Supplementary Material

Measuring the Path Towards Malaria Elimination

Table of Contents

1 Additional Elements of Discussion

Areas with *Plasmodium vivax*

The majority of countries aiming for malaria elimination are likely to be endemic for *P. vivax* (*1*, *2*). Even in Africa where the prevalence of Duffy-negative blood in the population severely restricts *P. vivax* transmission, it is thought that it will become increasingly epidemiologically important as the area approaches local elimination (*2*).

The current framework is designed to quantify transmission of falciparum malaria as it assumes that all new cases come from new infections. However, new cases of *P. vivax* can occur from a relapse of a previous infection. This is not accounted for in our approach, although the method is still applicable to *P. vivax* malaria. Relapse cases are similar to imported cases in that they are not caused by local transmission. If we knew the proportion of cases due to local transmission (i.e., cases which are neither imported nor relapsed), we could obtain unbiased estimates of R (average number of new persons infected by a person with malaria given current control interventions). In practice, however, it is not possible to determine if an autochthonous case is due to a relapse of infection or to local transmission. If we assume that all autochthonous cases of *P. vivax* are due to local transmission, we will overestimate the level of local transmission and R. The approach is therefore conservative: if the hypothesis $R \geq 1$ can be rejected under these conservative assumptions, we can be confident that $R < 1$. So, when using the method for *P. vivax*, we lose power if a substantial proportion of cases come from relapsed infections but the test remains valid and conservative.

In areas endemic for multiple species of malaria, all data should ideally be stratified by species. This will allow the epidemiology of the different infections to be disentangled and enable more precise estimates of falciparum malaria. If this is not possible, then all cases of malaria can be analyzed together, as the test remains valid irrespective of the ratio of the different species.

Spatial Heterogeneity

The question of defining the most appropriate geographical scale at which transmission should be assessed is critically important for any method aiming to ascertain levels of malaria transmission, as the infection is highly spatially heterogeneous. Too large a scale and countrywide estimates may mask the presence of more localized pockets of endemic transmission. For example, if data come from cities (where transmission is low) and other parts of the country (where transmission is high) are underrepresented, assessment of "country-wide" transmission will be biased and reflect the situation in the cities. Conversely, too small a geographical scale may make conclusions difficult because of a low number of cases. Compared with other methods, we believe that the approach we present here is well positioned to tackle the challenge of spatial heterogeneity. Indeed, other methods may typically require detailed data that may be difficult to gather at a fine spatial resolution. In contrast, standard techniques from geostatistics can be used to map the probability of a case's being imported from data consisting of the address of cases and their travelling status. This will allow the development of spatial models that reconstruct how the probability of a case's being imported varies spatially and will derive from that the map of the levels of local transmission.

Case-to-Case Heterogeneity In Infectivity

Further work is also needed to generate more accurate estimates of *k*, the case-to-case heterogeneity in infectivity. This will enable the power of the test to be improved (see section 2.3 below). Recent advances in parasite genetic sequencing (*3*) should make it easier to recreate chains of transmission and to get more precise estimates of *k* for a particular epidemiological setting. Methods such as those outlined in Section 3 of these supplementary materials can also be used to refine estimates of *k*.

Removing Cases Detected During Outbreak Investigations from the Analysis

Surveillance of infectious diseases often mixes routine surveillance where cases are detected within the community with more detailed outbreak investigations, consisting of active case searching near surveillance cases. However, it is difficult to correctly account for cases found during outbreak investigations in our statistical framework. This is because this two-step data collection process is likely to generate a selection bias toward larger outbreaks. To get an idea of the problem, one can conceptually partition cases into distinct outbreaks each of a different size. If all cases have the same probability of being detected, then larger outbreaks are more likely to be detected and investigated (because they have more opportunities of having at least one case detected by routine surveillance). Investigated outbreaks are therefore expected to be atypically large, which may bias estimates of the current level of transmission upwards. This issue is more thoroughly discussed and explored in Cauchemez *et al*. (*4*). By removing cases detected via outbreak investigation from our analysis, we ensure that our estimates of *R* remain unbiased. If we included the cases identified by active case detection in our analysis, we would expect to overestimate levels of local transmission. This means that the test for $R \geq 1$ would be expected to lose power although it would remain conservative; that is, if the hypothesis $R \geq 1$ can be rejected despite the fact that *R* is overestimated, then we can be confident that $R < 1$. In practice, including cases found during outbreak investigations in Swaziland would have had little impact on our analysis: In 2010, for example, the proportion of imported cases used to estimate *R* would have moved from 36% to 33%.

2 A Method to Ascertain Status of Controlled Nonendemic Malaria

Here, we give a technical description of our method to ascertain the status of controlled nonendemic malaria. The work focuses on *Plasmodium falciparum*, as this is the main species of malaria in Swaziland (*5*) (in section 1 of these supplementary materials, we discuss the use of the method in areas where *P. vivax* circulates).

The method uses the proportion of imported cases among malaria cases detected through routine surveillance to test the hypothesis $R \geq 1$. The status of controlled nonendemic malaria is achieved if the hypothesis $R \geq 1$ can be rejected.

The method is based on the intuitive idea, already explored in (*4*, *6*, *7*), that, if most cases detected by surveillance are imported, levels of local transmission are low. When local transmission is known to be below sustainability levels (i.e., $R < 1$), a simple formula linking the proportion *p* of imported cases to *R* [$p = 1/(1 - R)$] has been used to derive point estimates of *R*. However, formalizing it into a sound statistical test for $R \geq 1$ is challenging because the variability in the proportion of imported cases depends on parameters for which estimates may not be available: the case-to-case variability in infectiousness (i.e., the

fact that few cases may be responsible for most transmissions) (*8*); the total number of imported cases (including those missed by surveillance); and the number of local cases generated by a single importation on a finite time period (i.e., chains of transmission on finite time periods may be censored). To tackle these challenges and to ensure our approach is conservative, we consider worst-case scenarios for each of these unknown.

2.1 Total number of cases generated by a single importation

Here, we use branching process theory to model the total number of cases generated by a single importation (i.e., one importation + subsequent number of local cases).

Following Lloyd-Smith *et al.* (*8*), the offspring distribution (i.e., number of persons infected by a case) is modeled with a negative binomial distribution with mean *R* and overdispersion parameter *k* (parameter *k* inversely characterizes the variation in the number of secondary cases caused by each case of malaria).

The total number of cases, *L*, generated by a single importation after an infinite time has probability (*9*):

$$\begin{cases} g\left(L \mid R,k\right) = \dfrac{\Gamma\left(\left(k+1\right)L-1\right)}{\Gamma\left(kL+1\right)\Gamma\left(L\right)} \dfrac{k^{kL+1}R^{L-1}}{\left(R+k\right)^{\left(k+1\right)L-1}} & \text{for } L<\infty \\ g\left(\infty \mid R,k\right) = 1-\sum_{L<\infty} g\left(L \mid R,k\right) & \text{otherwise} \end{cases} \quad [1]$$

In practice, however, we work on finite time periods, and the epidemic may still be ongoing at the point when we examine the data. This is accounted for by right censoring the distribution at a certain threshold *Z* (i.e., if the total number of cases is going to be larger than *Z*, we assume it is of size *Z*):

$$\begin{cases} L^{*} = L & \text{if } L < Z \\ L^{*} = Z & \text{if } L \geq Z \end{cases} \quad [2]$$

We present tests for different values of *k* and *Z*.

