Mapping the receptivity of malaria risk to plan the future of control in Somalia

Abdisalan Mohamed Noor,1,2 Victor Adagi Alegana,1 Anand Prabhakar Patil,3 Grainne Moloney,4 Mohammed Borle,4 Fahmi Yusuf,5 Jamal Amran,5 Robert William Snow1,2

ABSTRACT

Objectives: To measure the receptive risks of malaria in Somalia and compare decisions on intervention scale-up based on this map and the more widely used contemporary risk maps.

Design: Cross-sectional community Plasmodium falciparum parasite rate (PPR) data for the period 2007–2010 corrected to a standard age range of 2 to <10 years (PPR2–9) and used within a Bayesian space–time geostatistical framework to predict the contemporary (2010) mean PPR2–10 and the maximum annual mean PPR2–10 (receptive) from the highest predicted PPR2–10 value over the study period as an estimate of receptivity.

Setting: Randomly sampled communities in Somalia.

Participants: Randomly sampled individuals of all ages.

Main outcome measure: Cartographic descriptions of malaria receptivity and contemporary risks in Somalia at the district level.

Results: The contemporary annual PPR2–10 map estimated that all districts (n=74) and population (n=8.4 million) in Somalia were under hypoendemic transmission (<10% PPR2–10). Of these, 23% of the districts, home to 13% of the population, were under transmission of <1% PPR2–10. About 58% of the districts and 55% of the population were in the risk class of 1% to <5% PPR2–10. In contrast, the receptivity map estimated 65% of the districts and 69% of the population were under mesoendemic transmission (>10%–50% PPR2–10) and the rest as hypoendemic.

Conclusion: Compared with maps of receptive risks, contemporary maps of transmission mask disparities of malaria risk necessary to prioritise and sustain future control. As malaria risk declines across Africa, efforts must be invested in measuring receptivity for efficient control planning.

INTRODUCTION

Malaria receptivity is a measure of the intrinsic vector transmission potential of an area.1 Interest in measuring malaria receptivity has emerged following the resurgence of the malaria elimination agenda2 3 and the need to quantify the risks posed by human population movement leading to the reintroduction of transmission.4 5 However, understanding receptivity is equally important to decision-making for countries that are implementing control. In low stable endemic countries, national programmes need to understand the risks posed by withdrawal of interventions from areas that are historically high transmission.3 6 In unstable transmission areas where parasite exposure is highly seasonal and prone to climatic anomalies, targeting interventions to prevent the risk of epidemics are a priority.1 The empirical malaria risk maps that are commonly available to countries to support malaria control planning are those that represent the contemporary distribution of risk under...
control\textsuperscript{7,8} and are therefore of limited value in defining the epidemic potential or the receptive rebound risks of withdrawing or failing to sustain interventions.

Measuring receptivity for malaria control countries ideally requires empirical data from a period of no control and under the optimum transmission conditions. There is hardly any country in Africa that has remained universally control-naive over the past 100 years.\textsuperscript{9} Alternatively, nationally representative empirical data on malaria transmission during the pre-Roll Back Malaria (RBM) era, or before intervention scale-up reached critical thresholds for a given country, may represent the best approximation of receptivity. Furthermore, such information must be resolved to administrative decision-making units to which malaria resources are allocated to make them relevant for policy. In this study, we use community \textit{Plasmodium falciparum} parasite prevalence data from 2007 to 2010 within a model-based geostatistical (MBG) framework to develop contemporary and receptive risk maps and resolve endemicity to districts in Somalia.

\section*{METHODS}

\subsection*{Country context}
Somalia is divided into the three zones of North West (Somaliland), North East (Puntland) and Central South (figure 1). The northern zones are generally dry and hot, whereas the Central South zone has subtropical climate and is where the two major rivers of the country, the Shabelle and the Juba, are located.\textsuperscript{10,11} \textit{Anopheles arabiensis} is the dominant malaria transmitting vector throughout the country, although \textit{Anopheles funestus} is reported in Central South.\textsuperscript{12,13} \textit{P falciparum} is the dominant species of the malaria parasite.\textsuperscript{14–16} The presence of \textit{Plasmodium vivax} cases have also been reported with studies in Somaliland showing relatively low parasitemia rates.\textsuperscript{17} It is believed that there are no reports of \textit{Plasmodium malariae} cases in Somalia.\textsuperscript{18} The climate and geography of Somalia makes it conducive for malaria transmission throughout the year, with no clear dry season.\textsuperscript{19}

\begin{figure}[h]
\centering
\includegraphics[width=\textwidth]{figure1.png}
\caption{Zone, regional and district maps of Somalia showing the distribution of the age-standardised community \textit{Plasmodium falciparum} parasite rate (PR\textsubscript{2–10}) data (n=1558) assembled during the period 2007–2010 (including 54 surveys undertaken in 2011). The zones are CS, Central South; NE, North East; NW, North West. The thick black line show the zone boundaries, the thin black lines show the regional boundaries and the thin grey lines show the district boundaries. The blues lines show the location of the Juba (lower) and Shabelle (upper) Rivers.}
\end{figure}
high vivax antibody responses.\textsuperscript{17} The failure of the long rains in Somalia in 2010 combined with the below average rainfalls in previous two seasons have resulted in a severe drought.\textsuperscript{11}

Since 1991, Somalia has had no effective central government and has experienced frequent internal armed conflicts resulting in the breakdown of public services and political disintegration.\textsuperscript{18} Although the transitional federal government, formed in February 2004, is internationally recognised as the official government of Somalia,\textsuperscript{19} in reality the three zones currently function as semi-independent states. Consequently, there are three ministries of health, but the majority of healthcare is provided by international and national non-government organisations that have come together under the umbrella of Somalia Aid Coordinating Board, which was later reformed as the Somalia Support Secretariat.\textsuperscript{20}

