An empirical malaria distribution map for West Africa

Immo Kleinschmidt¹, Judy Omumbo², Olivier Briët³, Nick van de Giesen⁴, Nafomon Sogoba⁵, Nathan Kumaseu Mensah⁶, Pieter Windmeijer⁷, Mahaman Moussa³ and Thomas Teuscher³

¹ South African Medical Research Council, Congella, Durban, South Africa
² Kenya Medical Research Institute/Wellcome Trust Collaborative Programme, Nairobi, Kenya
³ West Africa Rice Development Association, Bouake, Côte D’Ivoire
⁴ Center for Development Research, Bonn University, Bonn, Germany
⁵ Malaria Research and Training Center, Université du Mali, Bamako, Mali
⁶ Navrongo Health Research Centre, Navrongo, Ghana
⁷ Alterra, Wageningen, The Netherlands

Summary

The objective of this study was to produce a malaria distribution map that would constitute a useful tool for development and health planners in West Africa. The recently created continental database of malaria survey results (MARA/ARMA 1998) provides the opportunity for producing empirical models and maps of malaria distribution at a regional and eventually at a continental level. This paper reports on the mapping of malaria distribution for sub-Saharan West Africa based on these data. The strategy was to undertake a spatial statistical analysis of malaria parasite prevalence in relation to those potential bio-physical environmental factors involved in the distribution of malaria transmission intensity which are readily available at any map location. The resulting model was then used to predict parasite prevalence for the whole of West Africa. We also produced estimates of the proportion of population of each country in the region exposed to various categories of risk to show the impact that malaria is having on individual countries. The data represent a very large sample of children in West Africa. It constitutes a first attempt to produce a malaria risk map of the West African region, based entirely on malariometric data. We anticipate that it will provide useful additional guidance to control programme managers, and that it can be refined once sufficient additional data become available.

keywords malaria prevalence, mapping, spatial statistics, Geographic Information Systems (GIS), malaria control

correspondence Immo Kleinschmidt, South African Medical Research Council, PO Box 17120 Congella, Durban 4013, South Africa. E-mail: kleinsci@mrc.ac.za

Introduction

Accurate knowledge of the distribution of malaria is an important tool in planning and evaluating malaria control (Snow et al. 1996). A report to the recently held first sub-Saharan African summit meeting on malaria cites a ‘dire lack of extensive and comparable data about malaria’, and calls, amongst other things, for more research on trends in incidence and prevalence, epidemic outbreaks and clinical epidemiology (Sachs 2000).

Global, continental and regional maps of malaria distribution in the past have been largely based on expert opinion (Molineaux 1988), and more recently on climatic suitability (Craig et al. 1999). Empirical maps based on malariometric data have hitherto been produced only at the country or district level (Snow et al. 1998; Thomson et al. 1999; Kleinschmidt et al. 2000). These have the advantage of approximate homogeneity of factors related to malaria control and health services, but they ignore the ‘wider picture’ of effects outside the political boundaries of the country being studied. As transmission intensity and the factors that determine it are rarely confined to these political boundaries, a country or district map is subject to inaccuracies due to spatial effects acting across such boundaries.

The recently created continental database of malaria survey results (MARA/ARMA 1998) provide the opportunity for producing empirical models and maps of malaria distribution.
distribution at a regional and eventually at a continental level. This paper reports on the mapping of malaria distribution for West Africa based on these data. With a total population of nearly 300 million people, sub-Saharan West Africa represents the region with the largest population exposed to high levels of malaria transmission intensity. More detailed knowledge of the distribution of malaria transmission intensity in this region can be used as a basis for more targeted malaria control and health service provision for a very large number of people.

Our objective was to produce a malaria distribution map that would constitute a useful tool for development and health planners in West Africa. We also produced estimates of the proportion of population of each country in the region exposed to various categories of risk to show the impact that malaria is having on individual countries.

