Spatial clustering of malaria and associated risk factors during an epidemic in a highland area of western Kenya

Simon Brooker [1], Sia^n Clarke [1], Joseph Kiambo Njagi [2], Sarah Polack [1], Benbolt Mugo [3], Benson Estambale [4], Eric Muchiri [3], Pascal Magnussen [5] and Jonathan Cox [1]

[1] Department of Infectious and Tropical Diseases, London School of Hygiene and Tropical Medicine, London, UK
[2] Division of Malaria Control, Ministry of Health, Nairobi, Kenya
[3] Division of Vector Borne Disease, Ministry of Health, Nairobi, Kenya
[4] Department of Medical Microbiology, College of Health Sciences, Nairobi, Kenya
[5] Danish Bilharziasis Laboratory, Charlottenlund, Denmark

Summary

The epidemiology of malaria over small areas remains poorly understood, and this is particularly true for malaria during epidemics in highland areas of Africa, where transmission intensity is low and characterized by acute within and between year variations. We report an analysis of the spatial distribution of clinical malaria during an epidemic and investigate putative risk factors. Active case surveillance was undertaken in three schools in Nandi District, Western Kenya for 10 weeks during a malaria outbreak in May–July 2002. Household surveys of cases and age-matched controls were conducted to collect information on household construction, exposure factors and socio-economic status. Household geographical location and altitude were determined using a hand-held geographical positioning system and landcover types were determined using high spatial resolution satellite sensor data. Among 129 cases identified during the surveillance, which were matched to 155 controls, we identified significant spatial clusters of malaria cases as determined using the spatial scan statistic. Conditional multiple logistic regression analysis showed that the risk of malaria was higher in children who were underweight, who lived at lower altitudes, and who lived in households where drugs were not kept at home.

keywords

malaria, household distribution, risk factors, spatial clustering, highland malaria, Kenya

Introduction

In common with most infectious diseases, malaria distribution within a geographical area is heterogeneous and can vary greatly between villages and households (Greenwood 1989; Gamage-Mendis *et al.* 1991; Carter *et al.* 2000).

These patterns of malaria reflect a composite of heterogeneities in vector distribution, human–vector contact and human host factors (Greenwood 1989). Identified risk factors for malaria include distance to known mosquito breeding sites, household construction, household crowding and personal protection measures against mosquito biting (Gamage-Mendis *et al.* 1991; Trape *et al.* 1992; Adiamah *et al.* 1993; Koram *et al.* 1995; Thompson *et al.* 1997; Snow *et al.* 1998; van der Hoek *et al.* 1998; Ghebreyesus *et al.* 1999, 2000; Thomas & Lindsay 2000; Clarke *et al.* 2002). In turn, these factors are proximally influenced by differences in environmental landscape (Rejmankova *et al.* 1995; Thomas & Lindsay 2000) and socio-economic status (Koram *et al.* 1995; Clarke 2001).

Disentangling the influence of these different factors is frequently hindered by a lack of detailed data relating to a full range of contextual factors together, and few studies have been performed which include both household and environmental landscape factors. This is particularly true for epidemic-prone areas in highland locations, despite the increasing interest in the epidemiology of highland malaria (Lindblade *et al.* 1999; Shanks *et al.* 2000; Hay *et al.* 2002). In these areas, transmission is unstable and the risk of disease tends to be equal across all age groups as populations have little or no immunity against *Plasmodium* spp. It remains unclear, however, whether the risk of

malaria during an epidemic is equal amongst all households or the degree to which risk is spatially clustered. The investigation of infectious disease clustering is receiving renewed interest, not least because of advances in geographical information systems (GIS) and spatial statistics, which allow for the quantification of the degree of clustering of infections. Such approaches have been used to investigate the spatial clustering of dengue (Morrison et al.

Tropical Medicine and International Health
volume 9 no 7 pp 757–766 july 2004

1998), LaCrosse encephalitis (Kitron et al. 1997) and sleeping sickness (Fe`vre et al. 2001), but their application to malaria has been limited (Schellenberg et al. 1998; Chadee & Kitron 1999; Ghebreyesus et al. 2003). An improved understanding of the spatial clustering of malaria and its determinants in highland areas may provide useful insights into local epidemic control (Carter et al. 2000).

In May–July 2002, Western Kenya experienced a number of malaria epidemics, following heavy rains earlier during May. This provided a unique opportunity to investigate the epidemiology of clinical malaria within epidemic-prone areas during an epidemic. In this study, we have mapped and analysed the household distribution of clinical malaria based on active case detection among school children in Kapkangani Location, a rural epidemic-prone area of the highlands in western Kenya. The objectives of our study are to evaluate spatial clustering of clinical malaria and to investigate putative risk factors.