The main funder of malaria control in Somalia is the Global Fund to Fight Aids TB and Malaria with the United Nations Children Fund (Unicef) as the principal recipient and the WHO as subrecipient.\textsuperscript{21, 22} In 2010, the second national malaria strategy was launched with universal scale-up of vector control, parasitological diagnosis, effective antimalarials, improved surveillance and epidemic preparedness and response as the main strategic approaches.\textsuperscript{22, 23} Since 2004, over US$77 million has been approved to Somalia by the Global Fund to Fight Aids TB and Malaria for malaria control resulting in the distribution of almost 1 million long-lasting insecticidal nets (LLINs).\textsuperscript{24}

\textbf{Community survey data}

The community \textit{P. falciparum} parasite rate (\textit{PfPR}) is the most commonly used indicator for mapping malaria transmission.\textsuperscript{7} This is because it is easy to measure, has a historical legacy and a predictable relationship with other measures of transmission intensity, such as entomological inoculation rate and the basic reproductive rate.\textsuperscript{25, 26} The \textit{PfPR} data used for the present study were assembled through the Food and Agriculture Organization—Food Security and Nutrition Analysis Unit (FAO-FSNAU) surveys undertaken regularly in Somalia.\textsuperscript{16, 27} These surveys were initially established to monitor the nutritional status of children <5 years of age and began in 2000.\textsuperscript{27} Investigations of malaria prevalence covering persons of all ages were only included from 2007\textsuperscript{16} and have been undertaken annually since covering most regions of Somalia. A detailed description of the sampling design is provided elsewhere.\textsuperscript{16} During the survey, respondents provided a finger prick blood sample that was examined for the presence of \textit{P. falciparum} infection using a rapid diagnostic test (Paracheck Pf, Orchid Biomedical Systems, Goa, India). Consent was obtained for all individuals before interview and separately for the malaria testing. Additional information was recorded on the date of survey and age and sex of participants. After all survey

data were assembled, each surveyed community was geocoded using combinations of global positioning systems, electronic gazetteers (Google Earth, Encarta and Alexandria) and other sources of longitude and latitude such as a settlement database collated by FAO-SWALIM.\textsuperscript{11}

\textbf{Assessment of ecological and climatic predictors of malaria risk}

The transmission intensity of malaria is influenced by climatic and ecological factors through their independent or combined effects on the survival of the Anopheles vectors and the Plasmodium parasites within the vector.\textsuperscript{28} The ecological and climatic factors used commonly to improve the precision of empirical malaria mapping are urbanisation, rainfall, temperature and distance to potential mosquito larva breeding sites.

Data at 1×1 km spatial resolution on urbanisation,\textsuperscript{29} annual mean precipitation,\textsuperscript{30, 31} and enhanced vegetation index\textsuperscript{32} were assembled. Maps of rivers, flood plains, reservoirs and coastal wetlands were assembled from the Global Wetlands and Lakes Database\textsuperscript{33} and Euclidean distances to these proximates of breeding sites were computed in ArcGIS 10 (ESRI Inc., New York, USA). As a metric for the effect of temperature on malaria transmission, a temperature suitability index (TSI) at a spatial resolution of 1×1 km was used. TSI was constructed using monthly temperature time series within a biological modelling framework to quantify the effect of ambient temperature on sporogony and vector surviv- orship and determine the suitability of an area to support transmission globally.\textsuperscript{34} The values of the underlying ecological and climatic covariates were extracted to each survey location using ArcGIS 10 Spatial Analyst tool. Distance to potential breeding sites was log-transformed before analysis because of its high positive skew. The covariates were then included in total-sets analysis, which is an automatic model selection process based on a generalised linear regression model and implemented using the \textit{bestglm} package in R.\textsuperscript{35, 36} This approach selects the best combination of the covariates based on the value of the Bayesian Information Criteria statistic,\textsuperscript{37} which selects the lowest Bayesian Information Criteria as the best model fit. Details of the ecological and climatic predictors of \textit{PfPR} and the results of the total-set analysis are provided in supplementary information 1.

\textbf{The space–time Bayesian geostatistical model for predicting \textit{P. falciparum} distribution in Somalia}

Space–time MBG methods offer the flexibility of predicting an outcome to any given year and location in a time series by harnessing fully both the spatial and temporal relationships of the data and generate uncertainties of the predictions from the full posterior distributions.\textsuperscript{38}

In this study, the assembled \textit{PfPR} data were standardised to the classical age range of 2 to <10 years using an algorithm based on modified catalytic conversion
models. The continuous surfaces of the age-standardised data \( P f_{PR2-10} \) were generated using a space–time MBG framework whereby Bayesian inference was implemented using the Markov Chain Monte Carlo algorithm. Details of model code and statistical procedures are provided in supplementary information.

In brief, the value of \( P f_{PR2-10} \) was modelled as a transformation of a spatiotemporally structured field superimposed with unstructured (random) variation on a regular 1 km grid from 2007 to 2010. The number of \( P falciparum \)-positive responses from the total sample at each survey location was modelled as a conditionally independent binomial variate given the unobserved underlying age-standardised \( P f_{PR2-10} \) value and a linear function of the climatic and environmental predictors. The unstructured component was represented as Gaussian distribution with zero mean. The spatiotemporal component was represented by a stationary Gaussian process with covariance defined by a spatially anisotropic version of the space–time covariance function proposed by Stein (2005).

To partly model seasonality, the covariance function was modified to allow the time-marginal model to include a periodic component of wavelength 12 months in the temporal covariance structure. Each survey was referenced temporally using the midpoint (in decimal years) between the recorded start and end months. For each grid location, samples of the annual mean of the full posterior distribution of \( P f_{PR2-10} \) for each year were generated. These \( P f_{PR2-10} \) samples were then used to generate continuous maps of the annual mean. To determine the probable maximal malaria risk, the highest value of predicted mean annual \( P f_{PR2-10} \) value at each 1 km grid location over the period 2007–2010 was extracted. These were then combined to generate a single map of maximum mean \( P f_{PR2-10} \).

