Mapping Malaria Risk in Bangladesh Using Bayesian Geostatistical Models

Heidi Reid,* Ubydul Haque, Archie C. A. Clements, Andrew J. Tatem, Andrew Valletty, Syed Masud Ahmed, Akramul Islam, and Rashidul Haque

Pacific Malaria Initiative Support Centre (PacMISC), University of Queensland, School of Population Health, Brisbane, Queensland, Australia; International Centre for Diarrhoeal Disease Research, Bangladesh (ICDDR,B), Dhaka, Bangladesh; Australian Centre for International and Tropical Health, Queensland Institute of Medical Research, Brisbane, Queensland, Australia; Emerging Pathogens Institute and Department of Geography, University of Florida, Gainesville, Florida; BRAC, Dhaka, Bangladesh

Abstract. Background malaria-control programs are increasingly dependent on accurate risk maps to effectively guide the allocation of interventions and resources. Advances in model-based geostatistics and geographical information systems (GIS) have enabled researchers to better understand factors affecting malaria transmission and thus, more accurately determine the limits of malaria transmission globally and nationally. Here, we construct Plasmodium falciparum risk maps for Bangladesh for 2007 at a scale enabling the malaria-control bodies to more accurately define the needs of the program. A comprehensive malaria-prevalence survey (N = 9,750 individuals; N = 354 communities) was carried out in 2007 across the regions of Bangladesh known to be endemic for malaria. Data were corrected to a standard age range of 2 to less than 10 years. Bayesian geostatistical logistic regression models with environmental covariates were used to predict P. falciparum prevalence for 2- to 10-year-old children (P/PR2–10) across the endemic areas of Bangladesh. The predictions were combined with gridded population data to estimate the number of individuals living in different endemicity classes. Across the endemic areas, the average P/PR2–10 was 3.8%. Environmental variables selected for prediction were vegetation cover, minimum temperature, and elevation. Model validation statistics revealed that the final Bayesian geostatistical model had good predictive ability. Risk maps generated from the model showed a heterogeneous distribution of P/PR2–10, ranging from 0.5% to 50%; 3.1 million people were estimated to be living in areas with a P/PR2–10 greater than 1%. Contemporary GIS and model-based geostatistics can be used to interpolate malaria risk in Bangladesh. Importantly, malaria risk was found to be highly varied across the endemic regions, necessitating the targeting of resources to reduce the burden in these areas.

BACKGROUND

Because of the wide range of vectors able to transmit malaria, each with unique ecological requirements, malaria is transmitted in a wide range of eco-epidemiological settings. The disease can be coarsely grouped into four such settings: malaria of deep forests, forest fringe, and hills; rural malaria attributable to water-resources development (e.g., irrigation and large dams); rural malaria attributable to wetlands, rivers, streams, coasts, and non-agricultural manmade habitats; and urban and peri-urban malaria.2 Within counties and even across relatively small areas, there may be variations in eco-epidemiological settings and thus, variations in malaria epidemiology.

The advancement of geographical information systems (GIS) and spatial statistics has greatly improved our understanding of malaria dynamics, including its dependence on ecological factors.2–9 More recently, Bayesian geostatistics have been embraced for disease mapping, with the advantages that both environmental covariates and spatial autocorrelation are able to be estimated simultaneously and that full posterior distributions are produced, which can be used to quantify uncertainties in parameters of interest (e.g., predicted prevalence of infection).10 Spatial prediction models have been used to produce malaria risk maps at national,2–8,11–18 sub-continental,19,19 and global scales.20,21

Within Bangladesh, malaria is recognized to occur in two different ecological zones, hilly forested areas and floodplain areas,22 with separate vectors thought to sustain transmission in the different zones.23 In 2006, the Global Fund for AIDS, Tuberculosis, and Malaria awarded Bangladesh $39.6 million US dollars (USD) to support the national malaria-control program.24 Three key objectives of the program are to provide effective diagnosis and treatment to 80% of estimated malaria cases by 2012, to promote the use of long-lasting insecticidal nets (ITNs) in 80% of households in these districts, and to use selective indoor residual spraying (IRS) for containment of outbreaks.24 To help guide the program, a comprehensive malaria survey was conducted in malaria-endemic areas in Bangladesh in 2007.25

The aim of this present study was to produce accurate, validated maps of predicted Plasmodium falciparum prevalence for 2- to 10-year-old children (P/PR2–10) in the endemic areas of Bangladesh and identify high-risk populations at a precision useful for national and sub-national level control priority settings. The applicability of the methods and the implications of the results are discussed in the context of Bangladesh’s malaria-control program, which faces challenges of high-risk areas in remote locations and the spread of multi-drug-resistant P. falciparum.26,27

