The spatial-temporal risk profiling of *Clonorchis sinensis* infection over 50 years implies the effectiveness of control programs in South Korea: a geostatistical modeling study

Hai-Yan Xiao,[a] Jong-Yil Chai,[b] Yue-Yi Fang,[c] and Ying-Si Lai[a,d,]*

[a]Department of Medical Statistics, School of Public Health, Sun Yat-sen University, Guangzhou, Guangdong, PR China
[b]Department of Parasitology and Tropical Medicine, Seoul National University College of Medicine, and Institute of Endemic Diseases, Seoul National University Medical Research Center, Seoul, 110-799, South Korea
[c]Institute of Parasitic Diseases, Guangdong Provincial Center for Disease Control and Prevention, Guangzhou, Guangdong, PR China
[d]Sun Yat-sen Global Health Institute, Sun Yat-sen University, Guangzhou, Guangdong, PR China

Summary

Background Over the past 50 years, two national control programs on *Clonorchis sinensis* infection have been conducted in South Korea. Spatial-temporal profiles of infection risk provide useful information on assessing the effectiveness of the programs and planning spatial-targeted control strategies.

Methods Advanced Bayesian geostatistical joint models with spatial-temporal random effects were developed to analyze disease data collecting by a systematic review with potential influencing factors, and to handle issues of preferential sampling and data heterogeneities. Changes of the infection risk were analyzed.

Findings We presented the first spatial-temporal risk maps of *C. sinensis* infection at 5×5 km^2 resolution from 1970 to 2020 in South Korea. Moderate-to-high risk areas were shrunk, but temporal variances were shown in different areas. The population-adjusted estimated prevalence across the country was 5.99% (95% BCI: 5.09–7.01%) in 1970, when the first national deworming campaign began. It declined to 3.95% (95% BCI: 2.88–3.95%) in 1995, when the campaign suspended, and increased to 4.73% (95% BCI: 4.00–5.42%) in 2004, just before the Clonorchiasis Eradication Program (CEP). The population-adjusted prevalence was estimated at 2.77% (95% BCI: 1.67–4.34%) in 2020, 15 years after CEP started, corresponding to 1.42 (95% BCI: 0.85–2.23) million infected people.

Interpretation The first nationwide campaign and the CEP showed effectiveness on control of *C. sinensis* infection. Moderate-to-high risk areas identified by risk maps should be prioritized for control and intervention.

Funding The National Natural Science Foundation of China (project no. 82073665) and the Natural Science Foundation of Guangdong Province (project no. 2022A1515010042).

Keywords: *Clonorchis sinensis*; Preferential sampling; Bayesian spatial-temporal joint model; High-resolution risk map; Control program

Introduction

Clonorchiasis, caused by infection with *Clonorchis sinensis*, is one of the most important food-borne trematodiases in humans.[1,2] Human beings are mainly infected through ingestion of raw or undercooked freshwater fish infected with *C. sinensis* metacercaria.[3,4] The parasite can damage the liver and biliary systems, leading to cholelithiasis, cholangitis, or even fatal cholangiocarcinoma.[1,2] Preventive chemotherapy, information, education, and communication (IEC), and environmental modification are the major control measures.[5,6] *C. sinensis* infection is endemic mainly in Asia, and countries reporting the cases include China, South Korea, Vietnam, and Russia.[4] Particularly, clonorchiasis is one of the major parasitic diseases in South Korea. In the 1970s, soil-transmitted helminth infections, *Trichostrongylus orientalis* infection, and *C. sinensis* infection were the leading intestinal parasitic diseases in the country.[7] After effective control

*Corresponding author. Department of Medical Statistics, School of Public Health, Sun Yat-sen University, Guangzhou, Guangdong, 510080, PR China.

E-mail address: laiys3@mail.sysu.edu.cn (Y.-S. Lai).

Research in context

Evidence before this study

Clonorchiasis, one of the most important foodborne trematodiases, brings heavy, but neglected health threats to people living in South Korea. We searched two general databases (PubMed and ISI Web of Science) and three South Korean databases (NAVER, KISS and RISS) from inception to April 30, 2022, with search terms "(Liver fluke* OR Clonorchi*) AND Korea" and "'간흡충' OR '간디스토마' OR 'Clonorchis sinensis' OR 'Clonorchiasis'", respectively. As we tried to identify all potentially relevant studies, we set no restrictions for publication time, language, and study design. We found high-resolution spatial-temporal risk maps of *C. sinensis* infection haven't been produced in South Korea. Additionally, there are two nationwide control programs regarding *C. sinensis* infection have been conducted in South Korea in the past 50 years. The first program was the national deworming campaign from 1969 to 1995, and the second was the Clonorchiasis Eradication Program (CEP) from 2005. However, we found no article conducting a comprehensive assessment of the effectiveness of the two programs. Only a few ones conducted assessment in localized areas or among selected populations.

Added value of this study

This is the first study to estimate the yearly spatial-temporal risk maps of *C. sinensis* infection at high spatial resolution (5 × 5 km^2) in South Korea over 50 years. Our estimations were based on rigorous Bayesian geostatistical joint modeling approach taking into account preferential sampling issue and data heterogeneities. In our analysis, we identified moderate- and high-risk areas in different periods, presented the temporal trends of infection risk in different areas, and estimated the overall trend of the population-adjusted prevalence, which provided a valuable reference for understanding the risk of *C. sinensis* infection in this country and implied the effectiveness of the two nationwide control programs.

Implications of all the available evidence

The risk maps and overall trend of the estimated prevalence implied the first national deworming program and the CEP were effective. Even though low average level of infection risk in the country, several areas in river basins of the central eastern and the southern regions were still in moderate infection risk, which should be prioritized for control and intervention. All findings provide important information on spatial-targeting control and preventive strategies on *C. sinensis* infection in South Korea.

programs, the prevalence of these worm infections was mostly below 1% in recent days, except for *C. sinensis*, suggesting more focus should be paid on control and prevention of this disease.[8] The total medical expense was approximately 222 million won in 2018, ranking fourth among all parasitic diseases in the country.[9]

According to the results of the national surveys, the raw prevalence of *C. sinensis* infection in South Korea were 4.6%, 1.8%, 2.6%, 2.7%, 2.2%, 1.4%, 2.9%, and 1.9% in the year 1971, 1976, 1981, 1986, 1992, 1997, 2004, and 2012, respectively,[8] suggesting relatively low prevalence across the country. However, the infection risk in different areas show diversity in space and time.[10,11] Particularly, high prevalence was found in river basins.[12,13] There are two nationwide control programs regarding *C. sinensis* infection in South Korea. The first was the national deworming campaign from 1969 to 1995 by the Korea Association of Health Promotion, with the aim to eradicate intestinal worm infections.[14] The program included nationwide biannual school-based mass fecal screening and anthelmintic administrations on children tested positive of parasitic infection, education for the students and the public, and sanitation improvement.[15] A report on this campaign suggested a significant decreasing trend in egg-positive rates of *C. sinensis* among students across the country,[15] however, area- or regional specific changes were not reported in this study. The second program was the