Estimates of *k* in malaria are hard to derive. The spatial distribution of mosquitoes is often highly focal (*10*), particularly as an area approaches local elimination and transmission is restricted to smaller and smaller "hot spots" or populations at risk (*11*). Although this heterogeneity is increasingly being quantified (*12*, *13*), the number of cases caused by each case is far from clear and will depend on the level of endemicity, control interventions, and local mosquito behavior among others. In light of this uncertainty, we consider a scenario with high overdispersion (*k* = 0.2) and a scenario with limited heterogeneity (*k* = 1). When *R* = 1, this equates to 80% of all new infections coming from either 14% (*k* = 0.2) or 30% (*k* = 1) of cases. The lower the value of *k*, the greater the variance of *R* estimates (*4*). Therefore, we take *k* = 0.2 as the baseline scenario, as it may be seen as a worst-case scenario (*10*, *14*) and is therefore conservative. This is consistent with independent estimates of *k* calculated from the time-series data (see section 3.4).

We now want to determine sensible values for the threshold *Z* in a context where surveillance data are grouped by year. Remember that this threshold is introduced to capture the effect of right censoring that occurs in ongoing epidemics. A more refined way

to capture such censoring consists in analyzing the detailed epidemic time series, and we apply such approach in section 3 of these supplementary materials. Here, however, we aim to provide simpler approximations for the development of our operational tool that only relies on the proportion of imported cases for a given year. We approximate the situation of right censoring (i.e., the number of cases in a chain of infections is censored because transmission is still ongoing at the time of the analysis) with the following approach:

- We simulate chains of transmission from one imported case with R = 1.01, i.e., just above the critical level and with a generation time that is estimated from data (see section 3.2).
- We derive the probabilistic distribution of the size of chains of transmission that are still ongoing after 1 year.

We obtain a median size of 207 (2.5% and 97.5% percentile, 58–700) for k = 0.2 and 82 (30–258) for k = 1. The lower the Z value, the harder it is to distinguish between $R \geq 1$ and $R < 1$ and so, the smaller the power of our test (Figure S1 and Table S1). So, we can be conservative by choosing a small value for Z. This is what we do here, by specifying Z = 30 in our baseline scenario. We also consider Z = 15, 50, 100, 1000 in our sensitivity analyses (Figure S1 and Table S1). More refined approaches could be considered (for example, using the exact distribution of the size of chains derived above). Such approaches would have a better power but would be more dependent on the validity of assumptions made about the generation time and overdispersion parameter k.

2.2 Statistical procedure

We assume that each of the cases is classified either "imported" or "local." Furthermore, we assume that n cases have been independently detected by routine surveillance. It is possible that additional secondary cases were detected during epidemiological investigations (active case detection) that may have followed up the detection of the n surveillance cases. However, here, we ignore these secondary cases and restrict the analysis to the n surveillance cases (see section 1 of these supplementary materials for a discussion of this point).

We aim to use the proportion of detected cases that are classified as "imported" to test the hypothesis H_0: $R \geq 1$. This is achieved with the following procedure.

Known number of imported cases

Assume first that m, the total number of imported cases (including those that have not been detected) is known. Given m, we can derive the probabilistic distribution of the proportion of detected cases that are imported under the assumption that R = 1. This is achieved via simulations, with the following algorithm.

For simulation s = 1,…, 15,000:

- For imported case i = 1, …, m , draw the total number of cases generated by that case S_{si} from its probabilistic distribution (equations [1] and [2]).

- Calculate the total number of cases

$$S_s = \sum_i S_{si}$$

- Assuming that each case has the same probability of detection, draw the number B_s of detected cases that were imported from its distribution, which is hypergeometric (with n draws among m imported and $S_s - m$ local cases).

From these 15,000 simulations, we derive the 95% percentile $p_R\left(n \text{ detected}, m \text{ imported}\right)$ of the proportion of detected cases that were imported. If the observed proportion is above this value, we reject H_0: $R \geq 1$.

Unknown number of imported cases

In practice, the number of undetected imported cases is unknown, so the total number of imported cases *m* is also unknown. To derive a test that is conservative and robust to such uncertainty, we explicitly consider different scenarios for the total number of imported cases (*m* = 10, 15, 20, …, 100, 150, …, 500 introductions) and derive the test threshold value $p_R\left(n \text{ detected}, m \text{ imported}\right)$ for each of these scenarios. A test threshold value that is robust to uncertainty in the number of imported cases is

$$p_R\left(n \text{ detected}\right) = \max_{m=10,\ 15,\ 20,\ldots,\ 100,\ 150,\ldots\ 500} p_R\left(n \text{ detected}, m \text{ imported}\right)$$

2.3 Summary tables to ascertain controlled nonendemic malaria

The proportions of imported cases above which the hypothesis H_0: $R \geq 1$ can be rejected, as a function of the number of detected cases, are shown in Fig S1 and Table S1. Results are presented for different thresholds (*Z*) after which the size of chains of transmission is supposed to be censored (*Z* = 30 corresponds to the baseline scenario). This is done under the conservative assumption of important overdispersion in the offspring distribution (*k* = 0.2). Table S2 presents the same information but for hypothesis H_0: $R \geq 0.5$. Table S3 shows how power would improve in contexts where overdispersion in the offspring distribution is known to be more limited (*k* = 1). So, obtaining precise estimates of *k* will be important to improve power further, but our approach has the benefit that it remains valid and conservative as long as $k \geq 0.2$ (which covers most infectious diseases).

The methods are also equally applicable to other infections approaching local elimination.

Figure S1. *P*roportions of imported cases above which the hypothesis H_0: $R \geq 1$ can be rejected, as a function of the numbers of detected cases. Results are presented for different thresholds (*Z*) after which the size of chains of transmission is supposed to be censored, be it 15 (red), 30 (blue), 50 (green), 100 (purple) or 1000 (orange). This is done under the conservative assumption of important overdispersion in the offspring distribution (k = 0.2).

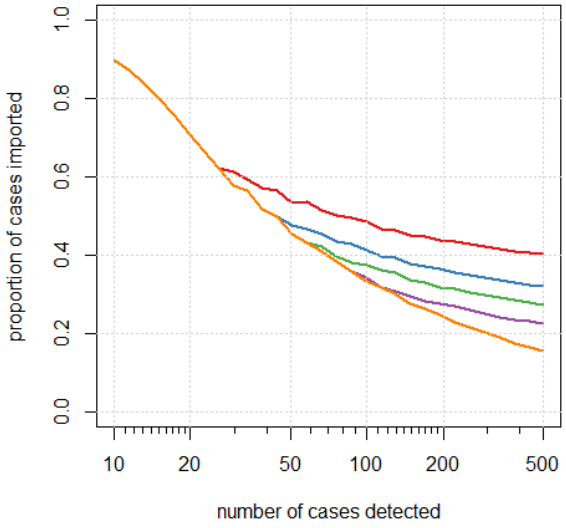


Table S1. *P*roportions of imported cases above which the hypothesis H_0 : $R \geq 1$ can be rejected, as a function of the numbers of detected cases (rows). Results are presented for different thresholds (*Z*) after which the size of chains of transmission is supposed to be censored (columns). This is done under the conservative assumption of important overdispersion in the offspring distribution (k = 0.2).

