Use of Bayesian geostatistical prediction to estimate local variations in *Schistosoma haematobium* infection in western Africa

Archie CA Clements,[a] Sonja Firth,[a] Robert Dembelé,[b] Amadou Garba,[c] Seydou Touré,[d] Moussa Sacko,[e] Aly Landouré,[e] Elisa Bosqué-Oliva,[f] Adrian G Barnett,[g] Simon Brooker[h] & Alan Fenwick[f]

Objective To predict the subnational spatial variation in the number of people infected with *Schistosoma haematobium* in Burkina Faso, Mali and the Niger prior to national control programmes.

Methods We used field survey data sets covering a contiguous area 2750 × 850 km and including 26 790 school-age children (5–14 years old) in 418 schools. The prevalence of high- and low-intensity infection and associated 95% credible intervals (CrIs) were predicted using Bayesian geostatistical models. The number infected was determined from the predicted prevalence and the number of school-age children in each km².

Findings The predicted number of school-age children with a low-intensity infection was 433 268 in Burkina Faso, 872 328 in Mali and 580 286 in the Niger. The number with a high-intensity infection was 416 009, 511 845 and 254 150 in each country, respectively. The 95% CrIs were wide: e.g. the mean number of boys aged 10–14 years infected in Mali was 140 200 (95% CrI: 6200–512 100).

Conclusion National aggregate estimates of infection mask important local variations: e.g. most *S. haematobium* infections in the Niger occur in the Niger River valley. High-intensity infection was strongly clustered in western and central Mali, north-eastern and north-western Burkina Faso and the Niger River valley in the Niger. Populations in these foci will carry the bulk of the urinary schistosomiasis burden and should be prioritized for schistosomiasis control. Uncertainties in the predicted prevalence and the numbers infected should be acknowledged by control programme planners.

Une traduction en français de ce résumé figure à la fin de l'article. *Al final del artículo se facilita una traducción al español.* الترجمة العربية لهذه الخلاصة في نهاية النص الكامل لهذه المقالة.

Introduction

An accurate estimate of the proportion of a population affected by a disease is important for prioritizing the control of that disease relative to another and for allocating resources to control or prevention programmes. Since high-quality surveillance data for developing countries is frequently lacking, calculations of the burden of a disease are often based on prevalence estimates from cross-sectional surveys, which are rarely randomized or representative of the whole population. For schistosomiasis, caused by trematodes of the genus *Schistosoma*, a 2003 review reported that 207 million people were infected globally and 779 million people were at risk, the majority in sub-Saharan Africa.[1] For most countries in the region, the numbers were derived from prevalence estimates contained in a 1989 report[2] that provided national aggregate data even though the prevalence of schistosomiasis is known to be geographically heterogeneous. In addition, previous reports have ignored uncertainties in prevalence estimates and in the size and spatial distribution of the population at risk.

Recently, empirical maps of tropical infectious diseases have been used to improve estimates of the populations infected and at risk at the continental or global level[3–5] and, increasingly, to plan and target control programmes. Advances in the production of these maps include geostatistical prediction of the prevalence of infection with *Schistosoma haematobium* (the aetiological agent of urinary schistosomiasis),[6] of other parasitic infections[7–10] and of coinfections[11] using Bayesian methods.[12] The Bayesian approach is advantageous because the effect of covariates and spatial heterogeneity, or clustering, can be modelled simultaneously and the uncertainty of predictions can be assessed.

While the prevalence of schistosome infection has been used for planning large-scale control programmes and for estimating the disease burden, the intensity of the infection is more informative for estimating morbidity, such as urinary tract lesions and anaemia,[13–15] and plays a greater role in driving transmission than prevalence. Therefore, maps showing the average infection intensity, or the distributions of low- and high-intensity infection, might provide more

[a] University of Queensland, School of Population Health, Herston Road, Herston, Qld, 4006, Australia.
[b] Programme National de Lutte Contre la Schistosomiase, Ministère de la Santé, Bamako, Mali.
[c] Programme National de Lutte Contre la Bilharziose et les Géohelminthes, Ministère de la Santé Publique et de la Lutte Contre les Endémies, Niamey, Niger.
[d] Programme National de Lutte Contre la Schistosomiase, Ministère de la Santé, Ouagadougou, Burkina Faso.
[e] Institut National de Recherche en Santé Publique, Bamako, Mali.
[f] Schistosomiasis Control Initiative, Imperial College, London, England.
[g] Institute for Health and Biomedical Innovation, Queensland University of Technology, Kelvin Grove, Qld, Australia.
[h] London School of Hygiene and Tropical Medicine, London, England.
Correspondence to Archie Clements (email: a.clements@uq.edu.au).
(*Submitted: 11 September 2008 – Revised version received: 28 January 2009 – Accepted: 4 February 2009 – Published online: 27 July 2009*)

effective tools than prevalence maps for estimating the disease burden and developing intervention strategies. An important statistical issue in analysing the intensity of parasitic infections is overdispersion, or aggregation, which occurs when most individuals have few parasites but a few individuals have many.[16] Bayesian models have previously been used in the spatial analysis[17–19] and prediction[20] of the intensity of parasitic infections, due for example to *Wuchereria bancrofti* and *Schistosoma mansoni*, with overdispersion in individuals' parasite or egg counts being modelled using the negative binomial distribution.

Burkina Faso, Mali and the Niger, three contiguous countries in the Sahelian zone of western Africa, recently conducted, coordinated, national cross-sectional, school-based, parasitological surveys.[21] These surveys were unprecedented in their size and covered approximately 2750 km × 850 km, 26 790 school-age children and 418 schools. We aimed to use survey data to predict subnational spatial distributions of the prevalence of low- and high-intensity *S. haematobium* infection and to use the prediction maps to calculate the numbers of school-age children infected or at risk. We also aimed to estimate uncertainties in the predicted prevalence and the numbers infected.

Methods

Selection of schools and children

The most prevalent parasitic infection in all three study countries is *S. haematobium*. Programmes supported by the Schistosomiasis Control Initiative were primarily designed to control urinary schistosomiasis[21] and our analysis includes only survey data on *S. haematobium*. Ethical approval for data collection was obtained from St Mary's Hospital Research Ethics Committee in the United Kingdom, the National Public Health Research Institute's (INRSP) scientific committee in Mali, the Ministry of Health's ethics and scientific committees in Burkina Faso and the Ministry of Health's ethical committee in the Niger.

Sample sizes were calculated using historical data from Mali[22]. It was decided to survey 87 schools in Burkina Faso, 226 in Mali and 215 in the Niger, and to include 60 children at each school. Ultimately, only 418 of the 528 schools were surveyed because remote, sparsely-populated areas had to be excluded for logistical reasons. Geographical coverage was maximized using different spatial stratification methods in the three countries. In Mali and the Niger, sample frames that contained the location of all communities were used. Spatial stratification was performed by overlaying a 1-decimal degree square grid on these countries in the ArcView version 9 (ESRI, Redlands, CA, United States of America) geographical information system. Communities were selected from the cells using simple random selection and, if they had more than one school, the study school was selected using simple random selection when the survey team arrived. In Burkina Faso, lists of schools were available for each province but were not georeferenced. The number of schools selected in each province was weighted according to the area of the province. School sampling was then done in each province using simple random selection.