Clonorchiasis Eradication Program (CEP) implemented by the Korea Centers for Disease Control and Preventions (KCDC) from 2005, with the goal to reduce the infection rate by diagnosing, treating, and educating residents in endemic areas.[14,16,17] Since the two programs were launched in different areas at different time and with different implementation efforts, it is important to evaluate the effectiveness of the programs comprehensively in different areas across the country,[18] which has not be done to our knowledge. Only a few ones conducted assessment in localized areas or among selected populations, such as Kim et al.'s study on changes of egg-positive rates of *C. sinensis* among students before and after the first national program,[15] Park et al.'s study showing a downward trend in infection rate of *C. sinensis* among riverside residents from 2004 (10.79%) to 2008 (8.53%),[11] and Oh et al.'s study suggesting a decreasing trend of prevalence in Sancheong-gun from 2004 (31.1%) to 2007 (14.9%).[19]

Understanding the spatial-temporal distribution of *C. sinensis* infection at high-resolution is important for assessment of control programs and development of spatial targeted control and preventive strategies. However, such risk maps have not been produced in South Korea, and simple statistical description of historical survey data is not enough to obtain such information. Although analysis of the temporal changes of infection risk in areas or different area levels could inform the

spatial-temporal distribution of *C. sinensis* infection, however, such analysis was not yet done across the country. Bayesian geostatistical modeling is one of the most rigorous approaches to produce high-resolution disease risk maps,[20] which has been applied in various studies on food-borne trematodiases in several countries or regions.[21,22] The approach models point-referenced disease data with potential risk factors (e.g., socioeconomic and environmental factors) and spatial-temporal random effects, thus estimates disease risk in areas without observed data.[23] Recently, more advanced Bayesian geostatistical models were developed to deal with preferential sampling issue and data heterogeneities,[22] such as jointly analyzing diseases data at both point- and areal levels,[24] and joint modeling of disease data which were partially missed the total number examined.[22]

In this study, we aimed to understand the spatial-temporal distribution of *C. sinensis* infection in South Korea, based on which, changes of the infection risk during the study period across the study region was assessed. We collected available geo-referenced disease data and data of potential influencing factors, following which, an advanced Bayesian geostatistical joint model was built and yearly risk maps of *C. sinensis* infection in South Korea at 5×5 km^2 were produced.

Methods

Ethics statement

This work was based on survey data of clonorchiasis, which were derived from peer-reviewed published literatures. Statements of ethics approval were included in the original sources. All data in this study were aggregated at point- or areal level and did not contain any identifiable information at the individual or household levels. So, there are no specific ethical issues warranted.

Disease data

We did a systematic review to collect data related to prevalence of *C. sinensis* infection in South Korea (registered in the International Prospective Register of Systematic Reviews No. CRD42021234803), and reported it following the PRISMA guidelines (Appendix A).[25] Two general databases (PubMed and ISI Web of Science) and three South Korean databases (NAVER, KISS and RISS) were searched from inception to April 30, 2022, with search terms "(Liver fluke* OR Clonorchi*) AND Korea" and "'간흡충' OR '간디스토마' OR 'Clonorchis sinensis' OR 'Clonorchiasis'", respectively. There were no restrictions on publication time, language, and study design. Additionally, we considered other potential literatures, such as reports from governments or Ministry of Health, documents from research groups, and relevant books.

A protocol is listed in Appendix B, with clear criteria for inclusion, exclusion, and extraction of data. Briefly, we included prevalence-related community-based surveys (i.e., with information on number of examined and number of positive, or information on prevalence) conducted from 1970 onwards, reporting data at provincial level and below, such as administrative divisions of level one (ADM1: province, etc.), level two (ADM2: city, etc.), and point-level (village, etc.). Studies were excluded if they were in-vitro investigations, or absence of human studies, or with specific study designs (e.g., case-control studies) or specific population groups (e.g., patients), or with study locations/areas not clearly identified, or with ineffective or unidentified diagnostic methods (e.g., direct smear), or with small sample size (less than 10). Firstly, we screened titles and abstracts to identify potentially relevant articles based on criteria for inclusion and exclusion. Then, the full-text review was conducted on relevant papers from the first screening round. 20% of randomly selected irrelevant papers from each round were re-checked for quality control.

Two independent reviewers assessed the quality of each included literature using an adapted nine-point quality assessment checklist (Appendix C).[22] We followed the GATHER checklist for data extraction (Appendix D). All included data were georeferenced by Google Maps (https://www.google.com/maps/) and entered into a database with detailed information (e.g., literature information, survey information, location information, and disease-related data), then verified by two independent reviewers. For surveys reported intervals of prevalence instead of exact observed values, we assigned the midpoints of the intervals as the observed prevalence. If studies adopted multiple diagnostic methods and reported the observed prevalence accordingly, the one with the most sensitive diagnostic method was recorded. If survey year was missing, we approximated it with the year of publication minus one, which was the median of the duration from surveys conducted to results published in the collected survey data.

Socioeconomic, environmental, and demographic data

Referring to similar geostatistical studies,[21,22,26] potential influencing factors, including the socioeconomic factors (i.e., human influence index [HII], urban extents, and travel time to the nearest big city) and environmental factors (i.e., distance to the nearest open water bodies, soil moisture, elevation, land surface temperature [LST] in the daytime, LST at night, normalized difference vegetation index [NDVI], land cover, and annual precipitation), and the demographic data were obtained from open-access databases, as listed in Appendix E. Data were integrated to a regular grid of 5×5 km^2 spatial resolution. As high-resolution data on potential influencing factors covering the full study period

(1970–2020) were unavailable, the average value of each factor within each grid pixel over available years was taken. LST and NDVI were averaged over the period of 2000–2020. Land cover was summarized by the most frequent category within each pixel over the period of 2001–2020. Besides, similar classes of land cover were re-grouped into six categories, that is (i) forests, (ii) grasslands and shrub, (iii) croplands, (iv) wet areas, (v) buildings, and (vi) barren areas. As the demographic data at high-spatial resolution were only available from 2000 onwards, population for previous years were estimated based on the available population data of the nearest year and the population growth rates obtained from United Nations' World Population Prospects,[27] with formula $P_m = P_n * e^{(m-n)*rate}$. Here P_m and P_n are the population counts in year m and year n ($m>n$), respectively.

