Malaria Stratification, Climate, and Epidemic Early Warning in Eritrea


Abstract. Eritrea has a successful malaria control program, but it is still susceptible to devastating malaria epidemics. Monthly data on clinical malaria cases from 242 health facilities in 58 subzobas (districts) of Eritrea from 1996 to 2003 were used in a novel stratification process using principal component analysis and nonhierarchical clustering to define five areas with distinct malaria intensity and seasonality patterns, to guide future interventions and development of an epidemic early warning system. Relationships between monthly clinical malaria incidence by subzoba and monthly climate data from several sources, and with seasonal climate forecasts, were investigated. Remotely sensed climate data were averaged over the same subzoba geographic administrative units as the malaria cases. Although correlation was good between malaria anomalies and actual rainfall from ground stations (lagged by 2 months), the stations did not have sufficiently even coverage to be widely useful. Satellite derived rainfall from the Climate Prediction Center Merged Analysis of Precipitation was correlated with malaria incidence anomalies, with a lead time of 2–3 months. NDVI anomalies were highly correlated with malaria incidence anomalies, particularly in the semi-arid north of the country and along the northern Red Sea coast, which is a highly epidemic-prone area. Eritrea has 2 distinct rainy seasons in different parts of the country. The seasonal forecasting skill from Global Circulation Models for the June/July/August season was low except for the Eastern border. For the coastal October/November/December season, forecasting skill was good only during the 1997–1998 El Niño event. For epidemic control, shorter-range warning based on remotely sensed rainfall estimates and an enhanced epidemic early-detection system based on data derived for this study are needed.

INTRODUCTION

Malaria control in Eritrea. Eritrea is a malaria epidemic-prone country in the horn of Africa. It is divided into three main topographical regions (the western lowlands, central highlands, and eastern coastal lowlands) with different rainfall seasonality patterns. There are six administrative regions (zobas) with 58 districts (subzobas) (Figure 1) and ~1,500 villages. The estimated population is 3.5 million.

A survey in 2002 estimated malaria prevalence at 2%, with some villages having up to 30% prevalence.1 Eritrea has had a successful malaria control program over the last 8 years since a devastating epidemic following the heavy rains associated with the 1997–1998 El Niño event.2 An intensive mixture of interventions has been applied,3 including widespread coverage with free insecticide-impregnated mosquito nets to all malaria risk areas,3 indoor residual spraying with DDT and malathion in some areas, provision of prompt treatment by village health agents, source reduction, and chemical or biologic larval control.4

To improve epidemic control in climate sensitive regions, the World Health Organization (WHO) has proposed a framework for the development of integrated malaria early warning systems (MEWS).5,6 based on vulnerability monitoring, seasonal climate forecasting, environmental and meteorological monitoring, and epidemiologic surveillance. Here we explore the potential for using these indicators in the development of a malaria early warning system in Eritrea.

Vulnerability assessment. Malaria risk maps derived from climate-driven malaria transmission models have been used to indicate areas of stable and unstable malaria7 and thereby indicate areas vulnerable to climate related epidemics. Early attempts to create a malaria risk map based on the modeled relationship between village-based prevalence data in Eritrea and a wide range of environmental and climatic predictors was found to be unsatisfactory.8,9 In this paper, we report on a new methodology based on routine incidence data, designed to indicate the intensity and seasonality of transmission at the district level.

Climate and environmental monitoring. In most of Eritrea, the majority of the rainfall occurs during July/August. In the eastern portion of the country, where a steep escarpment degrades toward the Red Sea, rainfall is considerably lower and peaks in October/January. Both ground station measures and satellite-derived estimates of rainfall are potential proxies for use by malaria control staff. Comparisons of the two sources in neighboring Ethiopia indicated that certain satellite-derived rainfall estimates (RFE) products may provide sufficiently accurate rainfall information in the region to be used routinely by malaria control.10

There is a strong relationship between normalized difference vegetation index (NDVI), a measure of environmental “greenness” and rainfall both spatially and temporally over East Africa, as well a demonstrated relationship between NDVI, rainfall, and the El Niño–Southern Oscillation (ENSO).11–13

Seasonal climate forecasts. For certain geographic areas and seasons, climate forecasts offer a degree of predictability of climate fluctuations at a seasonal (i.e., ~3 month) lead time.14 Predictability is highly dependent on the extent to which regional climate for a given season is determined by sea surface temperature (SST) patterns of global oceans; in particular SSTs in the tropical Pacific (i.e., the ENSO). There is
some indication that the summer rainfall in Eritrea is related to ENSO and therefore potentially predictable.

