, A., Kong, H. J., Jung, K. W., Won, Y. J., Park, S., Park, S. J., & Hong, S. T., 2010. Descriptive epidemiology of cholangiocarcinoma and clonorchiasis in Korea. *Journal of Korean Medical Science* 25(7), pp. 1011-1016.

Smith, B. J. & Cowles, M. K., 2007. Correlating point-referenced radon and areal uranium data arising from a common spatial process. *Journal of the Royal Statistical Society Series C-Applied Statistics* 56(3), pp. 313-326.

Smith, B. J., Yan, J., & Cowels, M. K., 2008. Unified geostatistical modeling for data fusion and spatial heteroskedasticity with R package ramps. *Journal of Statistical Software* 25(10), pp. 1-21.

Soares Magalhães, R. J., Biritwum, N. K., Gyapong, J. O., Brooker, S., Zhang, Y., Blair, L., Fenwick, A., & Clements, A. C., 2011. Mapping helminth co-infection and co-intensity: geostatistical prediction in ghana. *Plos Neglected Tropical Diseases* 5(6), p. e1200.

Speich, B., Ali, S. M., Ame, S. M., Bogoch, I. I., Alles, R., Huwyler, J., Albonico, M., Hattendorf, J., Utzinger, J., & Keiser, J., 2015. Efficacy and safety of albendazole plus ivermectin, albendazole plus mebendazole, albendazole plus oxantel pamoate, and mebendazole alone against *Trichuris trichiura* and concomitant soil-transmitted helminth infections: a four-arm, randomised controlled trial. *The Lancet Infectious Diseases* 15(3), pp. 277-284.

Speich, B., Ame, S. M., Ali, S. M., Alles, R., Huwyler, J., Hattendorf, J., Utzinger, J., Albonico, M., & Keiser, J., 2014a. Oxantel pamoate-albendazole for *Trichuris trichiura* infection. *New England Journal of Medicine* 370(7), pp. 610-620.

Speich, B., Knopp, S., Mohammed, K. A., Khamis, I. S., Rinaldi, L., Cringoli, G., Rollinson, D., & Utzinger, J., 2010. Comparative cost assessment of the Kato-Katz and FLOTAC techniques for soil-transmitted helminth diagnosis in epidemiological surveys. *Parasites & Vectors* 3, p. 71.

Speich, B., Utzinger, J., Marti, H., Ame, S. M., Ali, S. M., Albonico, M., & Keiser, J., 2014b. Comparison of the Kato-Katz method and ether-concentration technique for the diagnosis of soil-transmitted helminth infections in the framework of a randomised controlled trial. *European Journal of Clinical Microbiology & Infectious Diseases* 33(5), pp. 815-822.

Spiegelhalter, D. J., Best, N. G., Carlin, B. R., & van der Linde, A., 2002. Bayesian measures of model complexity and fit. *Journal of the Royal Statistical Society Series B-Statistical Methodology* 64(4), pp. 583-616.

Sripa, B., Kaewkes, S., Intapan, P. M., Maleewong, W., & Brindley, P. J., 2010. Food-borne trematodiases in Southeast Asia epidemiology, pathology, clinical manifestation and control. *Advances in Parasitology* 72, pp. 305-350.

Steinmann, P., Keiser, J., Bos, R., Tanner, M., & Utzinger, J., 2006. Schistosomiasis and water resources development: systematic review, meta-analysis, and estimates of people at risk. *The Lancet Infectious Diseases* 6(7), pp. 411-425.

Steinmann, P., Utzinger, J., Du, Z. W., & Zhou, X. N., 2010. Multiparasitism a neglected reality on global, regional and local scale. *Advances in Parasitology* 73, pp. 21-50.

Stensgaard, A. S., Vounatsou, P., Onapa, A. W., Simonsen, P. E., Pedersen, E. M., Rahbek, C., & Kristensen, T. K., 2011a. Bayesian geostatistical modelling of malaria and lymphatic filariasis infections in Uganda: predictors of risk and geographical patterns of co-endemicity. *Malaria Journal* 10, p. 298.