Methods

Study area

The present study was conducted in Kapkangani Location (0 14 ¢ –0 22 ¢ N, 34 54 ¢ –35 3 ¢ E), a rural part of Nandi District, western Kenya. Investigations were undertaken in three schools: Kiborgok in Kiborgok sub-location, and Koibem and Kabaskei in Chepkomia sub-location

(Figure 1). Kiborgok sub-location is situated on the Nandi Escarpment where elevation ranges from 1650 to 2050 m. Chepkomia sub-location is lower down the escarpment to the west of South Nandi Forest, running along the River Yala at elevations ranging from 1700 to 1900 m. Rainfall (annual average of 2428 mm) is seasonally bimodal, with the long rains occurring from March to May and the short rains from October to December. Average annual minimum and maximum temperatures are 12.2 and 23.6 C, respectively (unpublished data from the adjacent Kaimosi tea estate meteorological station).

Malaria transmission is acutely seasonal with peaks occurring 2–3 months after the peak rains in April–May, although the extent of the malaria burden varies considerably from year to year. This temporal pattern is similar to other areas of western Kenya (Hay *et al.* 2002). Early entomological research indicated that throughout the district, anthropophilic *Anopheles gambiae* s.l. (98%) was the principal malaria vector in the area, with *An. funestus* (2%) playing a minor role (Roberts 1964). Recent investigations report similar findings (Shililu *et al.* 1998; Minakawa *et al.* 2002a).

The population of the area consists of indigenous Kapsigi people and numerous Luhya settlers who have moved from the lowland areas of western Kenya and have purchased land during the past 30 years. The economy is primarily rural subsistence agriculture, with some families growing tea as a cash crop. Other economic opportunities include casual labour on local tea estates. This population is serviced by Kapkangani Government Health Centre, which has a catchment area of about 20 000 people and a catchment radius of about 20 km.

Figure 1
Map of study area, Kapkangani Location, Nandi District in western Kenya. The locations of sampled households, schools, health centre, land classification types, River Yala and elevation contours are shown. Kiborogok school is positioned north of the health centre, Koibem is south of the health centre and Kabaskei is the most southern school.

Selection of schools and methods of surveillance

Schools within the catchment area of the health centre were identified and classified according to ethnic mix. In order to

minimize the effect of differential immunity, host genetic heterogeneity and/or travel to malaria-endemic areas on the risk of malaria, only schools where almost all pupils were of indigenous Kapsigi descent (including Nandi and Kapchukin peoples) were eligible for inclusion in the study. The three schools closest to the health centre fulfilling these ethnicity criteria were selected. A series of meetings were held with teachers, parents and community leaders to explain the purpose and methodology of the study that participation was voluntary and children were able to withdraw from the study at any time. All children in classes three to seven were enrolled in the study, after written parental consent. The incidence of malaria amongst children was monitored over a 10-week period in May–July 2002, which corresponded to the time of peak malaria transmission. Data were collected through a system of active case detection at the three schools, supplemented by continuous passive case detection at Kapkangani Health Centre.

Ethical approval was obtained from the Ethical Review Board of Kenyatta National Hospital, Nairobi and the Danish Central Ethical Committee.

Active surveillance of malaria cases

Each school was visited by the surveillance team (nurse, laboratory technician and field assistant) two to three times per week to identify children with clinical episodes of malaria. To maximize case detection, class teachers identified and recorded the names of any children who were ill or absent from school each morning. Any children reporting fever or other malaria-related symptoms, or absenteeism because of illness were notified to the surveillance team for follow-up and screening, either in school or at home. Absentees were visited at home. A morbidity questionnaire was completed to include age, sex, history and duration of fever, other presenting signs and symptoms, and whether the child had received any prior treatment. A fingerprick blood sample was taken from any child satisfying one or more of the following screening criteria; (i) one or more of the following symptoms suggestive of malaria within the previous 24 h (reported fever, chills/shivering, rigours, vomiting, malaise, or generalized body pain) or (ii) demonstrable axillary temperature
‡
37.5
C. Giemsa-stained thick and thin
blood films were prepared and the number of asexual parasites per 200 leucocytes were counted. Schoolchildren with clinically diagnosed episodes were treated with sulphadoxine/pyrimethamine on the day of survey. Other conditions requiring treatment were referred to the health centre.