Assessing uncertainty of model predictions

As a first step to understanding the uncertainty around the predictions of \( P f_{PR2-10} \) using the Bayesian geostatistical model, the continuous mean maps were accompanied by estimates of the posterior standard deviation (SD). For the maximum mean \( P f_{PR2-10} \) map, the posterior SDs associated with the selected mean value was used. To allow for a scaled comparison of the uncertainty of the 2010 \( P f_{PR2-10} \) annual mean \( P f_{PR2-10} \), the coefficient of variation, which is a measure of dispersal around the mean, was computed as the ratio of the SD to the mean. Higher values of the coefficient of variation suggest increasing uncertainty of model predictions. In addition, a spatially representative validation set of \( P f_{PR2-10} \) survey data were also selected using a spatially declustered sampling algorithm. The annual predictions were then repeated in full using the remaining data to predict mean \( P f_{PR2-10} \) at the validation locations. The ability of the model to predict point values of \( P fPR \) at unsampled locations was quantified using two simple summary statistics: the mean prediction error (MPE) and the mean absolute prediction error (MAPE). The MPE provides a measure of the model bias, while the MAPE is a measure of the average accuracy of individual predictions.

Defining district-level malaria endemicity and population at risk

A 2010 population surface for Somalia at 100 km resolution with a spatial resolution of 50 m was provided by the AfriPop project. Using this map, estimates of the total population of each 1 km pixel to which mean \( P f_{PR2-10} \) was predicted were computed in ArcGIS 10. The 1 km grid squares with attached population estimates were further classified separately by the mean \( P f_{PR2-10} \) and the maximum mean \( P f_{PR2-10} \). To weight endemicity for population distribution, only those grid squares with population were retained and from these the mean \( P f_{PR2-10} \) was computed for each district. Based on the aggregate mean \( P f_{PR2-10} \), districts were then classified using a modification of the classical malaria endemicity classification. The hypoendemic class \( (\leq 10% P f_{PR2-10}) \) was split into \(< 1% P f_{PR2-10}, 1% \to 5% P f_{PR2-10}, 5% \to 10% P f_{PR2-10} \). The endemicity class of \(< 1% P f_{PR2-10} \) represents the threshold at which an area is considered to be under low endemic control and a decision for sustaining control or aiming for elimination can be made. The hyperendemic \( (\geq 50% \to 75% P f_{PR2-10}) \) and holoendemic \( (> 75% P f_{PR2-10}) \) classes were also combined, while the meso-endemic class \( (> 10% \to 50% P f_{PR2-10}) \) remained unchanged. The number of districts and population by these endemicity classes were then summarised based on both the 2010 mean \( P f_{PR2-10} \) (contemporary risk) map and the maximum mean \( P f_{PR2-10} \) (receptive risk) map.

RESULTS

Predictions of mean annual \( P f_{PR2-10} \) to 2010 and maximum mean posterior \( P f_{PR2-10} \)

A total of 1558 \( P falciparum \) community surveys (figure 1) in which 103 593 persons were examined were assembled for the period 2007–2011 in Somalia. The majority of the data were located in the Central South zone where most of the population live. Survey data were collected across nine different months over the 5 years, with the majority of data (76%) assembled in the months of November, December, May and June corresponding to the peak malaria seasons in Somalia. The results of the total-set analysis showed that the model with urbanisation, precipitation, enhanced vegetation index and distance to main water bodies and flood plains as the best fit in predicting \( P fPR \), and these variables were subsequently included in the malaria prediction model (supplementary information 1). TSI was not selected as a statistically strong predictor of \( P falciparum \) prevalence in Somalia.

The continuous 2010 malaria endemicity map for Somalia \( P f_{PR2-10} \) showed the majority of locations were
Figure 2  
(A) Map of the posterior annual mean $PPR_{2-10}$ prediction to 2010 (contemporary) at $1 \times 1$ km grid location in Somalia. 
(B) Map of the maximum mean $PPR_{2-10}$ prediction (receptive) at $1 \times 1$ km grid location as computed from the posterior annual mean $PPR_{2-10}$ prediction for each year from 2007 to 2010. 
(C) Map of the coefficient of variation (the SD/the mean $PPR_{2-10}$ prediction) of the contemporary prediction at $1 \times 1$ km grid location. 
(D) Map of the coefficient of variation at $1 \times 1$ km grid location of the receptive prediction. 

The thick black lines show the zone boundaries, the thin black lines show the regional boundaries and the thin grey lines show the district boundaries. Higher coefficient of variation of the predictions suggests higher uncertainties of the $PPR_{2-10}$ predictions. The scale bar for the continuous $PPR_{2-10}$ ends at 50, which is the upper limit of mesoendemic transmission. 

The blues lines show the location of the Juba (lower) and Shabelle (upper) Rivers.

predicted to have parasite prevalence of <5% indicating largely hypoendemic transmission (figure 2A). The majority of areas in North East and North West zones were predicted to be under \( P/PR_{2-10} <1\% \). In contrast, the maximum annual mean \( P/PR_{2-10} \) map showed a substantially different risk landscape, with the majority of the Central South zone having \( P/PR_{2-10} \) of >10% and an a maximum predicted mean of 38%, suggesting that peak malaria transmission in all of Central South zone and southern parts of North East zone is mesoendemic (figure 2B). In the northern zones, maximum mean risks were predicted to be predominantly between 5% and <10% \( P/PR_{2-10} \).

The MPE and MAPE associated with the full space–time geostatistical model was 4.8% and 0.2%, respectively. The 2010 annual mean \( P/PR_{2-10} \) predictions were associated with higher coefficients of variation compared with the maximum mean \( P/PR_{2-10} \) predictions, although the difference was moderate (figure 2C, D). In both maps, uncertainty appeared highest in northern zones where data in space and time were fewest.