Significant correlations between malaria incidence and observed SSTs (including time-lagged SSTs) have been observed. Building on these observations, seasonal climate forecasts produced from the real-time operational multimodel ensemble DEMETER forecast systems have started to contribute to the development of malaria early warning in Botswana.

Surveillance data. In many African countries, malaria data are collected at health facilities but often not looked at until they are summarized in the form of routine annual reports. Recently, many countries including Eritrea have attempted to improve the collection and use of health facility data, by updating health management information systems (HMIS). The aim is to improve estimates of the burden of disease at the national level while enhancing quality of care, accountability and decision making at the local level. Weekly surveillance data are particularly valuable for the detection of unusual situations, possibly an epidemic, which may require a prompt response.

Unfortunately, to date, there is very little information that feeds back to the district level for improved decision making, thus undermining the incentives for improved reporting. Furthermore, biases in data reported that result from quality of service, drug availability, physical and social access, and health-seeking behavior feed the widespread distrust of passive surveillance data collection as a tool in routine malaria decision making. Furthermore, for clinically diagnosed malaria cases (as opposed to those confirmed by laboratory diagnosis), there is often considerable overestimation of the case numbers as well as delay in appropriate treatment of nonmalaria fever patients. Despite these limitations, many countries have no alternative but to make best use of the information available through passive health surveillance systems. Geographical and seasonal patterns of malaria transmission are usually clearly evident from such data.

In Botswana, where the HMIS is not only functioning well but where all suspected malaria cases are routinely laboratory-confirmed, a strong correlation is observed between clinical and confirmed cases. Thus in epidemic-prone areas, routine surveillance of clinical cases may be useful for determining past thresholds and for the early detection of epidemics. A recent study from Kenya has shown that the quality of routinely collected health surveillance data from the HMIS can be dramatically improved through statistical methodologies that impute missing data using geospatial techniques.

METHODS

Surveillance data. Eritrea’s National Health Management Information System (NHMIS) was revised in 1998 with support from USAID and John Snow, Inc. (Boston, MA). This revision involved computerization of a previously existing paper system for monthly reporting from each health facility, standardized coding of case definitions, coding of health facilities and administrative areas, reduction in numbers of age groups reported, and establishment of reporting timelines. There were no major changes in malaria case definitions or unit of reporting (health facility).

For this study, the numbers of outpatient clinical malaria cases were extracted from the NHMIS Access database by month and health facility for the years 1998–2003. Data for 1996–1997 were obtained from similar Malaria Control Program records stored in Excel by health facility and month.

The NHMIS listed 325 health facilities in 1998. We restricted this to 242 sites by exclusion of private doctors, worksite clinics, national referral hospitals in Asmara, and exclusion of 3 nonfunctioning health facilities. The number of health facilities in the system had increased to 383 by 2003, of
which many were new small private clinics as well as one new National Referral Hospital. In this analysis, we use only data for the original 242 selected facilities. Administrative (subzoba and zoba) boundary files were obtained from the National Statistics and Evaluation office, Asmara, Eritrea, and visualized in ArcView 3.3. Malaria cases were summed over both age groups (under and over 5 years) by subzoba and month (1–9 facilities per subzoba, average 4.3). Malaria incidence per 1000 persons per month by subzoba was estimated using the year 2000 population estimates in the NHMIS.

