Grosso *et al. Malaria Journal* (2025) 24:173
https://doi.org/10.1186/s12936-025-05415-5

Malaria Journal

PERSPECTIVE

Open Access

An integrative review of the combined use of mathematical and statistical models for estimating malaria transmission parameters

Alessandro Grosso[1*], Niel Hens[2,3] and Steven Abrams[1,2]

Abstract

Background Characterizing malaria burden and its evolution is complicated by the high levels of spatio-temporal heterogeneity and by the complexity of the transmission process.

Main body This manuscript presents an integrative review of the combined use of mathematical and statistical models to estimate malaria transmission parameters. Therefore, this work aims to provide a solid methodological foundation for the estimation of transmission intensity and other relevant quantities. A perspective covering both mathematical and statistical models to appraise commonly used metrics is adopted and subsequently their inclusion as parameters in compartmental models as well as their estimation from available data is discussed. The current review argues in favour of a more widespread consideration of the Force of Infection (FOI) as a malaria transmission metric. Using the FOI dispenses the analyst from explicitly describing vector dynamics in compartmental modelling, simplifying the system of differential equations describing transmission dynamics. In turn, its estimation can be flexibly performed by solely relying on host data, such as parasitaemia or serology, avoiding the need for entomological data.

Conclusion The present work argues that the interaction between mathematical and statistical models, although previously exemplified by others, is underappreciated when modelling malaria transmission. Orienting the exposition around the FOI provides an illustration of the potential borne by the existing methodology. A connection between the two modelling frameworks warrants better scrutiny, as it leads to the possibility of exploiting the full range of modern statistical methods.

Keywords Malaria transmission, Mathematical models, Statistical models, Force of infection

Background

Assessing the current level and (spatio-)temporal changes in malaria transmission is of great importance for evaluating the feasibility (*ex ante*), impact (*ex post*) and cost-effectiveness of any strategy aiming at malaria control or elimination [1].

Transmission intensity is defined by the World Health Organization (WHO) as the "*frequency with which people living in an area are bitten by anopheline mosquitoes carrying human malaria sporozoites.*" [2]. It is characterized by spatio-temporal heterogeneity, with its level being related to a large range of factors encompassing

*Correspondence:
Alessandro Grosso
alessandro.grosso@uantwerpen.be
[1] Global Health Institute, Department of Family Medicine and Population Health, University of Antwerp, Antwerp, Belgium
[2] Data Science Institute, Interuniversity Institute for Biostatistics and statistical Bioinformatics, Hasselt University, Hasselt, Belgium
[3] Centre for Health Economics Research and Modelling Infectious Diseases, Vaccine and Infectious Disease Institute, University of Antwerp, Antwerp, Belgium

entomological (e.g., mosquito population density), environmental (e.g., seasonal variation in temperature and rainfall), biological (e.g., immunity), and other elements (e.g., heterogeneous biting) [3]. Moreover, the transmission process itself is highly complex, developing over several phases and biological cycles involving vector and host, thus offering a wide range of possibilities for intervention. This implies that a transmission metric can focus on dynamics of the vector, the host, or the parasite itself, so that an observable change in the value of a malaria transmission metric can be caused by a change in one (or more) of these factors.

The presence of all these intricacies makes it paramount to explicitly define criteria and methods required for an optimal (that is, "resource conscious") and robust estimation of malaria transmission intensity or other related epidemiological quantities, such as current parasite levels, which are closely related [4, 5]. The present work aims at meeting these needs by providing a methodological overview on the estimation of malaria transmission from a perspective covering *both* mathematical and statistical modelling techniques. Therefore, the current review is termed 'integrative', stressing the importance of combining both approaches and providing specific details concerning their use. The integration of statistical models into the mathematical modeling framework dates back to the seminal work by Sir Ronald Ross [6], who proposed to iterate between what he defined as a posteriori and a priori methods, nowadays often referred to as statistical and mathematical models, respectively, to estimate relevant parameters governing disease dynamics and their transmission process. Since Ross's time, the "toolbox" available to researchers interested in understanding the dynamics of malaria, from a mathematical and statistical point of view, has expanded considerably [7–9]. The present work argues that combining both approaches provides the best way forward to enjoy all these developments simultaneously, opening up a realm of possibilities to the analyst that the two frameworks, taken individually, would not provide [10]. In effect, the biological "facts" that one wants to translate into a set of mathematical equations may not be well understood beforehand, however, assumptions thereon could be contrasted with data through statistical models [11]. Next to that, estimates of key epidemiological parameters could be derived from models for which the underlying assumptions are tested to be in line with observed data. In particular, this work focuses on highlighting how the interaction between mathematical and statistical models can be achieved by considering the Force of Infection (FOI), commonly used in infectious disease epidemiology [12, 13], as a malaria transmission metric. Despite the fact that FOI has been considered as a measure of malaria transmission, its use is less common than the ubiquitous use for close contact and sexually transmitted diseases.

Previous work, such as the review by Tusting and coauthors [14], provides a comprehensive overview and evaluation of malaria metrics, and as such was used as a relevant starting point for our evaluation. However, it does not deal explicitly with either issues related to the estimation of such metrics, or with their role as parameters in mathematical models of disease transmission. On the other hand, the reviews by Mandal et al. [15] and Reiner et al. [9], focus more specifically on mathematical models of malaria transmission, although they do not deal with the possibility of estimating relevant epidemiological quantities using statistical models, while the paper by Smith et al. [16] presents a historical perspective on the so-called "Ross-Macdonald" model [17, 18] and variations thereof.

Building on a review covering the estimation of the FOI at large [19], and in the spirit of recent contributions to the scientific literature [20] summarizing methods for the analysis of serological data, the present work is defined as a *methodological integrative review* thereby synthesizing mathematical and statistical modelling approaches to the estimation of the FOI. By presenting existing methodologies and referring to applications in the context of malaria, the aim is to provide a cohesive and structured understanding of the topic. Despite the fact that several malaria review papers describe various themes, the literature on the subject of malaria transmission does not feature a literature review, to the best of our knowledge, addressing the estimation of the FOI based on mathematical and statistical models. Moreover, it was not possible to identify other publications simultaneously focusing on malaria metrics that are integrated as direct or indirect parameters of mathematical (compartmental) models and deriving estimates thereof from relevant data by exploiting the connection between mathematical and statistical models to do so.

The overview starts by introducing the FOI in Sect. 2, placing it in the wider context of metrics of common use Sect. 3 highlights how the FOI can be successfully integrated in mathematical models of malaria disease dynamics as well as appropriately estimated via statistical models dealing with the complexities highlighted above. Lastly, Sect. 10 is concerned with presenting earlier approaches that conducted an integrated analysis within the framework of malaria. These will act as 'building blocks" towards a general framework integrating mathematical and statistical modelling approaches, with which the review concludes. This framework should help connect the different highlighted perspectives, providing insights into their strengths, limitations, and synergies.

Malaria metrics and the FOI

Considering the structure of the malaria transmission cycle, it is possible to identify three possible measurement levels: (1) the human level, encompassing any clinical sign or symptom (e.g., the *spleen rate*, now obsolete, or presence of fever) or humoral immune response, gauged by serological data [21]; (2) the parasite level, identified for instance by microscopy on blood samples; and (3) the vector level, including entomological measurements. From now on, the term "parasitological" is used to refer to the measurement of the parasite load in a blood sample and "serological" to refer to the presence of antibodies to the malaria parasite in a blood sample. Common metrics employing data from each of these measurement levels are now briefly introduced.

First, detection of parasites in a blood sample can be used to compute the Parasite Rate (PR), a common metric for prevalence and endemicity, measuring the proportion of individuals harbouring malaria parasites in their blood at a given time. Serological data can be used to compute the Seroconversion Rate (SCR), expressing the rate of conversion of individuals in a cohort from the seronegative to the seropositive stage. In a sense, the SCR captures the switch of individuals from a non-infected to an infected state over a certain unit of time. However, the SCR is obtained from the analysis of antibody data, thus expressing *past* infection, rather than incidence [21]. The SCR can be used to reconstruct *history* of exposure, the more so if measurements taken on a large cohort of individuals across most (if not all) age groups are available and considering antibodies capturing short- and long-term response to infection. Last, the Entomological Inoculation Rate (EIR) counts the number of infectious bites per person per unit time, by combining measures of the "human biting rate" (i.e., the number of bites per person per year), and of the "sporozoite rate" (that is, the proportion of mosquitoes carrying sporozoites in the salivary glands, thus being potentially infectious). In turn, this requires complex experimental or field measurement settings, and its accuracy will critically depend on the accuracy of its constituent components, whose measurement still suffers from a lack of standardization [1, 22]. EIR provides a good summary of exposure to infectious bites, but it is difficult to translate to incidence or clinical disease, also considering the known discrepancy between transmission intensity and efficiency [3, 23].

The FOI measures the per capita number of infections per unit time or, equivalently, represents the instantaneous probability of acquiring infection, that is, the rate (or hazard) of infection [24, 25]. Importantly, the FOI counts all new infections, whether symptomatic or asymptomatic. This concept has been recently expanded to take into account the availability of genetic data so that the new concept of *molecular* Force of Infection, or mFOI can be conceptualized as a metric counting all new parasite clones acquired per unit of time. The main principle behind both metrics remains that of capturing the level of transmission by *counting incident infections*. As a result, one has to first define "infection" in an individual, so that the computation of the FOI can be performed. For example, assume that malaria parasite positivity is considered as the variable expressing (new) infection among parasite negative individuals at baseline. The FOI then expresses the rate of occurrence of infection (estimated based on sampled individuals), i.e., the so-called hazard of infection similar to the hazard function defined in survival analysis (e.g., hazard of death or relapse in oncological trials). Although the FOI has been considered constant, generally the force of infection depends on both measurable and unmeasurable factors, including host-specific features (e.g., age), calendar time (including seasonality), and environmental conditions (e.g., living conditions), to name a few.