The surveys were conducted from 2004–2006. The survey team determined the school's coordinates using a global positioning system. If there were fewer than 50 boys or girls in a school, all individuals of that sex were selected because compliance was more difficult if a minority were excluded. If there were more than 50 boys or girls, 30 individuals of that sex were selected using systematic random sampling. Urine and stool samples were collected from each child and processed using standard parasitological methods. The 10-ml urine samples were passed through a filter and examined by microscope in the field. The *S. haematobium* egg count was recorded and entered into a Microsoft Access database (Microsoft, Redmond, WA, USA).

Predicting the prevalence of infection

Spatial prediction was based on Bayesian geostatistics.[23] Rather than modelling egg counts using the negative binomial distribution, we used a multinomial model in which the egg count was categorized as representing: (i) no infection, (ii) low-intensity infection (i.e. 1–50 eggs per 10 ml urine), or (iii) high-intensity infection (i.e. > 50 eggs per 10 ml urine). There were two reasons: (i) expediency, given that in Burkina Faso extremely high egg counts were recorded as > 1000 eggs per 10 ml urine, meaning that the upper tail of the distribution was truncated, and (ii) to facilitate future estimation of the burden of schistosomiasis because existing evidence for related morbidity is based on stratified egg counts, often using WHO definitions of low and high intensity.[24]

It has been shown that age and sex are associated with the prevalence of urinary schistosomiasis, probably due to physiological differences in susceptibility.[25,26] Distance from a perennial inland water body is a plausible risk factor for exposure to schistosomes and subsequent infection because transmission requires contact with the aquatic habitat of intermediate host snails of the *Bulinus* genus. Distances were derived from an electronic perennial inland water body map obtained from the Food and Agriculture Organization of the United Nations (UN). The effect of temperature and rainfall on the distribution of *Bulinus* snails is reviewed in Rollinson et al.[27] Satellite-derived mean values for the land surface temperature and normalized difference vegetation index (a proxy for rainfall) for 1982–1998 were obtained from the National Oceanographic and Atmospheric Administration's Advanced Very High Radiometer. The initial candidate set of variables included the individual participant variables of sex and age (categorized as 5–9 years and 10–14 years) and the school-level ecological variables of distance from a perennial inland water body, land surface temperature and normalized difference vegetation index. We tested nominal and ordinal multinomial regression models and found that a nominal model provided a better fit. Variables were selected using fixed-effects multinomial regression models in the Stata/SE 10.0 statistical package (StataCorp, College Station, TX, USA). The normalized difference vegetation index was excluded because Wald's P was > 0.2. All remaining variables were selected for inclusion in the spatial model. This model and details of how predictions were made are presented in Box 1. The outputs of Bayesian models, including parameter estimates and spatial predictions, are termed posterior distributions. These distributions fully represent uncertainties associated with

estimated values. We summarized the posterior distributions in terms of the posterior mean and 95% credible interval (CrI), within which the true value occurs with a probability of 95%.

Calculating the number infected

An electronic population surface for the study area was obtained from the Global Rural–Urban Mapping Project (GRUMP) alpha version[28] and imported into ArcView. In practice, GRUMP is a 30-arc second (1-km²) population raster data set that combines year 2000 census data at a subnational level with an urban-extent mask. In GRUMP, the population is redistributed using an algorithm that assumes a greater proportion is located in urban areas.[29]

Country-specific population growth rates and the proportions of the population in given sex and age groups (i.e. 5–9 years and 10–14 years) were obtained from the UN Population Division publication *World population prospects*[30] and were used to generate population surfaces for 2005. Surfaces representing the mean and lower and upper 95% CrI limits of the predicted prevalence in each age–sex group were multiplied by the projected 2005 population surfaces using the Spatial Analyst Extension of ArcView, thereby giving the predicted number infected per km² pixel. These numbers were summed for each country.

Calculating the number at risk

By design, the multinomial model gave a predicted prevalence that was non-zero at all locations, even at those where field data indicated the absence of infection. A receiver operating characteristic analysis was conducted to determine the optimal threshold (i.e. where sensitivity = specificity) of the combined predicted prevalence (i.e. low- plus high-intensity infection prevalence) that best discriminated between schools with a zero and non-zero observed prevalence. With this approach, a predicted prevalence threshold of 5.3% gave the best discriminatory performance. A mask, created in the geographical information system to exclude areas with a combined predicted prevalence ≤ 5.3%, was overlaid on the different population surfaces for each country to calculate the numbers and proportions at risk of infection in both the school-age and total population.

Box 1. **The Bayesian multinomial regression model with geostatistical random effects**

The spatial model was fitted in WinBUGS version 1.4 statistical software (Medical Research Council Biostatistics Unit, Cambridge, and Imperial College, London, United Kingdom). Individual raw survey data were aggregated into groups according to age, sex and location. Using three infection outcome groups (i.e. 1 = no infection, 2 = low-intensity infection, and 3 = high-intensity infection), we assumed

$$Y_{ijk} \sim \text{Multinomial}(p_{ijk}, n_{ijk}), \text{ and}$$

$$p_{ijk} = \frac{\phi_{ijk}}{\sum_k \phi_{ijk}},$$

where Y_{ijk} is the observed number of children at location i in age–sex group j and outcome group k, n_{ijk} is the number tested and p_{ijk} is the probability of infection. Here, ϕ_{ijk} can be thought of as the overall odds of being in a specific outcome group relative to not being infected. To give a reference value, ϕ_{ij1} (i.e. for the no-infection group) was constrained to equal 1. For the other two outcome groups, we fitted the nominal regression models

$$log(\phi_{ij2}) = \alpha_2 + \sum_{z=1}^{T} \beta_{z2} \times x_{zij} + \theta_{i2}, \text{ and}$$

$$log(\phi_{ij3}) = \alpha_3 + \sum_{z=1}^{T} \beta_{z3} \times x_{zij} + \theta_{i3},$$

where α_k is the outcome group-specific intercept, β is a matrix of T coefficients and x is a matrix of T covariates. The θ_{ik} are coefficients representing location-level geostatistical random effects for the prevalence of low- and high-intensity infection. They have a multivariate normal distribution, $\theta_{ik} \sim \text{MVN}(0, \Sigma_{ik})$, and the variance–covariance matrices Σ_{ik} are defined by isotropic powered exponential spatial correlation functions

$$\sum_{ik} = f(d_{ab}; \phi) = \exp\left[-(\phi d_{ab})\right]$$

where d_{ab} are the distances between pairs of points a and b, and ϕ is the rate of decay of spatial correlation per unit distance. Noninformative priors were used for the intercepts (i.e. uniform priors with bounds $-\infty$ and ∞) and coefficients (i.e. normal priors with a mean = 0 and a precision = 1×10^{-4}). The prior distribution of ϕ was also uniform, with upper and lower bounds set at 0.06 and 50. The values of the precision of the θ_{ik} were given noninformative gamma distributions.