**Imputation of missing values.** If complete, the 8 years (1996–2003) of monthly incidence data for 58 subzobas would have records for 5568 subzoba-months. There were 191 missing data points (subzoba-months), representing 3.4% of the total matrix, impeding the use of principal component analysis (PCA), which is dependent on complete data sets. Missing values were imputed through a sequence of multivariate step-wise regressions that estimated the variables with the smallest number of missing values first and proceeding until all missing values were estimated.26–28

**Stratification map using NHMIS data.** We used the ADDATI software developed for exploratory data analysis, including principal components and clustering analysis.29 Monthly malaria incidence estimates, computed as arithmetic average of values in that month during the 8-year period for each of the 58 subzobas, were used for the analysis. The processing sequence in ADDATI includes i) PCA followed by ii) nonhierarchical clustering adapted from Diday30 on the most significant principal components. PCA is a technique for simplifying a data set by reducing multidimensional data sets to lower dimensions in the data.31

The reduction in dimensions of the data set is necessary to run and accelerate the clustering analysis, which is a classification of objects into different groups such that each group shares a common trait. A total of 12 principal components were generated from the monthly average incidence data. Using eigenvalue and proportion of variance explained by each component criteria, we retained, for further analysis, the 6 first principal components, which explained 99.37% of the cumulative variance in the data.

The iterative nonhierarchical clustering method groups the subzobas into a specific user-defined number of clusters. ADDATI generates the cluster centers and allocates the subzobas to the clusters based on proximity to initial centers (“seeds”).29 The data were initially clustered into 2, 3, 4, 5, 6, 8, and 12 clusters for analysis. Clusters were visualized in ESRI ArcView 3.3.

**Climate datasets.** Although there is a variety of rainfall datasets for Eritrea, few are suitable for analyses of long-term trends and variability. After investigating a range of datasets, we concluded that 3 were useful for assessing climate variability in Eritrea during the study period (Table 1): Rain gauge data from meteorological stations (locations shown in Figure 2); merged gauge and satellite rainfall estimates (CMAP); and the vegetation index (NDVI) from NOAA AVHRR computed as a monthly maximum value composite. A fourth potential dataset (from the Climate Research Unit at the University of East Anglia, U.K.) consisted of interpolated meteorological station data but was only available up until 1998. Using the first 3 sources, the monthly anomalies (departure from long-term mean) of climate environmental variables were calculated. The monthly anomalies, both for climate environmental variables and for incidence data, are used for the analysis. Anomalies are used instead of recorded values because both climate variables and malaria incidences have a similar seasonal pattern. Statistical analysis of recorded values would indicate a very high correlation between climate and malaria incidences only due to the fact that the temporal evolution of climate variables and malaria incidences follow the same seasonality. To eliminate the seasonal cycle effect and study only the impact of variations in climate

<table>
<thead>
<tr>
<th>Table 1</th>
<th>Types and sources of observed or forecast climate data for Eritrea</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Climate data</strong></td>
<td><strong>Coverage</strong></td>
</tr>
<tr>
<td>Climate Prediction Center Merged Analysis of Precipitation (CMAP), version 0407</td>
<td>Complete coverage</td>
</tr>
<tr>
<td>Normalized Difference Vegetation Index (NDVI), version “c” product</td>
<td>Complete coverage</td>
</tr>
<tr>
<td>Interpolated rainfall gauge data</td>
<td>Complete coverage</td>
</tr>
<tr>
<td>SSTs</td>
<td>Global oceans</td>
</tr>
<tr>
<td>ECHAM 4.5</td>
<td>Global</td>
</tr>
</tbody>
</table>
with variations in malaria incidence, the long-term mean is subtracted from the recorded data.

The satellite-derived climate variables from the IRI data library were averaged over the same administrative boundaries as the incidence data, using subzoba boundaries from the National Statistics and Evaluation Office, Eritrea. Although it would have been preferable to use smaller boundaries, such as health facility catchment areas, these are not yet available.

**Seasonal climate forecasts.** We used rainfall output from a global circulation model (GCM) ECHAM 45 (Table 1) to predict September NDVI using forecasted June/July/August (JJA) rainfall and to predict January NDVI using forecasted October/November/December (OND) rainfall.

**RESULTS**

**Reporting rate.** The average reporting rate by month at the health facility level over the entire 8-year period was 76.9%. Of the 242 health facilities, the majority (155 [64.0%]) provided > 80% of required reports, 38 facilities (15.7%) provided 60–80% of reports, 20 (8.3%) provided 40–60% of reports, 19 (7.9%) provided 20–40% of reports, and only 10 (4.1%) provided < 20% of the required reports. In some cases, this is because the facility was not functioning for part of the time period, but detailed information on this is not available.