As a consequence of this, the FOI is only going to be as accurate and precise as the measure which is chosen to signal the presence of a new infection in an individual, for instance, the presence of the parasite in the bloodstream, and on the sensitivity and specificity of the diagnostic instruments used to detect it. Following Tusting and coauthors [14] in defining accuracy as the ability of a metric to provide estimates close to the true value of the quantity of interest, a notion with a direct link to the idea of "unbiasedness" in statistical literature. Precision instead covers issues surrounding sampling variability, as it expresses the ability to provide estimates with a level of certainty and to obtain similar estimates, under the same conditions, in a series of independent replications of the study.

For example, if parasite presence is taken as infection signal, it should be taken into account that parasitaemia is highly sensitive to any factor that affects the parasite density load and the blood sample analysis technique (and the intrinsic uncertainty surrounding its use). To overcome these difficulties, the WHO recommends collecting a large number of samples, with high frequency, but this could prove unfeasible especially in remote areas with high endemicity [26]. The rationale behind this advice is that, as parasitaemia can fluctuate, one would ideally want to have as many measurements as possible, taken as frequently as possible, to keep track of the evolution of infection. Taking measurements only on a monthly or quarterly basis, for instance, might lead to underestimating the number of infections because one (or more) might develop and end between two subsequent blood tests.

What is argued in the following sections is that such difficulties can be overcome via a fruitful interaction between mathematical and statistical models. First, representing transmission intensity using the FOI leads to a simplification in mathematical compartmental models, dispensing the analyst from the need to conduct complex and time-consuming entomological measurements. Then, it is discussed how it is in fact possible to take into account the influence of factors affecting any measure used to construct FOI (e.g., parasitaemia) via statistical modelling, which can also provide ways to correct for bias induced by measurement error.

Estimating malaria transmission intensity using the FOI

Having provided a taxonomy and review of some commonly used metrics, the overview continues by discussing how transmission intensity, as captured by the FOI, can be expressed in mathematical models and estimated from available data.

Mathematical modelling of malaria transmission

Mathematical models describe the current state and evolution of a biological system over time via a set of equations and/or through (micro)simulation, thereby expressing a phenomenon such as disease transmission in mathematical terms. Mathematical modelling provides a convenient and flexible way to link individual-level processes (e.g., infection), with population dynamics (e.g., incidence and prevalence of disease) [27, 28]. They allow researchers to condense the evidence available regarding the dynamics of, say, an epidemic, and can be used to reflect on which factors constitute the most important drivers of its persistence. In turn, this can lead to the formulation of context-specific public health policies.

Various types of mathematical models have been proposed in infectious disease literature ranging from deterministic and stochastic compartmental models [see, e.g., [24, 29]], over meta-population models [30] to individual- or agent-based models [31]. In the context of infectious disease spread, the use of compartmental models is most widespread, entailing compartmentalisation of the population according to different disease states relevant for the pathogen under study (for an excellent introduction into compartmental models, see [25]).

Mathematical modeling of host dynamics

The early emphasis on control and elimination strategies that focused on attacking the malaria vector led to a proliferation of compartmental models including a set of equations describing the dynamics of both *hosts and vectors*. However, it is possible to argue that what appears to be common practice is neither a necessary nor sufficient condition to model malaria transmission dynamics. In a "standard" mathematical model of malaria, a parameter similar to FOI is normally expressed by relating the"happening" [6] of infection to the encounter between vector and host, expressed via a set of parameters quantifying the probability of a host-vector contact, i.e., that of a vector bite being infectious, and that of an infectious bite actually leading to an infection. A classic example is offered by Macdonald's model [32], which follows a Susceptible-Infected-Susceptible (SIS) structure for human compartments.

The mathematical equation describing the evolution of the fraction of infected and infectious humans (denoted here by $i_h(t)$) reads as follows:

$$\frac{\mathrm{d}i_h(t)}{\mathrm{d}t} = abmi_m(t)[1 - i_h(t)] - \nu i_h(t), \tag{1}$$

with $\mathrm{d}/\mathrm{d}t$ used for total derivatives with respect to time t, $i_m(t)$ representing the fraction of infected mosquitoes in the population at time t, and ν denoting the recovery rate. The other parameters present in the equation correspond to the human biting rate a, the proportion of bites leading to infection b, and the ratio of female to male mosquitoes in the population, denoted by m. A comprehensive overview related to the notation of the model parameters displayed in this manuscript can be found in the glossary Table 2 in the Appendix A. This can be considered a vector-borne disease adaptation of the general Mass Action Principle (MAP), a concept borrowed from chemistry [see, e.g., [10]], which explicitly links the interaction between susceptible and infected/infectious subjects in a population to the spread of infectious diseases. Briefly, in its most simple form, the principle states that the FOI is proportional to the number (or density) of infected individuals, with a proportionality factor representing the rate of transmission upon a contact between a susceptible and infected individual. In case of a vector-borne disease, the simple MAP can be extended to encompass that the FOI for a human host is proportional to the number of infected mosquitoes, as vectors responsible for transmission, and vice versa when looking at the FOI for mosquitoes. In the present case, $\beta := abm$ represents the so-called (effective) transmission rate upon contact between a susceptible human host and an infectious mosquito. One of the main tenets of the MAP is the assumption of homogeneous mixing, that is humans and vectors mix in a random and uniform manner. Under homogeneous mixing of humans and mosquitoes, the FOI is equal to $\lambda(t) := \beta i_m(t)$. It is apparent that, when excluding the vector population from the system of differential equations, it is no longer possible to infer the FOI through direct application of the MAP. However, counting new infections over a certain period of time, as

previously described, thereby tracking the movement of humans out of the "Susceptible" compartment, enables the estimation of the FOI without *explicitly* accounting for vector dynamics. This leads to the following equation (where the index h, for humans, is kept just to highlight the link with Equation (1), but is dropped in the remainder of the manuscript):

$$\frac{\mathrm{d}i_h(t)}{\mathrm{d}t} = \lambda(t)[1 - i_h(t)] - \nu i_h(t). \quad (2)$$

Under this formulation, there is no need to collect entomological data for the estimation of a, b, m, and $i_m(t)$. The analyst can now focus on the direct estimation of the FOI, and can do so from any study that at the very least collected clinical, parasitological, or serological data (recalling Sect. 2, measurements at level (1) or (2)).

Illustrative examples

Some examples of studies which applied the formulation of the FOI introduced in the previous section are now introduced. For the purpose of keeping the exposition simple, here and in the remainder of this work either age-dependent disease dynamics under the assumption of *endemic equilibrium*, implying that the system is in steady state or time homogeneity is considered, or work in which time-varying though age-invariant transmission is referred to. More specifically, endemic equilibrium entails a disease incidence that is fluctuating over time though around an equilibrium value. From the perspective of a Lexis diagram [see, e.g., [10, 33]] describing the evolution of birth cohorts over "calendar" time and age, time homogeneity implies that both time and age collapse into one common dimension or a single scale. This is particularly helpful when analysing and interpreting age-specific data in order to infer past dynamics, at least in the absence of time-varying dynamics. As a result, under endemic equilibrium, the FOI solely depends on age. In the latter case of age-invariant transmission, the FOI is expressed as a function of (calendar) time (see example of Stadler et al. [34] below). It is common practice to use capital letters to express the *number* of individuals, and lowercase equivalents to express the proportion of individuals in a given compartment. The examples provided below conformed to the aforementioned convention, thereby adapting the notation used by the respective authors, at least in view of this capitalization rule. Other notation and terminology is preserved. A comprehensive overview of the notation and terminology with regard to the model parameters related to the models described in this paper can be found in the glossary included in the Supplementary Material (see Tables 1 and 2 in Appendix A). Despite the fact that simplifying assumptions with regard to either age or time are often made, a general mathematical model under time heterogeneity, hence describing both age- and time-dependent transmission dynamics, can be formulated as presented in Appendix C.2.

Gosling et al. [35] used a Susceptible-Infected-Susceptible (SIS) model to understand the differences in protective efficacy of a series of interventions among infants. Disease transmission is parameterized by an age-specific force of infection $\lambda(a)$ (under the assumption of time homogeneity), as follows:

$$\frac{\mathrm{d}I(a)}{\mathrm{d}a} = \lambda(a)[N(a) - I(a)] - [\chi(a) + \nu(a)]I(a). \quad (3)$$

Here, $I(a)$ refers to the number of infected individuals of age a, $N(a)$ is the number of individuals in age group a (thus $[N(a) - I(a)]$ stands for the number of susceptible individuals of age a), and $\chi(a)$ and $\nu(a)$ express the age-dependent rate of development of clinical disease (which is detected by surveillance and subsequently leads to treatment and recovery from infection), and the natural clearance rate, respectively.

The age-dependent force of infection $\lambda(a)$ is then estimated using data from clinical trials on intermittent preventive treatment of infants as the mean incidence of clinical malaria in the placebo group. The FOI was assumed to be invariant from the immunity level in the base case and was allowed to vary by the immunity level in a set of scenarios.

In a different study, Aguas and co-authors [36, 37] applied a mathematical model to assess how the development of immunity results in age-specific differences in the burden of clinical malaria, between different populations. Their model encompasses a compartment for susceptible individuals, defined as individuals who are completely immunologically naive, but divide infected individuals into patients suffering from clinical/symptomatic malaria (I_1), and patients suffering from asymptomatic infections (I_2). Assuming that immunity protects against clinical symptoms, but not against infection, the model allows individuals in I_1 to transition first to a recovered state (R), and from there to I_2 if re-infected, unless they completely lost their immunity over time. In the latter case, they revert to S and start the cycle again. The FOI $\lambda(a)$ is assumed to be common across infection types, and to depend on age, such as:

$$\lambda(a) = \lambda_0(1 - ce^{-za}), \quad (4)$$

with $\lambda_0(1 - c)$ constituting the lower limit (at age zero), and λ_0 the upper limit for the FOI, while z and c control the steepness and amplitude of the age-specific prevalence curve.