Stensgaard, A. S., Vounatsou, P., Onapa, A. W., Simonsen, P. E., Pedersen, E. M., Rahbek, C., & Kristensen, T. K., 2011b. Bayesian geostatistical modelling of malaria and lymphatic filariasis infections in Uganda: predictors of risk and geographical patterns of co-endemicity. *Malaria Journal* 10, p. 298.

Stolk, W. A., Stone, C., & de Vlas, S. J., 2015. Modelling lymphatic filariasis transmission and control: modelling frameworks, lessons learned and future directions. *Advances in Parasitology* 87, pp. 249-291.

Stoltzfus, R. J., Dreyfuss, M. L., Chwaya, H. M., & Albonico, M., 1997. Hookworm control as a strategy to prevent iron deficiency. *Nutrition Reviews* 55(6), pp. 223-232.

Stothard, J. R., Chitsulo, L., Kristensen, T. K., & Utzinger, J., 2009. Control of schistosomiasis in sub-Saharan Africa: progress made, new opportunities and remaining challenges. *Parasitology* 136(13), pp. 1665-1675.

Strunz, E. C., Addiss, D. G., Stocks, M. E., Ogden, S., Utzinger, J., & Freeman, M. C., 2014. Water, sanitation, hygiene, and soil-transmitted helminth infection: a systematic review and meta-analysis. *Plos Medicine* 11(3), p. e1001620.

Tang, C. C., Lin, Y. K., Wang, P. C., Chen, P. H., Tang, C. T., Chen, T. S., Lin, C. T., Huang, C. K., Chen, C. F., & Chen, S. H., 1963a. Clonorchiasis in south Fukien with special reference to the discovery of crayfishes as second intermediate host. *Chin Med J (Engl.)* 82, pp. 545-562.

Tchuem Tchuenté, L. A., 2011. Control of soil-transmitted helminths in sub-Saharan Africa: diagnosis, drug efficacy concerns and challenges. *Acta Tropica* 120(Suppl 1), p. S4-S11.

Tchuem Tchuenté, L. A., Behnke, J. M., Gilbert, F. S., Southgate, V. R., & Vercruysse, J., 2003a. Polyparasitism with *Schistosoma haematobium* and soil-transmitted helminth infections among school children in Loum, Cameroon. *Tropical Medicine & International Health* 8(11), pp. 975-986.

Tchuem Tchuenté, L. A., Southgate, V. R., Jourdane, J., Webster, B. L., & Vercruysse, J., 2003b. *Schistosoma intercalatum*: an endangered species in Cameroon? *Trends in Parasitology* 19(9), pp. 389-393.

Teesdale, C. H. & Amin, M. A., 1976. Comparison of the Bell technique, a modified Kato thick smear technique, and a digestion method for the field diagnosis of schistosomiasis mansoni. *Journal of Helminthology* 50(1), pp. 17-20.

Truscott, J., Hollingsworth, T. D., & Anderson, R., 2014. Modeling the interruption of the transmission of soil-transmitted helminths by repeated mass chemotherapy of school-age children. *Plos Neglected Tropical Diseases* 8(12), p. e3323.

Truscott, J. E., Turner, H. C., & Anderson, R. M., 2015. What impact will the achievement of the current World Health Organisation targets for anthelmintic treatment coverage in children have on the intensity of soil transmitted helminth infections? *Parasites & Vectors* 8, p. 551.

Umetsu, S., Sogo, T., Iwasawa, K., Kondo, T., Tsunoda, T., Oikawa-Kawamoto, M., Komatsu, H., Inui, A., & Fujisawa, T., 2014. Intestinal ascariasis at pediatric emergency room in a developed country. *World Journal of Gastroenterology* 20(38), pp. 14058-14062.