Estimates of reporting completeness by year were as follows:

1996: 52.1% of reports received, 59.9% of facilities reporting;
1997: 71.2% of reports received, 74.4% of facilities reporting;
1998: 78.1% of reports received, 84.7% of facilities reporting;
1999: 83.3% of reports received, 87.2% of facilities reporting;
2000: 77.4% of reports received, 92.1% of facilities reporting;
2001: 84.8% of reports received, 93.8% of facilities reporting;
2002: 86.4% of reports received, 92.6% of facilities reporting;
2003: 89.5% of reports received, 93.0% of facilities reporting.

As mentioned in the Methods section, there were a potential 5568 data points (subzoba-months) from the 58 subzobas and 96 months of study. There were 191 missing data points where no reports were received in a particular subzoba, representing 3.4% of the total matrix.

**Vulnerability mapping with surveillance data.** The 5-cluster map (Figure 3) was found to be most representative of what is expected in the field and is consistent with expert opinion. With 5 clusters we were able to capture differences in both intensity and the seasonal dynamics of malaria incidence. This is particularly important in Eritrea as the coastal region is dominated by OND rains as opposed to July/August/September (JAS) rains for the rest of the country. However, this analysis was able to differentiate distinct risk regions (groups of subzobas) within the Red Sea coastal region of Eritrea as well as different intensities of transmission within the areas susceptible to the OND rainfall season. One can argue that the difference in intensity (or malaria incidences) observed between the clusters could be related to the number of health facilities per subzoba, which could affect the coverage and hence the number of malaria cases reported from a subzobas. We analyzed the number of health facilities and number of reports received per year to verify whether this hypothesis could influence the intensity. Tables 2 and 3 show that there is indeed a difference in number of health facilities between clusters, but the average number of reports received per cluster and per year is very similar and does not influence the number of incident cases recorded (e.g., Cluster 2 has the highest number of health facilities and still is classified by a low incidence rate).

In addition to clarifying the seasonal pattern and intensity of transmission in different areas, the stratification map produced is useful for targeting and timing interventions, such as indoor residual spraying to the highest-risk areas.

**Climate/environmental predictors of malaria incidence.** The relationship of monthly malaria incidence anomalies to anomalies in climate/environmental data (and each anomaly value squared) for each subzoba was assessed using Spearman and Pearson rank correlations.

There is a good correlation between rainfall and malaria anomalies in some of the subzobas where rain gauge station data are available (22 out of 58 subzobas) with a lead time of 2 months (Figure 4). However, there is no information for large tracts of the country, suggesting that, although this data source may be used in specific districts, it is not suitable for national coverage.

**Satellite-derived estimates of rainfall.** Rainfall estimate anomalies were also moderately correlated with malaria incidence anomalies with an average lead time of 2–3 months. Although the strength of the relationship was less than that of the gauge data, this data source had the advantage of complete coverage of the country.

**Satellite-derived estimates of vegetation “greenness.”** NDVI anomalies were correlated with malaria incidence anomalies, particularly in the semi-arid north of the country and along the coast in Northern Red Sea (Figure 5). Areas of high correlation are largely those identified as “sparse vegetation” from the Africover land cover map produced by the Food and Agriculture Organization of the United Nations. An advantage of NDVI over rainfall is that it is indicative of soil moisture and is therefore capable of picking out areas where rainfall runoff may have a significant impact on malaria transmission.
Seasonal climate forecasts. We retrospectively forecast June/August rainfall data using a general circulation model forced with May SST anomalies and correlated the results with September NDVI over the Greater Horn of Africa including Eritrea. This forecast would normally be available toward the end of June. Accuracy is good over the northern part and eastern coast of the Greater Horn of Africa with correlation ranging from 0.4 to 0.7 (not shown). However,