Exclusive modelling of host dynamics can also be employed to investigate relevant features of transmission such as heterogeneity in the acquisition of infection. Corder et al. [38] used the same formulation for the FOI as Aguas and colleagues, and included it in a compartmental SIS model describing disease dynamics by subdividing the population into low and high risk individuals, defined by their risk factors $x_j > 0, j = 1, 2$ for the two risk levels. The authors also consider repeated exposure as a potential modifier of the infection risk, introducing a parameter $\sigma(w) = e^{-\alpha \cdot w}$, where w is the number of clinical *Plasmodium Vivax* malaria episodes suffered from in the past and α describes the rate of immunity development after repeated episodes of infection. As a result, the age-dependent FOI $\lambda(a)$ is modified by previous exposure and risk factors as follows:

$$\lambda(a, w, j) = x_j \sigma(w) \lambda(a) \tag{5}$$

The authors then expanded the SIS model formulation to include an additional compartment for asymptomatic individuals, assuming immunity covers the development of clinical disease, thus describing asymptomatic dynamics by $x_j[1 - \sigma(w)]\lambda(a)$.

On the topic of risk heterogeneity, exclusive modelling of host dynamics has also been employed to investigate the dynamics of relapses for patient infected with *Plasmodium vivax* [34]. Stadler and coauthors constructed a series of mathematical models to describe how after an initial period of drug-induced protection, individuals can become infected again, or suffer from a relapse due to re-activation of liver stage hypnozoites. In their work, different formulations for the parameter expressing relapse risk were compared in order to account for the possibility of heterogeneity in the hazard of reactivation, for instance, across space and between different members of the same population. An immediate link between this study and Equation (2) can be expressed as:

$$\lambda(t) = r_i(t) + \lambda_0 \tag{6}$$

where exit from the "Susceptible" compartment S was thus modelled as the sum of a fixed parameter n expressing a constant infection rate and a parameter $r_i(t)$ expressing relapse rate, potentially time-varying and changing by individual. Although their model does not directly consider individual differences in the risk of reinfection, extensions such as those applied to the relapse risk parameter could, in principle, be applied to the FOI to account for a wider range of observed and unobserved sources of heterogeneities.

Finally, the whole area of within-host modeling, dealing with pharmacological dynamics and development of drug resistance, focuses on models regarding the human host only. However, a description of these models is beyond the scope of this review paper. The interested reader can find a complete classification of malaria modelling types, and a list of interesting publications, as part of the Supplementary Material in [9] and an example of such within-host models is reported in Appendix C.1 of this paper.

Statistical modelling of malaria transmission data

Having highlighted the integration of the FOI in mathematical compartmental models, the next section discusses the topic of its estimation from available data using modern statistical methodologies.

Analysis of longitudinal malaria data

Longitudinal studies constitute a *gold standard* for assessing malaria risk over time. However, these data need to be analysed with caution, given the repeated measurements per individual over time. Furthermore, in the presence of (additional) data hierarchies (e.g., individuals clustered in households, villages) an appropriate account thereof is in order. Moreover, the estimation of relevant epidemiological parameters from longitudinal data should be done by using appropriate statistical techniques to accommodate all model complexities, including, for example, the fact that subsequent episodes of infection and clearance cannot be assumed to be independent of each other. This is something that, despite a positive trend in recent years, is still currently performed by just one in three published cohort studies [39].

Methods that can deal with correlated longitudinal data with additional data hierarchies are available, and include subject-specific generalized (linear) mixed-effects models (GLMMs) and the marginal Generalised Estimating Equations approach [40]. Furthermore, the standard Cox model has been extended to accommodate association and recurrent events when analysing (censored) time-to-event data, for example, by modelling association among observations using a latent frailty term [41], or by making assumptions about the dependence of future on past events (Andersen-Gill [42], Prentice-William-Peterson models [43]). The number of applications of such methods to malaria data is limited, but has been increasing in recent times (see [44–46] for examples of applied studies and [47–49] for methodological investigations).

In general, estimation of the FOI in infectious disease epidemiology dates back to the early work by Muench [50], which is discussed in more detail in Sect. 12 and was based on the assumption of a constant FOI. This restrictive assumption has been relaxed using more flexible

models, including both parametric approaches (e.g., polynomial regression models [51, 52], nonlinear models [53–56], and models with fractional polynomials [10]) and semiparametric (ie, spline-based methods and nonparametric models [10]). For more details, the reader is referred to Hens et al. [10], with a shorter historical perspective provided in [19].

Illustrative example

In recent years, such longitudinal data have been used for a number of different purposes, such as identifying residual sources of malaria transmission [57], investigating the development of immunity over time [58, 59], understanding drug effectiveness [60], performing impact evaluation after interventions such as mass drug administration [61, 62], or following the evolution of the disease in a certain area over time [63], including changes in serological status [64]. Nevertheless, longitudinal parasitaemia and serological data also provide the type of information needed for estimating the FOI, as was already highlighted Sect.2 in 2. More specifically, in order to estimate the prevalence of infection (e.g., seroprevalence or parasite prevalence), denoted as $i_h(t)$ in Equation (2), the analyst should consider the wide range of modern statistical techniques to accommodate measured and unmeasured sources of heterogeneity in disease spread, which are more difficult to account for in mathematical models (e.g., by multiplying the number of compartments or to extend the underlying structure of the model). Such a statistical modeling approach goes beyond "plugging in" a simple estimate for the prevalence, irrespective of its type, into the respective model equation(s).

As an illustration of a statistical modeling approach, the simplest possible analysis, based on a dichotomization of the outcome variable expressing the presence of infection is now described, thus leading to the use of models for binary outcome data such as those mentioned above. Define Y_{ij} as a binary variable that denotes the presence ($Y_{ij} = 1$) or absence ($Y_{ij} = 0$) of parasites (or antibodies) in the blood, for individual $i = 1, \ldots, N$ at the observation number (time point) $j = 1, \ldots, n_i$ (further hierarchical levels, such as clustering within the household or community, can be considered). Denote $\pi_{ij} = P(Y_{ij} = 1 \mid \boldsymbol{x}_{ij}, \boldsymbol{z}_{ij}, \boldsymbol{b}_i)$ as the probability that $Y_{ij} = 1$ given $\boldsymbol{x}_{ij}$, i.e., a $(1 \times (p+1))$-vector of covariate information with respect to p independent variables, and $\boldsymbol{z}_{ij}$ a vector containing information about a set of q individual-specific random effects, denoted by $\boldsymbol{b}_i = (b_{1i}, \ldots, b_{qi})$.

It is relevant to briefly mention the relation between π_{ij} and $i(t)$ in the compartmental models presented above. When π_{ij} depends on time, both express the (point) prevalence, at least in case of Y_{ij} representing *current* infection, that is the proportion of infected individuals at a specific point in time. Even in case of serological data comprising information on past infection, proportions of hosts in compartments of a more complicated mathematical model encompassing loss of immunity could be linked to a statistical model estimating π_{ij}. Needless to say, π_{ij} accounts for individual-specific covariate information, thereby strengthening the relationship further. This notion also underpins the link between mathematical and statistical models which is further illustrated in Sect. 10. For the moment, the notation π_{ij} is employed, as it is more commonly found in the statistical literature dealing with modelling disease prevalence.

Given the covariates and the random effects, the observations $(Y_{ij} | \boldsymbol{x}_{ij}, \boldsymbol{z}_{ij}, \boldsymbol{b}_i) \sim \text{Bernoulli}(\pi_{ij})$ are assumed independent, with conditional mean:

$$\text{E}(Y_{ij} | \boldsymbol{x}_{ij}, \boldsymbol{z}_{ij}, \boldsymbol{b}_i) = P(Y_{ij} = 1 \mid \boldsymbol{x}_{ij}, \boldsymbol{z}_{ij}, \boldsymbol{b}_i) = \pi_{ij}.$$

The GLMM framework allows for the modelling of such conditional mean by transforming it using a so-called "link function", i.e., $g(\cdot)$, and equating the transformed mean to a linear predictor $\eta(\cdot)$:

$$g(\pi_{ij}) = \boldsymbol{x}_{ij}\boldsymbol{\beta}^T + \boldsymbol{z}_{ij}\boldsymbol{b}_i^T =: \eta(\boldsymbol{x}_{ij}, \boldsymbol{z}_{ij}, \boldsymbol{b}_i),$$

where $\boldsymbol{\beta}$ represents a vector of unknown model parameters to be estimated from the data. The choice of the link function determines the interpretation of the parameter estimates.

Modelling the linear predictor is not necessarily restricted to basic forms of GLMMs, and more complex models can be envisaged accounting for nonlinearities in the age effect or other features such as seasonality. An example of such a model, the generalized additive mixed model (GAMM), is extensively discussed in the monograph by Wood [65].

Combining mathematical and statistical models

With the aim of arguing in favour of combining mathematical and statistical modelling approaches when estimating malaria transmission rates, some relevant examples from the available literature are highlighted, after which a general framework will be presented and discussed.

Illustration 1: linking malaria metrics

The structure of the malaria transmission cycle is such that transmission metrics are necessarily related to each other. For instance, given that the introduction of parasites is related to mosquito bites, there must be a relationship between the PR and EIR, and consequently between EIR, expressing transmission efficiency, and FOI expressing transmission intensity. This relationship can be investigated in the following ways:

O1 Mathematically, from mathematical models of transmission expressing the mechanistic form of such relationships,

O2 Statistically, by examining the association between estimates of malaria metrics collected as part of epidemiological studies (coupled estimates from the same study, or as part of a meta-analysis),

O3 By positing a mathematical model making such relationship explicit, and estimating its parameters statistically.

As an elucidation of (O1), let us introduce the first model relating EIR and prevalence by Ross [17, 66]. The model follows an SIS structure, and under the assumption that the EIR is constant and that people become susceptible again after clearance, the EIR-PR relationship corresponds to [67]:

$$\frac{\mathrm{dPR}(t)}{\mathrm{d}t} = b \cdot \mathrm{EIR}[1 - \mathrm{PR}(t)] - \nu \mathrm{PR}(t),$$

where ν corresponds to the clearance rate, $b \cdot \mathrm{EIR}$ to the infection rate, and PR represents the proportion of infected individuals in the population (i.e., individuals in compartment I). Solving this equation leads to the following equilibrium solution:

$$\frac{\gamma \mathrm{EIR}}{\gamma \mathrm{EIR} + 1}, \tag{7}$$

with γ equal to b/r being the time to clearance per infectious bite.

