Model development, data, and public health: A combined approach against malaria

Inauguraldissertation
zur
Erlangung der Würde eines Doktors der Philosophie

vorgelegt der
Philosophisch-Naturwissenschaftlichen Fakultät
der Universität Basel

von
Theresa Reiker

Basel, 2023

Genehmigt von der Philosophisch-Naturwissenschaftlichen Fakultät
auf Antrag von

Prof. Dr. Melissa A. Penny, Prof. Dr. Thomas A. Smith and Prof. Dr. Deirdre Hollingsworth

Basel, den 27. April 2021

Prof. Dr. Marcel Mayor
Dekan

"Lumos."

Harry Potter

UNIVERSITY OF BASEL

Abstract

Philosophisch-naturwissenschaftliche Fakultät

Doctor of Philosophy

Model development, data, and public health: A combined approach against malaria

by Theresa REIKER

The health burden of infectious diseases remains substantial. Within ever-changing public health landscapes, characterised by complex disease biologies and limited operational resources, appropriate control strategies are often difficult to identify. In recent years, there has been an explosive increase in the popularity of mathematical modelling to bridge evidence gaps and support (policy) decision-making. To ensure accurate predictions, e.g. on the public health impact of new interventions, models must be grounded in plausible assumptions and calibrated to diverse data on multiple epidemiological and biological relationships.

Through a comprehensive investigation of the modelling process from development to application, my research aims to prompt discussions about the role of infectious disease modelling in decision-making and about opportunities for modernisation. With application to malaria modelling, I present methodological advancements, structural analyses and discussions, and application case studies. This includes the development of a novel, machine learning-based calibration approach that outperforms previous methods. A generalisable framework for incorporating calibration data while accounting for contextual covariates is developed and applied to a database of *Pf*PR-incidence records. I subsequently discuss the link between model calibration decisions and the model's later uses in simulating epidemiological relationships. Taking the leap from model development to application, I assess the use of surveillance-response interventions for malaria elimination, addressing the various technical challenges of quantifying elimination itself. Finally, I shift perspective towards the potentials and pitfalls of using modelling to support decision-making.

The research presented in this thesis contributes to keeping malaria modelling up-to-date with computational methods and global health developments. Many of the principles presented here encompass general discussions of infectious disease modelling, and aim to encourage conversations about the place of modelling at the public health decision-making table.

Acknowledgements

Over the years, I have been lucky to have been supported, guided, and pushed by so many in so many different ways: Teachers, mentors, friends at home in Germany (though *home* has by now become a rather abstract concept), in the U.S., in Oxford and Cambridge and finally here, in Basel. You are so many who have inspired me, it would be nearly impossible to thank you all individually, and I would most certainly forget someone. Know, that I am truly grateful to every single one of you. You have been an invaluable part of this journey and it would have impossible without you, both on a professional, and personal level.

First and foremost, I would like to thank Thomas Smith and Melissa Penny for their supervision and guidance over the past years. Thank you, Tom, for granting me the opportunity of undertaking this PhD adventure. I feel privileged to have worked on your team. Thank you for your guidance and wisdom, in the fields of malaria, statistics, and most things really. Thank you, Melissa, for your curiosity and excitability for novelty, for your drive and scientific rigour. While you, Tom grounded me, you, Melissa pushed me far beyond what I deemed reachable. Together, you guided and shaped my PhD experience and I am truly thankful for that.

To my colleagues and collaborators inside and outside of Swiss TPH. Thank you for all the valuable discussions and criticism. Through you, I have been exposed to such vast amounts of expertise and knowledge from so many fields and angles - I hope some of it has transferred to me. My special thanks to Ewan Cameron, for your patience in explaining machine learning basics to a biologist - and for letting me butcher your code.

To my friends, especially Lydia Burgert and Fabienne Fischer, who have walked alongside me, pushed me, and sometimes dragged me forward, feet first. Thanks to you this city has become another *home* to me. You have been with me throughout ups and downs, have seen my energetic, best self, and the self that was sometimes close to breaking. It's easy to share the highs, but you were there through the lows. You always told me what I needed to hear, not only what I wanted to hear, and that is what makes true friendship. We have shared countless wonderful memories: The climbing, the hiking, running, and strength trainings (that have since turned absurdly long), the beer, the wine, the dinners. And all the fun we had together. These beautiful moments - you, my Basel family - are what kept me going. I will never forget when you said „and now, we'll finish that PhD of yours". I couldn't thank you enough for everything. And thanks to you, I now have a cat, Aya - who I should also thank. For the daily reality checks and reminders of the importance of fresh air. When I thought my thesis was insane, you showed me what true insanity looked like.

Also many thanks to Christos Kokaliaris, for the long walks and countless engaging discussions about philosophy, statistics, and life. Thank you for being a loyal (fun, totally not boring)

friend, who was always there for me. And to Iva Jugovic. We have shared such great moments over the years in Cambridge, in Basel, and of course, in Greece. I am truly blessed to have you in my life.

Für Mama und Papa. Danke, dass ihr mich von klein auf in meiner Neugier unterstützt habt. Dass ihr bei (fast) jeder verrückten Idee gesagt habt, „Lass das Kind mal machen“. Ich durfte so viel erkunden, hinterfragen, ausprobieren. So viel mit Ankündigung selbst auf die Nase fallen und so viel mit euch wieder aufstehen. Geht nicht, gab es nicht. Was möglich war, war machbar. Ihr habt mir so viel Freiheit geschenkt, auch wenn ihr so früh loslassen musstet. Ohne euch wäre ich heute nicht wo ich bin. Danke.

Contents

List of Figures

List of Tables

List of Abbreviations

ACD	**A**ctive **C**ase **D**etection
ACT	**A**rtemisinin **C**ombination **T**herapy
ABM	**A**gent-**B**ased **M**odel
AQ	**A**modia**q**uine
BO	**B**ayesian **O**ptimisation
BOINC	**B**erkeley **O**pen **I**nfrastructure for **N**etwork **C**omputing
BMGF	**B**ill and **M**elinda **G**ates **F**oundation
CI	**C**onfidence **I**nterval
CPU	**C**entral **P**rocessing **U**nit
CQ	**C**hloro**q**uine
DE	**D**ifference **E**quation
DHS	**D**emographic and **H**ealth **S**urvey
$\mathbf{E_{14}}$	Probability of **E**ffective Treatment of any Clinical Case within **14** days
EIR	**E**ntomological **I**nocculation **R**ate
GA	**G**enetic **A**lgorithm
GA-O	**G**enetic **A**lgorithm **O**ptimisation algorithm
GP	**G**aussian **P**rocess
GP-BO	**G**aussian **P**rocess-based **B**ayesian **O**ptimisation
GPSG	**G**aussian **P**rocess **S**tacked **G**eneralisation
GPSG-BO	**G**aussian **P**rocess **S**tacked **G**eneralisation-based **B**ayesian **O**ptimisation
GPU	**G**raphics **P**rocessing **U**nit
HIV	**H**uman **I**mmunodeficiency **V**irus
HPC	**H**igh **P**erformance **C**omputing
IBM	**I**ndividual-**B**ased **M**odel
IPTp	**I**ntermittent **P**reventive **T**reatment in **p**regnancy
IRS	**I**ndoor **R**esidual **S**praying
ITN	**I**nsecticide **T**reated **N**ets
LHS	**L**atin **H**ypercube **S**ampling
LLIN	**L**ong-**L**asting **I**nsecticidal **N**ets
MARS	**M**ultivariate **A**daptive **R**egression **S**plines
MDA	**M**ass **D**rug **A**dministration
MIS	**M**alaria **I**ndicator **S**urvey
ML	**M**achine **L**earning
MTIMBA	**M**alaria **T**ransmission **I**ntensity and **M**ortality **B**urden across **A**frica
NN	**N**eural **N**et
ODE	**O**rdinary **D**ifferential **E**quation
OR	**O**dds **R**atio
P.	*Plasmodium*
PAF	**P**population **Attributable** **F**raction
PCD	**P**assive **C**ase **D**etection
***Pf*PR**	***Plasmodium falciparum*** **P**arasite **R**ate
QN	**Q**uin**i**ne

RBM	**R**oll **B**ack **M**alaria
RCD	**R**eactive **C**ase **D**etection
RCT	**R**andomised **C**ontrol **T**rial
RDT	**R**apid **D**iagnostic **T**est
RF	**R**andom **F**orest
RR	**R**elative **R**isk
RSS	**R**esidual **S**um of **S**quares
SI(R)	**S**usceptible **I**nfected (**R**ecovered)
SG	**S**tacked **G**eneralisation
SMC	**S**easonal **M**alaria **C**hemoprevention
SP	**S**ulfadoxine-**P**yrimethamine
SVD	**S**ingular **V**alue **D**ecomposition
SVM	**S**upport **V**ector **M**achine
SVR	**S**upport **V**ector **R**egression
TPP	**T**arget **P**roduct **P**rofile
WHA	**W**orld **H**ealth **A**ssembly
WHO	**W**orld **H**ealth **O**rganization

Für Mama und Papa…

1 Introduction

1.1 The role of mathematical modelling in guiding public health decision-making

Despite global efforts, infectious diseases still pose a great burden on health worldwide (*1*). The as of 2021 still ongoing Covid-19 pandemic has made evident that even the most privileged societies are not immune to the threat posed by infectious diseases. Nor are they well enough prepared to combat them. The various, ever-changing control strategies employed by different countries in the early times of the pandemic illustrate the challenges of public health decision-making: With lives at stake, the public pressure on decision-makers is high, opportunities and implementation strategies are manifold, and a clear methodology for the translation of scientific evidence into actionable policy is often lacking. Public health decisions are expected to consider a variety of factors, to optimise the total public health impact, but also to account for economic, societal, and political consequences.

This pandemic has reminded many industrial countries of the havoc caused by infectious diseases. Yet, many infectious diseases, like HIV, tuberculosis, and malaria, pose a vast burden to many societies at all times. The last decades have shown that the successful implementation of health interventions for prevention and treatment can greatly reduce global morbidity and mortality. To ensure that limited resources available for control and elimination are appropriately invested, decision-making in health should be objective, transparent, and built on a foundation of scientific evidence. However, direct evidence from (expensive, time-consuming) randomised control trials (RCTs) and observational studies commonly summarises effect estimates of individual interventions in specific study populations (*2*). The resulting sparse, contextual data often provides insufficient evidence for policy decisions in a complex disease and public health landscape. This creates a gap between urgent public health needs and evidence-based policy decision-making. To weigh intervention strategies against each other, and to make informed decisions at population level, additional quantitative evidence is required.

Mathematical modelling can support bridging this knowledge and evidence gap. Detailed mathematical models of disease transmission dynamics can explicitly capture, account for, and extrapolate from setting- and time-specific contexts and explore a wide range of intervention combinations. Over the last century, mathematical models have become an invaluable tool in

understanding and analysing disease and intervention dynamics for many infectious diseases, such as for malaria (*3*–*5*), tuberculosis (*6*), influenza (*7*), dengue (*8*), and many others (*9*, *10*). Going beyond the data or statistical summaries, mechanistic mathematical models translate the biological and epidemiological processes, that characterise a disease, into mathematical functions. The parameters describing and linking these functions are calibrated by pooling available data and knowledge, covering an array of biological and epidemiological relationships, such as age-dependent prevalence or incidence patterns, or within-host pathogen dynamics. The joint calibration to diverse data under an explicit framework of how these data are (mathematically) linked aims to yield an approximation of the systems observed in the real world. Simulation experiments using well-calibrated models can generate predictions that provide additional evidence in answering biological and epidemiological questions.

Applications for well-developed infectious disease models are manifold (*2*). They include general investigations of disease dynamics, as well as responses to acute policy needs. Modelling can provide quantitative estimates on the long-term effectiveness and cost-effectiveness of an intervention (*2*), like estimating the impact and cost-effectiveness of the RTS,S/AS01 vaccine against malaria (*11*), or provide estimates on the comparative effectiveness of different existing or planned interventions (*2*). Such estimates can be used to draft national disease control strategies, such as for malaria in Tanzania (*12*), or develop target product profiles for new interventions against existing ones, such as defining requirements for antimalarial long-acting injectables versus existing seasonal chemoprevention (Burgert et al. 2021: in preparation, see Chapter 7). Modelling can further provide rapid early guidance for outbreak-response, such as assessing the epidemic potential of an outbreak, or aiding the development of emergency guidelines. Recent examples include assessing the benefit of travel restrictions during the Ebola outbreak 2014–2016 (*13*) and investigations surrounding the 2020 Covid-19 pandemic (*14*–*17*). The latter in particular illustrates the acute impact of modelling in guiding and shaping public health responses in multiple countries (*17*): The strategy devised by the United Kingdom had originally strived for herd immunity (early March 2020) (*18*). However, this was reversed after modelling work by Ferguson et al. showed the dramatic health consequences and rapid speed at which hospital capacities would be reached if the disease spread in an uncontrolled manner (*17*, *19*).

Modelling for public health impact is an ambitious endeavour. The development of models requires understanding a vast array of scientific disciplines, spanning biology, epidemiology, sociology, mathematics, and computer science. From the scientific biological and epidemiological principles that define a model's assumptions and must be translated into mathematical expressions, to diverse calibration data, to the development of calibration algorithms, and ultimately to conducting simulation experiments for public health impact, a holistic approach is needed. This requires collaboration between many stakeholders invested in the improvement of public health, including but not limited to: researchers from different disciplines, industry and product developers, general guideline developers, such as the World Health

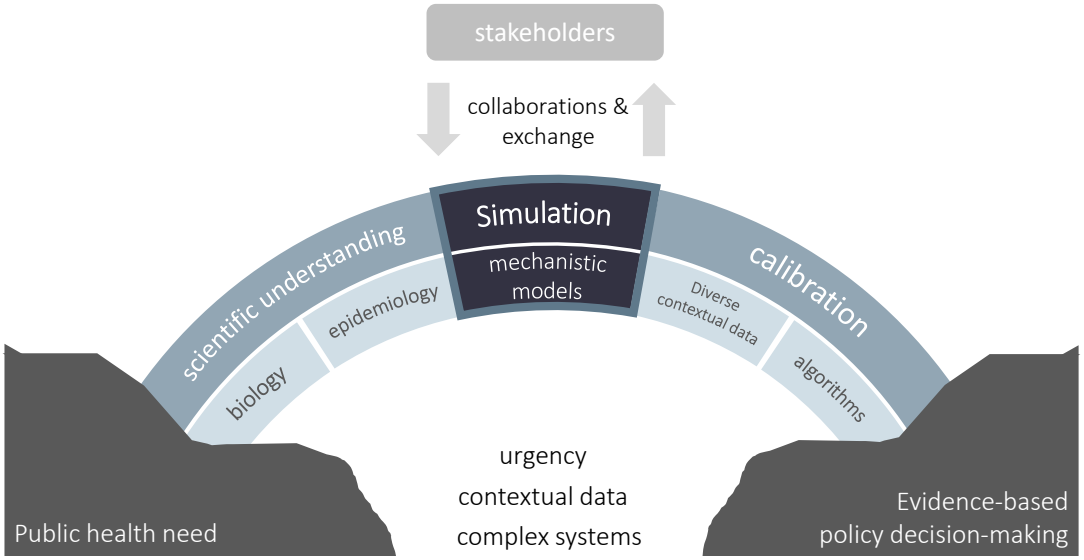

FIGURE 1.1: Mechanistic mathematical modelling can bridge the (knowledge) gap from public health needs towards evidence-based decision-making.
Mathematical models can support bridging evidence gaps in decision-making by mechanistically approximating the real world. They can address the urgency required in a public health context, extrapolate from contextual data, and capture complex systems and dependencies between different disease aspects. Mechanistic models build on a holistic scientific understanding of disease-related processes, which are translated into mathematical functions and provide a general framework. The associated parameters require calibration to diverse available data, such that the resulting model accurately captures the complex processes that characterise the disease. In continued exchange and collaboration between external stakeholders (e.g. funding bodies, regulators, and international and national policy organisations) and the research community, simulation experiments can generate supporting evidence for decision-making in health.

Organization (WHO) that decide whether and how modelling results should be incorporated in the guidelines on the use of new interventions for infectious diseases (*2*), funding bodies (such as research councils and foundations), national and international organisations, and governments on the implementation side. The basic principles of modelling for public health impact are summarised in Figure 1.1.

With application to malaria, this thesis takes a holistic approach to mathematical modelling for infectious diseases, spanning the data and algorithmic foundations of model calibration through to its public health applications. For malaria in particular, mathematical modelling has a long history of being a cornerstone for shaping policy, including (at times) for example the implementation of the Global Technical Strategy at global and national level (*20*). The parasite's complex life cycle involving multiple hosts, each with their own associated biologies, behaviours, and ecologies, yields a vast array of situations and possible intervention strategies that are impossible to evaluate through field experiments alone (*20*). In this chapter, section 1.2 provides an overview of the disease, its biology and epidemiology, while section 1.3 provides

context for the long-standing use of modelling in the field of malaria. This is followed by an introduction into the worlds of mathematical modelling and machine learning in sections 1.3 and 1.4, before combining these to set a joint stage for the research presented in this thesis in section 1.5.

1.2 A brief introduction to malaria epidemiology and control

Malaria remains one of the greatest challenges in global health to date with nearly half of the world's population at risk (*21*) and 228 million cases and 405 000 deaths reported in 2019 (*22*). The most vulnerable to severe outcomes are children under five years, accounting for 67% of global malaria-associated deaths [WHO World malaria report 2019]. Malaria is caused by the apicomplexan *Plasmodium* (*P.*) parasite and transmitted through the bites of infected *Anopheles* mosquitoes. In humans, approximately 80% of cases and 90% of all deaths are attributable to *P. falciparum*, with the other human-relevant species being *P. malariae*, *P. knowlensi*, *P. ovale* or *P. vivax* (*23*). As the most important human *Plasmodium* parasite for human malaria-attributable morbidity and mortality, this thesis is in reference to *P. falciparum* malaria, unless stated otherwise.

1.2.1 *Plasmodium falciparum* life cycle in the human and mosquito

As is characteristic of the majority of Apicomplexa, *Plasmodium* parasites have a complex life cycle involving both asexual and sexual reproduction in the human and the mosquito. Figure 1.2 briefly outlines the life cycle of *Plasmodium falciparum* malaria.

1.2.2 Pathenogenesis and human immunity

Presentations of *P. falciparum* malaria in humans range from asymptomatic carriage to severe systemic disease and death (*25*). The majority of infections result in uncomplicated malaria, especially when the infection is diagnosed and treated promptly (*28*). Symptoms of uncomplicated malaria include periodic fevers and chills coinciding with parasite release into the blood stream during the erythrocytic phase of infection, and non-specific, fever-associated symptoms, such as headache and fatigue (*25*). Severe malaria on the other hand is characterised by systemic inflammatory responses and can lead to coma or death (*29*). This is most common in children under five years and often takes the form of cerebral malaria or severe anaemia (*29*).

The presence and severity of symptoms depend on the infected individual's immune status. Individuals living in highly endemic or holoendemic regions face constant re-infection (*30*), resulting in a gradual build-up of natural acquired immunity over time. Consequently, most adults living in such regions experience little to no clinical symptoms upon infection, while children under five years (with little to no immunity) carry the majority of severe disease burden. Importantly, this immunity confers protection from clinical disease but not from

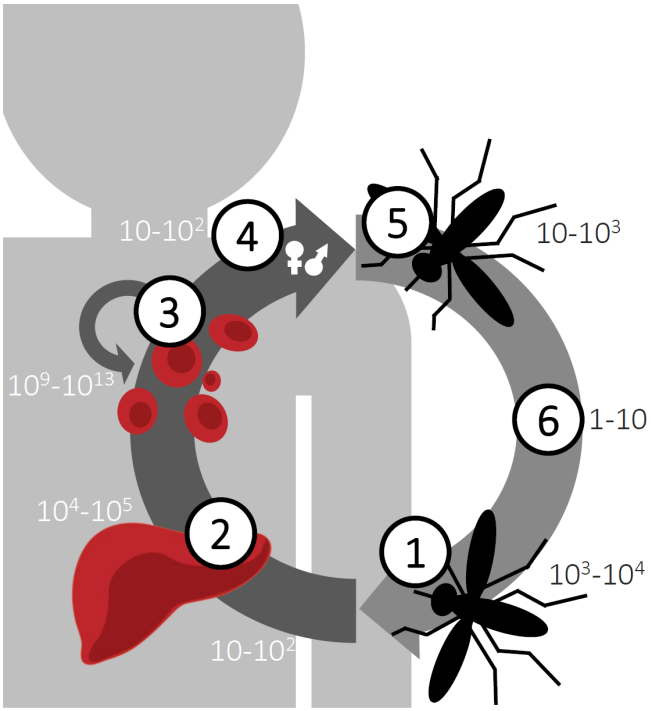

FIGURE 1.2: A schematic of the *P.* life cycle with estimated parasite numbers at each stage

More than 30 species of *Anopheles* (*An.*) mosquitoes serve as vectors to *P. falciparum*, in Africa most prominently *An. gambiae s.s.*, *An. arabiensis*, and *An. funestus*. Following the bite of an infected mosquito and injection into the human bloodstream (1), sporozoites enter hepatocytes (exo-erythrocytic or liver-stage, 2). Here, they undergo approximately 5-7 days of asexual division and growth (*24*), resulting in tens of thousands of first-generation merozoites. The resulting liver-schizonts rupture and merozoites are released into the bloodstream. The parasite subsequently penetrates red blood cells (erythrocytes) using their apical complex (a defining characteristic of the phylum), initiating the intraerythrocytic cycle (3). Merozoites divide asexually inside erythrocytes. At intervals of approximately 48 hours, the schizonts rupture, again releasing large numbers of merozoites. In symptomatic infections, this causes periodic waves of chills and fever in the infected human host, which are a defining clinical characteristic *P. falciparum* infection. If left untreated, the rate of multiplication in this stage is about a 10-fold increase every 2 days (*25*). The released next-generation merozoites infect more erythrocytes, continuing a cycle of infection, replication and release. After erythrocyte invasion, a certain number of merozoites may mature into gametocytes (4), awaiting ingestion by a mosquito during a blood meal (5). Upon fertilisation inside the mosquito gut (6), a zygote (ookinete) is formed, which subsequently infects mosquito midgut. Here, it transforms into a sporozoite-producing oocyst, which travels to the mosquito's salivary glands to again infect humans during the next blood meal, closing the cycle (1). (*26*, *27*).

infection. This poses challenges to disease control as it discourages care seeking, especially when access to treatment is poor. In the absence of curative treatment, infected individuals may still contribute to onwards transmission, threatening control and elimination efforts (*25*, *30*, *31*).

1.2.3 Malaria prevention, control and elimination

Malaria is both preventable and curable. In fact, most severe disease in children is a consequence of a lack of prevention or treatment or of treatment failure and thus avoidable (*25*). Nonetheless, malaria remains a major contributor to global mortality and morbidity and the road to control and elimination has been rocky.

In 1955, the Global Malaria Eradication Program (GMEP) was introduced by the World Health Organization (WHO) (*32*). Following early successes of the insecticide dichlorodiphenyltrichloroethane (DDT) and the discovery of chloroquine as an effective antimalarial drug, the GMEP envisioned global eradication within 10 years. However, 14 years later, the emergence of resistance against the first-line treatment, chloroquine, operational constraints, resurgence of transmission in previously eliminated regions, and financial constraints (*32*) led to a shift in strategic focus to sustained control before elimination. While today, the GMEP is largely viewed as a failure concerning its initial objectives and planned timeline, it did geographically limit the disease by eliminating malaria from most of Europe, North America, the Caribbean, and parts of Asia and Venezuela (*32*). However, by the beginning of the 1990s, the malaria-related global health situation had deteriorated due to a lack of financial support, widespread chloroquine resistance and repeated epidemic outbreaks (*32*). At this time, the world experienced an estimated 300-500 million reported cases and 1.5-2.7 million deaths annually, with 90% of cases occurring in Sub-Saharan Africa (*32*).

Following urges by the WHO and endorsement by World Health Assembly (WHA) for increased malaria control, the turn of the 21st century saw substantial and important improvements in malaria control efforts. A new, highly effective intervention was introduced: insecticide treated bednets (ITNs). In 2001, the WHO issued an official recommendation for the use of artemisinin combination therapies (ACTs) where resistance to standard antimalarials was present (*32*). ITNs, their successor long-lasting insecticide treated nets (LLINs), and the widespread use of ACTs as first-line treatment have since proven to be some of the most valuable tools against malaria (*33*). The scale up of interventions was facilitated by the foundation of The Global Fund for AIDS, tuberculosis and malaria and the U.S. President's Malaria Initiative, and their continued investments. As a result, total donor funding increased from less than 200 million US dollars per year in the early 2000s to over 1.8 billion by 2010 (*32*). In 2007, Bill and Melinda Gates put out a global call for malaria eradication, appealing for the start of *a new era* (*20*, *32*).

Overall, substantial progress in the global battle against malaria has been made since 2000. By 2010, a reduction in incidence by 17% and in mortality rates by 24% had been achieved and as of 2019, 17 of the 108 countries where malaria had been endemic in 2000 were no longer endemic. However, this progress has since stalled. With over 90% of cases, Sub-Saharan Africa carries the majority of the remaining disease burden (*22*). To reach malaria elimination,

the WHO puts particular emphasis on the importance of administering appropriate and sustainable interventions in low-transmission and recently eliminated settings, i.e. controlling residual transmission to avoid the past mistakes that led to resurgence (*34*).

However, adapting strategies from control towards elimination poses great practical challenges. Control relies predominantly on reducing morbidity and mortality through vector control and the effective treatment of clinical cases that are passively detected by the health care system (*35*). Elimination campaigns on the other hand, seek to interrupt transmission and prevent resurgence (*35*). This is achieved in two phases, the first consisting of the aggressive scale up and strict implementation of control interventions (*35*). Once very low transmission (a prevalence of around 1% or less) is achieved, additional tools are required (*35*). With decreasing transmission, increasing heterogeneity in incidence and transmission and spatiotemporal case clustering make blanket interventions cost-inefficient, increasingly ineffective and unsustainable (*35*). However, for elimination, residual transmission pockets must be detected and interrupted. This includes not only clinical cases but also asymptomatic infections that could, if left untreated, perpetuate transmission (*35*). In response, the WHO elimination guide (*36*, *37*) endorses a shift to highly targeted, locally adapted measures in the elimination phase. This includes the early detection of all cases, prevention of onward transmission from cases, management of malaria foci, and management of importation of malaria parasites.

To detect all cases, surveillance-response strategies, such as reactive case detection (RCD), are widely implemented (*35*, *38–41*). Assuming that cases are geographically clustered, surveillance-response strategies allow for targeted interventions on symptomatic cases who do not seek care as well as asymptomatic cases: When an index case of clinical malaria presents to a health facility, this triggers follow-up activities around the index case. Follow-up activities may include various parasite- or vector-focused responses including focal indoor residual spraying (IRS), bednet distribution, larval source management, presumptive treatment through mass drug administration (MDA), or screening and treatment (FSAT) (*42*). Although reactive strategies intuitively have high approval rates and have been implemented in many settings, clear guidelines on implementation and definitions of the terminology are lacking (*35*, *38–41*) and their effectiveness remains to be investigated. The quest for effective strategies for reaching elimination thereby presents a current public health need, where evidence from modelling would be valuable.

1.3 Mathematical models of infectious diseases

Especially in the field of malaria, mathematical modelling has become a popular and prominent tool in providing additional evidence for public health decision-making. Potential solutions to controlling transmission or elimination are so manifold that strategic scientific investigation

of all options by means of expensive, time-consuming clinical trials becomes impossible. The malaria parasite life cycle is complex, as are host immune responses to the parasite and the consequent pathogenesis (*20*), resulting in a vast array of interventions targeting different stages of the transmission cycle. Additionally, the suitability of (a combination of) interventions depends on the interplay of biological, ecological, epidemiological, demographic and systemic characteristics of a target setting. These include intensity, seasonal patterns and heterogeneity in transmission, vector ecology, resistance levels to antimalarial drugs and insecticides, the health care system, and the desired health goals. Mathematical modelling offers a solution to strategically evaluate and compare interventions and deployment strategies in a multitude of settings. An overview of the general structure of the modelling process from the model calibration to predictions is provided in Figure 1.3.

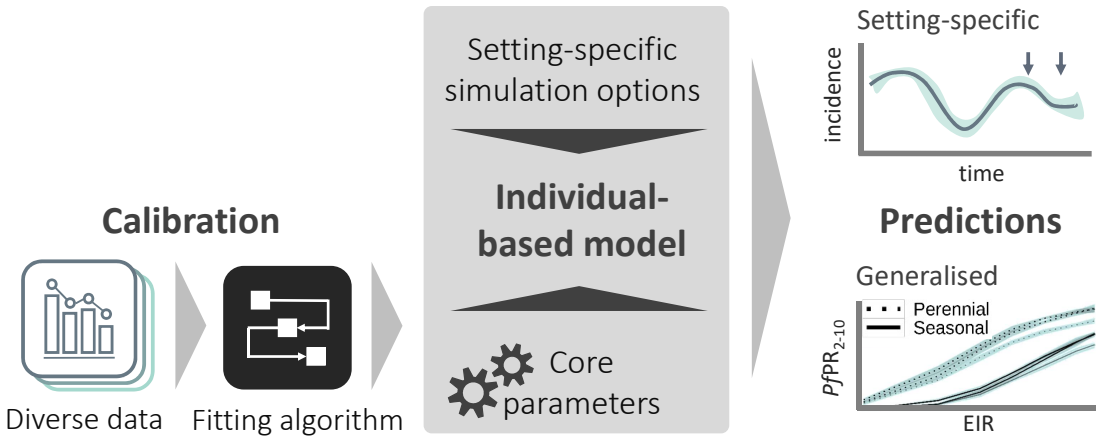

FIGURE 1.3: Overview of mathematical models: development, calibration, and application

Mathematical models are used to make predictions on disease trajectories. They possess two types of input parameters, a set of core parameters (e.g. about the natural history) and setting-specific, situational simulation options. The former determine the intrinsic behaviour of the model, i.e. the way the model captures the disease, akin to a tuned engine. The core parameters are generally fixed for all simulations of a specific disease using that model. The simulation options are determined depending on desired experiment. An optimal set of core parameters is yielded by calibration to diverse data with the aim to yield a model that best replicates some data on pre-determined relationships (objectives) of interest. Predictions from the calibrated model can be broadly grouped in two categories: Firstly, accurately capturing the natural history of the disease allows for proof-of-principle analyses and the disentanglement of general relationships. Secondly, the use of calibration data from different, explicitly describes settings enables setting-specific predictions (i.e. what happens if a specific intervention is implemented in a specific setting). With the ability to generate both generic and specific predictions, mathematical models can bridge gaps between partners and stakeholders with specific questions, e.g. on intervention strategies and evidence-based decision-making in health by extrapolating from sparse data. *Pf*$PR_{2\text{-}10}$ = *P. falciparum* parasite rate, EIR = entomological inoculation rate.

1.3.1 History of malaria modelling

The first series of mathematical models of malaria were developed in the late 19th century by Sir Ronald Ross, with the aim to understand and explain the relationship between components of the mosquito life cycle and malaria incidence in humans (*43*, *44*). In the case of malaria, the simplest compartmental models, such as the one developed by Ross, divide the population into *susceptible* and *infected* categories, where infected individuals (I) return to the *susceptible* (S) category on recovery. This results in the standard notation for such malaria models of S-I-S. The changes in the relative contribution of each category to the overall population are tracked using a system of linked differential equations. Ross' model was subsequently extended by MacDonald to explicitly include an S-I model for mosquitoes (*45*–*48*). These form the basis of compartmental population models and can be extended to include additional categories depending on the desired level of detail and the research question. This enables, for example, the capturing of additional stages of disease progression, geographic and demographic heterogeneity, different age groups, or (limited amounts of) stochasticity. The models developed by Ross and MacDonald mark an important milestone in the history of mathematical modelling for all infectious diseases as compartmental models are still widely used to date with increasing heterogeneity and complexity (*49*, *50*). However, at their core remains the fundamental concept of a population level, top-down approach to modelling, which can be insufficient to provide nuanced answers to policy questions.

1.3.2 Individual-based models

Advances in high-performance computing have enabled mathematical models of increasing sophistication and individual-based models (IBMs) that simulate individuals within populations as autonomous agents with heterogeneous characteristics and behaviours. Going beyond simpler (compartmental) models to capture stochasticity and heterogeneity in populations, disease progression, and transmission, IBMs can account for contact networks, individual care seeking behaviour, immunity effects, or within-human dynamics (*10*, *51*, *52*). This bottom-up approach to capturing population dynamics enables detailed predictions on population epidemic trajectories as well as the impact of interventions such as vaccines or new drugs (*10*, *51*). As such, IBMs provide opportunities for experimentation under relatively natural conditions without expensive clinical or population studies. Within the field of malaria, several IBMs have been developed over the last 15 years to support understanding of disease and mosquito dynamics (*53*–*55*), predict the public health impact or carry out economic analyses of (new) interventions (*11*, *33*, *56*, *57*), and investigate drug resistance (*58*). Many have had wide-reaching impact, influencing WHO policy recommendations (*59*–*62*) or strategies of national malaria control programs (*12*).

OpenMalaria

In 2006, Smith et al. published a series of papers on the development of *OpenMalaria*, a comprehensive IBM of malaria epidemiology and control (*63*). *OpenMalaria* features within-host parasite dynamics, the progression of clinical disease, development of immunity, individual care

seeking behaviour, vector dynamics and pharmaceutical and non-pharmaceutical antimalarial interventions at vector and human level (`github.com/SwissTPH/openmalaria.wiki.git`) (*52*, *63*, *64*). Details of *OpenMalaria* are provided Appendix A and in Chapters 2 and 4. In short, *OpenMalaria* captures different clinical presentations of malaria in human individuals, mosquito ecology across a range of species, and *P. falciparum* dynamics in both humans and mosquitoes. The tracking of blood stage parasite densities further enables case management actions depending on individual patient parasite densities and resulting simulated (clinical) events. Fourteen model variants for *OpenMalaria* exist (*64*) with differing assumptions on infection of humans, blood-stage parasite densities, infectiousness of humans to mosquitoes, incidence of morbidity, and mortality.

The calibration to diverse data, covering various epidemiological and biological relationships from a multitude of settings, is particularly challenging for IBMs. The tracking of individuals and inclusion of stochasticity often come at the cost of long simulation times and potentially large numbers of input parameters, making their calibration a challenge that requires elaborate algorithms. The following section outlines some of the caveats of calibration, addressing both data and algorithmic challenges.

1.3.3 Calibration data

The calibration data for IBMs often cover various epidemiological and biological relationships from a multitude of settings. *OpenMalaria* is calibrated to eleven epidemiological and biological relationships, including the age-specific prevalence and incidence patterns, mortality rates and hospitalisation rates. These relationships represent the natural history of malaria, which should be well captured in a model used for prediction of malaria transmission and intervention impact. A complete summary of the data used to calibrate *OpenMalaria* is provided in Appendix A

When calibrating to data, setting-specific simulations must be prepared. These must represent the studies that yielded the data in terms of transmission intensity, seasonal patterns, vector species, intervention history, case management, diagnostics, and data collection methods. The mirroring of field study characteristics in the simulation options aims to ensure that any deviation between simulation outputs and data can be attributed to calibrated parameters. This can prove challenging as calibration data commonly consist of historical data. In the absence of standing collaborations with researchers involved in the data collection and given the temporal distance to modern research, replicating the context of data collection can be difficult. Large amounts of data for the calibration of malaria models were supplied by the Garki project conducted by the WHO in 1969-1976 (*65*). This marked a major scientific endeavour to study the epidemiology and control of malaria in Nigeria and the first with the aim to systematically collate data to support the development of mathematical models. The main aims of this project were to quantify the impact of interventions on malaria transmission

(in particular MDA and IRS) and to collect enough data to develop and fit a mathematical model that could realistically simulate the transmission of malaria (*65*). Like the data collected during the Garki project, most data used for the calibration of *OpenMalaria* predates systematic (and successful) control efforts. As a result, most data was collected in medium to high transmission settings (with EIRs ranging from approximately 10 to nearly 600), while lower transmission and prevalence settings are underrepresented.

1.3.4 The calibration of individual-based models and the curse of dimensionality

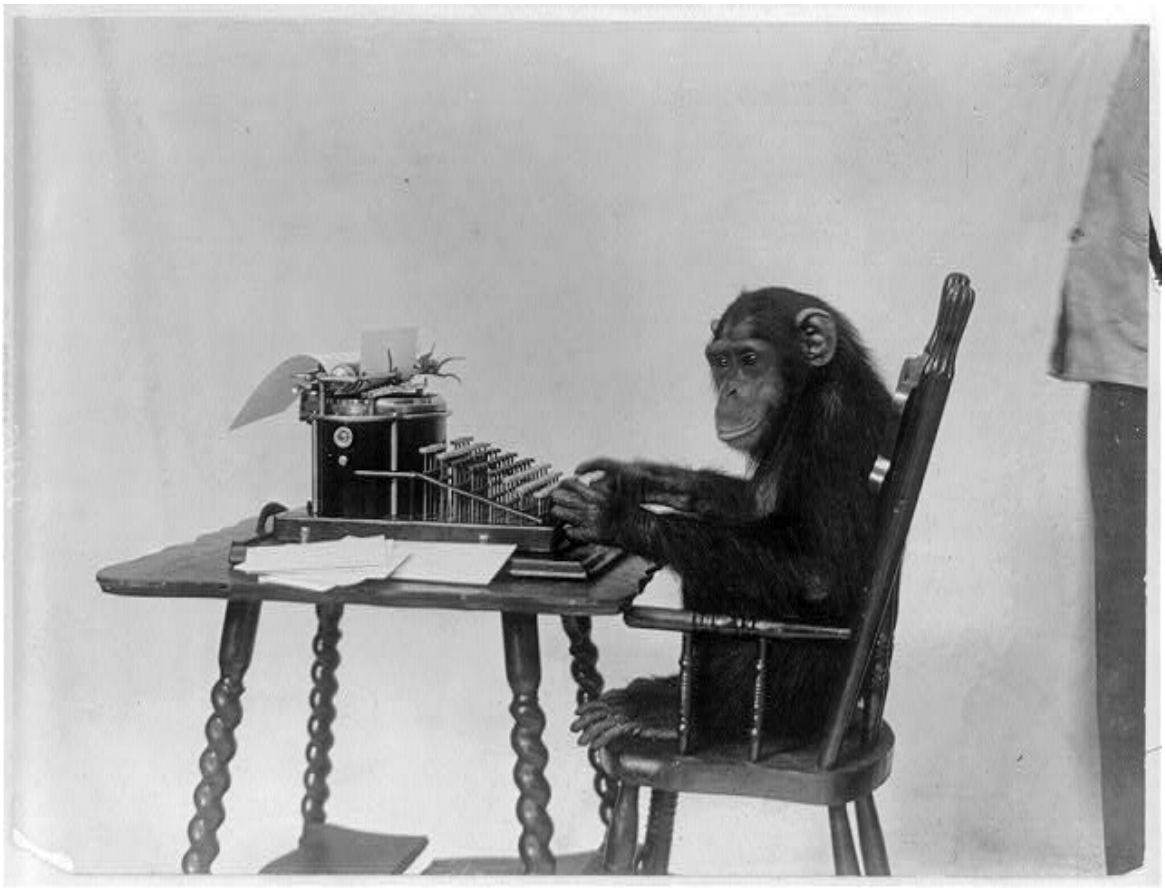

FIGURE 1.4: Chimpanzee at a typewriter
The random hitting of keys on a typewriter for an infinite amount of time will almost surely result in the complete texts of William Shakespeare. Image by the New York Zoological Society.

"To be or not to be, that is the question."

-William Shakespeare, *Hamlet*

The so-called infinite monkey theorem is a thought experiment that states that if a monkey were to randomly type at a typewriter for an infinite amount of time it would almost certainly at some point, by pure chance, type out any text, such as the complete works of William Shakespeare (Figure 1.4). This simple analogy illustrates the core challenge with the calibration

of IBMs (and other high-dimensional optimisation problems): Sequential testing of random combinations of letters (or here: numbers) by brute force would eventually yield the optimal or "*true*" solution. However, the larger the number of parameters that need to be (jointly) optimised, the larger the space, and the (exponentially) more evaluations are required. IBMs are often defined by dozens of parameters (23 for the *OpenMalaria* base model variant) and the required number of simulator evaluations to the *true* optimum would be near infinite. Additionally, IBMs often need to be calibrated with regards to multiple objectives (11 for *OpenMalaria*), which should ideally be evaluated in parallel for every parameter set to not introduce an artificial ranking of these objectives. This substantially increases the number of required evaluations (*9*). Therefore, many purely sampling-based calibration approaches that require sequential function evaluations, can be too slow for high-dimensional problems in irregular spaces, where only a limited number of function evaluations are possible.

1.4 Surrogate modelling and machine learning

Commonly faced questions in health research require sensitivity analyses (e.g. understanding the relative contribution of individual interventions to health outcomes), optimisation (e.g. finding the optimal intervention strategy for a specific setting or optimising drug properties), and risk or uncertainty quantification (to calculate e.g. the risk of elimination). As with the calibration of disease transmission models, all of these require a large number of simulations. However, each simulation using complex mechanistic (especially individual-based) models is a costly investment of computational demands and time. Therefore, conducting such analyses by simulation alone is highly resource intensive, time consuming, and often infeasible (*66*, *67*).

An innovative solution is offered by coupling expensive computer experiments to meta-models or *emulators* in a *surrogate modelling* framework (*66*). Surrogate models can roughly be divided into three categories (*68*):

1. Data-driven surrogates that approximate the input-output relationship by training on data generated by the simulator
2. Projection-based surrogates that reduce a problem's dimensionality by orthonormal vector projection (e.g. principal component analysis (PCA) or singular value decomposition (SVD))
3. Hierarchical (multifidelity) approaches that simplify the representation of the system e.g. by ignoring certain processes

In this thesis, I mainly focus on the first class of data-driven surrogates as they are directly linkable to simulation modelling and their potential in disease modelling remains largely unexplored. An emulator is a statistical or machine learning model that provides a computationally light secondary layer in the modelling process and allows for fast exploration of potential solutions. In brief, the surrogate is trained on samples of the simulator's input-output relationship to capture and generate fast empirical approximations of this relationship

without considering its inner workings. Examples of emulators differ in complexity and include polynomials (*69*), Gaussian processes (GPs) (*70*, *71*), support vector machines (SVMs) (*72*), and neural networks (NNs) (*73*). The trained emulator is generally much faster than the underlying model and can make rapid approximate predictions on likely simulation outcomes. For *OpenMalaria*, the comparison in computation speed can be within seconds for the emulator compared to 20 minutes or more for a single simulation (depending on the emulator and experimental setup). This makes expensive analyses affordable. By providing a faster way of assessing large areas of parameter space, joint simulation and emulation can be used to tackle research problems of higher complexity than traditional modelling in a time-efficient manner while retaining accuracy. For example, while optimisation using traditional modelling relies on either choosing the best of a fixed set of simulated options that can only be expanded by running additional time-consuming simulations, emulation can be used to augment sparse simulation results and to interpolate between them. Additionally, machine learning-augmented simulation modelling can be used for data mining and to provide deep insights into (non-linear) relationships between variables, for instance through (conditional) sensitivity analyses (*66*).

For additional context, the text in Box 1 provides a brief overview of the surrogate modelling process and common emulator functions.

BOX 1: Machine learning algorithms and Gaussian processes

The basic principle of surrogate modelling is to generate a supervised machine learning problem in which an emulator is trained to understand the relationship between input variables and simulation outputs. The costly simulator is originally evaluated for a range of experimental set-ups (usually in a space filling manner, e.g. using Latin hypercube sampling). The outputs (simulator observations) are stored, generating a training set consisting of experimental covariates and the simulation outputs. The emulator is subsequently trained to capture the input-output relationship and used to make out-of-sample predictions. As the input-output relationship is often not equally complex throughout parameter space and predictive uncertainties can be identified, the training set is usually enriched through adaptively sampling additional points at which simulations are required (*66*). **Gaussian processes** (GPs) surrogates (also called *Kriging*) are a particularly popular emulator that provides high flexibility as well as uncertainty quantification. In addition to reporting predictive uncertainty, if provided with replicate data points, GPs are also able to quantify the input uncertainty, the so-called *nugget effect*. Other emulator functions include random forests, support vector machines (SVMs), or neural nets (NNs). **Random forests** (RFs) consist of an ensemble of independent decision trees (generated through bagging and feature randomness), where each tree in the forest yields a class prediction which are combined into a joint random forest prediction (e.g. by majority rules), such that the joint prediction will be more accurate than that of any (errorsome) individual tree (*74*). Initially developed for classification problems, **Support vector machines** rely on transforming of the space such that classes of the response variables are (with some error) linearly separable by a hyperplane (*75*). The method has since been extended to regression problems (*76*, *77*) and is known for good generalisation and ability to handle nonlinearities (*76*). However, while powerful, SVMs scale poorly with increasing problem sizes as they require complex and memory intense matrix operations on the input space during construction (*78–80*), which can be very time consuming depending on the size of the problem and the programming they are implemented in. **Neural nets** rely on the (layered) dense interconnection of simple computational elements or *nodes* (for example, in its simplest form, the non-linearly transformed weighted sum of N inputs) (*81*). Due to their speed, NNs are often employed where many hypotheses are investigated in parallel and high computing rates are required despite drawbacks in interpretability (*81–83*). The appropriateness of employing different algorithms will depend on the problem at hand and needs to be assessed independently and on an individual basis for every problem.

1.5 Model development, data, and public health: A combined approach against malaria

Modelling the transmission and control of infectious diseases requires a comprehensive approach to maximise the value of predictions in supporting decision-making. From the collation of calibration data to the development of novel algorithms and the prediction of intervention success, this thesis addresses opportunities to advance current methodologies in infectious disease modelling with application to *P. falciparum* malaria.

More specifically, the objectives of this thesis are to

1. **Develop new, improved algorithms for solving high-dimensional optimisation problems**. In light of the complexity of disease simulators such as *OpenMalaria*, novel calibration methods are required to ensure an accurate representation of disease biology and epidemiology.
2. **Provide a formal framework for translating observational into *in silico* studies**. Diverse data form the backbone of model calibration. Model predictions are highly dependent on a careful collation and incorporation of this data that account for the context of collection.
3. **Collate data on an additional epidemiological relationship to incorporate into the calibration of *OpenMalaria***. The complexity of the age-specific prevalence (*Pf*PR)-incidence relationship is currently not explicitly incorporated during the calibration of malaria transmission models. Preparing the records from a recently published database of age-matched *Pf*PR-incidence for incorporation into calibration paves the way to improving predictive model performance.
4. **Implement novel surveillance-response intervention options in *OpenMalaria***. The development of RCD marks the first implementation of a surveillance-response intervention in *OpenMalaria*. This represents a much needed addition to *OpenMalaria* that enables the evaluation of new interventions required for malaria elimination.
5. **Discuss options on modelling disease elimination using experimental design strategies and machine learning emulators**. An analysis of the feasibility of reaching elimination in different transmission settings using only RCD provides insight to its isolated potential for reaching elimination. Further, it investigates different resource prioritisation, as many variations of intervention deployment strategies are implemented globally.

Chapters 2-4 focus on the calibration process from the development of improved calibration algorithms to data curation and a discussion of the implications of calibration decisions on predictions:

In Chapter 2, I present novel, machine learning-based approaches for model calibration. Using a Bayesian optimisation framework with different machine learning emulator functions I calibrate *OpenMalaria* simultaneously to multiple epidemiological relationships. This sets the

stage for fast and powerful solutions to high-dimensional optimisation problems, including, but not limited to the calibration of individual-based disease transmission models.