**District estimates of contemporary and receptive malaria risk**

According to the contemporary district malaria endemicity map based on the annual mean \( P/PR_{2-10} \) map of 2010, all districts in Somalia were under hypoendemic transmission (figure 3A). Of the 74 districts, an estimated 17 (23%) districts covering about 1.1 million people (13%) were in the <1% \( P/PR_{2-10} \) risk class (table 1 and figure 3A). The majority of the districts (58%) and population (55%) were in the risk class of 1%–<5% \( P/PR_{2-10} \) and the rest were under risks of 5%–10% \( P/PR_{2-10} \).

In contrast, the receptive risk map showed that there were no district under low stable endemic control (<1% \( P/PR_{2-10} \), and the majority of the districts (65%) and population (69%) were under the mesoendemic class (table 1 and figure 3B), with an upper district maximum mean estimate of 35% \( P/PR_{2-10} \). About 27% of the districts and 17% of the population were in the upper range of hypoendemicity (5%–10% \( P/PR_{2-10} \)). The rest of the districts and population were in the intermediate hypoendemic class of 1%–<5% \( P/PR_{2-10} \).

**DISCUSSION**

The malaria risk maps that are commonly available to countries in Africa to support malaria control planning are those that represent the contemporary distribution of risk.34–46 They have been developed primarily from geo-coded parasite rate survey data7 usually to predict risk to the most recent data year and therefore reflect transmission under scaled interventions during the era of the RBM partnership.51 52 In this study, we argue that, in addition to contemporary maps of malaria risks, low stable endemic control and unstable transmission countries require maps of receptivity to assess the risks of rebound and epidemics and decide on where to scale-up and/or sustain intervention coverage. To demonstrate this, we used community \( P/PR \) survey data from the period 2007–2010 within a space–time MBG framework to generate two continuous malaria risk maps for Somalia. One is a contemporary map of annual mean \( P/PR_{2-10} \) predicted to 2010 (figure 2A) and the other is the maximum annual mean \( P/PR_{2-10} \) map derived from the highest mean \( P/PR_{2-10} \) value predicted to a location in any year over the study period to approximate receptivity (figure 2B). We resolved these maps to the district, which is the malaria decision-making unit in Somalia, and classified them by endemicity using population-weighted mean \( P/PR_{2-10} \) (figure 3).

The efficacy and impact of malaria interventions on disease in an area are dependent on its intrinsic transmission potential.53–55 This is the theoretical basis upon which international guidelines for malaria control are formulated.1 56 One of the most important applications of malaria risk maps for control planning, therefore, is to inform the spatial targeting of the appropriate mixes of interventions.57 For Somalia, the results of the comparison of the contemporary and the receptive risk maps represents two very different transmission scenarios (table 1 and figure 3). The contemporary malaria risk map predicted that all of Somalia was under conditions of hypoendemic transmission (≤10% \( P/PR_{2-10} \)) in 2010 while a fifth of the districts under risks of <1% \( P/PR_{2-10} \) and the majority of the districts and population were in the intermediate hypoendemic transmission class of 1% to <5% \( P/PR_{2-10} \). Under these transmission conditions, targeted distribution of LLINs and indoor residual spraying aimed at control of residual foci are recommended, while intermittent presumptive treatment in pregnancy is not.58 Instead of being part of the broader monitoring and evaluation process, disease surveillance is also regarded as an intervention in of itself1 comprising high-quality passive case detection, case notification and active case detection in which all febrile cases within proximity of the index case are tested and those positive for malaria infection are radically treated.58 In the districts where transmission is <1% \( P/PR_{2-10} \), malaria elimination is considered to be technically feasible3 6 presenting an opportunity to re-orient the subnational strategy here towards elimination and undertake an assessment of its operational feasibility.5 In contrast, the receptive risk map predicted that over 65% of the districts and population were under mesoendemic transmission (>10%–50% \( P/PR_{2-10} \)) with the rest exposed to hypoendemic transmission. Using this map, in the hypoendemic districts LLINs and indoor residual spraying would be better targeted to foci of risk and in preparation for possible epidemics as universal coverage with these interventions is unlikely to be the most cost-effective. In those areas of receptive mesoendemic transmission, which comprise 65% of the districts, universal coverage with LLINs should be the sustained ambition.1 56

The two divergent potential national
malaria strategies emanating from the two different
descriptions of risk highlight the danger of relying on
contemporary risk maps to make decisions that require
the understanding of the intrinsic transmission potential
of malaria.

Available maps that describe pre-RBM distribution of
risks are either expert opinion maps \(^{59}\) or climate-based
deterministic transmission suitability models \(^{60}\) and not
driven by empirical data. Even where empirical data may
be available, in countries with unstable malaria trans-
mission susceptible to seasonal and anomalous climatic
variations such as the recent drought in Somalia, the
risks measured at one time point may not be representative
of the possible peak risk levels for that point.

Therefore, spatially nationally representative data over
several years are required to capture these variations and
estimate the highest possible transmission. In this study,
\(PfPR\) data for Somalia that is available over four
consecutive years has provided a unique opportunity to
develop a novel way of selecting the maximum predicted
risks within the time series. The resolution of risk levels
at the malaria resource decision-making unit also
represents a product that is likely to be of more policy
relevance to the national programme managers
compared with the more common pixel-level predictions
of risks.