An early instance of a statistical approach (O2) is the work by Beier and colleagues, who applied a linear model directly to pairs of EIR and PR estimates derived from cohorts of Kenyan children [68]. Recognizing that this model structure would yield poor predictions, particularly in regions approaching malaria elimination, they demonstrated in a subsequent study [69] that a log-linear relationship between PR and EIR offers a more accurate fit to their data.

As an example of (O3), consider the work by Smith and coauthors [3, 70]. EIR and FOI were long supposed to be linearly related, but the data did not support these conclusions [68, 71, 72]. Smith and colleagues therefore focused on nonlinear mathematical models which could explain, for instance, the observed discrepancy between intensity and efficiency. Other sources of individual heterogeneity must also be taken into account, such as immunity [73], heterogeneous biting [70] and the role of factors such as host size, which changes with age, or reduction in susceptibility to infectious inoculations [23]. Starting from the model proposed by Ross, the authors formulated SIS and SIRS (Susceptible-Infected-Recovered-Susceptible) mathematical models by changing a set of critical assumptions regarding the role that individual heterogeneity plays in susceptibility to biting and development of infection, immunity, and clearance. Each model led to a different formulation for the relationship between PR and EIR, which was then fit to coupled PR/EIR values using maximum likelihood techniques. Model selection was performed using Akaike's Information Criterion (AIC; [74]), and model fit was compared to Beier's log-linear model. For comparison with the previous equation, what resulted to be the best model, an SIS model with heterogeneous infection rates (distributed as a Gamma$(1, 1/d)$) and superinfection is reported, leading to the following relation:

$$\mathrm{PR}(t) = 1 - \left[1 + \frac{b\mathrm{EIR}(t)}{\nu d}\right] - d.$$

Amoah and colleagues [67] further modified the basic SIS model by allowing different infection and recovery rates for women and children, as well as superinfection. The authors reviewed a range of previous models and proposed a "logit-linear" extension of the original Beier "log-linear" formulation, critiquing the fact that in the log-linear model, as EIR goes to zero parasite rate goes to minus infinity, and approaches infinity as EIR goes to infinity:

$$\log\left[\frac{\mathrm{PR}(t)}{1 - \mathrm{PR}(t)}\right] = \alpha + b \cdot \log[\mathrm{EIR}(t)].$$

Overall, and unlike the work by Smith et al., Amoah and coauthors considered statistical models (log-linear and logit-linear) to outperform the functional relationships derived from mathematical transmission models, with the logit-linear model ultimately resulting in the best fit to the observed data. This gives an indication that the transmission models considered need further scrutiny.

Illustration 2: estimating the FOI using catalytic models

Catalytic models provide a way to characterise age-prevalence curves and fit data to estimate the FOI. They can be used to infer the acquisition of infection (and eventual immunity) given exposure to an infectious agent [8, 50]). In their original formulation, it was assumed that the infection would leave an exposure mark on a proportion k of the population, so it would be possible to distinguish between previously infected and uninfected individuals, as is, for example, the case for the antibody response

left by immunizing infections. Moreover, a constant FOI ("exposure rate") λ per unit of time was assumed, and, a fraction of the population was allowed to escape infection (l). The catalytic curve can be thus described as $\pi(t) = k(l - e^{-\lambda t})$ [see, e.g., [19]].

Catalytic models [75] were applied to malaria for the first time in the 1950 s by George Macdonald [76], in a simplified version which assumed $k = l = 1$ [76], thereby leading to:

$$\pi(t) = 1 - e^{-\lambda t}. \tag{8}$$

The link between the aforementioned compartmental models and the catalytic models is evident, since Equation (8) can be obtained directly by solving an SI (Susceptible-Infected) model for the proportion of infected, under the assumption of time homogeneity. Furthermore, Equation (8) corresponds to the expression for the cumulative distribution function of infection times in the case of an exponential time-to-event process [77].

An extension of the disease dynamics implies a more complicated relationship between prevalence and FOI, which could entail including a constant rate of "acquisition" of infection as well as of "reversion" from it, mirroring the extension of an SI model to account for return to the susceptible state in an SIS model structure:

$$\frac{\mathrm{d}i(t)}{\mathrm{d}t} = \lambda[1 - i(t)] - \nu i,$$

with λ and ν denoting the FOI and reversion rate, respectively, and $i(t)$ being the proportion of infected individuals at time t. This leads to the following analytical solution for $i(t)$:

$$i(t) = \frac{\lambda}{\lambda + \nu}[1 - e^{-(\lambda+\nu)t}], \tag{9}$$

thus allowing for inference regarding infection and reversion rates based on available data.

In the analysis of malariometric data, early applications of catalytic models can be found in [78, 79]. Pull and Grab [78] applied Muench's model for evaluating the risk of malaria infection in infants. Under the steady-state assumption and constant FOI, they analysed parasitaemia data obtained from a series of cross-sectional surveys. They estimated the FOI by applying the method of moments to Equation (8).

Bekessy and Molineaux [79] extended the approach to model longitudinal data for the simultaneous estimation of incidence and recovery. They proposed to model transitions from an individual's state (positive or negative) from time t to time $t + 1$ as a Markov process, with the probability of switching states assumed to follow a Binomial distribution. The solution of the corresponding system of equations is such that the probability of infection at time t corresponds to Equation (9). Parameter estimation was performed using Maximum Likelihood (ML) estimation [80].

Extensions to the standard reversible catalytic model include the work by Yukich and coauthors [81], who suggest integrating prevalence and health facility data, arguing that the estimation of the FOI via longitudinal studies requires long follow-up and large sample sizes, especially at low levels of transmission. However, information on clinical incidence is frequently readily available by collating data collected via routine healthcare systems, often reported in the form of Annual Parasite Index (API). Their work extends standard catalytic models by including the possibility of changing the duration of infection according to treatment, and also features a parameter τ, expressing the proportion of infections that are immediately treated and are not computed as part of the prevalence of the community. In this case, $\lambda(t) = (1 - \tau)\lambda_0(t)$, that is the FOI without treatment $\lambda_0(t)$ is rescaled by a factor $1 - \tau$. Moreover, the average clearance rate, μ_0 can be re-expressed as the sum of two components, one being the natural clearance rate, ν_0 and another, $\tau\lambda_0$, expressing the treatment rate. Equation (9) is thus modified into:

$$\frac{\mathrm{d}i(t)}{\mathrm{d}t} = (1 - \tau)\lambda_0(t)[1 - i(t)] - (\mu_0 + \tau\lambda_0(t))i(t),$$

which can be solved to express the FOI as:

$$\lambda_0(t) = \frac{\mu_0}{\frac{1-\tau}{i(t)} - 1}.$$

Surveillance data can then be used to calculate the value of treatment parameters and integrate with prevalence measures, such as those available from the Malaria Atlas Project [82]. Note how the estimation of the FOI, and ultimately its extension to allow for time-dependency, hinges on the value $i(t)$, which can be obtained by modelling prevalence and surveillance data using the statistical techniques mentioned in Sect. 7.

Even though the catalytic model was originally applied to parasitological data, it has in more recent times being championed for the analysis of serological data as well. The idea is that seroprevalence data, reflecting history of exposure and being less influenced by fluctuations such as those induced by seasonality and parasite densities, could provide a complementary way to estimate

transmission. A short summary is provided in Appendix D.1 and refer the interested reader to the review paper by Sepúlveda et al. [7].

General framework for analysis

A general framework for the estimation of the FOI, subsuming the illustrations previously outlined, is now formulated. The framework extends the idea at the core of a catalytic model, that is being able to derive an expression for the FOI in terms of the model parameters associated with an underlying mathematical model structure, constructed to resemble state of the art knowledge about the biological processes of interest, and employing statistical methods to estimate these parameters. However, there is no need to restrict one's attention to an SI or SIS model structure, or to assume that the FOI itself is constant. As an example, the formulation will be first derived from an SIR (Susceptible-Infected-Recovered) setting to highlight the flexibility of the framework, after which an SIS model adaptation is exemplified in the illustrative example below.

In terms of estimating the (relevant parameters associated with the) FOI, two distinct approaches are considerd (see below for a specific example thereof), which are defined as a *direct* or *indirect* parameterisation [10]. In the former case, a (parametric) model for the FOI is directly posited. The parameters describing such functional form enter the likelihood function for the available data (e.g., parasitaemia) directly, and their estimates are obtained, for instance, by direct ML estimation techniques. In the latter case, an estimate for the FOI is obtained by first modelling prevalence data, after which the estimated prevalence is used to estimate the FOI, conditional on the imposed (compartmental) mathematical model structure of choice and its implied link between the prevalence on the one hand and the FOI on the other hand (e.g., Yukich et al. [81]).

Consider again the case of a dichotomous response variable expressing presence or absence of infection (e.g., presence of parasites in the bloodstream, or previous positivity), let $Y_i, i = 1, \ldots, N$ for N individuals in the sample, and assume for illustration purposes that a disease follows an SIR model under time homogeneity, for which $\pi(a) = 1 - s(a) = 1 - e^{-\int_0^a \lambda(u)du}$. That is, once recovered, a person does not contract the infection again, hence, the proportion of people who are of a certain age a are or were infected is equal to one minus the proportion of people who have not yet experienced the infection. This is equivalent to the well-known concept of a survival or survivor function in survival analysis. Equivalently, the survival function expressing the probability of not having experienced the event of interest (i.e., infection in this case) at a specific time point is related to the cumulative (or integrated) hazard of infection $\lambda(a)$ up to age a [see, e.g., [83]].