In Chapter 3, I provide a framework for the incorporation of new calibration data that accounts for potentially confounding contextual covariates. This framework is applied to studies on the age-specific prevalence-incidence relationship of *P. falciparum* malaria in Africa for input into *OpenMalaria*. I provide the database of records and contextual information required for the inclusion of this epidemiological relationship into the calibration of *OpenMalaria*.

Chapter 4 provides a link between a model's development and its later uses. Building on insights and supplementary analyses from Chapters 2 and 3, I discuss the assumptions and choices made during the calibration of *OpenMalaria* and analyse their joint implications for prediction.

In Chapter 5, I develop surveillance-response interventions in form of RCD for *OpenMalaria* and provide a first assessment of its uses in achieving elimination. Chapter 5 further marks a (chronologically) first application of emulation to malaria modelling. Here, I use machine learning models to infer a continuous probability of malaria elimination from sparse simulation results, showcasing the value of machine learning-augmented simulation modelling.

Lastly, in Chapter 6, I deepen the discussion around the work presented in this thesis. I first provide a technical outlook that showcases data requirements and the value of coupling machine learning to simulation modelling. I conclude by providing a discourse of modelling reality, outlining persisting challenges, and discussing the value of modelling in public health decision-making.

2 Emulator-based Bayesian optimisation for efficient multi-objective calibration of an individual-based model of malaria

Theresa Reiker[1,2], Monica Golumbeanu[1,2], Andrew Shattock[1,2], Lydia Burgert[1,2], Thomas A. Smith[1,2], Sarah Filippi[3], Ewan Cameron[4,5,6], Melissa A. Penny[1,2*]

[1] Swiss Tropical and Public Health Institute, Basel, Switzerland
[2] University of Basel, Petersplatz 1, Basel, Switzerland
[3] Imperial College London, UK
[4] Malaria Atlas Project, Big Data Institute, University of Oxford, Oxford, UK
[5] Curtin University, Perth, Australia
[6] Telethon Kids Institute, Perth Children's Hospital, Perth, Australia

*Corresponding author

Email: melissa.penny@unibas.ch

Publication:

Reiker T, Golumbeanu M, Shattock A, Burgert L, Smith TA, Filippi S, Cameron, M Penny MA. *Emulator-based Bayesian optimisation for efficient multi-objective calibration of an individual-based model of malaria*. Nature Communications, Volume 12, Article number 7212 (2021)

2.1 Abstract

Individual-based models have become important tools in the global battle against infectious diseases, yet model complexity can make calibration to biological and epidemiological data challenging. We propose using a Bayesian optimisation framework employing Gaussian process or machine learning emulator functions to calibrate a complex malaria transmission simulator. We demonstrate our approach by optimizing over a high-dimensional parameter space with respect to a portfolio of multiple fitting objectives built from datasets capturing the natural history of malaria transmission and disease progression. Our approach quickly outperforms previous calibrations, yielding an improved final goodness of fit. Per-objective parameter importance and sensitivity diagnostics provided by our approach offer epidemiological insights and enhance trust in predictions through greater interpretability.

2.2 Introduction

Over the last century, mathematical modeling has become an important tool to analyse and understand disease-dynamics and intervention-dynamics for many infectious diseases. Individual-based models (IBMs), where each person is simulated as an autonomous agent, are now widely used. These mathematical models capture heterogeneous characteristics and behaviors of individuals, and are often stochastic in nature. This bottom-up approach of simulating individuals and transmission events enables detailed, robust, and realistic predictions on population epidemic trajectories as well as the impact of interventions such as vaccines or new drugs (*10*, *51*). Going beyond simpler (compartmental) models to capture stochasticity and heterogeneity in populations, disease progression, and transmission, IBMs can additionally account for contact networks, individual care seeking behavior, immunity effects, or within-human dynamics (*10*, *51*, *52*). As such, well-developed IBMs provide opportunities for experimentation under relatively naturalistic conditions without expensive clinical or population studies. Prominent recent examples of the use of IBMs include assessing the benefit of travel restrictions during the Ebola outbreak 2014–2016 (*13*) and guiding the public health response to the Covid-19 pandemic in multiple countries (*17*). IBMs have also been applied to tuberculosis (*6*), influenza (*7*), dengue (*84*), and many other infectious diseases (*10*). Within the field of malaria, several IBMs have been developed over the last 15 years and have been used to support understanding disease and mosquito dynamics (*53*–*55*), predict the public health impact or carry out economic analyses of (new) interventions (*11*, *33*, *56*, *57*); and investigate drug resistance (*58*). Many have had wide-reaching impact, influencing WHO policy recommendations (*11*, *59*, *60*, *62*) or strategies of national malaria control programs (*12*).

For model predictions to be meaningful, modellers need to ensure their models accurately capture abstractions of the real world. The potential complexity and realism of IBMs often come at the cost of long simulation times and potentially large numbers of input parameters, whose exact values are often unknown. Parameters may be unknown because they represent

derived mathematical quantities that cannot be directly measured or require elaborate, costly experiments (for example shape parameters in decay functions (*64*)), because the data required to derive them in isolation is incomplete or accompanied by inherent biases, or because they interact with other parameters.

Calibrating IBMs poses a complex high-dimensional optimisation problem and thus algorithm-based calibration is required to find a parameter set that ensures realistic model behaviour, capturing the biological and epidemiological relationships of interest. Local optima may exist in the potentially highly irregular, high-dimensional goodness-of-fit surface, making iterative, purely sampling-based algorithms (e.g., Particle Swarm Optimisation or extensions of Newton–Raphson) inefficient and, in light of finite runtimes and computational resources, unlikely to find global optima. Additionally, the *curse of dimensionality* means the number of evaluations of the model scales exponentially with the number of dimensions (*85*). As an example, for the model discussed in this paper, a 23-dimensional parameter space at a sampling resolution of one sample per 10 percentile cell in each dimension, would yield $10^{\text{number of dimensions}} = 10^{23}$ cells. This is larger than number of stars in the observable Universe (of order 10^{22} (*86*)). Furthermore, most calibrations are not towards one objective or dataset. For multi-objective fitting, each parameter set requires the evaluation of multiple outputs and thus multiple simulations to ensure that all outcomes of interest are captured (in the model discussed here epidemiological outcomes such as prevalence, incidence, or mortality patterns).

In this study, we applied our approach to calibrate a well-established and used IBM of malaria dynamics called *OpenMalaria*. Malaria IBMs in particular are often highly complex (e.g., containing multiple sub-modules and many parameters), consider a two-host system influenced by seasonal dynamics, and often account for multifaceted within-host dynamics. *OpenMalaria* features within-host parasite dynamics, the progression of clinical disease, development of immunity, individual care seeking behaviour, vector dynamics and pharmaceutical and non-pharmaceutical antimalarial interventions at vector and human level (`github.com/SwissTPH/openmalaria.wiki.git`) (*52*, *63*, *64*). Previously, the model was calibrated using an asynchronous genetic algorithm (GA) to fit 23 parameters to 11 objectives representing different epidemiological outcomes, including age-specific prevalence and incidence patterns, age-specific mortality rates and hospitalisation rates (*52*, *63*, *64*) (see Supplementary Notes A.1 and A.2 for details on the calibration objectives and data). However, the sampling-based nature and sequential function evaluations of GAs can be too slow for high-dimensional problems in irregular spaces where only a limited number of function evaluations are possible and valleys of neutral or lower fitness may be difficult to cross (*87*, *88*).

Other solutions to fit similarly detailed IBMs of malaria employ a combination of directly extracting parameter values from the literature where information is available, and fitting the remainder using multi-stage, modular Bayesian Markov Chain Monte Carlo (MCMC)-based

methods (*4*, *5*, *89–92*). For these models, multiple fitting objectives are often not addressed simultaneously. Rather, to our knowledge, most other malaria IBMs are divided into functional modules (such as the human transmissibility model, within-host parasite dynamics model, and the mosquito or vector model), which are assumed to be influenced by only a limited number of parameters each. The modules are then fit independently and in a sequential manner (*4*, *5*, *90–92*). Modular approaches reduce the dimensionality of the problem, allowing for the use of relatively straightforward MCMC algorithms. However, these approaches struggle with efficiency in high dimensions as their Markovian nature requires many sequential function evaluations (10^4~10^7 even for simple models), driving up computing time and computational requirements (*93*). Additionally, whilst allowing for the generation of posterior probability distributions of the parameters (*4*), the modular nature makes sequential approaches generally unable to account for interdependencies between parameters assigned to different modules and how their co-variation may affect disease dynamics.

Progress in recent years on numerical methods for supervised, regularised learning of smooth functions from discrete training data allows us to revisit calibration of detailed mathematical models using Bayesian methods for global optimisation (*94*). Current state-of-the art calibration approaches for stochastic simulators are often based around Kennedy and O'Hagan's (KOH) approach (*71*), where a posterior distribution for the calibration parameters is derived through a two-layer Bayesian approach involving cascade of surrogates (usually Gaussian processes, GPs) (*95*). A first GP is used to model the systematic deviation between the simulator and the real process it represents, while a second GP is used to emulate the simulator (*96*). However, this approach is computationally intense when scaling to high-dimensional input spaces and multi-objective optimisation. A fully Bayesian KOH approach is likely computationally heavy (*96*) for the efficient calibration of detailed malaria simulators like *OpenMalaria*. Single-layer Bayesian optimisation with GPs on the other hand have gained popularity as an efficient approach to tackle expensive optimisation problems, for example in hyperparameter search problems in machine learning (*97*, *98*). Assuming that the parameter-solution space exhibits a modest degree of regularity, a prior distribution is defined over a computationally expensive objective function by the means of a light-weight probabilistic emulator such as a GP. The constructed emulator is sequentially refined by adaptively sampling the next training points based on acquisition functions derived from the posterior distribution. The trained emulator model is used to make predictions over the objective functions from the input space with minimum evaluation of the expensive true (simulator) function. Purely sampling-based iterative approaches (like genetic algorithms) are usually limited to drawing sparse random samples from proposals located nearby existing samples in the parameter space. In contrast, the use of predictive emulators permits exploration of the entire parameter space at higher resolution. This increases the chances of finding the true global optimum of the complex objective function in question and avoiding local optima.

Here, we use a single-layer Bayesian optimisation approach to solve the multidimensional, multi-objective calibration of *OpenMalaria* (Figure 2.1 1). Employing this single-layer Bayesian approach further allows for the direct comparison to previous calibration attempts for *OpenMalaria* as the objective functions are retained. We prove the strength and versatility of our approach by optimizing *OpenMalaria*'s 23 input parameters using real-world data on 11 epidemiological outcomes in parallel. To emulate the solution space, we explore and compare two prior distributions, namely a GP emulator and a *superlearning* algorithm in form of a GP stacked generalisation (GPSG) emulator. We first use a GP emulator to emulate the solution space. Whilst GP emulators provide flexibility whilst retaining relative simplicity (*98*) and have been used previously as priors in Bayesian optimisation (*97*), stacked generalisation algorithms have not. They provide a potentially attractive alternative as they have been shown to outperform GPs and other machine learning algorithms in capturing complex spaces (*33*, *99*). The stacked generalisation algorithm (*99*) builds on the idea of creating ensemble predictions from multiple learning algorithms (*level 0 learners*). The cross-validated predictions of the level 0 learners are incorporated into a general learning system (*level 1 meta-learner*). This allows for the combination of memory-efficient and probabilistic algorithms in order to reduce computational time, whilst retaining probabilistic elements required for adaptive sampling. Here, we showcase the efficiency and speed of the Bayesian optimisation calibration scheme and propose a modus operandi to parameterize computationally intensive or complex mathematical models that harvests recent computational developments and is scalable to high dimensions in multi-objective calibration.

2.3 Results

2.3.1 Calibration workflow

The developed model calibration workflow approach is summarised in Figure 2.1a. In brief, goodness of fit scores were first derived for randomly generated, initial parameter sets. The goodness of fit scores were defined as a weighted sum of the loss functions for each of 11 fitting objectives. These span various epidemiological measures capturing the complexity and heterogeneity of the malaria transmission dynamics, including the age–prevalence and age–incidence relationships, and are informed by a multitude of observational studies (see the "Methods" section and Supplementary Note A.2). Next, GP and GPSG emulators were trained on the obtained set of scores and used to approximate the relationship between parameter sets and goodness of fit for each objective. After initial investigation of different machine learning algorithms, the GPSG was constructed using a bilayer neural net, multivariate adaptive regression splines and random forest as *level 0* learners and a heteroscedastic Gaussian process as *level 1* learner (Figure 2.1c, d, see the "Methods" section and supplement (Appendix A)). Using a lower confidence bound acquisition function based on the emulators' point and uncertainty predictions for proposed new candidate parameter sets, the most promising sets were chosen. These parameter sets were simulated and added to the database of simulations

for the next iteration of the algorithm. At the next iteration, the emulators are re-trained on the new simulation database and re-evaluated (Figure 2.1b). This iterative process of simulation, training and emulation was repeated until a memory limit of 1024 GB was hit. Approximately $130,000$ simulations were completed up to this point.

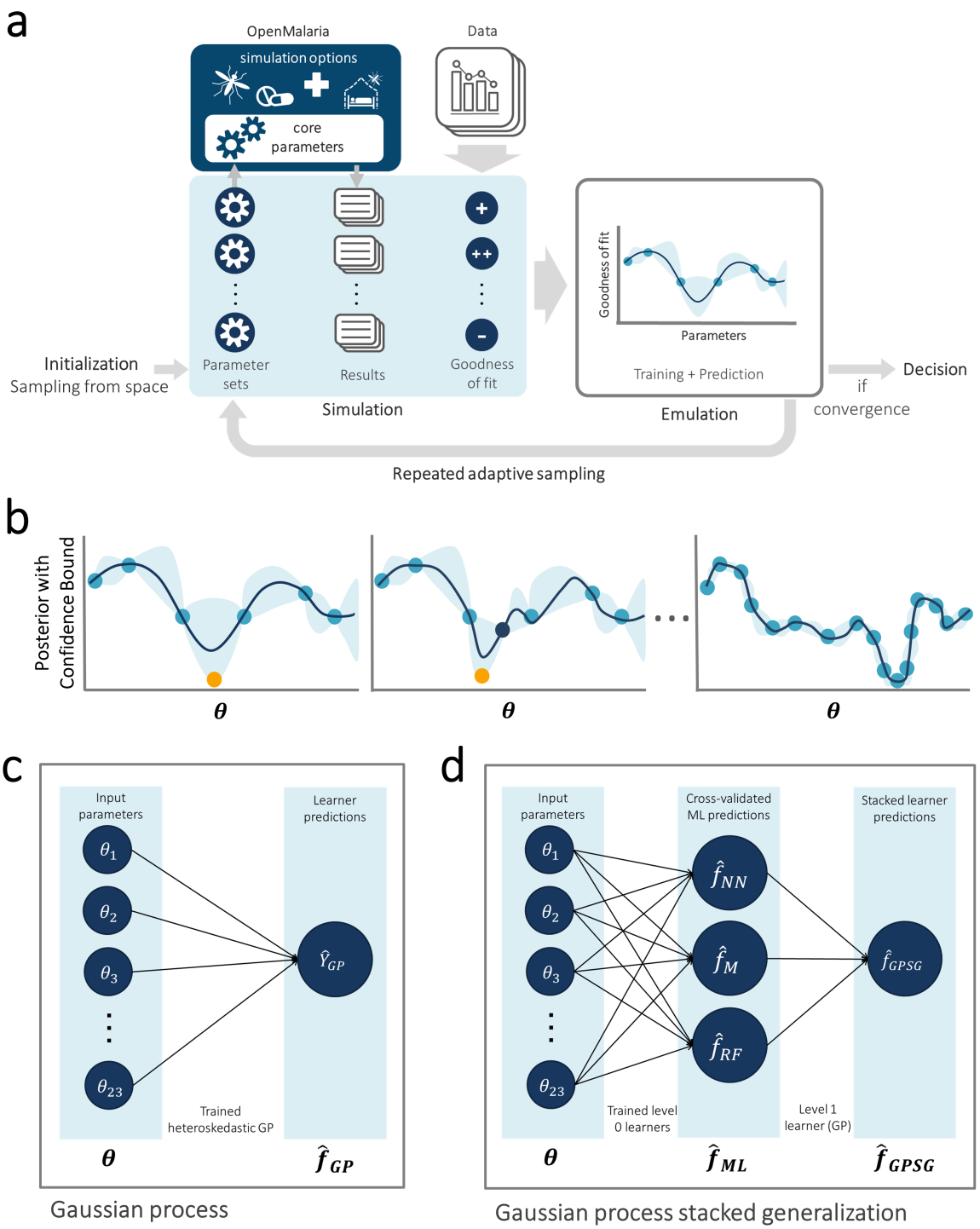

FIGURE 2.1: Overview of model calibration framework by Bayesian optimisation, acquisition function, and Gaussian process and machine learning emulators.
a. General Framework. The input parameter space is initially sampled in a space-filling manner, generating the initial core parameter sets (initialisation). For each candidate set, simulations are performed with the model, mirroring the studies that yielded the calibration data. The deviation between simulation results and data is assessed, yielding goodness of fit scores for each parameter set. (caption continued on the next page)

Figure 2.1 continued. An emulator (c or d) is trained to capture the relationship between parameter sets and goodness of fit and used to generate out-of-sample predictions. Based on these, the most promising additional parameter sets are chosen (adaptive sampling by means of an acquisition function), evaluated, and added to the training set of simulations. Training and adaptive sampling are repeated until the emulator converges and a decision on the parameter set yielding the best fit is made. **b. Acquisition Function.** The acquisition function (black line) is used to determine new parameter space locations, $\boldsymbol{\theta}$. $\boldsymbol{\theta}$ is a vector of input parameters (23-dimensional for the model described here) to be evaluated during adaptive sampling (blue dot for previously evaluated locations, orange dot for new locations to be evaluated in the current iteration). It incorporates both predictive uncertainty (blue shading) of the emulator and proximity to the minimum. **c. Gaussian process (GP) emulator.**A heteroscedastic Gaussian process is used to generate predictions on the loss functions, $\hat{\mathbf{f}}_{\mathrm{GP}}(\boldsymbol{\theta})$,for each input parameter set $\boldsymbol{\theta}$. **d. Gaussian process stacked generalisation (GPSG) emulator.** Three machine learning algorithms (level 0 learners: bilayer neural net, multivariate adaptive regression splines and random forest) are used to generate predictions on the individual objective loss functions $\hat{\mathbf{f}}_{\mathrm{NN}}$, $\hat{\mathbf{f}}_{\mathrm{M}}$, and $\hat{\mathbf{f}}_{\mathrm{RF}}$ (collectively $\hat{\mathbf{f}}_{\mathrm{ML}}$) at locations $\boldsymbol{\theta}$. These predictions are inputs to a heteroscedastic (level 1 learner) which is used to generate the stacked learner predictions $\hat{\mathbf{f}}_{\mathrm{GPSG}}$ and derive predictions on the overall goodness of fit $\hat{\mathbf{F}}_{\mathrm{GPSG}}$.

2.3.2 Emulator Performance

2.3.3 Algorithm performance (by iteration and time) and convergence

Both emulators adequately captured the input–output relationship of the calculated loss-functions from the simulator, with better accuracy when close to minimal values of the weighted sum of the loss functions, F (Figure 2.3b). This is sufficient as the aim of both emulators within the Bayesian optimisation framework is to find minimal loss function values rather than an overall optimal predictive performance for all outcome values. Examples of truth vs. predicted estimates on a 10% holdout set are provided in Figure 2.2a (additional plots for all objectives can be found in Supplementary Figures A.2-A.5). A *satisfactory fit* of the simulator was previously defined by a loss function value of $F = 73.2$ (*64*). The *previous best* model fit derived using the GA had a weighted sum of the loss functions of F=63.7 (*64*). *Satisfactory fit* was achieved by our approach in the first iteration of the GPSG-based Bayesian optimisation algorithm (GPSG-BO), and after six iterations for the GP-based algorithm (GP-BO) (Figure 2.2b). The *current best* fit was approximately retrieved after six iterations for the GPSG-BO algorithm and after nine iterations for GP-BO, and was improved by both algorithms after 10 iterations (returning final values $F = 58.3$ for GP-BO and 59.6 for GPSG-BO). This shows that the Bayesian optimisation approach with either of our emulators very quickly achieves a better simulator fit than obtained with a classical GA approach that was previously employed to calibrate *OpenMalaria*. Of the two emulators, the GP approach finds a parameter set associated with a better overall accuracy and the GPSG reaches *satisfactory* values faster (both in terms of iterations and time). A likely explanation for this is that the GPSG-BO is unable to propagate its full predictive variance into the acquisition function. Only uncertainty stemming from the level 1 probabilistic learner (GP) is therefore captured in the final prediction. This leads to underestimation of the full predictive variance, and a bias towards exploitation in the early stages of the GPSG-BO algorithm (as illustrated by early narrow sampling, see Supplementary Figures A.6-A.7).

Figure 2.2c shows examples of the posterior estimates returned by the optimisation algorithms in context of the log prior distributions for the parameters with the greatest effects on F (see also Figure 2.3c). All algorithms return parameter values within the same range and (apart from parameter 4), clearly distinct from the prior mean. The fact that highly similar parameter values are identified by multiple algorithms strengthens confidence in the final parameter sets yielded by the algorithms.

2.3.4 Optimal goodness of fit

The best fit parameter sets yielded by our approach are provided in Supplementary Table A.1. Importantly, after ten iterations of the GPSG-BO algorithm (˜7 days), and 20 iterations for the GP-BO algorithm (˜12 days), both approaches yielded similar values of the 11 objective loss functions, along with similar weighted total loss function values, and qualitatively

similar visual fits and predicted trends to the data (Figure 2.3a,b and Supplementary material Appendix A). WWe found this to be an unexpectedly fast result of the two algorithms. Details of the algorithm's best fits to the disease and epidemiological data are shown in Supplementary Figures A.8-A.18. Overall, several objectives had visual and reduced loss-function improvements, for example to the objective on the multiplicity of infection (Figure 2.3a).

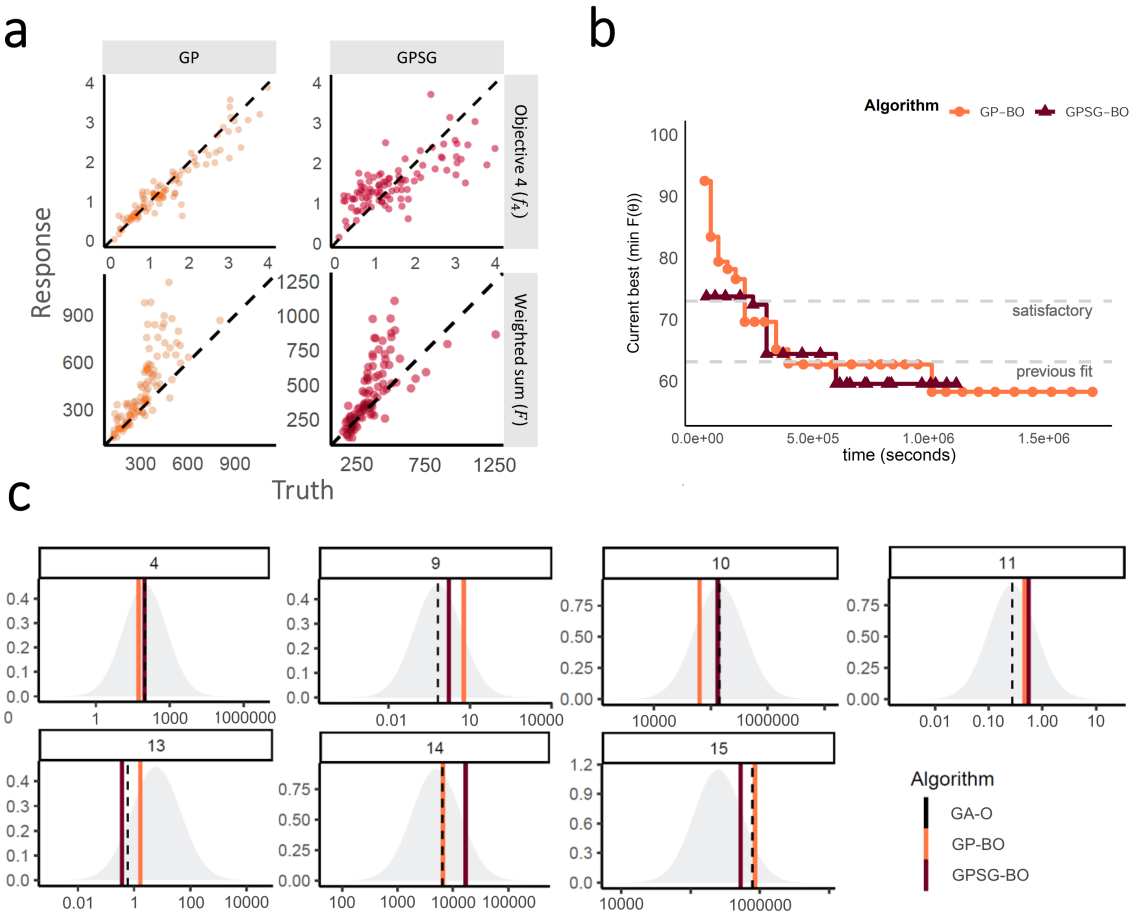

FIGURE 2.2: Emulator performance including predictions, convergence, and prior parameter distributions and posterior estimates.

a Example of emulator predictions vs true values on a 10% holdout set. Predictions are shown for the final iteration of each optimisation (orange dots for predictions in iteration 30 for GP-BO and red dots for predictions in iteration 23 for GPSG-BO). Here, emulator performances are shown for objective 4 (the age-dependent multiplicity of infection, f_4) and the weighted sum of loss functions over 11 objectives, $F(\boldsymbol{\theta})$. Plots for all other objectives are provided in the supplement (Appendix A). BO Bayesian optimisation, GP Gaussian process emulator, GPSG Gaussian process stacked generalisation emulator. **b** Convergence of the weighted sum of loss functions over 11 objectives ($F(\boldsymbol{\theta})$ associated with the current best fit parameter set by time in seconds. Satisfactory fit of OpenMalaria refers to a weighted sum of loss functions value of 73.2 (*64*). The previous best fit for *OpenMalaria* was achieved by the genetic algorithm and had a loss function value of 63.7. Our approach yields a fit F of 58.2 for GP-BO in iteration 21 within in $1.02e^6$s (˜12 days) and 59.6 for GPSG-BO in iteration 10 in $6.00e^5$ s (˜7 days). GP-BO Gaussian process emulator Bayesian optimisation, GPSG-BO Gaussian process stacked generalisation emulator Bayesian optimisation. **c** Example log prior parameter distributions (shown by the gray areas) and posterior estimates (vertical lines). The most influential parameters on the weighted sum of the loss functions are shown here in this figure (most influential parameters shown in Figure 2.3c). All other plots can be found in the supplement (Appendix A). The posterior estimates for GP-BO (orange line) and GPSG-BO (red line) are shown in relation to those previously derived through optimisation using a genetic algorithm (GA-O, dashed black line) for parameters $\theta_{4,9-11,13-15}$ (numbers in the panel labels)

2.3.5 Impact/parameter sensitivity analysis and external validation

An additional benefit of using emulators is the ability to understand the outcome's dependence on and sensitivity to the input parameters. To identify the most influential parameters for each of the 11 fitting objectives, we used the GP emulator trained on all available training simulation results from the optimisation process (R^2=0.53 [objective 7] - 0.92 [objective 3]) to conduct a global sensitivity analysis by variance decomposition (here via Sobol analysis (*100*)). Figure 2.3c shows Sobol total effect indices quantifying the importance of individual parameters and describing each parameter's contributions to the outcome variance for each objective. Our results indicate that most objectives are influenced by multiple parameters from different groups, albeit to varying degrees, thus highlighting the importance of simultaneous multi-objective fitting. Clusters of influential parameters can be observed for most objectives; for example, parameters associated with incidence of acute disease influence clinical incidence and pyrogenic threshold objectives. Some parameters have strong influence on multiple objectives, such as parameter 4, the critical value of cumulative number of infections and influences immunity acquisition; and parameter 10, a factor required to determine the pyrogenic threshold, which we find to be a key parameter determining infections progressing to clinical illness.

2.3.6 Algorithm validation

To test if our algorithms can recover a known solution, the final parameter sets for both approaches were used to generate synthetic field data sets, and our approaches were subsequently applied to recover the known parameter set. For the GP, 13 of the 23 parameters were recovered (Supplementary FigureA.19a).Those not recovered largely represented parameters to which the weighted loss function was found to be insensitive (Figure 2.3c). Thus, rather than showing a shortcoming of the calibration algorithm, this suggests a potential for dimensionality reduction of the simulator and re-evaluation of its structure.

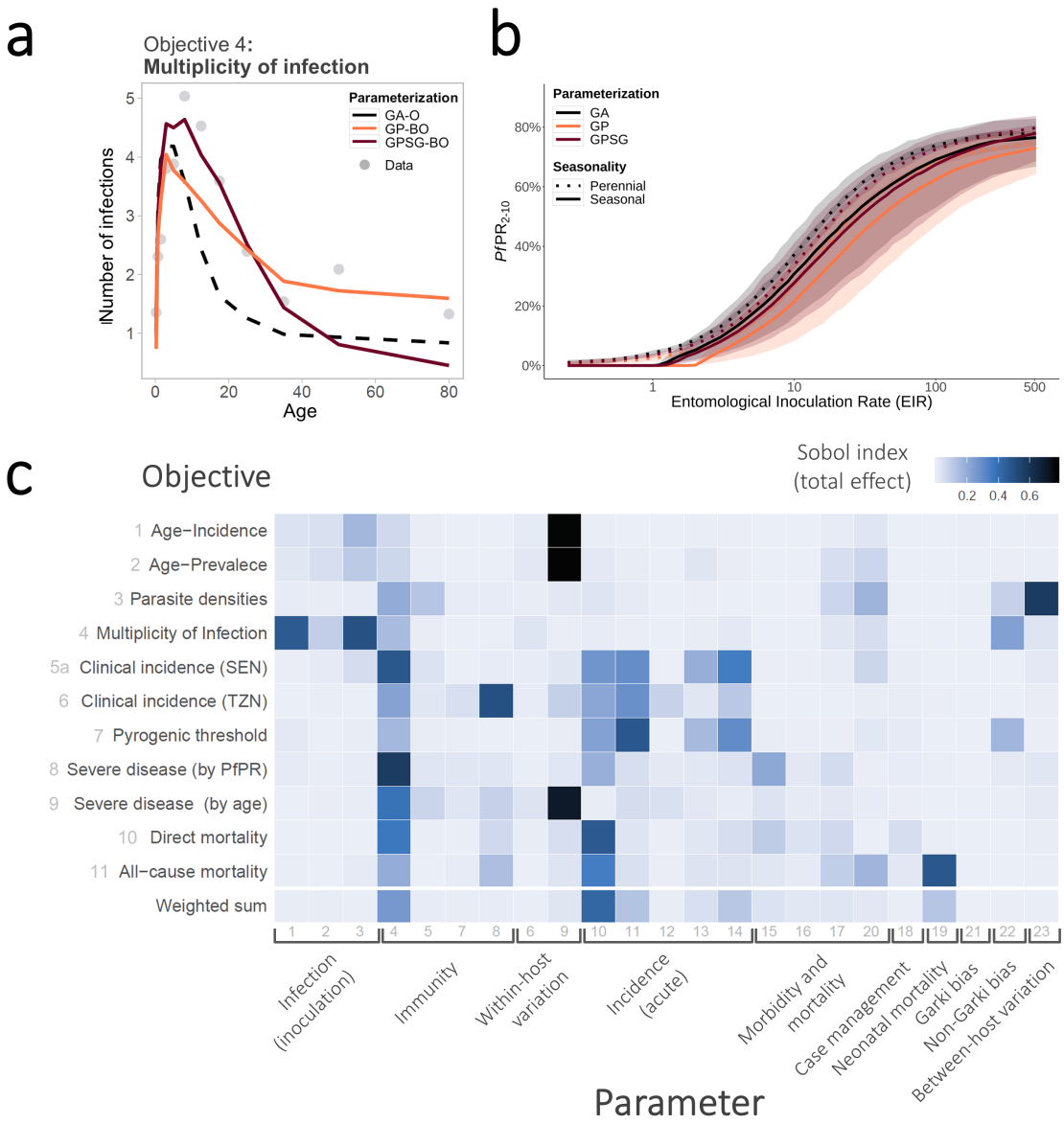

FIGURE 2.3: Exemplar plot of calibration and data for objective four "Multiplicity of infection", with exemplar epidemiological predictions of prevalence vs. EIR for the final calibration, and sensitivity of fitting objectives to each parameter.

a Multiplicity of infection by age. Comparison of simulator goodness of fit for objective 4, the age-specific multiplicity of infection (number of genetically distinct parasite strains concurrently present in one host). Simulations were carried out for the same random seed for all parameterisations and for a population size of $N = 5,000$. **b**. Simulated epidemiological relationship between transmission intensity (entomological inoculation rate, EIR) and *P. falciparum* prevalence rate ($PfPR_{2-10}$). Simulated epidemiological relationship between the transmission intensity (EIR in number of infectious bites per person per year) and infection prevalence in individuals aged 2–10 years ($PfPR_{2-10}$)under the parameterisations achieved by the different optimisation algorithms. Lines show the mean across 100 random seed simulations for a simulated population size $N = 10,000$ and the shaded area shows the minimum to maximum range. **c**. Parameter effects on the objective variance. Using the GP emulator, a global sensitivity analysis (Sobol analysis) was conducted. The tile shading shows the total effect indices for all objective functions and parameters grouped by function. SEN Senegal, TZN Tanzania.

2.3.7 Comparison of key epidemiological relationships and implications for predictions

The new parameterisations for *OpenMalaria* were further explored to assess key epidemiological relationships, in an approach similar to the four malaria model comparison in Penny et al. 2016 (*11*). We examined incidence and prevalence of disease, as well as incidence of mortality for multiple archetypical settings, considering a range of perennial and seasonal transmission intensity and patterns. The results are presented in Figure 2.3b and Supplementary Figures A.20-A.30. The new parameterisations result in increased predicted incidence of severe episodes and decreased prevalence for all transmission intensities (thus also slightly modifying the prevalence–incidence relationship). While we found that the overall implications for the other simulated epidemiological relationship were small, the differences in predictions for severe disease may carry implications for public health decision-making and warrant further investigations. We conclude that our new parameterisations do not fundamentally bring into question previous research conducted using *OpenMalaria*, but we do suggest re-evaluation of adverse downstream events such as severe disease and mortality.

2.4 Discussion

Calibrating IBMs can be challenging as many techniques struggle with high dimensionality, or become infeasible with long model simulation times and multiple calibration objectives. However, ensuring adequate model fit to key data is vital, as this impacts the weighting, we should give model predictions in the public health decision-making process. The Bayesian optimisation approaches presented here provide fast solutions to calibrating IBMs while improving model accuracy, and by extension prediction accuracy.

Using a Bayesian optimisation approach, we calibrated a detailed simulator of malaria transmission and epidemiology dynamics with 23 input parameters simultaneously to 11 epidemiological outcomes, including age-incidence and age-prevalence patterns. The use of a probabilistic emulator to predict goodness-of-fit, rather than conducting sparse sampling, allows for cheap evaluation of the simulator at many locations and increases our confidence that the final parameter set represents a global optimum. Our approach provides a fast calibration whilst also providing a better fit compared with the previous parameterisation. We are further able to define formal endpoints to assess calibration alongside visual confirmation of goodness of fit (*5*, *64*), such as the emulator's predictive variance approaching the observed simulator variance. The emulator's ability to quantify the input stochasticity of the simulator also enables simulation at small population sizes, contributing to fast overall computation times.

Despite the demonstrated strong performance of stacked generalisation in other contexts such as geospatial mapping (*33*, *99*, *101–104*), we found that using a *superlearning* emulator

for Bayesian optimisation was not superior to traditional GP-based methods. In our context using GPSG sped up convergence of the algorithm, but both approaches, GP and GPSG, led to equally good fits. Each approach does, however, have different properties with context-dependent benefits: The dimensionality reduction provided by GPSG approaches may lead to computational savings depending on the *level 0* and *level 1* learners. At the same time, only level 1 learner uncertainty is propagated into the final objective function predictions, which affects the efficacy of adaptive sampling and may lead to overly exploitative behaviour, where sampling close to the point estimate of the predicted optimum is overemphasised, rather than exploring the entire parameter space (see Supplementary Tables A.2 and A.3 on selected points). On the other hand, exploration/exploitation trade-offs for traditional GP-BO algorithms have long been examined and *no regret* solutions have been developed (*105*).

The methodology presented here constitutes a highly flexible framework for individual based model calibration and aligns with the recent literature on using emulation in combination with stochastic computer simulation experiments of infectious diseases(*106*). Both algorithms can be applied to other parameterisation and optimisation problems in disease modelling and also in other modelling fields, such as physical or mobility and transport models. Furthermore, in the GPSG approach, additional or alternative level 0 can be easily incorporated. Possible extensions to our approach include combination with methods to adaptively reduce the input space for constrained optimisation problems (*107*), or other emulators may be chosen depending on the application. For example, homoscedastic GPs, which are faster than the heteroscedastic approach presented here, may be sufficient for many applications (but not for our IBM in which heteroscedastic was required due to the stochastic nature of the model). Alternatively, the computational power required by neural net algorithms scales only linearly (compared with a nominal cubic scaling for GPs) with the sample size, and we envisage wide applications for neural net-based Bayesian optimisation in high dimensions. In our example, the bilayer neural net algorithm completed training and prediction within seconds whilst maintaining very high predictive performance. Unfortunately, estimating the uncertainty required for good acquisition functions is difficult in neural networks, but solutions are being developed (*23*, *108*). These promising approaches should be explored as they become more widely available in high-level programming languages. With the increased availability of code libraries and algorithms, Bayesian optimisation with a range of emulators is also becoming easier to implement.

The probabilistic, emulator-based calibration approach is accompanied by many benefits, including relatively quick global sensitivity analysis. As explored in this work, GP-based methods are easily coupled with sensitivity analyses, which provide detailed insights into a model's structural dependencies and the sensitivity of its goodness of fit to the input parameters. To the best of our knowledge, no other individual-based model calibration study has addressed this. In the case of malaria models, we have shown the interdependence of all *OpenMalaria* model components and a relative lack of modularity. In particular, within-host

immunity-related parameters were shown to influence all fitting objectives, including downstream events such as severe disease and mortality when an infection progresses to clinical disease. Thus, calibrating within-host immunity in the absence of key epidemiology and population outcomes can lead to suboptimal calibration and ultimate failure of the model to adequately capture disease biology and epidemiology.

We have employed a different approach to calibrating *OpenMalaria* compared with previous methods but reach broadly similar comparisons to the natural history of disease. We also attainted a slightly improved but similar goodness of fit, the main benefit being improved fitting times and the ability to measure parameter importance. Given the high number of influential parameters for each epidemiological objective in our parameter importance investigations, and the overlap between parameter–objective associations, we argue that, where possible, multi-objective fitting should be preferred over purely sequential approaches. Our approach confirms that using a parallel approach to parameterisation rather than a modular, sequential, one captures the joint effects of all parameters and ensures that all outcomes are simultaneously accounted for. To the best of our knowledge, no model of malaria transmission of comparable complexity and a comparable number of fitting objectives was simultaneously calibrated to all its fitting objectives. Disregarding the joint influence of all parameters on the simulated outcomes may negatively impact the accuracy of model predictions, in particular on policy-relevant outcomes of severe disease and mortality.

Despite providing relatively fast calibration towards a better fitting parameter set, several limitations remain in our work. We have not systematically tested that a global optimum has been reached in our approach, but assume it is close to a global minimum for the current loss-functions defined, as further iterations did not yield changes, and both the GP and GPSG achieved similar weighted loss function and parameter sets. We aimed to improve the algorithm to calibrate detailed IBM, but we did not incorporate new data, which will be important moving forward as our parameter importance and validation analysis highlights several key epidemiological outcomes on severe disease and mortality are sensitive to results.

The key limitations of Bayesian optimisation, particularly when using a GP emulator, are the high computational requirements in terms of memory and parallel computing nodes due to increasing runtimes and cubically scaling memory requirements of GPs. For this reason, we opted to not employ fully Bayesian KOH methods, which would double the number of GPs that would need to be run. Yet, memory limits may be reached before the predictive variance approached its limit. Furthermore, we chose an acquisition function with high probability to be *no regret* (*105*), but this likely overemphasizes exploration in the early stages of the algorithm considering the dimensionality of the problem and finite runtime. We opted here for pure exploitation every five iterations, but a more formal optimisation of the acquisition function should be explored. The GPSG approach presented here can partially alleviate this challenge, depending on the choice of learning algorithms, but the iterative nature and need

for many simulations remain. Memory-saving and time-saving extensions are thus worth exploring, such as incorporating graphics processing unit (GPU) computing or adaptively constraining the prior parameter space, dimensionality reduction, or addressing alternative acquisition functions. Additionally, as with all calibration methodologies, many choices are left to the user, such as the size of the initial set of simulations, the number of points added per iteration, or the number of replicates simulated at each location. There is no general solution to this as the optimal choices are highly dependent on the problem at hand, and we did not aim to optimize these. Performance might be optimised further through a formal analysis of all these variables, however the methodology here is already fast, effective, and highly generalizable to different types of simulation models and associated optimisation problems. Improving the loss-functions or employing alternative *Pareto front* efficiency algorithms was not the focus of our current study but would be a natural extension of our work, as would be alternative approaches to the weighting of objectives, which remains a subjective component of multi-objective optimisation problems (*109*).

A model's calibration to known input data forms the backbone of its predictions. The workflow presented here provides great advances in the calibration of detailed mathematical models of infectious diseases such as IBMs. Provided sufficient calibration data to determine goodness-of-fit, our approach is easily adaptable to any agent-based model and could become the modus operandi for multi-objective, high-dimensional calibration of stochastic disease simulators.

2.5 Methods

2.5.1 Preparation of calibration data and simulation experiments

Disease transmission models generally have two types of parameter inputs: core parameters, inherent to the disease and determining how its natural history is captured, and simulation options characterizing the specific setting and the interventions in place (Figure 2.1a). The simulation options specify the simulation context such as population demographics, transmission intensity, seasonality patterns, and interventions, and typically vary depending on the simulation experiment. In contrast, the core parameters determine how its epidemiology and aetiopathogenesis are captured. These include parameters for the description of immunity (e.g., decay of maternal protection), or for defining clinical severe episodes (e.g., parasitemia threshold). To inform the estimation of core parameters, epidemiological data on the natural history of malaria were extracted from published literature and collated in previous calibrations of *OpenMalaria* (*3*, *52*, *63*), which were re-used in this calibration round and detailed in the Supplementary material. These include demographic data such as age-stratified numbers of host individuals which are used to derive a range of epidemiological outcomes such as age-specific prevalence and incidence patterns, mortality rates, and hospitalisation rates.

Site-specific *OpenMalaria* simulations were prepared, representing the studies that yielded these epidemiological data in terms of transmission intensity, seasonal patterns, vector species, intervention history, case management, and diagnostics (*63*). The mirroring of field study characteristics in the simulation options ensured that any deviation between simulation outputs and data could be attributed to the core parameters. Age-stratified simulation outputs to match to the data include numbers of host individuals, patent infections, and administered treatments. A summary of the data is provided in the Supplementary Note A.2 .

2.5.2 General Bayesian optimisation framework with emulators

In our proposed Bayesian optimisation framework (Figure 2.1) we evaluated the deviation between simulation outputs and the epidemiological data by training probabilistic emulator functions that approximate the relationship between core parameter sets and goodness of fit. To test the optimisation approach in this study we considered the original goodness of fit metrics for *OpenMalaria* detailed in reference (*64*) and in Supplementary Note A.2, which uses either Residual Sum of Squares (RSS) or negative log-likelihood functions depending on the epidemiological data for each objective (*63*, *64*). The objective function to be optimised is a weighted sum of the individual objectives' loss functions.

We adopted a Bayesian optimisation framework where a probabilistic emulator function is constructed to make predictions over the loss functions for each objective from the input space, with a minimum amount of evaluations of the (computationally expensive) simulator.

We compared two emulation approaches. Firstly, a heteroskedastic GP emulator and secondly a stacked generalisation emulator (*99*). For approach 1 (GP-BO), we fitted a heteroskedastic Gaussian process with the input noise modelled as another GP (*110*) with a Matérn 5/2 kernel to account for the high variability in the parameter space (Figure 2.1c) (*97*, *111*). For approach 2 (GPSG-BO), we selected a two-layer neural network (*112–114*), multivariate adaptive regression splines (*115*), and a random forest algorithm (*74*, *116*) as level 0 learners.

With each iteration of the algorithm, the training was extended using adaptive sampling based on an acquisition function (lower confidence bound) that accounts for uncertainty and predicted proximity to the optimum of proposed locations (Figure 2.1b). As the emulator performance improves (as assessed by its predictive performance on the test set) we gain confidence in the currently predicted optimum.

2.5.3 Malaria transmission and disease simulator

We applied our novel calibration approach to *OpenMalaria* `github.com/SwissTPH/openmalaria.wiki.git`, an open source modelling platform of malaria epidemiology and control. It

features several related individual-based stochastic models of *P. falciparum* malaria transmission and control. Overall, the *OpenMalaria* IBM consists of a model of malaria in humans linked to a model of malaria in mosquitoes and accounts for individual level heterogeneity in humans (in exposure, immunity, and clinical progression) as well as aspects of vector ecology (e.g., seasonality and the mosquito feeding cycle). Stochasticity is featured by including between- and within-host stochastic variation in parasite densities with downstream effects on immunity (*63*). *OpenMalaria* further includes aspects of the health system context (e.g. treatment seeking behaviour and standard of care) (*52*, *63*) with additional probabilistic elements such as treatment seeking probabilities or the option for stochastic results of diagnostic tests. An ensemble of *OpenMalaria* model alternative variants is available defined by different assumptions about immunity decay, within-host dynamics, heterogeneity of transmission, along with more detailed sub-models that track parasite genetics, and pharmacokinetic and pharmacodynamics. The models allow for the simulation of interventions, such as the distribution of insecticide-treated nets (ITNs), vaccines, or reactive case detection (*50*, *117*), in comparatively realistic settings. Full details of the model and the history of calibration can be found in the original publications (*52*, *63*, *64*) and are summarised in Supplementary Notes A.1 and A.2. In our application, we use the term *simulator* to refer to the *OpenMalaria* base model variant (*64*).

2.5.4 Calibrating *OpenMalaria*: loss functions and general approach

Aim

Let $f(\boldsymbol{\theta})$ denote a vector of loss functions obtained by calculating the goodness of fit between simulation outputs and the real data (full details of loss function can be found in Supplementary Note A.2). In order to ensure a good fit of the model, we aim to find the parameter set $\boldsymbol{\theta}$ that achieves the minimum of the weighted sum of 11 loss functions (corresponding to the 10 fitting objectives) $F(\boldsymbol{\theta}) = \sum_{i=1}^{11} w_i f_i(\boldsymbol{\theta})$, where $f_i(\boldsymbol{\theta})$ is the value of objective function i at $\boldsymbol{\theta}$ and w_i is the weight assigned to objective function i :

$$\underset{\theta}{argmin} \left(\sum_{i=1}^{11} w_i f_i(\boldsymbol{\theta}) \right) \tag{2.1}$$

The weights are kept consistent with previous rounds of calibration and chosen such that different epidemiological quantities contributed approximately equally to $F(\boldsymbol{\theta})$ (see Supplementary Note A.2).

Step 1: Initialisation.

