From multiplicity of infection to force of infection for sparsely sampled *Plasmodium falciparum* populations at high transmission

Qi Zhan[1*], **Kathryn E. Tiedje**[2], **Karen P. Day**[2], **Mercedes Pascual**[3,4,5*]

***For correspondence:**
qizhan@uchicago.edu (QZ);
mp6774@nyu.edu (MP)

Present address: [§]Department, Institute, Country; [¶]Department, Institute, Country

[1]Committee on Genetics, Genomics and Systems Biology, The University of Chicago, Chicago, IL, USA; [2]Department of Microbiology and Immunology, Bio21 Institute and Peter Doherty Institute, The University of Melbourne, Melbourne, Australia; [3]Department of Biology, New York University, New York, NY, USA; [4]Department of Environmental Studies, New York University, New York, NY, USA; [5]Santa Fe Institute, Santa Fe, NM, USA

Abstract High multiplicity of infection or MOI, the number of genetically distinct parasite strains co-infecting a single human host, characterizes infectious diseases including falciparum malaria at high transmission. It accompanies high asymptomatic *Plasmodium falciparum* prevalence despite high exposure, creating a large transmission reservoir challenging intervention. High MOI and asymptomatic prevalence are enabled by immune evasion of the parasite achieved via vast antigenic diversity. Force of infection or FOI, the number of new infections acquired by an individual host over a given time interval, is the dynamic sister quantity of MOI, and a key epidemiological parameter for monitoring the impact of antimalarial interventions and assessing vaccine or drug efficacy in clinical trials. FOI remains difficult, expensive, and labor-intensive to accurately measure, especially in high-transmission regions, whether directly via cohort studies or indirectly via the fitting of epidemiological models to repeated cross-sectional surveys. We propose here the application of queuing theory to obtain FOI on the basis of MOI, in the form of either a two-moment approximation method or Little's law. We illustrate these methods with MOI estimates obtained under sparse sampling schemes with the recently proposed "*var*coding" method, based on sequences of the *var* multigene family encoding for the major variant surface antigen of the blood stage of malaria infection. The methods are evaluated with simulation output from a stochastic agent-based model, and are applied to an interrupted time-series study from Bongo District in northern Ghana before and immediately after a three-round transient indoor residual spraying (IRS) intervention. We incorporate into the sampling of the simulation output, limitations representative of those encountered in the collection of field data, including under-sampling of *var* genes, missing data, and usage of antimalarial drug treatment. We address these limitations in MOI estimates with a Bayesian framework and an imputation bootstrap approach. We demonstrate that both proposed methods give good and consistent FOI estimates across various simulated scenarios. Their application to the field surveys shows a pronounced reduction in annual FOI during intervention, of more than 70%. The proposed approach should be applicable to the many geographical locations where cohort or cross-sectional studies with regular and frequent sampling are lacking but single-time-point surveys under sparse sampling schemes are available, and for MOI estimates obtained in different ways. They should also be relevant to other pathogens of humans, wildlife and livestock whose immune evasion strategies

are based on large antigenic variation resulting in high multiplicity of infection.

Introduction

Despite substantial control and elimination efforts, falciparum malaria in high-transmission regions continues to be a major public health concern, as a cause of mortality among young children and considerable economic burden in many countries especially in sub-Saharan Africa (***World Health Organization, 2023***). It thus remains important to robustly evaluate the effects of intervention efforts in these regions, including on transmission intensity. Collectively, epidemiological studies have revealed a multiplicity of metrics that are associated with parasite transmission (***Tusting et al., 2014***). Chief among them is the force of infection (FOI), defined as the number of new *Plasmodium falciparum* infections acquired by an individual host over a given time interval. FOI has been found to reflect risk of infection and to be tightly associated with risk of clinical episodes (***Mueller et al., 2012***). While other metrics may describe the relationship between transmission intensity and the burden of malaria illness on global or continental scales (***Carneiro et al., 2010***; ***Beier et al., 1994***), FOI can relate local variation in malaria burden to transmission (***Mueller et al., 2012***). Despite being recognized as a key epidemiological parameter for malaria surveillance, FOI remains difficult, expensive, and labor-intensive to accurately measure, either directly or indirectly via the fitting of epidemiological models. As molecular tools for parasite genomics become more readily available, they are contributing to new approaches. In particular, molecular advances are providing a basis for estimating a sister "static" quantity, the multiplicity of infection (MOI), defined as the number of genetically distinct parasite strains co-infecting a single human host (***Zhong et al., 2018***; ***Chang et al., 2017***; ***Ruybal-Pesántez et al., 2022***; ***Tiedje et al., 2022***; ***Labbé et al., 2023***). We can therefore ask whether we can go further and on the basis of MOI obtain the dynamical, rate, quantity of force of infection?

Prior efforts to directly measure FOI included clearing infections and observing time to reinfection. Infant conversion rate, the time to a patent infection in uninfected children less than 1-year of age, was widely used in the 1950s to 1970s (***Macdonald, 1950***; ***Pull and Grab, 1974***). An alternative way was to perform detection of parasites by light microscopy for subjects following antimalarial treatment (***Msuya and Curtis, 1991***; ***Alonso et al., 2004***). In both cases, FOI was measured directly in either a naturally uninfected cohort (e.g. infants or uninfected immigrants) or by artificially creating a cohort of uninfected individuals through treatment with antimalarial drugs, which was then followed to estimate the attack rate or proportion acquiring infection over some time period. More recently, with the advent of molecular approaches to malaria epidemiology, direct and more precise measures of FOI can be generated by genotyping individual parasite infections (***Mueller et al., 2012***; ***Hofmann et al., 2017***). Individual clones can be tagged by a set of molecular markers which are polymorphic, a precondition for longitudinally tracking and differentiating individual parasite clones. PCR-based techniques also offer the added advantage of higher sensitivity by early infection detection, when parasite density is under the detection limit of microscopy (***Johnston et al., 2006***). It remains challenging however to accurately differentiate a truly new infection from the temporary absence from the peripheral blood of an old infection and its subsequent reemergence (***Daubersies et al., 1996***). This is primarily because some of these popular polymorphic molecular markers have a relatively low resolution for differentiating different strains from one another. Despite increased access to better techniques, the determination of molecular FOI remains challenging, labor-intensive, and costly, as it still involves closely monitoring and genotyping a large cohort over some period of time.

Alternatives to direct measurements involve cross-sectional surveys with FOI estimated by fitting simple epidemiological models, including Markov models corresponding to the reverse catalytic formulations parameterized from changes in prevalence with age (***Bekessy et al., 1976***; ***Muench, 1959***; ***Smith and Vounatsou, 2003***), or immigration-death models informed by the values of MOI

across age groups (***Felger et al., 2012***), or iterations between a generalized linear mixed model and a mathematical Susceptible-Infected-Susceptible model (***Mugenyi et al., 2017***). These model-fitting procedures require empirical data sampled regularly and frequently, such as age-stratified cohorts with a large number of individuals sampled six times a year, to account for the high heterogeneity and variation in FOI and infection duration (***Laishram et al., 2012***; ***Simpson et al., 2002***; ***Langhorne et al., 2008***; ***Childs and Buckee, 2014***; ***Chang et al., 2016***; ***Ashley and White, 2014***), both of which are dependent on the interplay between hosts immunity profiles and antigen composition of any infection (***Piper et al., 1999***; ***Molineaux et al., 2002***; ***Doolan et al., 2009***; ***Barry et al., 2007***). In addition, they may suffer from identifiability problems with more than one value of given parameters of interest resulting in similar likelihood or other model selection metrics (***Mugenyi et al., 2017***). Hence, these indirect measurements share limitations with direct ones.

As a result of the above-described challenges, FOI has not become a readily available epidemiological quantity across time and geographical locations. In contrast, various approaches have been proposed to estimate MOI from clinical samples, involving size-polymorphic antigenic markers, microsatellites, and panels of biallelic single nucleotide polymorphisms (SNPs) (***Felger et al., 1994***; ***Konaté et al., 1999***; ***Anderson et al., 2000b***; ***Daniels et al., 2008***; ***Chang et al., 2017***). Because Plasmodium parasites reproduce asexually during haploid stages within human hosts (***Guttery et al., 2012***), signatures of polymorphic genotypes are evidence of multiclonal infection. As methods for disentangling the molecular complexity of natural parasite infections across various transmission settings emerge and mature, MOI becomes much more commonly and easily surveyed across space and time. Although this quantity remains one of the most frequently used genetic metrics of parasite transmission (***Arnot, 1998***; ***Sondo et al., 2020***), it is by definition a number and not a rate.

A natural correlation exists however between MOI and FOI mediated by the duration of infection, which provides an opportunity to convert the former into the latter. This conversion has been particularly challenging because disease transmission and dynamics display heterogeneity in time and across individuals (***Laishram et al., 2012***; ***Simpson et al., 2002***; ***Langhorne et al., 2008***), duration of infection information is highly heterogeneous across a multiplicity of factors and mostly lacking (***Childs and Buckee, 2014***; ***Chang et al., 2016***; ***Ashley and White, 2014***; ***Bretscher et al., 2011***), and the majority of MOI estimates are obtained under sparse sampling schemes. Rather than cohort-studies or cross-sectional studies with regular and frequent sampling, most often single-time-point surveys are implemented, typically in wet (high-transmission) and dry (low-transmission) seasons (***Tiedje et al., 2022***; ***Abukari et al., 2019***). Although MOI estimates from those single-time-point cross-sectional surveys have been used to address a number of questions such as whether multiple infections protect against clinical attacks, and whether each additional infecting clone is associated with a higher prospective risk or a higher odd of treatment failure (***Earland et al., 2019***; ***Lee et al., 2006***), the sparse sampling scheme has limited their translation into a transmission rate.

In this work, we propose to specifically use these MOI estimates for the purpose of FOI inference. We present two mathematical modeling frameworks based on queuing theory for this purpose (***Choi et al., 2005***; ***Little and Graves, 2008***), and illustrate their application with empirical molecular data collected in the Bongo District of northern Ghana (***Tiedje et al., 2022***, ***2023***). The proposed methods can be easily scaled up and applied across different geographical locations and time. They are potentially applicable to the many geographical locations where cohort or cross-sectional studies with regular and frequent sampling are lacking but single-time-point surveys are available. We evaluate their performance with numerical simulation of an extended stochastic agent-based model (ABM) (***He et al., 2018***) for both a closed system and an open one under two different spatial scenarios, specifically two populations coupled via migration and a local population with migration from a regional pool, with either constant or seasonal transmission. We further consider cases of heterogeneous transmission in which a high-risk group of hosts receives the majority of infectious bites. We also examine scenarios in which the times of local transmission events follow different statistical distributions. We incorporate limitations representative of those encountered in the

collection of field data into the sampling of simulation output, including under-sampling of *var* genes, missing data, and usage of antimalarial drug treatment. We obtain MOI estimates with a Bayesian framework and a bootstrap imputation approach correcting for all the aforementioned errors. The Bayesian method accounts for the under-sampling of *var* genes for individual isolates. The bootstrap imputation approach accounts for the missing data error and the usage of antimalarial treatment by removing the samples for individuals who are under the impact of the two and then inferring their MOI estimates by sampling from those of the remaining population. We show that our Bayesian method and bootstrap imputation approach are robust to considerable fractions of individuals with their MOI information missing or under antimalarial treatment, comparable to those from our field surveys in northern Ghana. We demonstrate further that based on MOI estimates obtained in this way, both proposed methods give good and consistent FOI estimates across various simulated scenarios. After validating their performance with simulation output, we illustrate the application of the proposed methods to empirical data collected from an interrupted time-series study from Bongo District in northern Ghana that involved a three-round transient indoor residual spraying (IRS) intervention (***Tiedje et al., 2022***, ***2023***). We conclude with discussion of the relationship between FOI and another commonly used surrogate variable for transmission intensity, the entomological inoculation rate or EIR, defined as the number of infectious bites received by an individual over a given time period (***Shaukat et al., 2010***). We discuss how this relationship explains the known difficulty in relating transmission to malaria burden at more refined local scales, and that of achieving sustainable and considerable transmission reductions in high-transmission endemic regions. We further discuss how this relationship might be informative of malaria immunity, and within-host competitive interactions of co-infecting strains.

Results

The MOI estimates at the population level for surveys in Ghana resemble those under high-intensity and heterogeneous transmission from numerical simulation

While any polymorphic marker should be suitable in theory for estimating MOI, the common high number of multiclonal infections in high-transmission endemic regions limits the ability to accurately do so in practice (***Tiedje et al., 2022***; ***Ruybal-Pesántez et al., 2017***; ***Touray et al., 2020***; ***Labbé et al., 2023***). A multigene family known as *var* which encodes for hyper-diverse antigenic variation in the parasite provides one proposed solution, described in the Materials and Methods and referred to as "*var*coding" (***Ruybal-Pesántez et al., 2022***; ***Tiedje et al., 2022***). We recently extended the method to a Bayesian formulation to account for the measurement error from under-sampling or imperfect detection of *var* genes in the collection of empirical data (***Tiedje et al., 2023***), with details summarized in the Materials and Methods-Bayesian approach to MOI estimation with "*var*coding". Additionally, empirical surveys suffer from a missing data error due to factors including the detection limit of the selected method and low DNA quality, and they contain a fraction of individuals under antimalarial drug treatment during a previous window of time. Antimalarial treatment either prevents new infections or clears existing ones by shortening their duration, which potentially violates the assumptions of our two proposed methods for FOI inference. This is because the two methods rely on a known and fixed distribution of duration of infection for 1-5 year-old children. Because these children's immune profile is similar to that of naive hosts, for this specific sub-population class, the duration of infection can be approximated by that of malaria therapy patients with neurosyphilis and no history of prior exposure to malaria (***Collins and Jeffery, 1999***; ***Maire et al., 2006***). Details can be found in Materials and Methods-Inferring FOI from MOI estimates. We address both types of errors with a bootstrap imputation approach whose details are included in Materials and Methods-The under-sampling of infections in estimates of MOI from empirical surveys, and Materials and Methods-Antimalarial drug treatment of infections in estimates of MOI from empirical surveys.

Because we rely on MOI estimates to derive FOI ones, we first conduct here an investigation of

the MOI estimates obtained under the impact of each of these aforementioned errors, and of all combined, in comparison with true MOI values. We rely on the simulated output from our agent-based model (ABM) of malaria transmission for which true MOI values are known. The ABM is described in detail in previous studies of *P. falciparum*'s population structure based on *var* genes (***He et al., 2018***). We provide additional details on its formulation, including extensions implemented to the model for this work, and on the experimental design of simulation output in Appendix 1-Simulation data. See also Appendix 1-Figure 5A-C for illustrations of the experimental design.

Our analysis shows that MOI estimates obtained with the Bayesian method and the bootstrap imputation approach with all the aforementioned errors present, are close to the true MOI values. We measure the difference between estimated and true MOI by the difference in their mean value and the Kolmogorov-Smirnov statistic, which quantifies the distance in their cumulative distribution function. Results are included in the supplementary file 1-MOImethodsPerformance.xlsx, and illustrated in Appendix 1-Figure 6. The bootstrap imputation approach corrects well for the missing data error and the antimalarial drug treatment issue. The Bayesian method reduces the under-sampling of *var* genes error, resulting in a noticeable improvement relative to the original *var*coding method which relies on a constant repertoire size to convert the number of detected non-upsA DBLα types to the estimated MOI value, and hence does not consider the variation in repertoire size introduced by the under-sampling of *var* genes. See supplementary file 2-BayesianImprovement.xlsx, Appendix 1-Figure 7, and Appendix 1-A comparison between the Bayesian method and the original *var*coding one for further details.

The MOI estimates at the population level for simulation output under high-intensity and heterogeneous transmission are similar in range to the ones for surveys in Ghana (Appendix 1-Figure 6 and 8-9). The respective distributions of MOI estimates across the different surveys in Ghana and from the simulated outputs are not Poisson-distributed (Appendix 1-Test of deviation from Poisson homogeneity in MOI estimates, supplementary file 3-deviationFromPoissonTest.xlsx) (***Potthoff and Whittinghill, 1966***; ***Lloyd-Smith et al., 2005***). The deviation of empirical MOI estimates from a Poisson distribution indicates that either the arrival of inoculations deviates from a homogeneous Poisson process, or some latent heterogeneity causes inoculation duration distribution to be over-dispersed. Despite these conditions which challenge the conversion of MOI estimates to their corresponding FOI values, the proposed methods below are still applicable (see Materials and Methods-Inferring FOI from MOI estimates for details).

The Two-Moment Approximation and Little's Law approaches give good and consistent estimates for FOI across different simulated scenarios

We start with the result of the homogeneous exposure risk scenario for seasonal transmission and a closed system based on the two-moment approximation approach and Little's law. The times of local transmission events follow a Gamma distribution (Appendix 1-Figure 5C). Across pre-IRS, low-, mid-, and high-coverage IRS, the 95% confidence intervals of annual FOI per host and the full sampling distribution from bootstrap analysis are narrow and concentrated, containing, or very close to, the true FOI values (Figure 1). We infer FOI estimates based respectively on the true MOI values, on the MOI estimates with the Bayesian method and a bootstrap imputation approach correcting for all aforementioned errors, and on the MOI estimates with one source of error at a time corrected by either the Bayesian method or the bootstrap imputation approach, namely under the missing data error, the under-sampling of *var* genes, and the antimalarial treatment issue. The FOI estimates obtained with the two methods based on the true MOI values differ only slightly from the true FOI values. The FOI estimates based on the MOI estimates with the Bayesian method and a bootstrap imputation approach correcting for all errors exhibit a somewhat larger, but still small, difference from the true FOI values.