Number of detected cases	Threshold Z after which the size of chains of transmission is supposed to be censored				
	15	**30 (baseline)**	**50**	**100**	**1000**
10	0.90	0.90	0.90	0.90	0.90
20	0.70	0.70	0.70	0.65	0.65
30	0.60	0.57	0.57	0.57	0.57
40	0.58	0.53	0.50	0.50	0.50
50	0.54	0.48	0.46	0.46	0.46
60	0.53	0.47	0.43	0.43	0.42
70	0.51	0.44	0.40	0.40	0.40
80	0.50	0.43	0.39	0.38	0.38
90	0.49	0.42	0.38	0.36	0.35
100	0.48	0.41	0.37	0.34	0.34
150	0.45	0.38	0.34	0.29	0.28
200	0.44	0.37	0.32	0.28	0.25
250	0.42	0.35	0.31	0.26	0.22
300	0.42	0.33	0.29	0.25	0.20
350	0.41	0.33	0.28	0.24	0.19
400	0.41	0.33	0.28	0.23	0.17
450	0.40	0.32	0.28	0.23	0.16
500	0.40	0.32	0.27	0.22	0.15

***Table S2: P*roportions of imported cases above which the hypothesis H_0 : $R \geq 0.5$ can be rejected, as a function of the numbers of detected cases (rows).** Results are presented for different thresholds (*Z*) after which the size of chains of transmission is supposed to be censored (columns). This is done under the conservative assumption of important overdispersion in the offspring distribution ($k = 0.2$).

Number of detected cases	Threshold Z after which the size of chains of transmission is supposed to be censored			
	15	**30 (baseline)**	**50**	**100**
10	1.00	1.00	1.00	1.00
20	0.90	0.90	0.90	0.90
30	0.83	0.83	0.83	0.83
40	0.80	0.80	0.78	0.80
50	0.76	0.76	0.76	0.76
60	0.75	0.75	0.75	0.75
70	0.73	0.73	0.73	0.73
80	0.71	0.71	0.71	0.71
90	0.71	0.70	0.70	0.70
100	0.70	0.69	0.69	0.69
150	0.66	0.65	0.65	0.65
200	0.66	0.64	0.64	0.64
250	0.64	0.62	0.62	0.62
300	0.63	0.62	0.61	0.61
350	0.63	0.61	0.61	0.61
400	0.62	0.60	0.60	0.60
450	0.62	0.60	0.60	0.60
500	0.61	0.59	0.59	0.59

Table S3. *P*roportions of imported cases above which the hypothesis H_0 : $R \geq 1$ can be rejected, as a function of the numbers of detected cases (rows) and in a context where overdispersion in the offspring distribution is known to be limited ($k = 1$). Results are presented for different thresholds (Z) after which the size of chains of transmission is supposed to be censored (columns).

Number of detected cases	Threshold Z after which the size of chains of transmission is supposed to be censored				
	15	**30 (baseline)**	**50**	**100**	**1000**
10	0.60	0.60	0.60	0.60	0.60
20	0.45	0.40	0.45	0.40	0.45
30	0.40	0.37	0.37	0.37	0.37
40	0.40	0.35	0.33	0.33	0.33
50	0.38	0.32	0.28	0.28	0.28
60	0.37	0.30	0.27	0.27	0.25
70	0.36	0.30	0.26	0.24	0.24
80	0.35	0.29	0.25	0.23	0.23
90	0.33	0.28	0.24	0.21	0.21
100	0.33	0.27	0.24	0.21	0.20
150	0.31	0.25	0.22	0.18	0.17
200	0.30	0.24	0.21	0.17	0.15
250	0.29	0.23	0.20	0.16	0.13
300	0.29	0.22	0.19	0.15	0.12
350	0.28	0.22	0.19	0.15	0.11
400	0.28	0.22	0.18	0.15	0.10
450	0.28	0.21	0.18	0.14	0.10
500	0.27	0.21	0.17	0.14	0.09

2.4 Step-by-step guide to determine whether endemic transmission has been halted

To encourage the use of the method, we have developed a simple Excel spreadsheet that uses routine surveillance data to determine whether there is evidence that endemic transmission has been halted. The work is extended to allow more stringent targets to be tested (R = 0.1, 0.2, …, 0.9). Here follows a step-by-step guide outlining how the tool should be used.

Box 1: Step-by-step guide to implement the method

i. Collate routine surveillance data by year.

ii. Remove any case identified by outbreak investigation.*

iii. For each case detected through routine surveillance, determine whether the case was autochthonous or imported from abroad. For the definition of what constitutes an imported case, please refer to the latest World Health Organization guidelines (*15*).

iv. Remove from the analysis any case with an unknown status imported/autochthonous.

v. Enter the total number of cases selected for analysis into box C8 of the Excel tool. Note that the tool requires between 10 and 500 cases a year.

vi. Enter the total number of imported cases selected for analysis into box C9 of the Excel tool.

vii. Boxes C11 to C20 will populate automatically stating whether there is evidence that R is less than the corresponding value or whether the result was inconclusive. For example, if box C11 is "TRUE," then there is evidence that $R < 1$ and that endemic transmission has been halted. If box C16 is "TRUE," then there is evidence that $R < 0.5$ for that period. If an "ERROR" message appears, check that the number of cases is 10 or more and that it is no greater than the number of imported cases.

*If this is not possible, local transmission may be overestimated. This means that the test for $R \geq 1$ may lose power although it will remain conservative (see section 1 of these supplementary materials for a discussion of this point).

3 Using Malaria Epidemic Time Series to Evaluate Seasonal Variations in Transmissibility

In this section, we describe how malaria epidemic time series can be analyzed to evaluate seasonal variations in transmissibility. In terms of public health impact, there may be only a minor difference between countries with $R < 1$ throughout the year; and those where R gets above 1 for a short high-season period. This is because a high season of 3 months (such as is observed in Swaziland) represents at most two or three generations of malaria cases (see section 3.2). As a result, R is reduced before long chains of transmission can be formed. In such situations, it may nonetheless be important to separate R estimates for high and low seasons to evaluate the effectiveness of local interventions and to predict their impact in countries with different seasonal patterns. This technique also provides a method of validating our estimates of parameter k which was used in section 2. Data consist of weekly counts of imported and local incident cases.

In subsection 3.1, we present a mathematical model for malaria importations and local transmission. Fitting this model to data requires characterizing the generation time of malaria in the area under study. This is done for Swaziland in subsection 3.2. In subsection 3.3, we present model specifications and explain how model parameters are estimated from the data. Subsection 3.4 presents the results and sensitivity analyses.

3.1 A model for malaria importations and local transmission

3.1.1 Latent importation and transmission process

We denote M_t and L_t the total numbers (i.e., detected + undetected) of imported and local incident cases in week t. The total number of incident cases in week t is $T_t = M_t + L_t$.

We assume that the total number of imported incident cases M_t in week t follows a Poisson distribution with mean λ_t^I:

$$M_t \sim \text{Pois}\left(\lambda_t^I\right)$$

The modeling of local transmission is largely inspired from Cori *et al.* (*16*). We denote R_t the instantaneous reproduction number in week t [the instantaneous reproduction number is defined as the average number of people someone infected in week t would expect to infect if conditions remained unchanged (*17*)]. The infectivity profile of infected individuals is characterized by a probability distribution $\omega(u)$, where u is the number of weeks elapsed since infection (this distribution is derived for Swaziland in subsection 3.2 below). The total infectivity of infected individuals in week t is $\Lambda_t = \sum_{u>0} T_{t-u}\omega(u)$.