METHODS

Datasets. Malaria survey data. Data analyzed had been collected during a 2007 malaria-prevalence survey across the malaria-endemic thanas (fourth administrative unit) in Bangladesh23 (Figure 1). The endemic thanas lies within 13 of 64 districts (third administrative unit) in Bangladesh and represents approximately 10% of the country’s population. Based on existing knowledge of malaria prevalence in the country, the survey was designed using a multi-stage cluster sampling technique more fully described in Haque and others.25 Blood samples from 9,750 people (0-92 years of age) were tested, with diagnosis based on the results of rapid diagnostic tests (RDTs). The average Plasmodium prevalence was 3.97%, with 90% of cases being P. falciparum.25 The 9,750 individuals could be geo-located to 354 villages.
Because of the development of immunity, *Pf*PR is age-dependent, rising during early childhood, peaking in older children, and falling through adolescence and adulthood under stable endemic transmission. The assembled *Pf*PR data were, therefore, standardized to the classical age range of 2–10 years using an algorithm based on catalytic conversion models first adapted for malaria by Pull and Grab and modified by Smith and others.

**Environmental variables.** A number of ecological and climatic factors affect both the extrinsic life cycle of the malaria parasite and that of the *Anopheline* vectors. Based on published literature, variables thought pertinent to malaria transmission in Bangladesh were selected for statistical analysis.

A measure of vegetation cover was selected for analysis as the principle malaria vector in Bangladesh; *Anopheles dirus* (*An. dirus*) is known to be a forest-breeding mosquito. Data concerning vegetation cover were obtained from the GlobCover Land Cover product, which is derived from satellite imagery from 2005 to 2006. Vegetation was labeled according to the United Nations (UN) Land Cover Classification System at a resolution of 300 m. For this analysis, vegetation cover was dichotomized into forested, which included a number of more specific forest categories, and not forested, which included all other vegetation categories. Enhanced vegetation index (EVI) data were generated using methods described by Scharlemann and others.

Increases in altitude are commonly associated with cooler temperatures, thus providing a less suitable environment for malaria transmission, whereas increased precipitation has been found to support malaria transmission by creating additional water bodies for malaria vectors to complete their life cycle.
their lifecycle.\textsuperscript{38,39} Data for elevation, minimum temperature, and precipitation were obtained from the WorldClim Project at 1 km\textsuperscript{2} resolution.\textsuperscript{40} Peak malaria season in Bangladesh is from May to August,\textsuperscript{41} coinciding with the monsoon season. Monthly precipitation and temperature were obtained for 6 months over the period from April to September. Values were averaged to give mean minimum temperature and monthly precipitation over this monsoonal period.

Rural and remote populations are purported to face different malaria risks compared with those in urban areas,\textsuperscript{42–46} particularly with respect to access to health care and prophylactic measures, including bed nets. To further investigate this relationship between malaria prevalence and remoteness, accessibility (measured as travel time in minutes) to the nearest settlement of size > 50,000 people was calculated using the methods outlined in Uchida and Nelson,\textsuperscript{47} which are based on GIS data of road networks, road quality, land cover, settlements, and topography.

Using the spatial analyst extension of the GIS software ArcView version 9.3 (ESRI, Redlands, CA), data on vegetation, precipitation, minimum temperature, EVI, and access were extracted for the 354 village locations.

Population data. Gridded population data for 2005 at a resolution of 30-arc second (approximately 1 km\textsuperscript{2}) were obtained from the Global Rural-Urban Mapping Project (GRUMP)\textsuperscript{48} and imported into ArcMap. A 2008 study found GRUMP to be the most accurate spatial population dataset available.\textsuperscript{49} The 2005 population data was projected forward to 2007 using the UN determined growth rate of 1.7\% per annum for Bangladesh over this period.\textsuperscript{50}

Analysis. Assessing spatial autocorrelation. Statistical analyses of spatial autocorrelation in the prevalence data were done using semivariograms in R version 2.9.0 (The R foundation for Statistical Computing, Vienna) with the GeoR package.

Analyzing ecological variables. Backwards step-wise regression analysis was conducted on environmental variables (EVI, elevation, precipitation, minimum temperature, forest cover, and access) using the Stata/SE Version 10 (Stata Corp., College Station, TX) statistical software package to determine their inclusion into the final spatial prediction models. Those variables with a P < 0.1 were retained.