Statistical analysis

We transformed multi-categorical predictors into dummy variables and standardized continuous ones with mean zero and standard deviation one. To avoid collinearity, Pearson's correlation or Spearman correlation (for those did not satisfy the assumption of linearity identified by scatter plots) between each pair of continuous predictors was calculated (Appendix M). If correlation coefficients greater than 0.8, we chose the ones more meaningful or with better data quality among pairs. To build a parsimonious model, Bayesian variable selection was adopted to find the best set of predictors. First, to identify the best functional form (continuous or categorical) of continuous predictors, we converted them to three-level categorical ones according to preliminary, exploratory, graphical analysis.[21] Bayesian geostatistical models were built with either form as the independent variable. The one with the lowest log score was selected as the best functional form.[28] Secondly, the backward elimination approach was applied in our study to select appropriate combinations of potential predictors. To be noted, the diagnostic methods (i.e., Kato-Katz, formalin-ether concentration technique [FE], or the combined method of FE and Kato-Katz) were kept in all combinations, according to the previous studies suggesting observed infection prevalence different between different methods.[22,29–31]

We adopted an advanced multivariate Bayesian geostatistical joint modeling approach to analyze the point- and areal level survey data together,[24] taking into account disease survey data reporting both the number of examined and positive, and those only reporting prevalence, and handling the issue of preferential sampling as well.[22] In brief, we assumed a binomial distribution for disease data reported both the number of examined and positive, and a beta distribution for those with only prevalence reported. A logit scale of prevalence was modeled with a linear combination of fixed effect predictors, spatial-temporal random effects and exchangeable non-spatial random effects. For areal level survey data, the average values of predictors and the average spatial-temporal random effects of pixels within the corresponding areas were assigned. The spatial-temporal random effects capturing the temporal correlation between different survey years and the spatial correlation between locations were assumed to arise from a zero-mean Gaussian distribution, with the covariance matrix the Kronecker product of a spatial matrix and a temporal one. The former was assumed following a stationary Matérn covariance function and the latter with auto-regressive order 1 (AR1). Additionally, regular temporal knots were set and latent spatial-temporal random fields were approximated by employing the B-spline basis function, to reduce the computational burden. The models were built under a Bayesian framework and INLA-SPDE approach was used for model fitting.[32] Data-driven model fitting was done based on all included prevalence data in different survey years and across different locations or areas. FE-based yearly prevalence at each pixel was further estimated based on posterior distributions of model parameters and spatial-temporal random effects. As local surveys other than the national surveys were probably from endemic areas, preferential sampling may occur.[33] To overcome the issue, we treated local survey locations as preferential sampling ones and assumed them following a marked log-Gaussian Cox process. A joint model, extended from Diggle et al.'s approach, was built,[33] linking the infection risk of the disease and the local survey locations by sharing the spatial-temporal random components. More details were shown in Appendix F.

We used the 5-fold out-of-sample cross-validation approach for model validation. To evaluate the performance of the model, we calculated mean error (ME), mean absolute error (MAE), mean square error (MSE), the percentage of observations covered by 95% Bayesian credible intervals (BCIs) of posterior estimated prevalence, and the area under the receiver-operating characteristic (ROC) curve (AUC).[22,34] Details on the model validation were shown in Appendix G. Additionally, sensitivity analysis was conducted to assess the effects of several issues (Appendix H).

A regular grid of 5×5 km^2 spatial resolution was overlay across South Korea, resulting in 5,946 pixels. The infection risk for each pixel each year from 1970 to 2020 was estimated using Bayesian kriging. Low, moderate, and high infection risk areas (with prevalence rates <5%, 5–20%, and >20%, respectively) were defined based on the WHO's recommendations and other literature.[6,35] All the statistical process was done in R (version 4.0.4) and risk maps were produced using ArcGIS (version 10.2). The country-level infection prevalence was calculated with population-weighted pixel-level infection risk.
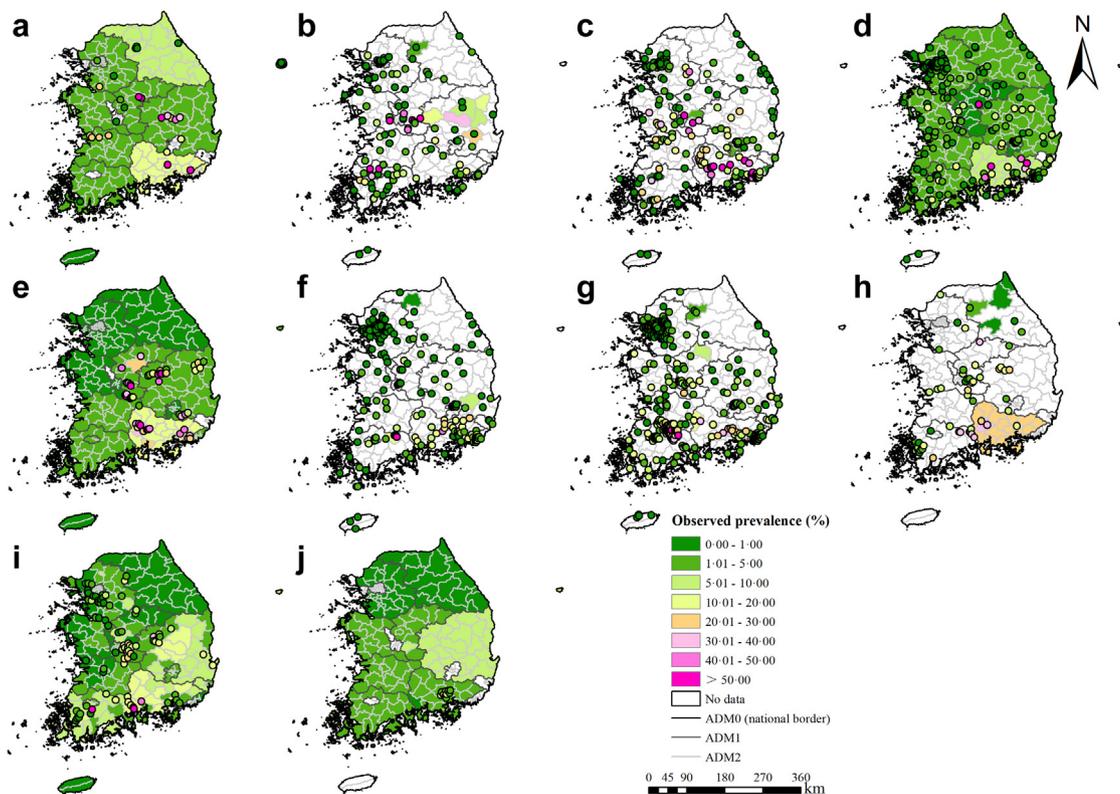
Fig. 1: Survey locations and observed prevalence of *C. sinensis* infection in South Korea. (a) 1970–1974, (b) 1975–1979, (c) 1980–1984, (d) 1985–1989, (e) 1990–1994, (f) 1995–1999, (g) 2000–2004, (h) 2005–2009, (i) 2010–2014, and (j) 2015 onward.

Role of the funding source

The funders played no role in the design or analysis of this study.

Results

Data summaries

A total of 12,838 records were identified through databases, while an additional 6 records were stemmed from other methods (Appendix I). 108 records remained according to the inclusion and exclusion criteria, resulting in 75 surveys at 15 ADM1-level divisions, 219 surveys at 185 ADM2-level divisions, and 1,360 surveys at 1,143 point-referenced locations. According to the quality assessment checklist, 94.44% of the included records showed good quality, with scores no less than 6 (Appendix C). The geographic distribution of observed prevalence of *C. sinensis* infection and the overview of survey characteristics were presented in Fig. 1 and Table 1, respectively. A large proportion of surveys (22.69%) were conducted between 2000 and 2004. Kato-Katz method was the most commonly used diagnostic method (74.50%).

