ACCEPTED MANUSCRIPT

A micro-epidemiological analysis of febrile malaria in Coastal Kenya showing hotspots within hotspots

Philip Bejon, Thomas N Williams, Christopher Nyundo, Simon I Hay, David Benz, Peter W Gething, Mark Otiende, Judy Peshu, Mahfudh Bashraheil, Bryan Greenhouse, Teun Bousema, Evasius Bauni, Kevin Marsh, David L Smith, Steffen Borrmann

DOI: http://dx.doi.org/10.7554/eLife.02130

Cite as: eLife 2014;10.7554/eLife.02130

Received: 19 December 2013
Accepted: 1 April 2014
Published: 24 April 2014

This PDF is the version of the article that was accepted for publication after peer review. Fully formatted HTML, PDF, and XML versions will be made available after technical processing, editing, and proofing.

This article is distributed under the terms of the Creative Commons Attribution License , which permits unrestricted use and redistribution provided that the original author and source are credited.

Stay current on the latest in life science and biomedical research from eLife.
Sign up for alerts at elife.elifesciences.org

Title: A micro-epidemiological analysis of febrile malaria in Coastal Kenya showing hotspots within hotspots

Authors: Philip Bejon[1,2,*], Tom N Williams[1,3], Christopher Nyundo[1], Simon I Hay[4], David Benz[4], Peter W Gething[4], Mark Otiende[1], Judy Peshu[1], Mahfudh Bashraheil[1], Bryan Greenhouse[5], Teun Bousema[6,7], Evasius Bauni[1], Kevin Marsh[1,2], David L Smith[8], Steffen Borrmann[1,9,10]

Affiliations:

1. Kilifi KEMRI-Wellcome Trust Collaborative Research Programme, Kenya
2. Centre for Clinical Vaccinology and Tropical Medicine, University of Oxford, UK
3. Imperial College London, UK
4. Spatial Ecology and Epidemiology Group, Dept of Zoology, University of Oxford, UK
5. Dept Medicine, University of California, San Francisco, CA, USA
6. London School of Hygiene and Tropical Medicine, UK
7. Radboud University Nijmegen Medical Centre, the Netherlands
8. John Hopkins Malaria Research Institute, Baltimore, Maryland, USA
9. Institute for Tropical Medicine, University of Tübingen, Germany
10. German Centre for Infection Research, Tübingen, Germany

* Corresponding author: Philip Bejon, KEMRI-Wellcome Trust Research Programme, Kilifi, Kenya, PO Box 230. Tel +254 724 445 743. Email: pbejon@kemri-wellcome.org

Abstract: Malaria transmission is spatially heterogeneous. This reduces the efficacy of control strategies, but focusing control strategies on clusters or "hotspots" of transmission may be highly effective. Among 1,500 homesteads in coastal Kenya we calculated a) the fraction of febrile children with positive malaria smears per homestead, and b) the mean age of children with malaria per homestead. These two measures were inversely correlated, indicating that children in homesteads at higher transmission acquire immunity more rapidly. This inverse correlation increased gradually with increasing spatial scale of analysis, and hotspots of febrile malaria were identified at every scale. We found hotspots within hotspots, down to the level of an individual homestead. Febrile malaria hotspots were temporally unstable, but 4km radius hotspots could be targeted for one month following one month periods of surveillance.

Introduction

The transmission of infectious disease often shows substantial heterogeneity (Woolhouse, Dye et al. 1997). Malaria transmission is determined by mosquito ecology and behavior, which is in turn determined by rainfall, hydrology, soils, human behavior and population distributions, and a range of other social, biotic and abiotic factors. Heterogeneity of malaria transmission is apparent at global scale (Gething, Patil et al. 2011), regional scale (Kleinschmidt, Omumbo et al. 2001; Noor, Gething et al. 2009), and at fine scale in, for instance, Mali (Gaudart, Poudiougou et al. 2006), Ghana (Kreuels, Kobbe et al. 2008), Ethiopia (Yeshiwondim, Gopal et al. 2009) Kenya (Brooker, Clarke et al. 2004; Ernst, Adoka et al. 2006; Bejon, Williams et al. 2010) and Tanzania (Bousema, Drakeley et al. 2010). This spatial heterogeneity makes transmission relatively resilient to indiscriminate control efforts, but also provides an opportunity to engage in targeted

45 malaria control on clusters of transmission (or "hotspots"), a strategy that is predicted to be
46 highly effective (Dye and Hasibeder 1986; Woolhouse, Dye et al. 1997).
47
48 We have previously identified hotspots of malaria using active surveillance (Bejon, Williams et
49 al. 2010). Others have identified hotspots using passive surveillance in health facilities linked to
50 demographic surveillance systems (Ernst, Adoka et al. 2006). Passive surveillance is more
51 readily scaled up, but may be biased by variations in access to health care facilities and socially-
52 determined health seeking behavior (Sumba, Wong et al. 2008; Franckel and Lalou 2009). The
53 incidence of febrile malaria presenting to health care is thus biased by access to care. This bias
54 may be countered by using the malaria positive fraction (MPF) among children with fever (also
55 termed "slide positivity rate" in some publications (Jensen, Bukirwa et al. 2009)). The MPF
56 includes all febrile children presenting to the dispensary as the denominator, hence controlling
57 for access to health care, in contrast to incidence for which all children in the community are
58 included in the denominator. The MPF is less likely to show systematic spatial bias with
59 distance from the health facility since parental accounts of illness have not been found to
60 discriminate malaria from non-malarial fever (Luxemburger, Nosten et al. 1998; Mwangi,
61 Mohammed et al. 2005), and diagnostic testing is not available outside the dispensary.
62
63 We present data from demographic surveillance linked to passive case detection in Pingilikani
64 dispensary in Kilifi District, coastal Kenya. Data are collected from 1,500 homesteads within an
65 8km radius followed for 9 years. We analyse the spatial heterogeneity of malaria cases in order
66 to determine the temporal and spatial scales of case clustering so as to inform targeting in
67 malaria control programmes. We also excluded visits with specific symptoms such as skin

infections or cutaneous abscesses, otitis media and gastroenteritis (>4 episodes diarrhea per day) that might have been the primary motivation for seeking health care rather than fever per se.

Results

Among ~20,000 remaining febrile presentations from ~1,500 different residences, 54% were positive for *Plasmodium falciparum* on blood smear examination. Using homestead as our unit of analysis, we found that the incidence of dispensary attendance declined with distance from the dispensary (on average -0.040 (95%CI 0.036-0.044) and -0.041 (95%CI 0.037-0.046) episodes per child year for each km for malaria smear positive and negative attendees, respectively). MPF was not found to vary significantly by distance of residence from the dispensary (from MPF=0.50, 95%CI 0.47 to 0.54 at <2km distance to MPF=0.52, 95%CI 0.47 to 0.57 at 6-7km, p=0.7).

The spatio-temporal distribution of MPF by homestead is shown in video 1 (slow speed) and video 2 (fast speed). The visual impression from these clips suggests marked spatial variation, with some geographical areas showing persistently high MPFs, and other areas showing more marked temporal variation. Temporally stable spatial heterogeneity would be expected to lead to spatial heterogeneity in the acquisition of immunity, which may be evidenced by variation in the age profiles of children with febrile malaria. We therefore tested this hypothesis as below.

Spatial heterogeneity in malaria risk and acquisition of immunity

MPF was inversely correlated with the average age of children with malaria (Spearman's rank correlation (r_s) =-0.16, p<0.0001 (Figure 1, a, b, c). This suggests that greater exposure to

malaria (i.e. high MPF) leads to more rapid acquisition of immunity as children grow up, hence predominantly younger children visiting the dispensary with febrile malaria. There was no evidence that this relationship was confounded by spatial clustering of age: the average age of children with non-malarial fever did not show spatial clustering (Moran's I=0.01, p=0.5 within 1 km and Moran's I=0.02, p=0.5 within 5 km) and was not associated with MPF (r_s=-0.02, p=0.4). We examined the effect of spatial scale at which this correlation occurred by imposing grids of increasing cell size on the study area, calculating r_s within each cell of the grid, and then estimating the mean r_s at each scale of grid (Figure 1d, blue lines). The mean r_s trended gradually away from 0 as the grid divisions became larger in scale. This pattern suggests gradual differentiation in transmission characteristics as the distance between homesteads included within a cell of the grid increases. We then examined the patterns seen on applying this analysis to simulated data. In order to exclude that this trend was a result of cells at fine-scale containing fewer homesteads, we ran permutations of the data using after randomly re-assigning spatial coordinates to the homesteads. These permutations show that a consistent correlation at r_s =-0.16 throughout the range of grid sizes, albeit with greater uncertainty with smaller cell size (Figure 1d, red lines). Hence the trend of a gradually increasing inverse correlation as the grid size increases does not appear to be explained simply by having fewer homesteads in each cell at fine scale. In order to determine the pattern that might be seen with specific spatial scales of clustering, we conducted further simulations by imposed patterns with specific scales on the spatial coordinates of the homesteads, in varying proportions with random noise using a gamma distribution. These simulations show that a specific scale of clustering produces "spikes" in r_s as the cell size varies, with the position of the spike coinciding with scale of the clustering (Figure 1 -figure supplement 1). Reducing the Signal:Noise ratio eventually obscured the "spikes" due to

a characteristic pattern, but only at the point where the overall correlation was no longer discernible (Figure 1 -figure supplement 2). Adding a gradient to the simulated characteristic scale attenuated but did not obscure the "spikes" (Figure 1 -figure supplement 3).

Hotspots within hotspots

Using the Bernoulli model in SaTScan(Kulldorff 1997) we identified a hotspot with a radius of 5.8km at $p<0.00001$ (Figure 2a) using the full dataset (for which n=20,702). However, on re-analysis of the children within this hotspot (in which n=5,300), we identified a further hotspot (with a radius of 0.76km) within the 5.8km hotspot ($p<0.00001$, Figure 2b). Then on further re-analysis of the homesteads within that 0.76km hotspot (within which n=1,406), we identified a third significant hotspot ($p=0.016$) which comprised a single homestead, in which there were 36 episodes of malaria compared with 3 malaria negative fevers (Figure 2d). When we selected a random 5km square area outside the original 5.8km radius hotspot, we identified a hotspot within this area a fourth hotspot with a 1.32km radius ($p<0.00001$, Figure 2c).

In order to further explore the scale of spatial clustering, we plotted the semivariogram (Figure 2 -figure supplement 1) and the log-log transformed semivariogram (Figure 2 -figure supplement 2). These plots suggested linear fits for the semivariogram, suggesting that spatial clustering occurred over a range of spatial scales.

Temporal Trends of Spatial Heterogeneity

We also examined temporal trends for individual homesteads (Figure 3). There was an inverse correlation between the mean MPF and the variance in MPF over the 10 year study period (r_s =-

137 0.61, $p<0.0001$, Fig 3a). The temporal trends for two subsets of homestead can be seen in figure
138 3b (stable high MPF) and figure 3c (unstable low MPF), suggesting that homesteads can be
139 characterized as stable high transmission homesteads or unstable low transmission homesteads.
140 Infant parasite rates have been proposed as a measure of transmission intensity that minimizes
141 the offsetting of acquired immunity in macro-epidemiological studies (Snow, Molyneux et al.
142 1996). We therefore hypothesized that the malaria positive fractions in children <1 yr of age
143 (hereafter "$MPF_{<1yr}$") would measure transmission intensity without the offsetting of acquired
144 immunity, and that unstable transmission would result in higher risk of malaria in older children.
145 To test this hypothesis, we calculated the mean $MPF_{<1yr}$ and the variance in $MPF_{<1yr}$ for each
146 homestead over the 9 years of follow up and tested the relationships between these metrics and
147 risk of malaria in older children in multivariable linear regression models.
148
149 In multivariable linear regression models $MPF_{<1yr}$ was strongly correlated with MPFs in children
150 in the 1-2 year-old and 2-3 year-old age group, but progressively less strongly correlated with
151 MPF in older children (Figure 3di). The regression coefficient was ~0.4 for 1-2 year olds,
152 meaning that each unit increase in $MPF_{<1yr}$ is associated with a 0.4 increase in the MPF for 1-2
153 year old children. On the other hand, the variance in $MPF_{<1yr}$ was not correlated with MPFs in 1-
154 2 or 2-3 year old children, but was progressively more strongly correlated with MPF in older
155 children (Figure 3dii). Hence there were high stable transmission homesteads, with
156 predominantly younger children getting febrile malaria, and low unstable transmission
157 homesteads, with increasing risk to older children. This pattern of high stable versus low
158 unstable transmission also occurs between regions or countries, and demonstrates a similarity
159 between the micro and macro-epidemiology of malaria (Hay, Smith et al. 2008).

Theoretical accuracy of targeted control undertaken at varying temporal and spatial scales. We then used our dataset to simulate the accuracy of targeting cases that a malaria control programme might achieve on conducting surveillance over a defined period of time followed by targeted control. We assumed that malaria control programmes would need to define *a priori* the period of time to use for surveillance, and also to select a spatial scale at which to define hotspots. For varying time periods and spatial scales, we determined the % of excess malaria cases within the targeted hotspots compared with the surrounding area in the period of time immediately following the simulated surveillance.

One week periods of surveillance (top left panel of figure 4) did not identify hotspots that are still present the following week at fine spatial scales (i.e. the plotted line indicates that the accuracy of targeting is 0% at scales of less than 1km). On the other hand, at larger spatial scales we found that one week periods of surveillance were more accurate, resulting in the targeting of areas with a 60% excess of new malaria cases compared with the surrounding area at a scale of an 8 km diameter. A similar pattern was seen for monthly periods of surveillance. Longer surveillance periods (e.g. 6 months) resulted in targeting areas with an excess of 20% malaria cases compared with the surrounding area over the range of spatial scales examined.

ITN use and spatial variation in risk

Mass distributions of Insecticide Treated Nets (ITNs) in the area began in 2006. ITN use was surveyed in 2009 and 2010. We found that children using ITNs had a reduced risk of malaria by logistic regression (i.e. OR=0.69, 95%CI 0.67 to 0.8, $p<0.001$), in keeping with previous

183 literature on the personal protection provided by ITN use (Lim, Fullman et al. 2011). On the other
184 hand we did not identify significant evidence that ITN use was clustered spatially (Moran's
185 I=0.02, p=0.5). Furthermore, adding ITN use as a covariate in SaTScan analysis to locate
186 hotspots had little effect on results; the addition of ITN use as a covariate changed the location
187 of the hotspot by 120m, and changed the predicted radius of the hotspot from 5.4 to 5.2km. On
188 re-analysis of the homesteads within the 5.4km hotspot, a further 0.87km hotspot was identified
189 the position and radius of which were not altered by the inclusion of ITN use as a covariate.
190 Finally, within this 0.87km hotspot the same 7 homesteads were identified as a hotspot
191 irrespective of the inclusion of ITN use as a covariate. We did not identify significant evidence
192 that ITN use correlated mean $MPF_{<1yr}$ (r_s = -0.04, p=0.04) or with the variance in $MPF_{<1yr}$ (r_s=-
193 0.01, p=0.7). Hence ITNs provided personal protection from malaria, but we were unable to
194 show that they explained the spatial micro-epidemiological patterns.
195

196 ## Discussion

197 We found that malaria cases were spatially heterogeneous in an 8km radius area of coastal
198 Kenya. The strongly significant inverse correlation between the malaria positive fraction (MPF)
199 and average age of children presenting with malaria suggests variable acquisition of immunity
200 between homesteads. Homesteads at high transmission intensity have a high MPF and a young
201 average age of malaria (with older children becoming immune and therefore not presenting to the
202 dispensary) whereas homesteads at low transmission intensity have a low MPF but an older
203 average age of malaria since older children are not becoming immune as rapidly. In theory, this
204 inverse correlation might have arisen because of heterogeneity at various spatial scales. For
205 instance, there might have been a block of homesteads all at high transmission in one half of the

study area (thus with high MPF and low average age) and a second block of homesteads at low transmission in the other half (with low MPF and high average age). On the other hand, the inverse correlation might have arisen because of a random distribution of "high" and "low" transmission intensity homesteads throughout the study area.

In order to determine at which spatial scale transmission was heterogeneous, we conducted an analysis where correlation coefficient was recalculated within each cell of a grid superimposed on the study area. The mean correlation coefficient of all cells was then presented as the cell size of the grid used was increased (Figure 1d). This analysis was done to identify the most influential geographical scale at which the inverse correlation was observed. In simulated data, we noted "spikes" where the inverse correlation was abruptly lost when the size of cells in the grid coincides with the size of the geographical "blocks" of homesteads that drove the inverse correlation, as seen in Figure 1, figure supplement 1. Similar spikes were seen after adding simulated noise and gradients in space over which the correlation varied (figure supplements 2 and 3). Real-world data would contain more complex sources of variation than we have simulated, and hence may not produce distinct spikes. Nevertheless, the analysis of these simulations suggests that discontinuities in the correlation between MPF and average age of malaria over cell size might be expected when clustering is at a specific spatial scale. In fact there was no such discontinuity in the function shown in figure 1d, indicating that the inverse correlation was present at every geographical scale examined within our study. It is likely that this pattern would extend at greater geographical scales, since a similar inverse correlation between the age distributions of malaria cases and transmission intensity can be seen on comparing countries and regions (Okiro, Al-Taiar et al. 2009).

The pattern of spatial heterogeneity is relevant to malaria control, since targeted disease control is predicted to be highly effective (Woolhouse, Dye et al. 1997). Spatial targeting is particularly appropriate for malaria "hotspots" (Coleman, Mabuza et al. 2009; Moonen, Cohen et al. 2010; Bousema, Griffin et al. 2012; Sturrock, Novotny et al. 2013) and many malaria control programs are already engaged in spatially-targeted intervention (Zhou, Githeko et al. 2010; Loha, Lunde et al. 2012). Our data showing clustering at varying spatial scales suggest that malaria control programs can expect to identify hotspots at many different geographical scales. We demonstrate that hotspots occur within hotspots, down to the level of a single homestead, and also that hotspots can be identified on "zooming in" on random areas outside the main hotspot (Figure 2c). These hotspots were based on analysis of a large dataset with adequate power, and were strongly significant based on the multiple permutations run in SaTScan, suggesting that type I statistical error is an unlikely explanation for our findings. The complexity of presenting "hotspots within hotspots" to a malaria control programme is further compounded by the temporal instability of the spatial pattern (Figure 3).

We therefore simulated the accuracy with which hotspots could be targeted using varying spatial scales and varying time periods of surveillance. We found that using data aggregated over one month of surveillance to define 4 to 8 km diameter hotspots would provide greatest accuracy, but this information is only relevant for one month before temporal instability necessitates further surveillance. One might therefore consider a continuous programme of parallel surveillance and targeting, where the surveillance data are examined at the end of each month to determine the

location to be targeted for the following month. Continuous surveillance would allow adaptive targeting of hotspots for the following month. Such a strategy might be employed all year round, or for a limited period of the year depending on local seasonality. (Cairns, Roca-Feltrer et al. 2012) Targeting at this spatial scale has the added practical advantage that it could be done with village-level location data and would not require fine-scale geo-positional data.

There are some caveats to this recommendation. Our observations are from a single site. Other sites should examine their local data to determine whether a similar targeting strategy is appropriate. Furthermore, some hotspots did show temporal stability. For instance, we identified a 6 km diameter hotspot south east of the dispensary that maintained a 30 to 60% increase in MPF compared with the surrounding area throughout the 9-year surveillance.

Children with positive microscopy slides for malaria presenting at the dispensary may have genuine febrile malaria, or alternatively may have chronic asymptomatic parasitaemia with co-incident non-malarial fever. Previous studies estimating malaria attributable fractions in the locality suggest 61% of the children in our analysis would have malaria as the proximate cause of their illness, with the other 39% having chronic asymptomatic parasitaemia with co-incident fever from another cause (Olotu et al, 2011). We have previously demonstrated that spatial heterogeneity is more temporally stable when analysed for asymptomatic parasitaemia rather than febrile malaria (Bejon et al, 2010). Targeting hotspots of asymptomatic parasitaemia would require community surveys rather than dispensary monitoring, which may need to be done less frequently than monitoring of febrile malaria episodes.

Furthermore MPF is not a comprehensive indicator of transmission intensity. Homesteads with consistently low average ages of febrile malaria are likely to be stable high transmission homesteads (such as those in subset p of Figure 3a) which amplify transmission in the areas surrounding them. Targeting such high transmission homesteads to interrupt transmission may be highly effective (Woolhouse, Dye et al. 1997). The stronger inverse correlation between MPF and average age of febrile malaria as spatial scale increases (Figure 1) suggests that the spatial heterogeneity of transmission is progressively more stable at more coarse spatial scales.

Malaria transmission is determined by mosquito ecology and behavior. Mosquito ecology may be determined by obvious geographical features such as altitude (Reyburn, Mbatia et al. 2005), cultivation practices (Lindsay, Wilkins et al. 1991), streams and dams (Ghebreyesus, Haile et al. 1999), wind direction (Midega, Smith et al. 2012) and mosquito searching behavior for hosts (Smith, Dushoff et al. 2004). Ecological models based on such features have been developed using frequentist techniques (Omumbo, Hay et al. 2005), Bayesian approaches (Craig, Sharp et al. 2007) and fuzzy logic (Snow, Gouws et al. 1998). However, the same ecological factor may act inconsistently in different geographical areas (Kleinschmidt, Sharp et al. 2001; Gemperli, Sogoba et al. 2006; Noor, Clements et al. 2008), and the effect of ecological factors is modified by fine-scale vector and host movement (Perkins, Scott et al. 2013). Our data suggests that the environmental factors determining malaria transmission operate at a range of spatial scales. We might speculate that mosquito breeding site density could be equally influenced by proximity to a large geographical feature such as a river, or to a micro-geographical feature such as a cow hoof-print (Sattler, Mtasiwa et al. 2005). Hence ecological models of malaria transmission will need to include data at a range of spatial scales in order to accurately predict malaria risk.

Methods

Approval for human participation in these cohorts was given by Kenya Medical Research Institute Ethics Research Committee, and research was conducted according to the principles of the declaration of Helsinki.

Study population

Pingilikani Dispensary is 40km to the North of Mombasa, in Kilifi Country, Coast Province, Kenya. The population relies mainly on subsistence farming and experiences all year round malaria transmission, with "long" and "short" rains each year causing two peaks in transmission. Estimates of the local EIR were 22-53 in 2003 (1), and 21.7 infective bites per person per year in 2010 (2). Between 2003 and 2011 data were collecting on all children (i.e. ≤15 years of age) attending the dispensary.

Demographic surveillance is conducted for the 240,000 people in a 900 square kilometre area in Kilifi County. Four-monthly enumeration rounds were conducted to identify births, deaths and migration (3). Each inhabitant is described by their family relationships and their homestead of residence, with geospatial coordinates, and assigned a unique personal identifier. These details were used to link children visiting Pingilikani dispensary to geospatial coordinates for the homestead of residence. During enumeration rounds in 2009-2011 ITN use per individual was established during visits to the homestead, as reported by a homestead representative.

We restrict analysis to within an 8km radius of the dispensary, which accounted for >96% of all visits to the dispensary, and excluded visits with specific symptoms such as skin infections or cutaneous abscesses, otitis media and gastroenteritis (>4 episodes diarrhoea per day) that might have been the primary motivation for seeking health case rather than fever per se. These latter exclusions combined accounted for 14% of all visits.

Malaria diagnosis and treatment

All children presenting for assessment (except those with trauma as their only concern) had finger-prick blood samples examined for malaria parasites. Thick and thin blood smears were stained with 10% Giemsa and examined at x1000 magnification for asexual *Plasmodium falciparum* parasites. 100 fields were examined before slides could be considered negative. Amodiaquine was the first-line anti-malarial from 2003 to 2005, when policy changed to Co-artemether.

Analysis

Fever was defined as either reported fever by the parents or measured fever, i.e. axilliary temperature ≥37.5◦C (Mackowiak, Bartlett et al. 1997). The malaria positive fraction (MPF) was calculated as the fraction of febrile children attending the dispensary with fever who were positive for malaria parasites by blood smear examination. MPF was aggregated by homestead. Multiple identifications of fever and parasitaemia in the same child within 21 days were considered a single episode.

342 The average age of febrile malaria was calculated as the arithmetic mean age at which children
343 visited the dispensary with fever and malaria parasites. Correlations between average age of
344 febrile malaria and MPF per homestead were calculated using spearman's rank correlation
345 coefficient. Grids of gradually increasing cell size were calculated using longitude and latitude
346 coordinates. Simulations were done using the distribution of homesteads identified in our study.
347 We applied a factor to MPF (positive) and average age (negative) to the homesteads within a
348 block of varying size to induce the appearance of clustering at a given spatial scale. Random
349 noise was added to these simulations using a gamma distribution. In the first round of
350 simulations we set the Signal:Noise ratio (i.e. the ratio between the factor applied to MPF and
351 average age versus the mean amplitude of the noise) to reproduce the r_s seen in the real data. In
352 the second round of simulations we varied the Signal:Noise Ratio as shown in individual panels,
353 and in the third round of simulations we introduced a gradient over which the correlation
354 emerged, where the factor applied to MPF and average age was tapered in a uniform way
355 towards 1 beginning at the edge of the simulated block.
356
357 Hotspots were defined using SaTScan software to calculate the spatial scan statistic (Kulldorff
358 1997). The software is freely available and can be downloaded from www.satscan.org. The
359 version used in this analysis was downloaded in November 2012, as v9.1 for a 64-bit system.
360 The spatial scan statistic uses a scanning window that moves across space. The scanning
361 windows are circles centred on each homestead, with a radius varied from inclusion of only the
362 single homestead it is centred on through to 30% of the population size. When using the
363 Bernoulli model, the software calculates the fraction of cases/ controls inside versus outside the
364 each possible scanning window, and selects the window giving the highest probability of a case

within the scanning window compared with the probability of a case outside the window. In our application of the Bernoulli model, cases were febrile children with parasitaemia and controls were febrile children without parasitaemia. The test of significance needs to take into account the whole process of selecting the optimal window rather than simply the comparison of inside versus outside the optimal window. This is achieved by running random permutations of the case/control data over the spatial co-ordinates of homesteads and determining the log-likelihood statistic for the model fit by the optimal window for each random permutation. The log-likelihood statistic for the real data is then compared with the statistics on the random permutations to derive a p value. We used 9999 replications in our study. The maximum hotspot size was set at 30% of the population, and the inference level for significance was set at 0.05. The main analysis was done without adjustment for covariates, and a secondary analysis was conducted for the 2009/2010 data with and without ITN use as a covariate. Kernel smoothing with a 1km radius is used for spatial display graphs, but all analyses of correlation are conducted on raw data without smoothing.

Semivariograms, Moran's I and linear regression models were run in Stata version 12 (StataCorp, Texas). Semivariograms were constructed using 0.1km intervals between 0.1km and 10km. Moran's I was assessed globally using cumulative bands of <0.1, <0.5, <1 and <2 and <5kms.

Conflicts of Interest: There are no conflicts of interest. The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

Acknowledgments: Peter D Crompton is thanked for helpful comments during manuscript drafting. The manuscript is published with the permission of the Director of KEMRI. PB is jointly funded by the UK Medical Research Council (MRC) and the UK Department for International Development (DFID) under the MRC/DFID Concordat agreement. Work in Pingilikani was funded by the German Research Foundation (DFG, Grant number SFB 544, A7) and by the Wellcome Trust. Bonston Piri and Epson Mwadori are thanked for their contributions in making the geospatial data available.

References:

Bejon, P., T. N. Williams, et al. (2010). "Stable and unstable malaria hotspots in longitudinal cohort studies in Kenya." Plos Med **7**(7): e1000304.

Bousema, T., C. Drakeley, et al. (2010). "Identification of hot spots of malaria transmission for targeted malaria control." J Infect Dis **201**(11): 1764-1774.

Bousema, T., J. T. Griffin, et al. (2012). "Hitting hotspots: spatial targeting of malaria for control and elimination." Plos Med **9**(1): e1001165.

Brooker, S., S. Clarke, et al. (2004). "Spatial clustering of malaria and associated risk factors during an epidemic in a highland area of western Kenya." Trop Med Int Health **9**(7): 757-766.

Cairns, M., A. Roca-Feltrer, et al. (2012). "Estimating the potential public health impact of seasonal malaria chemoprevention in African children." Nat Commun **3**: 881.

Coleman, M., A. M. Mabuza, et al. (2009). "Using the SaTScan method to detect local malaria clusters for guiding malaria control programmes." Malar J **8**(1): 68.

Craig, M. H., B. L. Sharp, et al. (2007). "Developing a spatial-statistical model and map of historical malaria prevalence in Botswana using a staged variable selection procedure." Int J Health Geogr **6**: 44.

Dye, C. and G. Hasibeder (1986). "Population dynamics of mosquito-borne disease: effects of flies which bite some people more frequently than others." Trans R Soc Trop Med Hyg **80**(1): 69-77.

Ernst, K. C., S. O. Adoka, et al. (2006). "Malaria hotspot areas in a highland Kenya site are consistent in epidemic and non-epidemic years and are associated with ecological factors." Malar J **5**: 78.

Franckel, A. and R. Lalou (2009). "Health-seeking behaviour for childhood malaria: household dynamics in rural Senegal." J Biosoc Sci **41**(1): 1-19.

Gaudart, J., B. Poudiougou, et al. (2006). "Space-time clustering of childhood malaria at the household level: a dynamic cohort in a Mali village." BMC Public Health **6**: 286.

Gemperli, A., N. Sogoba, et al. (2006). "Mapping malaria transmission in West and Central Africa." Trop Med Int Health **11**(7): 1032-1046.

Gething, P. W., A. P. Patil, et al. (2011). "A new world malaria map: Plasmodium falciparum endemicity in 2010." Malar J **10**: 378.

Ghebreyesus, T. A., M. Haile, et al. (1999). "Incidence of malaria among children living near dams in northern Ethiopia: community based incidence survey." Bmj **319**(7211): 663-666.

Hay, S. I., D. L. Smith, et al. (2008). "Measuring malaria endemicity from intense to interrupted transmission." Lancet Infect Dis **8**(6): 369-378.

Jensen, T. P., H. Bukirwa, et al. (2009). "Use of the slide positivity rate to estimate changes in malaria incidence in a cohort of Ugandan children." Malar J **8**: 213.

Kleinschmidt, I., J. Omumbo, et al. (2001). "An empirical malaria distribution map for West Africa." Trop Med Int Health **6**(10): 779-786.

Kleinschmidt, I., B. L. Sharp, et al. (2001). "Use of generalized linear mixed models in the spatial analysis of small-area malaria incidence rates in Kwazulu Natal, South Africa." Am J Epidemiol **153**(12): 1213-1221.

Kreuels, B., R. Kobbe, et al. (2008). "Spatial variation of malaria incidence in young children from a geographically homogeneous area with high endemicity." J Infect Dis **197**(1): 85-93.

Kulldorff, M. (1997). "A spatial-scan statistic." Communications in Statistics: Theory and Methods **26**: 1481-1496.

Lim, S. S., N. Fullman, et al. (2011). "Net benefits: a multicountry analysis of observational data examining associations between insecticide-treated mosquito nets and health outcomes." PLoS Med **8**(9): e1001091.

Lindsay, S. W., H. A. Wilkins, et al. (1991). "Ability of Anopheles gambiae mosquitoes to transmit malaria during the dry and wet seasons in an area of irrigated rice cultivation in The Gambia." J Trop Med Hyg **94**(5): 313-324.

Loha, E., T. M. Lunde, et al. (2012). "Effect of bednets and indoor residual spraying on spatio-temporal clustering of malaria in a village in south Ethiopia: a longitudinal study." PLoS ONE **7**(10): e47354.

453 Luxemburger, C., F. Nosten, et al. (1998). "Clinical features cannot predict a diagnosis of
454 malaria or differentiate the infecting species in children living in an area of low
455 transmission." Trans R Soc Trop Med Hyg **92**(1): 45-49.
456 Mackowiak, P. A., J. G. Bartlett, et al. (1997). "Concepts of fever: recent advances and lingering
457 dogma." Clin Infect Dis **25**(1): 119-138.
458 Midega, J. T., D. L. Smith, et al. (2012). "Wind direction and proximity to larval sites determines
459 malaria risk in Kilifi District in Kenya." Nat Commun **3**: 674.
460 Moonen, B., J. M. Cohen, et al. (2010). "Operational strategies to achieve and maintain malaria
461 elimination." Lancet **376**(9752): 1592-1603.
462 Mwangi, T. W., M. Mohammed, et al. (2005). "Clinical algorithms for malaria diagnosis lack
463 utility among people of different age groups." Trop Med Int Health **10**(6): 530-536.
464 Noor, A. M., A. C. Clements, et al. (2008). "Spatial prediction of Plasmodium falciparum
465 prevalence in Somalia." Malar J **7**: 159.
466 Noor, A. M., P. W. Gething, et al. (2009). "The risks of malaria infection in Kenya in 2009."
467 BMC Infect Dis **9**: 180.
468 Okiro, E. A., A. Al-Taiar, et al. (2009). "Age patterns of severe paediatric malaria and their
469 relationship to Plasmodium falciparum transmission intensity." Malar J **8**: 4.
470 Omumbo, J. A., S. I. Hay, et al. (2005). "Modelling malaria risk in East Africa at high-spatial
471 resolution." Trop Med Int Health **10**(6): 557-566.
472 Perkins, T. A., T. W. Scott, et al. (2013). "Heterogeneity, mixing, and the spatial scales of
473 mosquito-borne pathogen transmission." PLoS Comput Biol **9**(12): e1003327.

Reyburn, H., R. Mbatia, et al. (2005). "Association of transmission intensity and age with clinical manifestations and case fatality of severe Plasmodium falciparum malaria." Jama **293**(12): 1461-1470.

Sattler, M. A., D. Mtasiwa, et al. (2005). "Habitat characterization and spatial distribution of Anopheles sp. mosquito larvae in Dar es Salaam (Tanzania) during an extended dry period." Malar J **4**: 4.

Smith, D. L., J. Dushoff, et al. (2004). "The risk of a mosquito-borne infection in a heterogeneous environment." PLoS Biol **2**(11): e368.

Snow, R. W., E. Gouws, et al. (1998). "Models to predict the intensity of Plasmodium falciparum transmission: applications to the burden of disease in Kenya." Trans R Soc Trop Med Hyg **92**(6): 601-606.

Snow, R. W., C. S. Molyneux, et al. (1996). "Infant parasite rates and immunoglobulin M seroprevalence as a measure of exposure to Plasmodium falciparum during a randomized controlled trial of insecticide-treated bed nets on the Kenyan coast." Am J Trop Med Hyg **55**(2): 144-149.

Sturrock, H. J., J. M. Novotny, et al. (2013). "Reactive case detection for malaria elimination: real-life experience from an ongoing program in swaziland." PLoS ONE **8**(5): e63830.

Sumba, P. O., S. L. Wong, et al. (2008). "Malaria treatment-seeking behaviour and recovery from malaria in a highland area of Kenya." Malar J **7**: 245.

Woolhouse, M. E., C. Dye, et al. (1997). "Heterogeneities in the transmission of infectious agents: implications for the design of control programs." Proc Natl Acad Sci U S A **94**(1): 338-342.

Yeshiwondim, A. K., S. Gopal, et al. (2009). "Spatial analysis of malaria incidence at the village level in areas with unstable transmission in Ethiopia." Int J Health Geogr **8**: 5.

Zhou, G., A. K. Githeko, et al. (2010). "Community-wide benefits of targeted indoor residual spray for malaria control in the western Kenya highland." Malar J **9**: 67.

Rich Media File Legends:

Video 1

Each plotted point represents an individual homestead, where the color shading indicates the malaria positive fraction (MPF), with red shading for high MPF and blue shading for low MPF. Points change color each year.

Video 2

Each plotted point represents an individual homestead, where the color shading indicates the malaria positive fraction (MPF), with red shading for high MPF and blue shading for low MPF. Points change color each year. The frames are identical to those in video 1, but move more rapidly.

Figure Legends:

Figure Legend 1

Each plotted point represents an individual homestead, where the color shading indicates the malaria positive fraction (MPF) in panel a, or the average age of children who test positive for

malaria in panel b. Panel c shows the scatter plot for MPF versus average age (Spearman's rank correlation coefficient (r_s) =-0.16, $p<0.0001$). Panel d shows r_s (y axis) plotted against scale of analysis (x axis), where a grid with varying cell size is imposed on the study area, r_s is calculated within each cell and then the mean r_s presented, with 95% confidence intervals produced by boot-strap (blue solid and dashed lines, respectively), and the results of analysis of spatially-random permutations of the data with equivalent cell size are shown for comparison (red solid and dashed lines, respectively). The analysis shown in panel d was compared on simulations with varying simulated characteristic scales, signal:noise ratios and with added gradients (Figure supplements 1,2 and 3, respectively).

Figure 1 –figure supplement 1

Simulated data using imposed spatial clustering at specific scales are analysed to determine r_s (y axis) plotted against scale of analysis (x axis), where a grid with varying cell size is imposed on the study area, r_s is calculated within each cell and then the mean r_s presented, with 95% confidence intervals produced by boot-strap (blue solid and dashed lines, respectively). The six panels show the appearances of different imposed scales as shown in the sub-titles.

Figure 1 –figure supplement 2

Simulated data using imposed spatial clustering at specific scales are analysed to determine r_s (y axis) plotted against scale of analysis (x axis), where a grid with varying cell size is imposed on the study area, r_s is calculated within each cell and then the mean r_s presented, with 95% confidence intervals produced by boot-strap (blue solid and dashed lines, respectively). The six panels show the appearances using different signal:noise ratios.

Figure 1 –figure supplement 3

Simulated data using imposed spatial clustering at specific scales are analysed to determine r_s (y axis) plotted against scale of analysis (x axis), where a grid with varying cell size is imposed on the study area, r_s is calculated within each cell and then the mean r_s presented, with 95% confidence intervals produced by boot-strap (blue solid and dashed lines, respectively). The six panels show the appearances using gradients of varying spatial scales around the simulated clustering.

Figure Legend 2.

Each plotted point represents an individual homestead, where the color shading indicates the malaria positive fraction (MPF). Hotspots are identified using SATScan, using the whole study area (panel a), then repeated within the hotspot (panel b), within the hotspot of panel b (panel d), and then within a randomly chosen area outside the hotspot (panel c). The semi-variogram and log-log semi-variogram plot are shown in figure supplements 1 and 2, respectively.

Figure 2 –figure supplement 1

The semi-variogram is shown for MPF. A lowess smoothed line is superimposed on the data points.

Figure 2 –figure supplement 2

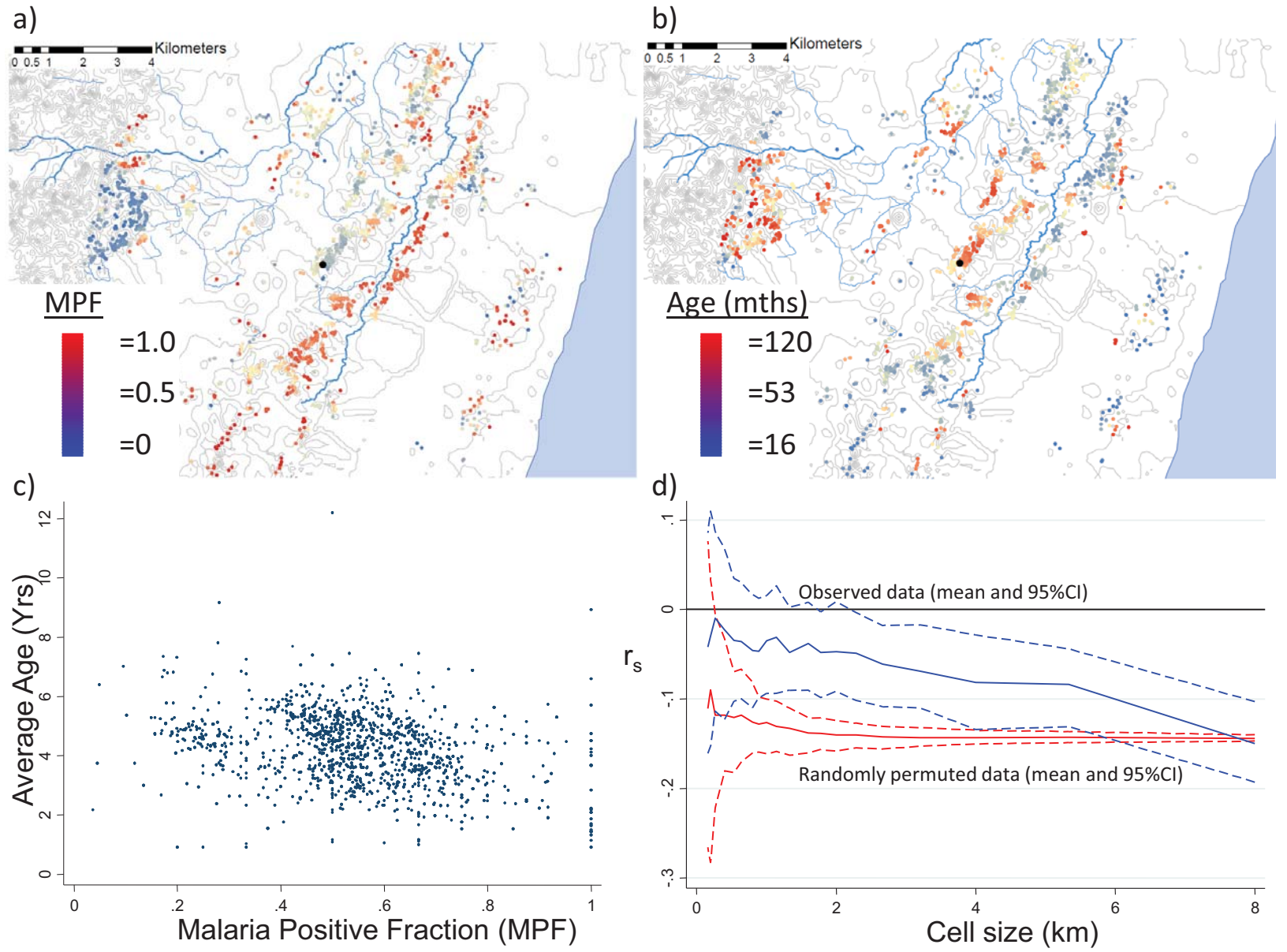
The log-log plot of the semi-variogram is shown for MPF. A lowess smoothed line is superimposed on the data points.

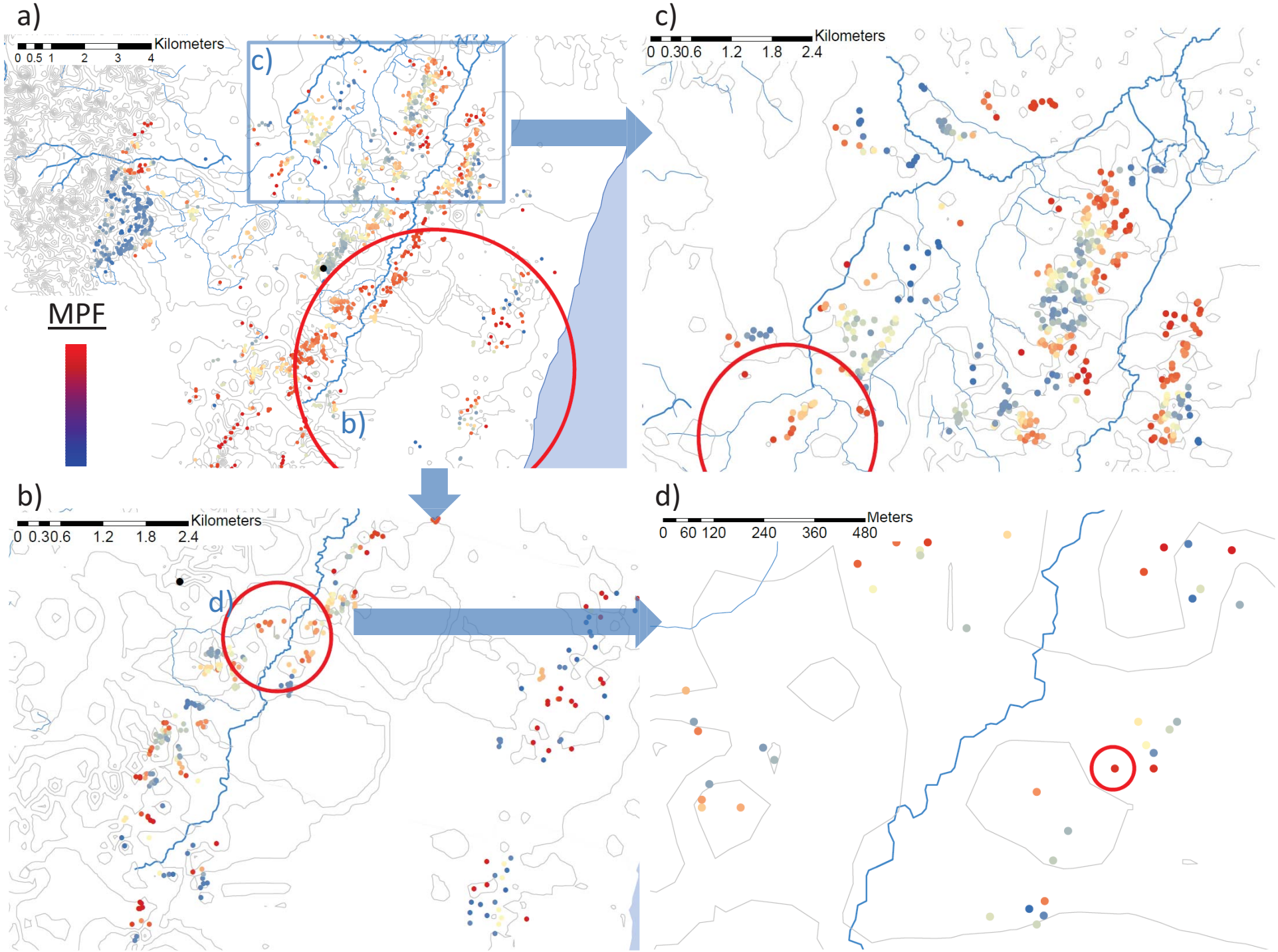
Figure Legend 3

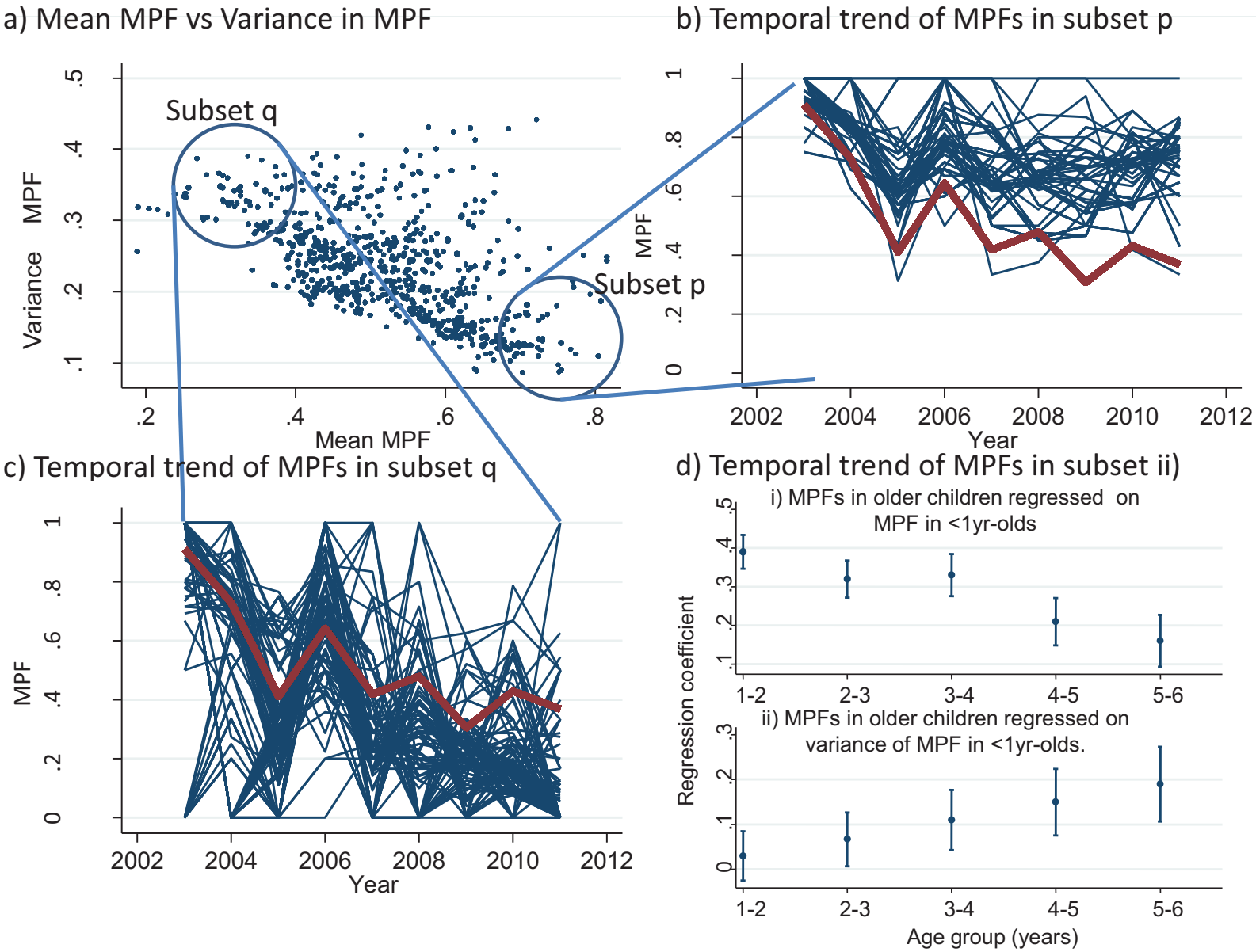
Panel a) shows the scatter plot of individual homesteads by mean malaria positive fraction (MPF) on the x axis vs variance in MPF on the y axis (r_s =-0.61, $p<0.0001$). A labelled blue circle indicates subset q (homesteads with high variance but low mean MPF) and subset p (homesteads with low variance and high mean MPF). The temporal trends for these two subsets are shown on panels b) and c), respectively. The median trend for the study area is shown in red. Panel d) shows the regression coefficients (y axis) for the malaria positive fractions (MPF) in older children when regressed on; i) the mean MPF in children <1yr of age ($MPF_{<1y}$) and ii) MPF in older children when regressed on the variance in $MPF_{<1y}$ over the 9 years of the study. Separate multivariable regression models (i.e. with mean $MPF_{<1y}$ and variance in $MPF_{<1y}$ as explanatory variables) are fit for each age group as shown on the x axis (excluding children <1yr of age, whose data are used to calculate $MPF_{<1y}$).

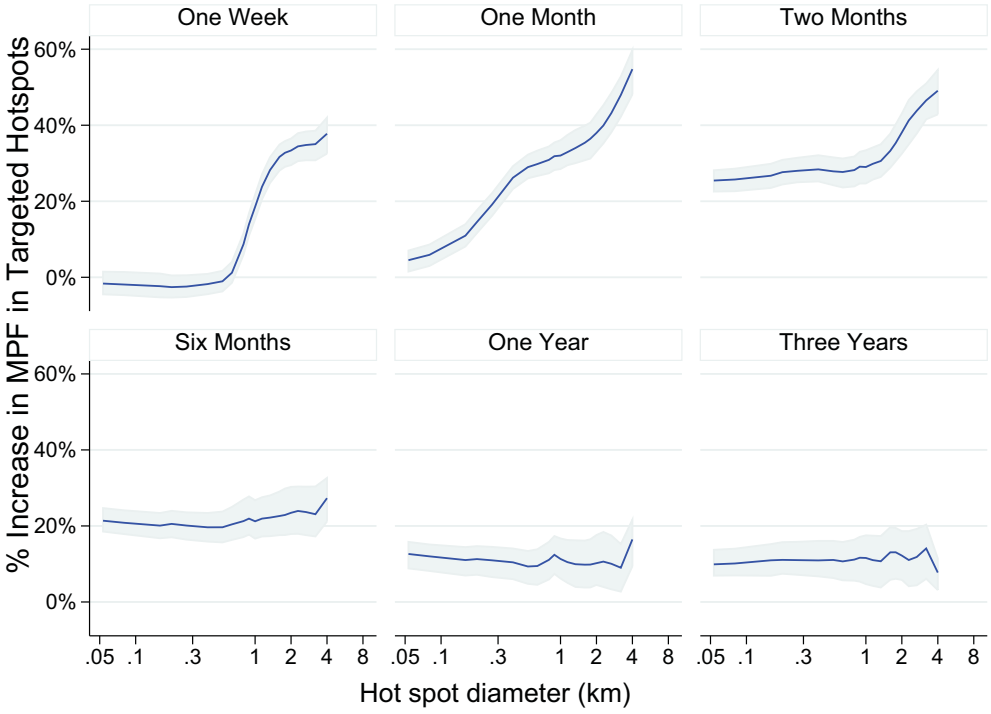
Figure Legend 4

The accuracy of varying strategies of hotspot identification is shown. Each panel is labeled with the time period of surveillance data used. The x axis shows the diameter of hotspot defined. In each case hotspots were selected to account for 20% of the homesteads in the area. The y axis shows the increase that would have been present assuming that they were targeted in the time period following their identification.


a)
Kilometers
0 0.5 1 2 3 4
MPF
=1.0
=0.5
=0
b)
Kilometers
0 0.5 1 2 3 4
Age (mths)
=120
=53
=16
c)
Average Age (Yrs)
0 2 4 6 8 10 12
0 .2 .4 .6 .8 1
Malaria Positive Fraction (MPF)
d)
r_s
.1 0 -.1 -.2 -.3
Observed data (mean and 95%CI)
Randomly permuted data (mean and 95%CI)
0 2 4 6 8
Cell size (km)


a)
Kilometers
0 0.5 1 2 3 4
c)
MPF
b)
c)
Kilometers
0 0.30.6 1.2 1.8 2.4
b)
Kilometers
0 0.30.6 1.2 1.8 2.4
d)
d)
Meters
0 60 120 240 360 480


a) Mean MPF vs Variance in MPF
Subset q
Subset p
Variance MPF
Mean MPF
b) Temporal trend of MPFs in subset p
MPF
Year
c) Temporal trend of MPFs in subset q
MPF
Year
d) Temporal trend of MPFs in subset ii)
i) MPFs in older children regressed on MPF in <1yr-olds
ii) MPFs in older children regressed on variance of MPF in <1yr-olds.
Regression coefficient
Age group (years)


One Week
One Month
Two Months
Six Months
One Year
Three Years
% Increase in MPF in Targeted Hotspots
60%
40%
20%
0%
.05
.1
.3
1
2
4
8
Hot spot diameter (km)