The mean number of local incident cases in week t is then given by the product of total infectivity and the instantaneous reproduction number $R_t \sum_{u>0} T_{t-u}\omega(u)$. Cori *et al.* (*16*) used a Poisson distribution to characterize the associated distribution:

$$L_t \sim Pois\left(R_t \sum_{u>0} T_{t-u}\omega(u)\right)$$

However, overdispersion in the offspring distribution may be substantial and may have important implications on the dynamics and for control (*8*). Here, we follow the seminal work of Lloyd-Smith *et al.* (*8*) and extend the model of Cori *et al.* to account for such overdispersion.

Assume first that total infectivity in week *t* is equal to $\Lambda_t = 1$. Following (*8*) , we can use a Negative Binomial distribution with mean R_t and overdispersion *k* to model the number L_t of local incident cases. This is a Negative Binomial distribution $NegBin\left(r_t, p_t\right)$ where $p_t = 1/\left(1+k/R_t\right)$ is the probability of success, $r_t = R_t\left(1-p_t\right)/p_t$ corresponds to the number of failures. The distribution has density

$$P\left(L_t = x\right) = \frac{\Gamma\left(x+r_t\right)}{x!\Gamma\left(x\right)}\left(1-p_t\right)^{r_t}\left(p_t\right)^{x}$$

More generally, for $\Lambda_t \neq 1$, we can use the additive properties of the Negative Binomial distribution to derive that the number L_t of local incident cases has a Negative Binomial distribution $NegBin\left(r_t\Lambda_t, p_t\right)$ with density:

$$P\left(L_t = x\right) = \frac{\Gamma\left(x+r_t\Lambda_t\right)}{x!\Gamma\left(x\right)}\left(1-p_t\right)^{r_t\Lambda_t}\left(p_t\right)^{x}$$

3.1.2 Underreporting

Only a proportion of imported and local cases are detected by the surveillance system. We denote M_t^O and L_t^O the observed numbers of imported and local incident cases in week t (i.e., cases detected by the surveillance system).

Conditional on the total numbers of imported and local incident cases, observed numbers have Binomial distributions:

$$M_t^O \mid M_t \sim Bin\left(M_t, \rho\right) \text{ and } L_t^O \mid L_t \sim Bin\left(L_t, \rho\right)$$

where ρ is the case-detection rate.

3.2 Deriving the generation time of malaria transmission in Swaziland

Here, we derive the distribution of the infectivity profile $\omega\left(u\right)$ characterizing infectivity of cases after infection and introduced in the previous section. The timelines of falciparum malaria transmission are relatively complex, with incubation periods in both the human and the mosquito. Continuing infection also varies over time so the temporal life-history of the parasite needs to be included in the derivation of $\omega\left(u\right)$.

The infectivity profile $\omega\left(u\right)$ can be approximated by the distribution of the generation time. The generation time is the mean time interval between infection of one person and infection of the people that individual infects. The mean generation time is denoted T_g . In malaria, the generation time remains poorly understood. Field evidence suggests that untreated

infections can last a long time (*18*, *19*), with recent molecular evidence suggesting an average duration of infection of approximately 140 days (*20*). These estimates were derived in areas of intense malaria transmission, so they may differ from regions approaching local elimination, where human immunity (and therefore parasite persistence) is likely to be different.

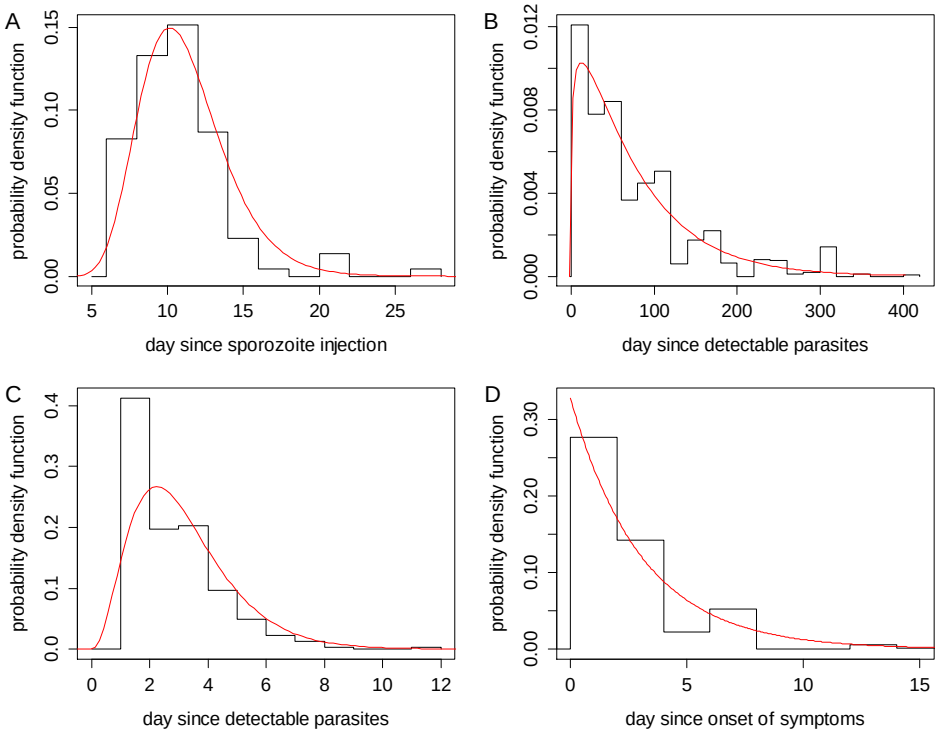
Treatment with a drug that kills transmission stages (gametocytes) will substantially reduce the generation time. In areas with access to effective antimalarial treatment, the generation time will then depend on the time till drug treatment, the type of drug used, adherence to treatment regimens, the proportion of the population who are asymptomatic (who therefore do not seek treatment) and the relative infectivity of asymptomatic versus symptomatic infections. The generation time will therefore be an amalgamation of biological factors governing parasite development and behavioral, social, and pharmacodynamics processes determining drug treatment and parasite clearance.

A list of the processes contributing to the generation time of treated and untreated cases of malaria are given in Table S4. The probability distribution of time taken for each process is either taken from the scientific literature or is estimated from experimental and Swaziland survey data. The majority of the human data comes from patients who were experimentally infected with *P. falciparum* as part of their neurosyphilis treatment (*21*). These patients were all malaria naive before infection so their immunological profiles are likely to be similar to those in areas approaching local elimination (assuming that the immunological process regulating parasite invasion and development are relatively short). A standard set of frequency distributions (normal, log-normal, gamma, Weibull, exponential) are fitted using maximum likelihood and compared using Akaike information criterion (AIC). Results are shown in Figure S2. Each shall be discussed in turn.

Table S4. List of the processes influencing the generation time of treated and untreated cases of malaria.