We continue with the result of the heterogeneous exposure risk scenarios in which a high-risk group ($\frac{2}{3}$ of the total population) receives around 94% of all bites whereas a low-risk group ($\frac{1}{3}$ of the total population) receives the remaining bites (Appendix 1-Figure 5C). Transmission is seasonal

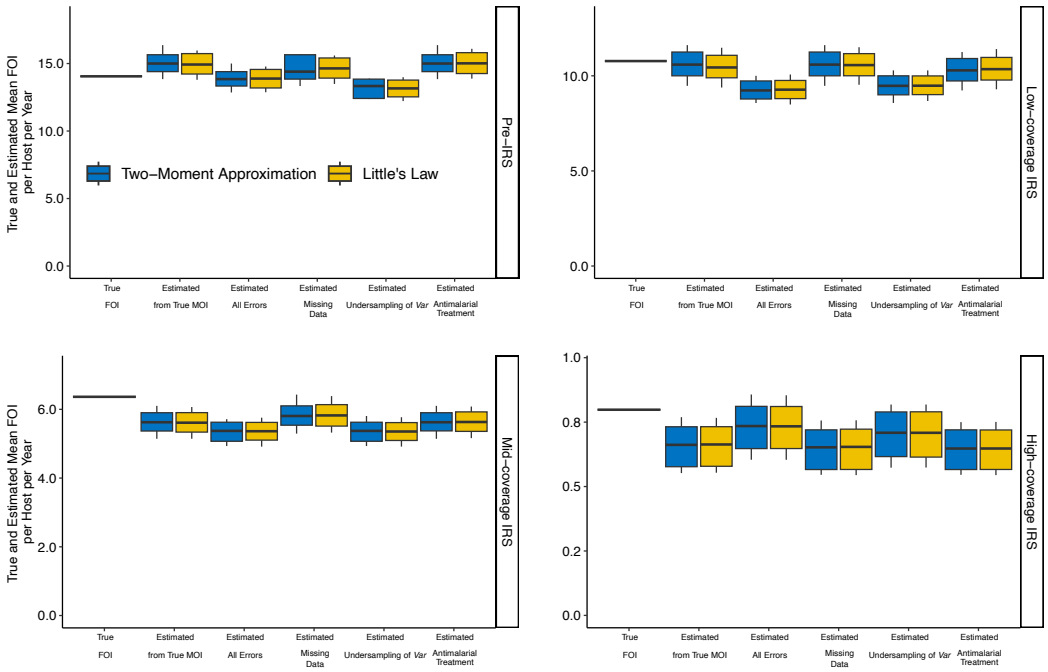


Figure 1. Confidence intervals for estimated mean FOI values in simulated risk scenarios with homogeneous exposure, before and during IRS interventions with three different coverage levels. The times of the local transmission events follow a Gamma distribution, and transmission is seasonal with the spatial setting corresponding to a closed system. We provide FOI estimates based respectively on the true MOI values, on the MOI estimates with the Bayesian method and a bootstrap imputation approach correcting for all aforementioned errors, and on the MOI estimates with one source of error at a time corrected by either the Bayesian method or the bootstrap imputation approach, namely under the missing data error, the under-sampling of *var* genes, and the antimalarial treatment issue. To obtain the true mean FOI per host per year, we divide the total number of infections acquired by the population by the total number of individual hosts in the population. Minimum, 5% quantile, median, 95% quantile, and maximum values are shown in the boxplot.

and the spatial setting corresponds to a semi-open system (Appendix 1-Figure 5A-B, and Appendix 1-Simulation data). The times of local transmission events are Gamma-distributed (Appendix 1-Figure 5C). Across pre-IRS, low-, mid-, and high-coverage IRS, the 95% confidence intervals and the full sampling distribution from bootstrap analysis are narrow and concentrated, containing, or very close to, the true FOI values (Figure 2).

The performance of the two methods across other simulated scenarios is demonstrated in Appendix 1-Figure 10-14. Details for deriving the confidence intervals are provided in Appendix 1-Confidence intervals for mean and variance.

The Two-Moment Approximation and Little's Law approaches give consistent FOI estimates for empirical surveys from Bongo District in northern Ghana

The two-moment approximation approach and Little's law also give consistent estimates of the FOI using the empirical MOI estimates from surveys in Ghana. We provide in Materials and Methods, further details and proposed solutions regarding the under-sampling of *var* genes, as well as antimalarial drug treatment and missing data due to factors such as the low detection limit of the selected method and low DNA quality, in obtaining empirical MOI estimates. Since PCR has a high, but yet not perfect power of infection detection, we further assume three levels of sensitivity, ranging from 0% of PCR-negative individuals carrying infection or a detection of 100% of the infected individuals (high detectability), to 5% of PCR-negative individuals carrying infections

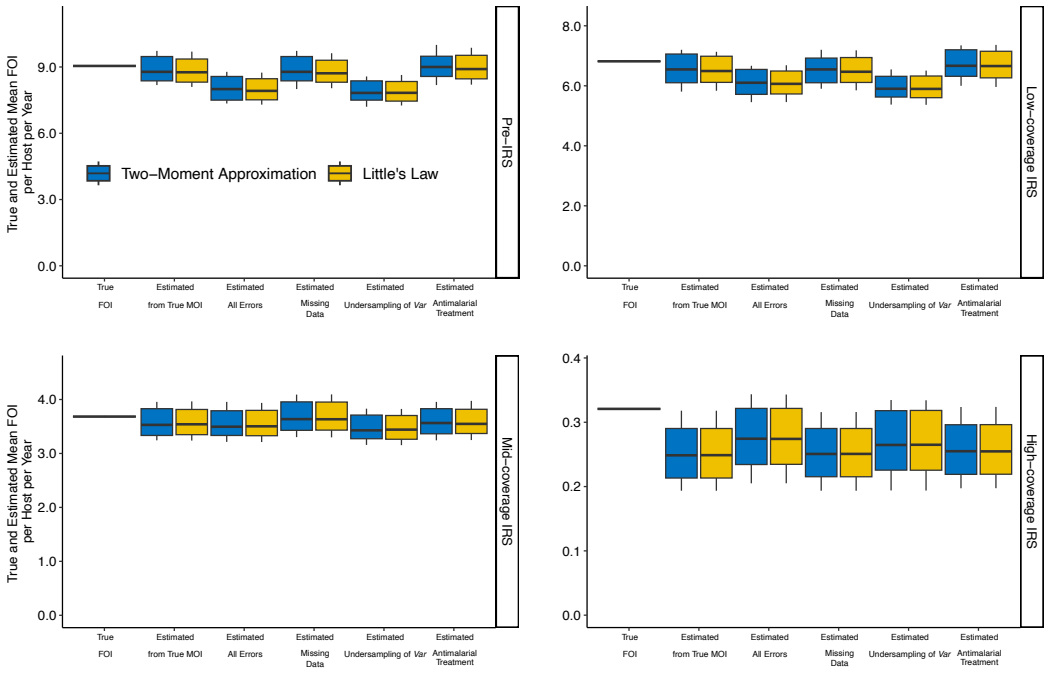


Figure 2. Confidence intervals for the estimated mean FOI values in simulated risk scenarios with heterogeneous exposure, before and during IRS interventions with three different coverage levels. The times of the local transmission events follow a Gamma distribution, and the transmission is seasonal with the spatial setting corresponding to a semi-open system. We provide FOI estimates based respectively on the true MOI values, on the MOI estimates with the Bayesian method and a bootstrap imputation approach correcting for all aforementioned errors, and on the MOI estimates with one source of error at a time corrected by either the Bayesian method or the bootstrap imputation approach, namely under the missing data error, the under-sampling of *var* genes only, and the antimalarial treatment issue. To obtain the true mean FOI per host per year, we divide the total number of infections acquired by the population by the total number of individual hosts in the population. Minimum, 5% quantile, median, 95% quantile, and maximum values are shown in the boxplot.

(mid detectability), to 10% of PCR-negative individuals carrying infections (low detectability). The impact of antimalarial treatment on infection status and duration, and therefore on MOI values, is difficult if not impossible to disentangle due to the complex biology of the parasite. We thus consider the following two extremes which should bound reality. Antimalarial treatment either does not impact treated individuals' infection status and MOI values, or it does affect both. For the former assumption (hereafter assumption 1), we include the infection status and MOI estimates of these treated individuals when inferring FOI. For the latter assumption (hereafter assumption 2), we discard their infection status and MOI estimates and sample instead their MOI values from the microscopy-positive individuals who were not under the missing data error or antimalarial treatment.

We start with the two-moment approximation approach. The estimated values of the annual mean FOI before IRS (Appendix 1-Figure 5E, the pre-IRS phase) are based on Survey 1 (S1; Oct 2012) and Survey 2 (S2; May/Jun 2013), undertaken at the end of the wet/high-transmission and dry/low-transmission seasons, respectively. Under assumption 1, the FOI estimates are respectively 5, 5.214286, and 5.367647, assuming respectively that 0%, 5%, 10% PCR-negative individuals carry infections. Under assumption 2, they are equal to 5.983607, 6.186441, and 6.293103, assuming respectively that 0%, 5%, 10% PCR-negative individuals carry infections. The estimated values for the year when IRS was discontinued (Appendix 1-Figure 5E, the post-IRS phase right after IRS was interrupted) are based on Survey 3 (S3; Oct 2015) and Survey 4 (S4; May/Jun 2016), undertaken at the end of the wet/high-transmission and dry/low-transmission seasons, respectively. The corresponding FOI estimates under assumption 1 are 0.7635983, 0.9864865, 1.140625, assuming respectively that 0%, 5%, 10% PCR-negative individuals carry infections. The corresponding FOI estimates under assumption 2 are are 1.46, 1.615044, 1.738095, assuming respectively that 0%, 5%, 10% PCR-negative individuals carry infections.

We then consider Little's law. The estimated annual mean FOI values before IRS under assumption 1 are 4.83786, 5.089157, 5.237725, for 0%, 5%, 10% PCR-negative individuals carrying infections respectively. For comparison, the estimated annual mean FOI values before IRS under assumption 2 are 5.92716, 6.099559, 6.224758. The corresponding estimated values immediately post-IRS under assumption 1 are 0.764853, 0.9870203, 1.14117. Those values under assumption 2 are 1.456373, 1.622026, 1.741665.

The two-moment approximation and Little's law approaches give consistent FOI estimates for empirical surveys from Bongo District in northern Ghana. There is a considerable reduction in FOI, of more than 80% or 70% under the two assumptions respectively, which indicates that the three-round IRS intervention although transient was efficient and strong. Reality is likely to lie somewhere in between the two bounding assumptions on the impact of antimalarial treatment on infection status and MOI values, so that such treatment may have a non-negligible effect on some individuals but not others. Regardless, the actual estimates of FOI are similar in magnitude and the trend in the change of FOI across intervention, and so the evaluation of the impact of such intervention, remain qualitatively the same.

The 95% confidence intervals and the full sampling distribution from bootstrap analysis are narrow and concentrated (Figure 3).

The inferred FOI and the directly measured EIR for surveys in Ghana are consistent with the relationship between the two measures from previous studies

We plot the measured mean annual EIR against the estimated mean annual FOI, and the transmission efficiency (the ratio between FOI and EIR) against the measured mean annual EIR from previous field studies, summarized in (***Smith et al., 2010***) (Figure 4). The estimation of FOI in these studies was primarily based on fitting a simple epidemiological model to age-stratified prevalence from cross-sectional parasitological studies (***Smith et al., 2010***; ***Pull and Grab, 1974***); that of EIR was based on diverse methods including exit bait collection, human bait collection, pyrethrum spray collection, night bite collection, and outdoor resting collection (***Smith et al., 2010***). The red

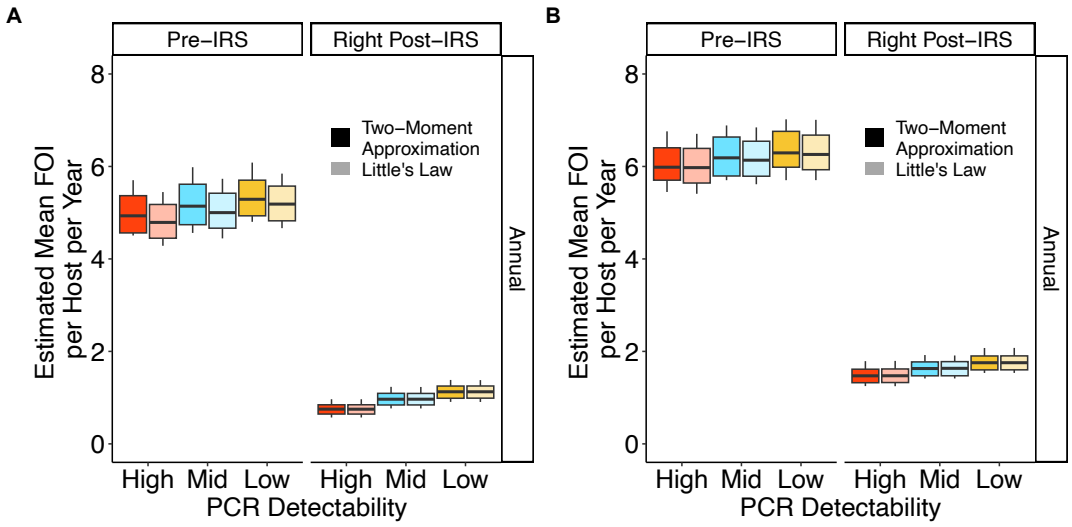


Figure 3. Confidence intervals for the estimated mean FOI values in Ghana surveys across a transient three-round IRS intervention based on the combination of surveys from the wet/high-transmission and dry/low-transmission seasons (A) The estimated FOI values when considering that antimalarial treatment does not affect infection status and MOI values in treated individuals (assumption 1). (B) The estimated FOI values when considering instead that antimalarial treatment affects infection status and MOI values in treated individuals (assumption 2). High, mid, and low detactability correspond to 0%, 5%, and 10% of PCR-negative individuals carrying infections respectively. Minimum, 5% quantile, median, 95% quantile, and maximum are shown in the boxplot.

line corresponds to the functional curve fitted to these data points as described in Materials and Methods-Conversion between FOI and EIR, as initially proposed in (***Smith et al., 2010***).

Based on these data pairs, we see that the values for mean annual FOI are well below an empirical limit of 20. Due to the highly non-linear pattern between FOI and EIR, there is no single constant factor that converts FOI to EIR (or the opposite) across different empirical settings. The transmission efficiency in high-intensity endemic regions is extremely low, with an average annual EIR reaching values of a few hundred to a thousand, whereas the average annual FOI takes values of only about 5-10. In future applications, when applying our proposed methods to obtain FOI estimates from MOI under sparse sampling schemes from specific geographical regions, we can rely on this functional curve to convert the FOI estimates to corresponding EIR values. There is however an inherently high variance associated with the conversion. Nevertheless, overall, our paired EIR (directly measured by the entomological team in Ghana (***Tiedje et al., 2022***)) and FOI values are reasonably consistent with the data points from previous studies, suggesting the robustness of our proposed methods.

The variance inferred reflects transmission intensity and heterogeneity across individuals

We have focused so far on estimation of the mean for FOI in both the simulation output and the empirical data. The two-moment approximation method further gives an estimate for the variance of the inter-arrival times of infections. We examine the variance inferred across all simulated scenarios. We find that regardless of the magnitude of the variance for FOI or the degree of heterogeneity in exposure risks across individual human hosts, the estimated mean FOI at the individual level by the two proposed methods, when aggregated across all individuals within the local population, reflects the total number of infections accumulated by that population.

Overall, seasonal runs have slightly larger variance than non-seasonal runs (Appendix 1-Figure 15). Runs with a heterogeneous transmission, in which a high-risk group of individuals receives

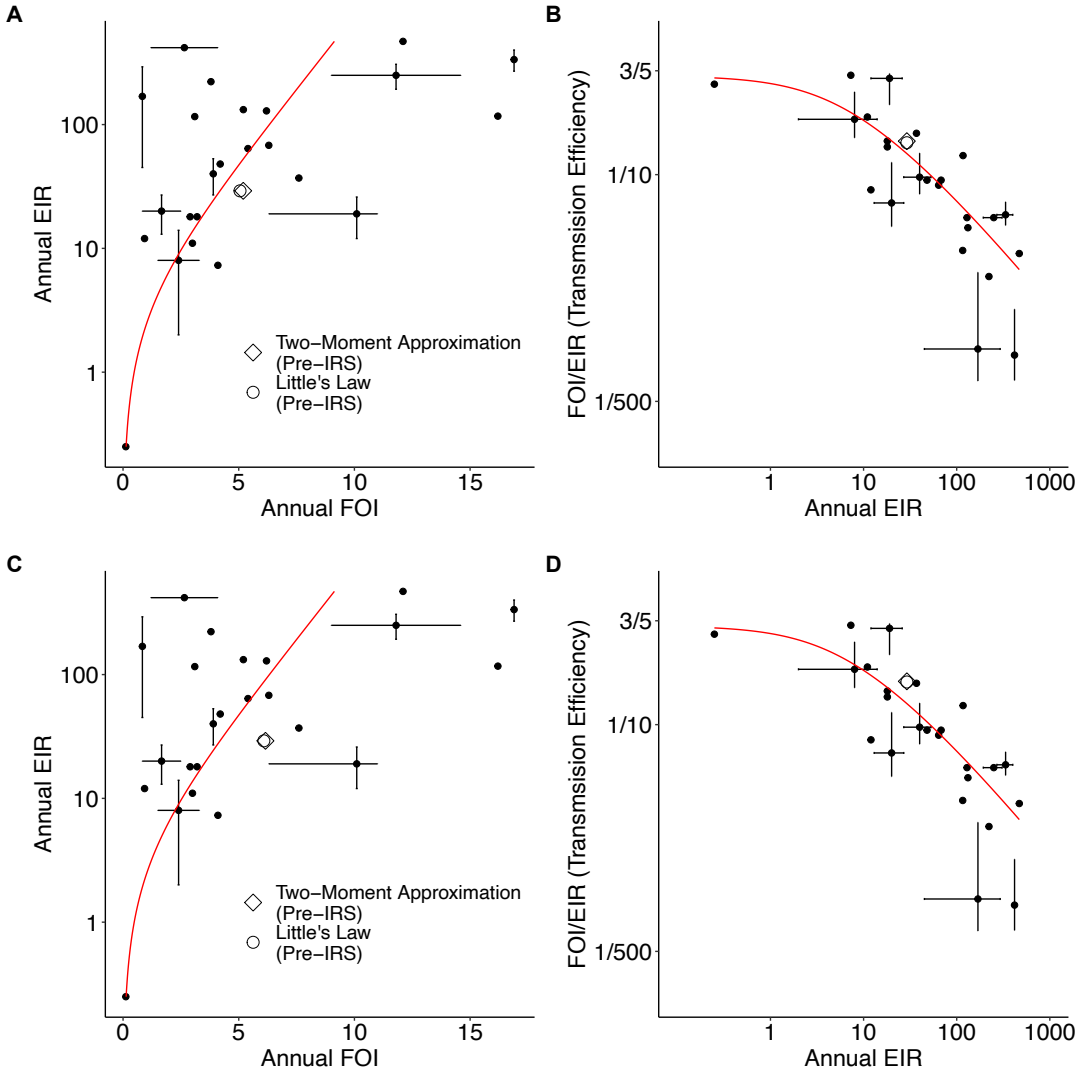


Figure 4. Saturation in FOI with EIR and the non-linear relationship between these two quantities from previous field studies. (A) and (B) includes our empirical estimates under the assumption that antimalarial treatment does not affect infection status and MOI values in treated individuals. (C) and (D) includes our estimates under the assumption that the treatment affects such quantities. The points correspond to paired EIR-FOI values from the literature, summarized in (***Smith et al., 2010***), and the crosses indicate the range of those values when several estimates of the EIR or the FOI were reported or estimated for the same location. Namely, a line or a cross was plotted with its center at the arithmetic mean and with legs that connect the center to each one of the estimates, and similarly for the scenario when ranges were reported for both variables. The red function curve is the best-fit to these paired EIR-FOI values from (***Smith et al., 2010***). The black hollow diamond and circle correspond to the Ghana data, for our respective estimates of FOI with the two methods and the EIR measure in the field by the entomological team (***Tiedje et al., 2022***).

more than 90% of the total bites, have slightly higher variance than those with a homogeneous transmission in which bites do not exhibit such variation. These findings are consistent with expectations since seasonality and heterogeneity in transmission should increase the dispersion of FOI across individuals, and consequently of MOI. The inferred variance correlates most significantly with the inferred mean. In particular, when the mean FOI is close to 0, the variance is large. This is intuitive as significant deviation from the close-to-zero mean FOI (or, large inter-arrival times, for example, at the order of several years) would be required to generate some fraction of individual hosts with MOI greater than zero.