A burn-in of 4000 Markov chain Monte Carlo iterations was used, followed by 1000 iterations during which values for the intercept and coefficients were stored. Diagnostic tests for convergence of the stored variables were carried out, including visual examination of history and density plots. Convergence was successfully achieved after 5000 iterations and the model was run for a further 10 000 iterations, during which the predicted prevalence at individual locations was stored for each age and sex group.

Predictions of the prevalence of low- and high-intensity infection were made at the nodes of a 0.15 × 0.15 decimal degree grid (approximately 18 km²) in WinBUGS using a spatial model and the *spatial.unipred* command, which used kriging to interpolate spatial random effects for low- and high-intensity infection. This assumes independence between prediction locations, as opposed to conditional predictions, and might lead to overestimation of uncertainties in the national-level estimates of numbers infected. However, conditional predictions are extremely intensive computationally and were considered as not being feasible in this study. Predicted prevalence was calculated by adding the interpolated random effect to the sum of the products of the coefficients for the fixed effects and the values of the fixed effects at each prediction location. For the individual-level fixed effects of sex and age, separate calculations were performed, in which the coefficient for the relevant age and sex group was added to the sum. The overall sum was then back-transformed from the logit scale to the prevalence scale, giving prediction surfaces that show the prevalence of low- and high-intensity infection in each age and sex group for all prediction locations.

Results

Predicted prevalence of infection

The total number of children aged 5–14 years included in the 2004–2006 surveys was 4808 in Burkina Faso, 14 586 in Mali and 7396 in the Niger. The raw prevalence of low-intensity infection was 10.0% (95% confidence interval, CI: 9.2–10.9) in Burkina Faso, 25.2% (95% CI: 24.5–25.9) in Mali and 10.1% (95% CI: 9.4–10.8) in the Niger. For high-intensity infection, the raw prevalence was 8.5% (95% CI: 7.7–9.3) in Burkina Faso, 11.4% (95% CI: 10.9–11.9) in Mali and 3.4% (95% CI: 3.0–3.8) in the Niger. A map of the raw prevalence of *S. haematobium* infection is presented in Fig. 1.

The spatial model is presented in Table 1. It can be seen from the 95% CrI of the quadratic term for the land surface temperature that the association between this variable and the prevalence of low- or high-intensity infection was not significant. However, the distance from a perennial inland water body was significantly and negatively associated with the prevalence of both low- and high-intensity infection. In addition, the prevalence of low- and high-intensity infection was significantly higher in boys and in children aged 10–14 years than in girls or children aged 5–9 years. The rate of decay of spatial correlation was higher for low-intensity infection than high-intensity infection and the variance of the spatial random effect (i.e. the sill in geostatistical terms) was higher for high-intensity infection than low-intensity infection, which indicates a stronger propensity for spatial clustering for high-intensity infection.

Separate prediction maps were produced for boys and girls and for ages 5–9 years and 10–14 years. The mean predicted prevalences of low- and high-intensity infection in boys aged 10–14 years (the highest prevalence group) are presented in Fig. 2 and Fig. 3, respectively, as illustrative examples. The maps for other age–sex groups (available from the corresponding author) showed the same spatial distribution but a lower predicted prevalence, reflecting the lower odds ratios (ORs) for infection in these groups. There were differences between the spatial distributions for low- and high-intensity infections: low-intensity infection was more widespread

Fig. 1. **Raw prevalence of *Schistosoma haematobium* infection in school-age children in Burkina Faso, Mali and the Niger, 2004–2006.**

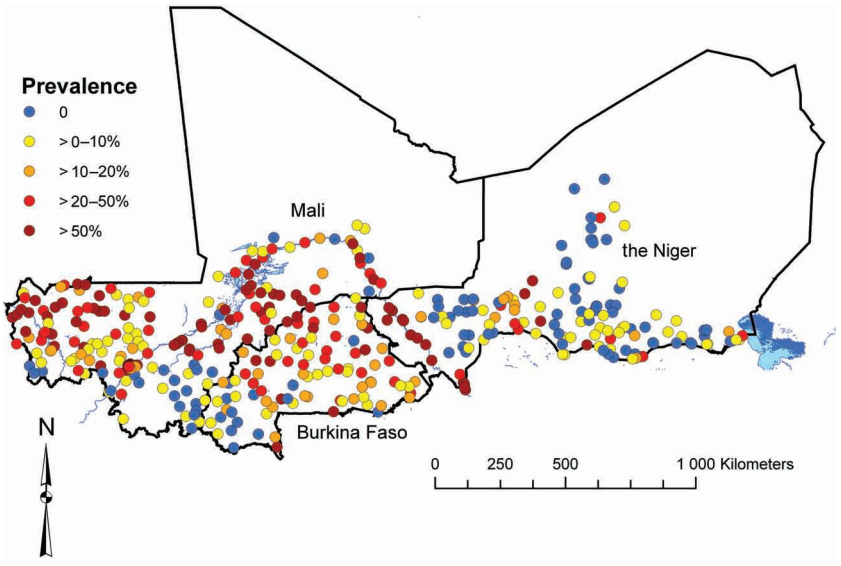


Image produced using ArcView version 9 (ESRI, Redlands, CA, United States of America) geographical information system.

Table 1. **Covariates and other parameters for the Bayesian multinomial regression model used to predict[a] the prevalence of low- (i.e. 1–50 eggs/10 ml urine) and high-intensity (i.e. > 50 eggs/10 ml urine) *Schistosoma haematobium* infection in school-age children in Burkina Faso, Mali and the Niger, 2004–2006**

Variable	Low-intensity infection Mean (95% CrI)	High-intensity infection Mean (95% CrI)
OR[b] for covariates		
Land surface temperature[c]	0.87 (0.65–1.24)	0.34 (0.12–0.57)
Land surface temperature squared[c]	1.11 (0.94–1.34)	1.00 (0.78–1.30)
Distance from a PIWB[c]	0.36 (0.24–0.50)	0.24 (0.11–0.52)
Age, 10–14 years vs 5–9 years	1.44 (1.31–1.58)	1.45 (1.25–1.65)
Sex, female vs male	0.80 (0.74–0.86)	0.49 (0.44–0.54)
Other model parameters[b]		
Intercept,[d] log odds scale	−3.04 (−3.62 to −2.51)	−4.60 (−5.13 to −3.39)
Rate of decay of the spatial correlation, ϕ[e]	2.03 (1.37–2.82)	1.49 (0.75–2.29)
Variance of the spatial random effect, σ^2	5.80 (4.53–7.94)	12.75 (8.92–22.27)

CrI, Bayesian credible interval; OR, odds ratio; PIWB, perennial inland water body.

[a] Spatial predictions were based on a combination of the environmental covariates of land surface temperature and distance from a PIWB and the geostatistical random effect, which was described by the parameters ϕ and σ^2 (the latter represents the sill of a traditional variogram). The individual-level covariates age and sex were used to adjust the predictions for the different age–sex profiles of the survey schools.

[b] The values of the ORs and parameters were derived from survey data by multiple regression.

[c] The scales of the ORs were standardized for land surface temperature and distance from a PIWB by orthogonalizing the data to have a mean of 0 and standard deviation of 1.

[d] The intercept represents the prevalence of infection for a male individual aged 5–9 years in a location with an average temperature and an average distance from a PIWB. Its value can be derived from the logit back-transformation and is approximately 4.5% for low-intensity infection and 1% for high-intensity infection.