To better understand the difference between the two aforementioned approaches (i.e., direct and indirect parametrizations) towards estimation of the FOI, one can compare their implied binomial log-likelihood functions:

Direct Parameterization

$$l(\boldsymbol{\theta}) = \sum_{i=1}^{N} y_i \log[1 - e^{-\int_0^a \lambda(u;\boldsymbol{\theta})du}] + (1 - y_i) \log[e^{-\int_0^a \lambda(u;\boldsymbol{\theta})du}]$$

Indirect Parameterization

$$l(\boldsymbol{\theta}) = \sum_{i=1}^{N} y_i \log[\pi(a_i;\boldsymbol{\theta})] + (1 - y_i) \log[1 - \pi(a_i;\boldsymbol{\theta})]$$

where, as above, $\pi(a_i;\boldsymbol{\theta})$ denotes, for example, the age-dependent parasite prevalence corresponding to observation i ($i = 1, \ldots, N$) depending on model parameters $\boldsymbol{\theta}$. It is evident to observe how the model for the direct parametrization features the model for $\lambda(a;\boldsymbol{\theta})$, making the dependence on the model parameters $\boldsymbol{\theta}$ explicit, directly in the likelihood, whereas for the case of an indirect parametrization, a model for $\pi(a_i;\boldsymbol{\theta})$ appears in the likelihood formulation.

As a result, in the case of a direct parametrization, maximization of the likelihood leads directly to the estimation of the FOI, $\widehat{\lambda}(a)$. An indirect parameterization instead requires modelling of the (parasite) prevalence first, so that, recalling the notation of generalized linear models introduction in Sect. 7 one starts with an estimate $\widehat{\pi}(a) = g^{-1}[\widehat{\eta}(a)]$, suppressing for brevity the dependence on additional covariates, model parameters $\widehat{\boldsymbol{\theta}}$ and (individual-specific) random effects (for unmeasured heterogeneity). In a second step, an estimate for the FOI is obtained as:

$$\hat{\lambda}(a) = \frac{\widehat{\pi}'(a)}{1 - \widehat{\pi}(a)}, \tag{10}$$

where $\pi'(a) = \mathrm{d}\pi(a)/\mathrm{d}a$ is the derivative of the prevalence with respect to age. Again, it is useful to stress the link with survival analysis [77]. More specifically, as pointed out before, the FOI represents the hazard of infection which can be expressed as the ratio of $f(t)$, the density function for the time to experiencing malaria infection, which is the change in prevalence $\pi'(a)$, and the survival function $S(t)$ equal to the proportion of susceptible individuals $[1 - \pi(a)]$.

Different formulations of the mathematical model lead to different solutions for the proportion of susceptible individuals, and thus to different formulations of the link between FOI and prevalence, which can be contrasted with available data to identify the best fitting model. At the same time, the form of $\pi'(t)$ and its relationship with $\pi(t)$ depends on the link function chosen for the statistical model of malaria prevalence. In general, for a binary response variable, the FOI takes the following form $\lambda(a) = \eta'(a)\delta[\eta(a)]$, with the second term determined by the choice of the link function [10].

Some examples illustrating the framework outlined above are introduced. Direct parameterization requires specifying a model for $\lambda(a)$, and examples of this were discussed in previous sections, such as the study by Aguas et al. [36] (see Equation (4)) or the work by Smith et al. [3]. For additional examples of mathematical models for the FOI, the interested reader is referred to the review by Hens and coauthors [19].

An example of an indirect parameterization used to estimate the FOI and applied to a longitudinal dataset of *P. falciparum* parasitaemia data [84] is presented here. Recalling the notation presented so far, the main aim is to be able to provide a mathematical link between $\lambda(t)$, the time-dependent FOI, and $\eta(\boldsymbol{x}_{ij}, \boldsymbol{z}_{ij}, \boldsymbol{b}_i)$, the linear predictor of a GLMM modelling available parasitaemia data. This can be achieved since GLMMs relate the mean of a dependent variable, such as point prevalence in case of a binary endpoint, to a linear predictor, relying on a chosen link function. In turn, the given structure of a mathematical model will provide a way to link the point prevalence to the FOI [10], by re-expressing the starting system of equations so that the FOI can be expressed as a function of other, known or estimable, quantities.

The work by Mugenyi and colleagues [84] offers an example of such analysis, where the following SIS system of equations in (11)

$$\begin{cases} \dfrac{\mathrm{d}s(t)}{\mathrm{d}t} = -\lambda(t)s(t) + \nu i(t) \\ \dfrac{\mathrm{d}i(t)}{\mathrm{d}t} = \lambda(t)s(t) - \nu i(t) \end{cases} \tag{11}$$

is re-expressed to formulate a relationship between FOI and PR:

$$\lambda(t) = \frac{i(t)\nu + i'(t)}{1 - i(t)}, \tag{12}$$

with $i'(t) = \mathrm{d}i(t)/\mathrm{d}t$ and ν being the rate of clearance, which is assumed to be constant over time (all other elements can be found in the Glossary). Although the previous derivations were done under the assumption of time homogeneity, implying that age and time dimensions collapse as pointed out earlier, in the work by Mugenyi and colleagues the parasite prevalence and implied force of infection is assumed to depend on both age and time, while adjusting for other measured covariates (e.g., related to treatment) and accommodating clustering at individual, household and district level. In their study, the authors provide estimates for the prevalence by analysing data on parasitaemia from a cohort of Ugandan children by making use of a GLMM accounting for clustering at individual and household level. This estimate for the (age- and time-dependent) prevalence is then plugged into (12), together with estimates of ν gathered from the literature, to obtain an estimate for the FOI.

Despite the limitations of the proposed SIS model structure for the specific case of modelling malaria transmission dynamics, the work constitutes an interesting starting point that warrants further development, which will have to deal, in particular, with the complexities when considering other more realistic model structures [see, e.g., [73]]. Other developments could target improvements in the flexibility of the relation between the prevalence and, for example, age and/or time, through the use of, for example, GAMMs [65].

Conclusion

This integrative review summarizes key concepts related to mathematical and statistical modeling of malaria transmission with the aim of stressing the importance of combining both frameworks. Despite the fact that this dates back to the early work by Ross [6], an integrated approach that allows estimation of model parameters from data is often lacking. Illustrative examples for each of the modelling frameworks separately and for a general framework combining the approaches are provided. Although various malaria metrics are proposed and used in the literature, this review mainly focuses on the use of the force of infection, thereby highlighting its close relationship with concepts in survival analysis and emphasizing its popularity in infectious disease epidemiology for close contact infections. Despite the additional complexity of vector-borne transmission of malaria, concepts like the mass action principle extend to the malaria context in a straightforward way as exemplified in this paper.

The current work offers a methodological perspective with respect to estimating the intensity of malaria transmission. After summarizing the state of the art concerning the nature, advantages, and disadvantages of the most commonly used metrics, a series of arguments favouring a more widespread use of the force of infection

was presented. Through the elucidation of previous approaches working towards methodological linkages, a unified framework centered on the interaction between mathematical and statistical models was introduced as a recommendation for future research. Such framework entails:

1. Expressing a (realistic compartmental) mathematical model;
2. Expressing a parameter of interest, as a function of other quantities estimable from observed data, such as point prevalence;
3. Estimating those quantities via statistical techniques, accommodating data complexities, taking into account relevant issues such as nonlinearities in space-time and (observed and unobserved) heterogeneities;
4. Replacing such estimates into the formula made explicit in point 2) to obtain a value for the desired quantity.

This framework provides a number of advantages in the specific context of estimating malaria *transmission* parameters, as it allows the focus to be on a set of equations expressing the dynamics of the host only, thus greatly simplifying the mathematical model structure and, by avoiding explicit reliance on the Mass Action Principle, eliminating the need for the collection and analysis of entomological observations.

This implies restricting our attention to the implementation of the following two metrics: the Force of Infection and the Seroconversion Rate. In fact, both express the rate of change from the state of susceptibility to the state of infectiousness, acting as a real-time transition "tracker". From a statistical point of view, the FOI can be flexibly estimated with a wide range of fully, semi-, or non-parametric techniques for handling the underlying infection measure, such as parasitaemia (or parasite density) [10, 84, 85].

In line with the work by Mugenyi et al. [84], further developments to the statistical modelling of parasitaemia data should be given even more prominence, in order to make the best possible use of longitudinal data, and assess the suitability of flexible modelling strategies, including Generalized Nonlinear (Mixed) Models, Generalised Additive Models, and splines-based approaches [19, 85], on top of a comprehensive assessment of heterogeneity, which should move beyond the sole consideration of available covariates, and include hidden variability in transmission, which can be assessed by inclusion of random effects or frailties [11, 86].

The choice of the mathematical model depends on the availability of entomological and human malaria data. More specifically, in case information regarding the vector population is available, this could be included in the mathematical model structure, thereby enabling a description of the connection between transmission dynamics in the vector and host populations. Despite the fact that this is possible, and many papers have studied such an explicit connection, the overarching methodology connecting mathematical and statistical modelling approaches as introduced in this paper, does not require that data on both human and vector populations are available. Based on assumptions with regard to the underlying transmission process, transmission intensity can be quantified in terms of the force of infection, and the effects of different vector control measures are indirectly quantified in terms of the reduced risk of malaria transmission for humans. As exemplified by the different examples included in the integrative review paper, estimation of the force of infection in the absence of data on transmission dynamics in the vector population simplifies the analysis considerably. Admittedly, the proposed general modelling framework has not been used to evaluate the impact of ignoring the vector population dynamics in relation to such control measures, nor for the quantification of effects of disregarding the vector completely on malaria metrics for the host in general. Although the methodology has been used in the absence of entomological information and solely based on, for example, parasitaemia data, a critical appraisal of the results when different data sources are available would be of utmost importance in the future and therefore provides an interesting avenue for further methodological and applied research.