Modeling fitting and validation

After variable selection, five variables were selected for the final geostatistical model (Table 2). The infection risk diagnosed by FE and the combined method of FE and Kato-Katz was 1.59 (95% BCI: 1.32–1.91), 2.08 (95% BCI: 1.86–2.33) times that of Kato-Katz method, respectively. The infection risk in areas with HII between 30 and 50 and those with a high HII (>50) was 0.80 (95% BCI: 0.65–0.999), 0.66 (95% BCI: 0.49–0.88) times that in areas with HII ≤30, respectively. Negative associations were found for the infection risk with the distance to the nearest open water bodies and annual precipitation. Increase in the distance to the nearest open water bodies or annual precipitation with one of its standard deviations was associated with 0.22 (95% BCI: 0.10–0.33) or 0.37 (95% BCI: 0.21–0.53) standard deviations decrease in the logit of the prevalence, respectively. NDVI was found positively correlated with the infection risk. Increase in NDVI of one of its standard deviations was associated with an expected increase in the logit of the prevalence of 0.14 (95% BCI: 0.001–0.27) of its standard deviations. Model validation showed that the model was able to correctly estimate 76.73% of locations within the 95% BCI, and the AUC of ROC was

Year of survey	1970–1974	1975–1979	1980–1984	1985–1989	1990–1994	1995–1999	2000–2004	2005–2009	2010–2014	2015 onward	Total
Number of articles	9	15	19	10	14	12	11	11	7	6	108
Number of surveys/locations[a]	38/37	129/124	184/180	228/225	88/73	231/229	375/373	60/59	287/267	33/19	1653/1322
Location type[a]											
ADM1	11/10	3/1	4/3	13/12	15/14	0/0	0/0	1/1	9/9	18/8	74/15
ADM2	0/0	7/7	6/6	2/2	5/5	3/3	4/4	7/6	180/178	5/1	219/185
Points	27/27	119/116	174/171	213/221	68/54	228/226	371/369	52/52	98/80	10/10	1360/1143
Diagnostic methods[a]											
Kato-Katz	21/20	100/97	166/165	220/218	37/37	199/199	290/290	6/6	209/209	0/0	1248/1067
FE[b]	17/17	24/23	13/13	3/3	21/21	10/9	81/80	46/46	78/58	33/19	326/248
Combined[c]	0/0	5/5	5/5	5/5	30/30	22/22	4/4	8/8	0/0	0/0	79/75
Average prevalence (%)[d]	17.33	7.29	10.42	3.68	24.31	3.37	7.53	11.67	4.60	3.72	7.40

[a]Presented as surveys/locations. [b]FE: formalin-ether concentration technique. [c]Reporting results based on diagnostic method combined with FE and Kato-Katz. [d]The mean prevalence calculated directly from all survey data.

Table 1: Overview of *C. sinensis* infection survey data in South Korea.

0.86, suggesting a reasonable capacity of prediction accuracy. The ME, MAE, and MSE were 0.36%, 4.50%, and 0.81%, respectively.

Spatial-temporal risk profiles

Moderate to high infection risk areas distributed mainly along the river basins (Fig. 2). Overall, moderate-to-high infection risk areas shrunk across the study period, but temporal variances were shown in different areas. In the first national deworming campaign period (1970–1994), there was a decreasing trend of infection risk in the western (e.g., Chungcheongnam-do, etc.) and densely populated northern regions (e.g., Seoul Teukbyeolsi, Incheon gwangyeoksi, Gangwon-do, and Gyeonggi-do, etc.). These regions also showed high probabilities of risk reduction (Fig. 3a). In the central (e.g., Chungcheongbuk-do and the western part of Gyeongsangbuk-do) and the southern regions (e.g., Gyeongsangnam-do, Jeollanam-do, Gwangju Gwangyeoksi, Ulsan Gwangyeoksi, and Busan Gwangyeoksi), the infection risk firstly showed a trend down until about 1985, and a gradual increase trend since then. In the gap period between the two national programs (1995–2004), the high infection risk areas in the central and southern regions shrunk to moderate risk ones. However, the moderate infection risk areas increased in the western and northern regions. The probabilities of risk reduction were high in the central regions but low in other regions (Fig. 3b). In the CEP period (from 2005 onward), the infection risk in most areas, particularly

Variable	Estimated median (95% BCI)	OR (95% BCI)	Prob (%)[a]
Diagnostic methods (Kato-Katz)[b]			
FE[c]	0.46 (0.28, 0.64)[e]	1.59 (1.32, 1.91)	>99.99
Combined[d]	0.73 (0.62, 0.85)[e]	2.08 (1.86, 2.33)	>99.99
Human influence index (≤30)[b]			
30–50	−0.22 (−0.43, −0.001)[e]	0.80 (0.65, 0.999)	0.03
≥50	−0.42 (−0.71, −0.12)[e]	0.66 (0.49, 0.88)	<0.01
Urban extents (Rural)[b]			
Urban	0.25 (−0.03, 0.52)	1.28 (0.97, 1.69)	0.95
Distance to the nearest open water bodies (km)	−0.22 (−0.33, −0.10)[e]	0.80 (0.72, 0.90)	<0.01
Annual precipitation (mm)	−0.37 (−0.53, −0.21)[e]	0.69 (0.59, 0.81)	<0.01
Normalized difference vegetation index	0.14 (0.001, 0.27)[e]	1.15 (1.001, 1.31)	0.97
Range (km)	54.73 (46.14, 64.36)	–	–
Spatial variance (σ_ϕ^2)	6.36 (5.20, 7.75)	–	–
Non-spatial variance (σ_{nonsp}^2)	0.60 (0.48, 0.74)	–	–
Temporal correlation coefficient (ρ)	0.08 (−0.06, 0.21)	–	–
Variance of beta-likelihood (σ_β^2)	0.02 (0.01, 0.02)	–	–
Tuning coefficient (α)	0.84 (0.74, 0.96)	–	–

[a]Posterior probability of OR >1. [b]In brackets, baseline values are reported. [c]FE: formalin-ether concentration technique. [d]Diagnostic method combined with FE and Kato-Katz. [e]Significant effect identified by not including zero in the 95% Bayesian credible interval (BCI) of posterior distribution of the corresponding coefficient.