[e] A ϕ of 2.03 for low-intensity infection indicates a range of spatial correlation, equivalent to the radius of clusters, of 1.48 decimal degrees (i.e. $3/\phi$, approximately 170 km at the equator) after accounting for the effects of the covariates.

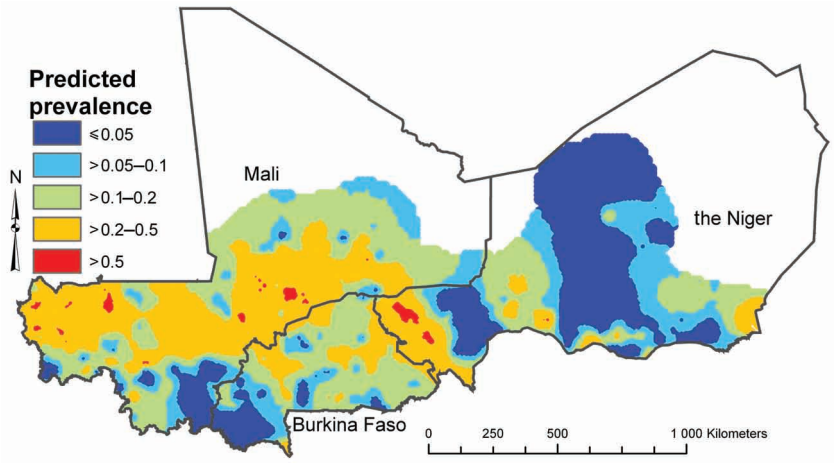
but the variation in prevalence was less extreme, with a predicted prevalence between 10% and 50% in most mapped areas of Burkina Faso (excluding the south-west), Mali (excluding the far south) and the Niger (excluding central regions). High-intensity infection had a more restricted spatial distribution, with defined clusters of high prevalence (> 50%) located in a mid-latitude band from western to central Mali, in northern and central Burkina Faso and in the Niger River valley in the Niger. There were large areas of low prevalence (< 5%) in southern, northern and eastern Mali, south-western Burkina Faso and most of the Niger, excluding the Niger River valley. The 95% CrI maps showed wide uncertainty in the predicted prevalence, though the prediction model was clearly able to exclude parts of each country as being at-risk areas for significant transmission of *S. haematobium*, as indicated by an upper 95% CrI limit for the predicted prevalence of < 5%). Fig. 4 and Fig. 5 (available at: http://www.who.int/bulletin/volumes/87/12/08-058933/en/index.html) illustrate the situation in boys 10–14 years of age; maps for other sex and age groups can be obtained from the corresponding author.

Number infected and at risk

The mean numbers of school-age children with low- and high-intensity infection in the three countries and their associated 95% CrIs are presented in Table 2. The estimated number of school-age children with low-intensity infection was 433 268 in Burkina Faso, 872 328 in Mali and 580 286 in the Niger and the number with high-intensity infection was 416 009 in Burkina Faso, 511 845 in Mali and 254 150 in the Niger. The 95% CrIs for the number infected in each age–sex group were very wide: e.g. the mean number of boys aged 10–14 years infected in Mali was 140 200 (95% CrI: 6200–512 100). Maps of the estimated number of boys aged 10–14 years with low- and high-intensity infection (Fig. 6 and Fig. 7, respectively) showed considerable within-country variation in the burden of schistosomiasis. This was also apparent for boys and girls of both age groups. As was observed with the predicted prevalence, the spatial patterns were identical for each age–sex group, though the proportion infected was uniformly lower for girls and younger boys (maps available from the corresponding author).

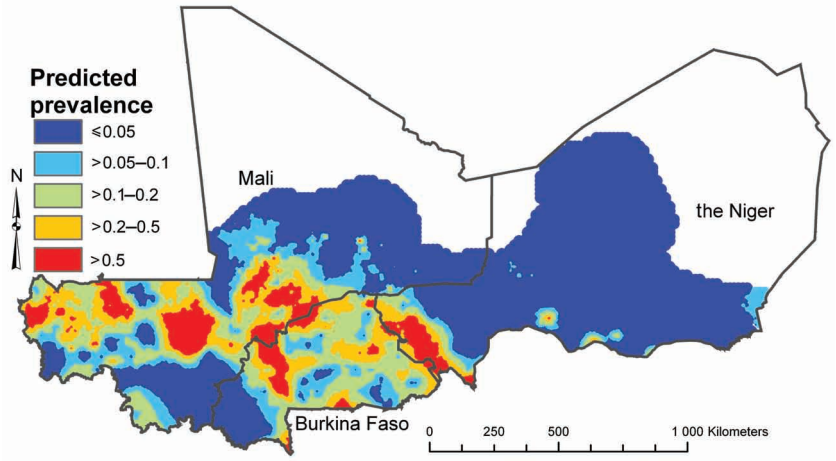
We estimated that 3.5 million, 3.4 million and 2.8 million school-age children and a total of 12.5 million (94.5% of the total population), 11.8 million (87.6% of the total) and 9.9 million (70.3% of the total) individuals of all ages were at risk of urinary schistosomiasis in Burkina Faso, Mali and the Niger, respectively. The maps (available from the corresponding author) show that there were parts of each country where people were not at risk and that could be excluded from active surveillance or nationally coordinated schistosomiasis control.

Fig. 2. **Prevalence[a] of low-intensity (1–50 eggs/10 ml urine) *Schistosoma haematobium* infection in boys aged 10–14 years in Burkina Faso, Mali and the Niger[b] in 2004–2006, as predicted using a Bayesian geostatistical multinomial regression model**



[a] The values presented are the posterior means obtained from the Bayesian geostatistical multinomial regression model and represent the most likely prevalence values.
[b] No predictions were made for the white areas on the map.
Image produced using ArcView version 9 (ESRI, Redlands, CA, United States of America) geographical information system.

Fig. 3. **Prevalence[a] of high-intensity (> 50 eggs/10 ml urine) *Schistosoma haematobium* infection in boys aged 10–14 years in Burkina Faso, Mali and the Niger[b] in 2004–2006, as predicted using a Bayesian geostatistical multinomial regression model**



[a] The values presented are the posterior means obtained from the Bayesian geostatistical multinomial regression model and represent the most likely prevalence values.
[b] No predictions were made for the white areas on the map.
Image produced using ArcView version 9 (ESRI, Redlands, CA, United States of America) geographical information system.