Appendix A Glossary

See Tables 1, 2

Table 1 Summary of main terms used in the manuscript

Term	Definition
Endemic equilibrium	Situation where disease transmission has stabilised, thus, as disease dynamics repeat regularly, calendar time becomes less relevant, and an individual's age really drives exposure to the disease.
Entomological Inoculation Rate (EIR)	Number of infectious bites an individual receives over a certain period of time.
Force of Infection (FOI)	The force of infection (λ) represents the per capita rate at which susceptible individuals become infected. In survival analysis terms, it corresponds to the hazard of infection.
Mass action principle (MAP)	The MAP expressed the notion that disease transmission is related to the number of susceptible and infectious individuals in a population, as well as to the rate of contact between the two groups.
Parasite Rate (PR)	Proportion of individuals testing positive for the presence of malaria parasite in their blood at a given time.
Parasitological data	Information obtained from tests analyzing a blood sample for the presence of parasites in the bloodstream.
Seroconversion rate (SCR)	Rate at which individuals transition from seronegativity to seropositivity status as a result of immune system reaction due to vaccination or infection.
Serological data	Information obtained from tests analysing a blood sample for the presence of antibodies, antigens, or other immune markers.

Table 2 Summary of notation for the parameters used in the manuscript. The equation reported corresponds to first usage

Notation	Parameter
a	Human biting rate (Equation (1))
b	Proportion of bites leading to infection (Equation (1))
$\boldsymbol{b_i}$	Individual specific random effects
I/i	Without a subscript, number/proportion of infectious individuals in the population, respectively. Subscripts are used to differentiate between humans, h and mosquitoes m when necessary (e.g., Equation (1)).
m	Ratio of female to male mosquitoes in the population (Equation (1))
N(a)	Number of individuals in age group a (Equation (3))
r_i	Relapse rate of individual i (Equation (6))
w	Number of clinical malaria episodes suffered in the past (Equation (5))
$\boldsymbol{x_{ij}}$	Vector of covariate information
Y_{ij}	Binary variable expressing presence or absence of parasite in the blood, for individual $i = 1, \ldots, N$ at time point $j = 1, \ldots, n_i$
$\boldsymbol{z_{ij}}$	Vector collecting information individual-specific random effects
α	Rate of development of immunity (Equation (5))
β	Effective transmission rate upon contact between a susceptible human host and an infectious mosquitoes (in Equation (1) $\beta = abm$)
γ	Time to clearance per infectious bite (Equation (7))
$\eta(.)$	Linear predictor of a regression model
$\boldsymbol{\theta}$	Any general set of model parameters
μ	Average clearance rate (in Equation (4.2) only)
ν	Natural recovery (clearance) rate (Equation (1))
$\pi(a)$	Age-dependent parasite prevalence. For the purpose of statistical modelling, it correponds to the probability that $Y_{ij} = 1$given covariates and random effects.
τ	Proportion of infections treated immediately (Equation (4.2))
χ	Rate of development of clinical disease (Equation (3))

Appendix B Literature review

In this appendix, additional details are provided concerning the methodological review of existing malaria literature relevant for the discussion presented in the main text. First, the scope and criteria used for this review are highlighted. Second, the search strategy is documented and described.

Scope and criteria guiding the review

In order to provide a comprehensive overview of the available literature, a combination of sources was deemed relevant. These include a set of key publications identified by the authors, as well as searches in the online databases PubMed. In line with the recommendations and guidance by [87], for each selected publication backward and forward citation searches were performed. The following

criteria for searching, screening, and selecting papers are considered:

- The article provides a direct estimation of the FOI, in the context of a statistical and/or mathematical model, together with an explicit description of the estimation method used.
- The article provides a description of the relationship between transmission metrics, using an underlying mathematical, statistical model, or statistical model informed by a mathematical transmission model.
- The article clearly aims at discussing or reviewing the characteristics and use of the FOI, alone or in combination with other metrics, in the context of malaria.

In order to be included in this review, papers had to fulfil at least one of these criteria.

Search strategy for the literature review

No restrictions were placed on the type of study included in the review or the year of publication, though only studies written in the English language were considered.

A summary list of key publications identified by the authors as a starting point for the review are reported here:

- Early contributions: [17, 71, 76, 88]
- Contributions based on the work by Muench [50, 75]: using cross-sectional survey data [78], using longitudinal data [79]
- Contemporary contributions: on the use of Parasite Rate: [81], on the relationship between metrics [3, 14, 16, 23, 70, 89]
- Previous literature reviews: [9, 15, 39]

In order to optimize the search strategy in PubMed, the most appropriate MeSH term were identified by first establishing:

- The disease of interest: Malaria
- The topic of interest: estimation of transmission

For the former component, the most appropriate starting point in the MeSH database for malaria is the disease itself, "Malaria". This retrieves 10 entries of which the most appropriate are "Malaria", "Malaria, falciparum", and "Malaria, Vivax". Other terms can be safely excluded, such as those referring to the malaria vaccine, to avian malaria, etc. As a consequence of the presence of terms not relevant for the purpose of this review, the default option of "exploding" the MeSH term to all its subheadings was employed. The most relevant subheadings appear to be the following: "analysis", "epidemiology", "statistics", "prevention and control", "transmission". It is possible to supplement MeSH terms with appropriately selected keywords, but given the MeSH terms seem to cover the topic comprehensively, this did not seem necessary.

For the latter component, it is important to identify appropriate words concerned with malaria transmission as captured by FOI. In the MeSH Dataset, the most directly related terms appear to be the following:

- Disease Transmission, Infectious
- Incidence

The relevant MeSH group appears to be group G03 (Biological Sciences - Environment and Public Health).

Based on prior knowledge of the relevant literature, it seemed sensible to supplement the searches with the following keywords to be looked for in the articles Title and Abstract:

- Force of infection
- Transmission Rate
- Attack Rate
- Effective Inoculation Rate
- Parasitological Inoculation Rate

The following search strategy was thus employed in PubMed:

(Malaria[MeSH Terms]) AND ("force of infection"[Title/Abstract] OR "attack rate"[Title/Abstract] OR "transmission rate" [Title/Abstract] OR "effective inoculation rate" [Title/Abstract] OR "parasitological inoculation rate" [Title/Abstract])

These results were supplemented by backward and forward citation searches from the key papers.

Appendix C Mathematical modelling

Within-host modelling example

The review by Reiner et al. [9] subdivides publications into those that modelled the vector explicitly, implicitly (i.e., via quantities such as vectorial capacity), or not at all [9]. Among the typology of models which did not consider vector dynamics explicitly, one category is those type of models that focus instead on "within-host" dynamics, which go beyond the focus of the current paper but constitute perhaps the most intuitive way in which vectors can be excluded without losing model accuracy.

An example is the study by [90], on the determinants and pharmacological dynamics of artemisin-resistant

malaria in the Cambodian setting. Their mathematical model encompasses a series of stages summarizing the development of the parasite inside the human host, according to the following sequence: Susceptible - Liver Stage - Blood Stage (non-infectious) - Blood Stage (infectious). In their model, the transition between the Susceptible and Infected stage were modelled as follows:

$$\frac{(1-\rho)\beta S_{dg}\sum_g\sum_d\sum_r(1-c_r)I_{rdg}}{N},$$

with ρ representing a reduction in transmissibility due to effective drug coverage, c_r representing the clearance rates (by treatment), S_{dg} being the number of susceptible by drug type and control strategy evaluated, I_{rdg} being the number of individuals in the Infectious blood stage, by resistance status, drug activity, and strategy evaluated, and finally, β representing the probability that a susceptible individual receives and infectious bite from an infectious vector who had previously bitten a randomly-chosen infectious individual. It is evident how the representation of the force of infection can be reconducted to the "classic" $\frac{\beta I}{N}$ structure, where all possible interactions across the groups identified by the specific study (in this case, drug type and control stategies) are considered.

Age- and time-dependent transmission dynamics

A general model to describe non-stationary disease dynamics (i.e., the disease is not in steady state or endemic equilibrium) requires keeping track of the age- and time-specific evolution in proportion of individuals in different compartments. Such dynamics are then translated into a system of Partial Differential Equations (PDEs) [29], describing the evolution of disease dynamics across different age groups and calendar times, e.g. extending the previous SIRS model even further:

$$\begin{cases} \frac{\partial s(a,t)}{\partial a}+\frac{\partial s(a,t)}{\partial t} &= -\lambda(a,t)s(a,t)+\gamma(a,t)r(a,t),\\ \frac{\partial i(a,t)}{\partial a}+\frac{\partial i(a,t)}{\partial t} &= \lambda(a,t)s(a,t)-\nu(a,t)i(a,t),\\ \frac{\partial r(a,t)}{\partial a}+\frac{\partial r(a,t)}{\partial t} &= \nu(a,t)i(a,t)-\gamma(a,t)r(a,t). \end{cases}$$

Solving a PDE system is complex, therefore these models are normally approximated by a tractable set of ordinary differential Equations (ODEs). An efficient way to do this is by the means of a so-called realistic age-structured model (RAS-model) [10, Chp. 16]. Under a RAS implementation, individuals are allowed to move across compartments over a time period of one year, assuming age groups of one year, after which they are moved to the next age group. RAS models better reflect aging and temporal dynamics than continuous age-structured (CAS) models in which individuals continuously transition and can grow old immediately.

Having simplified the original system of PDEs, it is possible to solve, analytically or numerically, the specific age-cohort system of (time-varying) differential equations to obtain a relation between the FOI and incidence (or prevalence), which is subsequently contrasted to available data, such as serological surveys, for the estimation of relevant parameters. Different model formulations can then be compared and assessed for model selection using standard statistical methods. For an example of this approach albeit in a different context, the reader is referred to the study by Ogunjimi and colleagues [91].

Alternatively, it is possible to posit *endemic equilibrium* [12, Chp.4], where the system has reached a steady state and only dependence on age is considered, thus dropping the derivatices with respect to time and the indices t. The system then reduces to a set of Ordinary Differential Equations (ODEs), which are more easily tractable and could in principle be solved, leading to the formulation of a direct link between the FOI and prevalence.