United Nations. World population prospects: the 2012 revision. http://esa.un.org/wpp (accessed 3 April 2016). New York: United Nations. (Database)

Utzinger, J., Bergquist, R., Shu-Hua, X., Singer, B. H., & Tanner, M., 2003. Sustainable schistosomiasis control--the way forward. *The Lancet* 362(9399), pp. 1932-1934.

Velasco-Forero, C. A., Sempere-Torres, D., Cassiraga, E. F., & Gomez-Hernandez, J. J., 2009. A non-parametric automatic blending methodology to estimate rainfall fields from rain gauge and radar data. *Advances in Water Resources* 32(7), pp. 986-1002.

Vercruysse, J., Albonico, M., Behnke, J. M., Kotze, A. C., Prichard, R. K., McCarthy, J. S., Montresor, A., & Levecke, B., 2011. Is anthelmintic resistance a concern for the control of human soil-transmitted helminths? *International Journal for Parasitology: Drugs and Drug Resistance* 1(1), pp. 14-27.

Viliamizar, E., Mendez, M., Bonilla, E., Varon, H., & deOnatra, S., 1996. *Ascaris lumbricoides* infestation as a cause of intestinal obstruction in children: experience with 87 cases. *Journal of Pediatric Surgery* 31(1), pp. 201-205.

Vounatsou, P., Raso, G., Tanner, M., N'Goran, E. K., & Utzinger, J., 2009. Bayesian geostatistical modelling for mapping schistosomiasis transmission. *Parasitology* 136(13), pp. 1695-1705.

Wang, W., Wang, L., & Liang, Y. S., 2012. Susceptibility or resistance of praziquantel in human schistosomiasis: a review. *Parasitology Research* 111(5), pp. 1871-1877.

Wang, X. B., Zhang, L. X., Luo, R. F., Wang, G. F., Chen, Y. D., Medina, A., Eggleston, K., Rozelle, S., & Smith, D. S., 2012. Soil-transmitted helminth infections and correlated risk factors in preschool and school-aged children in rural southwest China. *PLoS One* 7(9), p. e45939.

Wang, X. H., Zhou, X. N., Vounatsou, P., Chen, Z., Utzinger, J., Yang, K., Steinmann, P., & Wu, X. H., 2008. Bayesian spatio-temporal modeling of *Schistosoma japonicum*

prevalence data in the absence of a diagnostic 'gold' standard. *Plos Neglected Tropical Diseases* 2(6), p. e250.

Wang, Y. H., 1993. On the number of successes in independent trials. *Statistica Sinica* 3(2), pp. 295-312.

Warren, K. S., 1973. Regulation of the prevalence and intensity of schistosomiasis in man: immunology or ecology? *Journal of Infectious Diseases* 127(5), pp. 595-609.

Webster, R., Oliver, M. A., Muir, K. R., & Mann, J. R., 1994. Kriging the local risk of a rare disease from a register of diagnoses. *Geographical Analysis* 26(2), pp. 168-185.

WHO, 1995. Control of foodborne trematode infections. *WHO Technical Report Series* 849, pp. 1-157.

WHO, 2002a. Prevention and control of schistosomiasis and soil-transmitted helminthiasis: report of a WHO expert committee. *WHO Technical Report Series* 912, pp. 1-57.

WHO 2002b. *Report of the WHO informal consultation on the use of praziquantel during pregnancy/lactation and albendazole/mebendazole in children under 24 months (WHO/CDS/CPE/PVC/2002.4).* Geneva: World Health Organization.

WHO 2006. *Preventive chemotherapy in human helminthiasis: coordinated use of anthelminthic drugs in control interentions: a manual for health prefessionals and programme managers.* Geneva: World Health Organization.

WHO 2009. *Elimination of schistosomiasis from low-transmission areas: report of a WHO informal consultation (WHO/HTM/NTD/PCT/2009.2).* Geneva: World Health Organization.

WHO 2010a. *Informal Consultation on Expanding Schistosomiasis Control in Africa. Facilitating PZQ access to scale up schistosomiasis control.* Geneva: World Health Organization.