Discussion

Building on estimates of the multiplicity of infection under sparse sampling schemes, we have demonstrated the feasibility of converting this quantity to estimates of the force of infection. Across a wide range of numerical simulation scenarios with the extended stochastic agent-based model, the two-moment approximation approach and Little's law are shown to provide consistent and good FOI estimates, and to do so based on the MOI estimates obtained with the Bayesian method and bootstrap imputation approach in the presence of various types of errors commonly encountered in the collection of empirical data. The Bayesian method addresses in particular the undersampling of *var* genes and helps alleviate the underestimation of MOI estimates.

In high-transmission endemic regions which continue to pose major challenges to malaria elimination, clinical infections typically constitute a small fraction of the overall reservoir of infection behind the force of infection, of approximately 2% at the end of wet season and less than 1% at the end of dry one (***Tiedje et al., 2022***). Individuals can seek however antimalarial drug treatment in response to symptoms or the perception of transmission risk, which affects duration of infection and would interfere with our FOI estimates. Rather than attempting a correction of the duration of infection (which given the complexity of the biology involved would be difficult if not impossible at this stage), we implement instead the removal of the samples for treated individuals, and impute their MOI estimates with a bootstrap approach based on the remaining population. The proposed bootstrap imputation approach is robust to considerable fractions of individuals treated or missing MOI information, comparable to values from our field surveys in Ghana.

The application of both methods typically requires an evolving time series as it is observed over a long period of time. Because we do not have time series observation of MOI estimates for any specific individual host, we treat different individual hosts sampled at the same time point as if each one represented an observation of the process at a single time point and ignore any individual heterogeneity in their biology. This approximation turns out to be a sensible one and the bias due to interval edge effects when the system does not begin and end empty turns out to be small because the FOI estimates obtained with the two methods based on the true MOI values differ by only a very small amount from the true FOI values in the simulation output. Despite seasonality and heterogeneity in exposure risk, MOI estimates obtained under sparse sampling schemes are sufficient to recover the mean annual FOI. The inferred mean annual FOI when aggregated across all individuals faithfully reflects the total number of new *P. falciparum* infections acquired by the entire population. The inferred variance indicates transmission intensity and the degree of deviation from a homogeneous Poisson process, for the arrival of infections at the individual level. Thus, the flexibility and generality of these two proposed methods allow robust estimation of an important metric for malaria transmission across geographical locations and time, where single-time-point surveys are implemented, typically in wet (high-transmission) and dry (low-transmission) seasons, and information about FOI was previously unobtainable.

We illustrate the application of these two methods with molecular data from Bongo District in northern Ghana, a high-transmission endemic region for which estimation of multiplicity of infection and related FOI has been challenging. The three-round transient IRS intervention was strong and effective, resulting in a significant, more than 70% reduction in FOI.

Beyond describing basic malaria epidemiology and measuring outcomes of intervention, the FOI estimates obtained with these two proposed methods can inform process-based models for the population dynamics of complex infectious diseases. They can serve as values for priors when seeking to parameterize and validate more complex agent-based or equation-based models, in a way that reduces computational cost, improves efficiency, and reduce the degree of freedom or identifiability issues.

Although we have focused on infections caused by *P. falciparum*, our approach should be applicable to other *Plasmodium* parasites responsible for malaria transmission, including *P. vivax*, with distinctive life cycles. For *P. falciparum*, FOI is closely linked to the number of newly acquired infections from infectious mosquito bites, whereas for *P. vivax*, clones appearing in the blood-stream can either originate directly from recent infective mosquito bites or relapsing liver hypnozoites (***Chu and White, 2016***). Regardless of infection source, MOI estimates and related FOI ones should represent the number of effective infections which reach the blood stage.

As expected from estimating FOI based on MOI values, the estimation accuracy of the latter is crucial for that of the former. In high-transmission endemic regions, repertoires of *var* genes in co-infecting genomes can still be partially overlapping in their composition. This low overlap results in a reduced number of different *var* genes being sequenced and identified, which can lead to an underestimation of MOI. In addition, recent studies have identified a signature of parasite co-transmission from a single mosquito bite even in high-transmission endemic regions (***Nkhoma et al., 2020***; ***Wong et al., 2017***). This conclusion was based primarily on clinical or symptomatic infections whereas asymptomatic prevalence is high and constitutes the majority of transmission reservoir for malaria (***Andolina et al., 2021***; ***Lindblade et al., 2013***) in these regions. Hence, additional data needs to be collected on polygenomic relatedness across diverse epidemiological settings. These co-transmitted recombinant parasites are more related to each other than parasites from different mosquito bites, potentially resulting in a further reduced number of different *var* genes sequenced and identified (***Wong et al., 2022***; ***Nkhoma et al., 2020***; ***Wong et al., 2017***). Low or variable parasite densities resulting from small blood volumes, genomic DNA quality, clinical status, and/or within host dynamics, including synchronicity will all create sampling problems (***Peyerl-Hoffmann et al., 2001***; ***Bruce et al., 2000***; ***Okell et al., 2012***; ***Farnert et al., 1997***; ***Färnert et al., 2008***; ***Barry et al., 2021***; ***Hergott et al., 2024***). But these issues are common to all measures of MOI, as well as to all direct measures of FOI. Our previously proposed framework of MOI estimation could be extended in the future to account for these confounding effects in recovering the number of co-infecting strains.

A salient characteristic of malaria transmission is the saturation in FOI at high transmission, and the highly non-linear relationship between FOI and another central epidemiological measure related to transmission intensity, the entomological inoculation rate, EIR. Whereas EIR can reach values up to a few hundred or even a thousand per year, the annual FOI saturates rapidly and is typically kept below a limit of 20 (***Smith et al., 2010***). Although different choices are made on the way EIR and FOI are measured in the field, these cannot account for such drastically different magnitudes of the two quantities. For example, adults are used as bait to attract and trap mosquitoes, but the attack rates are typically evaluated in children, and there is an observed correlation between body size and biting rates. Transmission is highly inefficient in high-transmission endemic regions with high annual EIRs. Because FOI refers directly to detectable blood-stage infections whereas EIR concerns human-vector contact rates, the difference between these quantities is mediated primarily by immunity, or within-host dynamics, measurement bias, and heterogeneous transmission (***Donovan et al., 2007***; ***Macdonald, 1950***; ***John et al., 2005***; ***Doolan and Martinez-Alier, 2006***). The probability of transmission from an infectious mosquito bite is commonly used in mathematical models to mediate the difference between the two transmission quantities, as a general parameter encapsulating a variety of processes. Though historically paired EIR and FOI values are only available from regions under frequent sampling schemes, our methods, when applied to geographical locations under sparse sampling schemes, could generate further information on this

relationship and in so doing, provide further insights into within-host dynamics.

The saturation of FOI underlies the challenge of control and elimination efforts and the associated resilience of the transmission system, whereby relaxation of intervention leads to rapid rebounds in prevalence in high-transmission regions. In such regions of high EIR and FOI saturation, intervention efforts need to sufficiently reduce EIR to bring FOI below saturation. Only then further intervention efforts would have a significant impact by reducing the number of effective infections which advance to the blood stage. This apparent inefficiency of malaria transmission highlights why high-coverage interventions are often needed to achieve small observable impacts on individual exposure risk. Theory suggests a sharp nonlinear transition towards sustainable low transmission or elimination, including in a recent mathematical model highlighting the role of the high antigenic diversity of *P. falciparum* in high transmission (***de Roos et al., 2023***). The same molecular information used here to estimate MOI and FOI in Bongo District underlies such diversity. Our recent modeling work with the agent-based model on the resurgence dynamics of the transmission system below the saturation regime found that local systems can be highly resilient, and can compensate for the reduction in FOI following intervention with the heterogeneity in the duration of infection and more importantly the selection for infections with longer duration (***Zhan et al., 2024***). As a result, local systems can maintain during intervention the high prevalence and MOI levels comparable to original values.

The proposed methods are generally relevant to evaluate transmission intensity in many pathogens exhibiting multi-genomic infection as the result of large antigenic diversity including those encoding such variation by multigene families (***Deitsch et al., 2009***). More easily estimating changes in transmission intensity should contribute to the efficiency and evaluation of control programs across a wider range of infectious diseases.

Materials and Methods

High genetic diversity of *var* and the associated strain structure of limiting similarity

We present briefly here aspects of the biology of the malaria parasite *P. falciparum* that constitute the basis for the MOI estimation method used here to then apply the proposed FOI inference.

Individual human hosts acquire malaria infections through the infectious bites of mosquitoes, the rate of which can be either fairly constant or seasonal. In high-transmission endemic regions, individual human hosts remain susceptible to malaria re-infection (but not clinical disease) throughout their lifetime (***Doolan et al., 2009***). Indeed, the prevalence of asymptomatic infections is high within adults and older age groups in high-transmission regions, and these infections constitute a large transmission reservoir (***Andolina et al., 2021***; ***Lindblade et al., 2013***). High asymptomatic infections result not only from high transmission rates, but also from hosts' incomplete specific immunity conferred by high antigenic variation of the parasite (***Deitsch et al., 2009***).

A common strategy for immune evasion among parasites that exhibit such high asymptomatic prevalence under high transmission is to encode important variant surface antigens (VSAs) with multigene families (***Deitsch et al., 2009***). One important multigene family in the malaria parasite *P. falciparum* known as *var* encodes for the major VSA during the blood stage of infection, PfEMP1 (*Plasmodium falciparum* erythrocyte membrane protein 1) (***Zhang and Deitsch, 2022***; ***Baruch et al., 1995***; ***Smith et al., 1995***; ***Su et al., 1995***). Each parasite carries 50-60 *var* genes across its 14 chromosomes and these encode different variants of this protein. Parasites express these genes largely sequentially. Under high-transmission settings, the local parasite population exhibits a large pool of *var* gene variants documented to reach thousands to tens of thousands (***Day et al., 2017***; ***Tiedje et al., 2022***). This high level of diversity is enabled primarily by ectopic (mitotic) and meiotic recombination but also mutation (***Claessens et al., 2014***; ***Frank et al., 2008***; ***Freitas-Junior et al., 2000***; ***Bopp et al., 2013***). Migration can also introduce new variants and contribute to the extensive diversity of a local population. Negative frequency-dependent selection (NFDS) mediated by the

acquisition of specific immunity by hosts contributes to the maintenance of such a high *var* gene diversity, and to parasite population structure, with individual genomes (or strains or repertoires), and sets of genomes infecting different hosts (or isolates), exhibiting no, or extremely low, overlap of *var* gene composition. This pattern of limiting similarity has been shown to be both non-random and non-neutral (***Day et al., 2017***; ***He et al., 2018***).

As hosts typically have not seen many of the VSAs within a single strain, or across different circulating strains, infection duration lengthens which increases the fitness of the parasite by increasing its chance of getting transmitted (***Ruybal-Pesántez et al., 2022***). Multi-genomic infection is common, with infections within individual hosts consisting of different parasite clones that have arrived successively via different mosquito bites (superinfection) or simultaneously via one single mosquito bite containing multiple clones (co-transmission) (***Portugal et al., 2011***; ***Nkhoma et al., 2020***; ***Wong et al., 2017***; ***Volkman et al., 2012***; ***Anderson et al., 2000a***). The duration of an infection is influenced by the overlap between the constituent *var* genes of that infection and the immunological memory of the infected host. The measured duration of an individual infection with *P. falciparum* in similar settings from empirical observations has been reported to be highly variable (***Childs and Buckee, 2014***; ***Chang et al., 2016***; ***Ashley and White, 2014***; ***Bretscher et al., 2011***).

Empirical sequencing of *var* genes specifically concerns the DBLα tag, a small conserved ~450bp region within *var* genes which encodes for the immunogenic Duffy-binding-like alpha domain of PfEMP1 (***Tiedje et al., 2023***; ***Ruybal-Pesántez et al., 2017***, ***2022***; ***Day et al., 2017***). Bioinformatic analyses of a large database of exon 1 sequences of *var* genes showed a predominantly 1-to-1 DBLα-*var* relationship, such that each DBLα tag typically represents a unique *var* gene (***Tan et al., 2023***). We use DBLα and *var* interchangeably hereafter.

Bayesian approach to MOI estimation with "*var*coding"

The low to non-existent overlap of *var* repertoires enables estimation of MOI on the basis of the number of non-upsA DBLα types sequenced from an individual's isolate. The original method uses a constant repertoire length or number of non-upsA DBLα types in a parasite genome to convert the number of types sequenced in an isolate to the estimated MOI. As such, the method does not take into account the measurement error (Appendix 1-Figure 5D) in this length introduced by the sampling of *var* genes in an infection. We recently extended the method to a Bayesian formulation that does consider this error and provides a posterior distribution of MOI values for each sampled individual (***Tiedje et al., 2023***). We have previously documented the major steps of the Bayesian approach, compared two ways of obtaining an MOI distribution at the population level (either pooling the maximum a posteriori MOI estimates or calculating a mixture distribution), and examined the impact of different priors on the final MOI distribution at the population level. We provide in our analyses hereafter the estimated MOI distribution at the population level obtained from a mixture distribution and using an uniform prior for individuals.

We include an evaluation of this Bayesian *var*coding approach in Appendix 1-A comparison between the performance of the Bayesian approach and the original *var*coding method from the simulation output.

The under-sampling of infections in estimates of MOI from empirical surveys

The empirical MOI estimates in many epidemiological studies, including ours in Bongo District from northern Ghana, rely on individuals who are microscopy-positive (***Tiedje et al., 2023***, ***2022***). Given the sensitivity of microscopy, a significant fraction of individuals who carry infections are not detected. A subset of the Ghana surveys includes in addition submicroscopic infections detected by PCR (***Tiedje et al., 2023***, ***2022***). PCR is considerably more sensitive than microscopy, detecting a higher fraction, if not 100%, of individuals with *P. falciparum* infections. Using surveys for which both microscopy and PCR detection were used, we estimated a conversion factor between the proportion of hosts that are microscopy-positive and those that are PCR-positive within the targeted age group, i.e., children of 1-5 years old, which is 0.711329.

For surveys with available detection by both microscopy and PCR, we can directly calculate the number of individuals with undetected infections by microscopy, i.e., individuals who are microscopy-negative but PCR-positive. For surveys with detection only via microscopy, we use the estimated conversion factor to estimate the number of individuals who are microscopy-negative but PCR-positive. In addition, because the underlying sensitivity of PCR is high but not exactly known, we assume it varies within a reasonable range of values, from a relatively low detectability, i.e, 10% of all PCR-negative individuals carrying undetected infections, to a high and perfect detectability, i.e, none of PCR-negative individuals carrying undetected infections. We calculate the number of individuals with undetected infections by PCR for each sensitivity value respectively.

After calculating the number of individuals with undetected infections, including those who are microscopy-negative but PCR-positive and those who carry infections but are PCR-negative, we then sample values from the existing MOI estimates of microscopy-positive individuals who are not under antimalarial treatment (see the following section) to represent those missing MOI estimates.

In addition to the detection limit of microscopy and PCR, a fraction of microscopy-positive individuals may not have their *var* information sequenced and typed successfully due to factors including low DNA quality. Similarly, we sample values from the existing MOI estimates who are not under antimalarial treatment to represent those missing MOI estimates.

Antimalarial drug treatment of infections in estimates of MOI from empirical surveys

In addition to the missing data error described in the previous section, individuals can also seek and get antimalarial treatment in response to symptoms or the perception of transmission risk. In our field surveys from northern Ghana, more than 50% of children (1-5 years) responded that they had received an antimalarial treatment in the previous two-weeks (i.e., participants that reported they were sick, sought treatment, and were provided with an antimalarial treatment) prior to their blood samples being collected for the wet/high-transmission survey before IRS (i.e., Survey 1, Appendix 1-Figure 5E) (***Tiedje et al., 2022***). In contrast, the fraction is much lower for the dry/low-transmission survey before IRS or for the surveys collected right after IRS (i.e., Appendix 1-Figure 5E) (***Tiedje et al., 2022***).

To bound reality, which here is difficult if not impossible to determine, given the complexity of the biology, we consider two opposite extremes. For the first one (assumption 1), we assume that antimalarial treatment has no impact on treated individuals' infection status and MOI values. For the second one (assumption 2), we assume that antimalarial treatment affects infection status and MOI values in treated individuals. Thus, we remove the corresponding samples for treated individuals, and implement a bootstrap imputation approach based on the remaining population. We specifically sample values from the MOI estimates of microscopy-positive individuals who are neither under the aforementioned missing data error nor treated, to replace and represent MOI estimates for the treated individuals. This correction takes care of individuals who have used drugs and show either no infection (MOI=0) or infection (MOI>0). We show with numerical simulations that our bootstrap imputation approach is robust to considerable fractions of individuals treated, comparable to those from our survey in Ghana. Reality should lie somewhere in between these two assumptions, with antimalarial treatment affecting the infection status and MOI estimates in some individuals significantly but not in others.

The final distribution of MOI estimates at the population level, which includes values for microscopy-positive individuals, the imputed values for individuals with missing MOI information or the false negative individuals, the imputed values for individuals under malarial treatment (applicable for assumption 2), and the true zeros for uninfected individuals, is used for FOI inference.

Empirical surveys from Ghana

For this investigation we rely on empirical data collected from an interrupted time-series study in Bongo District northern Ghana, involving four age-stratified cross-sectional surveys of ~2,000 participants per survey between 2012 and 2016. This study was undertaken to investigate the impacts of a transient three-round IRS intervention, in combination with long-lasting insecticidal nets (LLINs), on the asymptomatic *P. falciparum* reservoir (***Tiedje et al., 2017***, ***2022***, ***2023***). The surveys were undertaken either at the end of the wet/high-transmission season (i.e., October) or at the end of the dry/low-transmission season (i.e., May/June). Additionally the study can be separated into two distinct study phases: (1) Pre-IRS: two surveys prior to the IRS intervention (Survey 1 October 2012; Survey 2 May/June 2013); and (2) Right post-IRS: two surveys immediately following the three-rounds of IRS (Survey 3 October 2015; Survey 4 May/June 2016) (Appendix 1-Figure 5E). Details on the study area/population, study design, malaria control interventions (i.e., IRS and LLINs), inclusion/exclusion criteria, data collection/generation procedures, routine surveys on whom in the population having taken antimalarial treatment in a previous window of time etc. have been previously described (***Tiedje et al., 2017***, ***2022***, ***2023***).