Methods and materials

Previous studies using the MARA database for the production of malaria distribution models have described methodological approaches that we essentially followed in this study (Snow et al. 1998; Craig et al. 1999; Kleinschmidt et al. 2000, 2001). In this paper we describe the methods and data used for this study, the results obtained and the implications for malaria control in West Africa. Further details relating to the methods and the results are contained in a technical report available from the corresponding author (Kleinschmidt et al. 1999).

Data

The entomological inoculation rate (EIR) (the number of sporozoite positive bites per person per time unit) would have been the ideal malariometric measure to model for the purpose of mapping the distribution of transmission intensity (Snow et al. 1996). As EIR is not widely available, we modelled parasite prevalence, which is far more commonly available and which is a reasonable proxy for EIR (Beier et al. 1999). Results from parasite prevalence surveys used for this analysis were restricted to those of childhood populations of less than 10 years of age, in order to avoid the effects of population immunity in endemic areas moderating the survey results.

The MARA/ARMA database of geographically referenced survey reports on malaria endemicity in sub-Saharan Africa has been described elsewhere (MARA/ARMA 1998). For this study all data relating to community based surveys between latitudes 1° and 22° north and longitudes 17° west and 16° east, in which at least 50 children between 1 and 10 years of age were examined for the presence of Plasmodium falciparum in blood smears, were extracted from the database. In a few instances where no further age breakdowns were available, surveys on populations between 1 and 15 years were also included. Surveys conducted during known epidemics were excluded, as were those that may represent biased samples, such as those that were restricted to school attenders only. Data from island populations were also excluded. The survey dates covered several decades from about 1970 onwards, and surveys conducted more than once at the same location were combined (summing numerators and denominators). An implicit assumption therefore is that malaria endemicity has remained relatively stable over this period, so that the surveys taken at different time points can be conceptually regarded as a cross-section of surveys, taken at many locations. A total of 450 data points resulted from this process, representing approximately one quarter of a million children surveyed for malaria parasites. The locations of these points are shown in Figure 1.

Distribution of malaria is governed by a large number of factors relating to the parasite, the vector and the host (Molineaux 1988). Predominant among these are climatic and environmental factors, particularly those that affect habitat and breeding sites of the anopheline vectors such as temperature, precipitation, humidity, presence of water, vegetation and man to vector contact. The data used in this study for modelling and mapping malaria parasite prevalence were long-term averages of monthly rainfall, monthly averages of daily minimum and maximum temperature (Hutchinson et al. 1995), normalized difference vegetation index (NDVI, FAO 1991), drainage density (Windmeijer & Andriesse 1993) and estimated population density (Deichman 1996). Monthly climate and vegetation data were aggregated into quarterly averages, from December onwards (to approximately coincide with the drier and wetter seasons, respectively).

Four agro-ecological zones (AEZ) were distinguished on the basis of the length of the growing period, i.e. the period when water is available for vegetative production on well-drained soils. This is a function of precipitation, evaporation and the amount of available water in the soil (FAO 1978). The definition of the zones is as follows: Equatorial Forest zone (> 270 days), Guinea Savanna zone (165–270 days), Sudan Savanna zone (90–165 days) and the Sahel zone (< 90 days), shown in Figure 1. Such zones are well established environmental entities with specific agricultural potential (FAO 1978).

Statistical modelling

For the purpose of this study, the data were divided into three groups corresponding to the AEZ described above, with Sahel and Sudan Savanna combined into one group.
A statistical model was derived for each of these three zone specific groups. This approach was based on the assumption that the factors affecting malaria risk, such as rainfall, would be different in the four AEZ. Parasite prevalence values varied from 0 to 100%. Of the total number of individuals surveyed, 48.8% tested positive. A variogram (Krige 1966; Carrat & Valleron 1992) of prevalence values showed that spatial dependence of the survey results extended over a distance of about 160 km.

Initial variable selection for each model was carried out by performing a stepwise procedure using a generalized linear model (GLM) with logit link function (Hosmer & Lemshow 1989; StataCorp. 1997) and with the parasite prevalence of a point being the response variable. The criterion for inclusion of a variable into the model was set to $P < 0.01$.