Table 2

<table>
<thead>
<tr>
<th>No. of health facilities</th>
<th>1. Very low incidence year around (Oct peak)</th>
<th>2. Low incidence year around (Oct peak)</th>
<th>3. Moderate incidence West (Oct peak)</th>
<th>4. High incidence (Oct peak)</th>
<th>5. Moderate incidence East (Jan peak)</th>
</tr>
</thead>
<tbody>
<tr>
<td>1996</td>
<td>42</td>
<td>103</td>
<td>38</td>
<td>24</td>
<td>35</td>
</tr>
<tr>
<td>1997</td>
<td>339</td>
<td>710</td>
<td>140</td>
<td>222</td>
<td>102</td>
</tr>
<tr>
<td>1998</td>
<td>373</td>
<td>942</td>
<td>268</td>
<td>248</td>
<td>236</td>
</tr>
<tr>
<td>1999</td>
<td>395</td>
<td>1012</td>
<td>334</td>
<td>239</td>
<td>289</td>
</tr>
<tr>
<td>2000</td>
<td>406</td>
<td>1089</td>
<td>352</td>
<td>244</td>
<td>329</td>
</tr>
<tr>
<td>2001</td>
<td>417</td>
<td>1037</td>
<td>284</td>
<td>219</td>
<td>288</td>
</tr>
<tr>
<td>2002</td>
<td>417</td>
<td>1106</td>
<td>335</td>
<td>260</td>
<td>326</td>
</tr>
<tr>
<td>2003</td>
<td>439</td>
<td>1108</td>
<td>405</td>
<td>265</td>
<td>381</td>
</tr>
</tbody>
</table>
over Eritrea the area of predictability was limited to the eastern border, and accuracy falls further from the sea (northwestern region).

Results show that good correlations (with coefficients between 0.6 and 0.8) are obtained in Eritrea for the period OND. However, correlation values dropped dramatically when the year 1997 is not taken into account, implying that good predictability in terms of rainfall could be achieved only during a very strong El Niño event.

**DISCUSSION**

Malaria stratification is a classification of areas according to the risk of malaria. It is a way to set priorities and target prevention efforts to the areas where they are most needed. It can highlight areas where the control program needs extra effort and helps to make the best use of resources. Changes in stratification over time can also help to indicate the areas at most risk of epidemics.

Because control activities are mostly organized by administrative units, such as subzoba, classification of these units is more useful to the program than a strict ecological stratification, where different environmental strata may overlap more than one subzoba. In future, risk stratification can be extended to smaller units within subzobas if desired. The availability from the International Research Institute for Climate and Society (IRI) of remotely sensed data averaged over the subzoba administrative unit, together with malaria cases in the same geographic units, was highly beneficial to the stratification process.

Our results show the ADDATI software, freely available for agroclimatic applications, can be a useful tool for malaria stratification. Although the program was intended for classification of satellite-derived time series land cover data into areas of similar spatiotemporal patterns in vegetation development, it is relatively easy to use for clustering districts with similar malaria incidence and seasonality.

Using only health-facility clinical malaria incidence data aggregated over administrative divisions we identified 5 clusters (groups of administrative units) of malaria risk based on both intensity and dynamics. The cluster classification provides guidance for targeting extra malaria control methods (such as indoor residual spraying) to the highest-incidence areas. It also clarifies the seasonality differences between regions and indicates optimum timing of interventions.

Stratification of malarious areas by average parasite incidence has been applied previously.\textsuperscript{33,34} Our current methodology is an improvement of that process because it identifies not only geographic areas with similar malaria intensity but also areas with similar temporal dynamics. It also relies only on readily available surveillance data on malaria incidence to stratify in areas with significant seasonal variation in incidence. Given that many malarious countries have established or improved their HMIS as repositories for surveillance and monitoring data, development of a stratification method that depends on such data has great potential.

**TABLE 3**

Average of reports per health facilities per cluster and per year

<table>
<thead>
<tr>
<th>No. of reports per facility, by year</th>
<th>1. Very low incidence year around (Oct peak)</th>
<th>2. Low incidence year around (Oct peak)</th>
<th>3. Moderate incidence West (Oct peak)</th>
<th>4. High incidence (Oct peak)</th>
<th>5. Moderate incidence East (Jan peak)</th>
</tr>
</thead>
<tbody>
<tr>
<td>1996</td>
<td>8.07</td>
<td>6.89</td>
<td>3.68</td>
<td>9.25</td>
<td>2.91</td>
</tr>
<tr>
<td>1997</td>
<td>8.88</td>
<td>9.15</td>
<td>7.05</td>
<td>10.33</td>
<td>6.74</td>
</tr>
<tr>
<td>2001</td>
<td>10.40</td>
<td>10.74</td>
<td>8.82</td>
<td>10.83</td>
<td>9.31</td>
</tr>
<tr>
<td>2003</td>
<td>10.45</td>
<td>10.76</td>
<td>10.66</td>
<td>11.04</td>
<td>10.89</td>
</tr>
</tbody>
</table>

**FIGURE 4.** Variance in malaria incidence predicted by rainfall from meteorological stations lagged by 2 months. The coefficient of determination ($R^2$) values indicate the proportion of variability in the malaria anomalies that is accounted for by the rainfall.