Inferring FOI from MOI estimates

Mathematical consideration of malaria transmission in relation to queuing theory

Mathematical models often consider that in a cohort of uninfected people acquiring and clearing infections independently, inoculations arrive according to a homogeneous Poisson process with rate equal to the mean FOI, with each inoculation exhibiting an exponentially distributed duration. It follows that at equilibrium, MOI is a Poisson variable with mean given by the mean FOI divided by the mean clearance rate (***Dietz et al., 1974***). In empirical settings, the arrival of inoculations often deviates from a homogeneous Poisson process due to various factors including seasonality and heterogeneity in exposure risk. Under heterogeneous biting of different hosts, the MOI distribution can be captured by a negative binomial distribution to account for a higher variance. There is no simple relationship in this case, however, between the parameters of the negative binomial distribution and the mean FOI and mean clearance rate. We perform a formal test for any deviation from Poisson homogeneity of the empirical and simulated MOI estimates, with details and results of the test provided in Appendix 1-Test of deviation from Poisson homogeneity in MOI estimates and supplementary file 3-deviationFromPoissonTest.xlsx.

Despite these complexities, the process of acquisition of infectious bites is structurally equivalent to stochastic queuing theory, which aims to explain how the length of a queue is related to the intensity of arrivals to the system, the systems' priority schedule, and the service and waiting times. Queuing theory is indeed widely employed across multiple applications including in communications, computer architecture, and operation management. The queuing system is usually modeled by a system of differential equations (with the Kolmogorov equations). Their structure depends on specific assumptions of the model, for example, single server vs. multiple servers, stationary vs. non-stationary arrivals, different priory schedules, and so on. The three main information components of a queuing system are the length of the queue, the intensity of arrivals to the system, and the service times, all of which are coupled. In theory, knowing any two components would allow inference about the third. A simple quantitative description of the process of malaria transmission would be analogous to customers arriving to, and departing from, a service facility. In other words, each individual host is equivalent to a service facility center with a given number of servers, i.e, the carrying capacity for blood-stage infections in the parasite. Moreover, each infection is equivalent to a customer, with infections arriving at an individual host over time and each infection lasting a given amount of time before clearance. The empirical MOI estimates provide information about the length of queues. Hypothetically, when the duration of infection is known, one could infer backwards the information about the rate of arrival of infections to individuals, in other words, the FOI.

However, the duration of a single infection is difficult to determine in field samples from en-

demic areas, where mono-clonal infections are rare but multi-genomic infections are common. Commonly-used polymorphic markers may not have a high power to differentiate different strains that are co-infecting individual human hosts (***Argyropoulos et al., 2023***). Reasons include that the diversity of those polymorphic markers can be limited, with two parasites sharing the same marker but having distinct sets of antigen-encoding genes. It is thus difficult to accurately measure the emergence and absence from the peripheral blood of any individual strain. Because *var* genes frequently undergo ectopic recombination and change the location of their homology blocks on different chromosomes across the genome, it is difficult to assign *var* genes to an individual genome and to a specific chromosomal location. The difficulty of phasing precludes the integrity of individual infections and thus the ability to isolate these, as well as the tracking of their corresponding first appearance in the blood and their clearance in time. Also, the duration of infection measured in field studies is highly variable. Examples include infections within individuals of different age groups and therefore different immune histories, but also infections from different geographical locations and times and thus circulating under different transmission intensity and different genetic diversity (***Childs and Buckee, 2014***; ***Chang et al., 2016***; ***Ashley and White, 2014***; ***Bretscher et al., 2011***).

We therefore propose the alternative strategy of focusing on the MOI estimates of the youngest age group in the population, that is, on those of the 1-5 year-old children. Because these children have received a much lower number of infections than the rest of the population, they have not yet seen much of the circulating diversity. It is thus safe to consider negligible the impact of immune memory accumulated from previous infections on the duration of a current infection. Therefore, infection duration in these children can be approximated by that of naive hosts. Estimates of the latter are available from a historical medical intervention in which patients with neurosyphilis were infected with malaria as a therapy. Between 1940 and 1963, 318 patients with syphilis were intentionally infected with a single strain of *P. falciparum* (***Collins and Jeffery, 1999***; ***Maire et al., 2006***). Data on fever and the number of malaria parasites in the blood were recorded. Because those patients had never been infected with *P. falciparum* malaria prior to the treatment, the documented infection duration is representative of naive infections, and can serve as a proxy for infection duration in young children. With MOI estimates and the duration of infection at hand, we can then infer the FOI experienced by these children using methods from queuing theory as described next. An additional motivation for focusing on 1-5 year-old children is that they are the most vulnerable group of individuals in the population, for whom mortality, severe and clinical cases are most prominent.

A two-moment approximation for a queue of finite capacity

Analysis of multi-server models in general is notoriously difficult. Exact results are only available for some special models, including the previously mentioned $M/M/c/k$ ones which require that both the inter-arrival times and service times are exponentially distributed. Additional examples of those special models include the $M/G/c/c$ (including the infinite-server $M/G/\infty$) and $GI/M/c/c+r$ queues which require that either the inter-arrival times or the service times are exponentially distributed. As such, those special models are often too restrictive to apply to real-world queuing situations in which arrivals do not follow a homogeneous Poisson process (i.e, for which the inter-arrival and service times are not exponentially distributed). Application of a two-moment approximation method is examined here for the steady state analysis as proposed by Kim and Cha (***Choi et al., 2005***), which considers the inter-arrival and service times as independent sequences of independent and identically distributed (i.i.d.) general random variables (r.v.'s). The approximations are obtained by replacing unknown arrival- and departure-average quantities by their corresponding (well-known) time-average counterparts, which are exact for exponential inter-arrival and service times. More mathematical details and derivations can be found in Appendix 1-The mathematical details for a two-moment approximation for a queue of finite capacity.

We vary here the mean and variance parameters for inter-arrival times across wide ranges. For

each combination of a specific mean and variance value, we calculate a probability density distribution for the queue length, i.e., a probability density distribution of MOI. We identify the combination of mean and variance values that maximizes the likelihood of observing the given distribution of MOI estimates in either the simulation output or the empirical data from Bongo District. We provide the details for deriving confidence intervals for the estimated mean and variance in Appendix 1-Confidence intervals for mean and variance.

In the empirical settings with seasonal transmission and heterogeneous exposure risks, the inter-arrival times of infection are likely to be complex and heterogeneous in time. By applying this two-moment approximation method, we treat those inter-arrival times as if they followed a general distribution of any kind with a specific mean and a specific variance to be inferred, and consider consecutive infections independent. Those assumptions are general and flexible enough to be applicable to a variety of empirical settings.

Finally, because the goal of this work is to utilize widely available MOI estimates sampled under sparse schemes across geographical space and time, alternative approaches such as fitting a non-homogeneous Poisson process or Gamma process to derive time-varying average arrival rates cannot be considered due to their requirements for densely sampled time-series data.

When applying this approach, one needs to specify values for the number of parallel servers (c) and the number of waiting places (r). The former corresponds to the maximum number or carrying capacity of blood-stage infections. The maximum MOI observed in the empirical data from Bongo is 20. Due to additional factors which result in a reduced number of *var* genes sequenced and typed but are not explicitly accounted for in current MOI estimation (see Discussion), the actual carrying capacity could be higher. For simplicity, we assume it to be 30 in the simulation. Without loss of generality, the chosen value for this quantity should not raise any issues for FOI inference based on simulation output, as long as one keeps it consistent when generating simulation output and when applying the two-moment approach to the simulation output. For FOI inference based on the empirical surveys from Bongo, we set this quantity to be either 25 or 30. Because the results of FOI inference are robust against those two values, we report the one with $c = 30$. Because FOI represents the number of effective infections which reach the blood stage and MOI is defined for blood-stage infections only, by default, the number of waiting places r should be set to be 0.

The mean arrival rate of infection from Little's law

The second method is known as Little's law (***Little and Graves, 2008***), which describes a relationship between the three main information components of queuing systems. This law states that the average number of items in a queuing system L equals the average arrival rate λ multiplied by the average waiting time of an item in the system, W. Reformulating Little's law in terms of malaria transmission, the average arrival rate of infection λ equals the average number of blood-stage active infections present in an individual L divided by the average duration of infection W.

$$\lambda = \frac{L}{W} \tag{1}$$

The relationship is remarkably simple and general. It requires that in the long run the average number of arrivals is less than or equal to the average number of exits, such that the queue does not overflow. The law applies however irrespective of the number of servers (here, the carrying capacity of blood-stage infections of hosts), the service time distribution (here, infection duration distribution), the distribution of inter-arrival times, the order of items' service, and whether each server has its own queue or a single queue feeds all servers. In theory, individual human hosts can experience "overflowing" under high-transmission settings with high EIR values, but evidence has suggested that the majority of these infectious bites do not result in detectable blood-stage infections. Because FOI by definition represents the rate of arrival of effective infections acquired by human hosts that have advanced to the blood-stage, and not the rate of arrival of all infections (which is known as EIR), we consider that the requirement approximately applies in the context

of FOI inference for malaria transmission. In other words, values of FOI do not include and hence should not be influenced by the loss of excessive arrivals due to various factors including immunity, measurement bias, and so on.

Little's law holds for an evolving time series as it is observed over a long period of time. Although we do not have time series of MOI estimates for any individual host, we treat different sampled individual hosts as if each one represented an observation of the process at a single time point. This is sensible because different individual hosts are likely to receive infections in an asynchronous manner, and therefore, the aggregation of MOI information across different individual hosts can function as a proxy for a time series of MOI values for an individual host.

Conversion between FOI and the Entomological Inoculation Rate, or EIR

As an indirect proxy for transmission intensity, malariologists typically measure EIR by counting how many infectious bites are received within a time interval by a human host seated in a fixed place (***Shaukat et al., 2010***). The resulting value of EIR is considered a standard metric of malaria transmission. Though FOI and EIR both reflect transmission intensity, the former refers directly to detectable blood-stage infections whereas the latter concerns human-vector contact rates. Studies reporting both the annual *P. falciparum* EIR and FOI estimates from detailed age-stratified prevalence in cross-sectional parasitological studies, have obtained very different magnitudes for these two quantities (***Smith et al., 2010***). We refer here to the number of infections (in the blood stage) per infectious bite (FOI/EIR) as transmission efficiency. Several studies have shown that malaria transmission is inefficient in high-intensity settings, and there has been a long-standing debate on the reason why. Heterogeneous biting, immunity, or measurement bias have been proposed as causes for explaining this discrepancy between EIR and FOI (***Donovan et al., 2007***; ***John et al., 2005***; ***Doolan and Martinez-Alier, 2006***; ***Smith et al., 2010***). Within host competition for resources between co-infecting strains and the resulting dominance of a subset of those strains, with the remaining others at low parasitemia, can also account for this discrepancy. Parasites may not be transmitted from an infectious mosquito to a human during the blood meal due simply to chance. We borrow a functional curve with estimated parameters under the assumption of heterogeneous transmission, which describes the highly non-linear relationship between these reported EIR-FOI pairs (***Smith et al., 2010***). The functional curve is in the format of the following equation with h denoting FOI and E denoting EIR. The corresponding parameters are as follows: $b = 0.55$, $\alpha = 4.6$ and $t = 43$ days.

$$h = \frac{log(1 + \alpha b E t)}{\alpha t} \tag{2}$$

The combination of these reported EIR-FOI pairs (***Smith et al., 2010***) and the functional curve provides a basis for converting from one quantity to the other.

The EIR information for a subset of surveys in Bongo District from northern Ghana was obtained (***Tiedje et al., 2022***). Together with the FOI estimates from the two methods from queuing theory, we have the EIR-FOI pairs for empirical surveys in Bongo District. Therefore, we can address whether our EIR-pairs are consistent with those previous historical collections of EIR-FOI pairs and the functional curve with the best-fit parameter values to those collections.

Appendix

The measurement error

We utilize a measurement error model in MOI estimation which accounts for the under-sampling or imperfect detection of *var* genes by sub-sampling *var* genes per strain, thus reducing the number of available distinct *var* genes per host. This measurement error model is based on the repertoire size distribution. The molecular sequences used for deriving the repertoire size distribution were previously sequenced from infections that are expected to be monoclonal (MOI = 1), i.e., hosts infected by a single *P. falciparum* strain, as they had less than or equal to 45 non-upsA DBLα types,

sampled during six cross-sectional surveys made from 2012 to 2016 in Bongo District from northern Ghana (Appendix 1-Figure 5D) (***Tiedje et al., 2023***, ***2022***; ***Pilosof et al., 2019***).

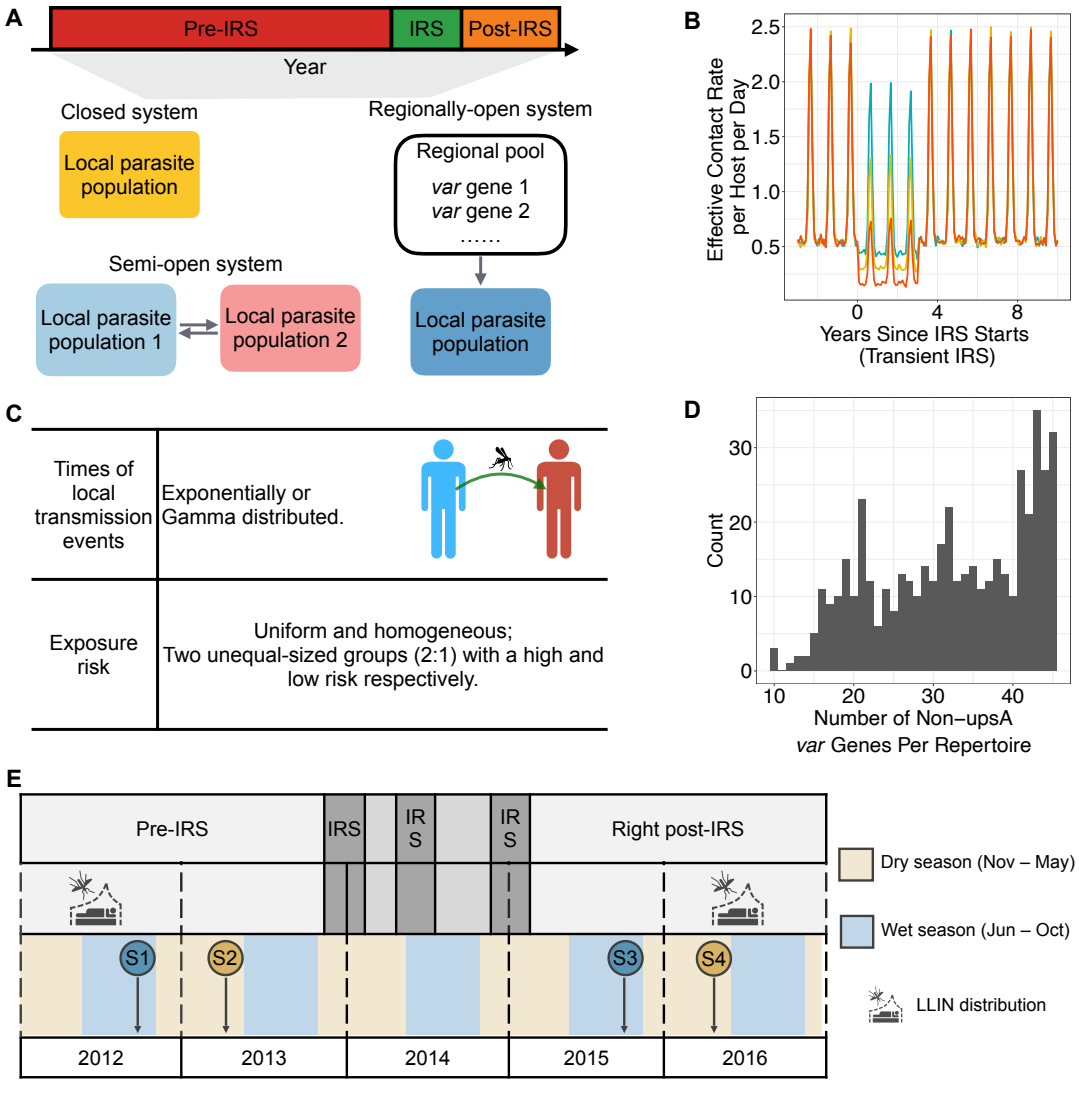
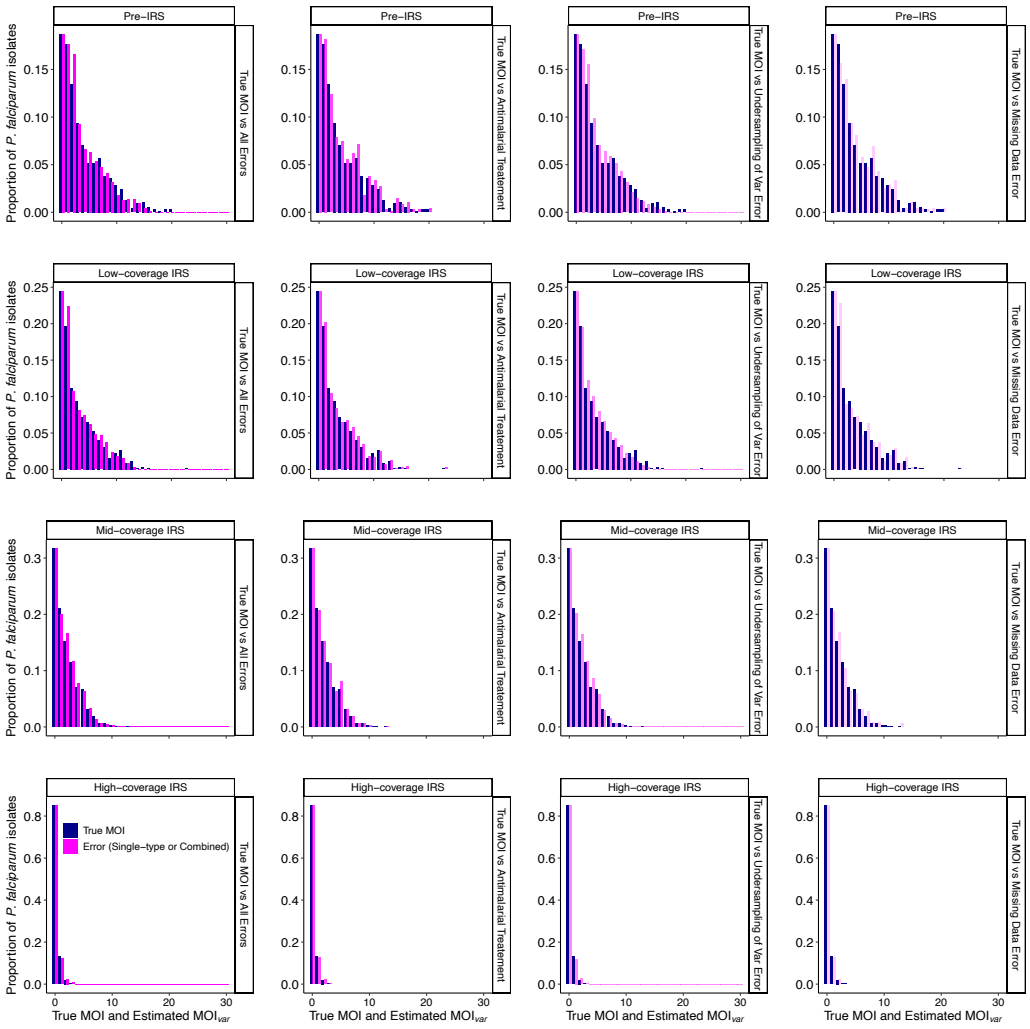