Untreated host	Treated host	Notation
Human Prepatient Period	Human Prepatient Period	X_1
Human Latent Period	Human Latent Period	p
Human Duration of Infectiousness	Human Duration of Infectiousness	X_2
Mosquito Latent Period	Mosquito Latent Period	n
Mosquito Duration of Infectiousness	Mosquito Duration of Infectiousness	X_3
	Human Patency to Symptoms Period	Y_1
	Human Symptoms to Treatment Period	Y_2
	Human Treatment to Parasite Clearance Period	Y_3

Figure S2. Processes influencing the generation time of falciparum malaria in a population receiving treatment. Histograms show observed data; red lines indicate the best fitting probability distribution. (**A**) The prepatient period; (**B**) patient infectiousness to mosquitoes; (**C**) onset of symptoms (temperature >101°F); and (**D**) time between onset of symptoms and starting treatment. For (A) to (C), we use data from the neurosyphilis data set (*21*); for (D), we use times to treatment from the Swaziland Demographic and Health Survey (DHS) for the time period of interest.



3.2.1 Human Prepatient Period

The time between infective mosquito bite and the appearance of asexual parasites in a person's peripheral bloodstream (as detected by microscopy) is called the prepatient period. It is estimated by fitting different distributions to data from 109 sporozoite-induced infections from the neurosyphilis data set (*21*). There was a minimum of 7 days before the first observable parasitemia. The distribution of prepatient periods is adequately described by a log-normal distribution $X_1 \sim \ln N\left(2.38, 0.254\right)$ (Figure S2A).

3.2.2 Human – Mosquito Transmission

The probability a human will transmit malaria to a blood-feeding mosquito depends on the time since infection. This can be estimated from the neurosyphilis data set where anopheline mosquitoes were fed on the skin of *P. falciparum*–infected patients different times after the appearance of asexual parasites (*21*). Mosquitoes were dissected after 7 to 8 days to determine the proportion of infected mosquitoes as measured by the presence of oocysts on the gut wall. There is a minimum of 9 days between the appearance of asexual parasites and the first mosquito infection (this period is denoted p). Thereafter the

proportion of infected mosquitoes is adequately described by a gamma distribution (Figure S2B). The distribution of times between patient parasitemia and mosquito infection in an untreated host can therefore be sampled from $X_2 \sim \text{Gamma}(1.19, 0.016)$ and by adding p.

3.2.3 Duration of mosquito infectiousness

The time taken for a blood-feeding mosquito to become infectious (the latent period or extrinsic incubation period) is typically of n = 10 days in malaria endemic areas (*22*). The average mosquito life expectancy is g = 9.5 days (*22*) and the maximum age of a mosquito is m = 30 days (*23*). It is assumed that mosquitoes die at a constant rate and are infectious for life (*22*). Therefore the time between blood feeding and infection of another host is assumed to have a truncated exponential distribution, $X_3 \sim \text{M}(g)$ taking values between n and m.

3.2.4 Duration of infectiousness to mosquitoes in treated and untreated cases

For an untreated patient the duration of infectiousness to mosquitoes will be determined by the shape of Figure S2B.

However for someone receiving treatment it will depend on the time till they develop symptoms, the time between the onset of symptoms and drug treatment and the time it takes for the drug to clear gametocytes. Each of these can be estimated separately. In the neurosyphilis study, the time Y_1 between patient infection and onset of symptoms (defined as temperature > 101°F) is shown in Figure S2C and is best described by a gamma distribution $Y_1 \sim \text{Gamma}(3.42, 1.08)$.

The distribution of times to treatment can be estimated from the Swaziland Demographic and Health Surveys (DHS). Data pooled from the study period are best described by an exponential distribution with an average of 3.05 days to treatment: $Y_2 \sim \text{M}(3.05)$ (Figure S2D). The DHS survey records time to treatment in children, which is likely to be quicker than that experienced by the overall population. This is adjusted for in the sensitivity analysis below.

Once treatment starts patients take a number of days before they are no longer infectious. Estimates of the effectiveness of different artemesinin-based combination therapy (ACT) treatment at reducing gametocyte density have been quantified using molecular methods (*24*). Results suggest gametocytes are killed at a constant rate and last for an average of 3 days from the start of treatment. For simplicity, it is assumed that human infectiousness to mosquitoes follows a similar pattern, although the relation between gametocyte density and mosquito infectivity is likely to be more complex (*25*). The time between the start of treatment and when a host is no longer infectious to mosquitoes is assumed to have an exponential distribution $Y_3 \sim \text{M}(3)$.

3.2.5 Generation time frequency distribution

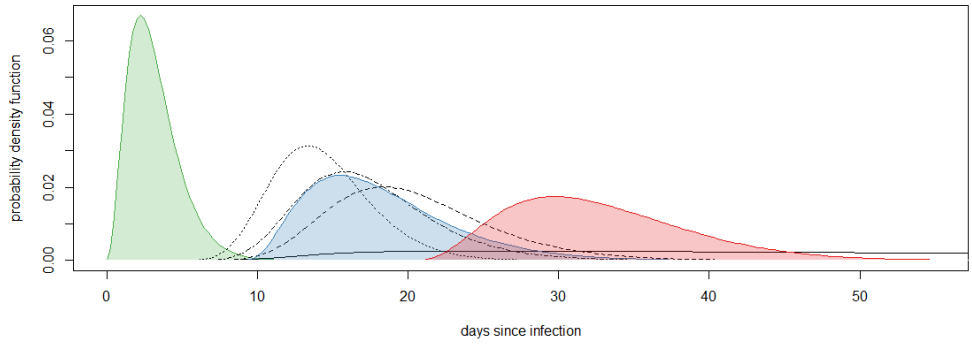
If we assume that the times described above are independent of one another, we derive the empirical distribution of the generation time for untreated and treated cases via simulation.

The generation time between a primary and secondary infection in an untreated population, U, can be generated using the following equation $U = \sum_{i=1}^{3} X_i + p + n$. The empirical density functions of U is denoted $h(t)$.

In a treated population, the generation time, V, can be generated using the same formula $V = \sum_{i=1}^{3} X_i + p + n$, but with the additional condition that the time taken for a mosquito to get infected ($X_2 + p$) must be less than the time taken for the patient to clear the infection, i.e., $(X_2 + p) \leq (Y_1 + Y_2 + Y_3)$. Simulated values that do not satisfy this condition do not contribute to the derivation of $f(t)$, the empirical density of V.

The mean generation time of a treated patient using the above parameters is 32.8 days compared to 102 days for an untreated case. A graphical representation of the life history of human-mosquito-human transmission is shown in Figure S3.

Figure S3. An illustration of the generation time of malaria in a population treated with ACT. All individuals are infected at time 0 through inoculation of sporozoites. The regions show passage of the parasite through the human, mosquito, and back to the human. The green area denotes the distribution of patient parasitemias, the blue region the distribution of mosquitoes becoming infected, and the red area is the times of secondary human infections (i.e., the infection-to-infection generation time). Black dotted line shows the distribution of the onset of symptoms, the black dot-dash lines the distribution of treatment, and the black dashed line the time of parasite clearance. The solid black line shows the relative infectiousness of mosquitoes in an untreated individual. In a treated patient, there is a brief period between a person being infectious to mosquitoes and clearing the infection, which shows how effective prompt treatment can be to transmission control. All curves are normalized.