Appendix D Statistical modelling

Analysis of serological data

Serological data have been mostly analysed by fitting catalytic models to them [21, 78] after identifying seronegative and seropositive individuals. Drakeley and colleagues [8, 21] fitted a reversible catalytic model, assuming a binomial error distribution and estimating parameters using maximum likelihood estimation, to antibody data from Kenya collected during two cross-sectional surveys of 250 individuals across 12 villages spanning different altitude levels. Seroprevalence correlated better with measures of altitude and EIR than point prevalence. Moreover, a separate analysis of short- and long-lived antibody responses, stratified by age group, made it possible to disentangle recent and distant patterns in transmission. The catalytic model is quite restrictive and more flexible models are typically required to describe disease dynamics, something which can be done by refining the underlying mathematical model. As an example, Bosomprah et al. [92] incorporates the "boosting" effect of continuous exposure, which can explain the long life of certain sets of antibody markers, by implementing a model featuring superinfection to explain seropositivity to malaria antigens. In addition to these modifications, it is possible to consider different structures for the changes in transmission over time, which can be assumed to undergo

a sharp drop at a certain time point, or consider, for instance, a linear or quadratic evolution in FOI between the start and the end of the study period. More recently, mixture models have been considered to use the full range of antibody titers [93, 94], and to avoid the (subjective) specification of fixed cut-off point(s) thereby di- (seronegative or -positive) or tri-chotomising (including an inconclusive/equivocal result when titer concentrations are between an upper and lower threshold value) the test result, thus leading to considerable information loss.

Abbreviations

AIC	Akaike information criteria
API	Annual parasite index
EIR	Entomological inoculation rate
FOI	Force of infection
GAMM	Generalised additive mixed model
GLMMs	Generalised linear mixed models
MAP	Mass action principle
mFOI	Molecular force of infection
ML	Maximum likelihood
PR	parasite rate
SCR	Seroconversion rate
SI	Susceptible-infected
SIR	Susceptible-infected-recovered
SIRS	Susceptible-infected-recovered-susceptible
SIS	Susceptible-infected-susceptible
WHO	World Health Organization

Author contributions

AG: Conceptualisation, Investigation, Writing - Original Draft; NH: Conceptualisation, Writing - Editing and Review, Supervision; SA: Conceptualisation, Writing - Editing and Review, Supervision.

Funding

AG is funded by a Bijzonder Onderzoeksfonds (BOF) grant, awarded by the University of Antwerp.

Data availability

The present work does not rely on data, code or other resources. Information which could help reconstructing the steps undertaken to perform the review have been included in the Appendix.

Code availability

Not applicable.

Declarations

Ethics approval and consent to participate

Not applicable.

Consent for publication

Not applicable.

Competing interests

The authors have no known competing financial interests or any other relationships that could have influenced the content or the work related to this manuscript, or lead to Conflict of interest.

Received: 19 November 2023 Accepted: 15 May 2025
Published online: 30 May 2025

References

1. Kelly-Hope LA, McKenzie FE. The multiplicity of malaria transmission: a review of entomological inoculation rate measurements and methods across sub-Saharan Africa. Malar J. 2009;8:19.
2. World Health Organization. WHO malaria terminology. Geneva: World Health Organization; 2024. https://www.who.int/publications/i/item/9789240038400.
3. Smith DL, Drakeley CJ, Chiyaka C, Hay SI. A quantitative analysis of transmission efficiency versus intensity for malaria. Nat Commun. 2010;1:108.
4. malERA Refresh Consultative Panel on Characterising the Reservoir and Measuring Transmission. malERA: An updated research agenda for characterising the reservoir and measuring transmission in malaria elimination and eradication. PLoS Med. 2017;14:11.
5. Rabinovich RN, Drakeley C, Djimde AA, Hall BF, Hay SI, Hemingway J, et al. malERA: an updated research agenda for malaria elimination and eradication. PLoS Med. 2017;14:11.
6. Ross R. An application of the theory of probabilities to the study of a priori pathometry. Part I Proc R Soc Lond A. 1916;92:638.
7. Sepúlveda N, Stresman G, White MT, Drakeley CJ. Current mathematical models for analyzing anti-malarial antibody data with an eye to malaria elimination and eradication. J Immunol Res. 2015;2015: 738030.
8. Drakeley C, Corran P, Coleman P, Tongren J, McDonald S, Carneiro I, et al. Estimating medium-and long-term trends in malaria transmission by using serological markers of malaria exposure. Proc Natl Acad Sci USA. 2005;102:14.
9. Reiner RC Jr, Perkins TA, Barker CM, Niu T, Chaves LF, Ellis AM, et al. A systematic review of mathematical models of mosquito-borne pathogen transmission: 1970–2010. J R Soc Interface. 2013;10:20120921.
10. Hens N, Shkedy Z, Aerts M, Faes C, Van Damme P, Beutels P. Modeling infectious disease parameters based on serological and social contact data: a modern statistical perspective. 1st ed. Germany: Springer Science & Business Media; 2012.
11. Abrams S, Hens N. Modeling individual heterogeneity in the acquisition of recurrent infections: an application to parvovirus B19. Biostatistics. 2015;16:1.
12. Anderson RM, May RM. Infectious diseases of humans: dynamics and control. 1st ed. United Kingdom: Oxford University Press; 1992.
13. Keeling MJ, Rohani P. Modeling infectious diseases in humans and animals. 1st ed. New Jersey: Princeton University Press; 2011.
14. Tusting LS, Bousema T, Smith DL, Drakeley C. Measuring changes in Plasmodium falciparum transmission: precision, accuracy and costs of metrics. Adv Parasitol. 2014;84:151–208.
15. Mandal S, Sarkar RR, Sinha S. Mathematical models of malaria-a review. Malar J. 2011;10:202.
16. Smith DL, Battle KE, Hay SI, Barker CM, Scott TW, McKenzie FE. Ross, Macdonald, and a theory for the dynamics and control of mosquito-transmitted pathogens. PLoS Pathog. 2012;8:4.
17. Ross R. Some quantitative studies in epidemiology. Nature. 1911;87:2188.
18. Macdonald G. The analysis of equilibrium in malaria. Tropical Dis Bull. 1952;49:9.
19. Hens N, Aerts M, Faes C, Shkedy Z, Lejeune O, Van Damme P, et al. Seventy-five years of estimating the force of infection from current status data. Epidemiol Infect. 2010;138:6.
20. Hay JA, Routledge I, Takahashi S. Serodynamics: a primer and synthetic review of methods for epidemiological inference using serological data. Epidemics. 2024;49: 100806.
21. Corran P, Coleman P, Riley E, Drakeley C. Serology: a robust indicator of malaria transmission intensity? Trends Parasitol. 2007;23:12.
22. Kilama M, Smith DL, Hutchinson R, Kigozi R, Yeka A, Lavoy G, et al. Estimating the annual entomological inoculation rate for *Plasmodium falciparum* transmitted by *Anopheles gambiae s.l.* using three sampling methods in three sites in Uganda. Malar J. 2014;13:111.
23. Smith T, Maire N, Dietz K, Killeen GF, Vounatsou P, Molineaux L, et al. Relationship between the entomologic inoculation rate and the force of infection for Plasmodium falciparum malaria. Am J Trop Med Hyg. 2006;75(Suppl 2):11–8.
24. Keeling M, Rohani P. Modeling Infectious Diseases in Humans and Animals. 1st ed. New Jersey: Princeton University Press; 2008.
25. Vynnycky E, White R. An introduction to infectious disease modelling. 1st ed. United Kingdom: Oxford University Press; 2010.