Let $D = 23$ denote the number of dimensions of the input parameter space $\boldsymbol{\Theta}$ and $W = 11$ the number of objective functions $f_i(\boldsymbol{\theta}), i = 1, \ldots, 11$. Prior distributions consistent with previous fitting runs (*64*) were placed on the input parameters. As each parameter is measured in different units, we sampled from the D -dimensional unit cube $\boldsymbol{\Theta}$ and converted these to quantiles of the prior distributions (*64*) (Supplementary Note A.2 and Figure

A.6). Previous research suggests that in high-dimensional spaces quasi-Monte Carlo (qMC) sampling outperforms random or Latin Hypercube designs for most function types and leads to faster rates of convergence (*118*, *119*). We therefore used Sobol sequences to sample 1,000 initial locations from $\boldsymbol{\Theta}$. The GP can account for input stochasticity of the simulator. For each sample, we simulated 2 random seeds at a population size of 10,000 individuals. Additionally, 100 simulations were run at the centroid location of the unit cube to gain information on the simulator noise. Using small noisy simulations with small populations speeds up the fitting as the noisy simulations are less computational expensive than larger population runs. Replicates were used to detect signals in noisy settings and estimate the pure simulation variance (*110*). Computational savings were later achieved through pre-averaging of replicates (*110*). The 2000 unique locations were randomly split into a training set (90%) and a test set (10%). All simulator realisations at the centroid were added to the training set.

Step 2: Emulation

2.1: Emulator Training

Each emulator type for each objective function was trained in parallel to learn the relationships between the normalized input space $\boldsymbol{\Theta}$, and the log-transform of the objective functions $f(\boldsymbol{\theta})$. In each dimension $d \in D$, the mean μ_d and standard deviation σ_d of the training set were recorded, $d = 1, \ldots, 23$.

2.2 Posterior prediction

We randomly sampled 500,000 test locations in $\boldsymbol{\Theta}$ from a multivariate normal distribution with mean $\boldsymbol{\theta}_{opt}$ and covariance matrix $\boldsymbol{\Sigma}$, where $\boldsymbol{\theta}_{opt}$ is the location of the current best location and $\boldsymbol{\Sigma}$ is determined based on previously all sampled locations, and scaled each dimension to mean μ_d and standard deviation σ_d . The trained emulators were used to make predictions $\widehat{\boldsymbol{F}(\boldsymbol{\theta})}$ of the objective functions $F(\boldsymbol{\theta})$ at the test locations. Mean estimates, standard deviations, and nugget terms were recorded. The full predictive variance at each location $\boldsymbol{\theta} \in \boldsymbol{\Theta}$ corresponds to the sum of the standard deviation and nugget terms. From this, we derived the weighted sum $\hat{F}(\boldsymbol{\theta}) = \sum_{i=1}^{11} w_i f_i(\boldsymbol{\theta})$, using weights $\boldsymbol{w}$ consistent with previous fitting runs (*64*) with greater weighting for further downstream objectives. The predicted weighted loss function at location $\boldsymbol{\theta}$ was denoted $\hat{F}(\boldsymbol{\theta})$ with a predicted mean $\hat{\mu}_F(\boldsymbol{\theta})$ and variance $\hat{\sigma}_F(\boldsymbol{\theta})$. Every 15 iterations, we increase the test location sample size to 5 million to achieve denser predictions.

Step 3: Acquisition.

We chose the lower confidence bound (LCB) acquisition function to guide the search of the global minimum (*120*). Lower acquisition corresponds to *potentially* low values of the weighted objective function, either because of a low mean prediction value or large uncertainty (*121*). From the prediction set at iteration t , we sample without replacement 250 new locations $\boldsymbol{\theta} = argmin_\theta \{\hat{\mu}_F(\boldsymbol{\theta}, t) - \sqrt{\nu \tau_t} \ \hat{\sigma}_t(\boldsymbol{\theta}, t)\}$, with the hyperparameter $\nu = 1$ and $\tau_t = 2log\left(T_t^{D/2+2} \pi / 3\delta\right)$, where T_t is the number of previous unique realisations of the simulator at iteration t, and $\delta = 0.01$ is a hyperparameter (*105*). We choose this method as with high probability it is *no regret* (*105*, *121*). With increasing iterations, confidence bound-based

methods naturally transition from mainly exploration to exploitation of the current estimated minimum. In addition to this, we force exploitation every 10 iterations by setting $\tau_t = 0$).

Step 4: Simulate.

The simulator was evaluated at locations identified in step 3 and the realisations were added to the training set. Steps 2-4 were run iteratively. The Euclidian distance between locations of current best realisations was recorded.

Step 5: Convergence.

Convergence was defined as no improvement in the best realisation, $argmin_{\boldsymbol{F}}\, \boldsymbol{F}$.

2.5.5 Emulator definition

We compared two emulation approaches. Firstly, a heteroskedastic GP emulator and secondly a stacked generalisation emulator (*99*) using a two-layer neural net, multivariate adaptive regression splines (MARS) and a random forest as level 0 learners and a heteroskedastic GP as level 1 learner:

Heteroskedastic Gaussian Process (hetGP).

We fitted a Gaussian process with the input noise modelled as another Gaussian process (*110*). After initial exploration of different kernels, we chose a Matérn 5/2 kernel to account for the high variability in the parameter space. A Matérn 3/2 correlation function was also tested performed equally. Each time the model was built (for each objective at each iteration), its likelihood was compared to that of a homoscedastic Gaussian process and the latter was chosen if its likelihood was higher. This resulted in a highly flexible approach, choosing the best option for the current task.

Gaussian Process Stacked Generalisation (GPSG).

Stacked generalisation was first proposed by Wolpert 1992 (*99*) and builds on the idea of creating ensemble predictions from multiple learning algorithms (level 0 learners). In *superlearning*, the cross-validated predictions of the level 0 learners are fed into a level 1 meta-learner. We compared the 10-fold cross-validated predictive performance of twelve machine learning algorithms on the test set. All algorithms were accessed through the mlr package in R version 2.17.0 (*122*). We compared two neural network algorithms (brnn (*113*) for a two layer neural network and nnet for a single-hidden-layer neural network (*123*), five regression algorithms (cvglmnet (*115*) for a generalised linear model with LASSO or Elasticnet Regularisation and 10-fold cross validated lambda, glmboost (*124*) for a boosted generalised linear model, glmnet (*115*) for a regular GLM with Lasso or Elasticnet Regularisation, mars for multivariate adaptive regression splines, and cubist for rule-and instance-based regression modelling), three random forest algorithms (randomForest (*116*), randomForestSRC (*125*) and ranger (*126*)), and a tree-like node harvesting algorithm (nodeHarvest (*127*)). Extreme gradient boosting and support vector regression were also tested but excluded from the comparison

due to its long runtime. Their performance was compared with regards to runtimes, and correlation coefficients between predictions on the test set and the true values. Based on these, we selected the two-layer neural network (brnn) (*114*), multivariate adaptive regression spline (mars) (*128*), and random forest (randomForest) (*116*) algorithms. This ensemble of machine learning models constituted the level 0 learners and was fitted to the initialisation set. Out-of-sample predictions from a 10-fold cross validation of each observation were used to fit the level 1 heteroskedastic GP. As in approach 1, we opted for a Matérn 5/2 kernel and retained the option of changing to a homoscedastic model where necessary.

2.5.6 Emulator performance

We ascertained that both emulators captured the input-output relationship of the simulator by tracking the correlation between true values f and predicted values $\hat{f}$ on the holdout set of 10% of initial simulations with each iteration (truth vs predicted R2 0.51-0.89 for GP vs 0.37-0.77 for GPSG after initialisation, see Supplementary Figure A.1). Transition from exploration to exploitation during adaptive sampling was tracked by recording the distribution of points selected during adaptive sampling in each iteration (Figures A.2 and A.3).

2.5.7 Sensitivity analysis

A global sensitivity analysis was conducted on a heteroskedastic GP model with Matérn 5/2 kernel that was trained on all training simulation outputs (n=5,400) from the fitting process. We used the Jansen method of Monte Carlo estimation of Sobol' sensitivity indices for variance decomposition (*129*, *130*) with 20,000 sample points and 1000 bootstrap replicates. Sobol' indices were calculated for all loss functions f as well as for their weighted sum F and in all dimensions. Whilst keeping the number of sample points to as low as possible for computational reasons, we ascertained that first-order indices summed to 1 and total effects >1. We further ensured that the overall results of the Sobol' analysis were consistent with the results of other global sensitivity analyses, namely the relative parameter importance derived from training a random forest (Figure A.32).

2.5.8 Synthetic data validation

Synthetic field data was generated by forward simulation using the final parameter sets from each optimisation process. The two optimisation algorithms were run anew using the respectively generated synthetic data to calculate the goodness of fit statistics. The parameter sets retrieved by the validation were compared against the parameterisation yielded by the optimisation process.

2.5.9 Epidemiological outcome comparison

We conducted a small experiment to compare key epidemiological outcomes from the new parameterisations with the original model and that detail in a four malaria model comparison in Penny et al. 2016 (*11*). We simulated malaria in archetypical transmission and seasonality

settings using the different parameterisations. The experiments were set up in a full-factorial fashion, considering the simulation options described in Table 2.1. Monitored outcomes were the incidence of uncomplicated, severe disease, hospitalisations, and indirect and direct malaria mortality over time and by age, prevalence over time and by age, the prevalence–incidence relationship, and the EIR–prevalence relationship. Simulations were conducted for a population of 10,000 individuals over 10 years.

TABLE 2.1: Full experimental design in setting archetypes

Experiments were run at 36% probability that an infected individual receives effective care within 14 days.

Number of stochastic realisation	Seasonality	Transmission (EIR)	Parameterisation
10	Perennial Seasonal (sinusoidal)	0.25, 0.5, 0.75, 1, 1.1, 1.25, 1.35, 1.5, 1.75, 2, 2.5, 3, 4, 5, 6, 7, 8, 9, 10, 12, 14, 16, 18, 20, 22, 25, 30, 35, 40, 45, 50, 64, 73, 80, 100, 128, 150, 200, 256, 512	GA GP-BO GPSG-BO

2.5.10 Software

Consistent with previous calibration work, we used *OpenMalaria* vversion 35, an open-source simulator written in C++ and further detailed in full in the supplement (Supplementary Text A.1), as well as *OpenMalaria* wiki (`https://github.com/SwissTPH/openmalaria/wiki`) or in the original publications (*52*, *63*, *64*). Calibration was performed using R 3.6.0. For the machine learning processes, all algorithms were accessed through the mlr package version 2.17.0(*122*). The heteroskedastic GP utilised the hetGP package under version 1.1.2 (*110*). The sensitivity analysis was conducted using the soboljansen function of the sensitivity package version 1.21.0 in R (*131*). All algorithms were adapted to the operating system (CentOS 7.5.1804) and computational resources available at the University of Basel Center for Scientific Computing, SciCORE, which uses a Slurm queueing system. The full algorithm code is available on GitHub and deposited in the zendo database under accession code `https://doi.org/10.5281/zenodo.5595100` and can be easily adapted to calibrate any simulation model. The number of input parameters and objective functions are flexible. Thus, to adapt the code to other simulators, code should be updated to run the respective model simulator, and tailored to user's operating system. Further requirements to adapt the workflow are sufficient calibration data, and a per-objective goodness of fit metric.

2.6 Data availability

All calibration data are detailed in reference (*63*). The data used for model fitting are available on GitHub and deposited in the zendo database under accession code `https://doi.org/10.5281/zenodo.5595100`. The data generated in this study and plotted in the main manuscript or supplement are publicly available and have been deposited

in the zendo database under accession code `https://doi.org/10.5281/zenodo.5552279`.

2.7 Code availability

Code is publicly available on GitHub and deposited in the zendo database under accession code https://doi.org/10.5281/zenodo.5595100.

2.8 Acknowledgements

We acknowledge and thank our colleagues in the Swiss TPH Disease Modeling unit. Calculations were performed at sciCORE (`http://scicore.unibas.ch/`) scientific computing core facility at University of Basel. The work was funded by the Swiss National Science Foundation through SNSF Professorship of M.A.P. (PP00P3_170702) supporting M.A.P., M.G., and L.B. T.R. was supported by Bill & Melinda Gates Foundation Project OPP1032350 to T.A.S. EC's research is supported by funding from the Bill and Melinda Gates Foundation to Curtin University (Opportunity ID: OPP1197730).

2.9 Author Contributions

M.A.P. and E.C. conceived the study. Algorithm development by E.C., T.R., M.A.P., and S.F. with implementation and preparation for sharing on GitHub by T.R. Loss functions by M.A.P. and T.A.S. Sensitivity analysis by T.R. with inputs from M.G. and L.B. First draft was written by T.R. and M.A.P., all authors contributed to writing and interpretation of results and approved the final manuscript.

3 Grounding predictive models in real world data: Why context matters

Theresa Reiker[1,2], Lydia Burgert[1,2], Ewan Cameron[3,4,5], Manuela Runge[1,2,6], Thomas A. Smith[1,2], Melissa A. Penny[1,2*]

[1] Swiss Tropical and Public Health Institute, Basel, Switzerland
[2] University of Basel, Petersplatz 1, Basel, Switzerland
[3] Malaria Atlas Project, Big Data Institute, University of Oxford, Oxford, UK
[4] Curtin University, Perth, Australia
[5] Telethon Kids Institute, Perth Children's Hospital, Perth, Australia
[6] Northwestern University, USA

*Corresponding author

Email: melissa.penny@unibas.ch

Publication: To be submitted to Trends in Parasitology: Review/Opinion

3.1 Summary

Health policy decisions consider a broad range of epidemiological, societal, and economic factors and provide general guidelines as well as solutions tailored to specific questions and settings. Decision-making in malaria is especially intricate because of the complex parasite life cycle, the multitude of interventions and their varying suitability in different contexts. Mathematical modelling is an important tool in navigating this complex public health landscape and its value for supporting evidence-based decision-making has been long recognised in the field of malaria. Applications range from supporting development of general guidelines' to real-time support during acute investigations.

Calibration to diverse data covering a broad portfolio of biological and epidemiological relationship is the backbone of any model and determines the value and credibility of predictions as evidence for informed decision-making. During calibration, the context of data collection must explicitly be accounted as many epidemiological relationships are context-dependent. This requires a comprehensive understanding of the data itself and the setting, and methods under which the data were collected and assembled. Here, we provide a generalizable framework for the integration of data into calibration.

This framework is applied to the *P. falciparum* prevalence -incidence relationship. Accurate incidence measures are difficult to obtain and cartographic interference from more readily available prevalence data often fails to account for the context dependence of the relationship. Through their mechanistic explicitness, mathematical models of malaria transmission can be used to more accurately capture the prevalence-incidence relationship and improve predictions. By providing a library of data and surrounding contextual covariates, the work presented here paves the way from implicit to explicit integration of this relationship into mathematical models.

In light of the increasing relevance of mathematical modelling and integration into decision-making and need for guidance on how data should be used, this manuscript provides a discussion on integrating data for model calibration to ensure credible predictions.

3.2 Background

3.2.1 The role of infectious disease modelling in public health decision-making

Health policy decisions consider a broad range of epidemiological, societal, and economic factors and provide general guidelines as well as solutions tailored to specific settings. Evidence-based decision-making in malaria is especially challenging as it has to address the urgent public health needs in the light of stalling progress in combatting malaria.The parasite's complex two-host life cycle can be interrupted through a variety of interventions, targeting different stages of the transmission cycle. These may be based on anti-parasitic drugs, vector control, or human behavioural adaptation. Their effectiveness depends on an interplay of biological, ecological, geographical, environmental, and socioeconomic factors.

Direct evidence from randomised control trials (RCTs) and other observational studies is generally available only as summarised effect or population impact estimates for specific preventive or therapeutic interventions in specific study populations. Potential interactions between interventions and contextual covariates are often ignored (*2*). The resulting data provides snapshots of the complex disease biology and intervention landscape necessary for elimination or control. Additional quantitative evidence from mathematical modelling can support decisions at population level by comparing and combining interventions.

Mathematical modelling has a long tradition in supporting public health decision-making on the control of infectious diseases. Starting with Bernoulli's 1760 analyses of smallpox transmission (*132*), to the early works of Ross and MacDonald on malaria at the turn of the 20th century, to today's breadth of complex disease transmission simulators for malaria (*4*, *5*, *63*), tuberculosis (*6*), dengue (*8*), and Covid-19 (*17*), the field has been experiencing an explosive increase in both popularity. Given needs to make public health decisions to reduce global mortality and morbidity, mathematical modelling has continued to provide evidence for malaria policy for over a decade (*56*, *59*, *60*). By retaining explicit descriptions of biological and epidemiological foundations, mechanistic mathematical models allow for approximations of experiments in field-like settings without the same use of time and resources (*60*, *133*). They can therefore complement observational studies and generate insights and extrapolations in ways that purely data-driven approaches cannot (*106*, *134*). This enables models to e.g. forward predict the long-term (comparative) effectiveness of interventions and guide funding decisions (*2*). Exemplary use cases include control strategy investigations for endemic diseases (*10*) or acute advice during outbreak investigations, such as real-time estimating the impact of travel restrictions during the West African Ebola Outbreak (*13*) or of non-pharmaceutical interventions during the COVID-19 pandemic in the United Kingdom (*17*).

A model suitable for providing evidence for decision making should synthesise available (data) evidence and contextualise it through biological and epidemiological assumptions (*135*). Models supporting decision-making must be adaptable and flexible, and be able to as

realiably as possible capture both general biological and epidemiological relationships as well as represent a multitude of specific geographical settings. The model must therefore capture a diverse range of biological and epidemiological relationships and provide a comprehensive summary of all aspects of the disease that carry implications for decision-making.

3.2.2 Calibration to diverse real-world data

Alongside valid structural assumptions, calibration to diverse high-quality data is a key backbone of any predictive model. Without diligent calibration, the most elaborate model cannot generate credible evidence for decision-making. To ensure applicability and adaptability of the model to diverse decisions processes, the calibration data must cover the various relationships of interest and be derived from a multitude of settings. In the case of malaria, this includes for example the age-prevalence and age-incidence relationships and age-dependence of the multiplicity of infection (*3*, *9*, *52*, *136*). Much of the data used for the calibration of infectious disease models of malaria stems from epidemiological studies carried out in the late 1980s to early 2000s. Advantages of this are that few interventions were in place at the time, providing a unique opportunity for a relatively *purist* view of disease transmission in a natural environment. However, this means that contextual information was often not systematically collated or has been lost.

Capturing the context of data collection

During calibration, the deviation between model predictions and data is minimised. For this, specific simulations must be created that realistically and comprehensively capture the settings and context of data collection. This includes information on the demography, efficiency of the local health care system, population care seeking behaviour, history of relevant disease control interventions, seasonal transmission patterns, and (if applicable) vector ecology like the abundance of different species or behaviours like biting patterns. By accounting for these confounders, any deviation between data and predictions should be attributable to the model parameters (*136*).

However, capturing specific settings is a challenge because the calibration data is often of historic nature and information has been lost over time. Without expert knowledge of the setting, relevant information on a range of topics must be collated from databases, the published literature (which may include publications in highly specialised journals or in local languages), epidemiological and demographic surveys reports, and other sources.

Here, we propose a generalisable framework for incorporating historical observational data into the calibration of infectious disease models while accounting for contextual covariates.

Specifically, we showcase the process of incorporating data on the age-specific prevalence-incidence relationship of *P. falciparum* malaria (*Pf*PR-incidence relationship) in Africa into *OpenMalaria*, an established individual-based stochastic simulation platform of malaria transmission and control. The resulting database of contextual information for this epidemiological relationship and experimental setups is publicly available. This sets the stage for experimentation in specific settings with *OpenMalaria* as well as other malaria simulators and lays the foundation for incorporating prevalence-incidence relationships into the calibration process. This provides an important step forward in improving the accuracy with which malaria simulators capture this relationship, which turn improves the quality and credibility of model-based evidence in decision-making.

3.3 A comprehensive translational framework

Our comprehensive framework captures the study- and setting-specific contextual covariates to be included in the modelling representation of each study. This framework is illustrated in Figure 3.1. Following our framework, simulations incorporate and reflect the transmission intensity and seasonality of malaria in the region as well as relevant vector ecology. Further, information on health care provision within and outside the study is included. First-line treatment and possible resistance to treatment are accounted for. Current and past control interventions are replicated in simulations. This information can be obtained from extensive searches of the primary and secondary literature and from public databases. Once all information is carefully collated, it is incorporated into the modelling environment as model inputs.

Figure 3.1 provides an overview of the information required to replicate observational studies on malaria. However, this information may be challenging to gather, or conflicting information may exist. The following sections provide an overview of the collection process. Additional, *OpenMalaria*-specific information is provided in Appendix B.

3.3.1 Transmission intensity and seasonality

With regards to gathering contextual information, data should be as unprocessed as possible to maximise objectivity. Raw data should be preferred over aggregated data and over modelled surfaces or statistical estimates. Using annual transmission intensity and seasonality as an example, 3.2 provides a schematic overview of how this information can be obtained and how different sources of information should be weighted. In our exemplar model, in *OpenMalaria*, accurately representing transmission requires providing an estimate of the annual transmission intensity (EIR) as well as the relative transmission each month. Firstly, any direct information (here on annual and monthly EIRs) should be gathered from the primary reference reporting the data. Then, the literature that was cited in the primary study should be consulted, followed by additional studies relating to the same or comparable locations. If

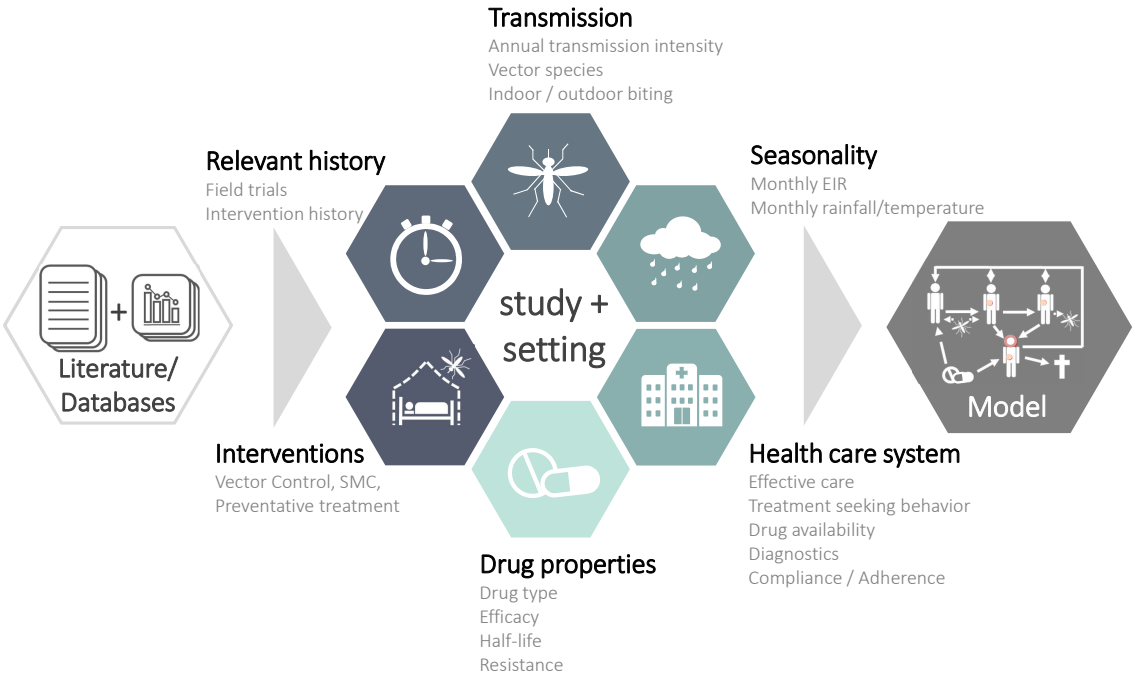

FIGURE 3.1: Translational framework.
For generating simulations that replicate the observational studies that generated the calibration data, the literature and databases must be searched for sufficient information to generate a holistic image of the context of data collection. These include information on transmission intensity and seasonality, specifics around the antimalarial care available to individuals, as well as interventions. Both the study and general, local context must be reflected. EIR= Entomological Inoculation Rate; SMC = Seasonal Malaria Chemoprevention.

information is insufficient, e.g. monthly EIR estimates are unavailable, the extended literature can be searched for direct proxy measures from which seasonality patterns can be inferred such as human biting rates. Lastly, indirect proxy measures that correlate with the measure of interest can be used. In case of relative monthly transmission, this could be information on rainfall or temperature data, which is available in public databases such as the World Bank Climate Change Knowledge Portal (`climateknowledgeportal.worldbank.org`). In general, information on EIR and seasonality should relate to locations as close to the study location as possible. As countries are often heterogeneous in transmission (but also in relation to other covariates), national averages should only be used if no other information is available. Similarly, any additional data is preferred to be in temporal proximity to the study period. This search process must be diligently repeated for all covariates of interest.

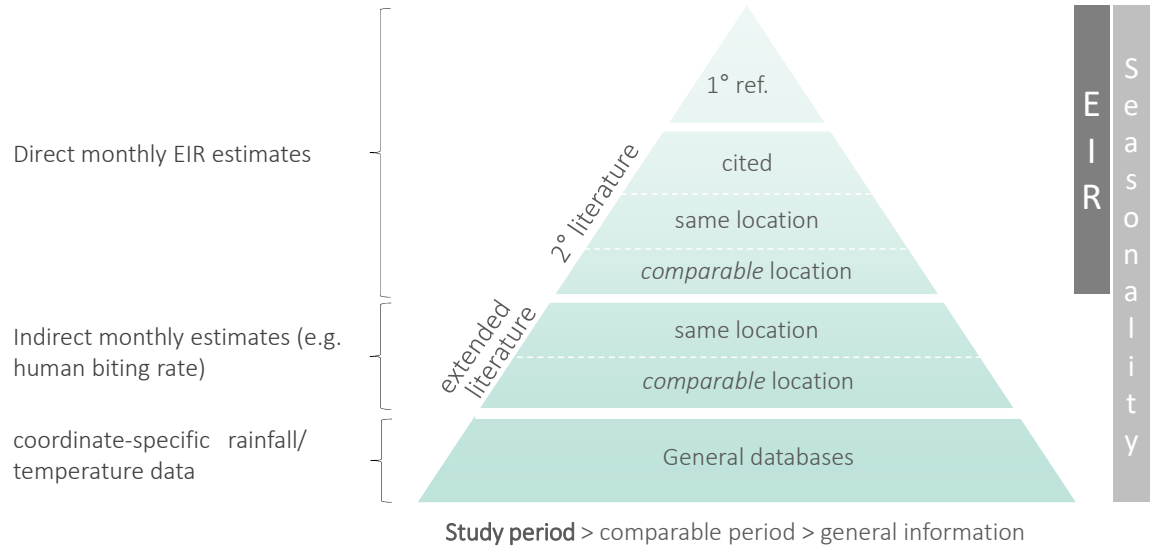

FIGURE 3.2: Case study: Hierarchy of information with application to the example of transmission intensity and seasonality.

A hierarchical framework ranking sources of information. This diagram shows a hierarchical framework for information on annual transmission intensity (EIR) and seasonality. Higher weighting should be attributed to estimates obtained from the primary reference or the directly cited literature. If no direct estimates are available, the extended literature should be screened for indirect (proxy) estimates from which seasonality can be inferred, such as monthly human biting rates. Lastly general databases could be consulted for more indirect proxy measures, e.g. for coordinate-specific monthly climate information.

3.3.2 Diagnostics and the care cascade

Care seeking behaviour, diagnostics and treatment carry important implications for morbidity, durations of infections, transmission and therefore the simulated epidemiological relationships. As *OpenMalaria* tracks parasite densities within the individual, diagnostic thresholds can be defined in parasites per microliter with given specificity according to the diagnostic tool (here: commonly thick blood smear microscopy).

While most models including *OpenMalaria* contain options for explicitly accounting for individual components of the care cascade, care seeking behaviour is, in *OpenMalaria*, most commonly defined through derivation from the effective care rate. This probability describes the probability of an infected individual receiving an effective curative malaria treatment within 14 days of infection and combines the probability that an infected individual seeks official care, adherence, compliance, treatment efficacy and potential resistance (*137*). Here, this value must account for active case detection (ACD) conducted during the study period, the background health system, and preventative chemotherapy. For studies conducted prior to the year 2000 (and thus prior to malaria awareness campaigns in the considered countries) we assumed low care seeking probabilities for uncomplicated malaria in the formal health sector unless indicated otherwise (*63*). For studies conducted after the year 2000, effective

care estimates were triangulated from the literature (e.g. (*137*)), country-specific Demographic Health Surveys or Malaria Indicator Surveys. In order to simulate the study conditions under which incidence was recorded, ACD was added at given and 100% coverage or at a coverage equivalent to the follow-up rate provided in the study. Additionally, efficacy estimates for first-line treatment and potential resistance levels in the study setting were extracted from the literature and factored into the final effective care estimates.

3.3.3 Interventions

For each setting, we consider its intervention history over the last ten years with regards to insecticide-treated net (ITN), distribution, long-lasting insecticide treated net (LLIN) use, indoor residual spraying (IRS), and intermittent preventive treatment in pregnancy (IPTp). Prior to the year 2000 we assumed no interventions unless indicated otherwise.

3.3.4 Acceptable levels of inaccuracies

In practice, it is often difficult to obtain direct, reliable estimates for every covariate. The researcher must therefore determine an acceptable level of indirectness or inaccuracy, which will depend on the covariate's likely influence on predictions. For example, monthly seasonality indices inferred only from rainfall data are likely somewhat inaccurate. However, if the annual transmission intensity is accurately captured in the simulations and the general seasonal pattern somewhat appropriately (e.g. highly seasonal vs. perennial transmission), the negative impact on predictions of the *Pf*PR-incidence relationship is likely low. This must be carefully considered on a case-by-case basis. These decisions will define a data quality threshold that must be met in order to retain a data record for calibration.

3.4 Case study: The malaria *Pf*PR-incidence relationship

Clinical incidence is an essential indicator to track progress on malaria control and elimination. However, accurate incidence measures are difficult to obtain, as explicitly measuring incidence requires long-term longitudinal studies with resource intense active case detection (*138*). For the case of clinical incidence, estimates often rely on routinely collected data (e.g. health facility records) that is prone to biases from incomplete reporting and relies on statistical adjustments (*53*, *138*). The sparsity, inaccessibility and unreliability of incidence data impose limitations on its value and usability in comparison to prevalence data. Prevalence (*Pf*PR) is more frequently measured and data more readily available. Therefore, cartographic approaches that estimate case incidence from densely available prevalence estimates and geospatial interpolation between sparse but reliable, matched incidence and prevalence estimates, are frequent (*139*). However, these methods rely on the ability to estimate incidence from *Pf*PR, which requires a deep, mechanistic understanding of this relationship (*53*, *140–142*). Purely

data-driven (statistical) approaches are limited in their ability to capture the complexity of the *Pf*PR-incidence relationship (*53*).

Detailed models of malaria transmission and control retain explicit mechanisms of the disease transmission and feature a broad range of contextual covariates. It has been previously shown that they can be calibrated specifically to capture the *Pf*PR-incidence relationship and generate age-specific predictions of incidence from *Pf*PR data (*53*). However, while outperforming purely data-driven approaches, substantial degrees of predictive uncertainty remain despite adjustment for frequency of ACD and diagnostic case definition thresholds (*53*).

Fig 3.3A shows the age-matched *Pf*PR-incidence relationship of all records relating to *P. falciparum* malaria in Africa included in (*53*). The raw relationship is highly valuable and shows only a weak positive correlation. In contrast, Figure 3.3B shows an exemplary selection of all data points relating to children aged <5 years and from studies conducted in Kenya, Mali, Senegal, or Tanzania. We conclude that the underlying setting-specific contextual covariates (such as transmission intensity, seasonality patterns or health system specifics) determine the *Pf*PR-incidence relationship and are thus important to be incorporated for subsequent analyses.

Two lessons can be drawn from these analyses: Firstly, it may prove valuable to integrate the *Pf*PR-incidence relationship into the calibration of detailed malaria models from the outset. Often, malaria transmission models incorporate the age-prevalence and age-incidence relationships separately (*3*, *4*), which may lead to poor depiction of age-specific *Pf*PR-incidence relationships. Secondly, age, care seeking probability, diagnostic thresholds and inferred transmission intensity may be insufficient to explain the variation in this relationship. During calibration, the deeper context of data collection should captured to account for the remaining variability. Explicitly incorporating the age-dependence and contextual complexity of the *Pf*PR-incidence relationship into disease simulators will improve their predictive accuracy and enable improved incorporation into decision-making processes that rely on accurate incidence quantification. In 2015, Battle et al. published a comprehensive database of age-matched *Pf*PR-incidence records (*140*). Here, we demonstrate the process of preparing these data for incorporation into the calibration of *OpenMalaria*.

3.4.1 Simulator of malaria transmission and control

OpenMalaria (`https://github.com/SwissTPH/openmalaria.wiki.git`) is an individual-based stochastic modelling platform of malaria transmission and control. OpenMalaria features within-host parasite dynamics, the progression to and of clinical disease, development of immunity, individual care seeking behaviour, vector dynamics and pharmaceutical and non-

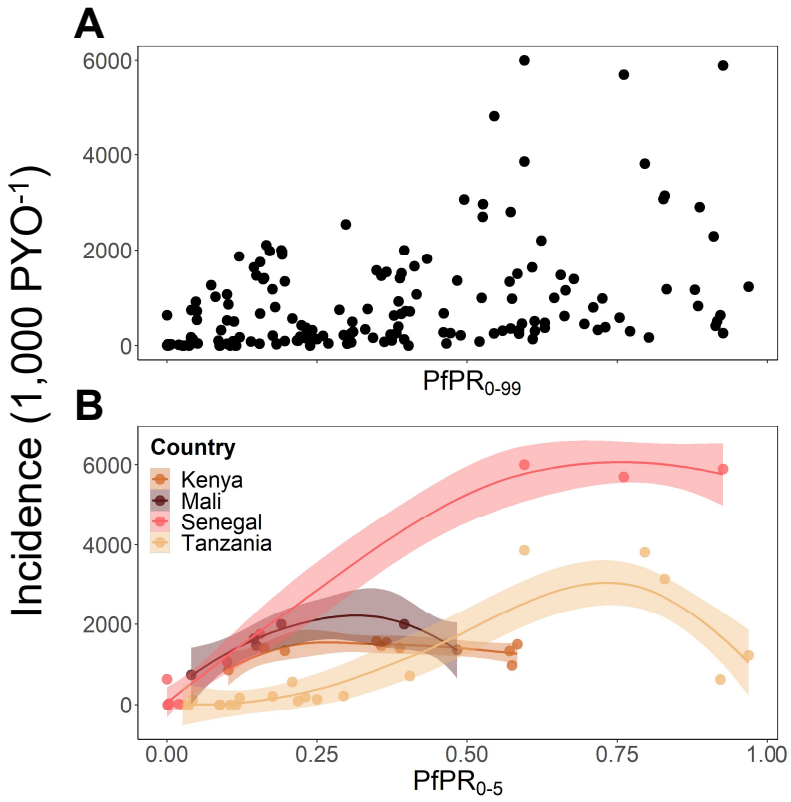

FIGURE 3.3: The *Pf*PR-incidence relationship in data records included by Cameron et al. 2015 (*53*).

A. Case incidence (events per 1,000 person-years observed) in relation to age-matched prevalence across all ages. The data from all selected records are jointly plotted here. **B**. Case incidence (events per 1,000 person-years observed) in relation to age-matched prevalence for children under 5 years old and obtained from studies conducted in Kenya, Mali, Senegal or Tanzania. These countries were selected as sufficient data were available for plotting.

pharmaceutical antimalarial interventions at vector and human level (*3*, *9*, *52*, *136*). Previously, the model has been calibrated to 11 objectives representing different epidemiological outcomes, including age-specific prevalence and incidence patterns, age-specific mortality rates and hospitalization rates (*3*, *9*, *52*, *136*).

3.4.2 Record Selection and Quality Control

Starting from the previously published database of age-matched prevalence-incidence records (*140*), we selected records relating to *P. falciparum* malaria in Africa based on the inclusion and exclusion criteria determined by Cameron et al. (*53*). Additionally, we required that prevalence records be available as direct measurements from cross-sectional studies rather than statistically estimated. Primary references reporting the prevalence and incidence records had to be publically available in English or French and all contextual covariates of interest had

to be retrievable. We selected 17 of the 32 records included by Cameron et al. (*53*), representing 10 separate studies. Exclusion reasons for the excluded studies are provided in the supplement.

We applied our translational framework and the scientific approaches for covariates in our case study to the *Pf*PR-incidence relationship records included by Cameron et al. (*53*). We discarded 15 out of the 32 original records for failing out our quality criteria. A list of the excluded records and the reason for exclusion are provided in Table B.1. For the remaining 17 records, we referred to 35 primary and secondary references. Additionally, multiple general databases and surveys including the Demographic Health Surveys and Malaria Indicator Surveys were consulted for all seven countries in which the studies were conducted.

3.4.3 Case study database: *Pf*PR-incidence records and contextual covariates

A summary of all records and our collated information is provided in Table 3.1. Further details including the data extracted from Battle et al. (*140*) and seasonality profiles for all records are provided in the supplement (Appendix B, see overview B.2).

3.5 Discussion

Calibrating disease transmission models to historical observational data is challenging as it relies on holistically representing the context under which the data were collected in a modelling framework. This involves building a complete picture of the time, location and methodology of data collection - information that can be difficult to retrieve given the historical nature of the data. Here, we present a generalizable framework for the incorporation of epidemiological data into the calibration of disease transmission models. We showcase an example of collating data on the *Pf*PR-incidence relationship and preparing it for incorporation into the calibration of an infectious disease transmission simulator, *OpenMalaria*. The resulting library provides sufficient information on contextual covariates to reproduce these source studies in most models of malaria transmission. This includes information on transmission intensity, seasonality patterns, health care system specifics and interventions. Our framework provides guidance on comprehensively representing data collection settings and can be modified or extended as required for other epidemiological relationships.

As case incidence informs the tracking of epidemiological trends and progress in control, which in turn influence decisions on funding of interventions, it is crucial that the models supporting the generation of evidence capture it well. We therefore encourage explicitly calibrating malaria models to the *Pf*PR-incidence relationship. Our work shows this relationship is strongly influenced by a variety of covariates including the sample population and the geographical, temporal, and methodological context of data collection. Therefore, this context must be explicitly accounted for during calibration to avoid models that are (computationally and mathematically) functional but potentially epidemiologically inadequate. Our case study

TABLE 3.1: Library of contextual information for age-matched *Pf*PR-incidence records

Study	Location	ACD period	duration	freq.	PfPR surveys	seasonality (REF)	EIR (Ref)	Vectors	diagnostic threshold	Effective care	1st line	resistance	interventions
Ba et al. (*143*)	Ndiop, Senegal	06/1993-05/1994	12m	1	monthly	highly seasonal (*144*)	63 (*144*)	AG (66%), AA (31%), AF (3%)(*144*)	2 p ml-1 (*143*)	0.04 (local, lack of information)	QN (*143*)	NA	NA
Bonnet et al. (*145*)	Koundou, Cameroon	03/1997-09/1998	18m	1	monthly	highly seasonal (*145*)	17.7 (*145*, *146*)	AM (47.7%), AG (47.2%), AF (5%) (*146*)	1000 p ml-1 (*145*)	0.04 (local, lack of information)	AQ (*145*)	0.23 (*145*, *147*)	NA
	Ebolakounou, Cameroon	03/1997-09/1998	18m	1	monthly	highly seasonal (*145*)	176.1 (*145*, *146*)	AG (100%)(Meunier et al. 1999)	1000 p ml-1 (*145*)	0.04 (local, lack of information)	AQ (*145*)	0.23 (*145*, *147*)	NA
Dicko et al. (*148*)	Douneguebougou, Mali	06/1999-11/1999; 06/2000-11/2000	12m	7	monthly	seasonal (*148*)	167.2 (1999), 137.3 (2000) (*148*)	AG (91.6%), AF (8.4%)(*148*)	Any patent (*148*)	0.268 (national, DHS 2001)	SP (*148*)	2-7%(*149*, *150*)	NA
	Sotuba, Mali	06/1999-11/1999; 06/2000-11/2000	12m	7	monthly	seasonal (*148*)	12.5 (1999), 3.64 (2000) (*148*)	AG (99.7%), AF (0.3%)(*148*)	Any patent (*148*)	0.268 (national, DHS 2001)	SP (*148*)	2-7% (*149*, *150*)	NA
Henry et al. (*151*)	Katiola "R0", Côte d'Ivoire	03/1997-02/1998	40d (8x5d)	1	19.04.1997, 27.4.1997, 11.6.1997, 3.9.1997, 19.10.1997, 26.11.1997, 21.1.1999	perennial with seasonal peaks (*151*)	158 (*151*)	AG (>98%)(*151*, *152*)	Age specific (see data) (*151*)	0.04 (local, triangulated from (*151*))	CQ (*151*)	11% (in Djebounoua 1993) (*153*, *154*)	NA
	Korhogo "R1", Côte d'Ivoire	03/1997-02/1998	40d (8x5d)	1	19.04.1997, 27.4.1997, 11.6.1997, 3.9.1997, 19.10.1997, 26.11.1997, 21.1.1999	perennial with seasonal peaks (*151*)	139 (*151*)	AG (>98%) (*151*, *152*)	Age specific (see data) (Henry et al. 2003)	0.04 (local, triangulated from Henry et a. 2003)	CQ (*151*)	11% (in Djebounoua 1993) (*153*, *154*)	NA
	Korhogo "R2", Côte d'Ivoire	03/1997-02/1998	40d (8x5d)	1	19.04.1997, 27.4.1997, 11.6.1997, 3.9.1997, 19.10.1997, 26.11.1997, 21.1.1999	perennial with seasonal peaks (*151*)	155 (*151*)	AG (>98%)(Zogo et al 2019, Henry et al. 2003)	Age specific (see data) (*151*)	0.04 (local, triangulated from Henry et a. 2003)	CQ (Henry et a. 2003)	11% (in Djebounoua 1993) (*153*, *154*)	NA
Lusingu et al. (*155*)	Mgome, Tanzania	04/2001-09/2001	6m	30	4/2001, 07/2001, 09/2001	seasonal (rainfall + 1 month)(*156*)	91 (*156*)	AG (90.2%), AF (9.8%) (*156*)	5000 p ml-1 (*155*)	0.04 (local, triangulated from Lusingu et al 2004)	SP (*155*)	70-80% (*157*, *158*)	ITNs: 7.1% (*155*)
	Ubiri, Tanzania	04/2001-09/2001	6m	30	04/2001, 07/2001, 09/2001	seasonal (rainfall + 1 month)(*156*)	0.9 (*156*)	AG (94.4%), AF (5.6%) (*156*)	1000 p ml-1 (*155*)	0.04 (local, triangulated from (*155*))	SP (*155*)	70-80% (*157*, *158*)	ITNs: 2%, prevention: 0.4% (*155*)
	Magamba, Tanzania	04/2001-09/2001	6m	30	04/2001, 07/2001, 09/2001	seasonal (rainfall + 1 month)(*156*)	0.03 (*156*)	AG (88.6%)), AF (11.4%)(*156*)	40 p ml-1 (*155*)	0.04 (local, triangulated from (*155*))	SP (*155*)	70-80% (*157*, *158*)	ITNs: 5.5%, IRS:1.2%, prevention: 0.4% (*155*)
Mwangi et al. (*159–161*)	Ngerenya, Kenya	05/1999-05/2001	24m	7	07/1999, 03/2000, 07/2000, 10/2000, 03/2001, 06/2001	seasonal (Mukombe/Kaoyeni average from (*162*))	10 (*159*)	AG (50%), AF (50%) (*163*)	Age specific (see data) (*159*)	0.04 (local, lack of information)	SP (*159*)	0.683	ITNs: 69% (children <10, 1999), 64% (population level, 1996), 96% (population level, 1993), <6% (pre-1993) (*159*, *164–166*)
	Chonyi, Kenya	05/1999-05/2001	24m	7	07/1999, 03/2000, 07/2000, 10/2000, 03/2001, 06/2001	seasonal (Assumed same as Ngerenya)	33.2 (triangulated from (*159*, *162*, *163*))	AG (95%), AF (5%) (*162*)	Age specific (see data) (*159*)	0.04 (local, lack of information)	SP (*159*)	0.683	6% ITNs in children <10 (1999) (*159*)
Saute et al. (*167*, *168*)	Manhiça, Mozambique	12/1996 - 07/1999	31m	7	02/1997, 10/1997, 10/1998, 02/1999	perennial with seasonal peak (Mendis et al 2000)	15 (*169*, *170*)	AG (27.7%), AF (72.3%) (*169*, *170*)	Any patent(*168*)	0.04 (local, lack of information)	CQ (*168*)	0.3 (*171*)	NA
Schellenberg et al. (*172*)	Ifakara, Tanzania	07/2000-06/2001	12m	7	weekly	perennial with seasonal peak (*173*)	29 (*173*, *174*)	AG (20%), AF (80%)(triangulated from (*173*))	Any patent citeschellen-bergIncidence-Clinical-Malaria2003	0.33 (local, triangulated from citeschellen-bergIncidence-Clinical-Malaria2003 (of measured fevers, one third were detected through PCD))	SP(*172*)	35.8% (64.2% efficacy) (*172*)	ITNs: 61% (children) (*172*)
Thomposon et al. (*175*)	Matola, Mozambique	12/1992-06/1995	31m	1	12/1993, 02/1994, 04/1994, 06/1994, 08/1994, 10/1994, 12/1994, 02/1995, 04/1995, 06/1995	perennial with seasonal peak (*176*)	20 (*175*)	AF (55%), AA (45%)(*176*)	Any patent (*175*)	0.04 (local, lack of information)	CQ (if failure on day 7: SP) (*175*)	11% (min of 2002-2005)	NA
Trape et al. (*177*)	Dielmo, Senegal	01/2007-07/2008	18m	1	10/2007	seasonal (*177*)	232 (2007), 155 (2008) (*177*)	AF (30%), AA (70%) (visual estimation from (*177*))	Any patent (*177*)	1 (local , (*177*))	ACT (*177*)	0 (NA)	NA

provides both a flexible guiding framework for the incorporation of this data and the data itself, paving the way to models that are able disentangle the heterogeneities observed in the *Pf*PR-incidence relationship. By adapting a similar methodology, this framework can easily be extended or modified for application to other epidemiological relationships and diseases. Importantly, for malaria this includes severe disease and mortality relationships. Notably for all-cause mortality these could for example include the malaria transmission intensity and mortality burden across Africa (MTIMBA) datasets: (*178–181*).
More broadly, we illustrate the intricacies and complexities of calibrating disease transmission simulators to historical data, which requires intense search efforts of the extended biological and epidemiological literature and databases to piece together the needed information. This process reveals the two core issues of calibration to historical data: Firstly, the trade-off between data quality and availability and secondly, the time required to generate a deep enough holistic understanding of the individual records to evaluate them. If this is glossed over, models and their predictions are at danger of being inaccurate or biased. It is therefore important to define and adhere to a strategy for addressing the calibration to historical data. Our work here aims to provide such a strategy for future reference by other research groups.

In disciplines where calibration data is mostly generated in controlled laboratory experiments, it has been argued that quality is more important than abundance (*182*). While data quality must be assured, the trade-off between quality and abundance is more nuanced in the field of disease modelling. The calibration of infectious disease transmission models relies largely on real-world observational data. In observational studies, external influences are less controllable than in lab experiments. This must be accounted for when using these data to inform models. In an ideal world, all data should be plentiful, accurate and unbiased. However, considering the scarcity of epidemiological data and the extensive resources required to conduct observational studies, we may be more lenient with quality criteria, so long as the shortcomings are considered and communicated: Sometimes, incorporating *some* data is better than none at all. For example, in three of the four big malaria simulators (EMOD DTL, MORU and OpenMalaria) the duration of natural malaria infection and within-host dynamics of *P. falciparum* malaria infection are calibrated to malariatherapy data (*89*, *183*, *184*). These data include parasite densities over time in malaria-naive African American individuals who were treated for neurosyphilis through infection with *P. falciparum* (*185*). This data exhibits substantial inter-individual variation in parasite densities and temporal dynamics of infection, dynamics observed in naive individuals may not be representative of the endemic settings that are commonly modelled and the data is of ethically questionable origin (*185*). Yet, as experimental infection without administering curative treatment is no longer possible, these are and will be the only data of its kind. It would also be possible to calibrate using the data from longitudinal studies of parasite genotypes (see e.g. (*186*)). This would, however, be an additional challenge. The use of malariatherapy data for calibrating within-host models of malaria illustrates the dilemma of sparse data in infectious diseases. Potential biases in single datasets enhance the importance of incorporating diverse data from different sources and

covering a multitude of relationships such that they jointly set a foundation for a reliable model.