The study has some limitations. Although Somalia
represents one of the few African countries with ubiqu-
ous \(PfPR\) data in space and time there are gaps in the
distribution of the data and uncertainty of the predic-
tions are partly a function of these. The validation tests,
however, show overall good predictive model perfor-
mance with overall bias of MPE of \(<5\%\) and a slight
average overprediction of about MAPE of \(0.2\%\). The
coefficient of variation, which is the ratio of SD to mean
\(PfPR\) appeared similar for both the 2010 mean
and the maximum mean maps with uncertainty highest in
northern zones where data in space and time were fewest
(figure 2C,D). In selecting the maximal mean risk as
predicted to a location in any year over the 4-year period,
we make assumptions as if the modelled predictions
were part of 4-year repeat ‘observations’ of \(PfPR\) in
that location. The basis for this is that if the mean esti-
mate at any 1x1 km location from the full posterior
distribution of the space–time model is a robust esti-
mate to the given time and location, then selected
maximum mean estimate is equally so. We suggest that
this is a plausible assumption but further efforts need to
be invested in probabilistically selecting the maximum
mean predictions from the series of usually spatially and
temporally uneven data. Any uncertainties in the
approach used to select the maximum annual mean
\(PfPR\) are, however, unlikely to be the source of the
major differences in endemicities when compared to the
annual mean \(PfPR\) for 2010. The \(PfPR\) data used in
this analysis were assembled during a period when access
to control interventions had increased in Somalia with
donor funding support. Although coverage of main

\[\text{Figure 3} \quad \text{District (n=74) maps of Somalia classified by endemicity using a population-weighted: (A) posterior aggregate annual mean } PfPR_{2-10} \text{ (contemporary) prediction to 2010 and (B) the maximum annual mean } PfPR_{2-10} \text{ (receptive) predictions over the period 2007–2010. The blues lines show the location of the Juba (lower) and Shabelle (upper) Rivers.}\]
vector control interventions remain modest, it is likely that some of observed data were influenced by these interventions and in parts of the country true receptivity may even be higher than estimated. Land-use changes due to urbanisation, large-scale agricultural schemes and hydroelectric power dam projects also act as modifiers of transmission and in mapping malaria risk these factors must be adjusted for. Urbanisation, which has been shown to reduce malaria transmission, was included in the analysis of receptivity for Somalia. The maps of water bodies and vegetation used in the analysis will to some degree capture any aquatic or agricultural land-use changes. However, due to the long civil war and the lack of a functioning central government, such changes have been limited.

To compute a single estimate of risk for a relatively large area, such as districts in Somalia, will always obscure some of the heterogeneity in malaria distribution within that area regardless of the methods used. Any decision to do so is therefore a compromise between the practical applications of such a classification and the potential loss of precision in risk estimation. Approaches that directly adopt the heterogeneous properties of the prevalence data to make statistically robust single estimates of mean PPR to an administrative unit are computationally and methodologically intensive but have the advantage of estimating the area-level uncertainty classification through joint simulation. In this study, we have used simpler approaches to partly capture the within-district heterogeneity in malaria risk when classifying them into a single endemicity class by first assigning pixel-level population to an endemicity class before aggregating to the district. Future efforts should therefore be invested in assembling detailed historical data on parasite prevalence to allow for a better understanding of receptivity and equip national programmes with reliable estimates of receptivity that will enhance better decision-making.

Acknowledgements The community survey parasite prevalence data were provided by the Food and Agricultural Organization—Food Security and Nutritional Analysis Unit (FAO-FSNAU) for Somalia. The authors are grateful for the support and diligence of all the FSNAU field staff who did a great job.

Table 1 A summary of districts (N=74) and population in 2010 (N=8.4 million) in Somalia classified by malaria endemicity

<table>
<thead>
<tr>
<th>Endemicity classification based on the 2010 annual mean PPR&lt;sub&gt;2-10&lt;/sub&gt; (contemporary predictions)</th>
<th>Endemicity classification based on the maximum mean PPR&lt;sub&gt;2-10&lt;/sub&gt; (receptive) predictions over the period 2007–2010</th>
</tr>
</thead>
<tbody>
<tr>
<td>Population-weighted mean PPR&lt;sub&gt;2-10&lt;/sub&gt;</td>
<td>Number (%) of districts</td>
</tr>
<tr>
<td>Hypoendemic</td>
<td>17 (23)</td>
</tr>
<tr>
<td>&lt;1% PPR&lt;sub&gt;2-10&lt;/sub&gt;</td>
<td>43 (58)</td>
</tr>
<tr>
<td>1% to &lt;5% PPR&lt;sub&gt;2-10&lt;/sub&gt;</td>
<td>14 (19)</td>
</tr>
<tr>
<td>Mesoendemic (&gt;10%–50% PPR&lt;sub&gt;2-10&lt;/sub&gt;)</td>
<td>0 (0)</td>
</tr>
<tr>
<td>Hyperendemic and holoendemic (&gt;50% PPR&lt;sub&gt;2-10&lt;/sub&gt;)</td>
<td>0 (0)</td>
</tr>
</tbody>
</table>

District classifications of endemicity were computed from population-weighted posterior annual mean PPR<sub>2-10</sub> predicted to 2010 (contemporary) and the maximum annual mean PPR<sub>2-10</sub> (receptive) predictions over the period 2007–2010.
under difficult conditions. We are grateful to all household members who participated in the survey and agreed to the malaria testing. We thank Jacob Ouko and Damian Kinyoki for their help with assembly of ancillary data. The authors are also grateful for comments on earlier drafts of the manuscript from Dr Emelde Oiko. This paper is published with permission from the Director of KEMRI.

Contributors AMN was responsible for overall scientific management, study design, data cleaning, analysis, interpretation, drafting and production of the final manuscript. VAA was responsible for data cleaning, geo-coding, analysis and contributed to the final manuscript. APP was responsible for coding of the MBG framework, developed the supplementary information on model specifications and contributed to final manuscript. GM and MB contributed to the survey design, data assembly and cleaning and contributed to final manuscript. FY and JA contributed to survey design, interpretation of results and contributed to final manuscript. RWS provided scientific guidance and contributed to the analysis, interpretation and preparation of final manuscript. All authors read and approved the final manuscript.

Funding Cross-sectional survey was funded by the FAO-FSNAU and partners. AMN is supported by the Wellcome Trust as an Intermediate Research Fellow (095127). RWS is supported by the Wellcome Trust as Principal Research Fellow (079080) that also funded support to APP. Programmatic support for this study was also provided through a Wellcome Trust Major Overseas Programme grant to the KEMRI/Wellcome Trust Research Programme (092654).