To account for spatial correlation in the data, we followed a previously documented (Kleinschmidt et al. 2001) iterative procedure for improving the specification of the covariance structure of the data using a generalized linear mixed model (GLMM) (Littell et al. 1996; SAS 1996). Deviance residuals were calculated for each statistical model that was derived from the initial GLM.

Semivariance (Carrat & Valleron 1992) of the deviance residuals of all pairs of observations was calculated and a variogram constructed to determine if there was evidence of residual spatial correlation, i.e. if the semivariance of pairs of residuals that are close together is markedly less than that of observations which are further apart. The parameters of the function that describes the relationship between semivariance and separation distance (the spatial model) is then used to specify the correlation structure of the data in the GLMM thereby taking account of any residual non-independence in the data. Allowing for spatial correlation may therefore lead to removal of some variables from the model due to the resultant inflation of the standard errors. Deviance residuals of the spatially adjusted model are calculated and a new variogram is constructed. This process is iterated until the variogram no longer changes – indicating that a covariance structure corresponding to the model residuals is adequately specified (Kleinschmidt et al. 2001; Appendix 1).

To improve the fit (i.e. reduce residual deviance), each variable that survived the above procedure was transformed into seven different fractional polynomials (Royston et al. 1999). The transformation producing the

---

Figure 1 Survey locations, and agro-ecological zones for West Africa.
biggest reduction in residual deviance was chosen if this
reduction in deviance exceeded 3.84 compared with the
untransformed variable. Transformations that were tried
for each variable $x$ were $1/x^2$, $1/x$, $1/x^{0.5}$, $\ln(x)$, $x^{0.5}$, $x^2$
and $x^3$.

Once the zone specific models had been derived, these
were used to produce a map based on the predictor
variables which are available as map images. The zone
boundaries represent a somewhat arbitrary cut-off, with
places near such a boundary sharing characteristics of the
zones on both sides of the boundary. Predictions of parasite
prevalence along a boundary between two zones were
therefore based on a weighted mean of the predictions
obtained from the models for the two adjoining zones, with
the weights dependent on the distances from the boundary.
This interpolation of predictions along zone boundaries
was carried out up to a distance of 160 km from each zone
boundary, as the previously constructed variogram showed
that spatial effects were limited to approximately this
distance.

To improve prediction in places where there is consid-
erable divergence between model predictions and obser-
vations in a local neighbourhood we used a previously
developed method (Kleinschmidt et al. 2000) based on
kriging (Krige 1966) of the residuals of the final model
predictions. A kriged map of deviance residuals is calcu-
lated, which is added to the predicted values on the logit
scale before transforming the result back to proportions.
The addition of kriged residuals will allow the map to
deviate from the model and move closer to the observed
values, if such deviation is supported by other observed
values in the neighbourhood. This improves the final map
in the sense that it does not deviate too severely from the
observations, which is particularly important if the model
does not adequately explain the observed variation in
transmission risk.

Our method therefore involves a combination of mod-
elling (predictions based on the values of climatic and
environmental variables at each location) and kriging
(interpolation of prevalence values at points between
observed survey locations). This has the effect that the map
predictions are primarily model driven in areas with a
paucity of points, whereas in areas with an abundance of
survey locations the map values will be primarily deter-
mined by the actual observed values at these points.

**Predicted population at risk**

We overlayed the final predicted prevalence map on a
population density map (Deichman 1996) to calculate the
population at risk for different endemicity categories for
each country, excluding urban areas.