Discussion

We used robust, contemporary statistical methods in a novel application to predict the spatial distribution of

Table 2. **Number of school-age children with low- or high-intensity *Schistosoma haematobium* infection in Burkina Faso, Mali and the Niger in 2005, as estimated using a Bayesian geostatistical model based on survey data from 2004–2006**

Age–sex group	Burkina Faso		Mali		the Niger	
	Estimated population, in thousands	No. infected, in thousands Mean[a] (95% CrI)	Estimated population, in thousands	No. infected, in thousands Mean[a] (95% CrI)	Estimated population, in thousands	No. infected, in thousands Mean[a] (95% CrI)
High-intensity infection						
Boys 5–9 years	1011.0	123.9 (1.4–603.5)	1063	150.6 (6.6–567.7)	1119.9	79.3 (1.6–342.6)
Boys 10–14 years	867.2	124.0 (1.5–563.0)	896.5	140.2 (6.2–512.2)	905.9	70.8 (1.4–303.1)
Girls 5–9 years	976.8	83.5 (0.8–463.0)	1057.2	114.3 (4.1–462.2)	1065.9	55.0 (1.0–257.7)
Girls 10–14 years	840.9	84.6 (0.9–441.9)	895.3	106.8 (3.9–421.2)	860.3	49.1 (0.9–226.1)
Low-intensity infection						
Boys 5–9 years	1011.0	112.0 (2.3–528.2)	1063	226.9 (18.4–687.8)	1119.9	157.4 (6.0–543.9)
Boys 10–14 years	867.2	115.9 (2.5–508.3)	896.5	218.2 (19.0–619.8)	905.9	146.2 (5.6–486.3)
Girls 5–9 years	976.8	100.3 (2.3–481.3)	1057.2	216.6 (18.5–659.9)	1065.9	143.4 (7.2–485.7)
Girls 10–14 years	840.9	105.1 (2.5–468.1)	895.3	210.7 (19.6–599.5)	860.3	133.2 (6.8–434.2)

CrI, Bayesian credible interval.
[a] The mean of the posterior distribution of Bayesian estimates.

S. haematobium infection. This resulted in estimates of local heterogeneity in high- and low-intensity parasitic infection that could be used in control programme planning. In 2003, Steinmann et al.[1] estimated the number of people infected with *Schistosoma* spp. was 7.8 million in Burkina Faso, 7.8 million in Mali and 3.2 million in the Niger by using prevalence estimates of 60.0%, 60.0% and 26.7% and assuming that the at-risk population was 100% of the estimated population, namely 13.0 million, 13.0 million and 12.0 million in the three countries, respectively. If we assume, as Steinmann et al. did, that the prevalence is the same for all age groups (which overestimates the prevalence since it is usually highest in school-age children),[31] our estimates would be 3.0 million (23.0% of the total) in Burkina Faso, 4.8 million (35.4% of the total) in Mali and 3.0 million (21.1% of the total) in the Niger. Clearly, the previously reported number infected was considerably overestimated for Burkina Faso and Mali. In addition, the number at risk was overestimated for all three countries. We have confidence in our conclusions because our data were recent, extensive, randomized and representative. Although our calculations excluded *S. mansoni* infection, which Steinmann et al. included, its prevalence in our surveys was very low in Burkina Faso and the Niger, at 0.5% and 0.3%, respectively, and moderately low in Mali, at 6.7%. More importantly, Steinmann et al.'s figures overlooked important heterogeneities in the spatial distribution of people infected or at risk of schistosomiasis, which our maps capture.

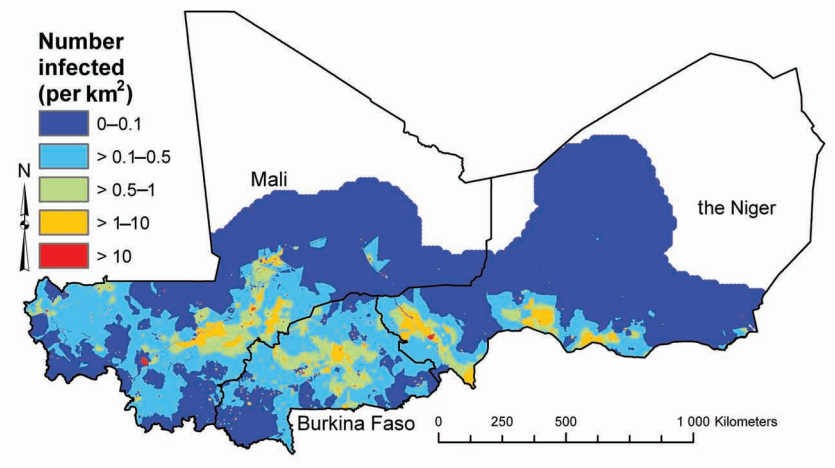
The uncertainties in the numbers infected with *S. haematobium* captured by our model were wide despite the large sample size and the geographical coverage of the data. It might be suggested that the large uncertainties shown in our prediction maps limit their utility for decision-making. However, since tools for representing uncertainty in spatial predictions now exist and there is still a need for disease maps in planning control programmes, it is beneficial to acknowledge such uncertainties when interpreting maps for disease control. Moreover, we did not include uncertainties in the projected population or those due to population migration between areas, which are considerable in these three countries. In addition, GRUMP can produce national population totals that do not match UN totals. In 2000, GRUMP population estimates for both Mali and the Niger differed from UN estimates by more than 5%,[28] contributing to additional uncertainty in our updated 2005 population estimates. We also assumed that population growth was even across different regions and age groups. The limitations of the ancillary data used in GRUMP calculations have been noted elsewhere,[29] but currently it is impossible to quantify the resulting inaccuracies.

Our surveys were spatially stratified, which ensured equal geographical coverage of areas of both high and low population density and made the precision of predicted prevalence estimates more even across the study area. However, less densely populated areas were more strongly represented than they would have been with a nonstratified approach. Because prevalence is likely to differ between low- and high-density areas, sample weighting is necessary to ensure accurate national prevalence estimates; our map calculations are essentially a sample-weighting approach. This is evident from the difference between raw prevalence estimates for low- and high-intensity infections combined, which were 18.5% for Burkina Faso, 36.6% for Mali and 13.5% for the Niger, and map-derived estimates, which were 23.0%, 35.4% and 21.1% in the three countries, respectively. The most striking difference was for the Niger. This arose because the prevalence was highest in the Niger River valley, which has the highest population density, and in the survey this area was underrepresented, from a population perspective, relative to low-density areas.

The different spatial distributions for low- and high-intensity infection are in agreement with Guyatt et al.,[16] who demonstrated that the relationship between overall prevalence and the prevalence of high-intensity infection varied geographically across Africa. While low-intensity infection was more widespread, it exhibited less spatial variability than high-intensity infection, which aggregated more into clusters.[16] Future control programmes will have the greatest impact if they focus on high-intensity infection clusters.