Competing interests None.

Ethics approval Ethical approval was provided through permission by the Ministry of Health Somalia, Transitional Federal government of Somalia Republic, Ref: MOH/WC/KA/146/07, dated 2 February 2007. Informed verbal consent was sought from all participating households and individuals. Participants who were positive for Plasmodium falciparum infection was treated with the correct dosages of the nationally recommended antimalarials.

Provenance and peer review Not commissioned; externally peer reviewed.

Data sharing statement Data from this study are not in the public domain.

REFERENCES

Supplementary Information (SI) 1: Predictors of age-standardised *Plasmodium falciparum* parasite rate (*PfPR*$_{2-10}$)

**SI 1.1 Urbanisation:** A surface of urbanisation in Somalia derived from a 100×100 m spatial resolution population surface developed from a combination of census, satellite imagery and land cover data$^{1,2}$ was used to define each survey location as urban or rural. Somalia population map was derived by a combination of simple and semi-automated processes involving the re-distribution of district level population estimates to a finer spatial scale using land cover and land use datasets$^3$. A refined land cover map was prepared by combining the settlement data with Africover data$^3$ and other ancillary GIS data on roads, rivers from Vector Map Level Zero (VMAP 0)$^4$ and the Landsat satellite imagery. A total of 22 land cover classes were formed and population density within these classes was calculated with urban and rural extents derived from the GEOterrain consultancy$^5$. These calculated densities were then used as weightings to redistribute population by settlement and in land cover types that were unaccounted for by existing settlement size data (Figure SI 1A).

**SI 1.2 Enhanced vegetation Index (EVI):** EVI is an index of intensity of photosynthetic activity and a good proxy for rainfall ranges from 0 (no vegetation) to 1 (complete vegetation). Monthly EVI surfaces at 1×1 km spatial resolution derived from the global Moderate Resolution Imaging Spectro-radiometer (MODIS) satellite imagery for the period 2001-2010 were downloaded from the MODIS website$^6$ and were used to compute annual mean EVI. These monthly maps were used to compute annual mean EVI for the years 2007 to 2010 (Figure SI 1B).

**SI1.3 Precipitation:** Monthly mean precipitation (mm) raster surfaces at 1×1 km resolution were downloaded from the WorldClim website$^7$ and used as a proxy for rainfall. The Worldclim database was compiled using weather stations data collected world-wide for the period 1950-2000 and interpolated using spline methods$^8$. The monthly mean precipitation was used to compute annual mean precipitation surfaces for the years 2007 to 2010 (Figure SI 1C).

**SI1.4 Temperature suitability Index (TSI):** As metric for the effect of temperature on malaria transmission, TSI at a spatial resolution of 1×1 km$^9$ was used. TSI was constructed using monthly temperature time series$^7$ within a biological modelling framework to quantify the effect of ambient temperature on sporogony and vector survivorship and determine the suitability of an area to support transmission globally$^9$. On a scale of increasing transmission suitability, TSI ranges from 0 (unsuitable) to 1 (most suitable) (Figure SI 1.1.1D).
SI 1.5 Potential mosquito breeding sites: To define potential breeding sites the Global Lakes and Wetlands databases (GLWD) developed as a partnership between the World Wildlife Fund and the Center for Environmental Systems Research, University of Kassel, Germany\textsuperscript{10} was used. The GLWD data contained 12 wetland classifications but for Somalia the following wetland classes were retained for analysis: lake, rivers, floodplains, reservoirs and coastal wetlands. Intermittent water bodies and brackish and saline water were removed. Euclidean distances to these wetlands were computed in ArcGIS 10 (ESRI Inc. NY, USA) (Figure SI 1E).

SI 1.6 Statistical analysis of the predictors of PfPR\textsubscript{2-10}

The values of the underlying ecological, climatic and population covariates describe above were extracted to each survey location using ArcGIS 10 Spatial Analyst tool. Distance to potential breeding sites was log-transformed before analysis because of its high positive skew. The covariates were then included in total-sets analysis which is an automatic model selection process based on a generalized linear regression model and implemented using the \textit{bestglm} package in R\textsuperscript{11,12}. This approach selects the best combination of the covariates based on the value of the Bayesian Information Criteria (BIC) statistic\textsuperscript{13} which selects the lowest BIC as the best model fit. The model excluding TSI showed the lowest BIC and therefore the best fit (Table SI 1.1.1).

\textbf{Table SI 1.1.1} Summary of regression analysis of covariates

<table>
<thead>
<tr>
<th>Covariate</th>
<th>Odds Ratio (95% CI)</th>
<th>P-value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Urbanisation</td>
<td>-0.88 (-1.02 - -0.69)</td>
<td>&lt;0.001</td>
</tr>
<tr>
<td>EVI</td>
<td>0.81 (0.19-1.44)</td>
<td>0.011</td>
</tr>
<tr>
<td>Precipitation</td>
<td>0.003 (0.002-0.004)</td>
<td>&lt;0.001</td>
</tr>
<tr>
<td>TSI</td>
<td>-0.68 (-0.89- 0.01)</td>
<td>&lt;0.312</td>
</tr>
<tr>
<td>Log of distance to wetlands</td>
<td>0.27 (0.24 – 0.31)</td>
<td>&lt;0.001</td>
</tr>
</tbody>
</table>
Figure SI 1 Somalia maps of 1x1 km spatial resolution of: A) population distribution showing the location of urban centres (in red); B) annual mean enhanced vegetation index (EVI); C) annual mean precipitation (mm); D) temperature suitability index for Plasmodium falciparum transmission; E) distance (km) to nearest water body.
References