Figure 5. (A) Each simulation follows three stages: a "pre-IRS" period during which the transmission in the local population reaches a stationary state, followed by an "IRS" intervention period of three years which reduces transmission rate (in what we call transient IRS), and a "post-IRS" period when transmission rates go back to their original levels. We let the system run for some years to reach a (semi-) stationary state before applying IRS interventions. After initial seeding, closed systems do not receive migrant genomes from the regional pool. Semi-open systems are explicit simulations of two individual local populations coupled via migration events. Regionally-open systems receive continuous migrant genomes from the regional pool throughout the simulation course. (B) Transmission intensity or effective contact rate (Appendix 1-Simulation data) varies as a function of time, from pre-, to during, to post-intervention. We simulate three different coverages of perturbations, ranging from low-coverage one which reduces the system's transmission by around 20% to high-coverage one which reduces the system's transmission by around 70-75%. (C) We incorporate heterogeneity in transmission across individual hosts by varying two things. First, we consider different statistical distributions for times of local transmission events: exponential and Gamma. Second, we consider homogeneous and heterogeneous exposure risks. For the latter, we set $\frac{2}{3}$ of the total population as being at high-risk and receiving more than 90% of all bites whereas the rest population receives only less than 10% of all bites. (D) The measurement error which describes potential sequencing errors of *var* genes: histogram of the number of non-upsA (i.e., upsB and upsC) DBLα *var* gene types per repertoire. The molecular sequences were previously sequenced from infections that are expected to be monoclonal (as they had less than or equal to 45 non-upsA DBLα types), sampled during six cross-sectional surveys made from 2012 to 2016 in Bongo District. (E) Study design showing the timing of the four age-stratified cross-sectional surveys conducted in Bongo District, Ghana at the end of the wet/high-transmission seasons (blue circles) and at the end of the dry/low-transmission seasons (gold circles). The study can be broken down into two phases: (1) Pre-IRS: Survey 1 (S1) October 2012 and Survey 2 (S2) May/June 2013; (2) Right post-IRS: Survey 3 (S3) October 2015 and Survey 4 (S4) May/June 2016. IRS were implemented against a background of widespread LLIN usage which were distributed across Bongo District between 2010 – 2012 and again in 2016 (***Tiedje et al., 2023***; ***Gogue et al., 2020***).



775

Figure 6. Comparison of the estimated MOI distribution under only one type of error among those commonly encountered in the collection of empirical data (missing data, under-sampling of *var* genes, and usage of antimalarial drugs) versus under all the types of the errors together and the distribution of true MOI values (indicated by different shades of the red). The scenarios shown here are simulated as semi-open systems under seasonal transmission. Individuals are under heterogeneous exposure risk (Appendix 1-Figure 5C) and the arrival of infection follows a Gamma distribution. The Bayesian *var*coding method and the bootstrap imputation approach are able to address all types of the errors. Results of the comparison across other simulated scenarios can be found in supplementary file 1-MOImethodsPerformance.xlsx

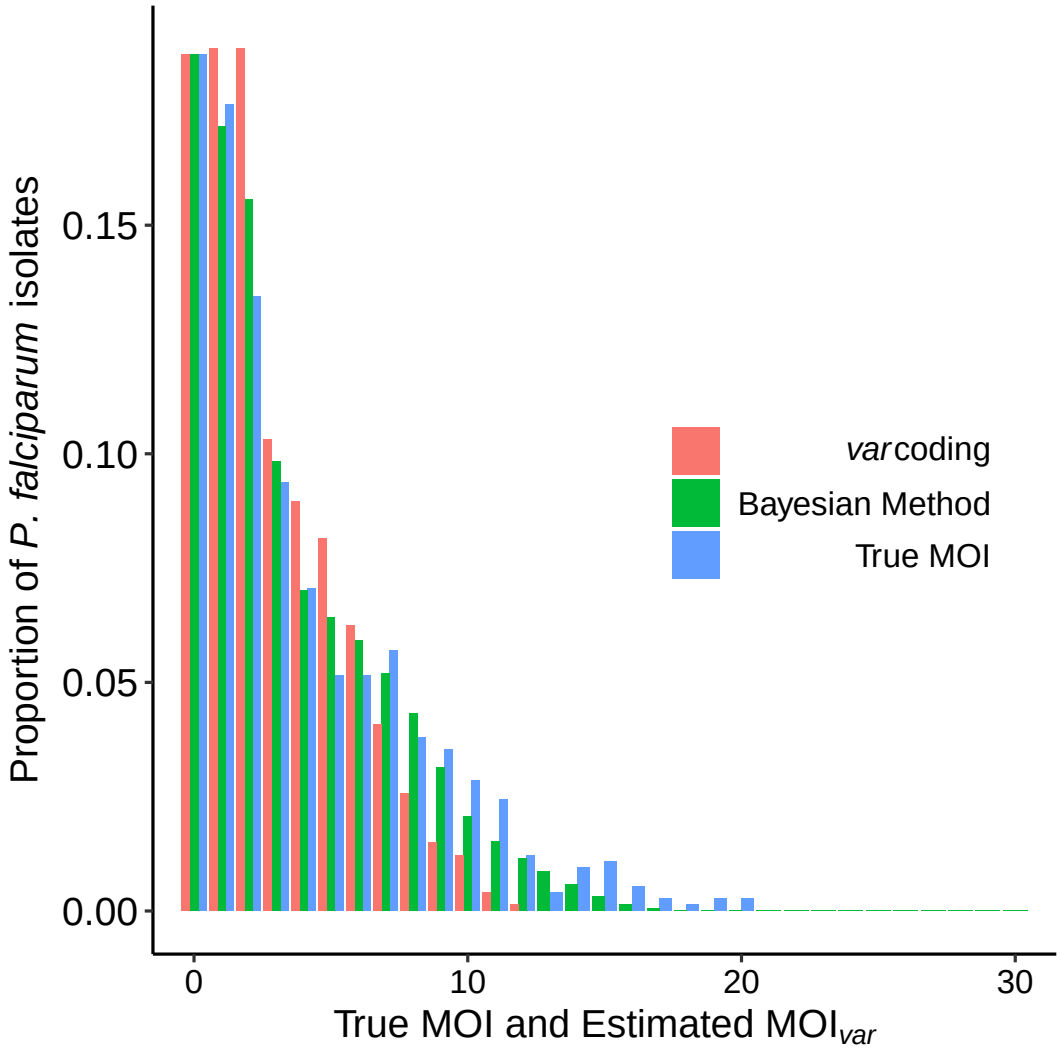


Figure 7. Comparison of the estimated MOI distribution based on the original *var*coding method, the Bayesian approach, and the true values. The scenarios shown here correspond to a semi-open system under seasonal transmission. Arrival of infection follows a Gamma distribution. Both the Bayesian *var*coding approach and the original method perform well, with the former showing a clear improvement in capturing the tail of high MOI values. Results of the comparison across the remaining different simulated scenarios can be found in supplementary file 2-BayesianImprovement.xlsx.

Simulation data

An extended *var* model and experimental designs

To evaluate the relative performance of the two proposed methods from queuing theory for FOI inference across different transmission settings, we generate simulation output using a computational model. This model is an extension of an agent-based, discrete-event, stochastic model in continuous time (***He et al., 2018***). The original implementation is adapted from the next-reaction method (***Gibson and Bruck, 2000***) which optimizes the Gillespie first-reaction method (***Gillespie, 1976***). It tracks the infection history and immune memory of each host. Parameters or symbols, and their values are summarized in the supplementary file 4-simParams.xlsx. We model a local population of 10000 individuals. All simulation runs are initialized with 20 infections to seed local population transmission. We let each simulation run for some years (200 years for closed systems, 150 years for open systems) to reach a (semi-) stationary state before introducing transmission-reducing interventions.

The size of a single parasite repertoire is assumed to be 45 here, based on the median number of non-upsA DBLα sequences identified in our 3D7 laboratory isolate (***Ruybal-Pesántez et al., 2022***; ***Tiedje et al., 2022***). This grouping of *var* genes is defined based on their semi-conserved upstream promoter sequences (ups) (correspondingly for upsA and non-upsA (upsB and upsC)) (***Gardner et al., 2002***; ***Lavstsen et al., 2003***; ***Kraemer et al., 2007***; ***Rask et al., 2010***). Although each parasite carries both types of *var* genes in a fairly constant proportion (***Ruybal-Pesántez et al., 2017***; ***Tiedje et al., 2023***; ***Ruybal-Pesántez et al., 2022***; ***Buckee and Recker, 2012***; ***Rask et al., 2010***), the MOI estimation method we consider here focuses on the non-upsA DBLα sequences as they are around 20 times more diverse and less conserved among repertoires than the upsA DBLα sequences. Although there is evidence indicating functional differences between the two groups, the groupings do not necessarily correlate with function. Within-group functional heterogeneity and variation, and cross-group functional similarity exist (***Claessens et al., 2012***; ***Kaestli et al., 2006***; ***Rottmann et al., 2006***). In addition, the mechanism underlying the fairly constant proportion of two structural groups in empirical samples across time and geographical locations remains to be properly understood. It is therefore unclear how to model functional divergence of *var* groups in a rigorous and parsimonious way, and whether assumptions made on the particular coarse-graining apply to the population dynamics of the disease. Therefore, for simplicity purposes, our model considers only one type of gene, corresponding to the non-upsA types.

Each *var* gene itself is represented as a linear combination of two epitopes (the parts of the molecule that act as antigens and are targeted by the immune system) (***Bull et al., 2008***; ***Larremore et al., 2013***; ***Rask et al., 2010***; ***He et al., 2018***).

The *var* genes in a repertoire are expressed sequentially and the infection gets cleared when the whole repertoire is depleted. The duration of expression of a *var* gene is proportional to the number of unseen epitopes by the infected host. The total duration of infection of a particular repertoire is proportional to the number of unseen epitopes by the infected host aggregated across all individual *var* genes of that repertoire. When a *var* gene is deactivated, the host adds the deactivated *var* gene epitopes to its immunity memory. Specific immunity toward a given epitope wanes at a certain rate, and re-exposure is therefore required to maintain it.

We simulate malaria dynamics with parameters representative of high-transmission endemic regions, with either constant or seasonal transmission, and across different spatial configurations, including closed, semi-open, and regionally-open systems. We apply perturbations in our model that represent interventions targeting the mosquito vector with indoor residual spraying (IRS) that involves the application of an insecticide to internal walls and ceilings of homes (***World Health Organization, 2015***). IRS effectively reduces mosquito population, and therefore, transmission rate and growth rate of the parasite population as a whole. We simulate three temporary IRS interventions of different coverages, ranging from low- to mid- to high- coverage respectively. The low-coverage intervention reduces the system's transmission by around 20%; the mid-coverage one reduces by around 45%; and the high-coverage one reduces by around 70-75%. Each IRS intervention lasts for 3 years.

Mosquitoes are not represented as explicit agents in the model but their role in transmission is implemented via contact events between human hosts. We consider an effective contact rate (hereafter, the transmission rate or transmission intensity) which determines the times of local transmission events. These times can follow different distributions, including exponential and Gamma ones. At these times, a donor and a recipient host are selected randomly and successful transmission occurs only if the donor carries active (blood-stage) infections and the liver stage of the recipient is below a specified carrying capacity of 30.

We implemented seasonality by multiplying a scaling constant by a temporal vector of 360 days, containing the daily number of mosquitoes over a full year. In other words, we have a basic vector representing the number of mosquitoes and a scaling constant encapsulating all other parameters involved in vectorial capacity. The product of both gives the effective contact rate. To obtain this temporal vector, (***Pilosof et al., 2019***) used a deterministic model for mosquito population dy-

namics from (***White et al., 2011***). The model was originally developed for *Anopheles gambiae* and consists of a set of ordinary differential equations describing the dynamics of 4 mosquito stages: eggs, larvae, pupae, and adults. Seasonality is implemented via density dependence at the egg and larva stages as a function of rainfall (availability of breeding sites). The values of the effective contact rate rather than the absolute number of daily mosquitoes are the key parameters for the purpose of this work.

For closed systems, after the initial seeding of local transmission from a regional pool of *var* genes, migration is discontinued. Semi-open systems are an explicit simulation of two individual populations coupled via migration. Regionally-open systems are a local population with migration from a regional pool. The regional *var* gene pool of a certain size acts as a proxy for regional parasite diversity, i.e., diversity from the aggregated individual populations in the region. Parasite genomes can migrate from the regional pool to the local population. Because each parasite genome is a repertoire with a given number of *var* genes, migrant genomes are assembled from the random sampling of this number of *var* genes from the regional pool.

Transmission can be homogeneous or heterogeneous across individual hosts. For the latter, we consider two groups of human hosts: the high risk group receives more than 90% of the total infection events, whereas the low risk group receives the remaining fraction. The two risk groups have different sizes, and the size of the high-risk group to that of the low-risk one is 2:1.

Test of deviation from Poisson homogeneity in MOI estimates

A great deal of research has addressed the statistical question of whether a count dataset exhibits significant deviation from a homogeneous Poisson distribution. We select the Potthoff-Whittinghill "index of dispersion" test (***Potthoff and Whittinghill, 1966***; ***Lloyd-Smith et al., 2005***), which is asymptomatically locally most powerful against the negative binomial alternative. For a dataset X with N elements, this statistic is given by $\frac{(N-1)*var(X)}{mean(X)}$ and its asymptotic distribution is chi-squared with $N-1$ degrees of freedom. A p-value is obtained by determining the cumulative density of the chi-squared $(N-1)$ distribution to the right of the test statistic, and represents the probability that the observed variance arose by chance from a Poisson distribution.

We summarize the p-value of the Potthoff-Whittinghill "index of dispersion" test in supplementary file 3-deviationFromPoissonTest.xlsx. The p-values are very close to 0, indicating that the deviation from a Poisson distribution is highly significant for both simulation output and empirical surveys from Bongo District. A few exceptions are observed when intervention is high-coverage pushing the system's prevalence to be very low and the majority of infections become mono-clonal.

The mathematical details for a two-moment approximation for a queue of finite capacity

Following (***Choi et al., 2005***), we describe here the consideration of the $GI/G/c/c+r$ queue with c (≥ 1) identical servers in parallel and r (≥ 0) waiting places. For an individual host, we consider c carrying capacity in blood stage and $r=0$ carrying capacity for the waiting space. Inter-arrival and service times of customers are independent sequences of independent and identically distributed (i.i.d.) general random variables (r.v.'s).

N represents the number of customers in the system at an arbitrary time. $N^A(N^D)$ denotes the number of customers that an arriving customer finds (that a departing customer leaves behind) in steady state. Customers who arrive to find $c+r$ customers in the system are assumed to depart immediately from the system with those $c+r$ customers left behind. P_n, P_n^A and P_n^D denote the probabilities that $N=n$, $N^A=n$, and $N^D=n$, respectively, for $0 \leq n \leq c+r$.

The previous probabilities will be expressed in terms of the following quantities:

$$a = E(A) = \frac{1}{\lambda}$$
$$b = E(S)$$
$$a_n^D = E[A_n^D], 0 \le n \le c + r$$
$$b_n^A = E[S_n^A], 0 \le n \le c + r$$
$$b_n^D = E[S_n^D], 0 \le n \le c + r$$

A_n^D with $0 \le n \le c + r$ is the residual inter-arrival time at the departure instant of a customer who leaves behind n customers in the system. $S_n^A(S_n^D)$ with $1 \le n \le c + r$ is the residual service time of a randomly chosen busy server at the arrival instant (the departure instant) of a customer who finds (leaves behind) n customers in the system.

A two-moment approximation for the steady-state queue-length distribution as follows:

$$\tilde{P}_n^A = \tilde{P}_n^D = \tilde{P}_0^A \Pi_{i=0}^{n-1} \frac{\tilde{\lambda}_i}{\tilde{\mu}_{i+1}}, 1 \le n \le c + r$$
$$\tilde{P}_n = \tilde{P}_n^A \tilde{\gamma}_n, 0 \le n \le c + r$$

where

$$\frac{1}{\tilde{\mu}_i} = \begin{cases} b - i(a - a_R), 1 \le i \le c \\ -c(a - a_R) + b_R, c + 1 \le i \le c + r \end{cases}$$

$$\frac{1}{\tilde{\lambda}_i} = \begin{cases} (i + 1)a_R, 0 \le i \le c - 2 \\ ca_R + b_R - b, c - 1 \le i \le c + r - 2 \\ ca, i = c + r - 1 \end{cases}$$

$$\tilde{\gamma}_i = \begin{cases} \lambda a_R, i = 0 \\ \lambda[\tilde{\mu}_i \frac{(a - a_R)}{\tilde{\lambda}_{i-1}} + a_R], 1 \le i \le c + r - 1 \\ \lambda[\tilde{\mu}_i \frac{(a - a_R)}{\tilde{\lambda}_{i-1}} + a], i = c + r \end{cases}$$

And, by normalization, $\sum_{n=0}^{c+r} \tilde{P}_n^A = 1$:

$$\tilde{P}_0^A = (1 + \sum_{n=1}^{c+r} \Pi_{i=0}^{n-1} \frac{\tilde{\lambda}_i}{\tilde{\mu}_{i+1}})^{-1}$$

Confidence intervals for mean and variance

Non-parametric bootstrap

Bootstrap datasets were generated by resampling with replacement from the original data. For each bootstrap dataset, the maximum likelihood estimates of the mean and the variance of FOI were determined, generating a bootstrap sampling distribution. We ran 200 bootstrap replicates because this number has been tested and proven to approach a similar coefficient of variation than a higher number of replicates (***Efron and Tibshirani, 1994***).