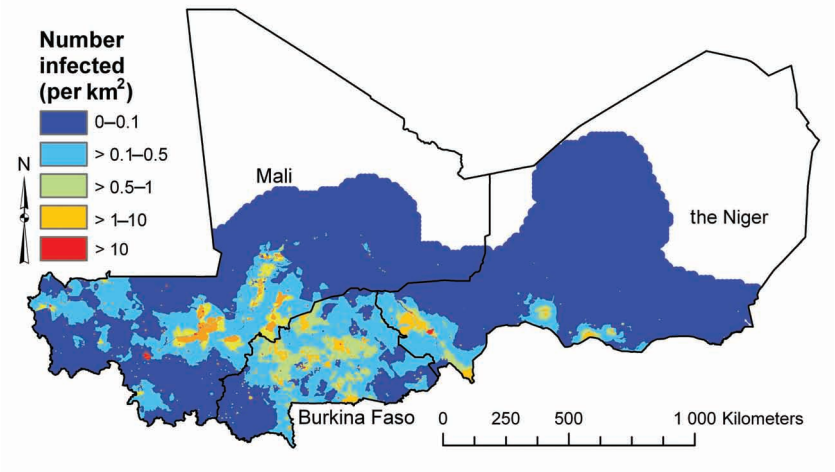
The maps presented in this report are currently being used by national programme managers as objective decision-support tools for geographically targeting existing resources more efficiently to high-risk communities. They have several other uses. First, national programme managers can use them to argue for resources from governments or international donors, particularly after funding from the Schistosomiasis Control Initiative ends. Second, the maps can be used to formulate and compare different disease control strategies by determining the likely impact on transmission and morbidity. And lastly, they can be used for advocacy and empowerment at the subnational level: local resource needs and priorities for schistosomiasis control, which are often subsumed in aggregate national data, can be presented to national programme managers and government officials. We encourage national programme managers in other countries, and those focussing on other diseases, to conduct spatially stratified disease surveys and undertake mapping of subnational disease distributions to provide evidence for more efficient targeting of resources. ■

Fig. 6. **Boys aged 10–14 years with low-intensity (1–50 eggs/10 ml urine) *Schistosoma haematobium* infection per square kilometre in Burkina Faso, Mali and the Niger[a] in 2005, as estimated using posterior means predicted by a Bayesian geostatistical multinomial regression model**



[a] No estimates were made for the white areas on the map.
Image produced using ArcView version 9 (ESRI, Redlands, CA, United States of America) geographical information system.

Fig. 7. **Boys aged 10–14 years with high-intensity (> 50 eggs/10 ml urine) *Schistosoma haematobium* infections per square kilometre in Burkina Faso, Mali and the Niger[a] in 2005, as estimated using posterior means predicted by a Bayesian geostatistical multinomial regression model**



[a] No estimates were made for the white areas on the map.
Image produced using ArcView version 9 (ESRI, Redlands, CA, United States of America) geographical information system.

Acknowledgements

We thank the children, parents, teachers and head teachers who participated in the surveys. We also thank the technicians and support staff who undertook the surveys and the members of national schistosomiasis control programmes who provided administrative and organizational assistance.

Funding: During the course of this study, the Schistosomiasis Control Initiative was supported by the Bill and Melinda Gates Foundation. Simon Brooker is supported by a Career Development Fellowship (081673) from the Wellcome Trust. Archie Clements receives salary support and is a visiting scientist at the Australian Centre for International and Tropical Health, Queensland Institute of Medical Research, Herston, Qld, Australia.

Competing interests: None declared.

Résumé

Utilisation de prédictions par géostatistique bayésienne pour estimer les variations locales de la prévalence des infestations par *Schistosoma haematobium* en Afrique occidentale

Objectif Prédire les variations spatiales au niveau infranational du nombre de personnes infestées par *Schistosoma haematobium* au Burkina Faso, au Mali et au Niger, avant la mise en place des programmes nationaux de lutte contre la schistosomiase.

Méthodes Nous avons utilisé un jeu de données d'enquête sur le terrain couvrant une zone contiguë de 2750 x 850 km et 26 790 enfants d'âge scolaire (5-14 ans), répartis dans 418 écoles. La prévalence des schistosomiases de forte et de faible intensité, ainsi que les intervalles de crédibilité à 95 % associés, ont été prédits à l'aide de modèles géostatistiques bayésiens. Le nombre de personnes infestées a été déterminé à partir de la prévalence prédite et du nombre d'enfants d'age scolaire par km².

Résultats D'après les prédictions de l'étude, le nombre d'enfants d'âge scolaire atteints d'une schistosomiase de faible intensité serait de 433 268 au Burkina Faso, de 872 328 au Mali et de 580 286 au Niger. S'agissant des enfants fortement infestés, les prédictions donnaient respectivement 416 009 cas pour le Burkina Faso, 511 845 pour le Mali et 254 150 pour le Niger. Les intervalles de crédibilité à 95 % étaient larges : par exemple, le nombre moyen de garçons de 10 à 14 ans infestés au Mali était de 140 200 (ICr à 95 % : 6200-512 100).

Conclusion Les estimations nationales agrégées masquent d'importantes variations locales : par exemple, la plupart des infestations par *S. haematobium* relevées au Niger étaient apparues dans la Vallée du Niger. Les cas d'infestation lourde étaient très fortement regroupés à l'Ouest et au centre du Mali, au Nord-est et au Nord-ouest du Burkina Faso et dans la Vallée du Niger, au Niger. Les populations de ces foyers supportent la plus grande part de la charge de schistosomiase urinaire et doivent être considérées comme prioritaires dans la lutte contre la schistosomiase. Les planificateurs de programmes de lutte contre cette maladie doivent être conscients des incertitudes qui pèsent sur les prédictions de la prévalence et du nombre de personnes infestées.

Resumen

Uso de modelos geoestadísticos bayesianos para predecir las variaciones locales de la infección por *Schistosoma haematobium* en África occidental

Objetivo Predecir la variación territorial subnacional del número de personas infectadas por *Schistosoma haematobium* en Burkina Faso, Malí y el Níger antes del inicio de los programas de control nacionales.

Métodos Usamos conjuntos de datos de encuestas sobre el terreno que abarcaron en total a 26 790 niños en edad escolar (5 a 14 años) de 418 escuelas repartidos en una zona de 2750 x 850 km. Mediante modelos geoestadísticos bayesianos se predijeron la prevalencia de infección de alta y baja intensidad y los correspondientes intervalos de credibilidad (ICr) del 95%. El número de personas infectadas se determinó a partir de la prevalencia predicha y del número de niños en edad escolar por km².

Resultados Las predicciones sobre el número de niños en edad escolar con infección de baja intensidad fueron de 433 268 en Burkina Faso, 872 328 en Malí y 580 286 en el Níger. El número de casos de infección de alta intensidad fue de 416 009, 511 845 y 254 150, respectivamente. Los ICr95% fueron amplios: p.ej., la media de muchachos de 10 a 14 años infectados en Malí fue de 140 200 (ICr95%: 6200–512 100).

Conclusión Las estimaciones totales nacionales de infección ocultan variaciones locales importantes: p.ej., en el Níger la mayoría de las infecciones por *S. haematobium* se dan en el valle del Río Níger. La infección de alta intensidad se concentra marcadamente en el centro y oeste de Malí, las zonas nororiental y noroccidental de Burkina Faso y el valle del Río Níger en el Níger. Las poblaciones de esos focos soportarán la mayor carga de esquistosomiasis urinaria y deberían ser un objetivo prioritario de la lucha contra la esquistosomiasis. Los planificadores de los programas de control deben tener en cuenta la incertidumbre asociada a la prevalencia predicha y las cifras de personas infectadas.

ملخص

استخدام التوقع الجغرافي الإحصائي لبيزان لتقدير التفاوتات المحلية في العدوى بالبلهارسيات الدموية في غرب أفريقيا

الهدف: للتنبؤ بالتفاوتات الفراغية على الصعيد دون الوطني لأعداد المصابين بعدوى البلهارسيات الدموية في بوركينا فاسو ومالي والنيجر قبل تنفيذ البرامج الوطنية للمكافحة.