Unfortunately, it is not always the case that publications are transparent on how well data was incorporated during model calibration. While the records themselves are often provided in the methodological explanations, *how* they were incorporated is often unclear. The presentation of a new model usually emphasises its functionality, highlighting its distinguishing features, elegant methodological advances and new possibilities provided by the model. In light of this, it is easy to forget that a model is grounded in reality by its calibration to data. If these foundations are fragile, the predictions are not necessarily credible. However, as we have illustrated here, setting a solid (data) foundation cam be time-consuming and a generalised approach does not exist. The lack of a generalized framework means that what is relevant for inclusion is decided on a case-by-case basis by the researcher performing the calibration. This compromises comparability between findings generated by different models and makes it difficult to discern whether differences in predictions are attributable to the model's assumptions, structural features, or calibration methods including how calibration data was incorporated.

We therefore emphasise the need for shared databases of calibration data. The ultimate goal of modelling in a public health context is to generate a sufficient evidence base for decision-making to improve population health. Shared, public databases on methodologies initiate dialogue and collaborations which improve research quality and ensure transparency around the possibilities as well as the limitations of modelling for public health division making. The resulting discussions on data and methodology must be openly communicated, including discussions around data and model uncertainties. For example, for each of the records presented here, we present point estimates of each of the contextual covariates rather than confidence or credible ranges. While practical because it limits the time required to research around each covariate, this is also crude. We hope that the large portfolio of records included for this and other relationships will outweigh the potential biases introduced by each individual record.

Overall, the results presented here carry important implications for model-based evidence in public health decision-making: the usefulness and credibility of model predictions relies on its calibration to data and on how well the contextual complexities of data collection are represented. We present a new dataset for the calibration of malaria simulators to the intricate *Pf*PR-incidence relationship. This data was collated under a clear methodological framework, is filtered for quality and provides contextual covariates for all records. Not capturing the context of collection results in an incomplete picture. This can be misleading because the calibration problem is still mathematically solvable but will yield results that may be biologically and epidemiologically inaccurate at best or at worst wrong predictions. Diversity in data and context and their accurate representation ensure flexible models that

can be adapted to both generalised and specific policy questions and meet the credibility requirements imposed by the weight of public health responsibility.

4 Data needs for calibration and implications for predictions: Insights for the future of *OpenMalaria*

In Chapters 2 and 3, I considered the two building blocks of model calibration: optimisation algorithms and real-world data. Building on supplementary analyses in context of the previous chapters, this chapter provides a link between a model's foundations in calibration and structure and its later uses in simulating epidemiological relationships and prediction. This chapter is not being considered for submission to a journal, however it represents an important basis of documentations and next steps for the future of *OpenMalaria*.

Section 4.1 provides background information on the development of *OpenMalaria*, from its original purpose and structural assumptions to its uses today. Section 4.2 connects the structural model assumptions to the choices made during calibration (on data and algorithms) and to their joint implications for prediction. Section 4.3 of this chapter provides a discussion of *OpenMalaria*'s power and limitations and more generally of the implications of the choices made during development on a model's predictive abilities. Finally, section 4.4 discusses requirements for the future.

4.1 *OpenMalaria*

4.1.1 Purpose of the model

In context of its original publication in 2006, *OpenMalaria* was described as *perhaps the most ambitious attempt ever made to apply mathematical modelling to understanding malaria* (Kevin Marsh, in the preface to Am. J. Trop. Med. Hyg. 75 (Supp2), 2006 (*187*)). It was originally developed for predicting the epidemiologic and economic effects of malaria vaccines at individual and population level (*63*). From the outset, it therefore featured options for simulating generalised pharmaceutical interventions at human level, including vaccines and chemotherapy. Since then, the applications of *OpenMalaria* have been expanded substantially. Additional functionalities allow for the simulation of dynamics of malaria in the mosquito vector and of vector interventions. Alternative model structures represent varying assumptions about immune decay, transmission heterogeneity, and treatment access (*64*). These stand to supplement the

core of the model, the simulation of transmission, infection, and disease progression of the human. Beyond investigations surrounding vaccine development, efficacy, effectiveness, and deployment strategies (*11*, *64*, *188–192*), *OpenMalaria* has a wide range of application and was used to model the impact of case management decisions and (non-vaccine) drug-based interventions (*59*, *193*, *194*), vector control (*195–200*), integrated strategies and national programs (*201–207*), and for transmission and burden of disease estimation (*208–210*) (see also: `github.com/SwissTPH/openmalaria/wiki/References`).

OpenMalaria is now being used to investigate a large variety of questions from epidemiological and biological basic research to informing public health policy (*49*, *60*). Modelling decisions around the core assumptions on biology and transmission of malaria influence downstream morbidity and mortality estimates that public health decision-making hinges on. It is vital to understand them.

4.1.2 Technical summary

To provide context for the rest of the chapter, this section gives a brief summary of the simulation of morbidity, mortality, immunity, and interventions in *OpenMalaria*. A more detailed, technical description is provided in Appendix A.

OpenMalaria features discrete individual-based stochastic simulations of malaria in humans in 5-day time steps. Every infection and individual are characterised by a set of continuous state variables, namely, parasite density, infection duration, and immune status. Key processes and relationships influencing the simulated course of infection and onwards transmission include the attenuation of inoculation, acquired pre-erythrocytic immunity, acquired blood-stage immunity, morbidity (acute and severe) and mortality (malaria-specific and indirect), anaemia, and the infection of vectors as a function of parasite densities in the human. Other model components include a vector model and a case management system. All individual components have previously been well documented (*3*, *64*). A visual summary of the model with references to further details on each component is provided in Figure 4.2.

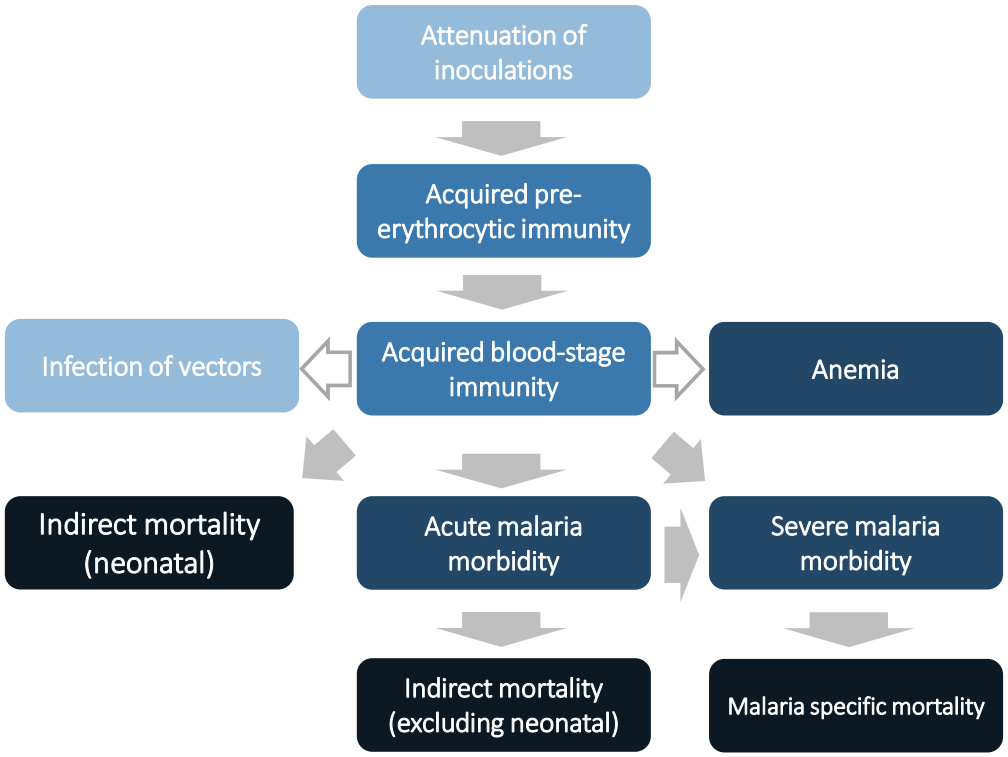

FIGURE 4.1: Visual summary of OpenMalaria with references to original publications on the model components

Adapted from Smith et al. 2006, Fig.3 (*63*). References from top to bottom and left to right: Attenuation of inoculations (*183*), Acquired pre-erythrocytic immunity (*183*), Infection of vectors (*211*, *212*), Acquired blood-stage immunity (*31*), Anemia (*213*), Indirect mortality (neonatal) (*214*), Acute malaria morbidity (*3*), Severe malaria morbidity (*215*), Indirect mortality excluding neonatal (*215*), Malaria specific mortality (*215*)

Infection of the human host

The course of infection in the human host in the absence of treatment is simulated in *OpenMalaria* as summarised in Figure 4.2.

Infection in the human occurs through the bite of an infectious mosquito. The seasonal pattern of entomological inoculation rate (EIR) determines the seasonal pattern of transmission. In the base model, the expected number of entomological inoculations experienced by an individual at any time is dependent on the annual transmission intensity and the individual's availability (determined by its age-dependent body surface area). The parasite densities in each individual are modified by natural or acquired immunity and interventions (*63*).

Clinical incidence, acute (uncomplicated) and severe morbidity

The duration and log density of each infection are sampled from normal distributions parameterised against malariatherapy data (*31*, *185*). Multiple concurrent infections in the same

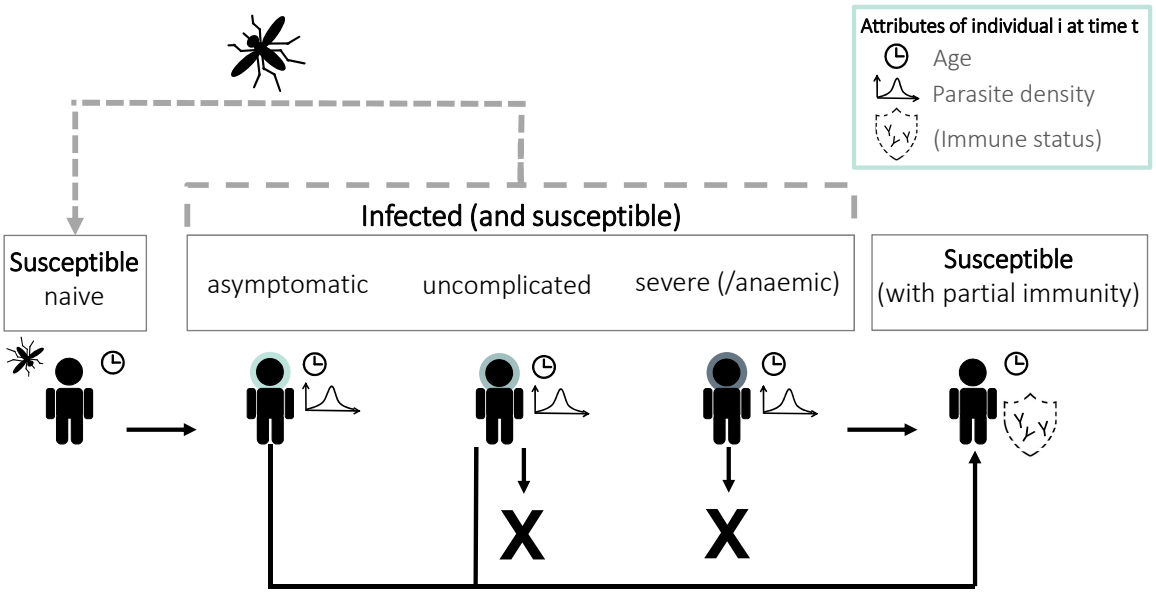

FIGURE 4.2: Simplified summary of the simulated course of infection in *OpenMalaria*

An individual is infected through the bite of an infectious mosquito. The individual may or may not develop symptoms of varying severity. The severity of disease depends on the individual's age and parasite densities. The parasite densities vary stochastically but on average decrease over the course of an untreated infection. Notably, there is no necessary progression from asymptomatic to symptomatic to severe. In fact, severe disease is more likely to occur early on during an infection. Individuals may die from malaria with probabilities dependent on age, severity and hospitalisation status. While infectious, the individual may also infect susceptible mosquitoes. Every (re-)infection leads to the build-up of (partial) immunity on recovery from an acute infection. Upon recovery, individuals may be re-infected but an individual's immune status influences its susceptibility, parasite densities on re-infection, and therefore the severity of symptoms. Already infected individuals may acquire additional infections with different strains of the parasites if bitten and infected again (multiplicity of infection). Thus, all individuals are susceptible, though some are more susceptible than others.

individual at the same time are possible. To simulate the clinical state of an individual at time t , for each five-day time step, independent samples from the simulated parasite density distribution are drawn for each concurrent infection and day. An episode of acute morbidity occurs in individual i at time t with a probability P_m that depends on the pyrogenic threshold (defined by multiple calibration target parameters) and the maximum density of the five daily densities sampled during the five-day interval (*3*).

Two different classes of severe episodes are considered by the model, denoted B_1 and B_2: The probability of progression to severe disease from an uncomplicated acute episode depends on the individual's clinical status as defined by its parasite density in one time step (episodes of class B_1) or the probability of a secondary insult occurring during an uncomplicated clinical

episode (episodes of class B_2). Both probabilities depend on multiple parameters that are targets of the calibration process (*215*).

Mortality

The predicted malaria mortality is the sum of the hospital and community malaria deaths. Malaria deaths in hospital are a random sample of admitted severe malaria cases, with the age-dependent hospital case fatality rate, derived from the data of Reyburn et al. (2004) (*215*, *216*) and the odds ratio for death in the community compared to death in in-patients is an age-independent constant to be calibrated (*215*). The model considers both direct malaria-attributable deaths and indirect (all-cause) mortality (*215*).

Immunity

OpenMalaria considers effects of innate and acquired immunity on parasite densities. In the base model variant, the effects of innate immunity, antigenic variation, and variant independent anti-merozoite immunity are implicitly captured by forcing average parasite densities to vary over the course of an infection, using a statistical model fitted to malariatherapy data. The resulting value of the expected (log) parasite density of infection j in individual i at time t is further adjusted to account for the effects of acquired immunity. The acquired immunity effects incorporate the cumulative density of asexual parasitaemia experienced by individual i since birth and the cumulative number of prior infections. The base model variant assumes no decay of acquired immunity (*31*). Infants are further protected through maternal antibodies. This effect is independent of maternal exposure and decays over time (*31*).

Interventions

OpenMalaria provides options for simulating a multitude of existing and hypothetical interventions targeting different stages of the parasite life cycle. These range from pharmaceutical interventions, such as vaccines with different modes of action (pre-erythrocytic, blood-stage, or transmission blocking), mass drug administration (MDA) or seasonal malaria chemoprevention (SMC), to vector control interventions (like insecticide treated nets (ITNs)/ long-lasting insecticidal nets (LLINs) or indoor residual spraying (IRS)). Within a simulation, an intervention introduces a change in the state of the simulation from the time of deployment and by a given magnitude (e.g. reduces transmission or changes the state of (a group of) simulated hosts). The addition of any new intervention requires an estimate of its effects different parts of the parasite life cycle. If available, calibration to real-world data is preferable (e.g. for RCD see (*50*) and Chapter 5. In the absence of real-world data, as is the case for interventions currently under development, intervention characteristics can be informed by preliminary clinical data or expert opinion.

4.2 The sensitivity of epidemiological predictions to calibration decisions: algorithms and data

The data needed for calibration are contingent on the model's intended uses. To guide public health decision-making on malaria, morbidity and mortality must be adequately captured. The choices made during the calibration of a complex disease transmission simulator like *OpenMalaria* may carry wide-reaching consequences for model predictions. This section provides a deeper analysis of *OpenMalaria*'s ability to capture disease biology and epidemiology through calibration to real-world data. On the example of severe disease and direct mortality, this section specifically investigates

1. the sensitivity of *OpenMalaria*'s goodness-of-fit and predicted epidemiological relationships to the calibration algorithms,
2. potential structural limitations of the model itself,
3. the implications of unreliable points in the calibration data, and
4. the role of uncertainty in data for stochastic events.

4.2.1 Incidence of severe disease: Structural insights and sensitivity to algorithm and data choices

Algorithm choice and structural insights.

Numerically, a similar overall goodness-of-fit was achieved by the different optimisation algorithms (see Chapter 2). Yet, the choice of algorithm impacts epidemiological predictions. Figure 4.3 shows *OpenMalaria*'s fit to Objective 8, the hospitalisation rate in relation to prevalence in children aged 0-9 years. Both of the Bayesian optimisation (BO) algorithms (see Chapter 2) converged towards similar predictions. However, the predicted severe disease incidence under the BO-derived parameterisations is almost twice as high as those under the previous parameterisation with GA-O (see Chatpter 2 and (*64*)). This is carried forward into the epidemiological predictions on the relationship of severe disease incidence by age (Figure 4.4 and Appendix A).

The sensitivity of *OpenMalaria*'s severe disease predictions to the calibration algorithm raises questions about how well its epidemiology is captured. The sensitivity analyses in Chapter 2 and varying final parameter estimates between optimisation algorithms reveal additional insights into the differences between the parameterisations (see Appendix A). The objectives relating to severe disease are highly sensitive to a parameter describing a critical value of the cumulative number of infections, which is used in describing the effect of the number of prior infections on parasite densities (denoted X_h^* in previous model documentations including Appendix A or θ_4 for calibration in Chapter 2, see Equation A.12). This parameter is approximately the same for all parameterisations and its value was recovered during validation exercises, providing evidence of high identifiability (Figures A.19 and A.31). Other

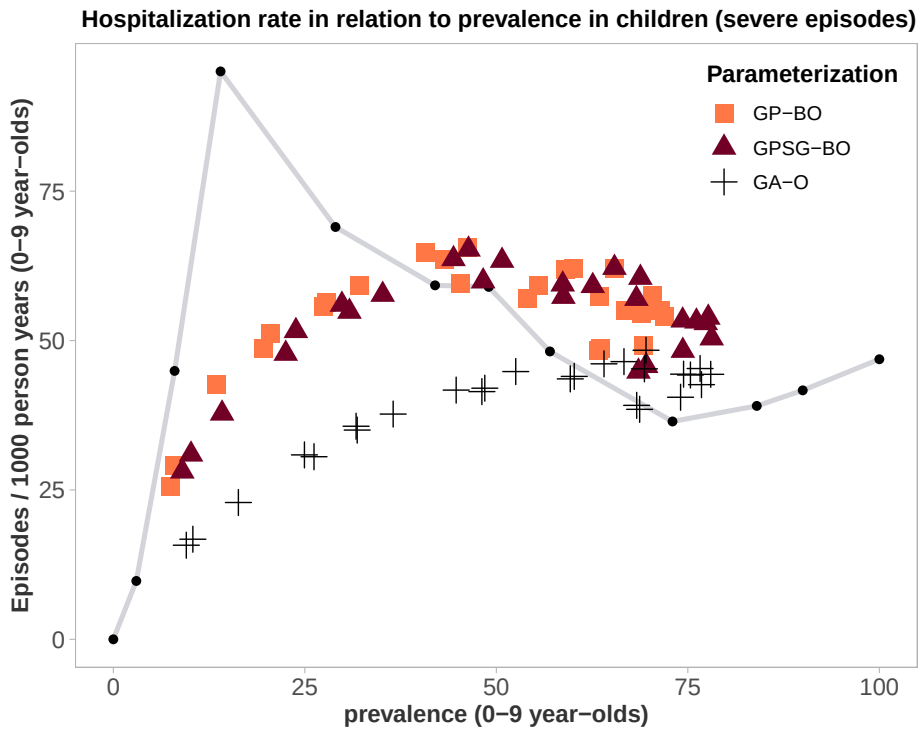

FIGURE 4.3: Objective 8: Hospitalisation rate in relation to prevalence in children

Final simulator fit to objective 8: *Hospitalisation rate in relation to prevalence in children* using the parameter sets from Gaussian process-based Bayesian optimisation (GP-BO) and Gaussian process stacked generalisation-based Bayesian optimisation (GPSG-BO) compared to the previous parameterisation (derived from optimisation with a genetic algorithm, GA-O). The data on the relative incidence of severe malaria-related morbidity and mortality in children <9 years old across different transmission intensities were originally collated by Marsh and Snow (1999) ((*217*). Measurements per age group were available as the relative risk of severe disease compared to a comparison age-group and the proportion/prevalence of severe episodes. A total of 26 entries on the relationship between severe malaria hospital admission rates and *P. falciparum* prevalence were included. Linear interpolation between data points was performed to obtain a continuous function relating hospital incidence rates to prevalence. To convert hospital incidence rates to community severe malaria incidence, the hospital admission rates was divided by the assumed proportion of severe episodes representing to hospital (48%). See A for a detailed description of the data.

influential parameters include a parasite growth parameter (α or θ_{10}, Equations A.25 and A.27), the parasitaemia threshold for severe disease of class B_1 ($Y^*_{B_1}$ or θ_{15}, see Equation A.28), and the critical age for comorbidity (a^*_F or θ_{17}, see Equation A.36), but none of these exhibited systematic differences between the BO and GA-O algorithms. Only for a parameter relating to the variance in parasite densities (σ^2_0 or θ_9 for calibration, see Equation A.14), did both BO algorithms yield higher values than GA-O. This, in combination with some observable differences in the predicted parasite densities (objective 3) between the BO parametrisations and GA-O may indicate that the differences severe disease estimates relate to differences in simulated

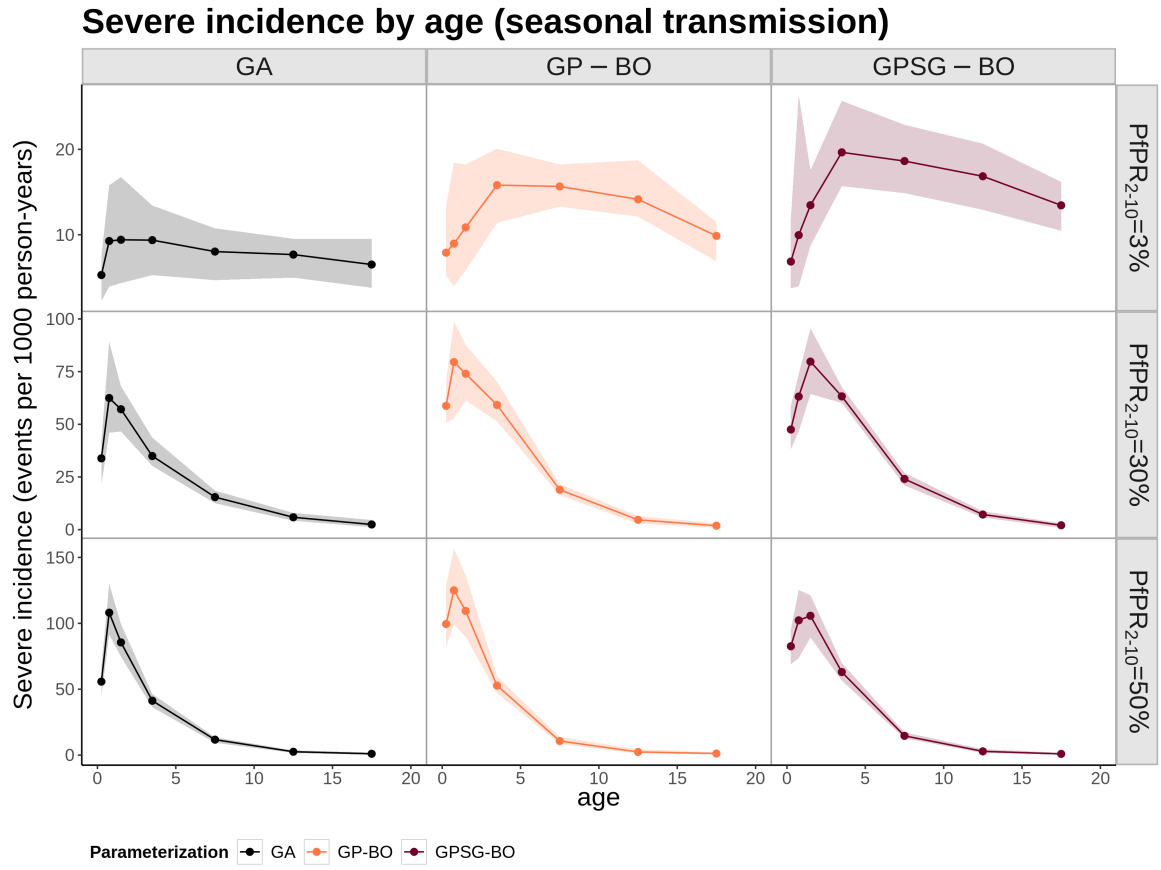

FIGURE 4.4: Yearly incidence of total severe malaria (seasonal)
Annual incidence of total severe malaria cases in a seasonal transmission setting as a function of age, displayed by transmission intensity (*Pf*PR$_{2\text{-}10}$) and parameterisation. Incidence is presented in terms of the yearly number of events per 1000 person-years. It is assumed that 48% of severe malaria cases seek official care at a heath care facility (hospital). The *Pf*PR$_{2\text{-}10}$ categories include simulated prevalences of 2.5-3.5% , 9-10% , 28-32% , and 47-53% labelled as 3% , 10% , 30% , and 50% , respectively.

parasite densities (see Appendix A, Figure A.10). However, the evidence is insufficient to draw well-founded conclusions and it is ultimately impossible to tell which parameterisation is most reflective of malaria biology. Within the calibration workflow, the parameter set with the lowest loss function value, F is chosen. However, a purely numerical decision may be biologically sub-optimal as the goodness-of-fit score is vulnerable to limitations in data quality.

Very low and very high prevalence settings.

Regardless of the optimisation algorithm, *OpenMalaria*'s ability to capture the data's relationship between prevalence and severe disease is limited (Figure 4.3), especially compared to its good fit for other objectives like the age-prevalence and age-incidence relationships (see Appendix A). Although the simulations for calibration to this objective covered a large range of EIRs (26 simulations with EIR$= 1 - 518$), neither very low (<10%) nor very high ($>$ 80%)

prevalences could be generated in children 0-9 years old. Possible explanations include structural limitations of the model itself, where the model assumptions and their mathematical translation are insufficient to capture the *true* disease biology, or insufficient calibration data for these settings or its incorporation.

In the quest for malaria elimination, mathematical models must be able to represent low-transmission settings well to justify their continued use in supporting decision-making. *OpenMalaria*'s fit to objective 8 on hospitalisation rates in relation to prevalence suggests that the model may be unable to capture such settings. The calibration to all other objectives relies mainly on data from medium to high transmission settings (see Appendix A. This is problematic because it makes any predictions on low or unstable transmission settings out-of-sample extrapolations of questionable accuracy. As increased malaria control efforts lead to decreases in transmission in many regions, more data on low transmission settings will become available that needs to be incorporated improve the robustness of predictions. The inclusion of the new *Pf*PR-incidence data provided in Chapter 3 into the calibration workflow could be a first step in this direction as it includes seven new record with an annual EIR of 20 or less.

4.2.2 Severe disease: Excluding unreliable data points

All parameterisations failed to reproduce the overall shape of the relationship between severe disease and prevalence as none captured the sharp peak in incidence at approximately 20%. This was already noted at the time of original publication (*215*). If the data point is truly representative of this relationship, *OpenMalaria* may be limited in its ability to depict natural infections. However, the data may not be a *true* representation of the actual relationship. It has been argued that this data point itself is unsuitable for inclusion in calibration as it stems from an area in Ethiopia where malaria is epidemic rather than endemic (*218*). Therefore, other research groups who developed comparably detailed malaria models decided to discard this point from calibration (*218*).

To assess the implications of including this data point for the calibration of *OpenMalaria*, additional calibration runs were performed dropping the Ethiopian data point from the calibration data. This recalibration was conducted using the GP-BO algorithm (GP-BO$_{\text{drop}}$), as this was the best performing calibration algorithm in Chapter 2.

An overall better goodness-of-fit value was achieved (weighted sum of the loss functions, $F = 54.3$ for GP-BO$_{\text{drop}}$ compared to 58.3 for GP-BO, 59.6 for GPSG-BO, and 63.7 for GA-O, Figure 4.5). This was to be expected as the goodness-of-fit statistic presents the weighted sum of the deviation between data and model predictions and one data point was dropped. For GP-BO$_{\text{drop}}$, a satisfactory fit was reached in iteration 3 after only 16

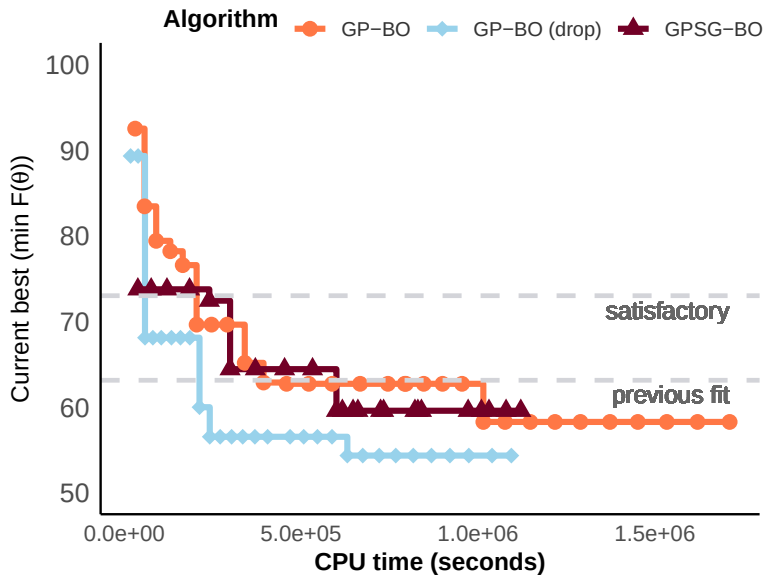

FIGURE 4.5: Comparison of algorithm convergence (GP-BO$_{drop}$)
Weighted sum of loss functions over 11 objectives associated with the current best fit parameter set by CPU time in seconds. Satisfactory fit of *OpenMalaria* refers to a weighted sum of loss functions value of 73.2 (as defined by Smith 2012 (*64*)). The previous best fit for *OpenMalaria* was achieved by the genetic algorithm had a loss function value of 63.7. GP-BO$_{drop}$ reaches its best fit of 54.3 in iteration 20 in $6.3e^5$ CPU seconds (~7 days) compared to 58.2 for GP-BO in iteration 21 in $1.02e^6$ CPU seconds (~12 days) and 59.6 for GPSG-BO in iteration 10 in $6.00e^5$ CPU seconds (~7 days). GP-BO$_{drop}$ = Gaussian process emulator Bayesian optimisation excluding contentious severe disease data point, GP-BO = Gaussian process emulator Bayesian optimisation, GPSG-BO = Gaussian process stacked optimisation emulator Bayesian optimisation.

hours, the GA-O fit was improved on within 2.5 days and the best fit was reached after approximately 7 days. GP-BO$_{drop}$ therefore reached lower values faster than GPSG-BO in the previous run. It is unlikely that this can be fully explained by the dropping of just one data point from the calculation. It is possible that the dropping of the contentious data point yields a severe disease-prevalence relationship that is more aligned with the biological assumptions made by *OpenMalaria* and that this speeds up convergence of calibration because it prevents the algorithm from trying to reproduce an unachievable severe disease relationship.

The fit for objective 8, which contained the discarded data point, was largely unchanged for GP-BO$_{drop}$ compared to previous calibration using GP-BO and GPSG-BO 4.6. However, the fit for some of the other objectives was affected. For example, the predicted relationship for the age-dependent multiplicity of infection changed, higher parasite densities were predicted for multiple of the calibration data sets, and direct mortality estimates were also increased. All model plots relating to the model fit are presented in Appendix C. This should not be a surprising result even though the dropped data point does not directly relate to the

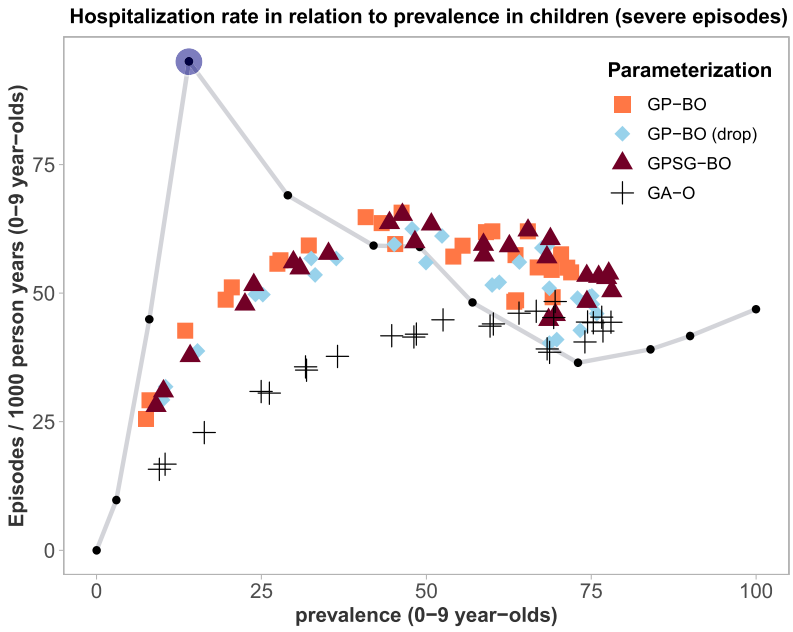

FIGURE 4.6: Hospitalisation rate in relation to prevalence in children (GP-BO$_{drop}$).
Final simulator fit using the parameter sets from GP-BO$_{drop}$, GP-BO and GPSG-BO compared to the previous parameterisation (derived from optimisation with a genetic algorithm, GA-O). The data point discarded for GP-BO$_{drop}$ is highlighted.

other affected relationships. There is strong interdependence of model parameters and the interconnectedness of the different model submodules. This further argues for simultaneous multi-objective calibration (as opposed to modular approaches) as implemented in the calibration of *OpenMalaria* (see Chapter 2, (*3*), and (*52*)).

It should be noted that the results presented here refer to only one realisation of the calibration process with each algorithm. It is therefore possible that the observed differences in overall loss function value, convergence time, and model fits are chance findings. Nonetheless the investigation around the implications of dropping the contentious severe disease data point from calibration may show either of two things: The currently used calibration data may be insufficient to identify one best parameterisation. There may exist multiple numerically similarly good fits and which one is achieved may be a matter of chance. Alternatively, the parametrisation to multiple of the fitting objectives may hinge on the presence or absence of just one contentious data point. Neither of these explanations are desirable and should prompt further investigations.

4.2.3 Mortality: uncertainty and scarcity

Direct, malaria-attributable mortality is captured by objective 10. The fit to this objective is shown in Figure 4.7. The data is characterised by great uncertainty and large confidence

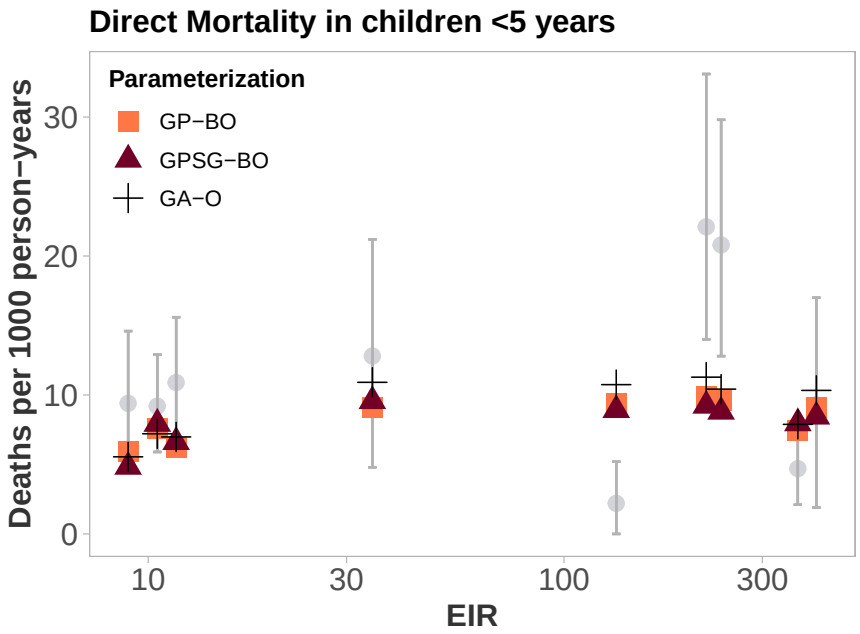

FIGURE 4.7: Objective 10: Direct mortality in children <5 years old
Final simulator fit to objective 10: *Malaria specific mortality in children (<5 years old)* in relation to annual transmission intensity (EIR) using the parameter sets yielded using GP-BO and GPSG-BO compared to the previous parameterisation (derived using optimisation with a genetic algorithm, GA-O). The data presents a subset of the data collated by Marsh and Snow (1999) (*217*), for which both malaria-specific mortality rates and seasonal transmission patterns were available. Mortality data were derived from verbal autopsy studies. There is one observation per study site and simulation scenario, and predicted values are for one survey at the end of 2 years. See A for a detailed description of the data.

intervals for every data point. On visual examination of the data, there is little evidence of a discernible relationship between EIR and malaria-related mortality in the data. While in the correct order of magnitude, the predicted relationship between EIR and mortality is flat, over- or underpredicting for some transmission intensities, yielding an overall poor model fit for this objective.

Yet, it is computationally possible to generate epidemiological predictions relating to mortality (Figure 4.8). Especially in seasonal low transmission settings ($Pf\text{PR}_{2\text{-}10}$<3%), the predicted relationship is highly variable between the different parameterisations and under the GA-O parameterisation shows large uncertainty bounds (4.8). The dataset for this objective includes 9 data points representing the malaria-specific mortality rates in children less than 5 years of age. From this scarce, uncertain calibration data, it is impossible to judge which (if any) parameterisation adequately captures direct mortality. It is therefore also not possible to infer which predicted epidemiological relationship is most representative of nature. In the absence of additional quantitative inputs or reference points, it is impossible to adequately evaluate

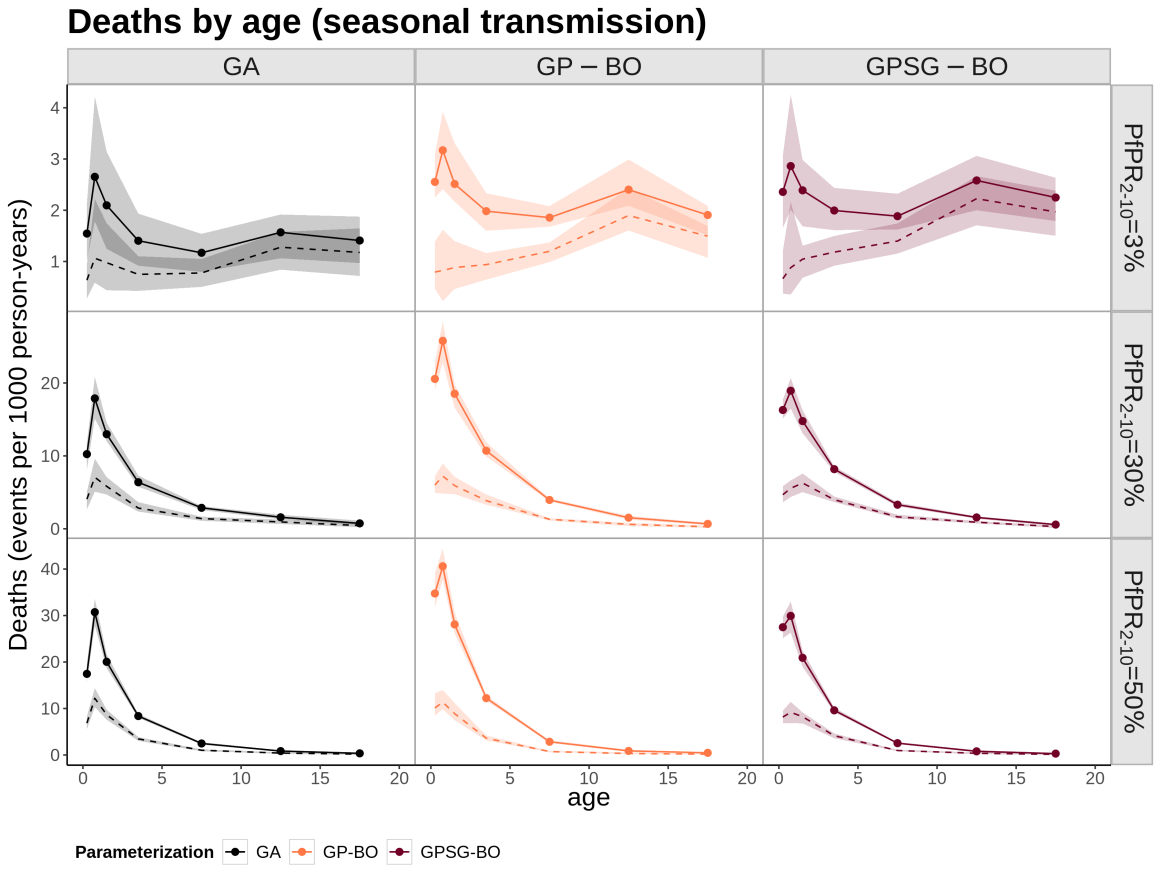

FIGURE 4.8: Yearly incidence of malaria-related deaths (seasonal)
Annual incidence of malaria-related deaths in a seasonal transmission setting as a function of age, displayed by transmission intensity ($Pf\text{PR}_{2\text{-}10}$) and parameterisation. Malaria mortality incidence is presented in terms of the yearly number of deaths in a population of 1000 individuals. The dashed estimates represent direct malaria deaths, and the solid all malaria deaths (including those attributable to co-morbidities). The $Pf\text{PR}_{2\text{-}10}$ categories summarise simulated prevalences of 2.5-3.5%, 9-10%, 28-32%, and 47-53% labelled as 3%, 10%, 30%, and 50%, respectively.

the predictions.

4.3 Insights on data needs and shortcomings

The currently used data on severe disease and both indirect and direct mortality is limited by scarcity and inherent uncertainties and overall insufficient. The data on the severe disease-prevalence relationship contains 10 data points (of which one is contentious) and direct mortality is calibrated to 11 data points with a large uncertainty for well-founded predictions. It is only possible to ensure that estimates are roughly in the correct order of magnitude.

When data is scarce, questionable, or unreliable, single data points can bias results and lead to biologically inaccurate calibration. The observable downstream consequences of inclusion decisions on the Ethiopian severe disease data point on the simulated epidemiological relationships (see Figures C.3-C.5 and C.11) mandate caution in defining quality criteria for calibration data. A guiding framework on possible decision points on inclusion and exclusion of data is provided in Chapter 3, but ultimately, no general rules exist and decisions must be made on a case-by-case basis influenced by abundance-quality trade-offs. The sensitivity of severe disease predictions to choices on data and algorithm are problematic as it ultimately limits the reliability and credibility of predictions. Further, all three datasets are limited to children (0-9 years old for severe disease, <5 years old for direct mortality, and the data on indirect mortality includes 1-11 month-old infants only). It is therefore questionable to what extend mortality or severe disease predictions and especially extrapolations outside the data, for example for other age groups, are justifiable.

The current calibration of *OpenMalaria* attributes higher weighting to the severe disease objective to balance out the data scarcity. Additional, reliable calibration data on the epidemiology of severe disease and direct and indirect mortality that has undergone careful quality control would be highly desirable to improve the robustness of *OpenMalaria*'s severe disease and mortality predictions. Notably for all-cause mortality, multiple additional candidate calibration datasets have become available since the original collation of calibration data for *OpenMalaria*. These include the malaria transmission intensity and mortality burden across Africa (MTIMBA) datasets ((*178–181*)). However, as illustrated in Chapter 3. It would be a challenging project to integrate these.

4.4 Requirements for the future

The development of a detailed disease transmission model that sufficiently and accurately replicates nature to support decision-making is an intricate and ambitious task. It hinges on the developers' biological and epidemiological understanding, technical ability, dedication, scientific diligence, and ability to filter out relevant information from a vast body of literature. In that sense, the development of *OpenMalaria* as a malaria model of unprecedented complexity, was ground-breaking and it remains a powerful tool in guiding thinking around malaria transmission and control. However, the previous three chapters have shown how the decisions made during the development of any model carry implications for its fit to data and ultimately its suitability and usefulness for predictions and supporting decision-making.

The potential shortcomings of a model can be mitigated through continued diligent model development and open communication around its assumptions. Calibration data must be handled with care. Chapter 3 has shown the dependence of the *Pf*PR-incidence relationship on contextual covariates. The case of the Ethiopian severe disease data point and the implications

of its inclusion or exclusion on model fit further illustrate the consequences data choices can have on predictions. Within the quality-abundance trade-off (see Chapter 3), sufficient quality and reliability of calibration data should be ensured and its incorporation should consider all potentially confounding covariates. To solidify the model's data foundations, it may be beneficial if tasks are performed by multiple researchers or research groups and their results compared. An additional step is external validation of a model and benchmarking against other comparable models. Lastly, wherever possible, the addition of more data can help increase model robustness. For *OpenMalaria*, this could in particular be achieved by including data from low transmission settings (e.g. the *Pf*PR-incidence library in Chapter 3), on severe disease (e.g. disaggregated or extended versions of the currently used data (*216*)) and on mortality (e.g. the MTIMBA datasets (*178–181*)).

The possible applications of modelling are plentiful and it is tempting to use it in all instances that are technically possible (i.e. yield numerical results). However, we must remind ourselves that there is no omnipotent model. How robust is *OpenMalaria* or any malaria model and how reliable its predictions? A general answer is difficult to provide. A model's suitability should evaluated on a case-by-case basis. For example, when calibrating to a specific setting (e.g for forecasting), it should first be shown that the model can capture the relevant disease epidemiology in the past before generating predictions for the future. Similarly, if a model is used to predict the impact of interventions, it must be ensured that these are also adequately captured through separate calibration of intervention properties to data, e.g. from clinical studies or RCTs. Broadly, we expect *OpenMalaria* predictions to be relatively accurate for comparative and relative analyses (such as relative prevalence or acute incidence reductions) in medium to high transmission settings. Substantial uncertainty remains especially for low transmission areas and around rare downstream outcomes. Predictions relating to severe disease or mortality should be handled with caution, as should any attempts at predicting absolute numbers (as these also strongly depend on the experimental set up). The analysis presented here focuses on how *OpenMalaria* captures the biology and epidemiology of malaria and does not address any intervention components of the model. Additional analyses should be conducted to investigate the ability of *OpenMalaria* (and other malaria models) to reflect the effects of intervention on transmission.