The active surveillance was supplemented by passive case surveillance in health centres, where treatment was provided free for schoolchildren enrolled in the study. At the end of the surveillance period, a cross-sectional survey among all schoolchildren was conducted to assess children's anthropometric status. Weight was measured to the nearest 0.1 kg using a Soehnle electronic balance (CMS Weighing Equipment; UK); height was measured to the nearest 0.1 cm using a portable fixed base stadiometer (CMS Weighing Equipment; UK). Anthropometric indices

were calculated using Anthro Software (Atlanta: CDC and Geneva: WHO) which uses the NCHS reference values. Height-for-age and weight-for-age were expressed as differences from the median in SD units or
z
-scores.
Children were classified as stunted and underweight if
z
-scores of height-for-age and weight-for-age respectively were <2SD below the NCHS median.

Definition of cases and community controls

A case of malaria was defined as a child with one or more of the screening symptoms and a parasite density threshold of >500 parasites/
l
l blood, following Bloland
et al.
(1999)
who used
P
500 parasites/
l
l for children aged >10 years in an endemic area of western Kenya. This is likely to be a conservative case definition when applied amongst a non-immune population in whom plasmodial infection is more likely to result in symptomatic illness. Controls were randomly drawn from amongst children who during the surveillance period were either a) well or b) symptomatic but slide-negative. Cases were matched to controls by year of age and school. We had originally planned to match every case to two controls. However, this was only possible in Kiborgok. In Chepkomia, the high number of cases meant that each case could only be matched to a single control.

Household mapping and household surveys

Homes of every school child enrolled in the study area were visited and the location and elevation of all households were determined using a hand-held Trimble GeoExplorer3 global positioning system (GPS; Trimble Navigation, Sunnyvale, CA, USA), which gives a positional accuracy within 5 m.

In addition, in the homes of cases and their age-matched controls, the household head or senior wife was interviewed in the local language to obtain data on household risk factors, and the room where the child slept was visited to record information on exposure factors that may affect mosquito–human contact. Exposure data recorded included roofing material, presence of open eaves and windows, ceiling, proximity to animal sheds, whether a child slept under a bednet and the use of methods of protection against mosquitoes such as insecticides, repellents and mosquito coils during the surveillance period. A pre-tested standardized question-

naire was used to record details of household socioeconomic characteristics including building materials and ownership of selected agricultural and household assets. The list of household assets and indicators was selected based on published literature and interviews/discussions with local key informants. These features were used to derive a wealth index, using the method of Filmer and Pritchett (2001), which has been shown to reliably measure economic status on the basis of asset ownership without the necessity of direct income or expenditure information. A principal component analysis (PCA) was used to determine the weights for an index of asset variables in order to calculate the wealth index (A_h) for each household, using STATA (v. 7.0; College Station, TX, USA). Specifically, $A_h = \sum R F_n (A_{nj}) A_n)/(S_n)$, where F_n is the scoring factor of nth asset, A_{nj} is the PCA score for nth asset of jth household, A_n and S_n are the mean and SD of the PCA score for nth asset. Variables entered into the PCA included: type of building materials used for roof and walls, presence of windows, presence of separate kitchen building, ownership of eight household assets (table, pressure lamp, mosquito net, iron, radio, clock, sofaset, bicycle) and eight agricultural assets (own land, tea bushes, cattle, sheep, donkey(s), wheelbarrow, ox-and-plough, tractor and/or other vehicle). The first principal component explained 20.7% of the variance in included variables and gave greatest weight respectively to ownership of a wheelbarrow, pressure lamp, cement

walls, iron and separate kitchen building. The resultant scores were divided into quintiles, so that each household could be classified in terms of relative socio-economic status (Armstrong-Schellenberg et al. 2003).

Land use/land cover

We image-processed satellite remote sensing data to derive thematic maps of principal land use/land cover types in the study area. Digital Landsat Enhanced Thematic Mapper (ETM+) data for 5th February 2001 were acquired, representing ecological conditions in advance of the main rainy season. The ETM+ sensor measures radiation reflected from the Earth's surface in a number of discrete spectral bands. From an ecological standpoint the most useful of these (bands 1–5 and band 7) cover the visible and near infrared portions of the electromagnetic spectrum and have a spatial resolution of 30 m. ETM data for the study area were geometrically corrected with reference to GPS ground control points using ENVI image processing software (Version 3.5; RSI Inc., Boulder, CO, USA). To produce coverages of land cover type we used a standard 'supervised' classification approach, where a maximum likelihood classification is performed to allocate each image pixel to one of a small number of known categories. The main classes identified included tea, primary forest, cleared forest and grassland. Subsequently, the distance of each household to the nearest area of forest and tea was determined using standard GIS functionality in Arc/Info (Version 7; ESRI, CA, USA).