26. World Health Organization. WHO Guidelines for Malaria, 3 June 2022. Geneva: World Health Organization; 2022. https://www.who.int/publications/i/item/guidelines-for-malaria.
27. Kretzschmar M, Wallinga J. Mathematical Models in Infectious Disease Epidemiology. In: Krämer A, Kretzschmar M, Krickeberg K, editors. Modern Infectious Disease Epidemiology. New York: Springer; 2009. p. 209–21.
28. Keeling M, Danon L. Mathematical modelling of infectious diseases. Br Med Bull. 2009;92:1.
29. Capasso V. Mathematical structures of epidemic systems. 1st ed. Germany: Springer Science & Business Media; 2008.
30. Hanski IA, Gaggiotti OE. Ecology, genetics and evolution of metapopulations. 1st ed. New York: Academic Press; 2004.
31. Railsback SF, Grimm V. Agent-based and individual-based modeling: a practical introduction. 1st ed. New Jersey: Princeton University Press; 2019.
32. Macdonald G. The Epidemiology and Control of Malaria. 1st ed. United Kingdom: Oxford University Press; 1957.
33. Keiding N. Statistical inference in the Lexis diagram. Philos Trans A Math Phys Eng Sci. 1990;332:1627.
34. Stadler E, Cromer D, Mehra S, Adekunle AI, Flegg JA, Anstey NM, et al. Population heterogeneity in Plasmodium vivax relapse risk. PLoS Negl Trop Dis. 2022;16:12.
35. Gosling RD, Ghani AC, Deen JL, Von Seidlein L, Greenwood BM, Chandramohan D. Can changes in malaria transmission intensity explain prolonged protection and contribute to high protective efficacy of intermittent preventive treatment for malaria in infants? Malar J. 2008;7:54.
36. Águas R, White LJ, Snow RW, Gomes MGM. Prospects for malaria eradication in sub-Saharan Africa. PLoS ONE. 2008;3:3.
37. White LJ, Maude RJ, Pongtavornpinyo W, Saralamba S, Aguas R, Van Effelterre T, et al. The role of simple mathematical models in malaria elimination strategy design. Malar J. 2009;8:212.
38. Corder RM, Ferreira MU, Gomes MGM. Modelling the epidemiology of residual Plasmodium vivax malaria in a heterogeneous host population: A case study in the Amazon Basin. PLoS Comput Biol. 2020;16:3.
39. Stanley CC, Kazembe LN, Mukaka M, Otwombe KN, Buchwald AG, Hudgens MG, et al. Systematic review of analytical methods applied to longitudinal studies of malaria. Malar J. 2019;18:254.
40. Fitzmaurice G, Davidian M, Verbeke G, Molenberghs G. Longitudinal data analysis. 1st ed. CRC Press; 2008.
41. Wienke A. Frailty models in survival analysis. 1st ed. Chapman and Hall/CRC; 2010.
42. Andersen PK, Gill RD. Cox's regression model for counting processes: a large sample study. Ann Stat. 1982;10:4.
43. Prentice RL, Williams BJ, Peterson AV. On the regression analysis of multivariate failure time data. Biometrika. 1981;68:2.
44. Seyoum D, Kifle YG, Rondeau V, Yewhalaw D, Duchateau L, Rosas-Aguirre A, et al. Identification of different malaria patterns due to Plasmodium falciparum and Plasmodium vivax in Ethiopian children: a prospective cohort study. Malar J. 2016;15:208.
45. Cairns M, Cheung YB, Xu Y, Asante KP, Owusu-Agyei S, Diallo D, et al. Analysis of preventive interventions for malaria: exploring partial and complete protection and total and primary intervention effects. Am J Epidemiol. 2015;181:12.
46. Rodriguez-Barraquer I, Arinaitwe E, Jagannathan P, Boyle MJ, Tappero J, Muhindo M, et al. Quantifying heterogeneous malaria exposure and clinical protection in a cohort of Ugandan children. J Infect Dis. 2016;214:7.
47. Ramjith J, Andolina C, Bousema T, Jonker MA. Flexible time-to-event models for double-interval-censored infectious disease data with clearance of the infection as a competing risk. Front Appl Math Stat. 2022;8:1035393.
48. Patson N, Mukaka M, Kazembe L, Eijkemans MJ, Mathanga D, Laufer MK, et al. Comparison of statistical methods for the analysis of recurrent adverse events in the presence of non-proportional hazards and unobserved heterogeneity: a simulation study. BMC Med Res Methodol. 2022;22:1.
49. Sagara I, Giorgi R, Doumbo OK, Piarroux R, Gaudart J. Modelling recurrent events: comparison of statistical models with continuous and discontinuous risk intervals on recurrent malaria episodes data. Malar J. 2014;13:293.
50. Muench H. Derivation of rates from summation data by the catalytic curve. J Am Stat Assoc. 1934;29:185.
51. Griffiths D. A catalytic model of infection for measles. J R Stat Soc C. 1974;23:3.
52. Grenfell BT, Anderson R. The estimation of age-related rates of infection from case notifications and serological data. Epidemiol Infect. 1985;95:2.
53. Farrington CP. Modelling forces of infection for measles, mumps and rubella. Stat Med. 1990;9:8.
54. Becker NG. Analysis of infectious disease data. 1st ed. Chapman and Hall/CRC; 1989.
55. Keiding N. Age-specific incidence and prevalence: a statistical perspective. J R Stat Soc A. 1991;154:3.
56. Trussell J, Hankinson R, Tilton J. Demographic applications of event history analysis. 1st ed. United Kingdom: Oxford University Press; 1992.
57. Andolina C, Rek JC, Briggs J, Okoth J, Musiime A, Ramjith J, et al. Sources of persistent malaria transmission in a setting with effective malaria control in eastern Uganda: a longitudinal, observational cohort study. Lancet Infect Dis. 2021;21:11.
58. Botwe AK, Oppong FB, Gyaase S, Owusu-Agyei S, Asghar M, Asante KP, et al. Determinants of the varied profiles of Plasmodium falciparum infections among infants living in Kintampo. Ghana Malar J. 2021;20:240.
59. Addy JW, Bediako Y, Ndungu FM, Valetta JJ, Reid AJ, Mwacharo J, et al. 10-year longitudinal study of malaria in children: Insights into acquisition and maintenance of naturally acquired immunity. Wellcome Open Res. 2021;6:79.
60. Wanzira H, Kakuru A, Arinaitwe E, Bigira V, Muhindo MK, Conrad M, et al. Longitudinal outcomes in a cohort of Ugandan children randomized to artemether-lumefantrine versus dihydroartemisinin-piperaquine for the treatment of malaria. Clin Infect Dis. 2014;59:4.
61. Eisele TP, Bennett A, Silumbe K, Finn TP, Porter TR, Chalwe V, et al. Impact of four rounds of mass drug administration with dihydroartemisinin-piperaquine implemented in southern province. Zambia Am J Trop Med Hyg. 2020;103(Suppl 2):7–18.
62. Bennett A, Porter TR, Mwenda MC, Yukich JO, Finn TP, Lungu C, et al. A longitudinal cohort to monitor malaria infection incidence during mass drug administration in Southern Province. Zambia Am J Trop Med Hyg. 2020;103(Suppl 2):54–65.
63. Trape JF, Tall A, Sokhna C, Ly AB, Diagne N, Ndiath O, et al. The rise and fall of malaria in a west African rural community, Dielmo, Senegal, from 1990 to 2012: a 22 year longitudinal study. Lancet Infect Dis. 2014;14:6.
64. Simmons RA, Mboera L, Miranda ML, Morris A, Stresman G, Turner EL, et al. A longitudinal cohort study of malaria exposure and changing serostatus in a malaria endemic area of rural Tanzania. Malar J. 2017;16:309.
65. Wood SN. Generalized additive models: an introduction with R. 1st ed. chapman and hall/CRC; 2006.
66. Ross R. The Prevention of Malaria. 2nd ed. United Kingdom: John Murray; 1911.
67. Amoah B, McCann RS, Kabaghe AN, Mburu M, Chipeta MG, Moraga P, et al. Identifying Plasmodium falciparum transmission patterns through parasite prevalence and entomological inoculation rate. Elife. 2021;10:e65682.
68. Beier JC, Oster CN, Onyango FK, Bales JD, Sherwood JA, Perkins PV, et al. Plasmodium falciparum incidence relative to entomologic inoculation rates at a site proposed for testing malaria vaccines in western Kenya. Am J Trop Med Hyg. 1994;50:5.
69. Beier JC, Killeen GF, Githure JI. Entomologic inoculation rates and Plasmodium falciparum malaria prevalence in Africa. Am J Trop Med Hyg. 1999;61:1.
70. Smith D, Dushoff J, Snow R, Hay S. The entomological inoculation rate and Plasmodium falciparum infection in African children. Nature. 2005;438:7067.
71. Davey T, Gordon R. The estimation of the density of infective anophelines as a method of calculating the relative risk of inoculation with malaria from different species or in different localities. Ann Trop Medn Parasitol. 1933;27:1.
72. Najera J. A critical review of the field application of a mathematical model of malaria eradication. Bull World Health Organ. 1974;50:5.
73. Filipe JAN, Riley EM, Drakeley CJ, Sutherland CJ, Ghani AC. Determination of the processes driving the acquisition of immunity to malaria using a mathematical transmission model. PLoS Comput Biol. 2007;3:12.
74. Akaike H. A new look at the statistical model identification. IEEE Trans Autom Control. 1974;19:6.

75. Muench H. Catalytic models in epidemiology. 1st ed. Harvard University Press; 2013.
76. Macdonald G. The analysis of malaria parasite rates in infants. Trop Dis Bull. 1950;47:10.
77. Klein JP, Moeschberger ML. Survival analysis: techniques for censored and truncated data. 1st ed. Springer; 2003.
78. Pull J, Grab B. A simple epidemiological model for evaluating the malaria inoculation rate and the risk of infection in infants. Bull World Health Organ. 1974;51:5.
79. Bekessy A, Molineaux L, Storey J. Estimation of incidence and recovery rates of Plasmodium falciparum parasitaemia from longitudinal data. Bull World Health Organ. 1976;54:6.
80. Pawitan Y. In All Likelihood: Statistical Modelling and Inference Using Likelihood. 1st ed. Oxford Science Publications; 2001.
81. Yukich J, Briët O, Bretscher MT, Bennett A, Lemma S, Berhane Y, et al. Estimating Plasmodium falciparum transmission rates in low-endemic settings using a combination of community prevalence and health facility data. PLoS ONE. 2012;7:8.
82. Gething PW, Patil AP, Smith DL, Guerra CA, Elyazar IR, Johnston GL, et al. A new world malaria map: Plasmodium falciparum endemicity in 2010. Malar J. 2011;10:1.
83. Collett D. Modelling survival data in medical research. 4th ed. Chapman and Hall/CRC; 2023.
84. Mugenyi L, Abrams S, Hens N. Estimating age-time-dependent malaria force of infection accounting for unobserved heterogeneity. Epidemiol Infect. 2017;145:12.
85. Held L, Hens N, D O'Neill P, Wallinga J. Handbook of infectious disease data analysis. 1st ed. CRC Press; 2019.
86. Coutinho FAB, Massad E, Lopez LF, Burattini MN, Struchiner CJ, Azevedo-Neto RSD. Modelling heterogeneities in individual frailties in epidemic models. Math Comput Model. 1999;30:1.
87. Hinde S, Spackman E. Bidirectional citation searching to completion: an exploration of literature searching methods. Pharmacoeconomics. 2015;33:5–11.
88. Macdonald G. Malaria transmission rates and infant parasite rates. Geneva: World Health Organization; 1950. WHO Expert Committee on Malaria, Malaria Conference in Equatorial Africa, Kampala, 1950. https://iris.who.int/handle/10665/64168.
89. Stuckey EM, Smith TA, Chitnis N. Estimating malaria transmission through mathematical models. Trends Parasitol. 2013;29:10.
90. Maude RJ, Pontavornpinyo W, Saralamba S, Aguas R, Yeung S, Dondorp AM, et al. The last man standing is the most resistant: eliminating artemisinin-resistant malaria in Cambodia. Malar J. 2009;8:31.
91. Ogunjimi B, Hens N, Goeyvaerts N, Aerts M, Van Damme P, Beutels P. Using empirical social contact data to model person to person infectious disease transmission: an illustration for varicella. Math Biosci. 2009;218:2.
92. Bosomprah S. A mathematical model of seropositivity to malaria antigen, allowing seropositivity to be prolonged by exposure. Malar J. 2014;13:1.
93. Yman V, White MT, Rono J, Arcà B, Osier FH, Troye-Blomberg M, et al. Antibody acquisition models: a new tool for serological surveillance of malaria transmission intensity. Sci Rep. 2016;6:19472.
94. Kyomuhangi I, Giorgi E. A unified and flexible modelling framework for the analysis of malaria serology data. Epidemiol Infect. 2021;149: e99.

Publisher's Note