Perhaps we ought to ask ourselves if we even truly need to find *the best* model. Perhaps, the quest for a perfect a model is futile and we should rather work on sensible, purpose-driven development, on understanding the power but also the limitations of *OpenMalaria* (and other models) and on communicating these clearly. *OpenMalaria*'s original purpose was to *stimulate thinking and lead to new insight and provoke new studies in both the field and the lab* (*187*). Perhaps we ought to return to this mindset and remind ourselves to be careful around attributing additional abilities to a model. Especially when used in supporting decision-making due diligence is required to ensure that the model adequately captures the changes to important indicators of clinical cases, severe disease, and mortality induced by control interventions or

health system changes. This needs to include both biological and epidemiological processes as well as intervention properties

Ultimately, developing a detailed disease transmission model such as *OpenMalaria* falls outside the best scientific practices that any one field of research preaches and in many regards is somewhat of an art. The development of a model is a delicate process that requires careful consideration at each step and synthesising as much information as possible. It is more than a mere technical problem. It rather requires thought and a holistic understanding of biology, epidemiology, mathematics, statistics and good programming practice. Every model has structural limitations and predictive uncertainties are inherent to both the data and the model itself. Nonetheless, models can help to navigate complex systems of biology, epidemiology, health systems, and public health and to answer questions that are too multifaceted for thought experiments. It also should be noted that the discussions around the assumptions and limitations of *OpenMalaria* presented in this chapter could be carried out equally critically for other malaria models, likely revealing some of the same shortcomings and dependencies. *OpenMalaria*'s documentation is (particularly) transparent with publicly accessible, collated calibration data and methodology and a list of publications relating to the development of intervention plugins (`github.com/SwissTPH/openmalaria/wiki/References`), which readily provides the information required for analyses such as this one. The excursions in this chapter should therefore not distract from the scientific value of *OpenMalaria* and other infectious disease models but ought to be a cautionary tale to work diligently and critically and not attribute false accuracy to predictions outside the models designed purpose.

5 Modelling reactive case detection strategies for interrupting transmission of *Plasmodium falciparum* malaria

Theresa Reiker[1,2] , Nakul Chitnis[1,2], Thomas Smith[1,2,*]

[1] Department of Epidemiology and Public Health, Swiss Tropical and Public Health Institute, 4051, Basel, Switzerland
[2] University of Basel, Petersplatz 1, Basel, Switzerland

*Corresponding author

Email: thomas-a.smith@unibas.ch

Keywords: simulation, targeting, elimination, surveillance, response.

Publication:

Reiker T, Chitnis N and Smith TA. *Modelling reactive case detection strategies for interrupting transmission of Plasmodium falciparum malaria*. Malaria Journal. 2019; 18: 259

5.1 Abstract

Background. As areas move closer to malaria elimination, a combination of limited resources and increasing heterogeneity in case distribution and transmission favour a shift to targeted reactive interventions. Reactive Case Detection (RCD), the following up of additional individuals surrounding an index case, has the potential to target transmission pockets and identify asymptomatic cases in them. Current RCD implementation strategies vary, and it is unclear which are most effective in achieving elimination.

Methods. *OpenMalaria*, an established individual-based stochastic model, was used to simulate RCD in a Zambia-like setting. The capacity to follow up index cases, the search radius, the initial transmission and the case management coverage were varied. Suitable settings were identified and probabilities of elimination and time to elimination estimated. The value of routinely collected prevalence and incidence data for predicting the success of RCD was assessed.

Results. The results indicate that RCD with the aim of transmission interruption is only appropriate in settings where initial transmission is very low (annual entomological inoculation rate (EIR) 1-2 or prevalence approx. $< 7 - 19\%$ depending on case management levels). Every index case needs to be followed up, up to a maximum case-incidence threshold which defines the suitability threshold of settings for elimination using RCD. Increasing the search radius around index cases is always beneficial.

Conclusions. RCD is highly resource intensive, requiring testing and treating of 400-500 people every week for 5-10 years for a reasonable chance of elimination in a Zambia-like setting.

5.2 Background

One of the great challenges of malaria control in the face of decreasing transmission is onwards-transmission by asymptomatic infections (*219*), which are not detected by traditional interventions. Furthermore, heterogeneity in incidence and transmission and spatial and temporal clustering of cases (*220*) make mass interventions cost-inefficient, increasingly ineffective and unsustainable in low-transmission settings. The World Health Organisation (WHO) elimination guide (*36*, *37*) addresses this need to adapt intervention strategies in the elimination phase to highly targeted, locally adapted measures to track down foci of transmission. Surveillance-based strategies, such as (re)active case detection (RCD) serve as key tools in following these aims and optimizing resource allocation.

In reactive strategies, when an index case of clinical malaria presents to a health facility, this triggers follow-up activities around the index case. Assuming that cases are geographically clustered and that presence or absence of symptoms is independent of surrounding cases, reactive interventions may thus allow targeted detection or intervention on symptomatic cases who do not seek care as well as asymptomatic cases. Although reactive strategies have high approval rates and have been implemented in a range of settings, in practice, many variations of strategies are employed globally. In low-prevalence urban areas of India, contacts of cases are screened (*221*). In Southern Province, Zambia, all individuals within a 140m radius of the index case are tested (*222*). RCD has also been implemented in 13 of 14 countries in the Asia Pacific region including China (*223*). China successfully adapted the so-called "1-3-7" strategy where malaria cases are reported within one day, their confirmation and investigation occurs within 3 days and the appropriate follow-up intervention to prevent onwards transmission occurs within 7 days. Follow-up interventions may include Indoor Residual Spraying (IRS) or RCD within the household (*223*, *224*). Little information on the detailed implementation strategies is available for other programmes in the Asia Pacific region (*225*). Overall, the range of possible surveillance-response combination strategies is too vast for systematically assessing different strategies across a wide range of settings. A comparison of RCD efficacy between the few existing field studies is difficult due to differences in contextual determinants (such as health system infrastructure, geography, demographic structure or transmission intensity), which may influence optimal strategies. In addition to this, none of the existing field studies have assessed the effect of RCD on transmission (*39*) and the sites in question are yet to reach elimination, making them unsuitable for deriving even case study estimates of the potential of different RCD strategies in achieving elimination. Even with a simple test-and-treat follow-up, the relative relevance of parameters, such as the numbers of index cases and follow-up individuals, and optimum strategies thus remain to be determined.

Previous work using a deterministic susceptible-infected-susceptible (SIS) model highlights the importance of prevalence at the beginning of the intervention and suggests that relative values of the number of index cases followed, ι , and the number of neighbours in the search radius, ν , affect equilibrium prevalence levels even when the total number of individuals screened (the product of ι and ν) is the same. The proportion of all infections found per unit time appears to be the main determinant of reduction in prevalence (*50*). These models highlight the main features of the dynamics of the system but cannot provide quantitative predictions applicable to specific settings as they do not incorporate stochastic events, the effects of seasonality, within host parasite dynamics, host immunity, dynamic numbers of index cases or dynamic testing rates. In this study, test-and-treat-based reactive case detection was implemented in *OpenMalaria*, a powerful, individual-based stochastic model that includes the above factors. *OpenMalaria* was parameterised using a data set from Southern Zambia, to provide a realistic setting for which the applicability of simpler models was evaluated. The relevance of different parameters in determining the proportion of runs where transmission is interrupted was assessed by carrying out simulations utilizing different RCD strategies across

entomological and health care settings.

Two large simulation experiments are reported in this paper. The first experiment is used to characterise settings where RCD alone can lead to interruption of transmission, assess which strategies are most efficacious in the different settings, and determine the sensitivity of success to programme specific parameters (follow-up capacity versus search radius). The second experiment is used to assess the time to interruption of transmission and determine whether success is predictable through routinely collected data. Because the simulations are stochastic, the results were analyzed using conventional statistical models and machine learning techniques, treating them as a large real-world experiment.

5.3 Methods

5.3.1 Transmission and disease model

The impact of RCD was simulated using *OpenMalaria* (`https://github.com/SwissTPH/openmalaria.wiki.git`), a modelling platform allowing for individual-based stochastic models of malaria dynamics in humans (*52*), linked to a periodically-forced deterministic model of malaria in mosquitoes (*54*). Further details about the model have been previously published (*63*). In brief, *OpenMalaria* captures different clinical presentations of malaria in individual humans as well as vector ecology across a range of species and *Plasmodium falciparum* dynamics in both humans and mosquitoes, allowing for simulations of interventions in comparatively realistic settings. Since blood stage parasite densities are tracked, the model allows for case management based on simulated events dependent on individual patient parasite densities. The simulated human population is updated every 5 days with multiple outcomes, including clinical incidence, the total number of infections, and the infectiousness to mosquitoes, which depends on the recent history of parasite densities. The original model required the fitting of 38 parameters of malaria epidemiology, either independently (13 parameters) for independent model components, such as within-host parasite dynamics, or through a combined fitting process (25 parameters). A total of 61 scenarios were constructed based on field data and correspond to field sites where the pattern of transmission and one or more epidemiological variables were known. Eleven[1] different objective functions (likelihood or sum of squares) were derived from these data sets, representing important epidemiological malaria relationships, such as age pattern of incidence of clinical malaria. A number of alternative model formulations exist.

[1]The publication relating to this Chapter refers to 10 objective functions (*117*). This is because the publication predates the calibration work of Chapter 2 and the notion of 10 objective functions was derived from the *OpenMalaria* literature. In reality, the overall goodness-of-fit score used during calibration is the weighted sum of 11 separately calculated loss function values. However, as Objectives 5 and 6 both describe age patterns of clinical incidence (for Dielmo and Ndiop, Senegal, and for Idete, Tanzania, respecively. See Chapter 2 and Appendix A), they are often summarised into one objective. For consistency with the technical information provided in Chapter 2, I refer to 11 objectives throughout this thesis.

OpenMalaria was parameterised in accordance with Zambia 2010 Southern Province census data with a simulated population size of $N = 10,000$, approximately equal to the population of a single health centre catchment. Seasonality was incorporated in the same pattern from Southern Zambian data as in previous publications (e.g. (*206*)). A warm-up period of one human life span was run to induce a stable level of immunity. During the warm-up period, simulations were run with forced transmission rates. Simulations were run for nominal calendar year $t = 2010$ until $t = 2035$ with introduction of RCD at $t = 2020$ for a period of 10 years. Monitoring was started with surveys in 5-day intervals from $t = 2017$.

5.3.2 Model of Reactive Case Detection

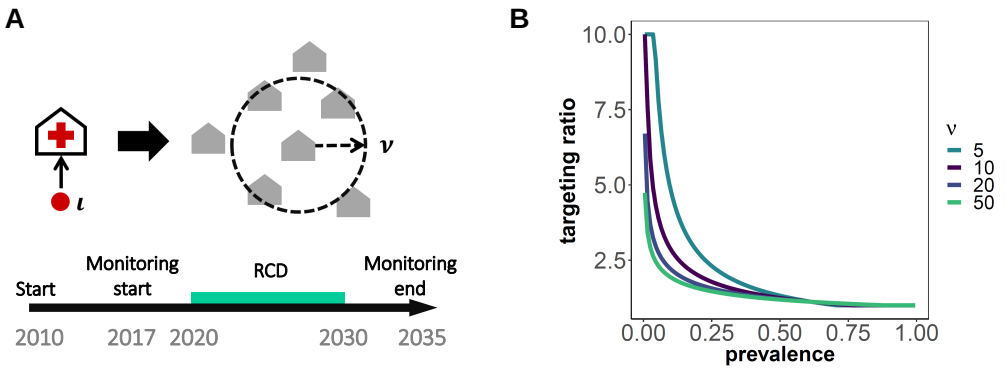

FIGURE 5.1: Basic concepts used in the models
A. Reactive case detection setup. B. The targeting ratio, τ as a function of prevalence, p and search radius ν .

RCD was modelled as a test-and-treat intervention added to and dependent on routine passive case detection and treatment included in the simulations for a period of ten years (Figure 5.1A). Within each 5-day time step, for all ι passively detected index cases tested with a rapid diagnostic test with a detection threshold 50 parasites per 1 μl and a specificity of 0.942 (*226*), an additional ν individuals (neighbours) are tested, and treated if infected. All treatments are simulated as leading to an immediate cure (but prophylactic effects were not considered).

OpenMalaria does not explicitly model the spatial pattern of infection. Instead, the effect of case clustering is captured by simulating treatment of all infections in $\nu\tau$ individuals selected at random, where the targeting ratio, τ , is the ratio of the prevalence amongst the ν individuals closest to the case, to the population prevalence, p . Equivalently, τ is the ratio of the size of a random sample that would be need to be tested (and treated if infected), to the number actually tested and treated, in order to achieve the same number of effective treatments (*50*).

The targeting ratio was estimated as a function of p and ν using Markov Chain Monte Carlo (MCMC) methods from data of a cluster-randomised trial carried out in the Zambian study area (*50*, *222*) and is defined as:

$$\tau\Big(\nu, p\Big) = exp\left(\left(-\alpha_1 log(p) + \frac{\alpha_2}{\nu} - \frac{\alpha_3}{\nu} log(p)\right)\frac{N-\nu}{N}\right) | \tau_{p,\nu} \geq 1$$

where N is the population of the health centre catchment. α_{1-3} are constants derived from field data: $\alpha_1 = 0.230$, $\alpha_2 = -1.395$ and $\alpha_3 = 2.874$. It should be noted that this definition of τ provides a better fit for low to medium values of τ and that high values are mathematically possible at low ν and p . Since these have not been observed in the field, τ was constrained to $\tau \leq 10$ (Figure 5.1B)

This yields an intervention coverage c of

$$c = \min\left(\frac{\tau_{p,v}\iota\nu}{N}, 1\right),$$

where ι is the actual number of index cases investigated per 5 day period. ι is dependent on the number of cases that present to a health facility and the maximum capacity of the RCD programme to follow up index cases, $i_{\max}$. Treatment failure in the community (for whatever reason) or diagnostic insensitivity are equivalent to a reduction in v. Coverage was calculated for each 1% prevalence interval and for $\iota = 1 \ldots \iota_{\max}$. The number of passively detected index cases that could be followed up per time interval, i.e. the capacity of the program, was constrained by fixing $\iota_{\max}$.

Simulation Experiments

Simulation Experiment 1: a total of 83,200 simulations were run in a full factorial experiment, considering 64 settings defined by different transmission intensities and case management levels for different intensities of RCD with one random seed each (Table 5.1). The simulated entomological inoculation rates (EIR) were all rather low. Entomological specifications including seasonality were parameterised as in previous simulations of southern Zambia (*206*). In accordance with the southern Zambian study site, a reference health care system was chosen such that the probability of effective treatment of any clinical case within 14 days (E_{14}) was 21.8% with a failure probability of 19.3% (*206*). This was converted to the health care system input parameter in *OpenMalaria* describing the probability that a case with an uncomplicated episode of malaria seeks care within 5 days as described previously (supplement to Penny *et al*, 2015 (*61*)). Eight different case management coverages (E_{14} levels) were simulated. The simulated importation rate was zero.

In each of the 64 settings, 1,300 RCD strategies were considered (Table 5.1). The 50 simulations per setting with ν =0 were control simulations where no RCD was performed. The simulated total number of infections, number of confirmed clinical cases, and treatments aggregated across all age groups and the simulated EIR, were tracked.

Simulation Experiment 2: a second simulation experiment was carried out to analyze time to zero prevalence (including both patent infections and those below the detection threshold) as a function of $\iota_{\max}$ and ν , replicating each simulation five times with different random seeds (Table 5.1), allowing us to analyse stochasticity in interruption of transmission. This experiment considered only the reference case management level.

TABLE 5.1: Setup of Simulation experiments

Variable	Description	Levels	Step size	No. of levels in experiment 1	No. of levels in experiment 2
EIR	Entomological Inoculation Rate in infectious bites per person per annum	1, 1.25, 1.5, 1,75, 2.0, 3.0, 4,0, 5.0	-	8	8
E_{14}	Probability (%) of effective treatment of any case within 14 days	13.9, **21.8**, 26.0, 36.7, 46.1, 54.6, 62.1 68.9% .	-	8	1
$\iota_{\max}$	Maximum number of index cases followed up in a 5 day period	1-50	1	50	50
ν	Number of neighbours tested, and treated if infected, for each index case.	0-50	2	26	26
Seeds	Seeds for random number sequence	-	-	1	5

The reference value (in bold) for E_{14} is the only value simulated in Experiment 2, where E_{14} is the probability of effective treatment of any clinical case within 14 days.

5.3.3 Analysis of simulation results

Because of the stochasticity featured in the model, the simulated dataset was analyzed as though it were a real-world experiment. The initial assessment was of whether and where RCD alone can lead to interruption of transmission as a proof of concept, considering also very low transmission settings (EIR = 1, 1.25). As transmission in such settings can be unstable, stochastic interruption of transmission may occur in control settings, so interruption of transmission in RCD simulations is not necessarily due to the RCD.

Predictors of interrupting transmission: The proportion of simulations in Experiment 1 where transmission was interrupted was aggregated across all interventions with EIR as an independent variable, stratified by case management level. The number of simulations summarised per data point, n, was 1300 ($\iota_{\max_{1-50}}$, $\nu_{0-50,2}$). First, the proportion of simulations where RCD interrupted transmission ($p = 0.0$) by the end of the monitoring period (t = 2035)

was calculated. Results were aggregated over all runs where RCD was implemented and where the control ($\nu = 0$, i.e. no RCD) did not reach interruption of transmission, regardless of RCD strategy. Thus the effectiveness of RCD was defined as the proportion $1 - \frac{e_1 n_0}{e_0 n_1}$, where e is the number of scenarios where transmission was interrupted, n is the corresponding total number of scenarios, and the subscripts indicate whether RCD was implemented (subscript 1) or not (subscript 0).

Each of the 64 settings was further assigned to one of three categories depending on whether transmission was interrupted in all, none, or only some simulations. This indicated in which settings success was dependent on the RCD strategy. The proportions or probabilities of interrupting transmission for a given setting and RCD strategy were estimated using a range of classification algorithms (Random Forest, Gaussian Process, Naïve Bayes and Support Vector Machines (SVM)). This provided smooth probability surfaces in the absence of replication, thus avoiding the need for massive numbers of simulations. The performances of these algorithms were compared using 10-fold cross validated mean area under the curve (AUC) and accuracy estimates with a two-third training/test split.

A formal variable importance analysis of the determinants of interruption of transmission was carried out by fitting a set of random forest classifiers to the simulation outcomes for settings where these outcomes were strategy-dependent. Two different random forest classifiers were used based on their inbuilt features, after finding their performance to be very similar (accuracy= 83-84%). The assessments considered the overall permutation importance of EIR, case management, $\iota_{\max}$, ν, and the derived variable $\iota_{max} \ / \ max\,(INC)$, the follow up capacity as a proportion of the maximum incidence across the simulation period. The permutation importance of each of the above variables was derived, in each case adjusted for the other variables.

A second approach measuring the overall incremental impact of RCD while adjusting for interruption of transmission in control simulations was to estimate the population attributable fraction, PAF. This is the proportion of simulations with interrupted transmission in Experiment 1 where transmission was interrupted because of the RCD, calculated as described in the supplementary information.

Median time to zero prevalence in years: this was computed for each scenario in Experiment 2 where transmission was interrupted in simulations for at least 3 out of 5 random seeds.

Prediction of RCD success: the simulations in Experiment 2 were also used to assess the suitability of routinely collected incidence and prevalence data in predicting RCD success (since EIR is not commonly measured in the field). Multiple single-variable logistic regressions

was used to determine compare the following variables as predictors of RCD success: pre-intervention EIR, incidence and the prevalence in the year before introduction of RCD, in the first year of the intervention period, and the relative reductions of incidence and prevalence, defined as $\frac{p_{2020}-p_{2019}}{p_{2019}}$ and $\frac{INC_{2020}-INC_{2019}}{INC_{2019}}$, where p_{2019} , INC_{2019} , p_{2020} , and INC_{2020} are prevalence and incidence in the last year before and the first year of RCD, respectively.

5.3.4 Software

The base model variant of *OpenMalaria* V36 (*52*) was used. The parameterisation process and model variants are detailed by Smith et al (2008 and 2012) (*52*, *64*). Scenarios were generated and all analysis was performed in R 3.4.1.The classification analysis was carried out using the mlr package in R. Classif.ranger was used to calculate the overall variable permutation importance, whilst the adjusted permutation importance was calculated using the classif.RandomForest.SCR classifier from the randomForestSCR package (`https://kogalur.github.io/randomForestSRC/`). Calculations were performed at sciCORE (`http://scicore.unibas.ch/`) scientific computing core facility at University of Basel.

5.4 Results

TABLE 5.2: Contingency analysis for interruption of transmission with RCD stratified by case management levels

	Transmission interrupted		Total				
E_{14} (%)	**RCD+ (e_1)**	**RCD- (e_0)**	**RCD+ (n_1)**	**RCD- (n_0)**	**Effectiveness**	**PAF (%)***	**p-value†**
13.9	4421	87	10000	400	0.51 (0.41, 0.59)	49.8 (39.7, 58.2)	<0.001
21.8 (REF)	5618	130	10000	400	0.42 (0.33, 0.50)	41.2 (32.4, 48.8)	<0.001
26.0	6023	121	10000	400	0.50 (0.42, 0.57)	48.8 (40.7, 55.8)	<0.001
36.7	6760	150	10000	400	0.44 (0.37, 0.51)	43.6 (36.1, 50.2)	<0.001
46.1	7586	218	10000	400	0.28 (0.21, 0.34)	27.4 (20.7, 33.5)	<0.001
54.6	7883	224	10000	400	0.29 (0.22, 0.35)	28.2 (21.8, 34.0)	<0.001
62.1	8510	191	10000	400	0.44 (0.38, 0.49)	42.9 (36.9, 48.4)	<0.001
68.9	8604	244	10000	400	0.29 (0.23, 0.35)	28.3 (22.6, 33.6)	<0.001
Total	55405	1365	80000	3200	0.38 (0.36, 0.41)	37.5 (35.0, 39.9)	<0.001

The total number of simulation runs at each case management level is 11400 with 11000 RCD+ and 400 RCD-. E_{14} is the probability of effective treatment of any clinical case within 14 days. REF indicates the reference scenario.

Transmission was interrupted in 68.2% of simulations in Experiment 1, 69.3% of simulations with RCD and 42.7% of control settings. The proportion of simulations with interruption of transmission where this could be attributed to the RCD (PAF) was 37.5% overall (95% CI 35.0, 39.9). Results stratified by case management level are presented in Table 5.2. The effectiveness of RCD ranges between 28% (21, 34) and 51% (41, 59) for different levels of case management. PAF estimates range between 27.4 (20.7, 33.5) at E_{14} = 46.1% and 49.8 (39.7, 58.2) at E_{14} = 13.9% . Overlapping confidence intervals and variability in PAF estimates yield no evidence of a trend for different case management levels. These results demonstrate that RCD can increase chances of interrupting transmission, but also highlight the stochasticity of this.

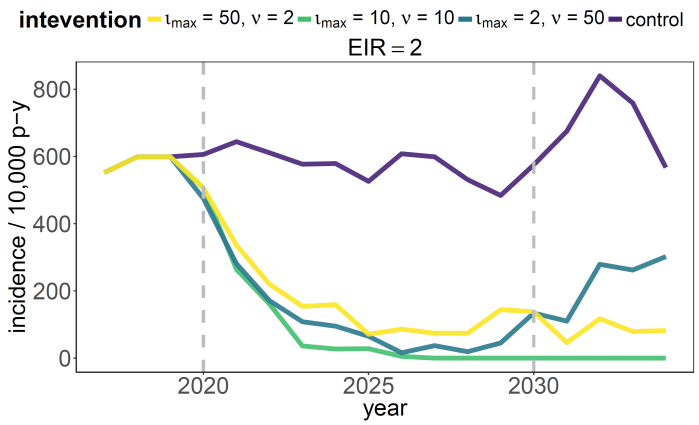

FIGURE 5.2: Example of annual incidence throughout the simulation period for different intervention strategies

The total screening capacity of RCD is kept constant at 100, but resources are differentially allocated between maximum number of index cases that can be screened, ι_{max} , and the individuals to be screened within the search radius, ν . Grey dashed lines denote the beginning and end of the intervention period. Case management is at reference level (E_{14} = 21.8% , where E_{14} is the probability of effective treatment of any clinical case within 14 days).

Three example simulations of time series of malaria annual incidence per 10,000 person-years throughout the simulation period are presented in Figure 5.2. The overall treatment capacity of the intervention, the product of ι_{max} and ν , was kept constant at 100 individuals per 5 days but resources were differentially allocated between the two parameters (50 and 2, 10 and 10, or 2 and 50 for ι_{max} and ν respectively). The fourth simulation corresponds to treatment of index cases only (zero follow up radius). Only the scenario with equal resource allocation ($\iota_{max} = \nu = 10$) reaches interruption of transmission by the end of the monitoring period, demonstrating the importance of appropriate resource allocation.

Figure 5.3A shows settings where transmission was not interrupted in controls categorised by proportion of simulations where transmission is interrupted. Each tile represents simulations of all RCD strategies for the given setting. Categories were assigned based on whether in all, some, or none of the simulations transmission was interrupted. *Pf*$PR_{0\text{-}99}$ values indicate mean prevalence prior to RCD.

For all settings where interruption of transmission is dependent on RCD strategy, strategy-dependent probabilities of interrupting transmission were predicted using Random Forest, SVM, Gaussian Process and Naïve Bayes Classifiers. Mean 15-fold cross-validated AUC was highest for Gaussian process classifier (range 0.70-0.98 across settings for 67% training), but predicted patterns were similar across all classifiers. The results are presented in Figure 5.3B and demonstrate the narrow range of settings where RCD strategy determines RCD success. The RCD strategy determined the probability of interrupting transmission only in

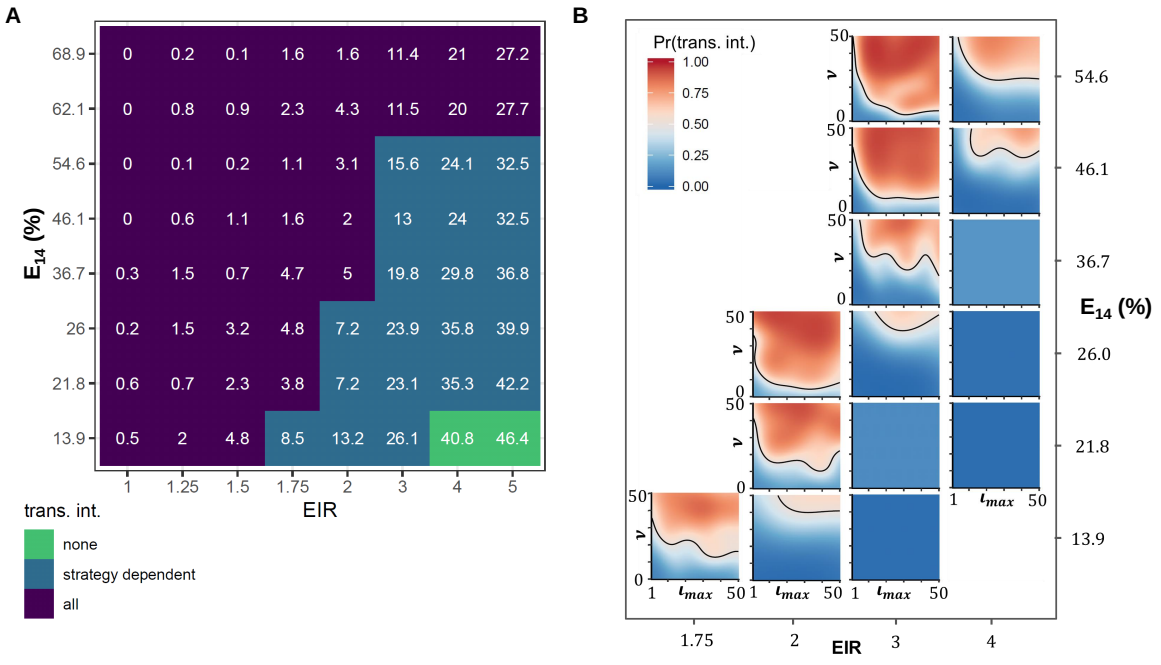

FIGURE 5.3: Dependence of prevalence and of probability of elimination on settings

A. Mean *Pf*PR$_{0\text{-}99}$ before RCD for different combinations of transmission potential and case management. Each pixel represents 250 different RCD strategies with the maximum number of index cases that can be followed up, $\iota_{\max}$, and the search radius, ν , ranging from 1 to 50 and 0 to 50 respectively. B. Predictions of probability of elimination for all strategies and settings. Probabilities were generated using a Gaussian Process classifier (c-classification, radial kernel) for each setting where it was previously identified that the elimination outcome is RCD strategy dependent. At EIR = 5 too few simulations reached elimination to generate contour plots. E_{14} is the probability of effective treatment of any clinical case within 14 days.

simulations with an initial *Pf*PR$_{0\text{-}99}$ of 7-20% , depending on case management levels (for the dependency of initial *Pf*PR$_{0\text{-}99}$ on EIR and case management is shown see Figure D.1 in the supplement, Appendix D). Below this range, transmission was generally unstable, above this range transmission was too high to be interrupted regardless of RCD intensity. Of the settings where interruption of transmission using RCD was possible, settings with higher transmission or lower management required more intense RCD. High case management alone can lead to 100% interruption of transmission even when transmission is moderate (*Pf*PR$_{0\text{-}99}$ = 28% , E_{14}>63%.

To further assess the differential effects of setting specific parameters, an individual sensitivity analyses for EIR was carried out, stratified by case management level. Figure 5.4 shows results for different initial EIR, different case management levels, and for the corresponding initial (all age) prevalence. There is a strong inverse relationship between EIR and RCD success. At the reference case management level, 50% of simulations with initial prevalence 10% led

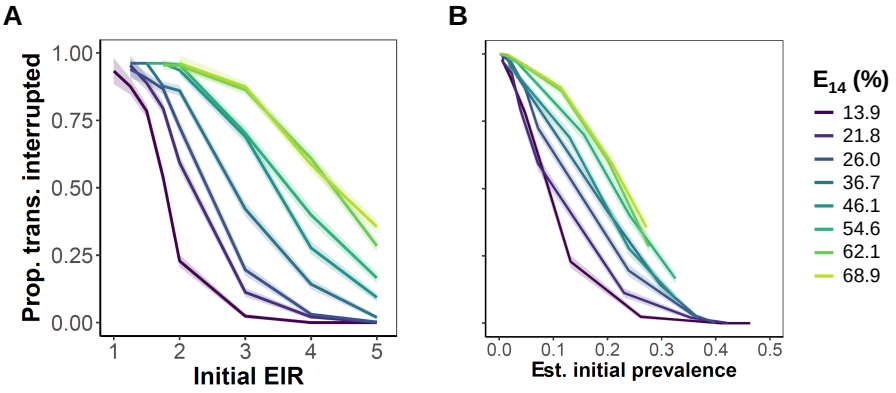

FIGURE 5.4: Proportion of runs where transmission is interrupted
The proportion of runs where transmission was interrupted only when RCD was implemented but not in the control setting, where the search radius is 0 ($\nu = 0$). A. Interruption of transmission by EIR. Proportion of simulations where p = 0 at t= 2035 (points) across all maximum follow-up capacities, $\iota_{\max}$ and ν . B. Interruption of transmission by initial prevalence. Combinations of EIR and case management were converted to mean initial prevalence from simulations. E_{14} is the probability of effective treatment of any clinical case within 14 days.

to interruption of transmission. Increasing case management can make elimination possible at somewhat higher transmission intensities. RCD with the aim of transmission interruption is thus only appropriate in settings where transmission is very low. Further analyses used machine learning approaches and estimates of the excess probability of transmission interruption that allowed for stochastic interruption of transmission in the absence of RCD.

Figure 5.5 shows the results of an importance analysis of the effects of $\iota_{\max}$ and ν on the probability of elimination. The results were obtained by fitting a random forest classifier to the simulated data and using an inbuilt function for variable importance. Setting specific parameters, case management and EIR, are the most important variables for elimination (panel A). Panels B-F show the covariate-adjusted variable importance on the probability of elimination. Any increase in search radius is associated with an increase in the proportion of eliminated simulations. Increasing the capacity to treat and follow up index cases ($\iota_{\max}$) is of great benefit initially, but quickly reaches an optimum and saturates at a threshold. Increasing case management shifts the saturation threshold to higher values of $\iota_{\max}$ but does not change the overall pattern observed with increases of the other parameters.

Further analyses to determine characteristics of the $\iota_{\max}$ threshold, were carried out. RCD success in each setting was analyzed in relation to the ratio of $\iota_{\max}$ to different quantiles of the distribution of 5-day incidence over the entire simulation period from 2020 to 2035. $\iota_{\max}$ as a proportion of the all-time maximum incidence, $\frac{\iota_{max}}{max(INC)}$, showed the strongest relationship with RCD success, suggesting that the ability to cope with maximum incidence

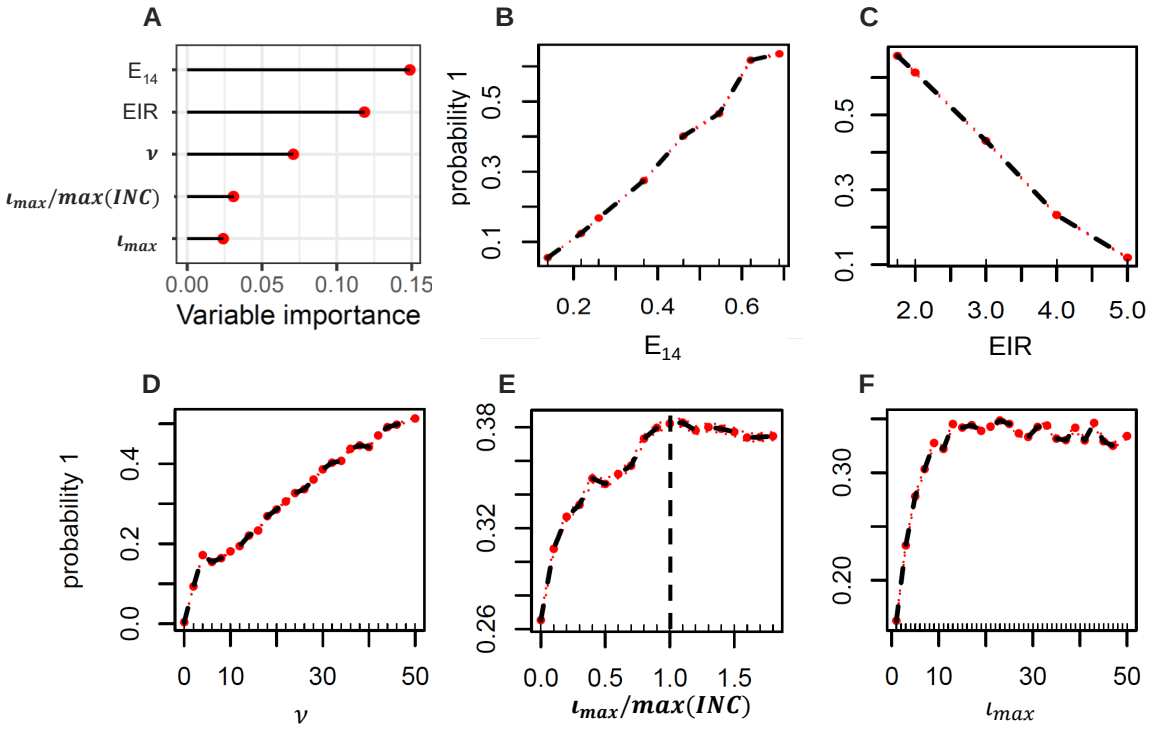

FIGURE 5.5: Permutation feature importance in settings with stable transmission (without elimination in controls)

A. Permutation variable importance. B-F. Covariate-adjusted variable importance . All plots were generated using a random forest classifier. "Probability 1" describes the probability of elimination. E_{14} is the probability of effective treatment of any clinical case within 14 days, $\iota_{\max}$ and ν describe the maximum number of index cases that can be followed up and the search radius respectively, and max(INC) describes the maximum 5-day case incidence throughout the simulation period.

during transmission peaks is the most important factor in choosing $\iota_{\max}$. It is important to note that the maximum incidence of the simulation period need not be reached prior to RCD implementation as stochastic fluctuations in incidence may lead to an increase in case incidence during the intervention period when case management is low. The adjusted importance analysis of this metric confirms that $\iota_{\max}$ should be set such that all index cases can be followed up (Figure 5.5F).

The contributions of case management and case follow up to interrupting transmission were assessed. Figure 5.6 shows the proportion of simulations where transmission is interrupted ($\frac{e}{n}$) and the proportion where this is attributable to RCD ($PAF\frac{e}{n}$) as a function of $\frac{\iota_{max}}{max(INC)}$. When stratifying by case management level, the PAF was found to decrease at high case management and high $\frac{\iota_{max}}{max(INC)}$, although the total proportion of simulations that reach elimination increases. High $\frac{\iota_{max}}{max(INC)}$ can be a result of both small RCD effort and low transmission. These results thus indicate that at low transmission, the relative contribution of case management

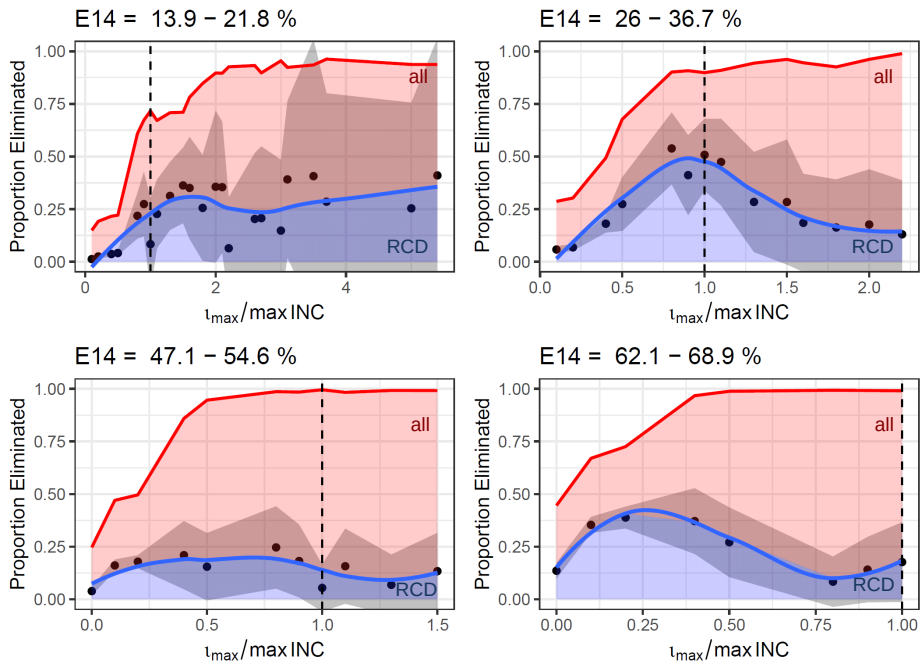

FIGURE 5.6: Proportion of eliminated simulations by relative follow up capacity

Blue represents the proportion of eliminated simulations where elimination is attributable to RCD (PAF), relative to the total proportion of eliminated simulations (red). Simulations were aggregated by relative follow up capacity in 10% intervals, thus each data point presents varying numbers of simulations with generally fewer simulations for ratios above 1. Mean PAFs (dots) and 95% CI (gray area) were calculated and a LOESS smoothing function with 0.75 span was fitted through the data (blue line). E_{14} is the probability of effective treatment of any clinical case within 14 days, $\iota_{\max}$ is the maximum number of individuals that can be followed up per five days, and max(INC) is the maximum 5-day incidence throughout the simulation period.

to elimination is higher than that of RCD (although RCD may provide some additional benefit).

These results demonstrate that ι_{max} is optimally chosen such that all index cases can be followed up, even during high transmission seasons. The relative contribution of RCD to interrupting transmission decreases when regular case management alone is strong (at high case management and low transmission). Still, RCD can provide small additional benefit.

Figure 5.7 summarises the time to interruption of transmission in years at reference case management level across intervention strategies and transmission levels. Points show the median years to interruption of transmission across five seeds at reference case management level. Presenting the median ensures that >50% of seeds over all simulations reach interruption and that this is not due to stochasticity. The line and shaded area show the conditional mean and confidence interval of the median, using a LOESS smoother. For EIR = 4 in panel

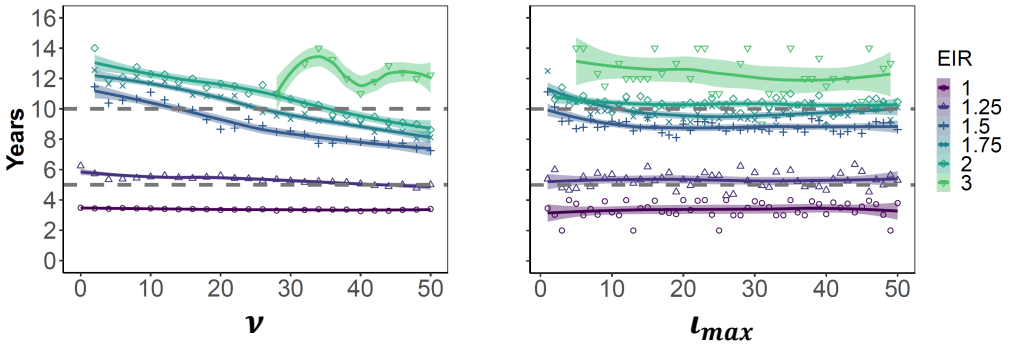

FIGURE 5.7: Median time to interruption of transmission
Results are disaggregated by the maximum number of individuals that can be followed up, $\iota_{\max}$, and the search radius, ν . The median time to elimination was calculated across all simulations (all values of ν and $\iota_{\max}$, respectively) of the given parameter value. Point estimates were plotted and a LOESS smoothing function was fitted. Median values could not be calculated for EIR>2 since not enough simulations reached interruption of transmission.

A, confidence intervals were too wide to be displayed. This suggests that transmission is interrupted towards the end of the intervention period, except for at very low transmission. A possible carry over effect beyond cessation of the programme was observed, as at higher EIRs elimination may be reached after the 10-year mark. The intervention strategy makes little difference except for increasing the search radius at EIR 1.5 – 2, approximately equivalent to an initial prevalence of 7% . Increasing $\iota_{\max}$ has little effect in determining time to interruption of transmission. Overall, there was a high degree of stochasticity.

Since EIR is difficult to measure in the field, a range of possible predictors of routinely collected data was chosen to assess their predictive ability. The predictive powers of mean incidence and mean annual prevalence the year before onset of the intervention period and in the first year of the intervention period were tested. The predictive ability of the relative reduction in incidence and prevalence in the first year of RCD was also assessed. Prevalence values were log transformed for the purpose of this analysis. The results presented in Figure 5.8 suggest that routine data such as incidence and prevalence are equally good in assessing the suitability of a site for RCD as EIR. 83.3% (both) and 86.1% and 84.8% of scenarios were correctly classified using incidence and log prevalence in the year before onset of RCD and the first year of RCD, respectively as covariates in a single variable logistic regression (Table 5.3). The cut off points for >50% probability of interrupting transmission with any strategy at reference case management in the fitted model were 711 per 10,000 person-years, equivalent to a mean of 10 cases per 5 days.

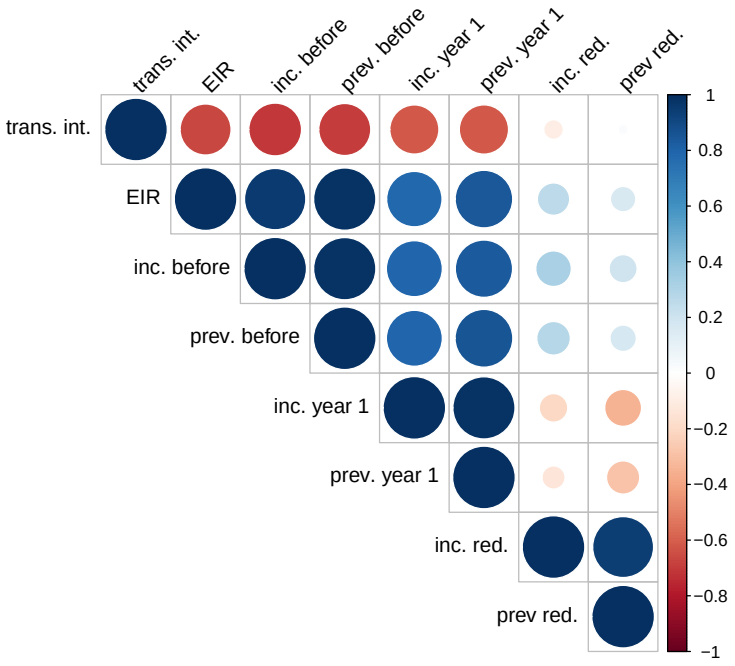

FIGURE 5.8: Correlation analysis of predictions of RCD success
Pearson correlation coefficients.

TABLE 5.3: Single-variable logistic regression results and % correctly classified runs when using the given model

	Estimated odds	95% CI	p-value	correctly classified (%)
EIR	0.082	(0.079, 0.088)	<0.0001	83.9
Incidence before	0.996	(0.996, 0.996)	<0.0001	83.3
log(prevalence before)∗	0.090	(0.086, 0.094)	<0.0001	83.3
Incidence year 1	0.990	(0.989, 0.990)	<0.0001	86.1
log(prevalence year 1)∗	0.038	(0.035, 0.040)	<0.0001	84.8
reduction in incidence	0.403	(0.367, 0.443)	<0.0001	63.4
reduction in prevalence	1.308	(1.164, 1.470)	<0.0001	62.8

∗ Per 1% increase

5.5 Discussion

A framework for implementing spatially targeted reactive interventions in *OpenMalaria* was developed by approximating a targeting ratio which captures spatial heterogeneity. Reactive case detection only leads to local interruption of malaria transmission in settings with low transmission potential, measured by either EIR or prevalence (initial EIR approx. 1-3 or prevalence approximately <7-19% for 50% probability of success, depending on case management levels). RCD was effective in 38% (95% CI 36, 41) of scenarios simulated, with a small trend towards greater effectiveness at lower case management. Large and overlapping confidence intervals highlight the stochasticity of interrupting transmission. Whilst the simulations use transmission intensity in terms of EIR and 5-day case management levels as inputs, they show that qualifying settings are also identifiable using routinely collected prevalence and

incidence data. It follows that settings can be classified into suitable or non-suitable for RCD, and predictions can be made such that RCD is not initiated where it is very unlikely to be successful. This is roughly consistent with another modelling study based on the same Zambian data, which suggests a prevalence threshold for RCD of approximately 10% (*227*). The relative success of RCD (in terms of reduction in incidence or prevalence) in the first year of implementation is not a good predictor of ultimate success.

A primary objective of this study was to determine which operational characteristics of RCD most influence the effectiveness. Considering the joint and individual influences of increasing capacity to follow up index cases and the search radius, a larger search radius is always beneficial. In contrast, increasing $\iota_{\max}$ only increases the probability of success up to a certain threshold, $\iota^*_{\max}$, with no additional value in increasing it beyond this. This threshold is related to the maximum number of cases presenting to be treated, and defines a cut-off point in case incidence beyond which the setting is not suitable for elimination via RCD altogether. For settings with low to intermediate case management levels, all index cases should be followed up. A natural approach for programme managers to use when considering introducing RCD, is to initiate it when the number of cases reporting has fallen to the level where the capacity exists to follow them all up. The simulations endorse this instinct, in contrast to previous modelling of RCD that used a deterministic compartmental susceptible-infected-susceptible (SIS) model with the same function for the targeting ratio (*50*). The SIS-models attributed greater importance to following up more index cases, even at the cost of a smaller ν across the entire parameter space, as the targeting ratio is much higher at small ν . In that analysis the upper bound of the possible number of index cases was set by the standing crop of infections; it was assumed that any pre-defined number of index cases less than this can be found.