Spatial clustering

Spatial analysis was used to explore the spatial pattern of malaria cases and help test hypotheses relating to the processes that may have given rise to the observed distributions. All households of schoolchildren were analysed, whether they were a case, control or not included in the case-control analysis. The Kulldorf spatial scan statistic was used to test whether malaria cases were distributed randomly over space, and if not, to identify significant spatial clusters (Kulldorff & Nagarwalla 1995). For this, we used the SaTScan™ software (http://satscan.org/). This programme uses a circular window moved systematically throughout the geographic space to identify significant clustering of cases. This window is centred on each of a number of possible locations throughout the study area and for each location, and the window size varies from 0 to a pre-defined upper limit. For the current analysis, the upper limit was specified as 50% of the study population, which allows both small and large clusters to be detected, while ignoring clusters that contain more than 50% of the population. For each location and size of the scanning window, a likelihood ratio test is conducted to test the hypothesis that there is an elevated rate of disease when compared with the distribution outside. The window size and location with the maximum likelihood is defined as the 'most likely' cluster (i.e. least likely to have occurred by chance). The distribution and

-value of the most likely and secondary clusters are determined by conducting Monte Carlo replications of the data set. SaTScan™ uses either a Poisson based or Bernoulli model. The latter is appropriate for 0/1 event data such as cases/non-cases, where non-cases are taken to represent the background distribution population. This approach was therefore selected for current analysis.

Risk factor analysis

Risk factor analysis was restricted to comparison of case-control sets (a subset of the data). Univariate analysis of all risk factors was conducted using logistic regression to estimate odds ratios (OR), with SE adjusted to account for within-household clustering of cases. In multivariate analysis, conditional multiple-logistic regression was employed. Analysis was conducted using STATA. Whether a child had always lived in the district was originally included in the questionnaire as a proxy measure for immune status. However, few children were born outside the district (18/284) and there was very little variation within the study population. Therefore, this variable was excluded in the analysis. Only one child was reported to be sleeping under a mosquito net and therefore this variable was also excluded from the analysis.

Results

A total of 129 incident cases were detected during a 10-week surveillance period between May and July 2002, yielding the following weekly incidence rates: 0.047/week in Koibem school; 0.032/week in Kabaskei and 0.013/week in Kiborogok. Cases were matched to 155 controls. The household distribution of malaria cases is shown in Figure 2, which indicates fewer cases occurred higher up the escarpment in Kiborgok than near the River Yala, suggesting evidence of an association between malaria and altitude.

To assess whether there are distinct spatial clusters in the distribution of malaria, we applied a spatial scan statistic separately for Kiborgok and Chepkomia. In Kiborgok, a single cluster of seven cases (1.68 expected) in seven households was identified (relative risk ¼ 4.16, P ¼ 0.058). In Chepkomia, a larger cluster of 17 cases (7.43 expected) in nine households was identified (relative risk ¼ 2.28, P ¼ 0.012). The geographical locations of these clusters are depicted in Figure 2.

Frequencies of risk factors amongst cases and controls, and associated univariate odds ratios are shown in Table 1. Neither age nor sex were identified as significant risk factors. Overall, 18.2% of the schoolchildren were stunted and 25.2% of children were underweight. There was no association between risk of malaria and stunting, but underweight children were significantly more prevalent among cases than controls. Malaria risk decreased with altitude, with significantly fewer cases occurring amongst children living above 1800 m compared with children living below 1750 m. Decreased risk of malaria was associated with increased distance from the forest fringe. Household socio-economic status was lower among the cases than controls. The practice of keeping medicines in the home was less common among families of cases than among families of controls. None of the child's room factors were associated with risk of malaria.

Based on the results of the univariate analysis and including variables with P-values <0.1, a conditional multiple-logistic regression model was developed in a backward stepwise fashion (Table 2). The risk of malaria was significantly reduced for children living in households at higher altitudes, whereas children who were underweight had a significantly higher risk of malaria. The practice of keeping malaria drugs at home had borderline significance. No significant interactions were detected among the factors included in the analysis.

By comparing the characteristics of cases identified in a spatial cluster – as identified by the scan-statistic with those of cases outside a cluster throughout the study area, the results of the spatial analysis provide further insights into risk factors for malaria. Households within an identified spatial cluster were positioned at lower altitudes than case households outside a cluster (1741 m *vs.* 1777 m, t-test: $t = 3.21$, $P < 0.001$). No other variables differed between cases identified in a cluster and cases identified outside.

2 km

Figure 2
Map of the household distribution of malaria cases identified during the active case detection in Koibem and Kabaskei, and Kiborgok (insert), and the location of significant clusters of cases as identified by the Kulldorf spatial scan statistic (large open circles).