WHO 2010b. *Working to overcome the global impact of neglected tropical diseases. First WHO report on neglected tropical diseases (WHO/HTM/NTD/2010.1).* Geneva: World Health Organization.

WHO 2011a. *Helminth control in school-age children: a guide for managers of control programmes. Second edition.* Geneva : World Health Organization.

WHO 2011b. *Report of a meeting to review the results of studies on the treatment of schistosomiasis in preschool-aged children (WHO/HTM/NTD/PCT/2011.7).* Geneva: World Health Organization.

WHO, 2011c. Schistosomiasis number of people treated, 2009. *Weekly epidemiological record* 86(9), pp. 73-80.

WHO 2012a. *Accelerating work to overcome the global impact of Neglected Tropical Diseases. A roadmap for implementation (WHO/HTM/NTD/2012.1).* Geneva: World Health Organization.

WHO, 2012b. Schistosomiasis: population requiring preventive chemotherapy and number of people treated in 2010. *Weekly epidemiological record* 87(4), pp. 37-44.

WHO 2013. *Sustaining the drive to overcome the global impact of neglected tropical diseases: second WHO report on neglected tropcal diseases (WHO/HTM/NTD/2013.1).* Geneva: World Health Organization.

WHO, 2014. Preventive chemotherapy: planning, requesting medicines, and reporting. *Weekly epidemiological record* 89(8), pp. 61-72.

WHO. Neglected Tropical Diseases: PCT databank.
http://www.who.int/neglected_diseases/preventive_chemotherapy/databank/en (accessed 4 April 2016). (Database)

WHO/EMRO 2007. *Inter-country meeting on strategies to eliminate schistosomiasis from the Eastern Mediterranean Region.* Muscat, Oman 6-8 November 2007.

Wichmann, D., Panning, M., Quack, T., Kramme, S., Burchard, G. D., Grevelding, C., & Drosten, C., 2009. Diagnosing schistosomiasis by detection of cell-free parasite DNA in human plasma. *Plos Neglected Tropical Diseases* 3(4), p. e422.

Woolhouse, M. E. J., 1991. On the application of mathematical-models of schistosome transmission dynamics .1. natural transmission. *Acta Tropica* 49(4), pp. 241-270.

Woolhouse, M. E. J., 1998. Patterns in parasite epidemiology: the peak shift. *Parasitology Today* 14(10), pp. 428-434.

World Bank 2005. *Repositioning nutrition as central to development: a strategy for large-scale action.* Washington: World Bank.

World Bank 2014. *South Asia economic focus, spring 2014: time to refocus.* Washington: World Bank.

World Bank. World DataBank - Poverty and Equity Database.
http://databank.worldbank.org/data/reports.aspx?source=poverty-and-equity-database (accessed 3 April 2016). (Database)

Wu, W., Qian, X., Huang, Y., & Hong, Q., 2012. A review of the control of clonorchiasis sinensis and *Taenia solium* taeniasis/cysticercosis in China. *Parasitology Research* 111(5), pp. 1879-1884.

Ximenes, R., Southgate, B., Smith, P. G., & Guimaraes, N. L., 2003. Socioeconomic determinants of schistosomiasis in an urban area in the Northeast of Brazil. *Revista Panamericana de Salud Pública* 14(6), pp. 409-421.

Xu, L. L., Jiang, B., Duan, J. H., Zhuang, S. F., Liu, Y. C., Zhu, S. Q., Zhang, L. P., Zhang, H. B., Xiao, S. H., & Zhou, X. N., 2014. Efficacy and safety of praziquantel, tribendimidine and mebendazole in patients with co-infection of *Clonorchis sinensis* and other helminths. *Plos Neglected Tropical Diseases* 8(8), p. e3046.

Xu, L. Q., Yu, S. H., Jin, Z. X., Yang, J. L., Lai, C. Q., Zhang, X. J., & Zheng, C. Q., 1995. Soil-transmitted helminthiases: nationwide survey in China. *Bulletin of the World Health Organization* 73(4), pp. 507-513.