Bayesian geostatistical model. Spatial prediction models were constructed based on the principle of model-based geostatistics\textsuperscript{51} using the Bayesian statistical software WinBUGS version 14.1 (MRC Biostatistics Unit, Cambridge, UK). Based on the RDT results, a binomial logistic regression model was constructed with the proportion infected at each surveyed location as the outcome. The predictor variables included two components: a deterministic component consisting of the fixed environmental effects and a stochastic (i.e., random effect) component with covariance in the random effects modeled using an isotropic, stationary spatial autocorrelation function. For comparison, two similar models were constructed, one without the deterministic component and one without the stochastic component. The deviance information criterion (DIC) statistic was calculated for the models to determine statistically if inclusion of both fixed and stochastic effects improved model fit (models with a lower DIC statistic are considered to show a better fit).

To predict the prevalence at unsampled sites within the endemic thanas, a grid of prediction locations was generated with locations spaced at 0.01 decimal degrees (approximately 1 km). Using the in-built spatial.unipred function in WinBUGS, the geostatistical random effect was interpolated to all prediction locations, and predicted prevalence was calculated by adding the random effect to the sum of the products of the coefficients for the covariates and the values of the covariates at each prediction location. The key outputs of this final analysis are probability distributions of the predicted prevalence at each of the unsampled locations, with each of distributions able to be summarized by a median, standard deviation (SD), and credible intervals (CrI). In this analysis, the median and lower 25\% and upper 75\% CrI will be presented.

Model validation and measures of uncertainty. Validation of predicted prevalence was undertaken by partitioning the data into five random subsets, running the model using four of five subsets, and validating the model with the remaining subset. Five separate models were run, each with a different subset excluded for validation. The accuracy of the prediction was determined in terms of area under curve (AUC) of the receiver-operating characteristic (ROC), with observed prevalence, dichotomized at 0 and > 0, taken as the comparator. This gave an indication of the ability of the model to discriminate between areas where transmission did and did not occur. Mean error and mean absolute error values were also calculated for the predicted prevalences to give an indication of prediction bias and overall precision respectively.

Defining at-risk population. Using the mean predicted \( P/PR_{2–10} \) various at-risk populations, > 0.5\%, > 1.0\%, and > 5.0\%, were defined. The adjusted GRUMP population surface was used to calculate population at risk and the predicted number of cases at a given point in time. With high-risk areas purported to occur in remote locations, the aforementioned accessibility surface was used to determine number of cases reached by targeting 80\% of the most affected populations (i.e., those with the highest predicted prevalences) or alternatively targeting 80\% of at-risk populations that are most accessible (i.e., closest to centers > 50,000 people).

RESULTS

The spatial distribution of \( P/PR_{2–10} \) in the 354 location villages surveyed is presented in Figure 1. A diagnostic variogram (Figure 2) revealed that spatial autocorrelation was a feature of age-standardized data with spatial autocorrelation observed at distances up to approximately 50 km. More simply, malaria prevalence at a given location is affected by the prevalence of neighboring locations (at distances of up to 50 km), necessitating the use of a geostatistical model to correct for this observed spatial dependence.

Forest cover, minimum temperature, and elevation were retained from the backwards step-wise regression analysis and included as fixed effects in the models. The model with both fixed and random effects performed marginally better than the model with no fixed environmental effects, with a DIC of 404.1 compared with 405.5 (Table 1). These two models had similar predictive performance, but the model including fixed environmental effects had marginally lower mean error and mean absolute error estimates (Table 2).

Maps of the median and lower 25\% and upper 75\% Bayesian credible interval for the posterior distributions of \( P/PR_{2–10} \) in Bangladesh are also presented in Figure 3. The distribution of \textit{P. falciparum} shows a large degree of heterogeneity ranging from 0\% to 50\%.
Approximately 6.2 million, 3.1 million, and 0.7 million people were calculated to be living in areas with a predicted PfPR2–10 of >0.5%, >1.0%, and >5.0%, respectively (Table 3). Within these different at-risk populations, targeting the 80% of the population closest to centers >50,000 people reached far fewer cases than if priority was given to reaching the 80% in highest prevalence areas.

DISCUSSION

The results of this analysis build on the findings of the Malaria Atlas Project (MAP), which previously defined the spatial limits of malaria within all endemic countries,²⁰²¹ to present a more detailed empirical description of malaria risk in Bangladesh. The findings indicate that malaria in the country exhibits a unique spatial pattern, with varying prevalence levels and significant and justifiable relationships to environmental variables. Importantly, there exist areas of relatively high transmission that should not be neglected because of their remoteness.