This paper takes a slightly different, programme planning focused approach by assessing the importance of the follow-up capacity rather than the actual number of followed up index cases. Cases that do not seek official care or are asymptomatic limit the potential number of index cases. When the case management level is increased, more index cases become available making it useful to adjust $\iota_{\max}$ upwards. Thus it would seem that case management limits the settings where elimination is possible because it imposes the threshold effect on $\iota_{\max}$

Despite providing a proof of concept that RCD can lead to local elimination, the results also suggest that successful RCD is highly resource intense and likely to be very costly (though no formal economic evaluation is presented). In most suitable settings RCD would have to be conducted for more than 5 years with a relatively aggressive strategy to yield a probability of interruption of transmission of >50%. In the area of the trial in Zambia, for example, the weighted mean EIR is estimated to be approximately 2.9 (*61*) and the case management (E_{14}) at the time where the RCD trial was conducted was about 21.8% (*206*). In such a setting, elimination with RCD is not feasible. Today, Zambia's case management (E_{14}) is estimated to be 34.7%. Such a setting would require a follow-up capacity of approximately 10 index cases

and search radius of > 35-40 individuals, i.e. a total screening capacity of 350-400 individuals per 5-day period would be necessary in order to reach interruption of transmission within 10 years with a probability of success of only 50%. On the contrary, if case management was further increased to $E_{14} = 46\%$ (*137*), following up 5 index cases with a 25-person radius, i.e. 125 individuals in total, would suffice for a 50% probability of interrupting transmission. If case management was to be increased to 62% this alone would lead to interruption of transmission (in all simulations). Strengthening access to care would also have further positive implications on population health and likely lower the burden of disease not just of malaria, but overall. Previous modelling studies having assessed RCD come to similar conclusions. An independent modelling study (*42*) suggests that in low prevalence settings, improving case management may be more impactful than RCD, although RCD may bring qualitative benefits to a setting. For example, RCD may improve visibility of community health workers and raise awareness of the disease in general.

The results indicate that policy decisions should prioritise improving access to care followed by appropriate treatment and follow up of all index cases. This can be explained through a greater importance of following up as many as possible definite cases that present to the health facility. All index cases are definitely cases (definite cases) whereas individuals in the search radius may or may not be cases at probability $p\tau\left(p,\nu\right)$. Targeting is stronger and τ is larger at small ν. The per-person probability of being a case in the radius around an index case is thus greater the smaller the radius. Increasing the number of index cases through improving access to care and treating and following up these index cases is therefore always more targeted than treating individuals in the search radius. RCD is thus only worthwhile where the number of cases detected through RCD $\left(min(\iota,\iota_{max})\nu p\tau\right)$ and in the limit, as $\iota_{max} \rightarrow \iota$, the number of cases detected per index case is greater than 1 $\left(1 < \nu p\tau\left(p,\nu\right)\right)$. This condition describes the threshold at which more hidden cases are found through RCD than index cases present to the health facility. If the aim is local elimination, the second condition for implementing RCD must of course be that interrupting transmission in the setting is itself feasible. As it is a highly stochastic event, this condition is more difficult to predict. Together, these two conditions define the narrow range of settings in which interrupting transmission through RCD alone is feasible. Where it is possible, it will thus generally be more effective to increase case management rather than implementing RCD. Based on these results, the following prioritisation is proposed: case management to increase number of index cases $\gg$ following up all index cases $\gg$ increasing the search radius. The effect of increasing the search radius also likely flattens off at some point, but not within the parameter space considered.

The results are based on Zambian parameterisation and thus the transferability of results across settings with different population densities remains to be confirmed. However by defining the search radius in terms of number of individuals rather than a physical distance, the findings presented here should be more translatable across settings at different population densities, assuming that vector movement and thus disease spread is dependent on host availability

more so than physical distance. Future research comparing RCD in different settings should explore this hypothesis. The study did not consider case importation. Simulations of Zambian settings with low coverage of case management (*50*) conclude that this is an important determinant of whether elimination is achieved, but simulations with *OpenMalaria* suggest that this is not the case if case management coverage is high (Smith et al. pers .com.). Whilst *OpenMalaria* does consider heterogeneous populations, including in their health seeking, the reactive intervention is applied to random individuals in the population.

The targeting ratio approach presented here provides a framework for implementing generic spatially targeted reactive interventions in *OpenMalaria*. Future work should adapt the framework to different settings by fitting the targeting ratio to data from different settings. However, as long as infections are acquired within the community similar patterns to those presented here can be anticipated, since the search radius is specified in terms of number of individuals, thereby making it independent of the population density. Future investigations may further include incorporating importation of cases as well as varying the timing of RCD and implementing RCD in combination with other (mass) interventions. If RCD is only successful at low transmission it would be well worth investigating adding a mass intervention at the beginning of RCD intervention course as well as employing RCD in scenarios where mass interventions have brought EIR down to < 3 infectious bites per person per year. Further, one may investigate employing RCD seasonally in the dry season when incidence is low. This may lead to stochastic elimination and save resources. This framework can also be adapted and used for wide ranges of spatially targeting interventions, such as RCD in different settings and reactive vector control.

5.6 Conclusions

Overall, the study demonstrates that RCD can increase the chances of stochastic elimination, but that it is a very resource intense, such that other interventions are likely more appropriate in most settings. In its final stages RCD leads to a sustained reduction in overall burden of malaria through strengthening the health care system so that imported infections are controlled. This stabilises the disease-free state (*50*), in contrast to *one-off* higher impact interventions, such as MDA that have strong immediate effects over a limited period but do not provide a sustained reduction in transmission. With MDA, even in the most favourable circumstances, persistence is highly stochastic depending on the size of the residual reservoir of infectious people exposed to mosquitoes. However, despite its potential impact RCD is a highly resource intense, long-term intervention that is inappropriate in many settings where resources are limited. In such settings, investments may be better made in improving the routine health care system.

5.7 Declarations

Ethical approval

Ethical approval for initial data collection and secondary analysis was obtained in Zambia from the Research Ethics Committee (REC) at the University of Zambia. Ethical approval was also obtained from the Tulane Institutional Review Board and PATH.

Consent for publication:

Not applicable.

Authors' contributions

TR, NC and TAS conceived of the study, designed the experiments and analyzed results. TR and TAS drafted the manuscript. All authors read and approved the final manuscript.

Competing interests

The authors declare that they have no competing interests.

Financial disclosure

TR, NC and TAS acknowledge funding by the Bill and Melinda Gates Foundation (# OPP1032350). The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

Availability of data and material

The data and underlying code are available on request from the authors

Acknowledgements

The authors would like to acknowledge helpful discussions with colleagues in the Infectious Disease Modelling Unit of Swiss TPH, in particular Melissa Penny and Monica Golumbeanu.

6 Discussion

Mathematical modelling has a long tradition in providing additional evidence to support public health decision-making against the spread of infectious diseases. The field has been experiencing a steady increase in both popularity and relevance, starting from Bernoulli's 1760 analyses of smallpox transmission and control (*132*), to the early works of Ross and MacDonald on malaria at the turn of the 20th century (*43*, *45–48*), to the modern breadth of complex disease transmission simulators for malaria (*3–5*), tuberculosis (*6*), dengue (*8*), and Covid-19 (*17*). While data science tools, including traditional statistical analyses and more recent machine learning approaches, are vital for detecting patterns hidden in large amounts of (observational) data, their explanatory power is oftentimes limited to the respective dataset. Much needed extrapolation to other settings to inform decision-making outside the original dataset's context therefore presents challenges. By enforcing strict mechanistic consistency under explicit biological and epidemiological assumptions, mathematical models can generate insights beyond the data in ways that purely data-driven approaches cannot (*106*, *134*).[1] The ability to forward predict long-term effectiveness or comparative effectiveness of interventions is of particular value in the evaluation of public health programs (*2*). In the interest of incorporating all available evidence, guideline developers such as the WHO should routinely consider modelling insights into their recommendations (see also section 6.6 of this Chapter).

However, the added layer of complexity (i.e. forcing explicit biological and epidemiological relationships) makes simulation modelling predictions especially sensitive to the assumptions and decisions made by developers and users. From the careful calibration to diverse and reliable data, to thoroughly planned and analysed simulation experiments, any decisions throughout the modelling process should be well thought through and transparently communicated. Through a comprehensive investigation of the modelling process from its foundations in calibration to its applications, this thesis provides methodological advancements and prompts discussions about the role of infectious disease modelling in public health. With application to malaria, this is achieved through:

1. **The development of new machine learning-based approaches for model calibration.** Specifically, I show that machine learning-based Bayesian optimisation algorithms offer a fast solution for solving high-dimensional optimisation problems and that they

[1] It should be noted that these assumptions need not be representative of nature – this depends on the developers' biological understanding. Nonetheless, predictions will adhere to the model's mechanistic framework. Therefore, assumptions should be carefully evaluated in their ability and appropriateness to depict nature.

outperform traditional, sampling-based calibration approaches (Chapter 2, Objective 1). For transparent sharing across disease modelling communities, these algorithms have been made publicly available.

2. **The epidemiological assessment of observational data and its conceptualisation for integration into the calibration process using a generalisable framework.** A framework for translating observational studies to *in silico* and for preparing data for calibration is provided in Chapter 3 (Objective 2). In an application case study, this framework is used to collate data on an epidemiological relationship not previously incorporated in the calibration of *OpenMalaria* (Chapter 3, Objective 3).

3. **An assessment of the implications of development decisions for predictions and possible applications.** To provide a link between model development and applications, I discuss the ability of *OpenMalaria* to reproduce epidemiological relationships and the dependence of (malaria) models and their predictions on the available calibration data (Chapter 4).

4. **An application case study involving the development of new surveillance-response interventions, and the machine learning-assisted initial assessment of their potential for elimination.** Chapter 5 shows the addition of RCD to *OpenMalaria* (Objective 4) and an assessment of its role on the road to elimination. Addressing experimental design and analysis strategies, structural model limitations, and data availability, the discussion around predicting elimination is furthered in section 6.4 (Objective 5).

The subsequent sections of this discussion provide broader context to the findings of this thesis. I discuss dependencies on data and assumptions, the value of joining simulation modelling and machine learning, the implications of decisions made by model developers and users for public health decision-making, and the limitations of modelling.

6.1 A model is only as good as the data it was calibrated to

Calibration to diverse high-quality data is the foundation of any model that aims to be useful for guiding decision-making in health. In the absence of data, the most elaborate calibration algorithm cannot identify the parameter values required for a biologically and epidemiologically sensible model and model predictions lose their value as evidence. Diligent calibration to diverse data, whose context is understood and explicitly captured, enables setting-specific predictions and extrapolations of generic disease dynamics (see Chapters 2, 3 and 4). Among other applications, a well-informed model can aid in the development of new tools, e.g. in guiding the formulation of target product profiles (TPPs) (*228*, *229*) or in evaluating use cases

for interventions (*230*), as well as provide guidance on national malaria control strategies (*231*).

Throughout this thesis, discussions have emerged around the sensible incorporation of data into modes, data scarcity, and the resulting uncertainty of model predictions. Chapter 2 illustrates the pivotal role of data at the core of model calibration. Without comprehensive data, there is no basis for assessing the ability of a model to depict the biology and epidemiology of the disease. Further, without data, calibration becomes impossible. Chapter 3 revealed the intricacies of incorporating data for calibration: The context of data collection must be holistically accounted for to avoid oversimplifying epidemiological relationships and potentially biasing predictions. Chapter 4 focused on the sensitivity of *OpenMalaria* predictions to the data quality standards set during calibration and the implications of scarcity and uncertainty. Chapter 5 illustrates the development of a new intervention plug-in for *OpenMalaria* based on data from a RCT (see also (*50*)). This chapter also highlights the uncertainties and limitation of predicting elimination that result from a lack of calibration data from settings with low or unstable transmission (see also Section 6.4 of this Chapter). Without strong data foundations, the robustness of model predictions and the potential of modelling-informed decision-making are compromised.

In Chapter 4, I highlight the effects of data scarcity on model predictions. Data scarcity increases the influence of individual data points, making model predictions more vulnerable to potential biases (see Chapter 4). Due to the interconnectedness of model components and interdependence of parameters, the decisions made regarding the calibration data of one objective can trigger a domino effect, leading to changed predictions on multiple epidemiological relationships. These changes may not be immediately noticeable and may require extensive, deeper analyses to be unveiled. For example, dropping the Ethiopian severe disease data points had no apparent effect on the relationship between prevalence and severe disease but carried implications for predictions on parasite densities and direct mortality estimates (Chapter 4). Large uncertainty bounds, such as those observed in the direct mortality data, can make the associated governing parameters unidentifiable and reduce the predicted epidemiological relationships to (somewhat informed) guesswork (Figure 4.7).

Generally speaking, the incorporation of additional data can decrease the influence of individual data points and balance out potential biases. Additional data for calibration and validation is clearly needed, particularly on severe disease, mortality and on low transmission settings. However, the continuous manual surveillance of a rapidly growing body of literature in the hope of finding relevant data is often infeasible. Modern, ML-based methodologies could help alleviate the operational challenges of continued development alongside application that arise from limited human and financial resources. For example, natural language processing can automatise the efficient mining of scientific databases for relevant newly published data (*232*, *233*). Automated translation can extend the search to publications in multiple languages or national reporting databases. Through the addition of automated

filters (whose accuracy would of course have to be validated), standing alert systems could be realised to varying or flexible degrees of elaboration. They could for example be tuned to only extract data from specific types of observational studies or exceeding a threshold sample size. Such alert systems could accompany model development during the application phase and ensure that the most recent data is continuously incorporated. This would further enable the real-time capturing of changes in epidemiology and would likely address the pressing need for additional data from low-transmission settings and both simplify and speed up the incorporation of new data.

However, blindly incorporating data on the grounds of availability is dangerous as it can compromise the quality of model predictions. Here lie the limits of automatisation. Ultimately, manual qualitative assessment by researchers will be required (as illustrated in Chapters 3 and 4). Decisions around data quality standards and the incorporation of contextual covariates require a comprehensive understanding of underlying assumptions and possible limitations. No objectively right or wrong methodology exists to address issues around scarcity, biases and uncertainties in calibration data. While standard approaches like sensitivity or parameter identifiability analyses can be viable tools to assess the value of added data sets for calibration, the interpretation of results within a broader epidemiological context is the responsibility of the researcher. Adjusting for contextual covariates in data collection further requires a deep understanding of the time, setting, and methodology. Underestimating the complexity of epidemiological relationships and their dependence on real settings that are characterised by a multitude of covariates can cause great inaccuracies in the model. In Chapter 3, I present a translational framework as well as a library of quality controlled data on the *Pf*PR-incidence relationship including contextual covariates for every record. These data are ready for incorporation into *OpenMalaria*, providing a timely contribution to addressing the nuances of this relationship. The framework and surrounding discussion can serve as a starting point for establishing generalised guidelines for the handling of calibration data for infectious disease. Alternatively to explicitly reflecting all contextual covariates in the simulations, researchers may choose to conduct additional pre-processing. For example, Griffin et al. introduced a scaling factor for differences in health care systems before incorporating severe disease data into their model of malaria transmission (*218*). However, statistical adjustments often rely on assumptions about the relationship between contextual covariates and outcomes e.g. linearity. Without a systemic mechanistic framework, such relationships may be unidentifiable or its complexity and multifacettedness may be underestimated. Capturing potentially confounding covariates in simulations allows for a more systemic approach where the joint influence of all covariates can be accounted for without additional assumptions about the shape of the relationship (outside the mechanistic assumptions of the model).

Overall, calibration data is a key determinant of the possible applications of a model - which questions can be answered, and within what parameter space the model yields accurate predictions. Ideally, any data used for calibration should be made publicly available along

with a description of decisions made during its incorporation. A ML-based standing alert system for new data, integrated into day-to-day modelling practice could address abundance issues by alleviating some of the challenges of detecting and exploiting new data. Nonetheless, a core problem of calibration is that as long as the algorithm is functional, it will yield some numerical solution, even if the data is incorrectly incorporated. The consequences for epidemiological simulations may be substantial and they may remain dangerously unrecognised if data integration methods are not granted sufficient focus in the communication and interpretation of predictions. It takes integrity to be able to fully recognise and stand by the limitations of one's model, especially after long-term continued development, often over the span of years. The explicit mechanistic assumptions allow for some extrapolation outside the calibration data, but the further we move away from the data, the more careful we ought to be. It is a developer's responsibility to be aware of and communicate the potentially wide-reaching consequences of decisions made during data integration. The discussions throughout this thesis provide an unparalleled deep dive into the inner workings of any one malaria transmission model (outside its original development). The epidemiological assessments of calibration data (Chapter 3) and functional analysis of its consequences for relationships simulated in *OpenMalaria* (Chapter 4), aim to provide the transparency required from models supporting decision-making. Together with *OpenMalaria*'s (particularly) open documentation (with publicly accessible, collated calibration data and methodology, and information on development of interventions), this degree of transparency should be a leading example for all models of infectious diseases. Only if a model's potentials, limitations and pitfalls are well understood, can the reliability and value of its predictions as evidence be evaluated. This includes the domain of plausible simulations, which is outlined by the calibration data, because even the most technically elegant model are merely speculative in the absence of sufficient data foundations.

6.2 Machine learning-augmented simulation modelling

Detailed individual-based simulation models, such as *OpenMalaria* can accommodate for the complexity of an infectious disease. However, long runtimes and the curse of dimensionality hinder a routine integration into public health decision-making (*68*). This can be alleviated through the coupling of simulation modelling to (machine learning) surrogates. At its core, machine learning-augmented simulation modelling enables faster predictions of model outputs and allows for derivative-free optimisation and (constrained) feasibility analyses. By speeding up runtimes and enabling more complex experiments, large arrays of intervention strategies can be quickly evaluated, making machine learning-augmented simulation modelling highly attractive for real time decision-support. While this has long been recognised in the engineering and physical sciences (*68*), it is somewhat surprising that it has not yet found wider application in infectious disease modelling. In fact, no well-known examples of combining of machine learning and simulation modelling of infectious diseases exist, outside the research conducted by disease modellers at Swiss TPH (concurrently to this

thesis) (*234*). This may be because the perceived complexity of an infectious disease simulator appears negligible in comparison to those employed in physics or engineering where a single simulation can take days or weeks (*66*). Further, adding another (*black box*) layer of complexity to already detailed and sometimes intransparent simulators makes the joint workings of simulator and emulator hard to interpret. As public health decision-making should be transparent, any additional complication requires thorough justification. However, the complexities of model calibration and public health decision-making (especially for complex diseases like malaria) require large, multidimensional experiments that warrant the use of surrogates.

The research presented in this thesis (Chapters 2 and 5) provides exemplary use cases of machine learning-augmented simulation modelling for calibration and public health application. Firstly, machine learning methods for calibration are shown to be faster than previous, purely sampling based approaches and yield an overall improved goodness-of-fit (Chapter 2). Secondly, sensitivity analyses reveal additional insights into input-output relationships. During the application case study of RCD in Chapter 5, sensitivity analyses provide insights into implementation strategies of RCD and the optimal allocation of limited resources. Lastly, by training machine learning emulators on simulation results, rare-event probabilities, like the probability of elimination can easily be estimated from sparse simulation results (Chapter 5). In all of these cases, the use of machine learning in combination with traditional simulation made large experiments possible and provided additional functionalities.

The problem of long runtimes could also be solved through computational changes such as improving model structure, its compilation, or incorporating GPU computing methods. For example, computational improvements made recent versions of *OpenMalaria* substantially faster. However, the applications of machine learning-augmented simulation modelling go beyond this. The generation of a smooth objective function surface, which can be coupled with gradient based non-linear optimisation methods (*68*), enables *formal* optimisation rather than choosing the best of a set of (sparse) samples (see Chapter 2). Jointly with the ability to draw constraints and requirements around the properties required of an intervention, this is highly valuable for supporting the development of new tools (*234*) as well as suggested policies on their deployment (see Chapter 5). Another application of (machine learning) surrogate models is their ability to serve as didactic tools to reveal opportunities for model simplification (*68*). For example, the sensitivity analyses shown in Chapter 2 identify parameters that may be irrelevant to the epidemiological predictions of *OpenMalaria*, or at least to those relationships captured during calibration.[2] The coupling of simulation modelling and machine learning may therefore aid in structural model development itself or initiate dialogue with scientists

[2]For example, the Sobol' analysis in Chapter 2 suggested that no objectives were sensitive to parameters α_m (θ_7), F_0 (θ_{16}), ϕ_1 (θ_{18}), or ν_0 (θ_{21}) (see Figure 2.3C). The comparison between parameter estimates between optimisation algorithms further suggests only weak identifiability of these parameters, in particular for parameters α_m (θ_7) and ν_0 (θ_{21}) (see Figure A.31).

conducting fundamental research to discuss causal relationships between biological parameters and processes.

Throughout this thesis, I have highlighted the potential benefits of coupling simulation modelling and machine learning (specifically in Chapters 2 and 5). Nonetheless, the additional layer of complexity and uncertainty introduced by an emulator requires clear communication on methodology and assumptions and their implications on predictions (both positive and negative). For guiding policy, it is absolutely necessary to characterise, quantify and propagate uncertainty, recognising both qualitative and quantitative components of both modelling layers (*235*). A proposed general framework for good practice on the combined use of machine learning and simulation modelling should therefore include experimental design, emulator evaluation, and use-cases:

Model choice and benchmarking. On a technical level, it has been demonstrated that no one emulation method performs best universally (*80*). Therefore, emulator performance must be formally addressed on a task-by-task basis. Quantitative performance metrics include the area under the curve (AUC) or accuracy for classification problems, or the mean squared error (MSE) for regression problems (*236*). The most suitable performance measure and machine learning model will depend on the application and on size and quality of the training set (*74*).

Sampling and experimental design. Emulators do not retain the forced logic and mechanistic assumptions of the simulator and can therefore not extrapolate but only interpolate (*68*). As a consequence, surrogates are not expected to perform well away from the design site and the sampled space may have to be increased to ensure good surrogate performance around the edges of the space of interest. Further, the over-sampling of rare events or in irregular areas of the space may be desirable. This logically prompts a discussion on stationary versus adaptive sampling strategies. Adaptive sampling based on the iterative acquisition of samples in areas where they are predicted to lead to improvement of the emulator's performance will be appropriate for most applications. Applied to predicting elimination this could involve adaptively oversampling the boundary region between certain elimination and certain endemicity. While a certain degree of adaptivity should be beneficial for most applications, this also increases the complexity and runtime of the overall experiment by forcing sequential iterations of simulation and emulation. The (computational) costs and benefits of an adaptive design should therefore be carefully evaluated.

Contextualisation. It is noteworthy that for applications like Bayesian optimisation, emulators are neither required to nor trained to be greatly accurate globally. Rather, they are trained to accurately identify areas of interest such as optimal regions for calibration and boundary regions for the prediction of elimination. Therefore, the trained emulator should only be applied to the problem it was trained for and not be used to extrapolate outside this.

Generally speaking, a governing principle of emulation should be to strive for simplicity wherever possible. This decreases the chance of coding errors and ensures transparency (*237*).

Traditional statistical methods should be used wherever they are sufficient to answer a research question to ensure interpretability and increase confidence in resulting public health decisions. Where machine learning surrogates can provide additional value, justification for the need and choice of emulation should be explicitly provided. Criteria to assess and demonstrate the usefulness of a surrogate should include the demonstrated increase in computational efficiency (in terms of average runtime) of a single emulation as well as the runtime for a combined (iterative) simulation-emulation approach compared to using simulation alone. The emulator should further provide demonstrated usefulness like enabling more thorough analyses or revealing extended conclusions to the research question. Lastly, the uncertainty introduced by the surrogate needs to be openly and clearly communicated and the surrogate itself validated. If these codes of conduct are followed, machine learning can provide a valuable complement to simulation modelling and decision-making.

6.3 Re-imagining the calibration of disease simulators

The calibration of detailed disease simulators, such as *OpenMalaria* is a challenging optimisation problem. It can only be solved using elaborate computational algorithms. It is noteworthy that even using the most complex algorithms, such problems will likely never be *truly solved* if by solving we mean finding a proven global optimum. That is because the complexity and multi-modality of the high-dimensional solution space is characterised by many local optima, making a global optimum difficult to find and even harder to prove, especially as simulation times are long. Computational resources and the time allocated to calibration are usually insufficient to fully explore the space. However, improving the computational methods of calibration itself by thinking outside the box of traditional sampling-based optimisation algorithms, such as Genetic algorithms or Markov-Chain Monte Carlo methods, can aid in *progressing* towards the true optimum. Advances in high performance computing and the increased availability and accessibility of machine learning algorithms in high level programming languages carry a great potential that should be exploited. Ideas can again be borrowed from disciplines that have long addressed complex optimisation problems, such as engineering, physics and computer science (e.g. for the tuning of hyperparameters of machine learning algorithms (*238*)).

In general, the use of probabilistic emulators to predict goodness-of-fit, rather than explicitly simulating every sparse sample, allows for quick approximations of the solution space and cheap evaluation of the simulator (or its likely performance) in many more locations. This increases the confidence that the final parameter set approximates the global optimum. In Chapter 2, I propose a novel approach to calibrate disease transmission models via a Bayesian optimisation framework employing machine learning emulator functions to guide a global

search over a multi-objective landscape. The new approaches (GP-BO and GPSG-BO[3]) outperformed previous calibrations using a sampling-based asynchronous genetic algorithm (*64*) both in terms of computing time and final goodness-of-fit. Comparing the performance of fast, exploitative GPSG-BO to that of traditional (*no regret*) GP-BO and that of the genetic algorithm suggests that the main crux of calibrating detailed simulators lies in the complexity and sparse possible sampling density of the solution space. For GPSG-BO, the efficacy of adaptive sampling is compromised by an underestimated full predictive uncertainty. This leads to overly exploitative behaviour early in optimisation, where sampling close to the point estimate of the predicted optimum is overemphasised, rather than exploring the entire parameter space. GPSG-BO so to say presents a *quick-and-dirty* approach compared to a thorough exploration of GP-BO (given the appropriate acquisition function). By extension, the possibility for denser sampling is why emulator-based Bayesian optimisation approaches outperform the sampling-based algorithms.

Possible extensions to the research presented in Chapter 2 fall into two categories: improving the calibration algorithm, and revisiting the calibration task itself.[4] The former promises multiple opportunities for extensions and improvements on speed. Even the emulation approach presented here only sparsely samples the space at five million samples in 23 dimensions. This is equivalent to a sampling resolution of only two samples in each dimension. It is therefore likely that the *true* final goodness-of-fit could be further improved. A possible solution would be to add an optimisation algorithm (e.g. stochastic gradient descent) to the emulator. However, the emulator is not optimised for performance at all output values, but to locate optima. It is therefore unclear to what extent adding an optimisation algorithm to the emulator would improve the method. Further, if the emulator optimisation was added at every iteration (in place of an acquisition function) this would drastically increase the required computation time and make the methodology inefficient and time-consuming. An alternative extension would be to employ an emulator where the computational requirements scale more favourably for large samples. Neural nets, for examples, scale linearly with the number of data points (*98*) (compared to the cubic scaling of Gaussian processes). While the uncertainty quantification required for acquisition is technically challenging for neural nets, pre-coded libraries in high-level languages exist. In fact, multiple implementations for efficient Bayesian optimisation using neural nets (*98*) written in *Python* are currently publically available through GitHub (such as *nn-bayesian-optimization* by the user *RuiShu*, *pybnn* by *automl*, or *BayesOpt* by *bpiyush*). Implementation in *Python* rather than *R* would further improve efficiency in parallelisation[5] and options for using GPU computing could further improve computing speed. Applying this to the calibration of disease simulators such as *OpenMalaria* would

[3]GP-BO = Bayesian optimisation using a heteroscedastic GP emulator; GPSG-BO = Bayesian optimisation using a GP stacked generalisation emulator, with NN, RF and MARS *level 0* learners and a heteroscedastic GP *level 1* learner. See 2

[4]The calibration task includes decisions around the number of calibration parameters or the weighting assigned to epidemiological objectives.

[5]Expert opinions gathered from discussions surrounding the research of this thesis suggests that in most implementations of parallelisation in *R*, the global environment (or at least the required objects) is copied into each parallel threads. This slows down computation itself and leads to vast increases in memory requirements. In

allow predictions at a much higher resolution, multiplying the number of possible prediction location compared to the current methodology.

Apart from algorithm improvements, future work should consider the fitting task itself and this should likely be prioritised. Currently, the only available solution to break the curse of dimensionality is to circumvent it by explicitly reducing the complexity of the problem itself (*66*). This can be achieved by either (progressively) limiting the ranges of parameter values, such that the space is sufficiently simple to be approximated by the emulator even when data is sparse, or by reducing the dimensionality itself by freezing variables (*66*). At first sight, this is, of course, undesirable as it might compromise the thoroughness with which we seek for the *true* optimum. However, the sensitivity analyses conducted in Chapter 2 reveal the non-identifiability and a lack of sensitivity of the overall goodness-of-fit to some *OpenMalaria* parameters. For example, no single objective (nor their weighted sum) was sensitive to the *Garki bias* parameter, a parameter relating to differences in data collection methodologies between settings. Revisiting the model structure itself and investigating the potential for fixing or removing parameters may be a laboursome task but would be highly desirable to simplify the optimisation problem itself. Further, the multi-objectivity and the weighting of these objectives should be further explored. The sensitivity analysis in Chapter 2 shows that a multi-objective parallel approach (as opposed to modular calibration) is required to account for the covariation of and interdependencies between parameters and model components. However, the weighting is currently manually forced. Implementing a dynamic approach such as Pareto fronts with a multi-objective *Expected improvement* acquisition function could provide a more objective, scientific framework to approaching this multi-objective calibration problem.

Overall, the research presented here and the use of machine learning for model calibration present substantial advances for the mathematical modelling of infectious diseases. This approach can also be used to address other complex optimisation problems outside calibration, such as questions surrounding the optimal allocation of interventions. For now, Bayesian optimisation-based algorithms are the present and immediate future of solving the multi-objective calibration of complex disease simulators. However, the algorithms and fitting approach itself can and should be continuously improved.

6.4 Predicting elimination

Despite stalling progress in malaria control, the global push for elimination remains. Malaria modelling must co-evolve with the changing public health landscape to ensure its continued usefulness for supporting decision-making. The analysis around achieving elimination

Python, implementations of parallelisation exist where each sub-thread can still access the global environment in the parent thread.

using RCD in Chapter 5 sets an important example of how machine learning-augmented simulation modelling can assist in the strategic evaluation of novel or currently discussed tools. Predicting the probability (or mathematical *risk*) of elimination is challenging for three reasons. Firstly, technical challenges exist surrounding the prediction of a rare, stochastic, binary output. Secondly, simulation models themselves may struggle structurally to capture elimination. Thirdly, there is a lack of data to support and validate a model's performance in capturing elimination. To address these challenges, it is worthwhile to broaden our horizons and consider solutions developed for conceptually similar problems in other disciplines with (occasionally longer) traditions of mathematical modelling. Examples include the calculation of failure risks in structural integrity analyses in engineering (*58*, *237*, *239–243*) or predicting the probability of floods in weather forecasting (*244*). The approaches employed to tackle these problems can serve as inspiration for developing new solutions to predicting elimination probabilities in disease modelling.

Rare-event probabilities are technically challenging to quantify through modelling unless large numbers of computationally expensive simulations are run. The experimental approach therefore plays a crucial role in quantifying elimination probabilities. Turning to the engineering literature on quantifying failure probabilities, generally two main approaches are suggested: Sampling-based Monte Carlo methods or approximation through surrogate modelling (*237*). For direct Monte Carlo sampling, large numbers of samples and replicates are required to calculate the proportion of simulations in the failure (or here: elimination) domain (*241*). The number of required samples is proportional to the inverse of the failure probability and a rule of thumb for reasonable accuracy in structural integrity analyses is to generate at least 10 failure samples (*241*). Applied to disease modelling, if the probability of elimination is 20% in an area of parameter space, 50 replicates would be required to derive sufficiently accurate estimates (*241*).[6] While the definition of a *sufficiently accurate* prediction may not be translatable between engineering and disease modelling, this illustrates the problem of deriving rare event probabilities through sampling the simulation model alone: The computational cost is high. The computational demands can be alleviated by employing mathematical extensions of Monte Carlo methods (such as line sampling or subset simulation(*237*)), but the requirement for exceedingly large numbers of simulator evaluations remains (*76*, *241*).

The preferable option is to turn to machine learning-augmented simulation modelling to limit the required number of expensive simulator evaluations (*241*) (see Chapter 5 and section 6.2). However, the usefulness of a surrogate-based solution depends on a combination of the emulation task, experimental design choices and *the right* surrogate. In Chapter 5, I employ RFs and SVMs trained on a static sample of *OpenMalaria* simulation outputs in a supervised classification problem. Machine learning allows interpolation between sparse simulation results and the generation of probability-of-elimination (*class probability*) heatmaps that would

[6] Explanation: If the probability of elimination is 20%, one in five simulations will reach elimination. If we define sufficient accuracy as having at least 10 failure samples, this yields a total requirement of at least 50 samples

otherwise have required many more simulations. At first glance the problem appears as one of binary classification, requiring a classification algorithm (as implemented in Chapter 5). However, for the quantification of rare event probabilities in engineering using support vectors, it has been shown that the problem ought to be treated as a regression rather than a classification task as this enables the assigning of a value to the class probability (*76*). Alternatively, GP algorithms should be assessed for their suitability because of their interpolation ability, flexibility to approximate arbitrary functions with high level of accuracy, and ability to simultaneously provide local uncertainty measures for model predictions (*243*, *245*). Additionally, straightforward GP-based sensitivity analyses can also reveal the importance of each parameter to a class outcome, in our case, elimination (*239*). In Chapter 5, I employ a stationary version of machine learning-augmented simulation modelling to simplify the workflow, where I uniformly sampled the space, and this initial training set is not subsequently extended. Performance could likely be improved through the over-sampling at the boundary region and through adaptive sampling. Adaptivity (e.g. adaptive importance sampling) would initially quickly establish the *safe* domain (non-elimination) and subsequently focus on the *failure* (elimination) domain. (*76*, *242*, *243*). Specifically, adaptive GP regression surrogates have been shown to perform well in the estimation of small failure probabilities (*237*) and would likely be promising in the quantification of elimination probabilities.

However, even the most elegant analytical and statistical approaches are dependent on an appropriate quantitative foundation. In the context of surrogate modelling for elimination, it must be ensured that the underlying simulation model can adequately capture elimination. This is contingent on the model structure and experimental covariates, as well as the data used for calibration. Deterministic models of disease transmission can establish very low stable endemicity and fail to capture stochastic extinction (*20*). These are therefore inappropriate for modelling elimination. Stochasticity, such as captured by *OpenMalaria*, is a crucial model feature for predicting elimination (*20*). Additionally, stochastic extinction depends on population size, an attribute that is not sufficiently addressed by most simulation models of malaria (*20*). Specifically, the ecological and population genetic consequences of small population sizes in both the human host and the parasite are often not captured. Small parasite population sizes, for example, mean that individuals face frequent exposure to the same genotypes, which results in a decreased probability of infection (*20*, *246*, *247*) and an acceleration towards elimination when transmission is very low. In Chapter 5, I circumvent *OpenMalaria*'s inability to explicitly capture small foci of transmission by using a targeting ratio, but this cannot account for population genetic fragmentation and associated biological and epidemiological consequences. Additionally, even if models explicitly captured these mechanisms, little data is available to support the calibration. By nature of the problem, elimination is rare. Few regions have eliminated malaria in recent years and publicly available data on elimination progress and interventions is extremely scarce. Countries that have successfully achieved elimination, like China, often employed a whole array of interventions that accompanied systemic changes including urbanisation, sociological, economical, and ecological changes. In light of so many simultaneous changes, it becomes impossible to discern which interventions can be attributed

with achieving elimination. These relationships and their respective roles in achieving the elimination of an infectious disease could only be disentangled with more data - which is not available. It is therefore unclear how well *OpenMalaria* or any other malaria transmission model capture low-transmission settings and elimination itself.

This ultimately yields the question how modelling can guide public health decision-making for elimination if current models struggle to predict elimination. For decision-making, it is inappropriate to mathematically force *precise* predictions if the underlying model is not equipped to answer the question. In this thesis, I propose an analytical framework on the prediction of elimination using simulation modelling in combination with surrogate models. This cannot solve the challenges associated with data scarcity surrounding elimination and the resulting qualitative uncertainty of *OpenMalaria* predictions in the elimination domain. However, the use of surrogates enables sensitivity analysis that can provide some insights into implementation strategies and the relative importance of different tools on the path to elimination, even if the probability of achieving it remains unknown.

6.5 Interventions for elimination: Modelling versus reality

Between 2000 and 2015, 19 previously malaria-endemic countries reached elimination and global malaria deaths fell by 50% (*248*). Affirming discussions around elimination and eradication, the WHO established a strategic advisory group to analyse the feasibility and expected cost of global eradication in 2016 (*248*). However, as of 2021, progress towards elimination has stalled, with no substantial reductions in malaria incidence over the last five years (*22*, *248*). With current interventions, it is estimated that there will still be 11 million cases of malaria in 2050. As the WHO continues to aim for eradication, the current strategy emphasises the need for new interventions, locally adapted solutions to control and elimination, and the strengthening of surveillance systems (*248*). Making the strengthening of surveillance systems a central pillar of eradication from the get-go aims to account for the final stages of elimination, which rely on interrupting transmission in the hardest-to-reach places and residual transmission pockets. In a successful surveillance-response system, it only takes one index case to trigger a focal follow-up intervention that will ideally target a complete, previously undetected transmission cluster. This explicit focus on interrupting residual transmission and preventing re-introduction comes as result of the hard lessons learned from the GMEP that saw resurgence after initial elimination in many places (*248*).

To provide strategic guidance around the risk of resurgence and elimination, mathematical models must be able to incorporate adaptive surveillance-response interventions. *OpenMalaria* can capture most antimalarial interventions currently in existence and offers a flexible framework for incorporating new interventions (such as new vector control strategies, treatments, chemoprevention methods or vaccines) in a tool agnostic manner. However, *OpenMalaria*

does not capture spatial patterns of infection and therefore it was not possible to model the focal (reactive) deployment of interventions. The unavailability of surveillance-response interventions limited the ability of *OpenMalaria* to support decision-making around the final stages of elimination. In Chapter 5, I present a methodology to circumvent the lack of spatiality using a *targeting ratio* 5.1. The effect of case clustering is captured by simulating an intervention (here: test-and-treat) on random individuals from the population, while assuming that these individuals have a higher-than-population-average probability of being infected because of their proximity to the index case. The targeting ratio framework is highly flexible. Provided data on the estimated relative effectiveness of an intervention as a function of distance from the surveillance-response trigger, this framework can be adapted to most other implementations of surveillance-response interventions. These could, for example, include reactive vector control. This research therefore presents a much needed addition to *OpenMalaria* that is essential for simulating elimination strategies and opens opportunities for guiding decision-making around suitable implementations of RCD.

Current implementation strategies of surveillance-response interventions vary greatly (*35*, *38–41*). In Chapter 5, I present an initial analysis of the potential of reaching elimination with different implementation strategies of test-and-treat-based RCD. I show that RCD leads to a sustained reduction in the overall malaria burden. By extrapolation, it could thus aid in controlling imported infections and stabilising the disease-free state. However, the results also highlight that reaching predicted elimination exclusively with RCD is highly resource intense and only feasible in few settings. Re-emphasising that the suitability of RCD is limited to the final stages of elimination, Chapter 5 mandates caution regarding the implementation of surveillance-response interventions in the interest of effectively investing valuable resources. Specifically, I highlight the importance of adequate access to care and suggested that this should be prioritised over increasing the surveillance response radius, which to some measure is both logical and intuitive. In fact, in Chapter 5, I provide a mathematical derivation that is founded in the decreasing probability of an individual being a case (and thus able to contribute to onwards transmission) between passively detected cases, index cases, and follow-up cases.

However, the weighting of (re-)actively against passively tested and treated individuals fails to take into account the complexity and operational reality of implementing either intervention. The comparison between health system strengthening and implementing RCD is hardly quantifiable. Simple increases of the modelling parameter *care seeking probability* in reality translates to a highly complex systemic challenge. Establishing access to formal care in the vicinity of every individual may require the stationing of community health workers or treatment centres. The logistics of setting up such additional points of care would not be in the hands of a single organisation or campaign focused on combatting a single disease (such as malaria), but be shared between organisations and campaigns targeting different diseases and the health authorities. As much of public health in this area is currently siloed by disease, the required involvement of many different stakeholders is likely to slow progress. Additionally,

it is not guaranteed that health system strengthening would reach asymptomatic or mildly symptomatic individuals without further incentives to seek care. The mathematical trading-off of parameters describing health system strengthening versus RCD is therefore short-sighted because it fails to take into account the differences in operational complexity associated with both interventions.

This analysis and the discussion around the results illustrate two things: Firstly, that on the road to elimination, RCD can be a valuable complement to an existing strong health care system and secondly, that it can be misleadingly easy to base public health suggestions solely on modelling results. Modellers should bear in mind the practical implications of their theoretical results and that what is modelled may be hard or impossible to translate into reality. This further emphasises the need for constant dialogue between modellers and those involved with policy development and implementation. The dialogue should address the technical aspects of modelling such as the development of a standardised modelling framework, the assessment of models (and their differences) and their suitability to answer to specific public health questions, as well as practical considerations of local and systemic conditions like operational feasibility.

6.6 Not all model evidence is created equal

Throughout this thesis, I illustrate and discuss different applications of mathematical modelling from capturing disease transmission to investigating intervention success and predicting elimination. I show that modelling has the potential to bridge evidence gaps for public health decision-making created by sparse contextual data representing complex systems. The ability of modelling to provide rapid answers addressing diverse and urgent public health questions is being increasingly recognised by guideline developers. Between 2007 and 2015, 23% of guidelines approved by the Guidelines Review Committee referred to modelling (*2*). However, the value of model predictions in a public health context is contingent on its assumptions and the data used for calibration. One criticism of modelling evidence is that the process of evidence generation is often perceived as overly complex *black box* guesswork with intransparent assumptions and implications (*2*). Unlike methodologies for observational studies, mathematical modelling is largely unregulated and the responsibility of clear communication around the correct interpretation of results lies largely with the modellers themselves (*20*). Models influencing policy are currently rarely formally assessed for quality, and quality criteria are often lacking (*2*). This presents a major hurdle to the incorporation of modelling evidence in decision-making. Its integration should therefore follow a systematic and transparent framework that is able to identify and evaluate relevant models on a case-by-case basis (*2*).

To ensure credibility in model-generated evidence, any model and modelling study must be openly assessed with regards to the experimental conceptualisation of the problem, model structure, calibration data, different dimensions of model uncertainty, methodological transparency and external validation efforts. Firstly, a model must be structurally suitable to analyse a question of public health relevance. While some extrapolation outside the calibration data is generally possible, the limitations of this must be recognised and openly communicated. For example, as explored in Sections 6.1 and 6.4 of this chapter, the value attributed to low-transmission and elimination predictions generated by malaria transmission simulators should be carefully considered. Secondly, uncertainty must be addressed in the evaluation of models. Uncertainty may be reducible or irreducible, the former referring to uncertainty stemming from knowledge gaps or the calibration data, and the latter presenting the inherent variability of the system (*235*). These classes of uncertainty can be difficult or impossible to discern. As uncertainty around predictions directly influences the interpretation of modelling evidence, clear communication surrounding the calibration data and its shortcomings are required. Chapters 3 and 4 address these for *OpenMalaria*.

Doubts surrounding the value attributable to evidence generated using one fixed model can be mitigated through ensemble modelling or multimodel approaches. No single omnipotent model exists. Rather, (in the field of malaria) a whole suite of detailed simulation models have been developed by different research groups in different contexts (*4*, *5*, *89–91*). These models are not harmonised in structure and have (in part) been calibrated to different data and using different methodologies. Differences in predictions between models reflect these modelling choices and as well as uncertainty in calibration data (*20*). This diversity bears both opportunities and challenges. Consensus modelling studies and ensemble approaches are both time consuming and technically challenging (*20*), but highly valuable for the exploration of ranges of possible outcomes. The models carry different characteristics and the shortcomings of one model can be balanced out by the strengths of another in ensemble modelling approaches. Ideally, consensus modelling will decrease the dependence of experimental outcomes on model development choices and increase robustness and confidence in the predictions. Examples of ensemble modelling efforts aiming to provide consensus answers to public health questions in malaria include modelling of the impact of the RTS,S vaccine (*11*), defining the age-specific malaria prevalence-incidence relationship (*53*), and development of an ensemble of *OpenMalaria* model variants with differing biological and epidemiological assumptions (*9*).

Despite individual research efforts in ensemble modelling (e.g. (*11*)), the urge for a general, systemic change towards harmonised multi-model approaches and methodological transparency with clear guidelines is strong (*2*, *135*). A recently proposed (but not widely enforced) framework for formal model assessment addresses many of the aspects outlined above, such as the need for models to be internally or externally validated or to explore variability in assumptions (*135*). However, the choice of what constitutes a suitable model is left to the researcher based on pre-defined inclusion and exclusion criteria (*135*). This is plausible, as the

suitability of a model to answer a public health question will depend on the question itself. However, while this framework provides guidance on the comparison of appropriate models concerning a specific question, the fundamental problem of how a *good model* ought to be defined remains largely unanswered. This has not been addressed in a holistic manner, despite being a challenge for all mathematical models within the field of malaria and outside.

A potential explanation why a holistic approach has not been developed so far is that in many instances, thorough model evaluation finds little space outside original model development in the output-driven research reality. In the dissemination of modelling studies (be it at conferences or in publications), the focus is usually on a model's potentials or specific applications. Publications provide technical details, but they are often hidden in supplementary material, which is not exposed to the same degree of scrutiny during the peer-review process as the main manuscript. As a result, many models are poorly documented.[7] Thus, understanding different modelling approaches and workflows and assessing and comparing underlying assumptions is laboursome and time consuming.

The conclusions of these considerations on evaluating model evidence are twofold: Firstly, research should urge for transparency and dialogue on model development and applied modelling methodologies. This thesis aims to provide a step towards this goal. Specifically, the works in Chapter 2 present a generalisable powerful calibration algorithm whose code is publicly available (`github.com/reikth/BayesOpt_Calibration`). Additionally, I provide a library of novel calibration data in Chapter 3, which contains sufficient information to be incorporated into the calibration of most malaria simulators. Secondly, wherever possible, consensus modelling should be preferable over single-model experiments to improve the robustness of predictions. To bypass the need for time-consuming technical analyses of models developed by other research groups, this should be achieved through collaborations. With many stakeholders involved, projects may become more difficult to manage, but it would be a worthwhile investment for the good of public health.

6.7 Conclusion: Setting the stage for a holistic future of mathematical modelling

Throughout this thesis, I showcase the benefits, limitations, and pitfalls of using modelling for public health decision-making. Altogether, the presented research constitutes powerful and combinable advancements in disease modelling and contributes to keeping *OpenMalaria* (and other malaria models) up-to-date with new computational methods and global health developments. Overarching concepts that guided my research were to consider the bigger picture of conducting research for a public health purpose, to question assumptions and

[7] *OpenMalaria* is a notable exception with a public Wiki that references all relevant papers relating to model development and calibration.

investigate their implications, to draw inspiration from other disciplines (e.g. 6.4), and to integrate modern machine learning methodologies where appropriate.

In the reality of public health decision-making, model predictions are often met with scepticism. An inherent criticism is their perceived *black box* nature, with heavy reliance on (sometimes intransparent) assumptions, input data, and methodologies (*2*, *249*). The dichotomy between the conviction felt by researchers and the scepticism of decision makers and the general public is a result of dysfunctional communication. All modelling is accompanied by inherent uncertainties and results can be sensitive assumptions made during development and application. However, this should not distract from the value of modelling as a support tool in decision-making where data is insufficient. There is thus a need for clear and transparent communication around model development, simulation methods, and the implications of results. This also requires researchers not to use previously developed models blindly. Their functional properties and decisions around development should first be diligently analysed. Based on this, a model's suitability to answer specific scientific questions should be assessed on a case-by-case basis before application.