The overall generation time will be determined by the proportion of malaria cases receiving treatment. This will depend on whether or not the patient is symptomatic. Symptomatic cases will tend to seek treatment, either through the routine health system or the private sector. These private sector distributors may or may not report back to the national health system (*26*), which might cause the proportion of cases recorded by routine surveillance (the case-detection rate) to be substantially lower than the proportion of cases receiving

treatment. Asymptomatic infections are less likely to receive treatment and, therefore, may continue to contribute to local transmission. For countries pushing for malaria elimination, the WHO recommends that outbreak investigations are carried out around every case identified by routine surveillance (*15*). This enables asymptomatic infections to be detected and treated. Outbreak investigations typically use rapid diagnostic tests (RDTs) to diagnose cases, although this may have imperfect sensitivity (*27*).

Asymptomatic patients will typically have lower gametocyte densities than symptomatic individuals, as the density of asexual parasites influences both fever and gametocyte production. This will likely reduce their infectiousness to mosquitoes. A recent study in Burkina Faso found that asymptomatic children infected on average 6.18% of mosquitoes (*28*). Comparing this with the infectivity from symptomatic infections from the neurosyphilis study, where on average 20% of mosquitoes were infected, suggests that asymptomatic hosts are on average 69% less infectiousness to mosquitoes. To account for this, we introduce a parameter ε which is the relative infectiousness to mosquitoes of asymptomatic to symptomatic infections, which has a value of 0.31.

In the context of Swaziland, routine surveillance and outbreak investigation are very efficient, and most cases of malaria receive treatment through the public health system (*5*, *26*, *29*). The case-detection rate is estimated from outbreak investigation data to 90%. Therefore, in Swaziland, it is reasonable to assume that:

1. All symptomatic cases and a certain proportion of asymptomatic cases are treated. In our baseline scenario, we assume that all cases receive treatment but consider, in a sensitivity analysis, scenarios where the proportion of untreated asymptomatic is δ = 5% and 10%. (This is only affecting the generation time distribution; see equation [3] below.)
2. The relative infectiousness of untreated asymptomatic cases (relative to untreated symptomatic cases) is ε = 0.31. (This is only affecting the generation time distribution; see equation [3] below.)
3. Untreated asymptomatic cases have the same generation time as untreated symptomatic cases.
4. Consistent with data from our outbreak investigation, we assume that the case-detection rate is ρ = 90% in our baseline scenario. In a sensitivity analysis, we also consider ρ = 50%.

In the context of Swaziland, under assumptions 1 to 3, the overall estimate of the generation time can be derived from the equation

$$T_g(t) = \frac{f(t)\delta + h(t)(1-\delta)\varepsilon}{\sum_u f(u)\delta + \sum h(u)(1-\delta)\varepsilon} \quad [3]$$

where ε is the relative infectiousness to mosquitoes of symptomatic to asymptomatic infections, δ is the proportion of malaria infections that get treated, $h(t)$ and $f(t)$ are the densities of the generation time for untreated and treated cases, respectively.

The generation time calculated above assumes that the time between onset of symptoms and treatment is, on average, 3 days. To see the epidemiological impact of a longer delay, the model is rerun with the assumption of an average of 6 days between symptoms and treatment. This delay is longer than the average time to treatment generally reported in the

literature (*30*, *31*), so the two scenarios should reflect the variability in generations that might occur in the field.

3.3 Model specifications and parameter estimation

3.3.1 Model specifications

Following (*32*), we take a flexible approach to represent temporal variations in the importation rate λ_t^I . This parameter is modeled with a stochastic diffusion from which variance is estimated (*32*). We assume that the overdispersion parameter *k* is constant over time.

Comment [BEPO1]: is this correct?

TSC: YES, it reads better

Each year, we estimate the instantaneous reproduction number R_t for the high and the low season (*33*). We assume that R_t is constant throughout each season. To test whether there is evidence that R_t has changed over the last three years, a model with a single high and low season R_t estimate is compared to one where each season has its own value.

3.3.2 Parameter estimation

We use particle Markov chain Monte Carlo (*34*) to estimate the parameters of our model from the surveillance data. We use 3,000 particles and 10,000 MCMC iterations per run, with a burn in of 500 and derive the posterior distribution of parameters, as well as the reconstructed trajectories of the latent infection process. We assume the overdispersion parameter *k* has a Uniform prior $U\left(\left[0,10\right]\right)$. For all other parameters, we use flat priors.

We use the deviance information criterion (DIC) for model comparison (*35*). Smaller values of the DIC indicate a better fit. A difference of 5 DIC units is considered to be substantial.

The first weeks of the observation period cannot be used for parameter estimation. Consider, for example, the first week of surveillance. Estimating transmissibility in that week would require knowing incident number of cases in preceding weeks (for a time period roughly covering the generation time distribution); but these data are unavailable. We therefore dedicate a first portion of the surveillance data (that covers most of the generation time distribution) to initiating the latent variables of the total numbers of imported and local incident cases, respectively. During this time period, consider, for example, M_t^0 —the observed number of imported incident cases in week *t*. We use the following formula to draw the total number of imported cases $M_t = M_t^0 + z$, where $z \sim NegBin\left(M_t^0, 1-\rho\right)$. A similar approach is used for the initialization of latent numbers of local incident cases. For the scenario where all cases are assumed to be treated (short generation time), the initialization period goes from October to December 2009, so that we can generate results for 3 years, between January 2010 and 2012. For the other scenarios where 5% or 10% of cases do not receive treatment, the initialization period is longer (from October 2009 till December 2010), and estimates are therefore only available for 2011 and 2012.

3.4 Results

Estimates of the importation rate and of the reproduction number are shown in Figures S4 and 1C, respectively. The fit of the model to data is presented in Figure S5. Sensitivity analysis is shown in Table S5 and Figure S6.

The model which allowed R_t to change every year (and season) gave a better fit (DIC = 1269.9) than the one with just a single estimate for all high and all low seasons (DIC = 1277.2), which indicates that, at the national scale, endemic transmission has significantly changed over the time period. Sensitivity analysis shows that estimates of R_t are robust to differences in the frequency distribution of generation times for treated and untreated malaria, the case-detection rate and the proportion of untreated infections (Figure S6).

Figure S4. Estimated rate of importation of malaria cases into Swaziland. Solid line indicates the best fitting model with shaded area denoting the 95% credible intervals.

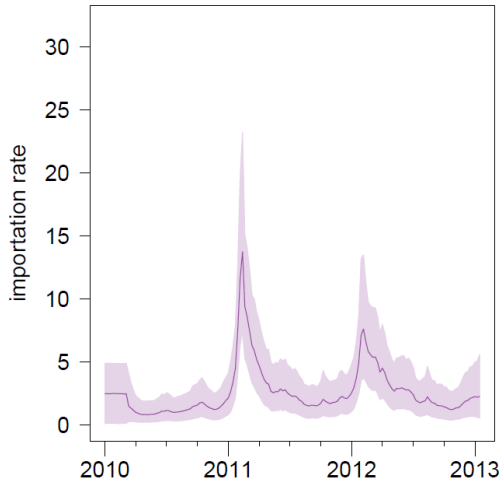


Figure S5. Model fits. (**A**) Number of imported cases and (**B**) number of locally acquired cases. Solid line indicates the posterior mean of estimates with shaded area denoting the 95% credible intervals. The points correspond to data.