The presentation of spatial variation in disease risk is one of the most important functions of spatial analysis,⁵² with a diagnostic semi-variogram the first step in this process. For malaria in Bangladesh, the observed spatial autocorrelation at distances up to 50 km is approximately one-half of that observed in Kenya¹⁷ but about 10 times that observed on a small island in Vanuatu.¹¹ The practical implication of this is that malaria control in Bangladesh will need to be targeted at a finer spatial scale than in Kenya but not with the same precision as that in Vanuatu.

The observed variation in risk can then be formally included in the geostatistical model to generate a continuous prediction surface, with this surface being the enabling link between the original point-prevalence survey and instructive outputs to guide the control program. For example, if the program managers decide that they will first target populations living in areas with prevalences above 5%, the predicted prevalence map shown in Figure 3 clearly demarcates their targeted area. The calculations in Table 3 are then able to inform program managers of the number of bed nets, for example, that need to be procured to cover 80% of the population in this area. The number of cases averted, however, will depend on which 80% of the population is provided with bed nets: those living closest to settlements >50,000 people or those living in the highest risk areas. This has important implications in meeting the first key objective of the control program, which is to effectively diagnosis and treat 80% of estimated malaria cases.

Targeting those higher endemicity regions located within forested areas will become particularly important as Bangladesh

Table 1

| Results of Bayesian geostatistical models for Bangladesh PfPR2–10 in 2007 |

<table>
<thead>
<tr>
<th>Model Description</th>
<th>Coefficient (posterior mean)</th>
<th>Odds ratio (posterior mean; 95% Bayesian credible intervals)*</th>
<th>DIC†</th>
</tr>
</thead>
<tbody>
<tr>
<td>Bayesian geostatistical model with environmental covariates and no random spatial effect</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>α (intercept)</td>
<td>−2.42 (−3.67 to −1.46)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Elevation</td>
<td>0.64 (0.06–1.05)</td>
<td>14.72 (1.29–82.45) OR/100 m</td>
<td></td>
</tr>
<tr>
<td>Minimum temperature</td>
<td>0.35 (0.07–0.82)</td>
<td>1.23 (1.02–1.31) OR/°C</td>
<td></td>
</tr>
<tr>
<td>Forest‡</td>
<td>1.93 (1.51–2.22)</td>
<td>6.87 (4.52–9.23)</td>
<td></td>
</tr>
<tr>
<td>DIC</td>
<td></td>
<td></td>
<td>593.7</td>
</tr>
<tr>
<td>Bayesian geostatistical model with spatial random effect but no environmental covariates</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>α (intercept)</td>
<td>−4.96 (−5.92 to −3.93)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Φ (rate of decay of spatial correlation)§</td>
<td>4.07 (1.19–8.12)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>σ² (variance of geostatistical random effect)¶</td>
<td>4.17 (2.38–10.0)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>DIC</td>
<td></td>
<td></td>
<td>405.2</td>
</tr>
<tr>
<td>Bayesian geostatistical model with spatial random effect and environmental covariates</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>α</td>
<td>−5.34 (−6.18 to −4.56)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Φ</td>
<td>7.96 (1.96–24.64)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>σ²</td>
<td>3.0 (1.72–6.67)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Elevation</td>
<td>0.34 (0.08–0.84)</td>
<td>4.17 (1.40–34.11) OR/100 m</td>
<td></td>
</tr>
<tr>
<td>Minimum temperature</td>
<td>0.77 (0.21–1.37)</td>
<td>1.29 (1.07–1.58) OR/°C</td>
<td></td>
</tr>
<tr>
<td>Forest</td>
<td>0.79 (0.15–1.52)</td>
<td>2.34 (1.61–4.73)</td>
<td></td>
</tr>
<tr>
<td>DIC</td>
<td></td>
<td></td>
<td>404.1</td>
</tr>
</tbody>
</table>

*Bayesian credible intervals can be interpreted as having a similar meaning to confidence intervals used in frequentist statistics.
†Models with a lower DIC statistic are considered to show a better fit.
‡Unforested is used as the reference category.
§The unit is change in spatial autocorrelation per decimal degree. A lower Φ indicates that spatial correlation occurs over longer distances (i.e., spatial clusters are larger).
¶A higher variance indicates a greater tendency to spatial clustering.
scales up control operations with the forest foci thought to serve as reservoirs from which the lowland and floodplain areas are reinfected. Movement of people to and from the forest not only provides a constant flow of malaria parasites from the forest to rural communities but also exposes immunologically naive individuals to high levels of transmission. To what extent the flow of malaria from the hilly, forested regions maintains transmission in the unforested, lowland areas is an important issue that warrants further research.