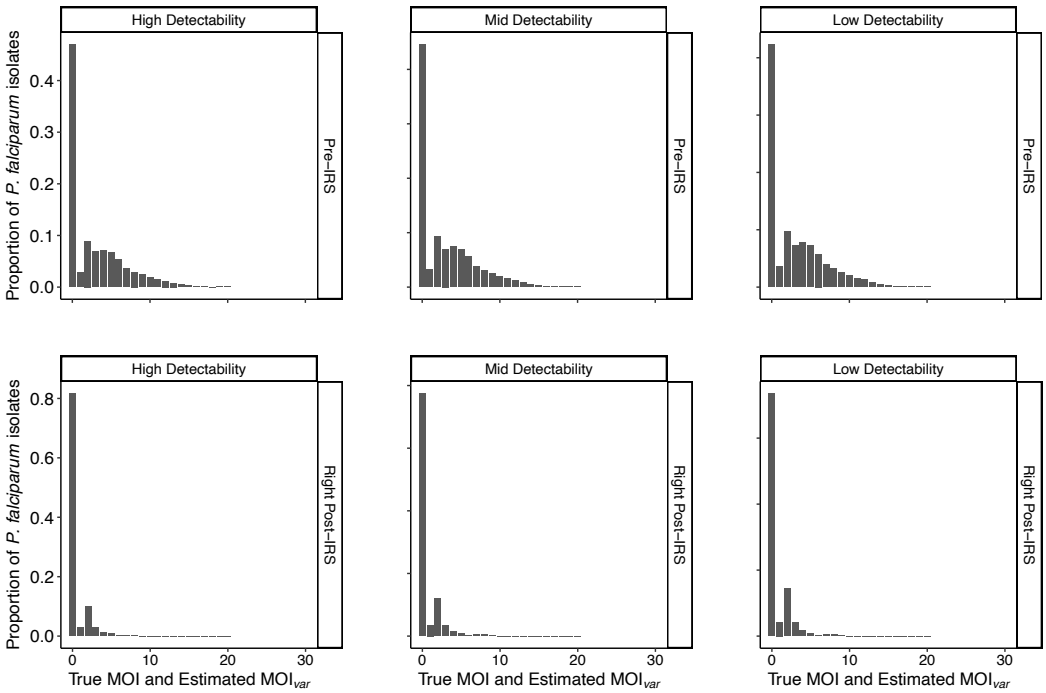
A comparison between the performance of the Bayesian approach and the original *var*coding method from the simulation output

We compare the performance of the original *var*coding method and the Bayesian formulation formally against true MOI values with a Kolmogorov-Smirnov Test. Specifically, we calculate the maximum vertical deviation between the two cumulative density function of MOI distribution obtained from either of the two methods against each other, and each against true values respectively. Additionally, we compare the mean value of MOI estimates obtained from the two methods and true MOI.

The Bayesian method shows improvement relative to the original *var*coding method which relies on a constant repertoire size to convert the number of detected non-upsA DBLα types to the estimated MOI value. We document the Kolmogorov-Smirnov statistics and the difference in the mean values of MOI estimates in supplementary file 2-BayesianImprovement.xlsx. This improvement is expected given consideration of variation in repertoire size due to the under-sampling of *var* genes. Both methods perform well for low or mid transmission scenarios (corresponding to high- or mid-coverage intervention scenarios respectively) when the majority of true MOI centers around low or intermediary values. The *var*coding method has decent performance for high transmission scenarios, with the Bayesian method showing further notable improvement in capturing the tail of high MOI values. The Bayesian method improves estimation accuracy for high MOI values in particular because the cumulative effect of the measurement error for multiple infections becomes more significant. In high-transmission endemic regions, a higher fraction of infected individuals harbors more infections, and thus have higher MOI values. The improvement can be more evident for these regions, including Bongo District, Ghana for which we obtained our empirical MOI estimates.

Both the Bayesian and the original version of *var*coding approach tend to underestimate MOI (see Discussion). Overall, the distribution of the estimated MOI from both methods is quite similar to that of the true MOI, across all simulated scenarios. The Bayesian MOI estimation is fairly accurate and robust in addressing the under-sampling of *var* genes error. Nevertheless, we note that the proposed approaches for estimating FOI can be based on MOI values obtained with any method and any polymorphic marker other than the *var* multigene family.

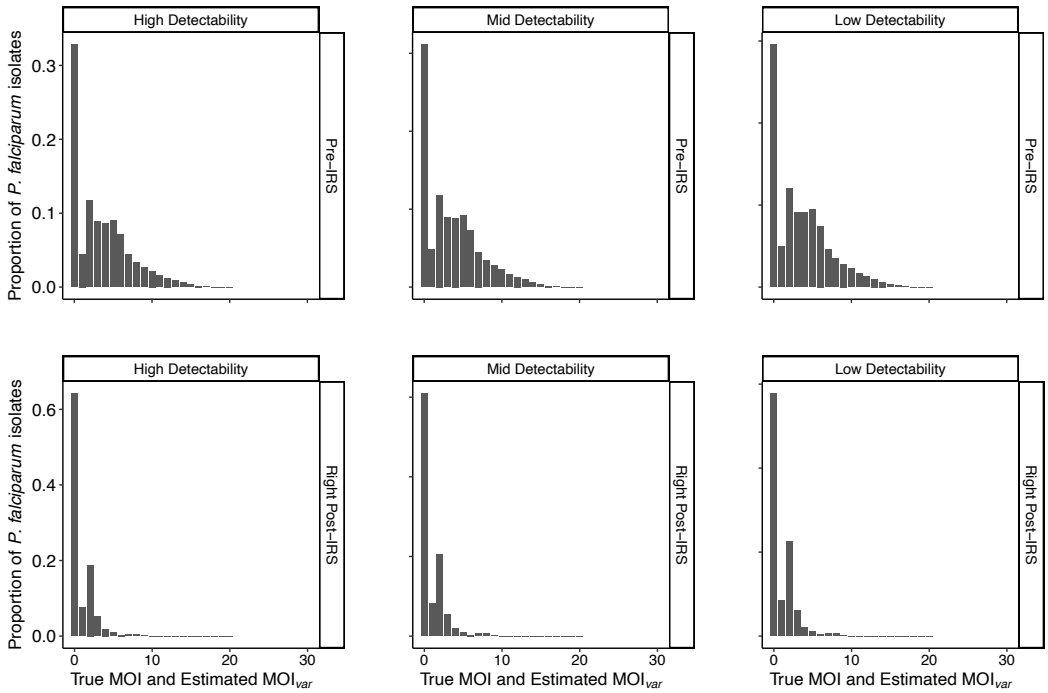
929



930

Figure 8. The distribution of MOI estimates based on the Bayesian *var*coding method and the bootstrap imputation approach for longitudinal surveys in Bongo District, from pre-IRS to right post-IRS. See Appendix 1-Figure 5E for the study design of the IRS intervention. The wet/high season survey for pre-IRS phase was collected in 2012 and denoted as S1 in Figure 5E. The dry/low season survey for pre-IRS phase was collected in 2013 and denoted as S2. The wet/high season survey for right post-IRS phase was collected in 2015, corresponding to S3. The dry/low season survey for right post-IRS phase was collected in 2016 and denoted as S4. As explained in the methods, the empirical surveys suffer from missing data, i.e., only microscopy-positive individuals have their *var* genes typed and sequenced. We quantify the number of individuals with missing MOI by utilizing the empirically measured detectability/sensitivity of microscopy relative to PCR, and assuming a range for the PCR detectability. We consider a high detectability in which PCR picks up all infected individuals, to a mid value in which 5% PCR-negative individuals actually carry infections, to a low value in which 10% PCR-negative individuals carry infections. The empirical surveys also suffer from missing data due to other factors than the detection limit of the selected method, for example, low DNA quality. We impute MOI values for all those individuals with missing MOI by sampling from MOI estimates of microscopy-positive individuals with their *var* sequenced and typed successfully who are not treated. In addition, some individuals may seek and get antimalarial treatment in a previous window of time. We consider here the extreme scenario that antimalarial treatment has no impact on treated individuals' infection status and MOI values, and use their infection status and MOI values directly for FOI inference.

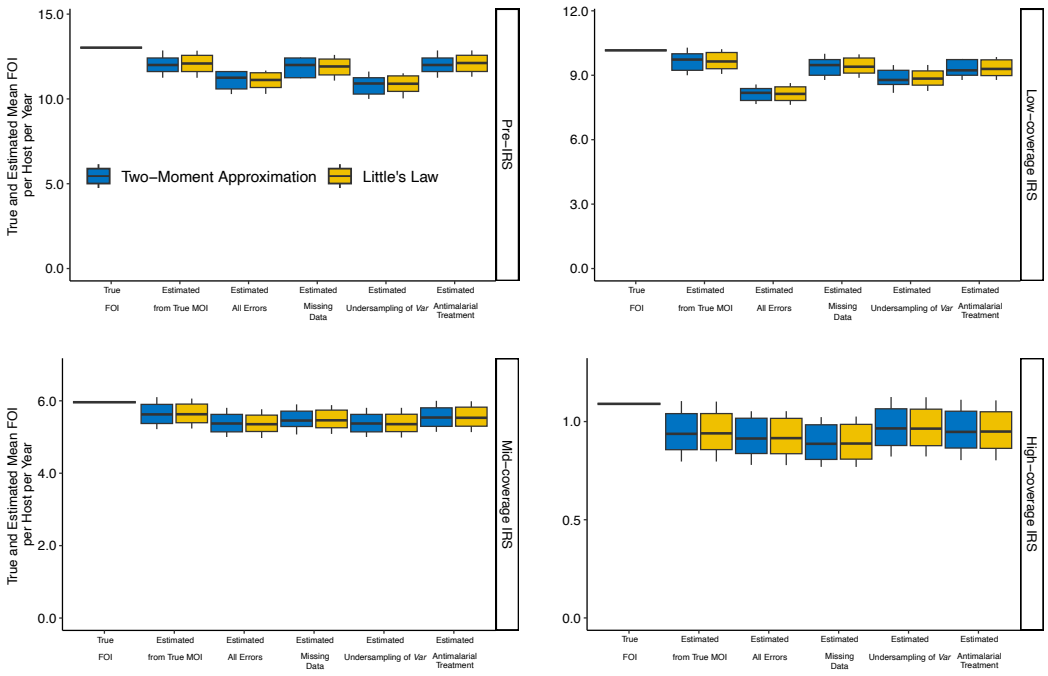
931



932

Figure 9. The distribution of MOI estimates based on the Bayesian *var*coding method and the bootstrap imputation approach for longitudinal surveys in Bongo District, from pre-IRS, to right post-IRS, to post-IRS/SMC. We assume here the other extreme scenario that antimalarial treatment shifts and changes treated individuals' infection status and MOI values completely. We thus impute MOI values for those treated individuals by sampling from MOI estimates of microscopy-positive individuals with their *var* sequenced and typed successfully who are not treated.

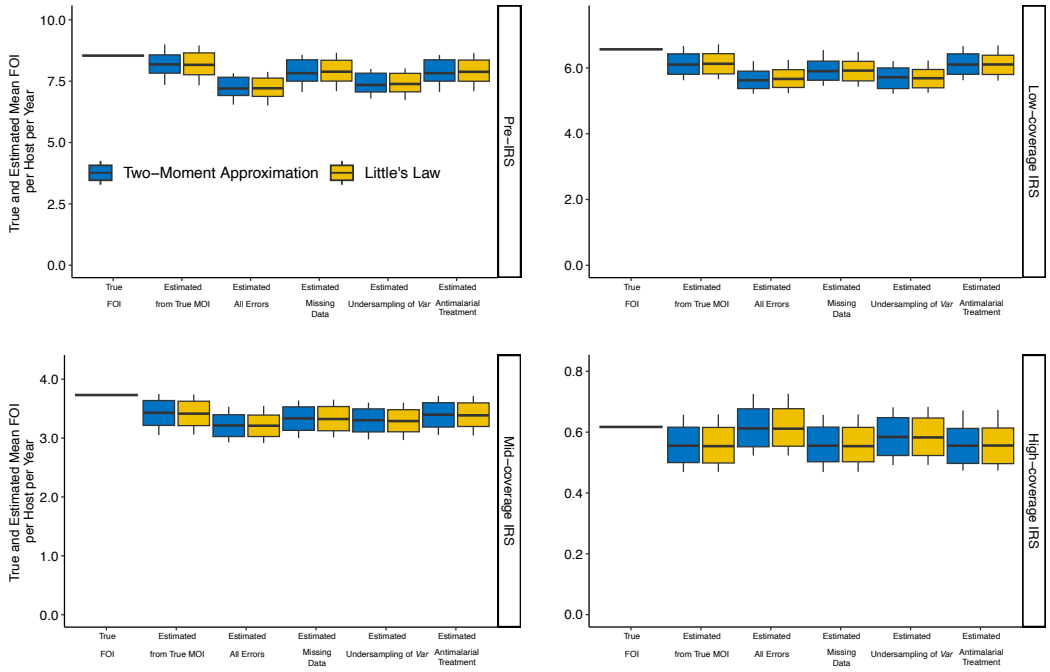
933



934

Figure 10. The true and estimated FOI by the two-moment approach and Little's law for additional simulated homogeneous exposure risk scenarios, under non-seasonal transmission, for a closed system. The times of the local transmission events are Gamma-distributed. The true mean FOI per host per year is calculated by dividing the total number of infections acquired by the population by the total number of individual hosts in the population. Minimum, 5% quantile, median, 95% quantile, and maximum values are shown in the boxplot.

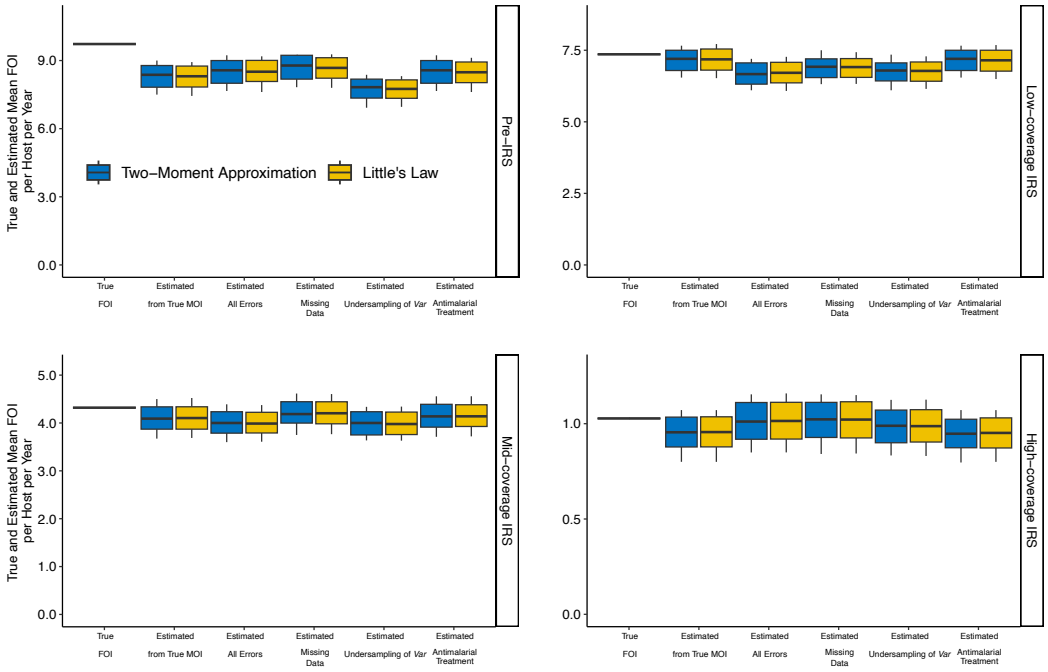
935



936

Figure 11. The true and estimated FOI by the two-moment approach and Little's law for simulated heterogeneous exposure risk scenarios under non-seasonal transmission in a semi-open system. The times of the local transmission events are Gamma-distributed. The true mean FOI per host per year is calculated by dividing the total number of infections acquired by the population by the total number of individual hosts in the population. Minimum, 5% quantile, median, 95% quantile, and maximum values are shown in the boxplot.

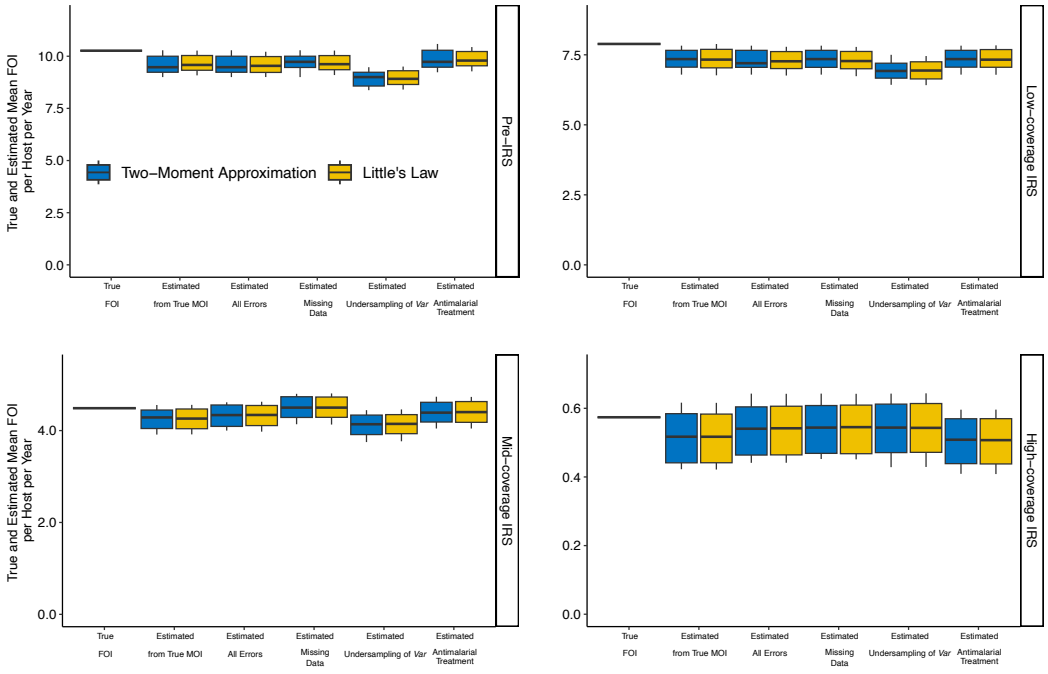
937



938

Figure 12. The true and estimated FOI by the two-moment approach and Little's law for simulated homogeneous exposure risk scenarios under seasonal transmission in a regionally-open system. The times of the local transmission events are Gamma-distributed. The true mean FOI per host per year is calculated by dividing the total number of infections acquired by the population by the total number of individual hosts in the population. Minimum, 5% quantile, median, 95% quantile, and maximum values are shown in the boxplot.