Table 2: Posterior summaries of model parameters for *C. sinensis* infection.
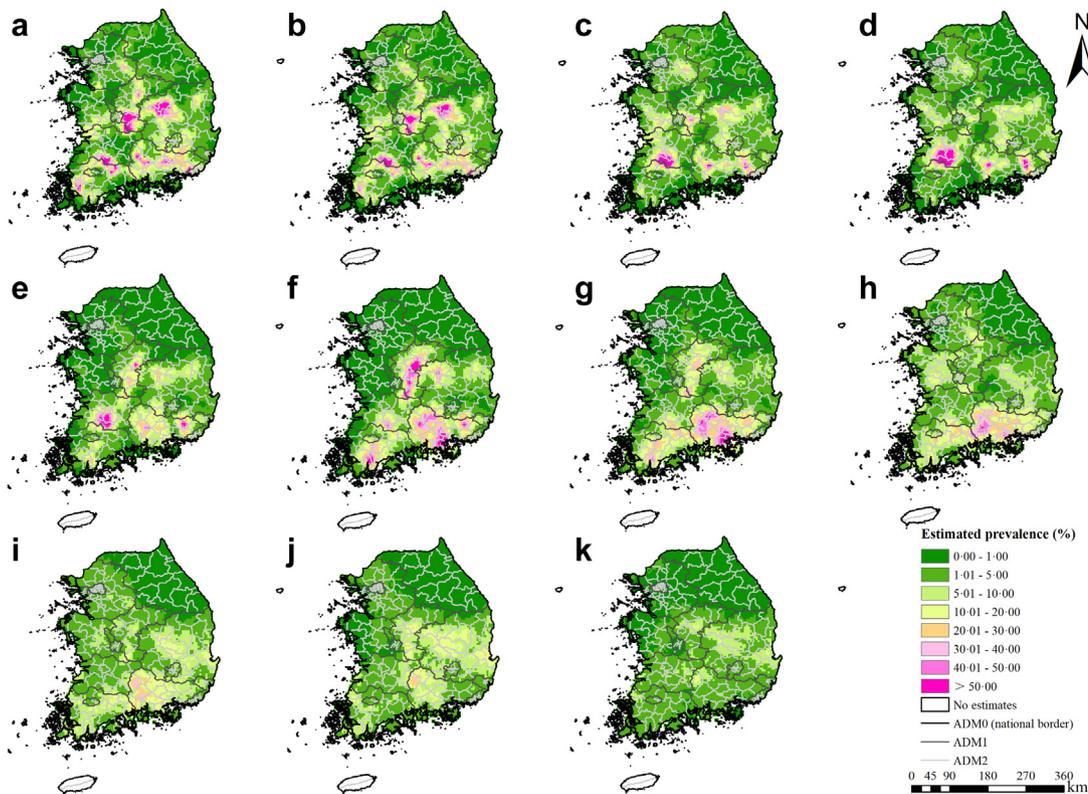
Fig. 2: Model-based estimated risk maps of *C. sinensis* infection across South Korea in selected years. Estimated prevalence was based on the median of the posterior estimated distribution of infection risk in (a) 1970, (b) 1975, (c) 1980, (d) 1985, (e) 1990, (f) 1995, (g) 2000, (h) 2005, (i) 2010, (j) 2015, and (k) 2020.

the ones in the southern part, decreased obviously, whereas a few moderate infection risk areas retained in central and eastern regions (e.g., the central and eastern parts of Gyeongsangbuk-do). The northern, the western, and the southern regions showed high probabilities of risk reduction (Fig. 3c). Besides, the maps for probabilities of infection risk >5% and >20% showed similar trends as Fig. 2 (Appendix J). High estimation uncertainty was mainly presented in the central and the southern regions (Appendix K).

The population-adjusted estimated prevalence
The population-adjusted estimated prevalence across the country showed an obvious decreasing trend until around 1990, followed by an upward until around 2005, and maintained a downward trend ever since (Fig. 4). The time-varying trend in the proportion of pixels with low to high risk is similar to Fig. 1. The proportion of high-risk areas declined from 1970 to 1988, then rose until around 1998, and finally maintained a slow downward trend, while the proportion of the areas at low to moderate risk showed the opposite overall trend, in which the proportion with prevalence <1% increased from 1970 to 1992, then fell to about 2007, and maintained an upward trend ever since. In 1970, the

population-adjusted estimated prevalence was 5.99% (95% BCI: 5.09–7.01%). 19.76% of areas were estimated at moderate risk and 8.29% at high risk, based on the median of the posterior distribution of infection risk. The prevalence declined to 3.95% (95% BCI: 2.88–3.95%) in 1995, the year of the first national campaign suspension, and then increased to 4.73% (95% BCI: 4.00–5.42%) in 2004, the year just before the CEP started. The population-adjusted prevalence was estimated at 2.77% (95% BCI: 1.67–4.34%) in 2020, 15 years after the CEP started. Based on the median of the posterior distribution of infection risk, the proportion of areas with moderate or high risk decreased to 10.12%. Additionally, we estimated that around 1.42 (95% BCI: 0.85–2.23) million people in the country were infected with *C. sinensis* in 2020.

Discussion
To our knowledge, we presented the first spatial-temporal risk maps of *C. sinensis* infection at spatial high-resolution (5 × 5 km²) in South Korea, using rigorous Bayesian geostatistical joint modeling approach. Our study identified moderate-to-high risk areas in different periods and presented the temporal
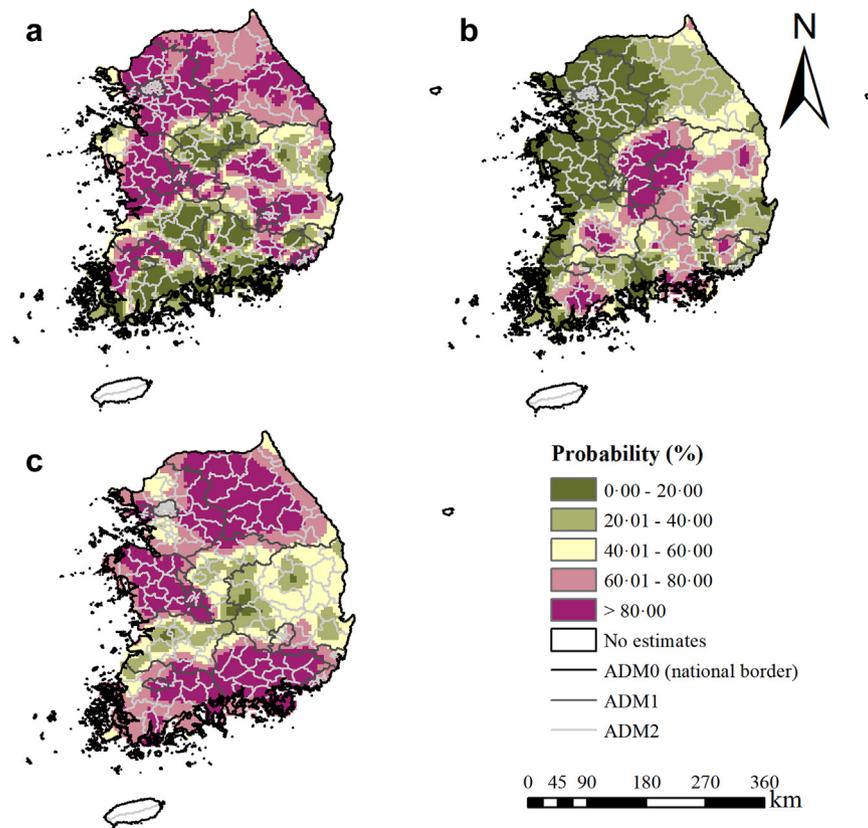
Fig. 3: The probability of risk reduction for *C. sinensis* infection across time periods. This probability was obtained by calculating the proportion of 500 posterior samples in each grid for which estimated prevalence of the latter year was lower than that of the former year. The probability (a) between 1970 and 1995, (b) between 1995 and 2005, and (c) between 2005 and 2020.

trends of infection risk in different areas, which provided an important reference for understanding the risk of *C. sinensis* infection in South Korea in the last 50 years and implied the effectiveness of national control programs.