Yang, G. J., Liu, L., Zhu, H. R., Griffiths, S. M., Tanner, M., Bergquist, R., Utzinger, J., & Zhou, X. N., 2014. China's sustained drive to eliminate neglected tropical diseases. *The Lancet Infectious Diseases* 14(9), pp. 881-892.

Yang, H. M., 2003. Comparison between schistosomiasis transmission modelings considering acquired immunity and age-structured contact pattern with infested water. *Mathematical Biosciences* 184(1), pp. 1-26.

Yang, H. M., Coutinho, F. A. B., & Massad, E., 1997. Acquired immunity on a schistosomiasis transmission model - fitting the data. *Journal of Theoretical Biology* 188(4), pp. 495-506.

Yangco, B. G., Delerma, C., Lyman, G. H., & Price, D. L., 1987. Clinical study evaluating efficacy of praziquantel in clonorchiasis. *Antimicrobial Agents and Chemotherapy* 31(2), pp. 135-138.

Yap, P., Du, Z. W., Wu, F. W., Jiang, J. Y., Chen, R., Zhou, X. N., Hattendorf, J., Utzinger, J., & Steinmann, P., 2013. Rapid re-infection with soil-transmitted helminths after triple-dose albendazole treatment of school-aged children in Yunnan, People's Republic of China. *The American Journal of Tropical Medicine and Hygiene* 89(1), pp. 23-31.

Yapi, R. B., Hürlimann, E., Houngbedji, C. A., Ndri, P. B., Silue, K. D., Soro, G., Kouame, F. N., Vounatsou, P., Furst, T., N'Goran, E. K., Utzinger, J., & Raso, G., 2014. Infection and co-infection with helminths and Plasmodium among school children in Côte d'Ivoire: results from a National Cross-Sectional Survey. *Plos Neglected Tropical Diseases* 8(6), p. e2913.

Yapi, Y. G., Briet, O. J., Diabate, S., Vounatsou, P., Akodo, E., Tanner, M., & Teuscher, T., 2005. Rice irrigation and schistosomiasis in savannah and forest areas of Côte d'Ivoire. *Acta Tropica* 93(2), pp. 201-211.

Yu, S. H., Xu, L. Q., Jiang, Z. X., Xu, S. H., Han, J. J., Zhu, Y. G., Chang, J., Lin, J. X., & Xu, F. N., 1994. Nationwide survey of human parasite in China. *The Southeast Asian Journal of Tropical Medicine and Public Health* 25(1), pp. 4-10.

Zhai S & Sun A, 2012. On the relationship between altitude and economy--the inspiration of altitude effects to the economic development of the Qinghai-Tibet plateau region. *Nationalities Research in Qinghai (in Chinese)* 23(2), pp. 152-159.

Zhang, P., Feng, Z. L., & Milner, F., 2007. A schistosomiasis model with an age-structure in human hosts and its application to treatment strategies. *Mathematical Biosciences* 205(1), pp. 83-107.

Zhang, Q. W., Huang F.Y., & Geng Y.J., 2009. Relationship between propagation of *Clonorchis sinensis* and ecology cultivation. *China Tropical Medicine* 9, pp. 1012-1125.

Zheng, C. L., 2009. *Research on custom about eating sashimi. MSc Thesis, Guangxi University for Nationalities*.

Zheng, Q., Chen, Y., Zhang, H. B., Chen, J. X., & Zhou, X. N., 2009. The control of hookworm infection in China. *Parasites & Vectors* 2, p. 44.

Zhou, X. N., Bergquist, R., & Tanner, M., 2013. Elimination of tropical disease through surveillance and response. *Infectious Diseases of Poverty* 2(1), p. 1.

Ziegelbauer, K., Speich, B., Mausezahl, D., Bos, R., Keiser, J., & Utzinger, J., 2012. Effect of sanitation on soil-transmitted helminth infection: systematic review and meta-analysis. *Plos Medicine* 9(1), p. e1001162.