Multiple challenges persist surrounding the role of mathematical modelling in supporting decision-making on the path to achieving malaria eradication. These concern the technicalities of mathematical modelling and communication around it, data scarcity, but also the broader public health context. While the technical challenges of predicting elimination and the urge for transparent communication and collaboration can be addressed, there are challenges beyond those that can be solved by modelling. Data scarcity, particularly on low transmission settings and elimination, limits the extent to which models can be parameterised to these. On the implementation side, progress towards elimination is threatened by the need to sustain national and international funding, political instability in high burden areas, the emergence and spread of ACT resistance of the parasite and insecticide resistance of the vector (*250*). Modelling is a valuable tool in supporting public health decisions required to address these challenges, e.g. investigating areas of the parameter space for target product profiles of new interventions and strategic deployment of interventions. Ultimately, however, computational and methodological advances are only one building block of modelling-supported decision-making in health. Modelling requires a breadth of diverse data, otherwise models remain but theoretical constructs. These data ought to be shared, code libraries ought to be public, and unified standards for best modelling practice and communication on uncertainties ought to be established. This thesis explores infectious disease modelling from its data foundations to algorithm development to applied modelling, and illustrates the requirement for multidisciplinary collaborations in the global fight against infectious diseases.

7 Publications, conferences and other projects

7.1 List of publications

Publications relating to the research of thesis

1. Reiker T, Chitnis N, Smith T. "Modelling reactive case detection strategies for interrupting transmission of Plasmodium falciparum malaria". Malaria Journal 18:259 (2019). DOI: 10.1186/s12936-019-2893-9
2. Reiker T, Cameron E, Golumbeanu M, Shattock A, Burgert L, Smith T, Filippi S, Penny M. "Emulator-based Bayesian optimization for efficient multi-objective calibration of an individual-based model of malaria". Nature Communications 12:7212 (2021). DOI: 10.1038/s41467-021-27486-z
3. Reiker T, Burgert L, Cameron E, Runge M, Smith T, Penny M. "Calibrating infectious disease models to real-world data: Context matters". In preparation for Trends in Parasitology (2022)
4. Chitnis N, Pemberton-Ross P, Yukich J, Hamainza B, Miller J, Reiker T, Eisele TP, Smith TA. "Theory of reactive interventions in the elimination and control of malaria". Malaria Journal 18:266 (2019). DOI: 10.1186/s12936-019-2882-z

7.2 List of conference contributions and talks

Presentations relating to the research of thesis

1. Reiker T, Chitinis N, Smith T. "Achieving Plasmodium falciparum malaria elimination using reactive case detection: A modelling study". Speed Talk and Poster. MIM Pan African Malaria Conference. Dakar, Senegal (2018)
2. Reiker T, Chitinis N, Smith T. "Modelling reactive case detection strategies for interrupting transmission of Plasmodium falciparum malaria". Poster. European Congress on Tropical Medicine and International Health. Liverpool, UK (2019)
3. Reiker T, Cameron E, Golumbeanu M, Filippi, S, Penny M. "Novel approaches in parameterising individual based models". Conference Presentation. Epidemics. Charleston, South Carolina, USA. (2019)

4. Reiker T, Cameron E, Golumbeanu M, Shattock A, Filippi, S, Smith T, Penny M. “Machine learning to parameterize individual-based models of malaria transmission”. Digitalization and Infectious Diseases. Basel, Switzerland (2020)

5. Reiker T, Burgert L, Runge M, Smith T, Penny M. “Translating observational studies for disease modelling”. American Society of Tropical Medicine and Hygiene, 69th Annual Meeting. Virtual / Toronto, Canada (2020)

7.3 Other

Scientific contributions concurrent with, but not relating to the research of thesis

1. Kuchenbaecker K, Telkar N, Reiker T, Walters R, Lin K et al. “The transferability of lipid loci across African, Asian and European cohorts". Nature Communications. 10(1):4330 (2019). DOI: s41467-019-12026-7

2. Shattock AJ, Neihus R, Reiker T, Golumbeanu M, Paurat M, Hay J, Lipsitch M, Bonhoegger S, Chitnis N, Penny MA. Combination packages of social distancing, testing, contact tracing, and facemasks to mitigate SARS-CoV-2 resurgence: a modeling study of Switzerland. In preparation. (2022)

3. Burgert L, Laager M, Reiker T, Golumbeanu M, Moehrle JJ, Penny MA. ”Model-informed target product profiles of long-acting-injectables for use as seasonal malaria prevention”. PLOS Glob Public Health 2(3): e0000211 (2022). DOI: 10.1371/journal.pgph.0000211

4. Burgert L, Reiker T, Lee TE, Moehrle JJ, Gobeau N, Penny MA. ”Antimalarial drug development and the missing indices for cure”. In preparation. (2022)

A Supplement to: Machine learning to calibrate individual-based infectious disease models

Supplementary Materials for

Machine learning approaches to calibrate individual-based infectious disease models

Theresa Reiker, Monica Golumbeanu, Andrew J. Shattock, Lydia Burgert, Thomas A. Smith, Sarah Filippi, Ewan Cameron, Melissa A. Penny.

Correspondence to: melissa.penny@unibas.ch

This file includes:

A.1 Supplementary text 1: Malaria Transmission Model

A.1.1 Main features

We test our calibration algorithm on *OpenMalaria*, an individual-based model of malaria dynamics. To provide context of the model's structure and the role of the fitted parameters (see supplementary text 1), we here briefly describe its main features and key equations. This description is adapted from that provided in Smith et al. 2012 (*64*) and Smith et al. 2006 (*3*). Full details of all model components can be found in *The American Journal of Tropical Medicine and Hygiene*, Volume 75, Issue 2 Supplement (2006).

OpenMalaria features discrete individual-based stochastic simulations of malaria in humans in 5-day time steps. Every infection and individual are characterised by a set of continuous state variables, namely, parasite densities, infection durations, and immune status. Key processes and relationships regarding the course of infection simulated by model include the attenuation of inoculations, acquired pre-erythrocytic immunity, acquired blood-stage immunity, morbidity (acute and severe) and mortality (malaria-specific and indirect), anemia, and the infection of vectors as a function of parasite densities in the human. Other model components include a vector model and a case management system. All individual components have previously been well documented (*3*, *64*). A visual summary of the model with references to further details on each component is provided in Figure A.1.

In our current recalibration only the original (base) model variant is used to test our new approach (*64*). Parameters estimated during the calibration process are highlighted and summarised in Table A.1 at the end of this section. Other parameter values were drawn from the literature or were calibrated to separated data: for example, the empirical parasite density model of Maire et al. 2006 (*31*) was calibrated to malariatherapy (*185*) data and not recalibrated at the population level.

A.1.2 Infection of the human host

The seasonal pattern of entomological inoculation rate (EIR) determines seasonal pattern of transmission and thus the parasite densities in the individual, modified by natural or acquired immunity and interventions (*63*).

Differential feeding by mosquitoes depending on body surface area

In the base model, the expected number of entomological inoculations experienced by individual i of age a at time t is

$$E_a(i,t) = \frac{E_{\max}(t)\, A(a(i,t))}{A_{\max}} \tag{A.1}$$

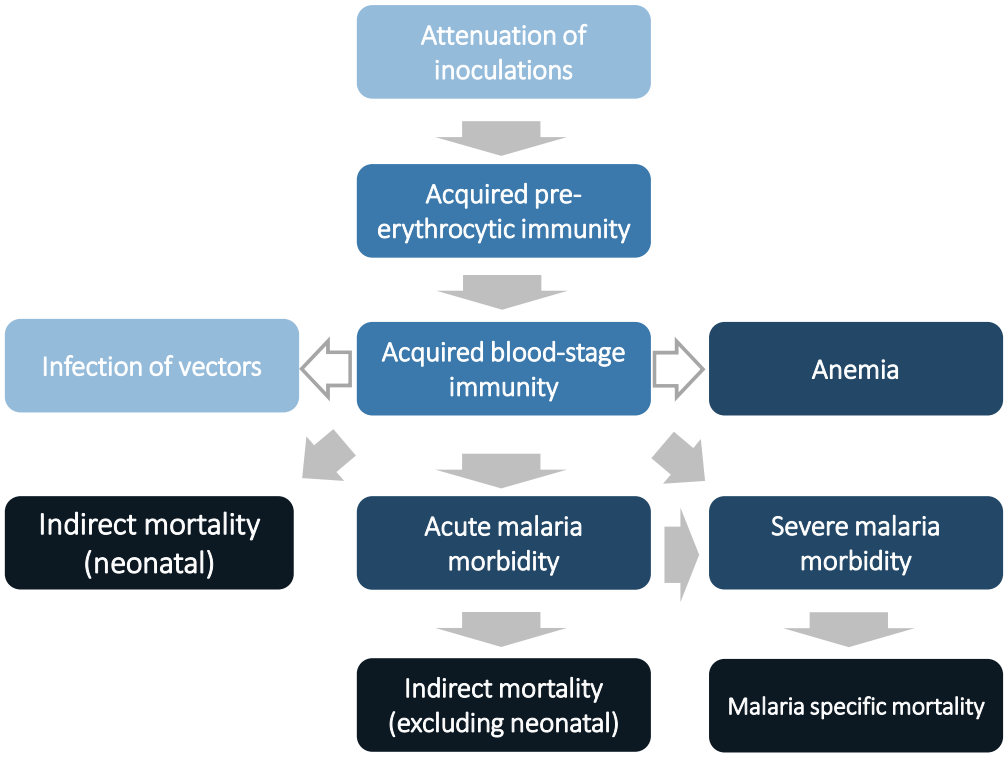

FIGURE A.1: Visual summary of OpenMalaria with references to original publications on the model components

Adapted from Smith et al. 2006, Fig.3 (*63*). References from top to bottom and left to right: Attenuation of inoculations (*183*), Acquired pre-erythrocytic immunity (*183*), Infection of vectors (*211*, *212*), Acquired blood-stage immunity (*31*), Anemia (*213*), Indirect mortality (neonatal) (*214*), Acute malaria morbidity (*3*), Severe malaria morbidity (*215*), Indirect mortality excluding neonatal (*215*), Malaria specific mortality (*215*)

where $E_{\max}(t)$ refers to the annual entomological inoculation rate (EIR) computed from human bait collections on adults and $A()$, is the individual's availability to mosquitoes, assumed to be proportional to average body surface area, depending only on age . $A(a(i,t))$ increases with age up to age 20 years where it reaches a value of $A_{\max}$ (the average body surface of people $\geq$ 20 years old in the same population).

The biting rate in relation to human weight is based on data from The Gambia published by Port and others (*251*), where the proportion of mosquitoes that had fed on a host were analysed in relation to the host's contribution to the total biomass and surface area of people sleeping in one mosquito net (*183*).

Control of pre-erythrocytic stages

The number of infective bites received per unit time for each individual i , adjusted by age, is given by Equation A.1 above. A survival function $S(i,t)$ defines the probability that the

progeny of an inoculation survives to give rise to a patent blood stage infection, i.e. the proportion of inoculations that result in infections or the susceptibility of individual i at time t . The force of infection is modelled as

$$\lambda\left(i,t\right)=S\left(i,t\right)E_{a}\left(i,t\right) \tag{A.2}$$

where $E_{a}\left(i,t\right)$ is the expected number of entomological inoculations endured by individual i at time t , adjusted for age and individual factors, and the number of infections $h\left(i,t\right)$ acquired by individual i in five-day time step t , follows a Poisson distribution:

$$h\left(i,t\right)\sim Poisson\left(\lambda\left(i,t\right)\right) \tag{A.3}$$

The susceptibility of individual i at time t , $S\left(i,t\right)$ is defined as:

$$S\left(i,t\right)=\left(S_{\infty}+\frac{1-S_{\infty}}{1+\frac{E_{a}(i,t)}{E^{*}}}\right)\left(S_{imm}+\frac{1-S_{imm}}{1+\left(\frac{X_{p}(i,t)}{X_{p}^{*}}\right)^{\gamma_{p}}}\right) \tag{A.4}$$

where S_{imm}, X_{p}^{*}, E^{*}, γ_{p} and S_{∞} are constants representing the lower limit of success probability of inoculations in immune individuals, critical value of cumulative number of entomologic inoculations, critical value of $E_{a}\left(i,t\right)$, steepness of relationship between success of inoculation and $X_{p}\left(i,t\right)$, and, the lower limit of success probability of inoculations at high where $E_{a}\left(i,t\right)$, respectively. Here

$$X_{p}\left(i,t\right)=\int_{t-a(i,t)}^{t}E_{a}\left(i,\tau\right)d\tau \tag{A.5}$$

S_{∞} and E^{*} are fixed to $S_{\infty}=0.049$, and $E^{*}=0.032$ inoculations/person-night and are detailed in (*183*).

Course of infection in the human host

The model for each individual infection j in host i comprises a time series of parasite densities. The base model for infection within humans is described in Maire et al. 2006 (*31*). In brief, the duration of each infection, $\tau_{\max}$ is sampled from

$$\ln\left(\tau_{\max}\left(i,j\right)\right)\sim Normal\left(5.13,0.80\right) \tag{A.6}$$

parameterised against malaria therapy data (*246*) and detailed in Maire et al. 2006 (*31*). In the absence of previous exposure or concurrent infections, the log density of infection j in host i at each time point, $\tau = 0, 1, \ldots, \tau_{\max}\,(i,j)$ is normally distributed with expectation

$$\ln\left(y_0\left(i,j,\tau\right)\right) = \ln d\left(i\right) + \ln y_G\left(\tau, \tau_{\max}\right) \tag{A.7}$$

where $y_G\left(\tau, \tau_{\max}\right)$ is taken from a statistical description of parasite densities in malariatherapy patients and $d\left(i\right)$ describes between-host variation with a log-normal distribution with variance σ_i^2 .

We consider the possibility of multiple concurrent infections in the same individual at the same time. Exposure to asexual blood stages is measured by

$$X_y\left(i,j,t\right) = \int_{t-a}^{t} Y\left(i,\tau\right) d\tau - \int_{t_{0,j}}^{t} y\left(i,j,\tau\right) d\tau \tag{A.8}$$

where $Y\left(i,\tau\right)$ is the total parasite density of individual i at time τ and $y\left(i,j,\tau\right)$ is the density of infection j in individual i at time τ and

$$X_h\left(i,t\right) = \int_{t-a}^{t} h\left(i,\tau\right) d\tau - 1 \tag{A.9}$$

In the presence of previous exposure and co-infection, the expected log density for each concurrent infection is then:

$$E\left(\ln\left(y\left(i,j,\tau\right)\right)\right) = D_y\left(i,t\right) D_h\left(i,t\right) D_m\left(i,t\right) \ln\left(y_0\left(i,j,\tau\right)\right) + \ln\left(\frac{D}{M\left(i,t\right)} + 1 - D_x\right) \tag{A.10}$$

where $M\left(i,t\right)$ is the total multiplicity of infection of in individual i at time t , and

$$D_y\left(i,t\right) = \frac{1}{1 + \frac{X_y(i,j,t)}{X_y^*}} \tag{A.11}$$

where $X_y\left(i,j,t\right) = \sum_{t-a}^{t} Y\left(i,t\right) - \sum_{t_{0,j}}^{t} y\left(i,j,\tau\right)$ (note that a continuous time approximation to this is given in the original publications (*31*, *183*) and hence measures the cumulative parasite load. Furthermore

$$D_h\left(i,t\right) = \frac{1}{1 + \frac{X_h(i,t)}{X_h^*}} \tag{A.12}$$

where, $X_h(i,t) = \sum_{t-a}^{t} h(i,\tau) - 1$, the number of inoculations since birth, excluding the one under consideration, which measures the diversity of inocula experienced by the host up to the time point under consideration.

$$D_m(i,t) = 1 - \alpha_m exp\left(-\frac{0.693a(i,t)}{a_m^*}\right) \tag{A.13}$$

which measures the effect of maternal immunity. X_y^* , X_h^* , D_x , a_m^* , and α_m are all constants estimated in the fitting process. These constants are described in Table A.1, or further in Maire et al. 2006 (*31*).

Variation within individuals described as $\sigma_y^2(i,j,\tau)$, where

$$\sigma_y^2(i,j,\tau) = \frac{\sigma_0^2}{1 + \frac{X_h(i,t)}{X_v^*}} \tag{A.14}$$

and σ_0^2 and X_v^* are constants, described in Table A.1.

The simulated density of infection j in individual i at time τ is then drawn from a normal distribution:

$$\ln(y(i,j,\tau)) \sim Normal\left(E(\ln(y(i,j,\tau))), \sigma_y^2(i,j,\tau)\right) \tag{A.15}$$

The total density of all infections in individual i at time t is then the sum of the densities of concurrent infections j

$$Y(i,t) = \sum_j y(i,j,\tau(i,j)) \tag{A.16}$$

Infectivity of the human host

The model infectivity of the human host is described in Ross 2006 where infectivity of individual i at time t is given by the distributed lag model:

$$\mathrm{Y}(i,t) = \beta_1 Y(i,t-2) + \beta_2 Y(i,t-3) + \beta_3 Y(i,t-4) \tag{A.17}$$

where t is in 5-day units and

$$\ln\left(y_g(i,t)\right) \sim Normal\left(\ln(\rho \mathrm{Y}(i,t)), \sigma_g^2\right) \tag{A.18}$$

where β_1, β_2, β_3, ρ, σ_g^2 are constants representing contributions of past infections to gametocyte densities. We define

$$\Pr\left(y_g\left(i,t\right) > \ y_g^*\right) = \Phi\left[\frac{\ln\left(\rho \mathrm{Y}\left(i,t\right)\right) - \ln\left(y_g^*\right)}{\sigma_g}\right] = \Phi\left[\frac{\ln\left(\mathrm{Y}\left(i,t\right)\right)}{\sigma_g} + \rho^*\right] \tag{A.19}$$

where Φ is the cumulative normal distribution, y_g^* is the density of female gametocytes necessary for infection of the mosquito, and $\rho^* = \frac{\ln(\rho) - \ln\left(y_g^*\right)}{\sigma_g}$ is constant (depending on the blood meal volume, gametocyte viability and system variability). Thus, the proportion of mosquitoes infected by individual i at time t is defined as

$$I_m\left(i,t\right) = \left[\Pr\left(y_g\left(i,t\right) > y_g^*\right)\right]^2 \tag{A.20}$$

and the probability of a mosquito becoming infected during any feed is

$$\kappa_u\left(t\right) = \eta \frac{\sum_i A\left(a\left(i,t\right)\right) I_m\left(i,t\right)}{\sum_i A\left(a\left(i,t\right)\right)} \tag{A.21}$$

where η is a constant scale factor and to be calibrated.

We define $\kappa_u^{(0)}\left(t\right)$ as the value of $\kappa_u\left(t\right)$ in the simulation of an equilibrium scenario to which an intervention has been applied. Let $E_{\max}^{(0)}\left(t + l_v\right)$ be the corresponding entomologic inoculation rate. $\kappa_u^{(1)}\left(t\right)$ and $E_{\max}^{(1)}\left(t + l_v\right)$ are the corresponding values for the intervention scenario. Then

$$E_{\max}^{(1)}\left(t + l_v\right) = \frac{E_{\max}^{(0)}\left(t + l_v\right) \kappa_u^{(1)}\left(t\right)}{\kappa_u^{(0)}\left(t\right)} \tag{A.22}$$

where l_v corresponds to the duration of the sporogenic cycle in the vector, which we approximate with two time steps (10 days). $\frac{E_{\max}^{(0)}(t+l_v)\kappa_u^{(1)}(1)}{\kappa_u^{(0)}(t)}$ is the total vectorial capacity)

A.1.3 Morbidity

In order to simulate the clinical state of individual i at time t , for each five-day time step 5 independent samples from the simulated parasite density distribution are drawn for each concurrent infection j .

Acute morbidity (uncomplicated clinical cases)

The model for an episode of acute morbidity was originally described in (3) and occurs in individual i at time t with probability

$$P_m\left(i,t\right) = \frac{Y_{\max}\left(i,t\right)}{Y^*\left(i,t\right) + Y_{\max}\left(i,t\right)} \tag{A.23}$$

where Y^* is the pyrogenic threshold and $Y_{\max}$ is the maximum density of five daily densities sampled during the five-day interval t .

The pyrogenic threshold changes over time following

$$\frac{dY^*(i,t)}{dt} = f_1(Y(i,t))\, f_2(Y^*(i,t)) - \overline{\omega} Y^*(i,y) \tag{A.24}$$

where $f_1(Y(i,t))$ is a function describing the relationship between accrual of tolerance and the parasite density $Y(i,y)$; $f_2(Y^*(i,t))$ describes the saturation of this accrual process at high values of Y^* and $\overline{\omega} Y^*(i,t)$ determines the decay threshold with first-order kinetics, ensuring that the parasite tolerance is short-lived (*3*).

Here $f_1(Y(i,t))$ is defined to ensure that the stimulus is not directly proportional to Y but rather that it asymptotically reaches a maximum at high values of Y :

$$f_1(Y(i,t)) = \frac{\alpha Y(i,t)}{Y_1^* + Y(i,t)} \tag{A.25}$$

At high values of Y^* , a higher parasite load is required to achieve the same increase:

$$f_2(Y^*(i,t)) = \frac{1}{Y_2^* + Y^*(i,t)} \tag{A.26}$$

Thus, the pyrogenic threshold Y^* is defined to follow

$$\frac{dY^*(i,t)}{dt} = \frac{\alpha Y(i,t)}{\left(Y_1^* + Y(i,t)\right)\left(Y_2^* + Y^*(i,t)\right)} - \overline{\omega} Y^*(i,t) \tag{A.27}$$

and the initial condition $Y^*(i,0) = Y_0^*$ at the birth of the host, where $\alpha, \overline{\omega}$ Y_0, Y_1^* and Y_2^* are targets of the calibration, and are defined in A.1.

Severe disease

The model for severe disease was described in Ross et al 2006 (*215*) and two different classes of severe episodes are considered by the model, B_1 and B_2 . $P_{B1}(i,t)$ is the probability that an acute episode (A) is of class B_1 and

$$P_{B_1}(i,t) = \Pr\left(H(i,t) \in B_1 | H(i,t) \in A\right) = \frac{Y_{\max}(i,t)}{Y_{B1}^* + Y_{\max}(i,t)} \tag{A.28}$$

where Y_{B1}^* is a constant to be calibrated and $H(i,t)$ is the clinical status of individual i at time t .

Class B_2 of severe malaria episodes occurs when an otherwise uncomplicated episode coincides with some other insult, which occurs with risk

$$F(a(i,t)) = \frac{F_0}{1 + \left(\frac{a(i,t)}{a_F^*}\right)} \tag{A.29}$$

where F_0 is the limiting value of $F(a(i,t))$ at birth and a_F^* is the age at which it is halved and both are to be calibrated.

The probability that individual i experiences an episode belonging to class B_2 at time t, conditional on there being a clinical episode at that time is

$$P_{B_2}(i,t) = \Pr(H(i,t) \in B_2 \,|H(i,t) \in A) = F(a(i,t)) \tag{A.30}$$

The age and time specific risk of severe malaria morbidity conditional on a clinical episode is then given by

$$P_B(i,t) = P_{B_1}(i,t) + P_{B_3}(i,t) - P_{B_1}(i,t) P_{B_2}(i,t) \tag{A.31}$$

Mortality

Malaria deaths in hospital are a random sample of admitted severe malaria cases, with age-dependent sampling fraction $Q_h(a)$, the hospital case fatality rate, derived from the data of Reyburn et al (2004) (*216*). The original model was described in Ross et al. 2006 (*215*).

The severe malaria case fatality in the community for age group a, $Q_c(a)$ is estimated as

$$Q_c(a) = \frac{Q_h(a)\phi_1}{1 - Q_h(a) + Q_h(a)\phi_1} \tag{A.32}$$

where ϕ_1 the estimated odds ratio for death in the community compared to death in in-patients is an age-independent constant to be calibrated and $Q_h(a)$ is the hospital case fatality rate. The total malaria mortality is the sum of the hospital and community malaria deaths.

The risk of neonatal mortality attributable to malaria (death in class D_1) in first pregnancies is set equal to $0.3\mu_{PG}$ where

$$\mu_{PG} = \mu_{\max}\left[1 - \exp\left(-\frac{x_{PG}}{x_{PG}^*}\right)\right] \tag{A.33}$$

where x_{PG} is related to x_{MG}, the prevalence in simulated individuals 20-24 years of age via

$$x_{PG} = 1 - \frac{1}{1 + \left(\frac{x_{MG}}{x_{MG}^*}\right)} \tag{A.34}$$

and x_{MG}^* and x_{PG}^* are constants.

An indirect death in class D_2 is provoked at time t, conditional on there being a clinical episode at that time with probability $P_{D2}(i,t)$ where

$$P_{D_2}(i,t) = \Pr(H(i,t) \in D_2 | H(i,t) \in A) \tag{A.35}$$

and

$$P_{D_2}(i,t) = \frac{Q_D}{1 + \left(\frac{a(i,t)}{a_F^*}\right)} \tag{A.36}$$

where Q_D is limiting value of $P_{D_2}(i,t)$ at birth and a_F^* is a constant to be calibrated. Deaths in class D_2 occur 30 days (six time steps) after the provoking episode.

TABLE A.1: Names and details of *OpenMalaria* core parameters

GA-O = Genetic algorithm optimisation, GP-BO = Gaussian process-based Bayesian optimisation, GPSG-BO = Gaussian process stacked generalisation-based Bayesian optimisation.

No.*	θ^+	Parameter	Meaning	Unit/ dimension	Prior	GA-O estimate (Smith et al. 2012, model R0001)(*64*)	New estimate **GP-BO** (Reiker et al.2020)	New estimate GPSG-BO (Reiker et al.2020)
1	–	$-\ln(1-S_\infty)$	S_∞ = Lower limit of success probability of inoculations at high $E_a(i,t)$	Proportion	–	0.051	0.051	0.051
2	–	E^*	Critical value of $E_a(i,t)$	Inoculations/ person-night	–	0.032	0.032	0.032
3	1[a]	S_{imm}	Lower limit of success probability of inoculations in immune individuals	Proportion	$\exp(N(\log(0.14),2))$	0.138	0. 196	0.036
4	3	X_p^*	Critical value of cumulative number of entomologic inoculations	Inoculations	$\exp(N(\log(1514),2))$	1,514.4	1,954.8	4,972.2
5	2	γ_p	Steepness of relationship between success of inoculation and $X_p(i,t)$	Dimensionless constant	$\exp(N(\log(1),1))$	2.037	1.291	1.871
6	23	σ_i^2	Variation between hosts on parasite densities (variance of log-normal distribution)		$\exp(N(\log(10.17),0.6))$	10.174	11.729	9.689
7	5	X_y^*	Critical value of cumulative number of parasite days	Parasite-days/ μL x 10^{-7}	$\exp\left(N\left(\log\left(3.52x10^7\right),2\right)\right)$	3.516	593.661	1.216
8	4	X_h^*	Critical value of cumulative number of infections	Infections	$\exp(N(\log(97.3),2))$	97.335	54.082	89.759
9	7	$\ln(1-\alpha_m)$	α_m = Maternal protection at birth	Dimensionless	$-\log(1-Beta(8,2))$	2.330	1.770	1.266
10	8	a_m^*	Decay of maternal protection	Per year	$\exp(N(\log(1.8),0.5))$	2.531	1.279	1.551
11	9	σ_0^2	Fixed variance component for densities	$[\ln(density)]^2$	$\exp(N(\log(0.66),2))$	0.656	5.838	1.440
12	6	X_v^*	Critical value of cumulative number of infections for variance in parasite densities	Infections	$\exp(N(\log(5),1))$	0.916	3.959	7.226
13	14	Y_2^*	Critical value of $Y^*(i,t)$ in determining increase in Y^*	Parasites/ μ L	$\exp(N(\log(5000),1))$	6,502.26	6,560.08	13,485.57
14	10	α	Factor determining increase in $Y^*(i,t)$	$Parasites^2 \mu L^{-2} day^{-1}$	$\exp(N(\log(142602),1))$	142,602	63,220.5	119,502
15	22	ν_1	Density bias (non Garki)	Dimensionless	$\exp(N(\log(0.177),0.6))$	0.177	0.123	0.159
16	–	σ_2	Mass action parameter	Dimensionless		1	1	1
17	18	$\log\phi_1$	Case fatality for severe episodes in the community compared to hospital	Log odds	$\exp(N(\log(2.09),0.3))$	0.736	0.340	0.285

continued on next page

continued from previous page

18	20[b]	Q_D	Co-morbidity intercept relevant to indirect mortality	Proportion	$\exp(N(\log(0.019), 1))$	0.019	0.019	0.023
19	19[c]	Q_n	Non-malaria intercept for infant mortality	Deaths / 1000 live births	$\exp(N(\log(49.5), 1))$	49.539	46.5095	40.163
20	21	ν_0	Density bias (Garki)	Dimensionless	$\exp(N(\log(4.79), 0.2))$	4.796	3.739	5.618
21	15	$Y^*_{B_1}$	Parasitaemia threshold for severe episodes type B_1	Parasites/ μ L	$\exp(N(\log(250000), 0.8))$	784,456	849,046	484,122
22	–	–	Immune penalty		–	1	1	1
23	–	–	Immune effector decay		–	0	0	0
24	16[d]	F_0	Prevalence of co-morbidity/susceptibility at birth relevant to severe episodes (B_2)	proportion	$\exp(N(\log(0.092), 0.5))$	0.097	0.078	0.094
25	11	$\frac{\log 2}{\omega}$	Y^* (pyrogenic threshold) half-life	Years	$\log(2)$ / $\exp(N(\log(2.52), 1))$	0.275	0.468	0.516
26	13	Y^*_1	Critical value of parasite density in determining increase in Y^*	Parasites/ μ L	$\exp(N(\log(6), 2))$	0. 597	1.665	0.477
27	–	–	Asexual immunity decay		–	0	0	0
28	12	Y^*_0	Pyrogenic threshold at birth	Parasites/ μ L	$\exp(N(\log(296.3), 1))$	296.302	90.938	201.671
29	–	–	Idete multiplier	Dimensionless	–	2.798	2.799	2.799
30	17	a^*_F	Critical age for co-morbidity	Years	$\exp(N(\log(0.225), 0.8))$	0.117	0.138	0.087

* Parameter number assigned for simulations in *OpenMalaria* scenarios, some parameters here are used in model variants and not in the base model. Listed for completeness; +Parameter number θ_i assigned for the optimisation problem. θ is drawn from the unit cube and determines the quantiles of the prior for the parameter value. [a] quantile = $\theta * 0.8372102$. [b] quantile = $\theta * 0.9999991$. [c] quantile = $\theta * 0.9986755$. [d] quantile = $\theta * 0.999963$.

A.2 Supplementary text 2: Calibration Approach and Data Summary

A comprehensive epidemiological calibration dataset was collated in order to parameterise *OpenMalaria*, an individual-based model of malaria transmission dynamics. This calibration dataset covers a total of 11 different epidemiological relationships (or objectives for fitting) that span important aspects of the natural history of malaria. Data were collated from different settings (see Table A.5 for summary) and were detailed in the original model descriptions (*3*, *63*) and a later parameterisation (*64*). A total of 61 simulation scenarios were setup to parameterise *OpenMalaria*, constructed to simulate the study surveys and study sites that yielded the calibration dataset. The study site observations were replicated in *OpenMalaria* by reproducing the timing of the surveys and their endpoints (such as prevalence and incidence) and matching simulation options to the setting with regards to transmission intensity and seasonality, vector species, treatment seeking behaviour and anti-malarial interventions. The objectives and data are further detailed below.

The parameter estimation process is a multi-objective optimisation problem with each of the epidemiological quantities in Table A.5 representing one objective. The aim of the optimisation is to find a parameter set that maximises the goodness of fit by minimizing a loss statistic computed as the weighted sum of the loss functions for each objective. Building a weighted average reduces the multiple loss terms to a single overall loss statistic, defined as:

$$F(\theta) = \sum_i w_i \sum_j f_{ij}(\theta) \tag{A.37}$$

where $f_{ij}(\theta)$ is the loss function for parameter vector θ , epidemiological quantity i and dataset j , and the weights w_i were chosen so that different epidemiological quantities contribute approximately equally to $F(\theta)$.

For the current calibration, we utilised the loss functions from Smith et al. 2012 (*64*), the loss function $f_i(\theta)$ for each objective i use either (negative) log-likelihoods or Residual Sum of Squares (RSS) with an unknown minimum. We did not updated these loss-functions in order to compare to our previous approaches.

The likelihood functions are given by

$$L(\theta|x_1,\ldots,x_n) = g(x_1,\ldots,x_n|\,\theta) = \prod_{i=1}^{n} g(x_i|\theta) \tag{A.38}$$

where the observed values are $x_1,\ldots,x_n$ and the model parameters θ . In practice, it is easier to work with the log likelihood, namely

$$\log L(\theta|x_1,\ldots,x_n) = \sum_{i=1}^{n} \log g(x_i|\theta) \tag{A.39}$$

The loss functions $f_i(\theta)$ used for each objective are detailed in the following sections.

A.2.1 Objectives: Epidemiological data and loss functions

Below we described each fitting objective in terms of the data (setting, surveys, observations, references) along with the associated loss function and original references. Table A.6 provides an overview of the 61 simulation scenarios used for calibration, and which objective they contribute to.

Age pattern of incidence after intervention

Data The data used for the calibration of objective 1 (Age pattern of incidence) consists of eight cross-sectional surveys of infection rates by age and EIR in Matsari village, capturing 12 age groups each. Matsari village was monitored entomologically for four years (Nov

1970 - Nov1973) during the Garki Project and multiple anti-malaria interventions were administered (*65*). From October 1970 to March 1972 (the baseline/pre-intervention phase), eight cross-sectional malariologic surveys of the whole village population and intensive entomologic surveillance (human bait collection of mosquitoes and dissections of the mosquito salivary glands for sporozoites) were carried out. The latter was used to estimate a baseline transmission intensity of 67 inoculations per person per year (EIR) and to derive seasonal transmission patterns. Mid-1972 marked the beginning of the intervention phase, during which an additional eight surveys were carried out at 10-week intervals (surveys 9-16). During this time, indoor residual spraying with Propoxur was carried out comprehensively in the village, along with mass treatment of the population with Sulfadoxine-pyrimethamine at 10 week-intervals immediately after assessment of individuals' parasitologic status. The experimental setup is summarised in Figure 3 of Smith et al 2006 (*183*). Incidence data (number of patent infections and number of hosts by age) from surveys 9-16 was used for our calibration.

Sites and scenario numbers: Matsari, Nigeria (30)

Original reference detailing data and model fits: Smith TA, Maire N, Dietz K, Killeen GF, Vounatsou P et al. Relationship between the entomological inoculation rate and the force of infection for Plasmodium falciparum malaria. Am J Trop Med Hyg. Volume 75, No. 2 Supplement. 2006 (183)

Loss function: Binomial Log Likelihood We denote the Binomial log likelihood for this objective to be

$$f_1(\theta) = logL(\theta) = \sum_{j=1}^{s}\sum_{k=1}^{a} P_{j,l} \log\left(\widehat{p_{j,\mathrm{k}}}\right) + \left(H_{j,k} - P_{j,k}\right) \log\left(1 - \widehat{p_{j,k}}\right) \tag{A.40}$$

where a is the number of age groups, s the number of surveys, $p_{j,k}$ the scenario data number of parasite positive hosts and $H_{j,k}$ the scenario data number of hosts for age group k and survey j . Parameter $\widehat{p_{j,k}}$ is associated with the model predictions and is given by

$$\widehat{p_{j,k}} = \widehat{P_{j,k}} \;/\; \widehat{H_{j,k}} \tag{A.41}$$

where $\widehat{P_{j,k}}$ are the predicted number of parasite positive hosts and $\widehat{H_{j,k}}$ the predicted number of hosts for age group k and survey j .

Age patterns of prevalence

Data The data used for the calibration of objective 2 (age-patterns of prevalence) consists of six cross-sectional malariology surveys conducted in the Rafin Marke, Matsari, Sugungum villages in Nigeria 1970-1972 (12 age groups each, part of the Garki Project during the pre-intervention period) (*65*), Navrongo in Ghana 2000 (12 age groups) (*252*) and Namawala 1990-1991 (*253*) and Idete in Tanzania (11 and 6 age groups, respectively) 1992-1993 (*254*). In all study sites, annual transmission intensity (EIR) and seasonal patterns were assessed using light trap

or human night bait collections and dissections of the salivary glands (see Figure 2 in Maire et al. 2006 (*31*)). In all sites except Idete, the health system at the time of the surveys treated only a small proportion of the clinical malaria episodes. In the Idete, the village dispensary was assumed to treat approximately 64% of clinical malaria (based on the published literature). During simulation, prevalence was defined by comparing each predicted parasite density with the limit of detection used in the actual study.

Sites and scenario numbers: Sugungum, Nigeria (24); Rafin-Marke, Nigeria (28); Matsari, Nigeria (29); Idete, Tanzania (31); Navrongo, Ghana (34); Namawala, Tanzania (35)

Original reference *detailing data and model fits*: *Maire N, Smith TA, Ross A, Owusu-Agyei S, Dietz K, et al. A model for natural immunity to asexual blood stages of Plasmodium falciparum malaria in endemic areas. Am J Trop Med Hyg. Volume 75, No. 2 Supplement. 2006 (31)*

Loss function: Binomial Log Likelihood We denote the binomial log likelihood for each scenario of this objective to be

$$f_2(\theta) = logL(\theta) = \sum_{j=1}^{s}\sum_{k=1}^{a} P_{j,k}\log(p_j, \mathrm{k}) + (H_{j,k} - P_{j,k})\log(1 - p_{j,k}) \tag{A.42}$$

where a is the number of age groups, s the number of surveys, $P_{j,k}$ the scenario data number of parasite positive hosts and $H_{j,k}$ the scenario data number of hosts for age group k and survey j. Parameter $p_{j,k}$ is associated with the model predictions and is given by

$$p_{j,k} = \widehat{P_{j,k}} / \widehat{H_{j,k}} \tag{A.43}$$

where $\widehat{P_{j,k}}$ are the predicted number of parasite positive hosts and $\widehat{H_{j,k}}$ the predicted number of hosts for age group k and survey j .

Age patterns of parasite density

Data The same data sources as for objective 2 (age pattern of prevalence) were used for calibration of objective 3 (age pattern of parasite density). Parasite densities in sites that were part of the Garki project (Sugungum, Rafin-Make and Matsari, Nigeria) were recorded by scanning a predetermined number of microscope fields on the thick blood film and recording how many had one or more asexual parasites visible. These were converted to numbers of parasites visible by assuming Poisson distribution for the number of parasites per field and a blood volume of 0.5 mm^3 per 200 fields. In the other studies (Idete and Namawala, Tanzania and Navrongo, Ghana), parasites were counted against leukocytes and converted to nominal parasites/microliter assuming the usual standard of 8,000 leukocytes/microliter. The biases in density estimates resulting from these different techniques were accounted for by multiplying the observed parasite densities with constant values estimated for Garki (ν_0) and non-Garki (ν_1) studies to rescale them to the values in malariatherapy patients (*255*).

Sites and scenario numbers: Sugungum, Nigeria (pre-intervention, 24); Rafin-Marke, Nigeria (pre-intervention, 28); Matsari, Nigeria (pre-intervention, 29); Idete, Tanzania (31); Navrongo, Ghana (34); Namawala, Tanzania (35)

Original reference detailing data and model fits: *Maire N, Smith TA, Ross A, Owusu-Agyei S, Dietz K, et al. A model for natural immunity to asexual blood stages of Plasmodium falciparum malaria in endemic areas. Am J Trop Med Hyg. Volume 75, No. 2 Supplement. 2006* (*31*)

Loss function: Log-normal log likelihood For objective 3 (age pattern of parasite densities) we denote the log-Normal log likelihood for each scenario to be

$$f_3(\theta) = \log L(\theta) = n(\log(\rho) - \log(\sigma)) - 0.5RSS/\sigma^2 \tag{A.44}$$

where n is the number of observations in the data set, $\rho = \exp(-0.5\log(2\pi))$, a constant from the log-normal likelihood, RSS is the residual sum of squares given by

$$RSS = \sum_{j=1}^{s}\sum_{k=1}^{a}\left(\frac{\widehat{Y_{j,k}}}{\widehat{P_{j,k}}} - \log(\nu) - \frac{Y_{j,k}}{P_{j,l}}\right)^2 \tag{A.45}$$

and σ is the standard deviation given by

$$\sigma = \sqrt{RSS/(n-1)} \tag{A.46}$$

Here, ν is the appropriate density bias, which is a fitting parameter, a is the number of age groups, s is the number of surveys, $P_{j,k}$ the scenario number of parasite positive hosts, and $Y_{j,k}$ the sum of the log densities, $\widehat{P_{j,k}}$ the predicted number of parasite positive hosts and $\widehat{Y_{j,k}}$ the predicted sum of the log densities for age group k and survey j . The density bias are fitting parameters ν_0 and ν_1 .

Age pattern of number of concurrent infections

Data For objective 4 (age pattern of number of concurrent infections), the dataset from Navrongo, Ghana (also used in the calibration of objectives 2 and 3) is used to calibrate to the total numbers of distinct parasite infections in one individual in each age group, and at each survey. Distinct infections were detected by polymerase chain reaction-restriction fragment length polymorphism in the sampled individuals.

Sites and scenario numbers: Navrongo, Ghana (34)

Original reference *detailing data and model fits*: *Maire N, Smith TA, Ross A, Owusu-Agyei S, Dietz K, et al. A model for natural immunity to asexual blood stages of Plasmodium falciparum malaria in endemic areas. Am J Trop Med Hyg. Volume 75, No. 2 Supplement. 2006 (31)*

Loss function: Poisson Log Likelihood Assuming that both the data and the simulations are Poisson distributed about the correct value and thereby also allowing for over-dispersion, we denote the Poisson log likelihood for each scenario to be for the objective of age pattern of number of concurrent infections to be

$$f_4(\theta) = logL(\theta) = \sum_{j=1}^{s}\sum_{k=1}^{a} -Pn_{j,k}\log\left(Pn_{j,k} \,/\, \lambda_{j,k}\right) + Pn_{j,k} - \lambda_{j,k} \tag{A.47}$$

where a is the number of age groups, s the number of surveys, $Pn_{j,k}$ the scenario data total patent infections for age group k and survey j . Parameter $\lambda_{j,k}$ is associated with the model predictions and is given by

$$\lambda_{j,k} = \frac{\widehat{Pn_{j,k}}}{\widehat{H_{j,k}}} H_{j,k} \tag{A.48}$$

where $\widehat{Pn_{j,k}}$ are the predicted total of patent infections and $\widehat{H_{j,k}}$ the predicted number of hosts for age group k and survey j and $H_{j,k}$ is the scenario data number of hosts for age group k and survey j .

Age pattern of incidence of clinical malaria

Data Two distinct datasets representing three study sites (Table A.2) were used for the calibration of objective 5 and objective 6 (age pattern of incidence of clinical malaria). For Objective 5, the dataset contains data on the age pattern of clinical episodes in the villages of Ndiop and Dielmo in Senegal (*256*, *257*). During the study period of July 1990 - June 1992, the village populations were visited daily to detect and treat any clinical malaria attacks with quinine. Cases were detected by reporting of symptoms (fever) during daily active case detection and subsequent thick blood smear microscopy. Only symptomatic individuals (axillary temperature $\geq$ 38.0° C or rectal temperature $\geq$ 38.5° C). Due to the active case detection and rapid treatment all symptomatic episodes are assumed to be effectively treated in these villages during the study period. No effective treatment of clinical malaria was assumed prior to the study period. The annual patterns of transmission were replicated as reported by Charlwood et al (1998) (*258*). A proportion P_t =35.75% are assumed to be treated effectively in Idete. As all individuals reporting to the village dispensary were treated presumptively with chloroquine, this proportion corresponds to the proportion of episodes reported to the village dispensary.

Sites and scenario numbers: Ndiop, Senegal (232), Dielmo, Senegal (233)

Original reference detailing data and model fits: *Smith TA, Ross A, Maire N, Rogier C, Trape J-F et al. An epidemiologic model of the incidence of acute illness in Plasmodium falciparum malaria. Am J Trop Med Hyg. Volume 75, No. 2 Supplement. 2006 (3)*

Loss function: RSS-biased We denote a loss function based on biased residual sum of squares:

$$f_5(\theta) = \sum_{j=s_1}^{s} \sum_{k=1}^{a} R^2 \tag{A.49}$$

where a is the number of age groups, s the number of surveys, s_1 the initial survey number, and R is the residual given by

$$R = I_{i,j} - \frac{\widehat{C_{j,k}}}{\left(\widehat{H_{j,k}}\right)} \frac{1}{\mu} \tag{A.50}$$

where $I_{j,k}$ is the observed recorded incidence rate, $\widehat{C_{j,k}}$ are the predicted total cases (severe and uncomplicated), $\widehat{H_{j,k}}$ the predicted number of hosts for age group k and survey j and μ is a bias related to the scenario. For scenarios 232 and 233 (representing Ndiop and Dielmo, Senegal) this bias is $\mu = 5$ indicating the duration in years for which episodes are collected. For scenario 49 in Objective 6 (Idete, Tanzania) the bias is $\mu = 0.357459$ and represents the proportion of episodes reported to the village dispensary.

TABLE A.2: Summary of study data set for objective 5: Age pattern of incidence of clinical malaria

Scenario No.	Study site	Age groups	Observations
232	Ndiop, Senegal	22	One per age group
233	Dielmo, Senegal	22	One per age group
49	Idete, Tanzania	4	One per age group

Age pattern of incidence of clinical malaria: infants

Data Objective 6 (age pattern of incidence of clinical malaria in infants) is informed by a dataset on incidence that contains passive case detection data on the age-incidence in infants recorded at the health centre in Idete, Tanzania from June 1993-October 1994 (*254*). The annual patterns of transmission were replicated as reported by Charlwood et al (1998) (*258*).

Sites and scenario numbers: Idete, Tanzania (49))

Original reference: *Smith TA, Ross A, Maire N, Rogier C, Trape J-F et al. An epidemiologic model of the incidence of acute illness in Plasmodium falciparum malaria. Am J Trop Med Hyg. Volume 75, No. 2 Supplement. 2006 (3)*

Loss function: RSS-biased The loss function for Objective 6 is the same as Objective 5. For scenario 49 (Idete, Tanzania) the bias is $\mu = 0.357459$ and represents the proportion of episodes reported to the village dispensary.

Age pattern of threshold parasite density for clinical attacks

Data Objective 7 (Age pattern of threshold parasite density for clinical attacks), uses the dataset from Dielmo, Senegal (see objective 5) for calibration. The pyrogenic threshold in the (*OpenMalaria*) predictions is output as the sum of the log threshold values across age groups. The pyrogenic threshold per age group is given as the parasite:leucocyte ratio for recorded incidence of disease. To adjust these densities to the same scale as that used in fitting the simulation model to other datasets, the parasite:leukocyte ratios were multiplied by a factor of 1,416 to give a notional density in parasites/microliter of blood. This number was derived as follows: Parasites were counted against leukocytes and converted to nominal parasites/microliter assuming the usual (though biased) standard of 8,000 leukocytes/microliter. The biases in density estimates resulting from these different techniques was accounted for by multiplying the observed parasite densities with constant values estimated for Garki (ν_0) and non-Garki (ν_1) studies to rescale them to the values in malariatherapy patients (*255*). The value 1416 comes from

$$8000\nu_1 \tag{A.51}$$

where the original $\nu_1 \approx 0.18$.

Sites and scenario numbers: Dielmo, Senegal (234)

Original reference detailing data and model fits: *Smith TA, Ross A, Maire N, Rogier C, Trape J-F et al. An epidemiologic model of the incidence of acute illness in Plasmodium falciparum malaria. Am J Trop Med Hyg. Volume 75, No. 2 Supplement. 2006 (3)*

Loss function: RSS-biased (log) For the objective 7 (Age pattern of threshold parasite density for clinical attacks) we denote a residual sum of squares loss function given by (13) with

$$f_7(\theta) = \log\left(Y^*_{j,k}\right) - \frac{\widehat{Y^*_{j,k}}}{\widehat{H_{j,k}}} - \log(\mu) \tag{A.52}$$

where Y^* is the observed pyrogenic threshold, $\widehat{Y^*}$ are the predicted sum log pyrogenic threshold, $\widehat{H_{j,k}}$ the predicted number of hosts for age group k and survey j and is a bias related to the scenario. Here, this bias is related to the log parasite/leucocyte ratio and thus $\mu = 1/(8000\nu_1)$ where ν_1 is the non-Garki density bias.