الطريقة: استخدم الباحثون مجموعة معطيات خاصة بالمسوحات التي تغطي إحدى المناطق التي تعاني من السراية ومساحتها 2750 × 850 كيلومتراً ويعيش فيها 26 790 من الأطفال في سن المدرسة (بعمر 5 – 14) عاماً يدرسون في 418 مدرسة. وقد كان معدل انتشار العدوى المنخفضة الكثافة والشديدة الكثافة والمترافق مع 95% من فواصل الثقة، قد حسب باستخدام نماذج جغرافية إحصائية لبيزان. وقد حدد الباحثون عدد المصابين بالعدوى انطلاقاً من معدل الانتشار المتوقع ومن عدد الأطفال في سن المدرسة في كل كيلومتر مربع.

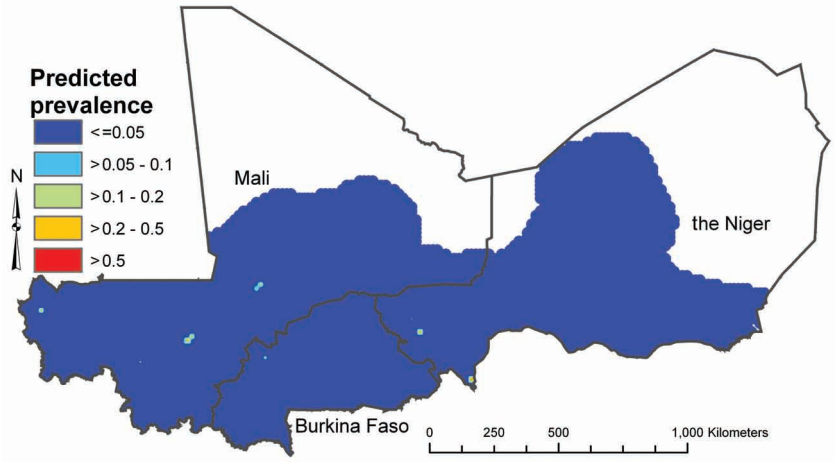
الموجودات: بلغ العدد المتوقع للأطفال في سن المدرسة، المصابين بعدوى منخفضة الكثافة، 268 433 في بوركينا فاسو و328 872 في مالي، و286 580 في النيجر. أما عدد المصابين بعدوى عالية الكثافة فقد كان في بوركينا فاسو 009 416 وفي مالي 845 511 وفي النيجر 150 254، وكان مجال الثقة واسعاً حيث بلغ عدد الأولاد الذين تتراوح أعمارهم بين 10 – 14 عاماً والمصابين بالعدوى في مالي 200 140 (بفاصلة ثقة 95%، إذ تراوحت بين 6200 و100 512).

الاستنتاج: إن التقديرات الوطنية المتجمعة للعدوى تخفي التفاوتات المحلية الهامة، فمعظم عداوى البلهارسيات الدموية في النيجر تحدث في وادي نهر النيجر. وقد كان هناك تجمع عالي الكثافة في غرب ووسط مالي، وشمال شرق وشمال غرب بوركينا فاسو ووادي نهر النيجر في النيجر. ويتحمل السكان في هذه المناطق مجمل عبء داء البلهارسيات البولي وينبغي أن يعطوا الأولوية في مكافحة داء البلهارسيات. وإن عدم التيقن في التنبؤ بمعدل الانتشار وفي تقدير عدد المصابين بالعدوى ينبغي أن يؤخذ في الاعتبار من قِبَل القائمين على تخطيط برامج المكافحة.