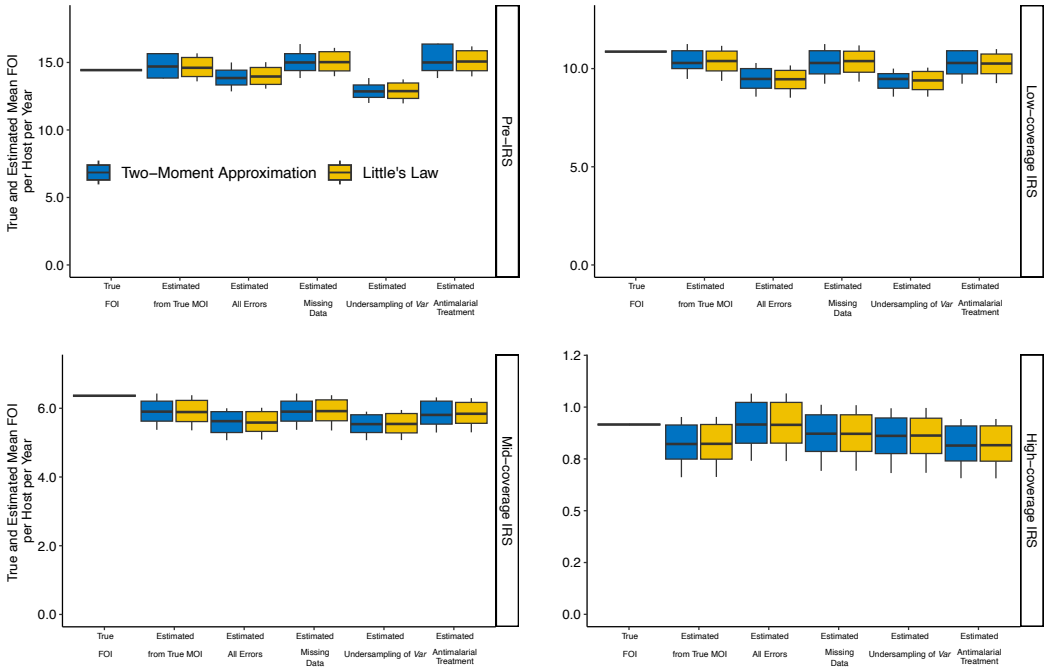
939



940

Figure 13. The true and estimated FOI by the two-moment approach and Little's law for simulated homogeneous exposure risk scenarios under non-seasonal transmission in a regionally-open system. The times of the local transmission events are Gamma-distributed. The true mean FOI per host per year is calculated by dividing the total number of infections acquired by the population by the total number of individual hosts in the population. Minimum, 5% quantile, median, 95% quantile, and maximum values are shown in the boxplot.

941



942

Figure 14. The true and estimated FOI by the two-moment approach and Little's law for simulated homogeneous exposure risk scenarios and seasonal transmission. The spatial setting corresponds to the closed system. The times of the local transmission events follow a exponential distribution. The true mean FOI per host per year is calculated by dividing the total number of infections acquired by the population by the total number of individual hosts in the population. Minimum, 5% quantile, median, 95% quantile, and maximum values are shown in the boxplot.

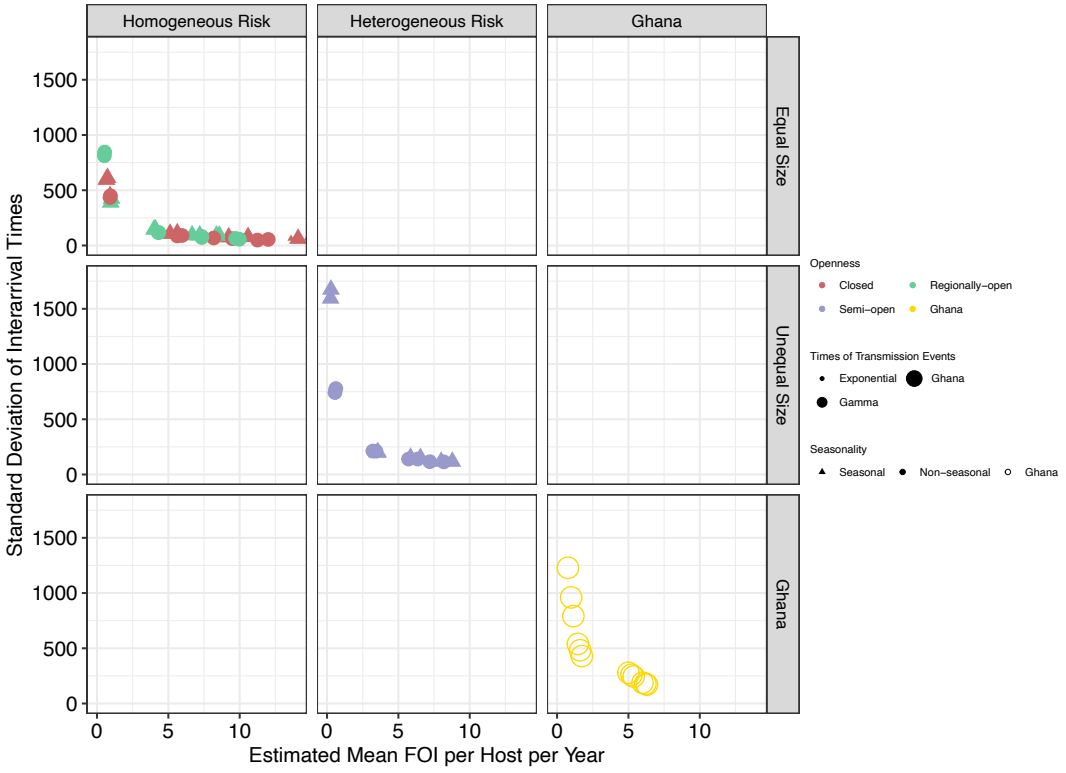


Figure 15. Magnitudes of the estimated variance by the two-moment approximation method for different simulation scenarios and the field data in Bongo District, Ghana.

Data and code availability

The sequences utilized in this study are publicly available in GenBank under BioProject Number: PRJNA 396962. All additional data associated with this study are available in the main text, the supplementary information, or upon reasonable request to the authors. For information on the simulation code and analysis scripts, please see https://github.com/qzhan321/FOI.

Acknowledgments

We wish to thank the participants, communities, and the Ghana Health Service in Bongo District, Ghana for their willingness to participate in the study of empirical data. We would like to thank the field teams in Bongo for their technical assistance in the field, as well as the laboratory personnel at the Navrongo Health Research Centre for their expertise and for undertaking the sample collections and parasitological assessments. This research was supported by Fogarty International Center at the National Institutes of Health through the joint NIH-NSF-NIFA Ecology and Evolution of Infectious Diseases award R01-TW009670 to K.P.D and M.P; and the National Institute of Allergy and Infectious Diseases, National Institutes of Health through the joint NIH-NSF-NIFA Ecology and Evolution of Infectious Diseases award R01-AI149779 to K.P.D and M.P. We appreciate the support of the Research Computing Center at the University of Chicago through the computational resources of the Midway cluster.

References

Abukari Z, Okonu R, Nyarko SB, Lo AC, Dieng CC, Salifu SP, Gyan BA, Lo E, Amoah LE. The Diversity, Multiplicity of Infection and Population Structure of P. falciparum Parasites Circulating in Asymptomatic Carriers Living in High and Low Malaria Transmission Settings of Ghana. Genes (Basel). 2019; doi: doi:10.3390/genes10060434.

Alonso PL, Sacarlal J, Aponte JJ, et al. Efficacy of the RTS,S/AS02A vaccine against *Plasmodium falciparum* infection and disease in young African children: randomised controlled trial. The Lancet. 2004; 364(9443):1411–1420. doi: https://doi.org/10.1016/S0140-6736(04)17223-1.

Anderson TJ, Haubold B, Williams JT, et al. Microsatellite markers reveal a spectrum of population structures in the malaria parasite *Plasmodium falciparum*. Mol Biol Evol. 2000; doi: 10.1093/oxfordjournals.molbev.a026247.

Anderson TJC, Haubold B, Williams JT, Estrada-Franco§ JG, Richardson L, Mollinedo R, Bockarie M, Mokili J, Mharakurwa S, French N, Whitworth J, Velez ID, Brockman AH, Nosten F, Ferreira MU, Day KP. Microsatellite Markers Reveal a Spectrum of Population Structures in the Malaria Parasite Plasmodium falciparum. Molecular Biology and Evolution. 2000 10; 17(10):1467–1482. https://doi.org/10.1093/oxfordjournals.molbev.a026247, doi: 10.1093/oxfordjournals.molbev.a026247.

Andolina C, Rek JC, Briggs J, Okoth J, Musiime A, Ramjith J, Teyssier N, Conrad M, Nankabirwa JI, Lanke K, Rodriguez-Barraquer I, Meerstein-Kessel L, Arinaitwe E, Olwoch P, Rosenthal PJ, Kamya MR, Dorsey G, Greenhouse B, Drakeley C, Staedke SG, et al. Sources of persistent malaria transmission in a setting with effective malaria control in eastern Uganda: a longitudinal, observational cohort study. The Lancet Infectious diseases. 2021; doi: https://doi.org/10.1016/S1473-3099(21)00072-4.

Argyropoulos DC, Tan MH, Adobor C, Mensah B, Labbé F, Tiedje KE, Koram KA, Ghansah A, Day KP. Performance of SNP barcodes to determine genetic diversity and population structure of *Plasmodium falciparum* in Africa. Front Genet. 2023; doi: 10.3389/fgene.2023.1071896.

Arnot D. Clone multiplicity of Plasmodium falciparum infections in individuals exposed to variable levels of disease transmission. Transactions of the Royal Society of Tropical Medicine and Hygiene. 1998; 92(6):580–585. https://www.sciencedirect.com/science/article/pii/S0035920398907738, doi: https://doi.org/10.1016/S0035-9203(98)90773-8.

Ashley EA, White NJ. The duration of Plasmodium falciparum infections. Malar J. 2014; doi: doi:10.1186/1475-2875-13-500.

Barry A, Awandu SS, Tiono AB, Grignard L, Bousema T, Collins KA. Improved Detectability of Plasmodium falciparum Clones with Repeated Sampling in Incident and Chronic Infections in Burkina Faso. Am J Trop Med Hyg. 2021; doi: 10.4269/ajtmh.21-0493.

Barry AE, Leliwa-Sytek A, Tavul L, Imrie H, Migot-Nabias F, Brown SM, McVean GAV, Day KP. Population Genomics of the Immune Evasion (var) Genes of Plasmodium falciparum. PLoS Pathog. 2007; doi: https://doi.org/10.1371/journal.ppat.0030034.

Baruch DI, Pasloske BL, Singh HB, et al. Cloning the P. falciparum gene encoding PfEMP1, a malarial variant antigen and adherence receptor on the surface of parasitized human erythrocytes. Cell. 1995; doi: 10.1016/0092-8674(95)90054-3.

Beier JC, Oster CN, Onyango FK, Bales JD, Sherwood JA, Perkins PV, Chumo DK, Koech DV, Whitmire RE, Roberts CR, Diggs CL, Hoffman SL. *Plasmodium falciparum* incidence relative to entomologic inoculation rates at a site proposed for testing malaria vaccines in western Kenya. The American journal of tropical medicine and hygiene. 1994; doi: https://doi.org/10.4269/ajtmh.1994.50.529.

Bekessy A, Molineaux L, Storey J. Estimation of incidence and recovery rates of Plasmodium falciparum parasitaemia from longitudinal data. Bull World Health Organ. 1976; .

Bopp SER, Manary MJ, Bright AT, Johnston GL, Dharia NV, Luna FL, et al. Mitotic Evolution of *Plasmodium falciparum* Shows a Stable Core Genome but Recombination in Antigen Families. PLoS Genet. 2013; doi: 10.1371/journal.pgen.1003293.

Bretscher MT, Maire N, Chitnis N, Felger I, Owusu-Agyei S, Smith T. The distribution of Plasmodium falciparum infection durations. Epidemics. 2011; doi: https://doi.org/10.1016/j.epidem.2011.03.002.

Bruce MC, Galinski MR, Barnwell JWea. Genetic diversity and dynamics of *plasmodium falciparum* and *P. vivax* populations in multiply infected children with asymptomatic malaria infections in Papua New Guinea. Parasitology. 2000; doi: 10.1017/s0031182099006356.

Buckee CO, Recker M. Evolution of the multi-domain structures of virulence genes in the human malaria parasite, *Plasmodium falciparum*. PLoS Comput Biol. 2012; doi: 10.1371/journal.pcbi.1002451.

Bull PC, Buckee CO, Kyes S, Kortok MM, Thathy V, Guyah B, Stoute JA, Newbold CI, Marsh K. *Plasmodium falciparum* antigenic variation. Mapping mosaic *var* gene sequences onto a network of shared, highly polymorphic sequence blocks. Molecular microbiology. 2008; doi: https://doi.org/10.1111/j.1365-2958.2008.06248.x.

Carneiro I, Roca-Feltrer A, Griffin JT, Smith L, Tanner M, Schellenberg JA, Greenwood B, Schellenberg D. Age-patterns of malaria vary with severity, transmission intensity and seasonality in sub-Saharan Africa: a systematic review and pooled analysis. PloS one. 2010; doi: https://doi.org/10.1371/journal.pone.0008988.

Chang HH, Childs LM, Buckee CO. Variation in infection length and superinfection enhance selection efficiency in the human malaria parasite. Sci Rep. 2016; doi: https://doi.org/10.1038/srep26370.

Chang HH, Worby CJ, Yeka A, Nankabirwa J, Kamya MR, Staedke SG, Dorsey G, Murphy M, Neafsey DE, Jeffreys AE, Hubbart C, Rockett KA, Amato R, Kwiatkowski DP, Buckee CO, Greenhouse B. THE REAL McCOIL: A method for the concurrent estimation of the complexity of infection and SNP allele frequency for malaria parasites. PLoS Computational Biology. 2017; doi: https://doi.org/10.1371/journal.pcbi.1005348.

Childs LM, Buckee CO. Dissecting the determinants of malaria chronicity: why within-host models struggle to reproduce infection dynamics. Journal of the Royal Society, Interface. 2014; doi: https://doi.org/10.1098/rsif.2014.1379.

Choi DW, Kim NK, Chae KC. A Two-Moment Approximation for the GI/G/c Queue with Finite Capacity. INFORMS Journal on Computing. 2005; 17(1):75–81. https://doi.org/10.1287/ijoc.1030.0058, doi: 10.1287/ijoc.1030.0058.

Chu CS, White NJ. Management of relapsing Plasmodium vivax malaria. Expert Rev Anti Infect Ther. 2016; doi: 10.1080/14787210.2016.1220304.

Claessens A, Adams Y, Ghumra A, et al. A subset of group A-like *var* genes encodes the malaria parasite ligands for binding to human brain endothelial cells. Proc Natl Acad Sci USA. 2012; doi: 10.1073/pnas.1120461109.

Claessens A, Hamilton WL, Kekre M, D OT, Faizullabhoy A, Rayner JC, Kwiatkowski D. Generation of Antigenic Diversity in *Plasmodium falciparum* by Structured Rearrangement of *Var* Genes During Mitosis. PLoS Genet. 2014; doi: https://doi.org/10.1371/journal.pgen.1004812.

Collins WE, Jeffery GM. A retrospective examination of sporozoite- and trophozoite-induced infections with *Plasmodium falciparum* in patients previously infected with heterologous species of Plasmodium: effect on development of parasitologic and clinical immunity. Am J Trop Med Hyg. 1999; doi: 10.4269/tropmed.1999.61-036.

Daniels R, Volkman SK, Milner DA, et al. A general SNP-based molecular barcode for Plasmodium falciparum identification and tracking. Malar J. 2008; doi: https://doi.org/10.1186/1475-2875-7-223.

Daubersies P, Sallenave-Sales S, Magne S, Trape JF, Contamin H, Fandeur T, Rogier C, Mercereau-Puijalon O, Druilhe P. Rapid turnover of Plasmodium falciparum populations in asymptomatic individuals living in a high transmission area. The American journal of tropical medicine and hygiene. 1996; doi: https://doi.org/10.4269/ajtmh.1996.54.18.

Day KP, Artzy-Randrup Y, Tiedje KE, Rougeron V, Chen DS, Rask TS, Rorick MM, Migot-Nabias F, Deloron P, Luty AJF, Pascual M. Evidence of strain structure in *Plasmodium falciparum var* gene repertoires in children from Gabon, West Africa. Proceedings of the National Academy of Sciences. 2017; doi: https://doi.org/10.1073/pnas.1613018114.

Deitsch KW, Lukehart SA, Stringer JR. Common strategies for antigenic variation by bacterial, fungal and protozoan pathogens. Nature reviews. 2009; doi: https://doi.org/10.1038/nrmicro2145.

Dietz K, Molineaux L, Thomas A. A malaria model tested in the African savannah. Bull World Health Organ. 1974; .

Donovan MJ, Messmore AS, Scrafford DA, Sacks DL, Kamhawi S, McDowell MA. Uninfected mosquito bites confer protection against infection with malaria parasites. Infection and immunity. 2007; doi: https://doi.org/10.1128/IAI.01928-06.

Doolan DL, Dobaño C, Baird JK. Acquired immunity to malaria. Clin Microbiol Rev. 2009; doi: doi:10.1128/CMR.00025-08.

Doolan DL, Martinez-Alier N. Immune response to pre-erythrocytic stages of malaria parasites. Curr Mol Med. 2006; doi: 10.2174/156652406776055249.

Earland D, Buchwald AG, Sixpence A, Chimenya M, Damson M, Seydel KB, Mathanga DP, Taylor TE, Laufer MK. Impact of Multiplicity of Plasmodium falciparum Infection on Clinical Disease in Malawi. The American Journal of Tropical Medicine and Hygiene. 2019; doi: https://doi.org/10.4269/ajtmh.19-0093.

Efron B, Tibshirani RJ. An Introduction to the Bootstrap (1st ed.); 1994. doi: https://doi.org/10.1201/9780429246593.

Farnert A, Snounou G, Rooth I, Bjorkman A. Daily dynamics of *Plasmodium falciparum* subpopulations in asymptomatic children in a holoendemic area. The American journal of tropical medicine and hygiene. 1997; doi: https://doi.org/10.4269/ajtmh.1997.56.538.

Felger I, Tavul L, Kabintik S, Marshall V, Genton B, Alpers M, Beck HP. Plasmodium falciparum: Extensive Polymorphism in Merozoite Surface Antigen 2 Alleles in an Area with Endemic Malaria in Papua New Guinea. Experimental Parasitology. 1994; 79(2):106–116. https://www.sciencedirect.com/science/article/pii/S0014489484710708, doi: https://doi.org/10.1006/expr.1994.1070.