**Results and discussion**

Significant explanatory variables for the model for the
Sahel and Sudan Savanna zones were: average monthly
rainfall from March to May, average minimum tempera-
ture from September to November and from December to
February, average maximum temperature from March to
May and from September to November, average vegetation
index from March to May and drainage density. For the
model for the Guinea Savanna zone the significant vari-
ables were average monthly rainfall from September to
November, average vegetation index from December to
February, and from March to May, average minimum
temperature from December to February and from June to
August, average maximum temperature from September to
November, difference in maximum monthly and minimum
monthly vegetation index, drainage density and population
density. Finally, the model for the Forest zone contained
average maximum temperature from September to
November and from June to August, and average monthly
rainfall from September to November. As all the models
are multiple variable models, each variable is corrected for
all the other variables in the model. The relationship
between these quantities and parasite prevalence is com-
plex, and we give details of model coefficients and their
plausibility in the technical report (Kleinschmidt et al.
1999).

Figure 2 shows the final map of predicted risk of malaria
infection for children under 10 years during a location’s
main malaria season that was predicted from our models
after processing the predictions in the way described above.
The grouping of the map predictions into the four
categories of risk shown in the map are the same as were
used for a country level malaria map for Mali
(Kleinschmidt et al. 2000).

Our data contained a handful of points ($n = 21$) that
could be regarded as urban on account of their 1995
population density being above 386/km$^2$ (US Bureau of
the Census 1995). Average parasite prevalence in these
‘urban’ surveys was 45.1%, compared with a mean of
46.7% for non-urban surveys (two-sample $t$-test,
$P = 0.77$). This result was not sensitive to the particular
population density cut-off chosen for the definition of
urban sites, and it was true in all three zones. It was only
in the Guinea Savanna zone that there was a significantly
higher prevalence for points with population densities
below 1 per sq km after adjusting for other factors in
the model. Despite this lack of evidence in the MARA
database for lower parasite ratios in urban areas, we
considered our data too unrepresentative of urban areas
to make any predictions there. Urban areas were therefore
excluded from the prediction map, and from the
population at risk calculations. It is quite likely that some surveys were taken in places which were rural outskirts of urban areas at the time of the surveys, but which are now urban. Whilst climatic factors might justifiably have been regarded as constant over the time that the surveys were conducted, this assumption is almost certainly not uniformly valid for population density, and this may be the reason for it not featuring more prominently as a significant explanatory variable.

Comparing our final map predictions with the observed prevalence values of the 450 surveys, 77.6% (349/450) of the surveys were correctly classified, i.e. the predicted prevalence category agreed with the observed prevalence category ($\kappa = 0.62, p < 0.0001$). Of the points where there was a disagreement between the observed and predicted prevalence categories, only three were misclassified by more than one category value.

A visual comparison of our map with the climatic suitability map by Craig et al. (1999) shows many similar features, which is not surprising as climatic factors were involved in the production of both maps. Visual comparisons of our map with previous expert opinion maps (Haworth 1988; Wernsdorfer & McGregor 1988) and with a map of Mali derived from MARA data (Kleinschmidt et al. 2000) also show broad agreement. We should caution that there were several countries in the regions which were either poorly covered by surveys, or not at all. We are optimistic that this situation will improve in future and this will allow a more accurate map to be produced. However, in the meantime our map predictions for these areas are entirely based on our models that were derived from data from neighbouring countries. This may still give reasonable predictions for smaller countries or those that are surrounded by countries with an abundance of data points, but it is bound to give inaccurate estimates for countries on the periphery of our map window, such as Niger. We excluded Niger from the calculation of populations at risk (Table 1) for this reason. Most of Nigeria, and the central parts of Ghana, also suffered from a sparse coverage of points, and hence the predictions in these regions are model dependent rather than interpolation driven. A current shortcoming in our modelling methodology is the fact that we are unable to give an estimation error for the various parts of the map.