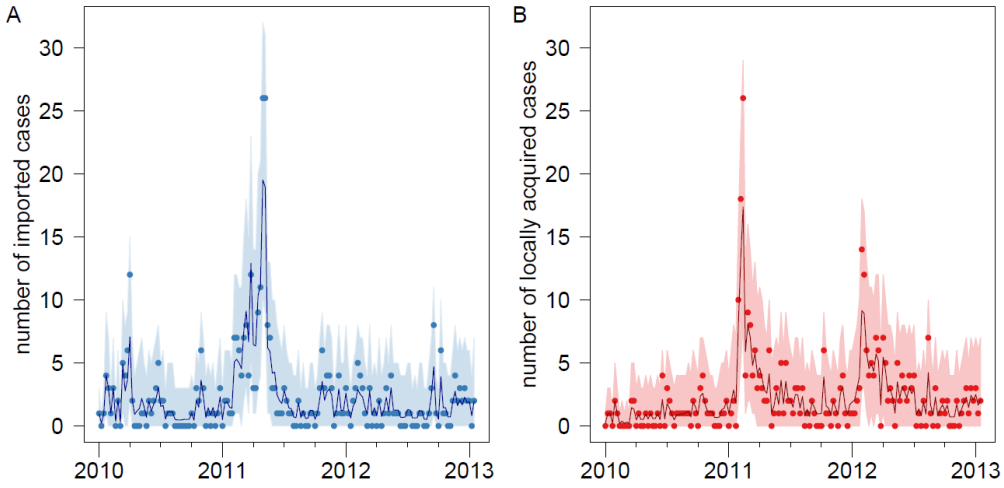
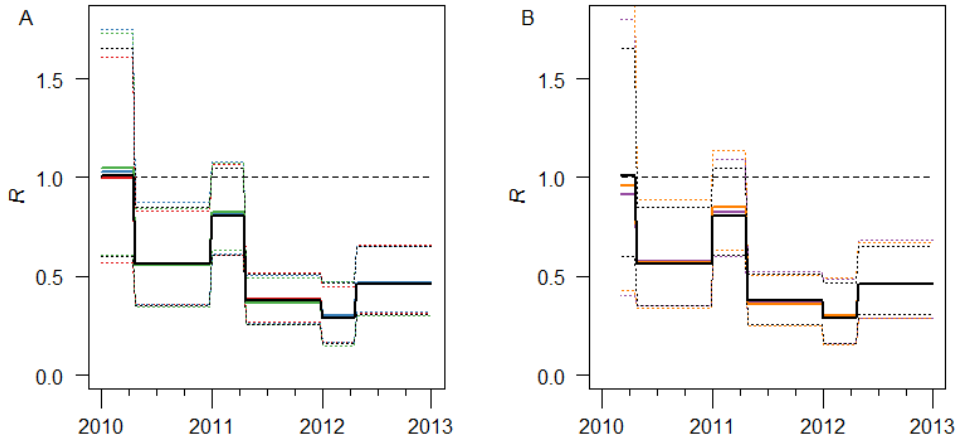


Table S5. Model sensitivity analysis. Parentheses show 95% credible intervals. Lower DIC indicates a better fit. DIC values from different proportion of the population receiving treatment are not comparable, as each were fit to different sized data sets.

Proportion of untreated cases δ	Case-detection rate ρ	Annual variations in *R*	Mean delay onset to treatment (days)	Mean generation time	Volatility of the importation rate	Overdispersion parameter *k*	DIC*
0	0.5	no	6	35.2	0.31[0.21,0.45]	0.21[0.12,0.38]	1277.9
			3	32.9	0.31[0.21,0.44]	0.21[0.12,0.42]	1278.5
		yes	6	35.2	0.31[0.2,0.46]	0.3[0.16,0.89]	1270.8
			3	32.9	0.31[0.21,0.45]	0.29[0.16,0.88]	1270
	0.9	no	6	35.2	0.29[0.19,0.44]	0.46[0.22,1.52]	1280.1
			3	32.9	0.29[0.18,0.44]	0.48[0.24,1.45]	1277.2
		yes	6	35.2	0.3[0.19,0.44]	0.99[0.32,6.92]	1271.6
			3	32.9	0.29[0.19,0.44]	1.03[0.34,9.05]	1269.9
0.05	0.5	no	6	38.5	0.31[0.19,0.46]	0.18[0.11,0.34]	1252.7
			3	36.2	0.31[0.21,0.46]	0.18[0.11,0.37]	1250.8
		yes	6	38.5	0.32[0.22,0.47]	0.24[0.13,0.55]	1246.8
			3	36.2	0.32[0.21,0.47]	0.25[0.13,0.67]	1246.8
	0.9	no	6	38.5	0.3[0.2,0.45]	0.4[0.21,1.08]	1246.4
			3	36.2	0.3[0.2,0.45]	0.41[0.21,1.35]	1246.6
		yes	6	38.5	0.3[0.19,0.45]	0.71[0.28,7.08]	1241.8
			3	36.2	0.3[0.2,0.46]	0.62[0.24,7.28]	1241.5
0.1	0.5	no	6	42.0	0.31[0.22,0.46]	0.18[0.11,0.33]	1251.9
			3	39.8	0.32[0.22,0.47]	0.18[0.11,0.35]	1250.3
		yes	6	42.0	0.32[0.2,0.47]	0.24[0.13,0.61]	1246.9
			3	39.8	0.31[0.22,0.47]	0.23[0.13,0.59]	1246
	0.9	no	6	42.0	0.3[0.18,0.45]	0.41[0.21,4.52]	1244.5
			3	39.8	0.3[0.19,0.45]	0.4[0.21,1.68]	1245.8
		yes	6	42.0	0.3[0.19,0.45]	0.71[0.27,6.04]	1240.3
			3	39.8	0.31[0.2,0.46]	0.58[0.27,4.73]	1240.7

Figure S6. Sensitivity analysis comparing the reproduction number, R_t, estimates fit using different parameters for (A) the mean delay onset to treatment and case-detection rate, and (B) the proportion of untreated cases. Solid lines show the mean estimate; dotted lines indicate the 95% credible intervals. In both panels, the black line denotes the best-fitting model with the short delay onset to treatment and 90% case-detection rate and 0% untreated infections. In (A), all lines assume 0% untreated cases and have the following parameters: red line, short-delay onset to treatment and 50% case-detection rate; blue line, long-delay onset to treatment and 50% case-detection rate and; green line, long-delay onset to treatment and 90% case-detection rate. In (B), all lines have the same parameters with the exception of the proportion of untreated cases, be it 5% (purple line) or 10% (orange line). When not visible the lines are overrun by the best-fitting model. The R_t estimates are relatively insensitive to changes in parameters.