Curriculum vitae

Yingsi Lai

Date and place of birth	October 1, 1984 in Beihai, Guangxi, P.R. China
Nationality	Chinese
E-mail	yingsi.lai@unibas.ch, laiyingsi@163.com
Current address	Leuengasse 8, 4057 Basel, Switzerland

EDUCATION

Sep.2012-Now	PhD studies in Epidemiology Swiss Tropical and Public Health Institute (Swiss TPH)/ University of Basel, Basel, Switzerland Thesis: Bayesian geostatistical and mathematical models to assess the geographical distribution of neglected tropical diseases (supervision: PD DR. P. Vounatsou and Prof. Dr. J. Utzinger)
Sep.2010-Feb.2012	Master of Science (MSc) in infection biology and epidemiology Swiss TPH/University of Basel, Basel, Switzerland Thesis: Geographical distribution and burden of soil-transmitted helminth infections in China (supervision: PD DR. P. Vounatsou and Prof. Dr. J. Utzinger)
Sep.2003-Jul.2008	Bachelor of Medicine in Preventive Medicine School of Public Health, Peking University, Beijing, China
Sep.2005-Jul.2008	Bachelor of Science in Mathematics and Applied Mathematics (double major) School of Mathematical Sciences, Peking University, Beijing, China

PROFESSIONAL EXPERIENCE

Mar.2012-Aug.2012	Research Assistant Biostatistics Unit, Epidemiology and Public Health Department, Swiss TPH
Feb.2010-Aug.2010	Research Assistant The George Institute China, Beijing, China
Sep.2008-Dec.2009	Research Assistant State Key Laboratory of Biological Membrane and Membrane Bioengineering, Peking University, Beijing, China

PRESENTATIONS & CONFERENCES

Jul.2014	"A Systematic Review and Geostatistical Meta-Analysis on

	Schistosoma Infection in Sub-Saharan Africa" (oral), in conference Parasitology and Tropical Medicine 2014, Zurich, Switzerland
Jun. 2013	Attendance in conference Spatial Statistics 2013, Columbus, Ohio, USA
Dec.2012	"Bayesian modelling of *Schistosoma mansoni* age-prevalence curves" (poster), in conference Bayesian methods in Biostatistics and Bioinformatics 2012, Barcelona, Spain

PUBLICATIONS

Lo NC, **Lai YS**, Karagiannis-Voules DA, Bogoch II, Coulibaly JT, Bendavid E, et al (2016) Re-evaluating global guidelines for preventive chemotherapy against schistosomiasis and soil-transmitted Helminthiasis: a cost-effectiveness modeling study. *Lancet Infect Dis*, in press.

Lai YS, Biedermann P, Ekpo UF, Garba A, Mathieu E, Midzi N, et al (2015) Spatial distribution of schistosomiasis and treatment needs in sub-Saharan Africa: a systematic review and geostatistical analysis. *Lancet Infect Dis*, doi: 10.1016/S1473-3099(15)00066-3.

Assaré RK, **Lai YS**, Yapi A, Tian-Bi YN, Ouattara M, Yao PK, et al (2015) The spatial distribution of *Schistosoma mansoni* infection in four regions of western Côte d'Ivoire. *Geospat Health*, doi: 10.4081/gh.2015.345.

Lai YS, Zhou XN, Utzinger J, Vounatsou P (2013) Bayesian geostatistical modelling of soil-transmitted helminth survey data in the People's Republic of China. *Parasit Vectors*, doi:10.1186/1756-3305-6-359.

Zhang HB, Li RC, Xu M, Xu SM, **Lai YS**, Wu HD, et al. (2013) Ultrastructural uncoupling between T-tubules and sarcoplasmic reticulum in human heart failure. *Cardiovasc Res*, doi: 10.1093/cvr/cvt030.

Wu HD, Xu M, Li RC, Guo L, **Lai YS**, Xu SM, et al. (2012) Ultrastructural remodelling of Ca2+ signalling apparatus in failing heart cells. *Cardiovasc Res*, doi: 10.1093/cvr/cvs195.