Figure 2 Predicted prevalence of *P. falciparum* in children aged 2–10 years for West Africa. *Differing map resolutions have caused some digitization error along the coast, causing some coastal urban areas not to show on the map.*
The proportion of population in each country exposed to each of the four risk categories varies considerably between countries in the region (Table 1). For example, the population living in areas with less than 30% prevalence make up 17% of the population of the entire region, with high proportions of the population in this category living in Mauritania (50%), Guinea Bissau (30%), Mali (31%) and Senegal (23%). Some of these could be populations with low levels of immunity and it can reasonably be expected that exceptional rainfall will cause significant morbidity in all age groups. Often such areas are remote and interventions are hampered by poor health service infrastructures. On the other hand, populations in areas with predicted prevalences above 30% (categories 3 and 4 on the map) are more likely to have some measure of immunity with young children and pregnant women being the groups most vulnerable to morbidity and mortality due to malaria. According to our map, 58% of the population of West Africa (168 million people) fall into this category. In Côte d’Ivoire, Togo, Burkina Faso, Sierra Leone and Liberia 70% or more of the population is exposed to this level of transmission intensity.

Although the highest prevalence category, namely 70–100%, occupies a considerable area on the map, the proportion of population living in these areas is reasonably small in all countries except Togo. In the West African region as a whole, about 16 million people are exposed to this high level of transmission intensity. Marsh and Snow (1999) suggested that vector-contact reducing measures such as insecticide treated materials (ITM) may change severe-disease patterns of malaria and consequently case-fatality in high endemicity settings. The introduction of ITMs on a large scale should be accompanied by more intense monitoring efforts in such circumstances. Our prediction map helps to identify areas where such long-term morbidity monitoring might need to accompany ITM deployment.

Ideally, we would like to have a map that clearly identifies two types of areas requiring two quite distinct types of intervention packages. These would be epidemic prone areas, and areas with stable malaria endemicity. In areas of unstable malaria transmission, surveillance efforts, the stocking of efficacious insecticides such as DDT for in-house spraying as well as appropriate and affordable diagnosis and treatment algorithms play a primary role. In holoendemic areas, on the other hand, rapid diagnosis and treatment, intermittent treatment during pregnancy, behavioural aspects related to the large scale use of ITMs and innovative strategies to ensure the availability of high

### Table 1 Predicted percentage of population at risk by country and risk category (excluding urban populations)

<table>
<thead>
<tr>
<th>Country†</th>
<th>Predicted prevalence of &lt; 10%</th>
<th>Predicted prevalence of 10–30%</th>
<th>Predicted prevalence of 30–70%</th>
<th>Predicted prevalence above 70%</th>
</tr>
</thead>
<tbody>
<tr>
<td>Benin</td>
<td>0</td>
<td>5</td>
<td>43.4</td>
<td>12</td>
</tr>
<tr>
<td>Burkina Faso</td>
<td>0</td>
<td>17</td>
<td>76</td>
<td>0</td>
</tr>
<tr>
<td>Cameroon</td>
<td>1</td>
<td>16</td>
<td>58</td>
<td>2</td>
</tr>
<tr>
<td>Côte d’Ivoire</td>
<td>0</td>
<td>4</td>
<td>75</td>
<td>0</td>
</tr>
<tr>
<td>Gambia</td>
<td>0</td>
<td>8</td>
<td>44</td>
<td>0</td>
</tr>
<tr>
<td>Ghana</td>
<td>1</td>
<td>15</td>
<td>46</td>
<td>17</td>
</tr>
<tr>
<td>Guinea</td>
<td>1</td>
<td>12</td>
<td>57</td>
<td>3</td>
</tr>
<tr>
<td>Guinea Bissau</td>
<td>2</td>
<td>30</td>
<td>13</td>
<td>0</td>
</tr>
<tr>
<td>Liberia</td>
<td>0</td>
<td>1</td>
<td>81</td>
<td>2</td>
</tr>
<tr>
<td>Mali</td>
<td>3</td>
<td>28</td>
<td>66</td>
<td>1</td>
</tr>
<tr>
<td>Mauritania</td>
<td>20</td>
<td>30</td>
<td>6</td>
<td>1</td>
</tr>
<tr>
<td>Nigeria</td>
<td>0</td>
<td>8</td>
<td>48</td>
<td>8</td>
</tr>
<tr>
<td>Senegal</td>
<td>1</td>
<td>22</td>
<td>41</td>
<td>2</td>
</tr>
<tr>
<td>Sierra Leone</td>
<td>0</td>
<td>0</td>
<td>79</td>
<td>2</td>
</tr>
<tr>
<td>Togo</td>
<td>0</td>
<td>0</td>
<td>39</td>
<td>38</td>
</tr>
<tr>
<td>Entire region</td>
<td>2.4</td>
<td>14.8</td>
<td>52.7</td>
<td>5.4</td>
</tr>
<tr>
<td>Total population at risk</td>
<td>7 006 869</td>
<td>42 941 669</td>
<td>152 779 264</td>
<td>15 698 929</td>
</tr>
</tbody>
</table>