Felger I, Maire M, Bretscher MT, Falk N, Tiaden A, Sama W, Beck HP, Owusu-Agyei S, Smith TA. The Dynamics of Natural *Plasmodium falciparum* Infections. PLoS ONE. 2012; doi: https://doi.org/10.1371/journal.pone.0045542.

Frank M, Kirkman L, Costantini D, Sanyal S, Lavazec C, Templeton TJ, Deitsch KW. Frequent recombination events generate diversity within the multi-copy variant antigen gene families of *Plasmodium falciparum*. International journal for parasitology. 2008; doi: https://doi.org/10.1016/j.ijpara.2008.01.010.

Freitas-Junior LH, Bottius E, Pirrit LA, et al. Frequent ectopic recombination of virulence factor genes in telomeric chromosome clusters of *P. falciparum*. Nature. 2000; doi: 10.1038/35039531.

Färnert A, Lebbad M, Faraja L, Rooth I. Extensive dynamics of Plasmodium falciparum densities, stages and genotyping profiles. Malar J. 2008; doi: 10.1186/1475-2875-7-241.

Gardner MJ, Hall N, Fung E, White O, Berriman M, Hyman RW, Carlton JM, Pain A, Nelson KE, Bowman S, Paulsen IT, James K, Eisen JA, Rutherford K, Salzberg SL, Craig A, Kyes S, Chan MS, Nene V, Shallom SJ, et al. Genome sequence of the human malaria parasite *Plasmodium falciparum*. Nature. 2002; doi: https://doi.org/10.1038/nature01097.

Gibson MA, Bruck J. Efficient Exact Stochastic Simulation of Chemical Systems with Many Species and Many Channels. The Journal of Physical Chemistry A. 2000; doi: 10.1021/jp993732q.

Gillespie DT. A General Method for Numerically Simulating the Stochastic Time Evolution of Coupled Chemical Reactions. Journal of Computational Physics. 1976; doi: http://dx.doi.org/10.1016/0021-9991(76)90041-3.

Gogue C, Wagman J, Tynuv K, et al. An observational analysis of the impact of indoor residual spraying in Northern, Upper East, and Upper West Regions of Ghana: 2014 through 2017. Malar J. 2020; doi: https://doi.org/10.1186/s12936-020-03318-1.

Guttery DS, Holder AA, Tewari R. Sexual development in Plasmodium: lessons from functional analyses. PLoS pathogens. 2012; doi: https://doi.org/10.1371/journal.ppat.1002404.

He Q, Pilosof S, Tiedje KE, Ruybal-Pesántez S, Artzy-Randrup Y, Baskerville EB, Day KP, Pascual M. Networks of genetic similarity reveal non-neutral processes shape strain structure in *Plasmodium falciparum*. Nature Communications. 2018; doi: https://doi.org/10.1038/s41467-018-04219-3.

Hergott DEB, Owalla TJ, Staubus WJ, et al. Assessing the daily natural history of asymptomatic Plasmodium infections in adults and older children in Katakwi, Uganda: a longitudinal cohort study. The Lancet Microbe. 2024; doi: 10.1016/S2666-5247(23)00262-8.

Hofmann NE, Karl S, Wampfler R, Kiniboro B, Teliki A, Iga J, Waltmann A, Betuela I, Felger I, Robinson LJ, Mueller I. The complex relationship of exposure to new Plasmodium infections and incidence of clinical malaria in Papua New Guinea. eLife. 2017; doi: https://doi.org/10.7554/eLife.23708.

John CC, Moormann AM, Pregibon DC, et al. Correlation of high levels of antibodies to multiple pre-erythrocytic *Plasmodium falciparum* antigens and protection from infection. Am J Trop Med Hyg. 2005; doi: https://doi.org/10.4269/ajtmh.2005.73.222.

Johnston SP, Pieniazek NJ, Xayavong MV, Slemenda SB, Wilkins PP, Silva AJd. PCR as a confirmatory technique for laboratory diagnosis of malaria. J Clin Microbiol. 2006; doi: 10.1128/JCM.44.3.1087-1089.2006.

Kaestli M, Cockburn IA, Cortés A, Baea K, Rowe JA, Beck HP. Virulence of malaria is associated with differential expression of *Plasmodium falciparum var* gene subgroups in a case-control study. J Infect Dis. 2006; doi: 10.1086/503776.

Konaté L, Zwetyenga J, Rogier C, Bischoff E, Fontenille D, Tall A, Spiegel A, Trape JF, Mercereau-Puijalon O. 5. Variation of Plasmodium falciparum msp1 block 2 and msp2 allele prevalence and of infection complexity in two neighbouring Senegalese villages with different transmission conditions. Transactions of The Royal Society of Tropical Medicine and Hygiene. 1999 02; 93(Supplement$_1$) : 21 – –28., 10.1016/S0035-9203(99)90323-1.

Kraemer SM, Kyes SA, Aggarwal G, Springer AL, Nelson SO, Christodoulou Z, Smith LM, Wang W, Levin E, Newbold CI, Myler PJ, Smith JD. Patterns of gene recombination shape *var* gene repertoires in *Plasmodium falciparum*: comparisons of geographically diverse isolates. BMC Genomics. 2007; https://doi.org/10.1186/1471-2164-8-45.

Labbé F, He Q, Zhan Q, Tiedje KE, Argyropoulos DC, Tan MH, Ghansah A, Day KP, Pascual M. Neutral vs. non-neutral genetic footprints of Plasmodium falciparum multiclonal infections. PLoS Computational Biology. 2023; https://doi.org/10.1371/journal.pcbi.1010816.

Laishram DD, Sutton PL, Nanda N, Sharma VL, Sobti RC, Carlton JM, Joshi H. The complexities of malaria disease manifestations with a focus on asymptomatic malaria. Malaria journal. 2012; https://doi.org/10.1186/1475-2875-11-29.

Langhorne J, Ndungu FM, Sponaas AM, Marsh K. Immunity to malaria: more questions than answers. Nat Immunol. 2008; https://doi.org/10.1038/ni.f.205.

Larremore DB, Clauset A, Buckee CO. A network approach to analyzing highly recombinant malaria parasite genes. PLoS Comput Biol. 2013; 10.1371/journal.pcbi.1003268.

Lavstsen T, Salanti A, Jensen ATR, Arnot DE, Theander TG. Sub-grouping of *Plasmodium falciparum* 3D7 *var* genes based on sequence analysis of coding and non-coding regions. Malaria journal. 2003; https://doi.org/10.1186/1475-2875-2-27.

Lee SA, Yeka A, Nsobya SL, Dokomajilar C, Rosenthal PJ, Talisuna A, Dorsey G. Complexity of Plasmodium falciparum infections and antimalarial drug efficacy at 7 sites in Uganda. The Journal of infectious diseases. 2006; https://doi.org/10.1086/501473.

Lindblade KA, Steinhardt L, Samuels A, Kachur SP, Slutsker L. The silent threat: asymptomatic parasitemia and malaria transmission. Expert Rev Anti Infect Ther. 2013; 10.1586/eri.13.45.

Little JDC, Graves SC. Little's Law. In: Chhajed D, Lowe TJ, editors. *Building Intuition* International Series in Operations Research & Management Science, Springer; 2008.p. 81–100. https://ideas.repec.org/h/spr/isochp/978-0-387-73699-0_5.html, 10.1007/978-0-387-73699-0.

Lloyd-Smith JO, Schreiber SJ, Kopp PE, Getz WM. Superspreading and the effect of individual variation on disease emergence. Nature. 2005; 10.1038/nature04153.

Macdonald G. The analysis of malaria parasite rates in infants. Tropical diseases bulletin. 1950; .

Maire N, Smith T, Ross A, Owusu-Agyei S, Dietz K, Molineaux L. A model for natural immunity to asexual blood stages of *Plasmodium falciparum* malaria in endemic areas. Am J Trop Med Hyg. 2006; 10.4269/ajtmh.2006.75.19.

Molineaux L, Träuble M, Collins WE, Jeffery GM, Dietz K. Malaria therapy reinoculation data suggest individual variation of an innate immune response and independent acquisition of antiparasitic and antitoxic immunities. Transactions of the Royal Society of Tropical Medicine and Hygiene. 2002; https://doi.org/10.1016/s0035-9203(02)90308-1.

Msuya FH, Curtis CF. Trial of pyrethroid impregnated bednets in an area of Tanzania holoendemic for malaria. Part 4. Effects on incidence of malaria infection. Acta Trop. 1991; 10.1016/0001-706x(91)90035-i.

Mueller I, Schoepflin S, Smith TA, Benton KL, Bretscher MT, Lin E, Kiniboro B, Zimmerman PA, Speed TP, Siba P, Felger I. Force of infection is key to understanding the epidemiology of *Plasmodium falciparum* malaria in Papua New Guinean children. Proceedings of the National Academy of Sciences of the United States of America. 2012; https://doi.org/10.1073/pnas.1200841109.

Muench H. Catalytic models in epidemiology. Harvard University Press. 1959; .

Mugenyi L, Abrams S, Hens N. Estimating age-time-dependent malaria force of infection accounting for unobserved heterogeneity. Epidemiology amp; Infection. 2017; 145(12):2545–2562. 10.1017/S09502

Nkhoma SC, Trevino SG, Gorena KM, Nair S, Khoswe S, Jett C, Garcia R, Daniel B, Dia A, Terlouw DJ, Ward SA, Anderson TJC, Cheeseman IH. Co-transmission of Related Malaria Parasite Lineages Shapes Within-Host Parasite Diversity. Cell host microbe. 2020; https://doi.org/10.1016/j.chom.2019.12.001

Okell LC, Bousema T, Griffin JT, Ouédraogo AL, Ghani AC, Drakeley CJ. Factors determining the occurrence of submicroscopic malaria infections and their relevance for control. Nature communications. 2012; https://doi.org/10.1038/ncomms2241.

Peyerl-Hoffmann G, Jelinek T, Kilian A, Kabagambe G, Metzger WG, von Sonnenburg F. Genetic diversity of *Plasmodium falciparum* and its relationship to parasite density in an area with different malaria endemicities in West Uganda. Trop Med Int Health. 2001; 10.1046/j.1365-3156.2001.00761.x.

Pilosof S, He Q, Tiedje KE, Ruybal-Pesántez S, Day KP, Pascual M. Competition for hosts modulates vast antigenic diversity to generate persistent strain structure in *Plasmodium falciparum*. PLoS Biology. 2019; https://doi.org/10.1371/journal.pbio.3000336.

Piper KP, Roberts DJ, Day KP. Plasmodium falciparum: analysis of the antibody specificity to the surface of the trophozoite-infected erythrocyte. Experimental parasitology. 1999; https://doi.org/10.1006/expr

Portugal S, Drakesmith H, Mota MM. Superinfection in malaria: Plasmodium shows its iron will. EMBO reports. 2011; 10.1038/embor.2011.213.

Potthoff RF, Whittinghill M. Testing for homogeneity. II. The Poisson distribution. Biometrika. 1966; https://doi.org/10.1093/biomet/53.1-2.183.

Pull JH, Grab B. A simple epidemiological model for evaluating the malaria inoculation rate and the risk of infection in infants. Bulletin of the World Health Organization. 1974; .

Rask TS, Hansen DA, Theander TG, Pedersen AG, Lavstsen T. *Plasmodium falciparum* erythrocyte membrane protein 1 diversity in seven genomes–divide and conquer. PLoS computational biology. 2010; https://doi.org/10.1371/journal.pcbi.1000933.

de Roos AM, He Q, M P. An immune memory-structured SIS epidemiological model for hyperdiverse pathogens. Proc Natl Acad Sci USA. 2023; 10.1073/pnas.2218499120.

Rottmann M, Lavstsen T, Mugasa JP, et al. Differential expression of *var* gene groups is associated with morbidity caused by *Plasmodium falciparum* infection in Tanzanian children. Infect Immun. 2006; 10.1128/IAI.02073-05.

Ruybal-Pesántez S, Tiedje KE, Pilosof S, Tonkin-Hill G, He Q, Rask TS, Amenga-Etego L, Oduro AR, Koram KA, Pascual M, Day KP. Age-specific patterns of DBLα var diversity can explain why residents of high malaria transmission areas remain susceptible to *Plasmodium falciparum* blood stage infection throughout life. International Journal for Parasitology. 2022; https://doi.org/10.1016/j.ijpara.2021.12.00

Ruybal-Pesántez S, Tiedje KE, Tonkin-Hill G, Rask TS, Kamya MR, Greenhouse B, Dorsey G, Duffy MF, Day KP. Population genomics of virulence genes of Plasmodium falciparum in clinical isolates from Uganda. Sci Rep. 2017; https://doi.org/10.1038/s41598-017-11814-9.

Shaukat AM, Breman JG, McKenzie FE. Using the entomological inoculation rate to assess the impact of vector control on malaria parasite transmission and elimination. Malaria journal. 2010; https://doi.org/10.1186/1475-2875-9-122.

Simpson JA, Aarons L, Collins WE, Jeffery GM, White NJ. Population dynamics of untreated Plasmodium falciparum malaria within the adult human host during the expansion phase of the infection. Parasitology. 2002; https://doi.org/10.1017/s0031182001001202.

Smith DL, Drakeley CJ, Chiyaka C, Hay SI. A quantitative analysis of transmission efficiency versus intensity for malaria. Nature Communication. 2010; https://doi.org/10.1038/ncomms1107.

Smith JD, Chitnis CE, Craig AG, et al. Switches in expression of *Plasmodium falciparum var* genes correlate with changes in antigenic and cytoadherent phenotypes of infected erythrocytes. Cell. 1995; 10.1016/0092-8674(95)90056-x.

Smith T, Vounatsou P. Estimation of infection and recovery rates for highly polymorphic parasites when detectability is imperfect, using hidden Markov models. Statistics in medicine. 2003; https://doi.org/10.1002/sim.1274.

Sondo P, Derra K, Rouamba T, Diallo SN, Taconet P, Kazienga A, Ilboudo H, Tahita MC, Valéa I, Sorgho H, Lefèvre T, Tinto H. Determinants of Plasmodium falciparum multiplicity of infection and genetic diversity in Burkina Faso. Parasites Vectors. 2020; https://doi.org/10.1186/s13071-020-04302-z.

Su XZ, Heatwole VM, Wertheimer SP, et al. The large diverse gene family *var* encodes proteins involved in cytoadherence and antigenic variation of *Plasmodium falciparum*-infected erythrocytes. Cell. 1995; 10.1016/0092-8674(95)90055-1.

Tan MH, Shim H, Chan Y, Day KP. Unravelling *var* complexity: Relationship between DBLα types and *var* genes in *Plasmodium falciparum*. Front Parasitol. 2023; 10.3389/fpara.2022.1006341.

Tiedje KE, Oduro AR, Agongo G, Anyorigiya T, Azongo D, Awine T, Ghansah A, Pascual M, Koram KA, Day KP. Seasonal Variation in the Epidemiology of Asymptomatic Plasmodium falciparum Infections across Two Catchment Areas in Bongo District, Ghana. Am J Trop Med Hyg. 2017; 10.4269/ajtmh.16-0959.

Tiedje KE, Oduro AR, Bangre O, Amenga-Etego L, Dadzie SK, Appawu MA, Frempong K, Asoala V, Ruybal-Pésantez S, Narh CA, Deed SL, Argyropoulos DC, Ghansah A, Agyei SA, Segbaya S, Desewu K, Williams I, Simpson JA, Malm K, Pascual M, et al. Indoor residual spraying with a non-pyrethroid insecticide reduces the reservoir of Plasmodium falciparum in a high-transmission area in northern Ghana. PLOS Global Public Health. 2022; https://doi.org/10.1371/journal.pgph.0000285.

Tiedje KE, Zhan Q, Ruybal-Pésantez S, Tonkin-Hill G, He Q, Tan MH, Argyropoulos DC, Deed SL, Ghansah A, Bangre O, Oduro AR, Koram KA, Pascual M, Day KP. Measuring changes in *Plasmodium falciparum var* census population size and structure in response to sequential malaria control interventions. *eLife*. 2023; https://doi.org/10.7554/eLife.91411.1.

Touray AO, Mobegi VA, Wamunyokoli F, Herren JK. Diversity and Multiplicity of P. falciparum infections among asymptomatic school children in Mbita, Western Kenya. Sci Rep. 2020; doi:10.1038/s41598-020-62819-w.

Tusting LS, Bousema T, Smith DL, Drakeley C. Measuring changes in *Plasmodium falciparum* transmission: precision, accuracy and costs of metrics. Advances in parasitology. 2014; https://doi.org/10.1016/B
0-12-800099-1.00003-X.

Volkman S, Neafsey D, Schaffner S, et al. Harnessing genomics and genome biology to understand malaria biology. Nat Rev Genet. 2012; 10.1038/nrg3187.

White MT, Griffin JT, Churcher TS, Ferguson NM, Basáñez MG, Ghani AC. Modelling the impact of vector control interventions on Anopheles gambiae population dynamics. Parasites Vectors. 2011; https://doi.org/10.1186/1756-3305-4-153.

Wong W, Griggs AD, Daniels RF, Schaffner SF, Ndiaye D, Bei AK, Deme AB, MacInnis B, Volkman SK, Hartl DL, Neafsey DE, Wirth DF. Genetic relatedness analysis reveals the cotransmission of genetically related Plasmodium falciparum parasites in Thiès, Senegal. Genome Med. 2017; https://doi.org/10.118
017-0398-0.

Wong W, Volkman S, Daniels R, Schaffner S, Sy M, Ndiaye YD, Badiane AS, Deme AB, Diallo MA, Gomis J, Sy N, Ndiaye D, Wirth DF, Hartl DL. RH: a genetic metric for measuring intrahost *Plasmodium falciparum* relatedness and distinguishing cotransmission from superinfection. PNAS Nexus. 2022; https://doi.org/10.1093/pnasnexus/pgac187.

World Health Organization. Indoor Residual Spraying: An Operational Manual for Indoor Residual Spraying (IRS) for Malaria Transmission Control and Elimination. World Health Organization; 2015.

World Health Organization. World malaria report 2023. World Health Organization; 2023.

Zhan Q, He Q, Tiedje K, Day KP, Pascual M. Hyper-diverse antigenic variation and resilience to transmission-reducing intervention in falciparum malaria. medRxiv. 2024; https://doi.org/10.1101/2024.02.

Zhang X, Deitsch KW. The mystery of persistent, asymptomatic *Plasmodium falciparum* infections. Curr Opin Microbiol. 2022; 10.1016/j.mib.2022.102231.

Zhong D, Koepfli C, Cui L, Yan G. Molecular approaches to determine the multiplicity of *Plasmodium* infections. Malaria Journal. 2018; https://doi.org/10.1186/s12936-018-2322-5.