**FIGURE 5.** Correlation between monthly NDVI and monthly malaria incidence. The coefficient of determination ($R^2$) values indicate the proportion of variability in the malaria anomalies that is accounted for by the NDVI.
Although the problem of underreporting of disease incidence from underserved areas is always present, this may be compensated by the availability of much larger and more detailed datasets than are usually available from cross-sectional prevalence surveys. Problems of inconsistent reporting can be minimized by restricting analysis only to health facilities with reliable reporting, ideally of confirmed malaria cases.

The dataset used for stratification also provides necessary outcome data for assessing the effectiveness of different control methods used in the country over the last 8 years. However, because malaria anomalies and rainfall/NDVI are clearly associated in Eritrea, it is imperative that an analysis of the impact of interventions take into account climate/environment variability.

The routine data is also a rich source of past information at the subzoba and health facility level for generating epidemic threshold values for early detection of abnormal numbers of cases. This task is currently in progress.

The climate/environmental variables used in this report can be routinely monitored, and in fact some are already used within Eritrea for food security and locust monitoring. These environmental variables have the potential to be partially predicted using seasonal climate forecasts, although this predictability is unlikely to be as good as other parts of east and southern Africa.25,36

There is a known trade-off between timing and accuracy in any such warning system. In our example:

a. Health surveillance using epidemic thresholds detects the early phase of epidemics and provides a couple of weeks’ warning of an epidemic peak.

b. Environmental monitoring of NDVI provides concurrent prediction of malaria anomalies.

c. Rainfall from gauge data provides prediction of malaria anomalies with a lead time of 2–3 months for some subzobas with meteorological stations.

d. Rainfall from merged gauge and satellite data (CMAP) predicts September or January NDVI (when peak malaria occurs) with 2–3 months’ lead time.

e. General circulation models predict rainfall with a lead time of 2 months and NDVI with a lead time of 4 months, but with reasonable accuracy only in El Niño years for most parts of Eritrea.

Stratification and definition of the relationship between climate and malaria are significant steps toward development of an early warning system. Once a picture of the “expected” number of monthly (or, perhaps in the future, weekly) malaria cases in a subzoba or cluster has been built up, together with the expected climate variables in the same areas, all of the necessary components of an early warning system are in place.

Received August 21, 2006. Accepted for publication April 16, 2007.

Acknowledgments: We thank the Minister of Health for Eritrea, Mr. Saleh Meky, the Director General for Health, Mr. Berhanu Tensae, and the head of Disease Prevention and Control, Dr. Goitom Mebratu, for their support. We are very grateful to Kiros Sereke, Afwerki Araia, Andemariam WeldeMicael, Solomon Neguse, Meles GhebreYesus, Asrat GhebreLui, Mehari Zerom, Asemelash Geggziabher, Helen Fekadu, Solomon Mengistu, Yohannes Bein, David Sintasath, and Dr. Josephat Shililu of the NMCPI and Ezra Kidane, Amanuel Kifle, and Samuel Goitom from the NHMIS. We thank Matthew Lynch of USAID, Lisa Lukang of CIESIN, and Tsegay Wolde-Georgis of IRI for input and support, and also the National Statistics and Evaluation Office, Asmara, Eritrea, and the Water Resources Department, Ministry of Lands, Water and Environment, for provision of GIS data.

Financial support: This study was funded by the USAID Environmental Health Project, Contract HRN-1-00-99-00011-00 sponsored by Office of Health, Infectious Disease and Nutrition, Bureau for Global Health, USAID, Washington, D.C. 20523. The following organizations supported the malaria control program of Eritrea during the period of study: The World Bank HAMSET project; USAID; WHO; Italian Cooperation; UNICEF.