5. GeoTerraImage. [www.geoterraimage.com](http://www.geoterraimage.com), Accessed 15 February 2012.


Supplementary Information (SI) 2: Model-based Geostatistical Procedures

Below are the details of the MBG framework used to develop the contemporary and receptive malaria risk maps for Somalia. The MBG procedure was adopted from Gething et al 2011\(^1\) where further model details, especially on age-standardisation of parasite rate data, can also be found. The generic model code is available on Malaria Atlas Project \(P.\) falciparum Cartographic code link\(^2\) and has been adapted for Somalia. Model fitting was achieved using Markov chain Monte Carlo (MCMC)\(^3\)\(^,\)\(^4\)

**SI 2.1 The MBG presentation**

Each of the \(N_i\) individuals in sample \(i\) was assumed \(P.\) falciparum positive with probability \(\hat{k}_i P'(x_i, t_i)\), so the number positive \(N_i^+\) was distributed binomially:

\[
N_i^+ | N_i, P'(x_i, t_i) \overset{\text{ind}}{\sim} \text{Bin}(N_i, \hat{k}_i P'(x_i, t_i)) \quad \text{S2.1}
\]

The coefficient \(P'(x_i, t_i)\) was modelled as a Gaussian process. The factor \(\hat{k}_i\) converted \(P'(x_i, t_i)\) to the probability that individuals within the age range reported for study \(i\) were \(P.\) falciparum positive, and that the infection was detected, thereby accounting for the influence of age on the probability of detection\(^5\). The age-standardisation factor \(\hat{k}_i\) in each population was assumed drawn independently from a distribution \(D_k\) whose parameters were the lower \(A_{L,i}\) and upper \(A_{U,i}\) ages reported in study \(i\):

\[
\hat{k}_i | A_{U,i}, A_{L,i} \overset{\text{ind}}{\sim} D_k(A_{U,i}, A_{L,i}) \quad \text{S2.2}
\]

The form of \(D_k\) is described in Gething et al 2011\(^5\).

\(PR_{2-10}\) is the \(P.\) falciparum parasite rate for individuals between ages 2 (2.00) and 10 (9.99). Its value at an arbitrary location \(x\) and time \(t\) is the product of \(P'(x, t)\) and another age-standardisation factor, \(k_{2-10}\), distributed as \(D_k(2, 10)\):

\[
PR_{2-10}(x, t) = P'(x, t) k_{2-10}(x, t) \quad k_{2-10}(x, t) \overset{\text{ind}}{\sim} D_k(2, 10) \quad \text{S2.3}
\]

The factor \(k_{2-10}\) converted \(P'(x, t)\) to the probability that individuals between ages 2 and 10 at location \(x\) are \(P.\) falciparum positive. The age-standardisation factor \(\hat{k}\) of a survey is the product of the age-standardisation factor \(\bar{k}\) associated with the same place, time and age range and the sensitivity of the survey.

\[1\]
The coefficient $P'(x, t)$ at arbitrary location $x$ and time $t$ was modelled as the inverse-logit function applied to a random field $f$ evaluated at $(x, t)$, plus an unstructured (random) component $\epsilon(x, t)$.

$$P'(x, t) = \logit^{-1}(f(x, t) + \epsilon(x, t)) \quad S2.4$$

The components $\epsilon(x, t)$ were assumed independent and identically distributed for each location $x$ and time $t$ and a standard diffuse but proper prior with expectation 0.25 was assigned to their variance $V$.

$$\epsilon(x, t) | V \overset{iid}{\sim} N(0, V) \quad S2.5$$

$$\frac{1}{p} \sim \text{Gamma}(3, 12) \quad S2.6$$

The random field $f$ was modelled as a Gaussian process characterised by its mean and covariance functions:

$$f(x, t) | \beta, \tau, \phi_x, \phi_t, \lambda, \psi, \rho, \upsilon \sim \text{GP}(\beta, C) \quad S2.7$$

The mean function was defined as $\mu = \beta X$, where $X = 1, X_1(x), ..., X_n(x)$ was a vector consisting of a constant and $n = 20$ environmental covariates indexed by spatial location $x$, and $\beta = \beta_0, \beta_1, ..., \beta_n$ was a corresponding vector of regression coefficients. The covariance of the field was modelled using a version of the spatiotemporal covariance function recently recommended by Stein$^6$ (equation 2.12):

$$C(x_i, t_i; x_j, t_j) = \tau^2 \gamma(0) \left(\frac{\Delta x}{\Delta t}\right)^{\eta(\Delta t)} K_{\gamma}(\Delta x) \left[\frac{\Gamma(\gamma(\Delta t)}{\Gamma(\gamma(\Delta t) + 1)}\right],$$

$$\gamma(\Delta t) = \frac{1}{2p+2(1-p)}(1+v)e^{-\Delta t/2(1+v)}\cos(2\pi \Delta t),$$

$$K_{\gamma} \text{ is the modified Bessel function of the second kind of order } \gamma, \text{ and } \Gamma \text{ is the gamma function [7,8]}. \quad S2.8$$

Spatial distance between a pair of points $x_i$ and $x_j$ was computed as great-circle distance $D_{GC}(x_i, x_j)$ multiplied by a factor that depends on the angle of inclination $\theta(x_i, x_j)$ of the vector pointing from $x_i$ to $x_j$. $\theta$ was computed as if latitude and longitude were Euclidean coordinates (on a cylindrical projection) to allow for anisotropy:
\[ \Delta x = 2\sqrt{\gamma(\Delta t)} \frac{D_{GC}(x_i, x_j)\sqrt{1 - \psi^2 \cos^2(\theta(x_i, x_j) - \lambda)}}{\phi_x} \]  

When \( \Delta x = 0 \) (that is, for points at the same location but different times), the covariance function reduces to

\[ \rho + (1 - \rho) \left[ (1 - v)e^{-|\Delta t|/\phi_t} + v \cos(2\pi \Delta t) \right] \]  

As temporal separation increases, the covariance approaches a limiting sinusoid \( \tau^2[\rho + (1 - \rho)v \cos(2\pi \Delta t)] \) rather than zero. When \( \Delta t = 0 \), on the other hand (for points at different locations but the same time), it reduces to a standard exponential form with range parameter \( \phi_x \sqrt{2} \). Unlike standard sum-product models, this covariance function does not have problematic ridges along its axes\(^6\).