* Percentages do not sum to 100% as urban populations have been excluded and parts of some countries lie outside the map ‘window’.
† Excluding Niger.
quality first line treatments at home might be considered high priority by country control programmes. Our map cannot provide such clear division into endemic and epidemic areas, but it can be used to guide such decisions.

It is well known that malaria transmission intensity exhibits strong spatial heterogeneity even at a local level. It is therefore likely that the map may be at variance with local experience in some places. Where this occurs, it ought to motivate further investigation through well conducted local surveys.

A possible source of variation that is not determined by natural factors such as climate and drainage density may be differences in socio-economic development, which have played a part in malaria control and eradication elsewhere, probably coinciding with other factors (Bruce-Chwatt & de Zuleta 1980; Packard 1984; Molineaux 1988; Wernsdorf & Wernsdorfer 1988). Socio-economic development could reduce malaria transmission in a variety of ways. For example, increases in household income of women and poverty reducing measures in general have the potential to reduce exposure to malaria and to improve health seeking behaviour and quality of treatment. However, socio-economic development in a high transmission tropical setting could equally increase malaria transmission because of changes such as forest clearing or the migration of people with little or no immunity into areas of high endemicity. We have been unable to model such factors in our analysis due to the fact that such data for the entire region are currently not available with adequate spatial resolution. It is highly likely that there are other unmeasured, perhaps more local factors that determine variation in parasite prevalence.

A further source of variation that has not been taken into account in this study is variation in prevalence by season and by age (M S Sissoko, O Briet, M Sissoko et al. unpublished observation). The impact of these factors will differ according to the endemicity level of an area. It was our opinion that the differentiation that was available within the results of many surveys was inadequate to stratify the data by these factors.

A regional malaria risk map, such as the one produced in this study, will allow planners to assess the possible health impacts of measures aimed at improving food security through the promotion of large scale irrigation and wetland management projects. Elsewhere in Africa such developments have significantly increased malaria infection and morbidity in epidemic prone areas of unstable malaria (Ghebreyesus et al. 1999). However, the same agricultural production methods are unlikely to affect the malaria risk profile of rural populations living in areas characterized by high parasite prevalences (Faye et al. 1995; Dossou-Yovo et al. 1998).

Finally, the map will also help guide public health research managers in identifying appropriate study environments for intervention trials as well as assist with the identification of populations potentially benefiting from new interventions.

The data we used represent a very large, albeit imperfectly, sampled population of children in West Africa. This study is a first attempt to produce a malaria risk map of the West African region, based entirely on malariometric data. We anticipate that it will provide useful additional guidance to control programme managers, and that it can be refined once sufficient additional data become available.

Acknowledgements

This regional analysis of the MARA database was conducted with the aid of a grant (centre file 00397) from the International Development Research Centre (IDRC), Ottawa, Canada. We gratefully acknowledge support for the MARA project by the Multilateral Initiative on Malaria (MIM), the IDRC and Roll Back Malaria. We also thank the West Africa Rice Development Association (WARDA) for providing the necessary infrastructure to host the workshop. Particular thanks are due to Mylene Kamara (WARDA) for producing the population at risk calculations and to Nicole Teuwsen for assisting with GIS work.

References


Malaria distribution map for West Africa