An additional challenge for the national control program is the flow of drug-resistant malaria across the border from Myanmar. Areas of elevated transmission along the border region with Myanmar, which can be identified from the risk maps (Figure 3), present strategic locales for both the surveillance of imported malaria cases and drug-resistant malaria. Any attempt to control malaria in Bangladesh is jeopardized by unmonitored migration across the border.

An earlier study identified the most influential environmental parameters for the distribution of *An. dirus* to be forest cover, altitude, rainfall, and temperature. The current analysis replicated the positive associations with forest cover, altitude, and temperature but did not find a significant relationship with rainfall, possibly because of high and relatively consistent levels of rainfall across the study region over the monsoonal period. The positive association between malaria risk and increasing elevation, although unusual for some vector species, is to be expected for *An. dirus*, with higher altitudes providing denser forest cover and thus, a more favorable ecological environment.

As GIS and geostatistical techniques become more sophisticated, there is a need to foster better links between malaria-program managers and researchers such that the most useful data are generated and that these data are accessible and usable outside the research community. Strong research-program links in Kenya led to the development of a 2009 malaria-risk map for Kenya, which is now the cornerstone for planning the scale up of malaria interventions such as ITNs and malaria diagnostics in the country beyond 2009. Additionally, risk maps and geo-spatial data are increasingly being used to support malaria elimination in the southwest Pacific. With closer ties to national programs, the results of model-based geostatistics can be better used.

In conclusion, geographic targeting has tremendous potential to enhance the effectiveness of the national malaria-control program in Bangladesh. In Bangladesh, the areas of relatively high malaria transmission in the hilly, forested region pres-

Table 2

<table>
<thead>
<tr>
<th>Model</th>
<th>AUC*</th>
<th>Mean error (% prevalence)</th>
<th>Mean absolute error ‡ (% prevalence)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Bayesian geostatistical model with spatial random effect and environmental covariates</td>
<td>0.78 (0.74–0.82)</td>
<td>−0.11</td>
<td>5.06%</td>
</tr>
<tr>
<td>Bayesian geostatistical model with spatial random effect but no environmental covariates</td>
<td>0.77 (0.73–0.81)</td>
<td>−0.16</td>
<td>5.64%</td>
</tr>
</tbody>
</table>

*Area under curve (AUC) between 0.5 and 0.7 indicates a poor discriminative capacity, 0.7–0.9 indicates a reasonable capacity, and >0.9 indicates a very good capacity.
† Mean error is a measure of the bias of predictions (the overall tendency to over or under predict).
‡ Mean absolute error is a measure of overall precision (the average magnitude of error in individual predictions).
ent an immediate starting point for malaria-control activities. Targeting this area will ensure that interventions reach those most susceptible and potentially, control the flow of malaria to lowland low-risk areas. Furthermore, the presented P/PR estimates may provide baseline information against which epidemiological changes can be compared as the malaria-control program in Bangladesh is scaled up.

Received March 10, 2010. Accepted for publication June 30, 2010.

Acknowledgments: We are extremely grateful for the expertise of those involved with BRAC and International Centre for Diarrhoeal Disease Research, Bangladesh (ICDDR,B) who designed and carried out the 2007 survey and the participation of the local villagers, without whom this study would not have been possible. We would also like to thank the MAP for providing algorithms to age standardize the survey data and for their comments and guidance. U.H., S.M.A., A.I., and R.H. designed and carried out the original parasitological survey. H.R., A.C.A.C., and A.J.T. conducted the geostatistical analysis. H.R. wrote the provisional manuscript. All authors read and approved the final manuscript.

Financial support: The original survey was implemented with the support of Global Fund for AIDS, Tuberculosis, and Malaria (GFATM). The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

Disclaimer: No conflicts of interest to declare.

Authors’ addresses: Heidi Reid, Archie C. A. Clements, and Andrew Valletty, Pacific Malaria Initiative Support Centre (PacMISC), University of Queensland, School of Population Health, Brisbane, Queensland, Australia, E-mails: heidilouiserieid@gmail.com, a.clements@sph.uq.edu.au, and a.valletty@sph.uq.edu.au. Ubydul Haque and Rashidul Haque, International Centre for Diarrhoeal Disease Research, Bangladesh (ICDDR,B), Dhaka, Bangladesh, E-mails: ubydul@icddrb.org and rhaque@icddrb.org. Andrew J. Tatem, Emerging Pathogens Institute and Department of Geography, University of Florida, Gainesville, FL, E-mail: andy.tatem@gmail.com. Syed Masud Ahmed and Akramul Islam, BRAC, Dhaka, Bangladesh, E-mails: ahmed.sm@brac.net and aislam@theunion.org.

REFERENCES