### SI 2.2 Prior Specification

The square root of the partial sill \( \tau \) and the spatial range parameter \( \phi_x \) were assigned skew-normal priors:

\[ \log \tau | \mu_\tau, V_\tau, \alpha_\tau \sim \text{Skew-Normal}(\mu_\tau, V_\tau, \alpha_\tau) \]  

\[ \log \phi_x | \mu_\phi, V_\phi, \alpha_\phi \sim \text{Skew-Normal}(\mu_\phi, V_\phi, \alpha_\phi) \]  

and their specification is described further below.

The standard “one-over-x” prior for the temporal scale parameter \( \phi_t \) resulted in collapse to zero, a common artefact when data do not contain strong information. A relatively vague but proper prior, which has an expectation of ten years, was used instead.

\[ \phi_t \sim \text{Exponential}(0, 1) \]  

A uniform prior was assigned to the direction of anisotropy parameter \( \lambda \) and to the square of the “eccentricity” parameter \( \psi \), which controls the amount of anisotropy,

\[ \lambda \sim \text{Uniform}(0, \pi) \]  

\[ \psi^2 \sim \text{Uniform}(0, 1) \]
a uniform prior was assigned to the temporal parameters governing the amplitude of the sinusoidal components of the mean.

Although standard priors such as the improper "flat" prior were assigned to most of the basic model parameters, subjective skew-normal priors were specified for the range and partial sill parameters \( \tau \) and \( \phi_x \).

### 2.3.2 Model implementation

Before the MCMC loop began, the marginal likelihood was approximated using standard Monte-Carlo integration for several values of \( P'(x_i, t_i) \). That is, values for the model parameters \( \alpha_y, b_y, c_i \) and \( s_i \) and the age distribution \( S_i \) were drawn from their posterior predictive distributions, then expression (S3.1) was evaluated to obtain \( k_y \) then the binomial probability was evaluated for several values of \( P'(x_i, t_i) \). The probabilities resulting from many such draws were averaged. Inside the MCMC loop, the marginal likelihood function for arbitrary values of \( P'(x_i, t_i) \) was evaluated by interpolation.

### 2.3.3 Age Correction Model

The survey population parameters \( S_0, S_i \) and \( \nu \) are independent of the relative PfPR parameters \( P', \alpha_y, c_y, b_y, s_y, \mu_A, \sigma \) and \( R \) given the data, so these two groups of parameters were inferred using separate MCMC algorithms.

In the MCMC for the age distribution parameters, the survey populations' age distributions \( S_i \) were updated using Gibbs steps. The concentration parameter \( \nu \) was updated using random-walk Metropolis steps. The typical age distribution \( S_0 \) was represented as a normalized sequence of gamma random variables, and these variables were updated one at a time using random-walk Metropolis steps.

In the MCMC for the relative PfPR parameters, the distributional parameters \( \mu_A, \sigma \) and \( R \) were updated jointly using the method of Haario, Saksman and Tamminen. The parameters \( P', \alpha_y, c_y, b_y \) and \( s_i \) were updated jointly for each population \( i \) using the same method.

### 2.3.4 Space-time Predictions

The posterior distribution consisted of \( \{ \theta(l); l = 1, ..., m \} \) samples from the posterior of the parameter set \( \theta = \{ \beta, \tau, \phi_x, \phi_t, \lambda, \psi, \rho, k, V \} \) and a corresponding \( \{ f(x_i, t_i(l)); l = 1, ..., m \} \) samples from the posterior of the space-time random field at each of the \( n \) data locations \( \{(x_i, t_i); i = 1, ..., n \} \).

For every \( l \)th sample, the conditional distribution of the annual mean of the space-time random field

\[ f(x_i, t_i(l)) \]
was predicted at each prediction location \(x_j\) on the nodes of a regular 1x1 km grid within the spatial limits of stable \(P. falciparum\) transmission\(^{14}\). The distribution of the annual mean \(f(x_j)_{(l)}\) for each of the 12 prediction locations \(\{t = 2007_{Jan}, ..., 2007_{Dec}\}\) was modelled as the joint multivariate normal distribution of the 12 predicted monthly values \(\{t = 2007_{Jan}, ..., 2007_{Dec}\}\) at each of the 12 predicted locations \(x_j\) and at each of the \(n\) data locations respectively, \(C_{DP}\) and \(C_{DD}\) were the data-to-prediction and data-to-data covariance matrices respectively, and \(p(x, t)\) was the vector of \(n\) data values. The 12 x 12 variance-covariance matrix \(\hat{\Sigma}(x_j)_{(l)}\) was computed using:

The value of the 1\(^{st}\) sample of \(\hat{\Sigma}(x_j)_{(l)}\), the variance of the unstructured component \(\epsilon(x, t)\), was then added to the diagonal of the matrix \(\hat{\Sigma}(x_j)_{(l)}\) and 1000 draws were made randomly from the distribution specified in equation S2.34. The draws represented samples from the posterior distribution of \(f(x_j)\) and were subject to an inverse logit transform and then multiplied by the 1\(^{st}\) sample of the age-standardisation parameter \(\hat{k}_{2-10}(l)\) to form the 1\(^{st}\) sample from the posterior distribution of the predicted mean annual 2010 \(P.f\) endemicity surface at location \(x_j\):

This procedure was repeated for every \(P.f\) sample to form the set \(\{P.f_{2-10}(x_j)_{(l)}; l = 1, ..., m\}\) of \(m\) samples for each prediction location. The point estimate of \(P.f\) endemicity at each location was defined as the mean of this set, whilst the probability of membership to each class was computed as the proportion of these samples falling within each class definition.

References: