

# Mapping malaria risk in low transmission settings: challenges and opportunities

Hugh JW Sturrock, Adam F Bennett, Alemayehu Midekisa, Roly D Gosling, Peter W Gething, Bryan Greenhouse

Correspondence to: [Hugh.Sturrock@ucsf.edu](mailto:Hugh.Sturrock@ucsf.edu) (Sturrock, H.J.W.)

## Abstract

As malaria transmission declines, it becomes increasingly focal and prone to outbreaks. Understanding and predicting patterns of transmission risk becomes an important component of an effective elimination campaign, allowing limited resources for control and elimination to be targeted cost-effectively. Malaria risk mapping in low transmission settings is associated with some unique challenges. This article reviews the main challenges and opportunities related to risk mapping in low transmission areas including recent advancements in risk mapping low transmission malaria, relevant metrics and statistical approaches and risk mapping in post elimination settings.

## Malaria risk mapping

The sophistication of approaches to mapping malaria risk has grown dramatically over the past 15 years [1-5]. This has enabled spatio-temporal patterns of risk to be quantified with progressively more accuracy and at finer levels of detail, allowing for improved national and global estimates of burden and of populations at risk of malaria, facilitating resource allocation. Understanding spatio-temporal patterns of risk is particularly important in low transmission settings, where malaria transmission becomes increasingly rare, often clusters into hotspots [6-9], and results in outbreaks. In order to remain cost-effective, programs in these settings need to transition from universal coverage of interventions to a more targeted approach. This shift in approach is especially relevant and important for malaria elimination programs which often face dwindling resources as the malaria burden declines.

Metrics and methods to map malaria risk in moderate and high transmission settings are well established. Far less attention has been given to low transmission settings, where traditional metrics used for risk mapping become less useful. An estimated 25% (811 million) of the total global population at risk of malaria in 2014 live in one of the 35 malaria eliminating countries [10, 11], with many more living in low transmission areas of countries in the control phase. Risk mapping in low transmission areas can support improved targeting of remaining clusters of malaria and help programs reduce malaria burden and eliminate the disease. However, the issue of how to produce risk maps in low endemic and elimination settings requires attention. It is also not clear how malaria risk mapping should be conducted in areas once local transmission has been interrupted. Yet, risk maps are needed in settings where malaria has been eliminated but remain receptive to malaria (areas that have the potential for malaria transmission because the vector is present) or are vulnerable to transmission (areas that have a high rate of importation of malaria parasites in human migrants). Here we review available metrics for evaluating malaria risk in low transmission and post elimination settings, and describe the assumptions, limitations and optimal modeling approaches when using each to estimate and predict risk.

## **Available transmission metrics for low transmission risk mapping**

### *Entomological Inoculation Rate*

There are a number of indicators which act as proxies for transmission (Table 1). Many of these measure different aspects of transmission and are related to each other, albeit in complex and nonlinear ways [12]. The entomological inoculation rate (EIR) is considered by some to be the gold standard for estimating transmission. In practice, measurements of EIR are labor intensive and difficult to standardize, requiring capture of sufficient numbers of mosquitos and subsequent examination for sporozoites. EIR is particularly challenging, if not impossible, to estimate accurately in low transmission settings, where the density of mosquitoes can be extremely low [13, 14]. Finding a malaria vector, yet alone an infected vector, is near impossible in many low transmission settings. Other mosquito based indicators, such as vector presence, density, and human biting rate are not directly correlated to transmission although they are useful for program decision making [12].

### *Parasite rate*

Analyses of spatio-temporal patterns of malaria transmission have commonly focused on the use of parasite rate (PR, infection prevalence) data [1, 5, 15]. Parasite rate is relatively easily captured via cross-sectional surveys, using microscopy, rapid diagnostic tests (RDT), or molecular diagnostics. Furthermore, studies have shown a predictable relationship between PR and EIR [16]. As transmission declines, however, the sample sizes required to generate risk maps with acceptable uncertainty become unfeasibly large, even if surveys are targeted to areas of transmission or when more sensitive molecular diagnostics are used to detect infections [17, 18]. For example, the Swaziland malaria indicator survey (MIS) conducted in 2010 collected blood samples from 4 330 individuals, with only one *Plasmodium falciparum* infection found by polymerase chain reaction (PCR) [19]. The point at which PR is no longer a suitable indicator for measuring and mapping transmission is dependent on a number of factors including the financial and operational implications of sample sizes required; historically when prevalence is <3% alternative metrics, such as clinical incidence, have been advised [20].

#### *Serology*

Serological methods - the detection of human antibodies against malarial parasites – have the potential to address some of the limitations of PR in low transmission settings. Antibody responses to *Plasmodium* antigens provide a record of exposure to parasites, providing information about past as well as present transmission. As such, evaluation of antibody responses in a community provides a more sensitive measure of exposure that can be mathematically converted into estimates of transmission and may allow for more precise estimates and/or smaller sample sizes than those obtained from PR in low and very low transmission settings. Studies have shown that antibody seroconversion rates, derived from age-stratified seroprevalence data, correlate well with EIR across a wide range of transmission and are able to detect changes in transmission over time [21, 22]. Newer analytical methods have recently been proposed to take full advantage of information reflecting antibody titer often available in the laboratory data, instead of reducing responses to binary values by choosing a cutoff, the choice of which is not always straightforward [23, 24]. These methods may offer increases in precision, but require further validation and are dependent on being able to consistently transform raw data (e.g. optical density or mean fluorescence intensity) into relative quantification of antibody titer.

A limitation of most established serological methods is that the traditional antibodies examined (e.g. apical membrane antigen 1 (AMA 1) and merozoite surface protein 1 (MSP1)) do not distinguish very recent exposure (e.g. within the last year) from more distant exposure. Changes in exposure over time may leave their mark on the shape of the relationship between seroprevalence and age, but these are

most easily detectable if changes occur over a relatively discrete time period and exposure is well distributed across age groups. Thus, while established serological methods are useful for inferring historic and medium-term risk of exposure to parasites, their utility for mapping current risk is dependent on the spatio-temporal stability of malaria transmission, i.e. the persistence of transmission at a given location over time. This is because evidence of historic exposure is not useful in detecting transient hotspots, by definition, and also because individuals' higher cumulative exposure within a stable vs. transient hotspot is more likely to result in a detectable antibody signal. Evidence from moderate transmission settings suggests that at least some transmission 'hotspots' are relatively stable over time [8, 25, 26]. Similar evidence from low transmission settings is lacking, although data from Swaziland, a country on the verge of malaria elimination, suggests that while local cases appear to be confined to the wetter, low lying areas of the country, there is considerable geographic variation between years. While serological 'hotspots' identified during the 2010 MIS have accurately predicted some areas of previously unidentified risk, many cases also appear outside of these areas (Figure 1).

One approach to extend the utility of serological measures for predicting transmission is to model them with environmental and climatological variables, as done recently in Ethiopia [27]. That said, as established serological markers are unable to cleanly distinguish between historic and recent exposure, it remains unclear at what temporal resolution the data represent. New serological markers and analytical methods that are able to more precisely quantify exposure over different timescales, including measurement of very recent exposure, offer an exciting potential to better understand spatio-temporal patterns of risk [28, 29]. Different antigenic targets may also provide qualitatively different information. For example, antibody responses against a parasite protein exclusively produced during the preerythrocytic stage of the life cycle will be more likely to reflect the timing of mosquito to human transmission, while responses against blood stage parasites will also be influenced by the duration of chronic infections.

The use of serology for mapping current transmission risk in low transmission settings shows promise, but requires further evaluation and development of sensitive and specific markers of very recent exposure. That said, serology has the unique attribute of being able to provide retrospective information on transmission patterns. Once a sero-marker for recent exposure in low transmission settings can be rigorously validated, serology can easily and cheaply be incorporated into monitoring and evaluation efforts and can be used to evaluate the impact of interventions [30, 31].

*Clinical surveillance data*

116 An alternative to the aforementioned approaches is to use clinical surveillance data, such as annual  
117 parasite index (API) which refers to the numbers of clinical cases of malaria reported at health facilities  
118 per person per year. This metric has the advantage that it is routinely collected as part of passive  
119 surveillance systems, in contrast to EIR, PR, and serology which require some level of dedicated active  
120 surveillance activities. In moderate to high transmission settings the relationship between API and  
121 transmission intensity is confounded by the relationship between exposure and acquired immunity. In  
122 low transmission settings, unless there have been very rapid declines in transmission, the majority of the  
123 population are likely to have little clinical immunity against symptomatic disease. In these settings API is  
124 likely to be more closely associated with transmission intensity [32]. Additionally, the continuous  
125 collection of clinical surveillance data means that maps can be updated in real time, providing up to date  
126 information on where transmission is currently occurring. This is particularly useful in environments  
127 where transmission is not spatially stable over time, such as those prone to epidemic transmission.

128 There are several factors that affect the accuracy with which API reflects transmission risk. First, routine  
129 case data often don't differentiate between clinically diagnosed cases and those confirmed by diagnostic  
130 tests. Fortunately, since the launch of WHO's Test Treat Track (T3) campaign  
131 ([http://www.who.int/malaria/publications/atoz/test\\_treat\\_track\\_brochure.pdf?ua=1](http://www.who.int/malaria/publications/atoz/test_treat_track_brochure.pdf?ua=1)), the proportion  
132 of cases that are correctly confirmed via diagnosis with RDT or microscopy is increasing.

133 Second, information on whether identified cases are acquired locally or are imported is often not  
134 available. Since risk maps ideally reflect local transmission, failure to make this distinction when  
135 generating risk maps could lead to spurious results. This issue is particularly important in very low  
136 transmission settings, where the proportion of cases that are imported may be substantial [33, 34].

137 Where cases are classified by origin, e.g. as estimated by recent travel history, it may be possible to use  
138 mathematical modeling to estimate receptivity and malariogenic potential [35], although these methods  
139 require further validation and a mechanism to incorporate infections undetected by existing surveillance  
140 systems. Incorporating information on human movement may provide a mechanism to estimate  
141 importation rates and adjust risk maps generated using passive and/or active surveillance data  
142 accordingly. Such an approach has been taken at the national scale in Kenya and Namibia [36, 37].

143 However, estimating international human movement is challenging and movement patterns of the  
144 population at large may not be representative of those individuals importing malaria infections. The  
145 ability to track parasite movement by genotyping may soon offer an additional solution to understand

146 parasite population structures and origins of infections, but work to establish the optimal sampling  
147 strategies, genetic markers, and analytical methods is in its infancy [38].

148 Third, data reporting is often incomplete, leaving gaps in space and time. Recent technological  
149 innovations, such as the introduction of mobile technology (Short Message Service (SMS), Unstructured  
150 Supplementary Service Data (USSD) and internet), have helped to improve completeness and timeliness  
151 of reporting, but these tools are not yet widely implemented [39-41]. Where data are incomplete at  
152 random (spatially and temporally), this can be accounted for during the modeling process [42, 43]. If,  
153 however, data are missing in a more systematic way, analyses need to be interpreted with caution or  
154 restricted to areas for which the data quality allow.

155 Fourth, since not everyone with malaria will seek treatment, due to individuals being asymptomatic or  
156 unwilling or unable to seek treatment when symptomatic, passive surveillance data typically  
157 underestimate true burden. As treatment seeking behavior may vary spatially, applying blanket  
158 corrections to country level data is likely to be overly simplistic. More detailed spatial patterns of  
159 treatment seeking behavior can be estimated from data collected as part of population level surveys  
160 [44, 45]. Such data allow a more accurate estimate of the denominator from which clinical incidence is  
161 estimated, in turn improving risk estimates. Where estimates of the population denominator are  
162 thought to be uncertain, the test positivity rate (proportion of malaria tests that are positive) (TPR) may  
163 be a useful alternative metric to use. While open to influence from variations in testing rates between  
164 facilities, and trends in non-malarial fevers, TPR has shown to have a good correlation with other  
165 transmission metrics [46, 47].

166 Fifth, while some countries now collect and report clinical incidence data from community health  
167 workers, routine case data are often only available in aggregate form, i.e. at the level of the health  
168 facility or district. This is occasionally true for other types of data as well, but is a particular issue for  
169 incidence data since these data collection systems have not been purposefully designed to map risk.  
170 Consequently, researchers and control programs are often limited to modeling and predicting incidence  
171 at this spatial scale [39, 48, 49]. While mapping at coarse scales offers insight into broader scale  
172 transmission patterns of malaria, this approach ignores the inherent heterogeneity of transmission  
173 within health facility catchments or districts, precluding identification and targeting of hotspots and  
174 associated risk factors at smaller scales. Sturrock et al. [45] adapted a Bayesian hierarchical approach  
175 originally applied to species distribution modeling [50] to produce fine scale malaria risk predictions  
176 from health facility level data in Swaziland [45]. A similar approach using log-Gaussian Cox process

models, which model the underlying fine scale risk as a spatially continuous intensity surface, has been applied to aggregate level data for other diseases and may be appropriate for mapping malaria as well [51, 52].

Where household or village locations of individual cases are available generating high resolution risk maps is theoretically more straightforward. Using random forests models [53], Cohen et al. produced high resolution risk maps of malaria in Swaziland using individual case data [54]. Tatem et al. took a similar approach to produce a risk map in Namibia [37]. For these types of models to be used, information on controls is required. As data on the locations on controls (i.e. individuals known to not have malaria) are not typically available, this requires the generation of “pseudo-controls” or background points. The selection of pseudo-controls requires careful thought and is a focus of investigation [55, 56]. Pseudo-controls can be generated by sampling points with probability proportional to the population density to reflect the distribution of the underlying population [54]. However, as noted above, spatial variation in treatment seeking behavior has to be accounted for to ensure that controls are representative of the population from which cases arose.

Another issue related to the use of models that require pseudo-controls is that prediction values represent a relative rather than absolute measure of risk and are therefore difficult to convert to or compare with traditional epidemiological metrics. Furthermore, this index of suitability is influenced by the number of background points used (Figure 2). This has been well described by others [57] and is a familiar issue in the analysis of traditional case-control data [58]. An alternative approach is to consider the case data as a realization of a point process and use models designed specifically for these types of data [59]. Point process models attempt to estimate the intensity surface from which the cases arise (see [60] for an excellent review) and therefore allow a direct measure of risk without the need for control data. While point process models have been used for the study and prediction of other diseases [61], to our knowledge they have yet to be applied to malaria data.

## **Risk mapping post-elimination**

In addition to mapping current risk based upon active and passive infection data, in areas approaching elimination and those that have recently eliminated, risk mapping is an important tool for understanding the susceptibility of an area to resurgence or reintroduction. The risk of reintroduction of malaria has historically been evaluated using the concept of “malariogenic potential”, which is itself the product of

receptivity and vulnerability [62-64]. Receptivity refers to the overall transmission potential of an area based upon the presence and abundance of anopheline vectors and ecological and climatic conditions permitting parasite development [63]. Vulnerability, on the other hand, refers to the propensity of an area to exposure to new infections, via imported infections or infected mosquitoes [34, 65].

Receptivity has traditionally been operationalized using the basic reproductive number,  $R_0$ , to represent underlying transmission potential in the absence of control, and  $R_c$ , the reproductive number under control. Both  $R_0$  and  $R_c$  are difficult to measure precisely, and are usually modeled based upon existing measurements of vectorial capacity, EIR, parasite prevalence, and incidence [64].

Mapping  $R_0$  or intrinsic transmission potential can be estimated using historical transmission data, as in Noor et al [66], who used historical maxima in parasite prevalence in northern Namibia to predict maximum transmission potential at 5x5km resolution. However, the authors were unable to incorporate data predating the introduction of vector control efforts and therefore likely underestimate true transmission potential. That said, this approach allows a proxy for historical transmission potential, especially in areas with historical data predating control. Changes in the intrinsic transmission potential over time due to climate, economic, ecological, and population changes can only be estimated if high resolution data on these factors can be linked in space and time to historical infection data.

Receptivity can also be estimated and mapped from measures of presence and density of anopheline mosquitoes, using the concept of vectorial capacity. Vectorial capacity is an expression of the number of potentially infective bites that originate daily from a case of malaria, and can be measured in a given area from the longevity and density of vectors. Romi et al [67] assessed the receptivity of central Italy using collections of larval and adult anophelines, models of seasonal climatic suitability, and calculation of vectorial capacity. Using these data the authors produced predictive weather-based maps of the distribution of adult anophelines. A number of other studies have produced maps of predicted anopheline density based upon models of larval data, human biting rates, and satellite derived remotely sensed data [68, 69].

One methodological issue that must be considered when modeling and mapping vector data is the sampling approach, given that vector data are traditionally preferentially sampled in locations where vectors are known to exist. Such preferential sampling is particularly common in low transmission settings where anopheline densities are very low and often highly clustered. Entomological sampling in such settings is therefore only cost-effective if targeted in space and time. If not accounted for during



the modeling process, this sampling bias can impact predictions [70, 71]. Using simulations and observational biomonitoring data from Spain, Diggle *et al.* approached this problem by considering the sampling locations as an inhomogeneous point process dependent on the spatial process of interest [71]. In turn, this allows unbiased estimation of the spatial process itself. Pati *et al.* applied similar methods to environmental data in eastern USA [72].

Given receptivity, the vulnerability of an area to reintroduction of malaria is dependent upon the rate of imported infections into an area. Introduction of parasites into receptive areas, i.e. areas with populations of viable mosquito vectors, can cause resurgence events, as seen in Greece, Sri Lanka, Turkmenistan, and Zanzibar [34]. Understanding which areas are most vulnerable (i.e. prone to importation of new parasites) is also vital for post-elimination control efforts. Basic estimates of vulnerability can be established using proximity to transmission areas as well as by recording the locations where imported infections reside. More complex estimates of parasite movement and vulnerability can be inferred from human movement data and transmission models [37, 73]. Survey data that include travel histories can provide an estimate of the proportion of individuals in an area traveling to other areas with transmission, which can then be used to map estimates of vulnerability at finer spatial scale. Cell phone data, where available, can help to generate estimates of population movement between areas including linkages between areas without transmission and those with transmission ([36, 37]). However these data currently only allow estimates of movement within national boundaries as many cell phone providers do not cross national borders, and their spatial scale is dependent upon the density of cell phone towers.

## **Concluding Remarks and Future Perspectives**

While the strengths and weaknesses of different potential metrics have been discussed, in practice, the optimal indicator for risk mapping will depend on the quality and reliability of the available data in any given setting. Despite improvements in the quality of routine case data over the last decade, in many settings the data are of insufficient quality to be reliable. Moving forward, where possible, programs

operating in low transmission settings should consider collecting information which can be used to provide location information on cases. For example, in Thailand and Zanzibar, village name is recorded at the facility and linked to a georeferenced database of villages [40]. In elimination provinces in the Solomon Islands and Vanuatu, cases are linked to a georeferenced database of households at the time of diagnosis [74]. In other low transmission settings, such as Swaziland, Zambia and Zimbabwe, cases are geolocated using GPS devices during active case investigation at household level [6, 75]. Collecting location information on controls, i.e. those who test negative, can also be useful and may help to better understand spatial variation in treatment seeking rates as well as map and predict test positivity rates.

Collecting information on travel history, which can be used to distinguish local from imported cases, is critical to distinguish patterns of receptivity from vulnerability. Ideally case classification should be done at the time of diagnosis so as to avoid missing data. While there is currently no standardized method to classify cases as local or imported, recent efforts to identify the origin of infections using parasite genotyping may offer a gold standard against which different approaches can be tested [76-78]

An additional approach to improve the utility of routinely collected data for risk mapping is to establish a network of sentinel surveillance sites. This would involve selection of a set of facilities at which enhanced data collection is conducted and rigorously monitored for quality, with additional training and support provided where required [79]. While this will result in sparser data spatially, improving the quality and amount of information collected at sentinel sites would allow for more accurate predictions from spatial models. Additionally, sentinel sites provide examples of best practices which can positively influence the wider health system. Sentinel sites could be randomly selected, or could be more carefully selected to provide an optimal spatial configuration for spatial modeling, as done in environmental monitoring [80].

In settings where both serology and case data are of good quality, it may be possible to combine these indicators to allow for a more complete picture of transmission. Similar efforts have been applied to schistosomiasis [81] and loiasis [82], whereby different indicators resulting from different diagnostic approaches were modeled together to produce predictions of infection prevalence.

For countries in both elimination and prevention of reintroduction stages, obtaining robust entomological data relating to the distribution and densities of vectors is central to the long term success of campaigns. Given the low densities of mosquitos, targeted surveillance at high risk times and

296 locations is likely to represent the most cost-effective approach, but requires careful interpretation and  
297 analyses to account for the inherent sampling bias.

298 Generating malaria risk maps in low transmission settings is challenging. Infection prevalence, the  
299 indicator most widely used to describe spatial patterns of transmission, becomes increasingly  
300 impractical due to sample size requirements. Vector based indicators are challenging for similar reasons.  
301 While current serological indicators are useful for describing spatial patterns of historic, long term risk,  
302 their utility for targeting resources for prevention of transmission are uncertain. That said, identification  
303 of antibody markers of very recent transmission offer exciting opportunities for more dynamic and finer  
304 temporal scale risk mapping. Undoubtedly, routine passive surveillance data will continue to play an  
305 extremely important role in describing and predicting patterns of low transmission risk, yet these data  
306 currently suffer from issues of quality and detail. While surveillance data are steadily improving across  
307 countries, there is a need to incorporate and consider a spatial dimension. Having mechanisms and  
308 protocols to assign spatial information to surveillance data, such as village or household location of  
309 individuals, extends the usefulness of surveillance data and opens up the possibility of describing and  
310 predicting transmission risk at high spatial resolution. As programs head towards and beyond  
311 elimination, and are faced with dwindling resources and rarer and increasingly clustered hotspots of  
312 transmission, capacity to map malaria risk and target resources becomes an invaluable resource to  
313 ensure program success.

314

315 **Table 1. Suitability of different indicators of malaria transmission for risk mapping in low transmission**  
 316 **settings**

Indicator	Advantages	Disadvantages	Potential suitability for low transmission risk mapping
Entomological inoculation rate	<ul style="list-style-type: none"> <li>Considered by some to be gold standard transmission metric</li> </ul>	<ul style="list-style-type: none"> <li>Operationally challenging and expensive to obtain</li> <li>Only provides picture at one point in time</li> </ul>	Low
Parasite prevalence	<ul style="list-style-type: none"> <li>Well established and widely used method with established statistical methods and comparisons over time</li> <li>Straightforward to obtain via cross-sectional surveys</li> </ul>	<ul style="list-style-type: none"> <li>Sample size requirement in low transmission often prohibitively large, especially to obtain spatial resolution required for risk mapping</li> <li>Only provides picture at one point in time</li> <li>Cross-sectional surveys can be expensive, done infrequently</li> </ul>	Low
Serology	<ul style="list-style-type: none"> <li>Ability to integrate information on exposure over time increases sensitivity</li> </ul>	<ul style="list-style-type: none"> <li>Methods to distinguish very recent from more distant exposure in low transmission settings</li> </ul>	Medium (with currently available tools) – High (if precise)

	<p>and may allow higher precision with smaller sample sizes than PR</p> <ul style="list-style-type: none"> <li>• Ability to estimate exposure over multiple time scales may provide information on recent risk as well as suitability of an area to historically sustain transmission (vectorial capacity)</li> <li>• Incremental cost to augment active surveillance data is very low</li> </ul>	<p>are not well established</p> <ul style="list-style-type: none"> <li>• Like PR, requires the expense of dedicated collection of samples and data e.g. from cross-sectional surveys, done infrequently</li> <li>• Existing markers may not be good indicators of current and/or future transmission</li> </ul>	<p>markers of very recent exposure are developed)</p>
Clinical incidence	<ul style="list-style-type: none"> <li>• Cheap – data routinely collected at health facility level</li> <li>• Provides data over continuous time</li> </ul>	<ul style="list-style-type: none"> <li>• Quality of data highly dependent on quality of diagnoses, completeness of reporting and variations in treatment seeking behavior</li> <li>• Could lead to spurious results without information on whether cases are local or imported</li> <li>• Location data often</li> </ul>	<p>High (if data are of sufficient quality)</p>

---

restricted to health  
facility limiting spatial  
resolution of risk maps

- Only useful if spatial  
distribution of  
symptomatic cases is  
reflective of overall  
transmission, which may  
be driven by  
asymptomatic reservoirs

---

317

318

319

320 **Figure 1. Serological hotspot and malaria cases in Swaziland 2010-2013.** Locations of passively detected  
321 local cases during the malaria season 2011-2013 and primary serological ‘hotspot’ detected in Swaziland  
322 from malaria indicator survey (MIS) data collected in 2010 (red ellipse). Each point represents a single  
323 case. Figure adapted from Hsiang et al. [19] and Sturrock et al. [45].

324

325 **Figure 2. A simulated example illustrating the implications of using varying numbers of pseudo-**  
326 **controls for risk mapping case data.** Each figure represents the same geographical area. A – a simulated  
327 population surface, B – simulated malaria risk (probability of there being a case), C – simulated case  
328 locations; D – a risk map (predicted probability of there being a case) generated using a random forest  
329 model with two covariates (simulated elevation and vegetation index – not shown) with 1 000 pseudo-  
330 controls; E – as per D with 5 000 pseudo-controls; F – as per D with 10 000 pseudo-controls. Note the  
331 difference in level of risk between the simulated malaria risk (B) and its predictions using increasing  
332 numbers of pseudo-controls (D-F).

333

334

335

## 336      **References**

- 337      1 Gething, P., *et al.* (2011) A new world malaria map: *Plasmodium falciparum* endemicity in 2010.
- 338      *Malaria Journal* 10, 378
- 339      2 Hay, S.I., *et al.* (2011) Estimating the global clinical burden of *Plasmodium falciparum* malaria in 2007.
- 340      *PLoS Medicine* 7, e1000290
- 341      3 Kleinschmidt, I., *et al.* (2000) A spatial statistical approach to malaria mapping. *Int J Epidemiol* 29, 355 -
- 342      361
- 343      4 Bennett, A., *et al.* (2013) Mapping malaria transmission intensity in Malawi, 2000-2010. *Am J Trop Med*
- 344      *Hyg* 89, 840-849
- 345      5 Bhatt, S., *et al.* (2015) The effect of malaria control on *Plasmodium falciparum* in Africa between 2000
- 346      and 2015. *Nature* 526, 207-211
- 347      6 Sturrock, H.J.W., *et al.* (2013) Reactive case detection for malaria elimination: real-life experience from
- 348      an ongoing program in Swaziland. *PLoS One* 8, e63830
- 349      7 Stresman, G., *et al.* (2010) A method of active case detection to target reservoirs of asymptomatic
- 350      malaria and gametocyte carriers in a rural area in Southern Province, Zambia. *Malaria Journal* 9, 265
- 351      8 Ernst, K.C., *et al.* (2006) Malaria hotspot areas in a highland Kenya site are consistent in epidemic and
- 352      non-epidemic years and are associated with ecological factors. *Malar J* 5, 78
- 353      9 Bejon, P., *et al.* (2014) A micro-epidemiological analysis of febrile malaria in Coastal Kenya showing
- 354      hotspots within hotspots. *Elife* 3, e02130
- 355      10 Newby, G., *et al.* (2016) The path to eradication: a progress report on the malaria-eliminating
- 356      countries. *Lancet* 387, 1775-1784
- 357      11 World Health Organization, G. (2015) World Malaria Report 2015.
- 358      12 Tusting, L.S., *et al.* (2014) Measuring changes in *Plasmodium falciparum* transmission: precision,
- 359      accuracy and costs of metrics. *Adv Parasitol* 84, 151-208
- 360      13 Oesterholt, M., *et al.* (2006) Spatial and temporal variation in malaria transmission in a low
- 361      endemicity area in northern Tanzania. *Malar J* 5, 98
- 362      14 Mbogo, C.M., *et al.* (2003) Spatial and temporal heterogeneity of *Anopheles* mosquitoes and
- 363      *Plasmodium falciparum* transmission along the Kenyan coast. *Am. J. Trop. Med. Hyg.* 68, 734-742
- 364      15 Noor, A.M., *et al.* (2014) The changing risk of *Plasmodium falciparum* malaria infection in Africa:
- 365      2000–10: a spatial and temporal analysis of transmission intensity. *Lancet* 383, 1739-1747
- 366      16 Smith, D.L., *et al.* (2005) The entomological inoculation rate and *Plasmodium falciparum* infection in
- 367      African children. *Nature* 438, 492-495
- 368      17 Hay, S., *et al.* (2008) Measuring malaria endemicity from intense to interrupted transmission. *Lancet*
- 369      *Infect Dis* 8, 369 - 378
- 370      18 Jovani, R. and Tella, J. (2006) Parasite prevalence and sample size: misconceptions and solutions.
- 371      *Trends Parasitol* 22, 214 - 218
- 372      19 Hsiang, M., *et al.* (2012) Surveillance for malaria elimination in Swaziland: a national cross-sectional
- 373      study using pooled PCR and serology. *PLoS One* 7, e29550
- 374      20 Yekutieli, P. (1960) Problems of epidemiology in malaria eradication. *Bull. World Health Organ.* 22,
- 375      669-683
- 376      21 Corran, P., *et al.* (2007) Serology: a robust indicator of malaria transmission intensity? *Trends in*
- 377      *Parasitology* 23, 575-582
- 378      22 Drakeley, C.J., *et al.* (2005) Estimating medium- and long-term trends in malaria transmission by using
- 379      serological markers of malaria exposure. *Proceedings of the National Academy of Sciences* 102, 5108-
- 380      5113
- 381      23 Pothin, E., *et al.* (2016) Estimating malaria transmission intensity from *Plasmodium falciparum*
- 382      serological data using antibody density models. *Malaria journal* 15, 1



24 Yman, V., *et al.* (2016) Antibody acquisition models: A new tool for serological surveillance of malaria transmission intensity. *Scientific reports* 6, 19472

25 Bejon, P., *et al.* (2010) Stable and unstable malaria hotspots in longitudinal cohort studies in Kenya. *PLoS Med* 7, e1000304

26 Mosha, J., *et al.* (2014) Hot spot or not: a comparison of spatial statistical methods to predict prospective malaria infections. *Malaria Journal* 13, 53

27 Ashton, R.A., *et al.* (2015) Geostatistical modeling of malaria endemicity using serological indicators of exposure collected through school surveys. *Am J Trop Med Hyg* 93, 168-177

28 Helb, D.A., *et al.* (2015) Novel serologic biomarkers provide accurate estimates of recent *Plasmodium falciparum* exposure for individuals and communities. *Proceedings of the National Academy of Sciences* 112, E4438-E4447

29 Ondigo, B.N., *et al.* (2014) Estimation of recent and long-term malaria transmission in a population by antibody testing to multiple *Plasmodium falciparum* antigens. *J. Infect. Dis.* 210, 1123-1132

30 Cook, J., *et al.* (2011) Serological markers suggest heterogeneity of effectiveness of malaria control interventions on Bioko Island, equatorial Guinea. *PLoS One* 6, e25137

31 Cook, J., *et al.* (2012) Sero-epidemiological evaluation of changes in *Plasmodium falciparum* and *Plasmodium vivax* transmission patterns over the rainy season in Cambodia. *Malaria Journal* 11, 86

32 Patil, A., *et al.* (2009) Defining the relationship between *Plasmodium falciparum* parasite rate and clinical disease: statistical models for disease burden estimation. *Malaria Journal* 8, 186

33 Cotter, C., *et al.* (2013) The changing epidemiology of malaria elimination: new strategies for new challenges. *Lancet* 382, 900 - 911

34 Sturrock, H.J.W., *et al.* (2015) Tackling imported malaria: an elimination endgame. *The American Journal of Tropical Medicine and Hygiene* 93, 139-144

35 Reiner, R.C., *et al.* (2015) Mapping residual transmission for malaria elimination. *eLife* 4:e09520

36 Wesolowski, A., *et al.* (2012) Quantifying the impact of human mobility on malaria. *Science* 338, 267 - 270

37 Tatem, A., *et al.* (2014) Integrating rapid risk mapping and mobile phone call record data for strategic malaria elimination planning. *Malaria Journal* 13, 52

38 Greenhouse, B. and Smith, D.L. (2015) Malaria genotyping for epidemiologic surveillance. *Proceedings of the National Academy of Sciences*, 201507727

39 Bennett, A., *et al.* (2014) A methodological framework for the improved use of routine health system data to evaluate national malaria control programs: evidence from Zambia. *Population health metrics* 12, 30

40 Ohrt, C., *et al.* (2015) Information systems to support surveillance for malaria elimination. *The American Journal of Tropical Medicine and Hygiene*

41 Zurovac, D., *et al.* (2012) Mobile phone text messaging: tool for malaria control in Africa. *PLoS Med* 9, e1001176

42 Gething, P., *et al.* (2008) Developing geostatistical space-time models to predict outpatient treatment burdens from incomplete national data. *Geogr Anal* 40, 167 - 188

43 Gething, P., *et al.* (2007) A local space-time kriging approach applied to a national outpatient malaria dataset. *Comput Geosci* 33, 1337 - 1350

44 Alegana, V., *et al.* (2012) Spatial modelling of healthcare utilisation for treatment of fever in Namibia. *Int. J. Health. Geogr.* 11, 6

45 Sturrock, H., *et al.* (2014) Fine-scale malaria risk mapping from routine aggregated case data. *Malaria Journal* 13, 421

46 Ceasay, S., *et al.* (2008) Changes in malaria indices between 1999 and 2007 in The Gambia: a retrospective analysis. *Lancet* 372, 1545 - 1554

47 Bi, Y., *et al.* (2012) Can slide positivity rates predict malaria transmission? *Malaria journal* 11, 1-8

431 48 Alegana, V.A., *et al.* (2014) Modelling the incidence of *Plasmodium vivax* and *Plasmodium falciparum*  
432 malaria in Afghanistan 2006-2009. *PLoS One* 9, e102304

433 49 Alegana, V.A., *et al.* (2013) Estimation of malaria incidence in northern Namibia in 2009 using  
434 Bayesian conditional-autoregressive spatial-temporal models. *Spatial and Spatio-temporal Epidemiology*  
435 7, 25-36

436 50 Keil, P., *et al.* (2013) Downscaling of species distribution models: a hierarchical approach. *Methods in*  
437 *Ecology and Evolution* 4, 82-94

438 51 Diggle, P.J., *et al.* (2013) Spatial and spatio-temporal log-Gaussian Cox processes: extending the  
439 geostatistical paradigm. *Statistical Science* 28, 542-563

440 52 Li, Y., *et al.* (2012) Log Gaussian Cox processes and spatially aggregated disease incidence data. *Stat.*  
441 *Methods Med. Res.* 21, 479-507

442 53 Prasad, A., *et al.* (2006) Newer classification and regression tree techniques: bagging and random  
443 forests for ecological prediction. *Ecosystems* 9, 181-199

444 54 Cohen, J., *et al.* (2013) Rapid case-based mapping of seasonal malaria transmission risk for strategic  
445 elimination planning in Swaziland. *Malaria Journal* 12, 61

446 55 Wisz, M. and Guisan, A. (2009) Do pseudo-absence selection strategies influence species distribution  
447 models and their predictions? An information-theoretic approach based on simulated data. *BMC*  
448 *Ecology* 9, 8

449 56 Barbet-Massin, M., *et al.* (2012) Selecting pseudo-absences for species distribution models: how,  
450 where and how many? *Methods in Ecology and Evolution* 3, 327-338

451 57 Phillips, S.J. and Elith, J. (2013) On estimating probability of presence from use–availability or  
452 presence–background data. *Ecology* 94, 1409-1419

453 58 King, G. and Zeng, L. (2013) Case-Control Studies, Inference in. In *Encyclopedia of Biopharmaceutical*  
454 *Statistics, Second Edition*, pp. 1-13, Taylor & Francis

455 59 Pullan, R.L., *et al.* (2012) Spatial parasite ecology and epidemiology: a review of methods and  
456 applications. *Parasitology* 139, 1870 - 1887

457 60 Renner, I.W., *et al.* (2015) Point process models for presence-only analysis. *Methods in Ecology and*  
458 *Evolution* 6, 366-379

459 61 Ahn, J., *et al.* (2014) A space-time point process model for analyzing and predicting case patterns of  
460 diarrheal disease in northwestern Ecuador. *Spat Spatiotemporal Epidemiol* 9, 23-35

461 62 Cohen, J., *et al.* (2010) How absolute is zero? An evaluation of historical and current definitions of  
462 malaria elimination. *Malar J* 9, 213

463 63 WHO (2007) *Malaria Elimination: a field manual for low and moderate endemic countries*. World  
464 Health Organization

465 64 Moonen, B., *et al.* (2010) Operational strategies to achieve and maintain malaria elimination. *Lancet*  
466 376, 1592 - 1603

467 65 WHO (2012) *Disease surveillance for malaria elimination: an operational manual*. WHO

468 66 Noor, A., *et al.* (2013) The receptive versus current risks of *Plasmodium falciparum* transmission in  
469 Northern Namibia: implications for elimination. *BMC Infect Dis* 13, 184

470 67 Romi, R., *et al.* (2012) Assessment of the risk of malaria re-introduction in the Maremma plain  
471 (Central Italy) using a multi-factorial approach. *Malar J* 11, 98

472 68 Dlamini, S.N., *et al.* (2015) Assessing the relationship between environmental factors and malaria  
473 vector breeding sites in Swaziland using multi-scale remotely sensed data. *Geospatial health* 10, 88-98

474 69 Machault, V., *et al.* (2012) Risk mapping of *Anopheles gambiae* s.l. densities using remotely-sensed  
475 environmental and meteorological data in an urban area: Dakar, Senegal. *PLoS One* 7, e50674

476 70 Gelfand, A.E., *et al.* (2012) On the effect of preferential sampling in spatial prediction. *Environmetrics*  
477 23, 565-578

71 Diggle, P.J., *et al.* (2010) Geostatistical inference under preferential sampling. *Journal of the Royal Statistical Society: Series C (Applied Statistics)* 59, 191-232

72 Pati, D., *et al.* (2011) Bayesian geostatistical modelling with informative sampling locations. *Biometrika* 98, 35-48

73 Le Menach, A., *et al.* (2011) Travel risk, malaria importation and malaria transmission in Zanzibar. *Sci Rep* 1, 93

74 Kelly, G., *et al.* (2013) A high-resolution geospatial surveillance-response system for malaria elimination in Solomon Islands and Vanuatu. *Malar J* 12, 108

75 Larsen, D.A., *et al.* (2015) Malaria surveillance in low-transmission areas of Zambia using reactive case detection. *Malaria journal* 14, 1-9

76 Patel, J.C., *et al.* (2014) Genetic evidence of importation of drug-resistant *Plasmodium falciparum* to Guatemala from the Democratic Republic of the Congo. *Emerg Infect Dis* 20, 932-940

77 Obaldia, N., *et al.* (2015) Clonal outbreak of *Plasmodium falciparum* infection in eastern Panama. *J. Infect. Dis.* 211, 1087-1096

78 Preston, M.D., *et al.* (2014) A barcode of organellar genome polymorphisms identifies the geographic origin of *Plasmodium falciparum* strains. *Nature communications* 5, 4052

79 Yukich, J.O., *et al.* (2014) A description of malaria sentinel surveillance: a case study in Oromia Regional State, Ethiopia. *Malaria journal* 13, 1-13

80 Melles, S.J., *et al.* (2011) Optimizing the spatial pattern of networks for monitoring radioactive releases. *Computers and Geosciences* 37, 280-288

81 Sturrock, H.J.W., *et al.* (2013) The use of bivariate spatial modeling of questionnaire and parasitology data to predict the distribution of *Schistosoma haematobium* in coastal Kenya. *PLoS Negl. Trop. Dis.* 7, e2016

82 Crainiceanu, C.M., *et al.* (2008) Bivariate binomial spatial modeling of *Loa loa* prevalence in tropical Africa. *J Am Stat Assoc* 103, 21-37