The spatial-temporal risk profiles and the temporal changes of population-adjusted overall risk provide evidence on the effectiveness of the control programs. In the first national deworming campaign period (1970–1994), the population-adjusted estimated prevalence declined obviously and the areas with <1% prevalence expanded, suggesting that the campaign seemed effective on control of *C. sinensis* infection. However, there was a slight increase in the proportion of moderate-to-high risk areas and the overall prevalence from 1990, which may be attributed to the reduced frequency of examination and treatment of intestinal worm infections from biannual to annual.[15] Beside, with the localization of praziquantel, the public began to believe that if they ate the fish and then took praziquantel, they would not be infected, thus the behavior of raw fish consumption increased.[36] The overall risk increased and the areas with prevalence <1% shrank

from 1995 to 2004, probably due to the fact that the nationwide antiparasitic campaign was suspended, while some small-scale local interventions carried out in this period, which may result in the reduction risk in the central areas.[37] The overall infection risk declined gradually and the moderate-to-high risk area shrunk obviously since 2005, implying the CEP seemed effective. In practice, the two national programs started in different times in different areas with different intensity of implementation, which may lead to different local trends of infection risk.[15,18]

Compared with results from the eight national surveys, we estimated slightly higher *C. sinensis* infection prevalence (Appendix L). The national surveys reported raw overall prevalence by calculating the proportion of positive individuals among all examined ones,[8] which may be influenced by sampling framework.[22] Differently, our estimates relied on rigorous Bayesian geo-statistical modeling of all available geo-referenced disease survey data along with important influencing predictors, thus was able to take into account high-resolution spatial heterogeneous of both disease infection risk and population density across the country
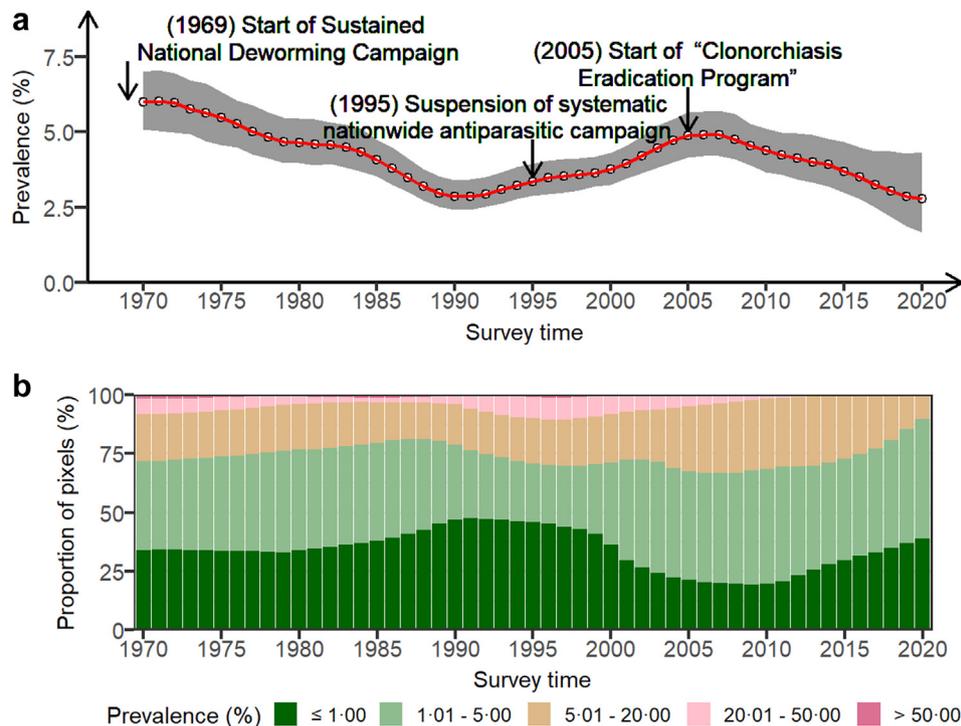
Fig. 4: Trends of estimated prevalence of *C. sinensis* infection in South Korea. (a) The trend of population-adjusted prevalence across the country, presented by the median and 95% Bayesian credible interval of the posterior distributions of population-adjusted estimated prevalence in each year. (b) The compositions of pixel-level areas with different levels of estimated prevalence (i.e., ≤1.00%, 1.01–5.00%, 5.01–20.00%, 20.01–50.00%, and >50.00%). The estimated prevalence was based on the median of the posterior distribution of infection risk in each grid.

when compiling country-level prevalence.[22] We estimated overall prevalence of 2.40% (95% BCI: 1.46–3.91%) of *C. sinensis* infection in South Korea in 2020, which was slightly lower than the results presented by Shin et al. (3.8% in 2020).[38] The latter study was based on survey data of residents living in the five major river basins, where the disease was endemic, while our results were based on population-weighted estimates across the country.

We did not provide estimates of infection risk for islands Jeju-do and Ulleung-gun, as sparse geo-referenced data were obtained, and they are quite a distance away from the mainland. It is improper to combine island and continent data for geostatistical modeling, due to spatial discontinuity.[39] Though, previous surveys suggested very low prevalence of *C. sinensis* infection in the two islands (0.00%–0.58% for Jeju-do from 1971 to 2012, and 0.4% for Ulleung-gun in 1973).[8,40] The temporal knots were set every 10 years to lower the computational burden. The results were satisfied using this time-knot setting, with reasonable model performance. And the temporal changes during the knots can be captured by the model, as the random effects of the years not being the time knots were approximated under the model by the data-driven fitting. Although we obtained annual risk map, we

presented the maps every 5 years, in order to facilitate the years of the two national programs started and suspended (e.g., 1995 and 2005), as well as for typographic aesthetics.

We identified several socioeconomic and environmental predictors significantly related with *C. sinensis* infection, which may provide insights in prevention and control of the disease. We found that surveys using FE or combined the two diagnostic methods (i.e., FE and Kato-Katz) tended to report higher prevalence than that using Kato-Katz method, in line with the previous studies.[30] Distance to the nearest water bodies was found negative associated with infection risk, consistent with knowledge that inhabitants living closer to waterbodies had more chance of freshwater fish consumption,[18,21] suggesting that residents living in the river basins should be prioritized for effective control and prevention of the disease. Our results showed that the infection risk was negatively associated with HII, indicating the disease was more common in areas with low levels of human activities, which tended to be remote, poorly educated, and economically underdeveloped.[22] Thus, the prevention *C. sinensis* infection in remote areas should not be neglected. Besides, the analysis showed that the model fit of the categorical form of HII (with log score

1.226) was slightly better than the linear form (with log score 1.228). The categorical and linear form of HII were put into the model separately for sensitivity analysis. The results of these two forms were similar, suggesting that HII in either form affected little on the final results. Following the variable selection procedure, HII of categorical form was selected in the final model. We also found that environmental factors, such as NDVI, and precipitation, were associated with infection risk, possibly due to the fact that environmental factors may influence the survival and reproduction of intermediate hosts, and thus affect the risk in the corresponding areas.[21]