References

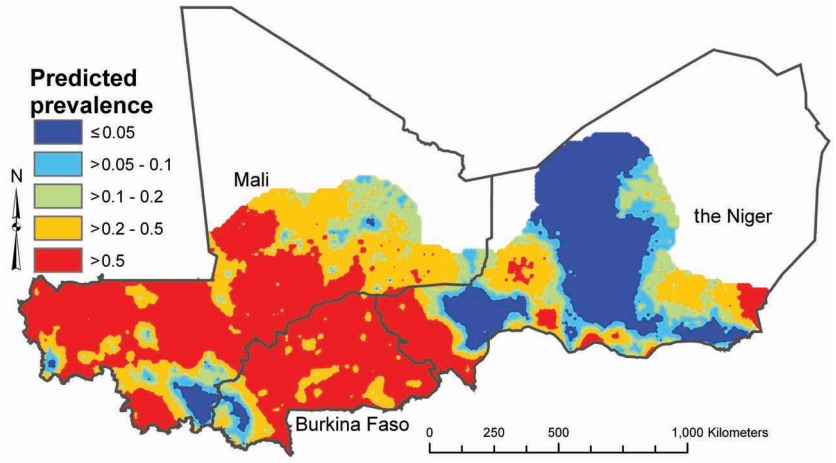
1. Steinmann P, Keiser J, Bos R, Tanner M, Utzinger J. Schistosomiasis and water resources development: systematic review, meta-analysis, and estimates of people at risk. *Lancet Infect Dis* 2006;6:411-25. PMID:16790382 doi:10.1016/S1473-3099(06)70521-7
2. Utroska JA, Chen MG, Dixon H, Yoon S, Helling-Borda M, Hogerzeil HV, et al. *An estimate of global needs for praziquantel within schistosomiasis control programmes*. Geneva: World Health Organization; 1989.
3. Brooker S, Clements AC, Bundy DA. Global epidemiology, ecology and control of soil-transmitted helminth infections. *Adv Parasitol* 2006;62:221-61. PMID:16647972 doi:10.1016/S0065-308X(05)62007-6
4. Guerra CA, Snow RW, Hay SI. Mapping the global extent of malaria in 2005. *Trends Parasitol* 2006;22:353-8. PMID:16798089 doi:10.1016/j.pt.2006.06.006
5. Hay SI, Guerra CA, Tatem AJ, Noor AM, Snow RW. The global distribution and population at risk of malaria: past, present, and future. *Lancet Infect Dis* 2004;4:327-36. PMID:15172341 doi:10.1016/S1473-3099(04)01043-6
6. Clements AC, Lwambo NJ, Blair L, Nyandindi U, Kaatano G, Kinung'hi S, et al. Bayesian spatial analysis and disease mapping: tools to enhance planning and implementation of a schistosomiasis control programme in Tanzania. *Trop Med Int Health* 2006;11:490-503. PMID:16553932 doi:10.1111/j.1365-3156.2006.01594.x
7. Gemperli A, Sogoba N, Fondjo E, Mabaso M, Bagayoko M, Briet OJ, et al. Mapping malaria transmission in West and Central Africa. *Trop Med Int Health* 2006;11:1032-46. PMID:16827704 doi:10.1111/j.1365-3156.2006.01640.x
8. Gemperli A, Vounatsou P, Kleinschmidt I, Bagayoko M, Lengeler C, Smith T. Spatial patterns of infant mortality in Mali: the effect of malaria endemicity. *Am J Epidemiol* 2004;159:64-72. PMID:14693661 doi:10.1093/aje/kwh001
9. Raso G, Matthys B, N'Goran EK, Tanner M, Vounatsou P, Utzinger J. Spatial risk prediction and mapping of Schistosoma mansoni infections among schoolchildren living in western Cote d'Ivoire. *Parasitology* 2005;131:97-108. PMID:16038401 doi:10.1017/S0031182005007432
10. Raso G, Vounatsou P, Gosoniu L, Tanner M, N'Goran EK, Utzinger J. Risk factors and spatial patterns of hookworm infection among schoolchildren in a rural area of western Cote d'Ivoire. *Int J Parasitol* 2006;36:201-10. PMID:16259987 doi:10.1016/j.ijpara.2005.09.003
11. Raso G, Vounatsou P, Singer BH, N'Goran EK, Tanner M, Utzinger J. An integrated approach for risk profiling and spatial prediction of Schistosoma mansoni-hookworm coinfection. *Proc Natl Acad Sci USA* 2006;103:6934-9. PMID:16632601 doi:10.1073/pnas.0601559103
12. Diggle P, Moyeed R, Rowlingson B, Thompson M. Childhood malaria in the Gambia: a case-study in model-based geostatistics. *Appl Stat* 2002; 51:493-506.
13. Vester U, Kardorff R, Traore M, Traore HA, Fongoro S, Juchem C, et al. Urinary tract morbidity due to Schistosoma haematobium infection in Mali. *Kidney Int* 1997;52:478-81. PMID:9264005 doi:10.1038/ki.1997.356
14. Abdel-Wahab MF, Esmat G, Ramzy I, Fouad R, Abdel-Rahman M, Yosery A, et al. Schistosoma haematobium infection in Egyptian schoolchildren: demonstration of both hepatic and urinary tract morbidity by ultrasonography. *Trans R Soc Trop Med Hyg* 1992;86:406-9. PMID:1440818 doi:10.1016/0035-9203(92)90241-4
15. Warren KS, Mahmoud AA, Muruka JF, Whittaker LR, Ouma JH, Arap Siongok TK. Schistosomiasis haematobia in coast province Kenya. Relationship between egg output and morbidity. *Am J Trop Med Hyg* 1979;28:864-70. PMID:484768
16. Guyatt HL, Smith T, Gryseels B, Lengeler C, Mshinda H, Siziya S, et al. Aggregation in schistosomiasis: comparison of the relationships between prevalence and intensity in different endemic areas. *Parasitology* 1994;109:45-55. PMID:8058368 doi:10.1017/S0031182000077751
17. Alexander N, Moyeed R, Stander J. Spatial modelling of individual-level parasite counts using the negative binomial distribution. *Biostatistics* 2000;1:453-63. PMID:12933567 doi:10.1093/biostatistics/1.4.453
18. Alexander ND, Moyeed RA, Hyun PJ, Dimber ZB, Bockarie MJ, Stander J, et al. Spatial variation of Anopheles-transmitted Wuchereria bancrofti and Plasmodium falciparum infection densities in Papua New Guinea. *Filaria J* 2003;2:14. PMID:14525619 doi:10.1186/1475-2883-2-14
19. Brooker S, Alexander N, Geiger S, Moyeed RA, Stander J, Fleming F, et al. Contrasting patterns in the small-scale heterogeneity of human helminth infections in urban and rural environments in Brazil. *Int J Parasitol* 2006;36:1143-51. PMID:16814294 doi:10.1016/j.ijpara.2006.05.009
20. Clements AC, Moyeed R, Brooker S. Bayesian geostatistical prediction of the intensity of infection with *Schistosoma mansoni* in East Africa. *Parasitology* 2006;133:711-9. PMID:16953953 doi:10.1017/S0031182006001181
21. Garba A, Toure S, Dembele R, Bosque-Oliva E, Fenwick A. Implementation of national schistosomiasis control programmes in West Africa. *Trends Parasitol* 2006;22:322-6. PMID:16690357 doi:10.1016/j.pt.2006.04.007
22. Gryseels B. *Mission d'Evaluation du report, Lutte contre la schistosomiasis au Mali*. Eschborn: Gesellschaft für Technische Zusammenarbeit; 1989.
23. Diggle P, Tawn J, Moyeed R. Model-based geostatistics. *Appl Stat* 1998; 47:299-350.
24. *Coordinated use of anthelminthic drugs in control interventions: a manual for health professionals and programme managers*. Geneva: World Health Organization; 2006.
25. Fulford AJ, Webster M, Ouma JH, Kimani G, Dunne DW. Puberty and age-related changes in susceptibility to schistosome infection. *Parasitol Today* 1998;14:23-6. PMID:17040685 doi:10.1016/S0169-4758(97)01168-X
26. Feldmeier H, Poggensee G, Krantz I. Puberty and age-intensity profiles in schistosome infections: another hypothesis. *Parasitol Today* 1998;14:435. PMID:17040837 doi:10.1016/S0169-4758(98)01313-1
27. Rollinson D, Stothard JR, Southgate VR. Interactions between intermediate snail hosts of the genus *Bulinus* and schistosomes of the *Schistosoma haematobium* group. *Parasitology* 2001;123 Suppl;S245-60. PMID:11769287 doi:10.1017/S0031182001008046
28. Global Rural–Urban Mapping Project (GRUMP) [Alpha version]. Palisades, NY: Socioeconomic Data and Applications Center, Columbia University. Available from: http://sedac.ciesin.columbia.edu/gpw [accessed on 19 June 2009].
29. Balk DL, Deichmann U, Yetman G, Pozzi F, Hay SI, Nelson A. Determining global population distribution: methods, applications and data. *Adv Parasitol* 2006;62:119-56. PMID:16647969 doi:10.1016/S0065-308X(05)62004-0
30. United Nations. *World population prospects: the 2006 revision*. New York, NY: United Nations, Department of Economic and Social Affairs, Population Division; 2007. Available from: http://www.un.org/esa/population/publications/wpp2006/wpp2006.htm [accessed on 19 June 2009].
31. Brooker S, Donnelly CA, Guyatt HL. Estimating the number of helminthic infections in the Republic of Cameroon from data on infection prevalence in schoolchildren. *Bull World Health Organ* 2000;78:1456-65. PMID:11196498

Fig. 4. **Map showing lower value for the Bayesian 95% CrI for the predicted prevalence of high-intensity (> 50 eggs/10 ml urine) *Schistosoma haematobium* infection in boys aged 10–14 years in Burkina Faso, Mali and the Niger[a] in 2004–2006, as derived using a Bayesian geostatistical multinomial regression model**



CrI, Bayesian credible interval.
[a] No predictions were made for the white areas on the map.
Image produced using ArcView version 9 (ESRI, Redlands, CA, United States of America) geographical information system.

Fig. 5. **Map showing upper value for the Bayesian 95% CrI for the predicted prevalence of high-intensity (> 50 eggs/10 ml urine) *Schistosoma haematobium* infection in boys aged 10–14 years in Burkina Faso, Mali and the Niger[a] in 2004–2006, as derived using a Bayesian geostatistical multinomial regression model**



CrI, Bayesian credible interval.
[a] No predictions were made for the white areas on the map.
Image produced using ArcView version 9 (ESRI, Redlands, CA, United States of America) geographical information system.