Comment [BEPO2]: links were added to references

4 References

1. C. A. Guerra *et al*., The international limits and population at risk of *Plasmodium vivax* transmission in 2009. *PLoS Negl. Trop. Dis.* **4**, e774 (2010). Medline doi:10.1371/journal.pntd.0000774
2. World Health Organization, "World malaria report: 2013" (WHO, Geneva, 2013).
3. M. Manske *et al*., Analysis of *Plasmodium falciparum* diversity in natural infections by deep sequencing. *Nature* **487**, 375–379 (2012). Medline doi:10.1038/nature11174
4. S. Cauchemez *et al*., Using routine surveillance data to estimate the epidemic potential of emerging zoonoses: Application to the emergence of US swine origin influenza A H3N2v virus. *PLoS Med.* **10**, e1001399 (2013). Medline doi:10.1371/journal.pmed.1001399
5. M. S. Hsiang *et al*., Surveillance for malaria elimination in Swaziland: A national cross-sectional study using pooled PCR and serology. *PLoS ONE* **7**, e29550 (2012). Medline doi:10.1371/journal.pone.0029550
6. C. Chiyaka *et al*., The stability of malaria elimination. *Science* **339**, 909–910 (2013). Medline doi:10.1126/science.1229509
7. J. M. Cohen, B. Moonen, R. W. Snow, D. L. Smith, How absolute is zero? An evaluation of historical and current definitions of malaria elimination. *Malar. J.* **9**, 213 (2010). Medline doi:10.1186/1475-2875-9-213
8. J. O. Lloyd-Smith, S. J. Schreiber, P. E. Kopp, W. M. Getz, Superspreading and the effect of individual variation on disease emergence. *Nature* **438**, 355–359 (2005). Medline doi:10.1038/nature04153
9. H. Nishiura, P. Yan, C. K. Sleeman, C. J. Mode, Estimating the transmission potential of supercritical processes based on the final size distribution of minor outbreaks. *J. Theor. Biol.* **294**, 48–55 (2012). Medline doi:10.1016/j.jtbi.2011.10.039
10. T. Smith, J. D. Charlwood, W. Takken, M. Tanner, D. J. Spiegelhalter, Mapping the densities of malaria vectors within a single village. *Acta Trop.* **59**, 1–18 (1995). Medline doi:10.1016/0001-706X(94)00082-C
11. H. J. Sturrock *et al*., Targeting asymptomatic malaria infections: Active surveillance in control and elimination. *PLoS Med.* **10**, e1001467 (2013). Medline doi:10.1371/journal.pmed.1001467
12. P. Bejon *et al*., Stable and unstable malaria hotspots in longitudinal cohort studies in Kenya. *PLoS Med.* **7**, e1000304 (2010). Medline doi:10.1371/journal.pmed.1000304
13. T. Bousema *et al*., Hitting hotspots: Spatial targeting of malaria for control and elimination. *PLoS Med.* **9**, e1001165 (2012). Medline doi:10.1371/journal.pmed.1001165
14. M. E. Woolhouse *et al*., Heterogeneities in the transmission of infectious agents: Implications for the design of control programs. *Proc. Natl. Acad. Sci. U.S.A.* **94**, 338–342 (1997). Medline doi:10.1073/pnas.94.1.338
15. WHO, "Disease surveillance for malaria elimination: An operational manual" (WHO, Geneva, 2012).
16. A. Cori, N. M. Ferguson, C. Fraser, S. Cauchemez, A new framework and software to estimate time-varying reproduction numbers during epidemics. *Am. J. Epidemiol.* **178**, 1505–1512 (2013). Medline doi:10.1093/aje/kwt133

17. C. Fraser, Estimating individual and household reproduction numbers in an emerging epidemic. *PLoS ONE* **2**, e758 (2007). Medline doi:10.1371/journal.pone.0000758

18. W. Sama, G. Killeen, T. Smith, Estimating the duration of *Plasmodium falciparum* infection from trials of indoor residual spraying. *Am. J. Trop. Med. Hyg.* **70**, 625–634 (2004). Medline

19. D. L. Smith, J. Dushoff, R. W. Snow, S. I. Hay, The entomological inoculation rate and *Plasmodium falciparum* infection in African children. *Nature* **438**, 492–495 (2005). Medline doi:10.1038/nature04024

20. M. T. Bretscher *et al*., The distribution of *Plasmodium falciparum* infection durations. *Epidemics* **3**, 109–118 (2011). Medline doi:10.1016/j.epidem.2011.03.002

21. W. E. Collins, G. M. Jeffery, A retrospective examination of sporozoite- and trophozoite-induced infections with *Plasmodium falciparum*: Development of parasitologic and clinical immunity during primary infection. *Am. J. Trop. Med. Hyg.* **61** (suppl.), 4–19 (1999). Medline doi:10.4269/tropmed.1999.61-04

22. D. L. Smith, S. I. Hay, Endemicity response timelines for *Plasmodium falciparum* elimination. *Malar. J.* **8**, 87 (2009). Medline doi:10.1186/1475-2875-8-87

23. J. D. Lines, T. J. Wilkes, E. O. Lyimo, Human malaria infectiousness measured by age-specific sporozoite rates in *Anopheles gambiae* in Tanzania. *Parasitology* **102**, 167–177 (1991). Medline doi:10.1017/S0031182000062454

24. T. Bousema *et al*., Revisiting the circulation time of *Plasmodium falciparum* gametocytes: Molecular detection methods to estimate the duration of gametocyte carriage and the effect of gametocytocidal drugs. *Malar. J.* **9**, 136 (2010). Medline doi:10.1186/1475-2875-9-136

25. T. S. Churcher *et al*., Predicting mosquito infection from *Plasmodium falciparum* gametocyte density and estimating the reservoir of infection. *eLife* **2**, e00626 (2013). doi:10.7554/eLife.00626

26. J. M. Cohen *et al*., Public health. Optimizing investments in malaria treatment and diagnosis. *Science* **338**, 612–614 (2012). Medline doi:10.1126/science.1229045

27. WHO, “Malaria rapid diagnostic test performance results of WHO product testing of malaria RDTs: Round 3 (2010–2011).” (WHO, 2011).

28. A. L. Ouédraogo *et al*., Substantial contribution of submicroscopical *Plasmodium falciparum* gametocyte carriage to the infectious reservoir in an area of seasonal transmission. *PLoS ONE* **4**, e8410 (2009). Medline doi:10.1371/journal.pone.0008410

29. WHO, “Roll Back Malaria Progress and Impact Series: No. 5 Focus on Swaziland ” (WHO, Geneva, Switzerland, 2012).

30. W. Piyaphanee *et al*., Emergence and clearance of gametocytes in uncomplicated *Plasmodium falciparum* malaria. *Am. J. Trop. Med. Hyg.* **74**, 432–435 (2006). Medline

31. I. Sutanto *et al*., The effect of primaquine on gametocyte development and clearance in the treatment of uncomplicated falciparum malaria with dihydroartemisinin-piperaquine in South sumatra, Western indonesia: An open-label, randomized, controlled trial. *Clin. Infect. Dis.* **56**, 685–693 (2013). Medline doi:10.1093/cid/cis959

32. J. Dureau, K. Kalogeropoulos, M. Baguelin, Capturing the time-varying drivers of an epidemic using stochastic dynamical systems. *Biostatistics* **14**, 541–555 (2013). Medline doi:10.1093/biostatistics/kxs052

33. J. M. Cohen *et al*., Rapid case-based mapping of seasonal malaria transmission risk for strategic elimination planning in Swaziland. *Malar. J.* **12**, 61 (2013). Medline doi:10.1186/1475-2875-12-61

34. C. Andrieu, A. Doucet, R. Holenstein, Particle Markov chain Monte Carlo methods. *J. R. Stat. Soc. Ser. B.* **72**, 269–342 (2010). doi:10.1111/j.1467-9868.2009.00736.x

35. D. J. Spiegelhalter, N. G. Best, B. R. Carlin, A. van der Linde, Bayesian measures of model complexity and fit. *J. R. Stat. Soc. Ser. B* **64**, 583–639 (2002). doi:10.1111/1467-9868.00353