Frankly, there are several limitations in our study. Specific information on implementation of the two programs was unavailable (e.g., coverages and launch time in different areas) in the model, which made it difficult to determine the correlation between the interventions of the programs and *C. sinensis* infection. Furthermore, the spatial-temporal changes of the disease risk may be influenced by other unknown factors, such as the behavior change in consumption of raw fish, and the implementation of local control programs. Therefore, the causality of changes of disease risk with the program interventions could not be confirmed. However, the overall trends of population-adjusted prevalence showed obvious downward trends since the two programs began, and an upward trend during the gap period between the two programs (Fig. 4). Besides, high probabilities were found in many regions that the estimated prevalence was lower after the programs implemented than that before the programs (Fig. 3). The above implies the effectiveness of the two national control programs. Secondly, the culturally rooted habit of raw fish consumption is one of the major reasons for *C. sinensis* infections.[4] However, surveys on eating habit with raw fish were conducted in only a few areas in South Korea,[10,41] which could not provide enough information on the distribution of this habit across the country. Due to data unavailability, we did not consider this factor in our model. Nevertheless, our model performance was good, with prediction accuracy 76.73% within 95% BCI and AUC under ROC 0.86, suggesting that our results were reliable. Similarly, the studies on the distribution of intermediate hosts, the key to transmission of the disease, were conducted only in some river basins,[42,43] which was difficult to provide useful information for risk estimation across the country. However, taking into account that the environmental factors (e.g., precipitation and NDVI) may potentially influence the human infection risk indirectly by affecting the growth environments of intermediate hosts, we considered those as potential influencing factors in the model referring to other literatures.[21,22] We also considered either linear or categorical form of these covariates, in order to capture the non-linear relationships.

Thirdly, as several surveys only reported the intervals of prevalence instead of the exact observed values, there may be a potential bias resulted from assigning the midpoint values of the intervals as observations.[8] Nevertheless, the sensitivity analysis suggested this effect was ignorable, as our final estimated posterior distribution of parameters and patterns of risk maps were similar with those assigning the lower or upper bounds of the intervals as observations in these surveys (Appendix H). Additionally, high-resolution data on potential influencing factors covering the full study period (1970–2020) were unavailable. Therefore, the average value of each factor over available years was taken, as *C. sinensis* infection is chronic, affected by environmental and economic factors over long time.[4,21] To assess how this way of data process may affect the outcomes, a sensitivity analysis was conducted. For factors with available data for many years (e.g., land surface temperature in the daytime and at night), we took the data of the corresponding year for the observed prevalence. When the specific years of data were unavailable, data of the nearest available year was substituted.[44] The results of sensitivity analysis showed similar as our final results (Appendix H), suggesting the outcomes were reliable. We also did a sensitivity analysis on the quality of literatures. When we excluded literatures assessed to be of very low quality (with scores 1–2), results were similar as our final results. If we further excluded the moderate low ones (with scores 3–4), there was a slightly higher estimation of the infection risk in the central region of South Korea in recent years (Appendix H). This may be driven by the relatively high observed prevalence before 2015, since only three literatures reported surveys in the central region after 2015, indicating a low observed prevalence in the region recently. Although the three documents had moderate low scores, we considered them providing reliable data, as they were local surveys reported in the press or cited by other literatures. The low scores of them were mainly due to the lack of detailed information on how surveys were conducted. After careful consideration, we included these data in the final analysis, to provide important information for recent years. In this study, we included surveys with different diagnostic techniques (with Kato-Katz and FE the major techniques adopted). However, several studies suggested that different diagnostic methods for *C. sinensis* infection showed different sensitivities and specificities.[30,31] Therefore, we included the diagnostic methods as a covariate to adjust the effect on the results referring to the previous studies.[22,45] Furthermore, we conducted a sensitivity analysis by firstly adjusting the observed prevalence under Kato-Katz to the prevalence under FE, using the relative ratio of the prevalence diagnosed by the two techniques reported by Cho et al.[31] Then we did the model fitting based on the adjusted prevalence. The results based on adjusting the observed prevalence were

similar as our final results, suggesting that our results were stable by both ways of dealing with different diagnostic methods. Additionally, fecal examination (e.g., Kato-Katz and FE) was difficult to discriminate the eggs of *C. sinensis* and minute intestinal flukes of the family Heterophyida (e.g., Heterophyidae *Metagonimus yokogawai*, *Metagonimus miyatai*, and *Metagonimus takahashii*) without using molecular diagnostic methods,[46] which may cause incorrect prevalence on *C. sinensis* infection in the endemic areas of *M. yokogawai* infection, such as Hadong-gun and Gokseong-gun.[47] More accurate diagnostic methods were encouraged in those areas.

In conclusion, we presented the first temporal risk maps of *C. sinensis* infection at high-resolution over 50 years across South Korea. The risk maps and overall trend of the population-adjusted estimated prevalence implied the first national deworming program and the CEP were effective. Even though low average level of infection risk in the country, several areas in river basins of the central eastern and the southern regions were still in moderate infection risk, which should be prioritized for control and intervention. All findings provide important information on spatial-targeting control and preventive strategies on *C. sinensis* infection in South Korea.

References

1 Na BK, Pak JH, Hong SJ. Clonorchis sinensis and clonorchiasis. *Acta Trop*. 2020;203:105309.
2 Qian MB, Utzinger J, Keiser J, Zhou XN. Clonorchiasis. *Lancet*. 2016;387(10020):800–810.
3 Qian MB, Jiang ZH, Zhou CH, Ge T, Wang X, Zhou XN. Familial assimilation in transmission of raw-freshwater fish-eating practice leading to clonorchiasis. *PLoS Neglected Trop Dis*. 2020;14(4):e0008263.
4 Qian MB, Zhou XN. Clonorchis sinensis. *Trends Parasitol*. 2021;37(11):1014–1015.
5 World Health Organization. Foodborne trematode infections [Accessed] https://www.who.int/health-topics/foodborne-trematode-infections#tab=tab_3; 2022.
6 Huang XH, Qian MB, Zhu GH, Fang YY, Hao YT, Lai YS. Assessment of control strategies against Clonorchis sinensis infection based on a multi-group dynamic transmission model. *PLoS Neglected Trop Dis*. 2020;14(3):e0008152.
7 The Ministry of Health and Social Affairs. *The Korean Association of parasite Eradication. Status of the first Korean intestinal parasite infection*. Seoul, Korea: The Ministry of Health and Social Affairs and The Korean Association of parasite Eradication; 1971.
8 Korea Centers for Disease Control and Prevention, Korea National Institute of Health. *National survey of the prevalence of intestinal parasitic infections in Korea, the 8th report*. Osong, Korea: Korea Centers for disease control and prevention, Korea National Institute of Health; 2013.
9 Kim JY, Yi MH, Yong TS. Parasitic infections and medical expenses according to Health Insurance Review Assessment claims data in South Korea, 2011-2018. *PLoS One*. 2019;14(11):e0225508.
10 Kim C, June KJ, Cho SH, Park KS, Lee HS, Park JY. Prevalence and related factors of clonorchiasis among five major riverside residents in South Korea. *J Korean Acad Community Health Nurs*. 2016;27(4):346–357.
11 Park JS. *Regional analysis of effectiveness of eradication project of clonorchiasis and its statistical test of significance*. Seoul, South Korea: Centers for Disease Control and Prevention; 2010.
12 June KJ, Cho SH, Lee WJ, Kim C, Park KS. Prevalence and risk factors of clonorchiasis among the populations served by primary healthcare posts along five major rivers in South Korea. *Osong Public Health Res Perspect*. 2013;4(1):21–26.
13 Jeong YI, Shin HE, Lee SE, et al. Prevalence of Clonorchis sinensis infection among residents along 5 major rivers in the Republic of Korea. *Korean J Parasitol*. 2016;54(2):215–219.
14 Hong ST, Yong TS. Review of successful control of parasitic infections in Korea. *Infect Chemother*. 2020;52(3):427–440.
15 Kim TJ, Chai JY, Jang HG. *Sustained national deworming campaign in South Korea 1969-1995*. Seoul, South Korea: KDI School of Public Policy and Management; 2014.
16 Park DS. Parasitic disease control service of PHCP. *J Korean Acad Community Health Nurs*. 2008;3(2):88–95.
17 Jo SH. Development of monitoring system and prevention of high risk population for clonorchiasis elimination of Korean main riverside area. *Public Health Weekly Report*. 2013;6(7):125–131.
18 Jeong JY, Lee JY, Chung BS, Choi Y, Alley AB, Kim HJ. A new method for estimating the prevalence of clonorchiasis in Korea A proposal to replace arbitrary riverside sampling. *Medicine (Baltimore)*. 2017;96(13):e6536.
19 Oh JK, Lim MK, Yun EH, et al. Control of clonorchiasis in Korea: effectiveness of health education for community leaders and individuals in an endemic area. *Trop Med Int Health*. 2014;19(9):1096–1104.
20 Karagiannis-Voules DA, Biedermann P, Ekpo UF, et al. Spatial and temporal distribution of soil-transmitted helminth infection in sub-Saharan Africa: a systematic review and geostatistical meta-analysis. *Lancet Infect Dis*. 2015;15(1):74–84.
21 Lai YS, Zhou XN, Pan ZH, Utzinger J, Vounatsou P. Risk mapping of clonorchiasis in the People's Republic of China: a systematic review and Bayesian geostatistical analysis. *PLoS Neglected Trop Dis*. 2017;11(3):e0005239.
22 Zhao TT, Feng YJ, Doanh PN, et al. Model-based spatial-temporal mapping of opisthorchiasis in endemic countries of Southeast Asia. *Elife*. 2021;10:e59755.
23 Gelfand AE, Banerjee S. Bayesian modeling and analysis of geostatistical data. *Annu Rev Stat Appl*. 2017;4:245–266.
24 Moraga P, Cramb SM, MK L, Pagano M. A geostatistical model for combined analysis of point-level and area-level data using INLA and SPDE. *Spat Stat*. 2017;21:27–41.
25 Page MJ, McKenzie JE, Bossuyt PM, et al. The PRISMA 2020 statement: an updated guideline for reporting systematic reviews. *BMJ*. 2021;372:n71.
26 Kokaliaris C, Garba A, Matuska M, et al. Effect of preventive chemotherapy with praziquantel on schistosomiasis among school-

aged children in sub-Saharan Africa: a spatiotemporal modelling study. *Lancet Infect Dis*. 2022;22(1):136–149.

27 United Nations. World population Prospects 2022. https://population.un.org/wpp/Download/Standard/Population/; 2022. Accessed June 10, 2022.

28 LI P. The conditional predictive ordinate for the normal distribution. *J R Stat Soc B*. 1990;52:175–184.

29 Li GS, Yang L, Nong CY, Huang J, Wu QH. Comparison of the detection rate of helminth eggs by Kato-Katz method and Formalin-ether concentration technique method. *Guangxi Prov Med*. 1998;4(4):60.

30 Uga S, Tanaka K, Iwamoto N. Evaluation and modification of the formalin-ether sedimentation technique. *Trop Biomed*. 2010;27(2):177–184.

31 Cho SY, Lee SH, Rim HJ, Seo BS. An evaluation of cellophane thick smear technique for mass stool examination. *Kisaengchunghak Chapchi*. 1969;7(1):48–52.

32 Krainski E, Gómez-Rubio V, Bakka H, et al. *Advanced spatial modeling with stochastic partial differential equations using R and INLA*. Boca Raton: CRC Press, Taylor & Francis Group; 2019.

33 Diggle PJ, Menezes R, Su TL. Geostatistical inference under preferential sampling. *J R Stat Soc Ser C Appl Stat*. 2010;59(2):191–232.

34 Brooker S, Hay SI, Issae W, et al. Predicting the distribution of urinary schistosomiasis in Tanzania using satellite sensor data. *Trop Med Int Health*. 2001;6(12):998–1007.

35 *Sustaining the drive to overcome the global impact of neglected tropical diseases: second WHO report on neglected tropcal diseases*. Geneva: World Health Organization; 2013.

36 Jung JH. *Transnational Network of Parasite Control Activities in Korea, 1950s~2000s*. Seoul: Seoul National University; 2020.

37 Lee GS, Cho IS, Lee YH, et al. Epidemiological study of clonorchiasis and metagonimiasis along the geum-gang (river) in okcheon-gun (county), Korea. *Korean J Parasitol*. 2002;40(1):9–16.

38 Shin HE, Baek SO, Lee YJ, Ju JW, Lee HI. The infection status of intestinal parasites and the degree of infection risk of freshwater fish in 2020. *Public Health Weekly Report*. 2021;6(10):1696–1706.

39 Santafé G, Adin A, Lee D, Ugarte MAD. Dealing with risk discontinuities to estimate cancer mortality risks when the number of small areas is large. *Stat Methods Med Res*. 2021;30(1):6–21.

40 Cho KM, Chang JK, Chang SJ, Rhee YS. Prevalence of intestinal parasites in ullung-do Island. *Yonsei Rep Trop Med*. 1973;4(1):50–58.

41 Shin HE, Lee MR, Ju JW, et al. Epidemiological and clinical parameters features of patients with clonorchiasis in the geum river basin, Republic of Korea. *Interdiscip Perspect Infect Dis*. 2017;2017:7415301.

42 Sohn WM, Na BK. Infections with digenetic trematode metacercariae in freshwater fishes from two visiting sites of migratory birds in gyeongsangnam-do, Republic of Korea. *Korean J Parasitol*. 2019;57(3):273–281.

43 Sohn WM, Na BK, Cho SH, Ju JW. Infection status with Clonorchis sinensis metacercariae in fish from yangcheon (stream) in sancheong-gun, gyeongsangnam-do, Korea. *Korean J Parasitol*. 2019;57(2):145–152.

44 Schmidt CA, Cromwell EA, Hill E, et al. The prevalence of onchocerciasis in Africa and Yemen, 2000-2018: a geospatial analysis. *BMC Med*. 2022;20(1):293.

45 Goovaerts P. Visualizing and testing the impact of place on late-stage breast cancer incidence: a non-parametric geostatistical approach. *Health Place*. 2010;16(2):321–330.

46 Xu M, Yin JH, Cao SK, Zhang XF, Shen YJ. Comparison of efficiency of Kato-Katz technique and PCR assay for detecting Clonorchis sinensis infection. *Zhongguo Xue Xi Chong Bing Fang Zhi Za Zhi*. 2019;31(2):165–168.

47 Bahk YY, Park YK, Na BK, et al. Survey on intestinal helminthic infection status of students in two counties, hadong-gun and goseong-gun, Korea. *Korean J Parasitol*. 2018;56(4):335–339.