Hospitalisation rate in relation to prevalence in children

Data Data on the relative incidence of severe malaria-related morbidity and mortality in children <9 years old across different transmission intensities were originally collated by Marsh and Snow (1999) (*217*) (Table A.6). Data measurements per age group were available as the relative risk of severe disease compared to age group 1 and the proportion/prevalence of severe

episodes. A total of 26 entries on the relationship between severe malaria hospital admission rates and *P. falciparum* prevalence were used to calibrate objective 8 (Hospitalisation rate in relation to prevalence in children), each represented in a separate simulation scenario, with one observation per scenario. These are summarised in Table A.3. To obtain a continuous function relating hospital incidence rates to prevalence, linear interpolation between data points was performed. To convert hospital incidence rates to community severe malaria incidence, the hospital admission rates was divided by the assumed proportion of severe episodes representing to hospital (48%). There was assumed to be no effective treatment of uncomplicated malaria episodes or malaria mortality.

Sites and scenario numbers: Bo, Sierra Leone (501); Niakhar, Senegal (502), Farafenni, The Gambia (503); Areas I-V, The Gambia (504-508); Gihanga, Burundi (509); Katumba, Burundi (510); Karangasso, Burkina Faso (511); Kilifi North, Kenya (512); Manhica, Mozambique (514); Namawala, Tanzania (515); Navrongo, Ghana (516); Saradidi, Kenya (517); Yombo, Tanzania (518); Ziniare, Burkina Faso (519); Matsari, Nigeria (520); ITC control, Burkina Faso (521); Mlomp, Senegal (522); Ganvie, Benin (523); Kilifi Town, Kenya (524); Chonyi, Kenya (525); Bandafassi, Senegal (526); Kongodjan, Burkina Faso (527)

Original reference detailing data and model fits: *Ross A, Maire N, Molineaux L and Smith TA. An epidemiologic model of severe morbidity and mortality caused by Plasmodium falciparum. Am J Trop Med Hyg. Volume 75, No. 2 Supplement. 2006 (215)*

Loss function: squared deviation The loss function is denoted as the log of residual sum of squares

$$f_8(\theta) = \left[\log\left(\frac{a_s \widehat{R_{k=1}}}{R^*_{k=1}}\right)\right]^2 \tag{A.53}$$

where a_s is the access to treatment of severe cases (0.48, estimated in base model), $\widehat{R_{k=1}}$ is the scenario predicted rate of severe episodes per 1000 person year for age group $k = 1$ (0-9 years), and parameter $R^*_{k=1}$ is the interpolated observed rate of severe episodes per 1000 person year given by

$$R^*_{k=1} = \frac{\left(\widehat{P_{k=1}} - P_l\right)}{(P_u - P_l)}(R_u - R_l) + R_t \tag{A.54}$$

where $\widehat{P_{k=1}}$ is the predicted prevalence summed over all surveys, P_u and P_l are the observed prevalences above and below the predicted prevalence $\widehat{P_{k=1}}$, respectively and R_u and R_l are the corresponding severe episode rates to the observed prevalences.

The predicted prevalence is given by

$$\widehat{P_{k=1}} = \frac{\widehat{Pt_{k=1}}/24}{\widehat{H_{k=1}}/24} \tag{A.55}$$

where $\widehat{P_{k=1}}$ is the total number of parasite positive predicted and $\widehat{H_{k=1}}$ are the total number of hosts (division by 24 to give mean values). The predicted rate of episodes per 1000 person year is given by

$$\widehat{R_{k=1}} = \frac{1000\ \widehat{S_{k=1}}/2}{\widehat{H_{k=1}}/24} \tag{A.56}$$

where $\widehat{S_{k=1}}$ is the number of severe cases predicted and with division by 2 to convert to from 2 years to 1 year and the division by 24 to give mean number of hosts.

TABLE A.3: Settings used for calibrating the incidence of severe malaria. (Adapted from Table 1 from Ross et al. 2006 (*215*))

Site	**EIR data**	
	Year	**EIR**
Burkina Faso		
ITC Control	1994-1995	389
Karangasso	1985	244
Kongodjan	1984	133
Ziniare	1994-1995	70
Burundi		
Gihanga	1983	205
Katumba	1982	13.6
Kenya		
Chonyi	1992-1993	50
Kilifi North	1992-1003	10.5
Kilifi Town	1990-1991	2.8
Saradidi	1986-1987	239
Senegal		
Bandafassi	1995-1996	363
Mlomp	1995	30
Niakhar	1995	11.6
Tanzania		
Namawala	1990-1991	329
Yombo	1992	234
The Gambia		
Area I-V	1991	$^{+}$
Farafenni	1987	8.9
Others		
Bo, Sierra Leone	1990-1991	34.7
Ganvie, Benin	1993-1995	11
Manhica, Mozambique	2001-2002	38
Matsari, Nigeria	1971	68
Navrongo, Ghana	2001-2002	418

$*$ EIR = entomological inoculation rate, ITC = control group of randomised trial of insecticide-treated curtains. $^{+}$Five sites with annual EIR between 1 and 10

Age pattern of hospitalisation: severe malaria

Data For objective 9 (Age pattern of hospitalisation), a subset of the data collated by Marsh and Snow (1999) (*217*) (see objective 8) is used. Detailed age-specific severe hospital admission rates were available for 5 of the sites (Table A.4). The patterns of incidence by age were summarised by age in 1-4 and 5-9 year-old children and compared with 1-11 year old infants by calculating the relative risk. Of the five sites, four were selected for fitting objective 9 based on the predicted prevalence. Baku, The Gambia was excluded as the very low (2%) prevalence here could not be matched.

Sites and scenario number(s): Area V, The Gambia (158); Saradidi, Kenya (167); Ganvie, Benin (173); Bandafassi, Senegal (176)

Original reference detailing data and model fits: *Ross A, Maire N, Molineaux L and Smith TA. An epidemiologic model of severe morbidity and mortality caused by Plasmodium falciparum. Am J Trop Med Hyg. Volume 75, No. 2 Supplement. 2006 (215)*

TABLE A.4: Age-specific period prevalence rates∗ of severe malaria, severe malaria, severe malaria anaemia and acute respiratory-tract infections from five communities in The Gambia and Kenya. (Adapted from Table 2 from Snow et al 1997 (*259*))

Estimate	Sukuta, The Gambia	Kilifi North, Kenya	Kilifi South, Kenya	Siaya, Kenya
Years of paediatric ward surveillance	1992-95	1990-95	1992-96	1992,1994-96
Person-years exposure to risk of children aged 0-9 yr	23468	52675	45967	40064
Rates				
All-cause malaria, age 1-11 mo	23.3 (17.8–28.9) [66/2830]	59.5 (53.2-65.9) [318/5342]	79.9 (71.6-86.4) [407/5152]	84.6 (76.4-92.8) [374/4420]
All-cause malaria, age 1-4 yr	35.3 (32.2-39.4) [372/10379]	41.7 (39.0-44.4) [905/21714]	17.4 (15.5-19.3) [321/18493]	18.8 (16.7-20.9) [312/16567]
All-cause malaria, age 5-9 yr	16.3 (13.8-18.8) [167/10259]	5.3 (4.4-6.2) [135 / 25619]	1.7 (1.2-2.2) [38/22322]	1.7 (1.1-2.3) [33/19077]
All-cause malaria, age 0-9 yr	25.8 (23.8-27.8) [605]	25.9 (24.5-27.2) (1363) +	16.7 (15.5-17.9) [766]	18.0 (16.7-19.3) [719]
Cerebral malaria 0-9 yr	2.6 (1.8-3.3) [61]	1.5 (1.2-1.8) [79]	0.8 (0.5-1.1) [36]	0.1 (0.0-0.2) [5]
Severe malaria anaemia, 0-9 yr	NA	5.0 (4.4-5.6) [262]	4.2 (3.6-4.8) [192]	3.7 (2.7-4.7) [50/13416]
All-cause ARI age 0-9 yr	8.4 (7.3-9.6) [198]	9.3 (8.5-10.1) [492]	8.3 (7.5-9.1) [380]	8.7 (7.8-9.6) [348]

∗ Period prevalence rather than incidence because precise matching of each community member to hospital admission was not possible. Rates as admission per 1000 children per year (95% CI). +Precise dates of birth unobtainable for five children. Defined as child admitted with primary diagnosis of malaria and Blantyre coma score of 2 or less. Defined in child with primary diagnosis of malaria and haemoglobin of 5.0g/dL or less on admission. Rates for Siaya derived from person-years exposure to risk and admissions for period Nov 1, 1994 to Oct 31, 1995

Loss function: Residual sums of squares for relative risk We denote a loss function based on residual sum of squares:

$$f_9(\theta) = \sum_{k=2,3} \left[\log \frac{\widehat{RR_k}}{RR_k} \right]^2 \tag{A.57}$$

where RR_k is the relative risk of severe episode for age group k compared to age group 1 and $\widehat{RR_k}$ is the predictive relative risk for age group k compared to age group 1. The predicted relative risk is given by

$$\widehat{RR_k} = \frac{\widehat{S_k}}{\widehat{H_k}} - \frac{\widehat{S_1}}{\widehat{H_1}} \tag{A.58}$$

where $\widehat{S_k}$ is the number of severe cases predicted for age group k and $\widehat{H_k}$ the total number of hosts for age group k .

Malaria specific mortality in children (< 5 years old)

Data For objective 10 (Malaria specific mortality in children (<5 years old)), a subset of the data collated by Marsh and Snow (1999) (*217*) (see objective 8) was used (*260*). Mortality data were derived from verbal autopsy studies in sites with prospective demographic surveillance and were adjusted for the effect of malaria transmission intensity on the sensitivity and specificity of the cause of death determination. The odds ratio for death of a case in the community relative to that in hospital was estimated by fitting to the malaria-specific mortality rates in children less than five years of age assuming the published hospital case fatality rate. Nine sites for which both malaria-specific mortality rates and seasonal transmission patterns were available were included for calibration.

There is one observation per study site and simulation scenario, and predicted values are for one survey at the end of 2 years.

Sites and scenario number(s): Bo, Sierra Leone (301); Niakhar, Senegal (302); Farafenni, The Gambia (303); Kilifi North, Kenya (312); Navrongo, Ghana (316); Saradidi, Kenya (317); Yombo, Tanzania (318); Bandafassi, Senegal (326); Kongodjan, Burkina Faso (327)

Original reference detailing data and model fits: *Ross A, Maire N, Molineaux L and Smith TA. An epidemiologic model of severe morbidity and mortality caused by Plasmodium falciparum. Am J Trop Med Hyg. Volume 75, No. 2 Supplement. 2006* (*215*)

Loss function: Residual sums of squares For objective 10 on malaria-specific mortality in children, the loss function minimises the log sum of squares

$$f_{10}(\theta) = \left[\log\left(\frac{\widehat{DMR_1}}{DMR_1}\right)\right]^2 \tag{A.59}$$

where DMR_1 is the observed direct mortality rate for age group 1 (0-5 years) and $\widehat{DMR_1}$ is the predicted direct mortality rate for age group 1. The predicted direct mortality rate is given by

$$\widehat{DMR_1} = \frac{\widehat{DD_1}}{2\widehat{H_1}} \tag{A.60}$$

where $\widehat{DD_1}$ is the number of direct malaria deaths cases predicted for age group 1 and $\widehat{H_1}$ the total number of predicted hosts for age group 1. The division by 2 is to convert to yearly rate as the survey was conducted at the end of 2 years.

Indirect malaria infant mortality rate

Data For objective 11 (indirect malaria infant mortality rate), a subset of the data collated by Marsh and Snow (1999) (*217*) (see objective 8) was used. These constitute a library of sites for which entomologic data were collected at least monthly and all-cause infant mortality rates (IMR) were available. There is one observation per scenario: all cause infant mortality rate (returned as a single number over whole intervention period).

Sites and scenario number(s): Bo, Sierra Leone (401); Niakhar, Senegal (402); Area V, The Gambia (408); Karangasso, Burkina Faso (411); Manhica, Mozambique (414); Namawala, Tanzania (415); Navrongo, Ghana (416); Saradidi, Kanya (417); Yombo, Tanzania (418); Mlomp, Senegal (422); Bandafassi, Senegal (426)

Original reference detailing data and model fits: *Ross A, Maire N, Molineaux L and Smith TA. An epidemiologic model of severe morbidity and mortality caused by Plasmodium falciparum. Am J Trop Med Hyg. Volume 75, No. 2 Supplement. 2006 (215)*

Loss function: Residual sums of squares The loss function minimises the log sum of squares:

$$f_{11}(\theta) = \left[\log\left(\frac{i\widehat{DMR_1}}{iDMR_1}\right)\right]^2 \tag{A.61}$$

where $iDMR_1$ the observed indirect mortality rate for age group 1 and $i\widehat{DMR_1}$ is the predicted indirect mortality rate for age group 1.

A.2.2 Tables A.5-A.6

TABLE A.5: Epidemiological quantities and data sources used for parameterizing models

Epidemiological quantity	Data sources	No. of scenarios	No. of data points*	Publication for fitting of base model	Prior	Weighting in GOF statistic	Scenario numbers	Loss vector number (f_i)	Loss function
Age pattern of incidence of infection after intervention	Molineaux and Gramiccia (1980)(65)	1	12	Maire et al 2006 (*31*)	Binomial	0.001	30	1	Binomial log-likelihood
Age patterns of prevalence of infection	Molineaux and Gramiccia (1980) (65)	6	563	Maire et al 2006 (*31*)	Binomial	0.001	24, 28, 29, 35, 34, 31	2	Binomial log-likelihood
Age patterns of parasite density	Molineaux and Gramiccia (1980) (65)	6	563	Maire et al 2006 (*31*)	Log Normal	0.01	24, 28, 29, 35, 34, 31	3	log likelihood
Age pattern of number of concurrent infections	Maire et al. 2006 (*31*); Owusu-Agyei et al 2002 (*252*)	1	12	Maire et al 2006 (*31*)	Poisson	0.01	34	4	Poisson log-likelihood
Age pattern of incidence of clinical malaria: age-specific Ndiop & Dielmo, Senegal	Trape and Rogier 1996 (*256*); Kitua et al 1996 (*254*)	2	26	Smith et al 2006 (*183*)	Log Normal	1	232, 233, 49	5	RSS
Age pattern of incidence of clinical malaria: infants Idete, Tanzania	Kitua et al 1996 (*254*)	1	4	Smith et al 2006 (*183*)	Log Normal	1	49	6	RSS
Age pattern of threshold parasite density for clinical attacks	Rogier et al 1996 (*89*)	1	13	Smith et al 2006 (*69*)	Log Normal	1	234	7	RSS
Hospitalisation rate in relation to prevalence in children	See Ross et al 2006 (*215*)	26	10	Ross et al 2006 (*215*)	Log Normal	2	501, 502, 503, 504, 505, 506, 507, 508, 509, 510, 511, 512, 514, 515, 516, 517, 518, 519, 520, 521, 522, 523, 524, 525, 526, 527	8	Squared deviation
Age pattern of hospitalisation: severe malaria	Marsh and Snow 1999 (*217*)	4	12	Ross et al 2006 (*215*)	Log Normal	2	158, 167, 173, 176	9	RSS
Malaria specific mortality in children (<5y)	Snow et al 1997 (*259*)	9	9	Ross et al 2006 (*215*)	Log Normal	1	301, 302, 303, 312, 316, 317, 318, 326, 327	10	Squared deviation logRate
All-cause infant mortality rate	See Ross et al 2006 (*215*)	11	11	Ross et al 2006 (*215*)	Log Normal	10	401, 402, 408, 411, 414, 415, 416, 417, 418, 422, 426	11	Squared deviation logRate

(a) Some scenarios are used to predict several outcomes, so the total of this column does not equal the total of 61 scenarios involved in fitting the models. (b) The number of data points is the sum over all scenarios and simulated survey periods of the number of age groups into which the data were disaggregated for comparison with the model predictions. (c) In relation to the EIR specified as a seasonal pattern. (d) Model predictions for this objective are compared with linear interpolations between the field data points. * The number of data points is the sum over all scenarios and simulated survey periods of the number of age groups into which the data were disaggregated for comparison with the model predictions. Table adapted from Table S1 in Smith et al 2012 (*64*).

TABLE A.6: Simulation scenarios for calibration

Scen. No.	Site/reference	Description	Objective(s)	Data Reference
24	Sugungum, Nigeria (pre-intervention phase)	8 cross sectional surveys of entire village population at 10-week intervals (4,487 blood slides)	Age-prevalence (2); Age-parasite densities (3)	Molineaux and Gramiccia. 1980 (*65*)
28	Rafin-Marke, Nigeria (pre-intervention phase)	8 cross sectional surveys of entire village population at 10-week intervals (2,593 blood slides)	Age-prevalence (2); Age-parasite densities (3)	Molineaux and Gramiccia. 1980 (*65*)
29	Matsari, Nigeria (pre-intervention phase)	8 cross sectional surveys of entire village population at 10-week intervals (2,963 blood slides)	Age-prevalence (2); Age-parasite densities (3)	Molineaux and Gramiccia. 1980 (*65*)
30	Matsari, Nigeria (intervention phase)	8 cross sectional surveys of entire village population at 10-week intervals (2,663 blood slides)	Age-incidence of patent infections (1)	Molineaux and Gramiccia. 1980 (*65*)
31	Idete, Tanzania	Surveillance of a rolling cohort of infants (1,382 blood slides over 16 months). Also 1 cross-sectional survey of 312 children 1-5 months	Age-prevalence (2); Age-parasite densities (3)	Kitua et al 1996 (*254*)
34	Navrongo, Ghana	6 age-stratified cross-sectional surveys at 2-month intervals (total 522 slides / DNA samples)	Age-prevalence (2); Age-parasite densities (3),Age-specific multiplicity of infection (4)	Owusu-Agyei S et al. 2002 (*252*)
35	Namawala, Tanzania	12 age-stratified cross-sectional surveys at 2-month intervals (3,901 blood slides)	Age-prevalence (2); Age-parasite densities (3)	Smith et al 1993 (*253*)
49	Idete, Tanzania	Passive case detection at the village dispensary over 15 months in 12 age groups.	Age Pattern of Incidence of Clinical Malaria in Idete in infants (5b)	Kitua et al. 1996 (*254*); Vounatsou et al. 2000 (*261*)
158	Area V, The Gambia	Hospitalisation rate by age	Age pattern of severe hospitalisation (8)	Snow et al. 1997 (*259*)
167	Saradidi, Kenya	21 cohorts each of approximately 50 children between 6 months and 6 years of age whose parasites were cleared and who were then followed up with 2 weekly surveys. Hospitalisation rate by age.	Age pattern of severe hospitalisation (8)	Beier et al. 1999 (*262*), Snow 1997 (*259*)
173	Ganvie, Benin	Hospitalisation rate by age.	Age pattern of severe hospitalisation (8)	Snow et al. 1997 (*87*)
176	Bandafassi, Senegal	Hospitalisation rate by age.	Age pattern of severe hospitalisation (8)	Snow et al. 1997 (*259*)
232	Ndiop, Senegal	Longitudinal study of 350 permanent residents over 2 years: Individual level active case detection three times a week (questionnaire + recording of symptoms) and parasitologic surveys twice a week; daily recording of new fever cases at compound level. By age group (9 groups)	Age pattern of incidence of clinical malaria (5a)	Trape JF and Rogier C. 1996 (*256*)
233	Dielmo, Senegal	Longitudinal study of 206 permanent residents over 2 years: Individual level active case detection three times a week (questionnaire + recording of symptoms) and parasitologic surveys twice a week; daily recording of new fever cases at compound level. By age group (9 groups)	Age pattern of incidence of clinical malaria by age (5a)	Trape JF and Rogier C. 1996 (*256*)
234	Dielmo, Senegal	Longitudinal study of 206 permanent residents over 2 years: Individual level active case detection three times a week (questionnaire + recording of symptoms) and parasitologic surveys twice a week; daily recording of new fever cases at compound level. By age group (9 groups)	Age Pattern of parasite density threshold for clinical attack (6)	Trape JF and Rogier C. 1996 (*256*)
301	Bo, Sierra Leone	Point estimate based on a 1-year longitudinal study covering 776 person-years	Direct Malaria Mortality (9)	Korenromp et al. 2003 (*204*)
302	Niakhar, Senegal	Point estimate based on 5-year longitudinal study covering 29,491 person-years [XML label: Diohine]	Direct Malaria Mortality (9)	Korenromp et al. 2003 (*204*)

continued on next page

continued from previous page

303	Farafenni, The Gambia	Point estimate based on 2-year longitudinal study covering 2,263 person-years [XML label: Tally Ya]	Direct Malaria Mortality (9)	Korenromp et al. 2003 (*204*)
312	Kilifi North, Kenya	Point estimate based on 3-year longitudinal study covering 20,679 person-years	Direct Malaria Mortality (9)	Korenromp et al. 2003 (*204*)
316	Navrongo, Ghana	Point estimate based on 1-year longitudinal study covering 1,065 person-years	Direct Malaria Mortality (9)	Korenromp et al. 2003 (*204*)
317	Saradidi, Kenya	21 cohorts each of approximately 50 children between 6 months and 6 years of age whose parasites were cleared and who were then followed up with 2 weekly surveys.	Direct Malaria Mortality (9)	Korenromp et al. 2003 (*204*)
318	Yombo, Tanzania	Point estimate based on 3-year longitudinal study covering 5,850 person-years	Direct Malaria Mortality (9)	Korenromp et al. 2003 (*204*)
326	Bandafassi, Senegal	Point estimate based on 6-year longitudinal study covering 8,488 person-years	Direct Malaria Mortality (9)	Korenromp et al. 2003 (*204*)
327	Kongodjan, Burkina Faso	Point estimate based on 5-year longitudinal study covering 1,271 person-years	Direct Malaria Mortality (9)	Korenromp et al. 2003 (*204*)
401	Bo, Sierra Leone	Point estimates of all-cause neonatal, post-neonatal, and infant mortality rates	All-cause mortality (10)	Barnish et al. 1993 (*263*)
402	Niakhar, Senegal	Point estimates of all-cause neonatal, post-neonatal, and infant mortality rates; XML label: Diohine	All-cause mortality (10)	INDEPTH Network, 2002 (*264*); Spencer et al. 1987 (*265*)
408	Area V, The Gambia	Point estimates of all-cause neonatal, post-neonatal, and infant mortality rates	All-cause mortality (10)	D'Alessandro et al. 1995 (*266*)
411	Karangasso, Burkina Faso	Point estimates of all-cause neonatal, post-neonatal, and infant mortality rates	All-cause mortality (10)	Duboz et al. 1989 (*267*)
414	Manhica, Mozambique	Point estimates of all-cause neonatal, post-neonatal, and infant mortality rates	All-cause mortality (10)	INDEPTH Network, 2002 (*264*)
415	Namawala, Tanzania	Point estimates of all-cause neonatal, post-neonatal, and infant mortality rates; Pre-intervention	All-cause mortality (10)	Armstrong-Schellenberg et al. 1999 (*268*)
416	Navrongo, Ghana	Point estimates of all-cause neonatal, post-neonatal, and infant mortality rates	All-cause mortality (10)	INDEPTH Network, 2002 (*264*)
417	Saradidi, Kanya	21 cohorts each of approximately 50 children between 6 months and 6 years of age whose parasites were cleared and who were then followed up with 2 weekly surveys.	All-cause mortality (10)	Spencer et al. 1987 (*265*)
418	Yombo, Tanzania	Point estimates of all-cause neonatal, post-neonatal, and infant mortality rates	All-cause mortality (10)	Premji Z et al. 1997 (*269*)
422	Mlomp, Senegal	Point estimates of all-cause neonatal, post-neonatal, and infant mortality rates	All-cause mortality (10)	Trape et al. 1998 (*270*)
426	Bandafassi, Senegal	Point estimates of all-cause neonatal, post-neonatal, and infant mortality rates	All-cause mortality (10)	INDEPTH Network, 2002 (*264*)
501	Bo, Sierra Leone	Point estimate of the severe malaria hospital admission rate and P.falciparum prevalence in children <9 years old.	Severe episodes by prevalence (7)	Marsh and Snow 1999 (*217*)
502	Niakhar, Senegal	Point estimate of the severe malaria hospital admission rate and P.falciparum prevalence in children <9 years old.XML label: Diohine (ca 20 km from Niakhar)	Severe episodes by prevalence (7)	Marsh and Snow 1999 (*217*)
503	Farafenni, The Gambia	Point estimate of the severe malaria hospital admission rate and P.falciparum prevalence in children <9 years old.XML label: Tally Ya (ca 15 km from Farafenni)	Severe episodes by prevalence (7)	Marsh and Snow 1999 (*217*)
504	Area I, The Gambia	Point estimate of the severe malaria hospital admission rate and P.falciparum prevalence in children <9 years old.	Severe episodes by prevalence (7)	Marsh and Snow 1999 (*217*)
505	Area II, The Gambia	Point estimate of the severe malaria hospital admission rate and P.falciparum prevalence in children <9 years old.	Severe episodes by prevalence (7)	Marsh and Snow 1999 (*217*)
506	Area III, The Gambia	Point estimate of the severe malaria hospital admission rate and P.falciparum prevalence in children <9 years old.	Severe episodes by prevalence (7)	Marsh and Snow 1999 (*217*)

continued on next page

		continued from previous page		
507	Area IV, The Gambia	Point estimate of the severe malaria hospital admission rate and P.falciparum prevalence in children <9 years old.	Severe episodes by prevalence (7)	Marsh and Snow 1999 (217)
508	Area V, The Gambia	Point estimate of the severe malaria hospital admission rate and P.falciparum prevalence in children <9 years old.	Severe episodes by prevalence (7)	Marsh and Snow 1999 (217)
509	Gihanga, Burundi	Point estimate of the severe malaria hospital admission rate and P.falciparum prevalence in children <9 years old.	Severe episodes by prevalence (7)	Marsh and Snow 1999 (217)
510	Katumba, Burundi	Point estimate of the severe malaria hospital admission rate and P.falciparum prevalence in children <9 years old.	Severe episodes by prevalence (7)	Marsh and Snow 1999 (217)
511	Karangasso, Burkina Faso	Point estimate of the severe malaria hospital admission rate and P.falciparum prevalence in children <9 years old.	Severe episodes by prevalence (7)	Marsh and Snow 1999 (217)
512	Kilifi North, Kenya	Point estimate of the severe malaria hospital admission rate and P.falciparum prevalence in children <9 years old.	Severe episodes by prevalence (7)	Marsh and Snow 1999 (217)
514	Manhica, Mozambique	Point estimate of the severe malaria hospital admission rate and P.falciparum prevalence in children <9 years old.	Severe episodes by prevalence (7)	Marsh and Snow 1999 (217)
515	Namawala, Tanzania	Point estimate of the severe malaria hospital admission rate and P.falciparum prevalence in children <9 years old. Pre-intervention	Severe episodes by prevalence (7)	Marsh and Snow 1999 (217)
516	Navrongo, Ghana	Point estimate of the severe malaria hospital admission rate and P.falciparum prevalence in children <9 years old.	Severe episodes by prevalence (7)	Marsh and Snow 1999 (217)
517	Saraddi, Kenya	Point estimate of the severe malaria hospital admission rate and P.falciparum prevalence in children <9 years old.	Severe episodes by prevalence (7)	Marsh and Snow 1999 (217)
518	Yombo, Tanzania	Point estimate of the severe malaria hospital admission rate and P.falciparum prevalence in children <9 years old.	Severe episodes by prevalence (7)	Marsh and Snow 1999 (217)
519	Ziniare, Burkina Faso	Point estimate of the severe malaria hospital admission rate and P.falciparum prevalence in children <9 years old.	Severe episodes by prevalence (7)	Marsh and Snow 1999 (217)
520	Matsari, Nigeria	Point estimate of the severe malaria hospital admission rate and P.falciparum prevalence in children <9 years old. Pre-intervention	Severe episodes by prevalence (7)	Marsh and Snow 1999 (217)
521	ITC control, Burkina Faso	Point estimate of the severe malaria hospital admission rate and P.falciparum prevalence in children <9 years old.	Severe episodes by prevalence (7)	Marsh and Snow 1999 (217)
522	Mlomp, Senegal	Point estimate of the severe malaria hospital admission rate and P.falciparum prevalence in children <9 years old.	Severe episodes by prevalence (7)	Marsh and Snow 1999 (217)
523	Ganvie, Benin	Point estimate of the severe malaria hospital admission rate and P.falciparum prevalence in children <9 years old.	Severe episodes by prevalence (7)	Marsh and Snow 1999 (217)
524	Kilifi Town, Kenya	Point estimate of the severe malaria hospital admission rate and P.falciparum prevalence in children <9 years old.	Severe episodes by prevalence (7)	Marsh and Snow 1999 (217)
525	Chonyi, Kenya	Point estimate of the severe malaria hospital admission rate and P.falciparum prevalence in children <9 years old.	Severe episodes by prevalence (7)	Marsh and Snow 1999 (217)
526	Bandafassi, Senegal	Point estimate of the severe malaria hospital admission rate and P.falciparum prevalence in children <9 years old.	Severe episodes by prevalence (7)	Marsh and Snow 1999 (217)
527	Kongodjan, Burkina Faso	Point estimate of the severe malaria hospital admission rate and P.falciparum prevalence in children <9 years old.	Severe episodes by prevalence (7)	Marsh and Snow 1999 (217)

Calibration data for objectives 2-4, age patterns of prevalence, parasite densities, and multiplicity of infection

A.3 Emulator performance

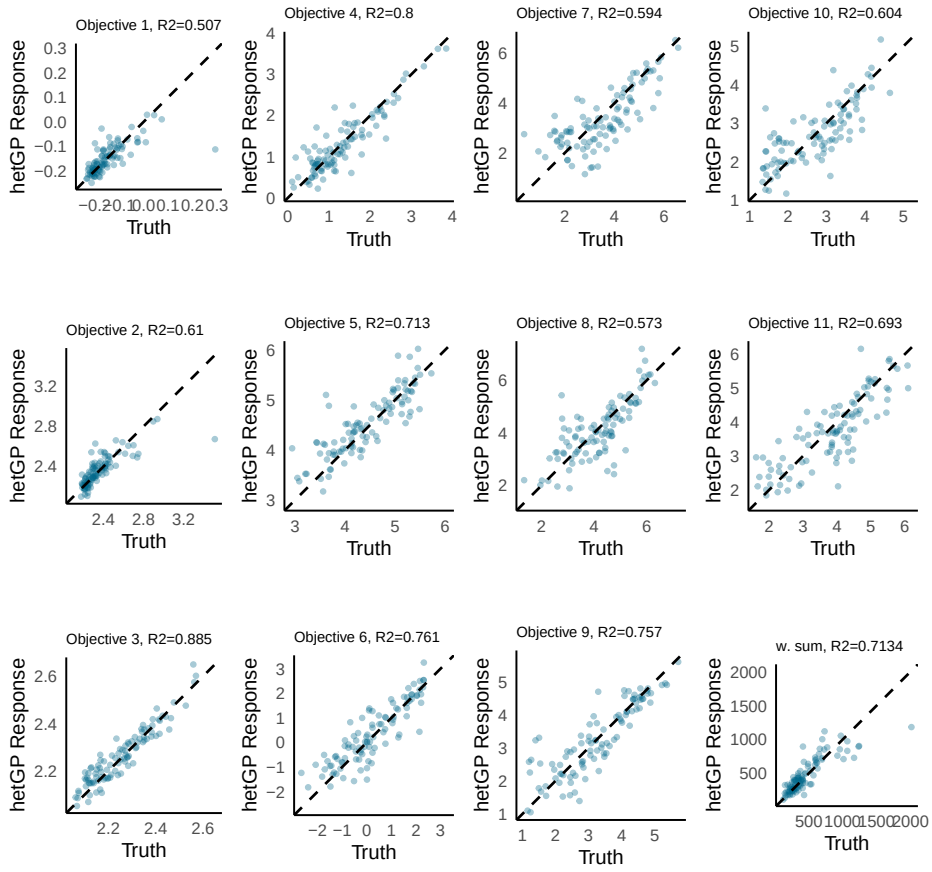

FIGURE A.2: GP emulator performance

Emulator predictions vs true values on a holdout set compromising 10% of initial samples in iteration 1. w.sum is the weighted sum F , of the 11 objectives. Here, predictions are generated as the weighted sum of individual objective predictions.

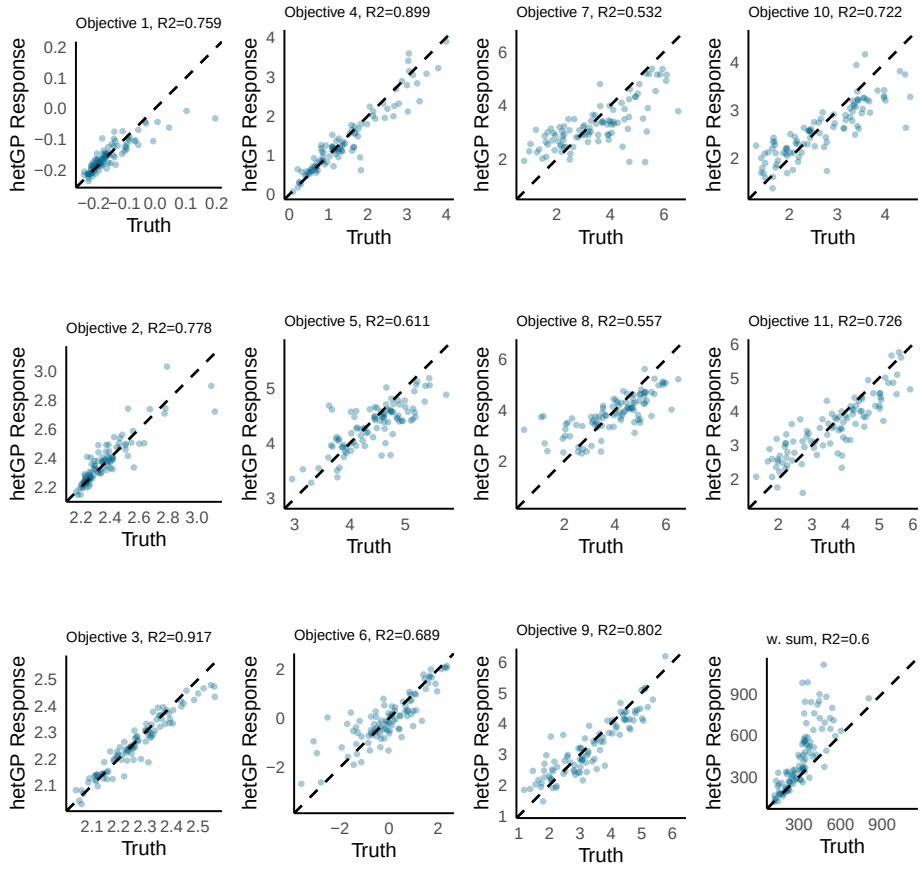

FIGURE A.3: GP emulator performance

Emulator predictions vs true values on a holdout set compromising 10% of initial samples in iteration 30 (final iteration). w.sum is the weighted sum F , of the 11 objectives. Here, predictions are generated as the weighted sum of individual objective predictions.

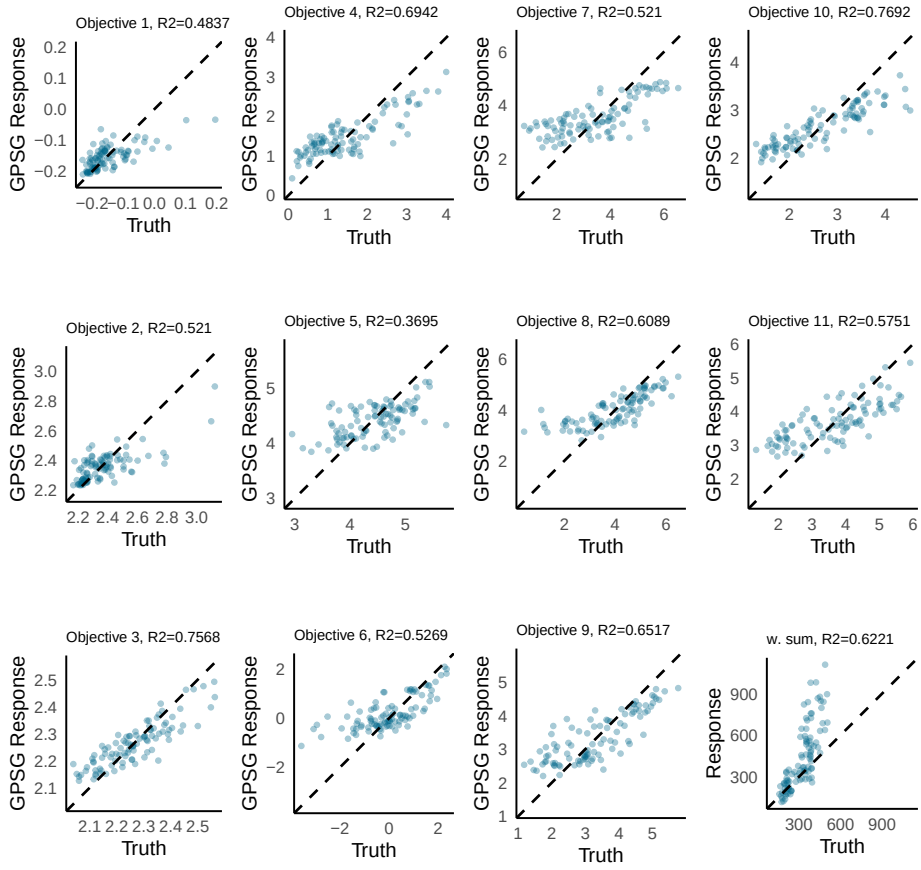

FIGURE A.4: GPSG emulator performance

Emulator predictions vs true values on a holdout set compromising 10% of initial samples in iteration 1. w.sum is the weighted sum F , of the 11 objectives. Here, predictions are generated as the weighted sum of individual objective predictions.

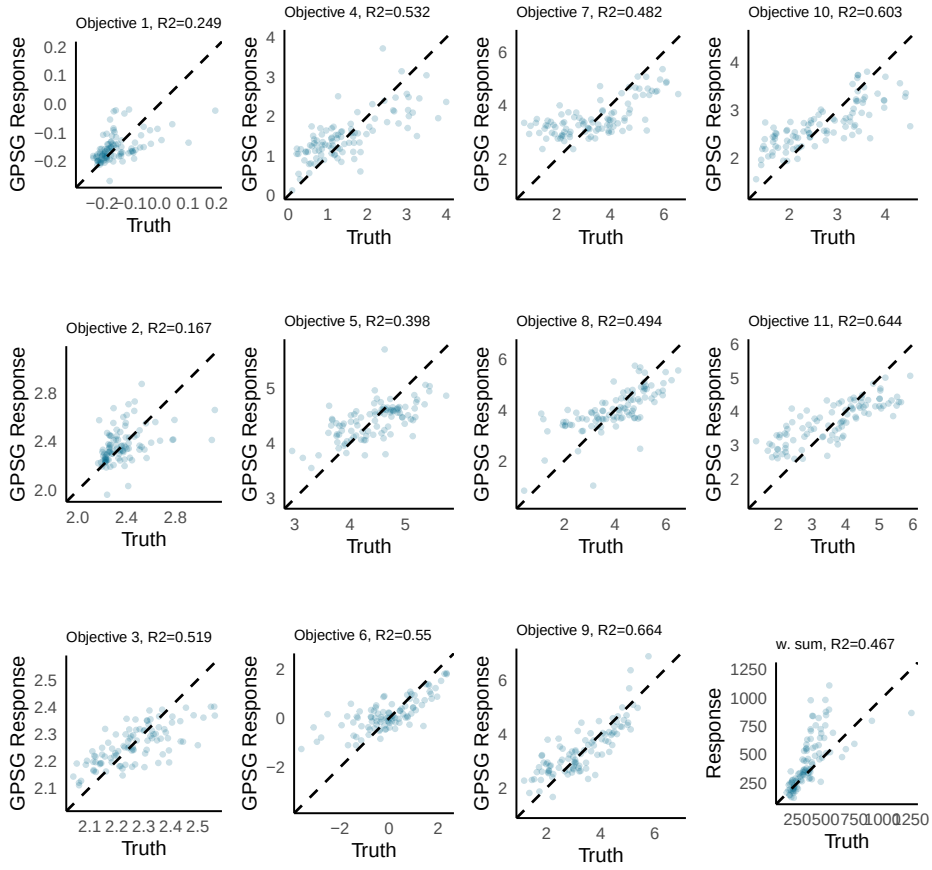

FIGURE A.5: GPSG emulator performance

Emulator predictions vs true values on a holdout set compromising 10% of initial samples in iteration 23 (final iteration). w.sum is the weighted sum F , of the 11 objectives. Here, predictions are generated as the weighted sum of individual objective predictions.

A.4 Adaptive sampling: selected points

A.4.1 GP-BO

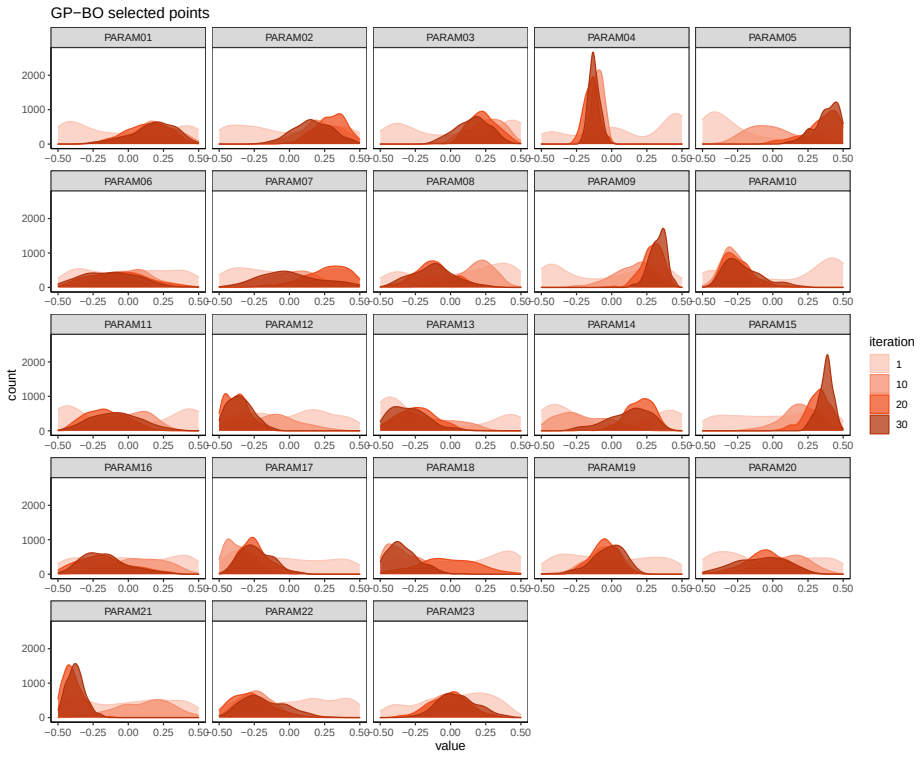

FIGURE A.6: GP-BO sampling behaviour

Values in each dimension of the points sampled during adaptive sampling of GP-BO algorithm in iterations 1,10, 20, and 30.

A.4.2 GPSG-BO

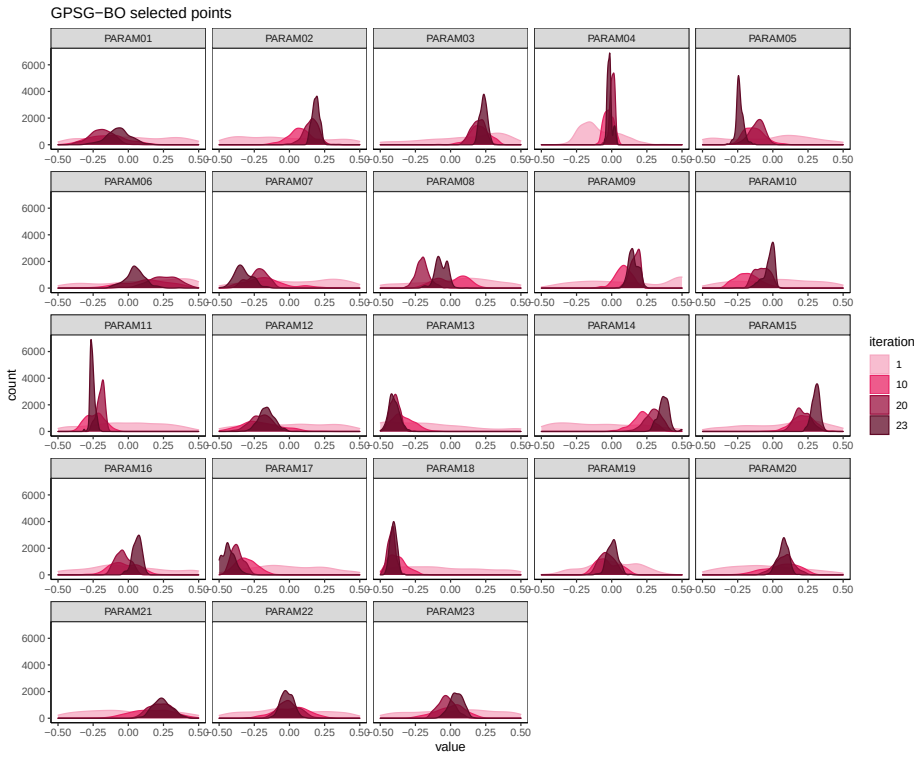

FIGURE A.7: GPSG-BO sampling behaviour

Values in each dimension of the points sampled during adaptive sampling of GPSG-BO algorithm in iterations 1,10, 20, and 23.

A.5 *OpenMalaria*: Final simulator fit

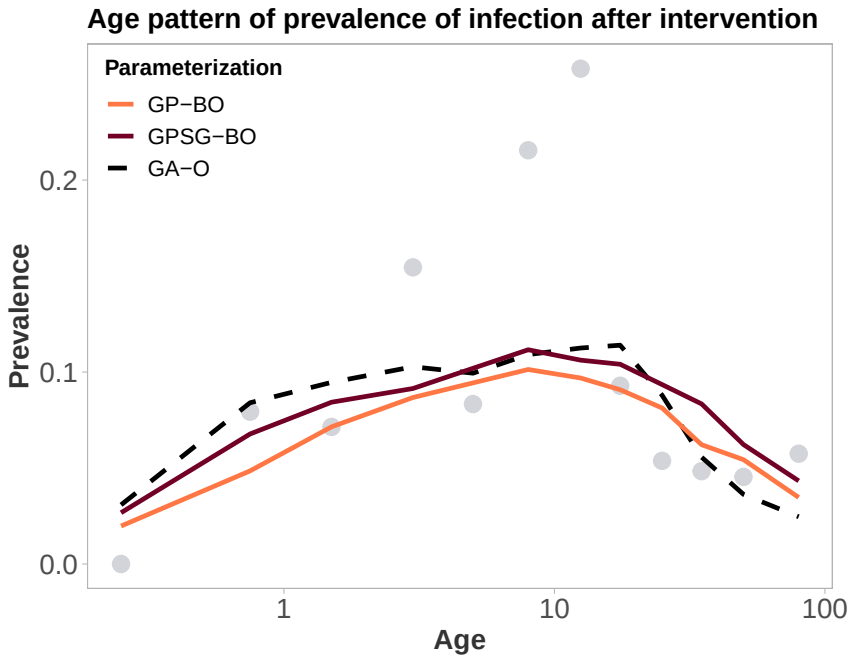

FIGURE A.8: Objective 1: Age pattern of prevalence in Matsari, Nigeria during the intervention

Final simulator fit using the parameter sets yielded using GP-BO and GPSG-BO compared to the previous parameterisation (derived using optimisation with a genetic algorithm, GA-O).

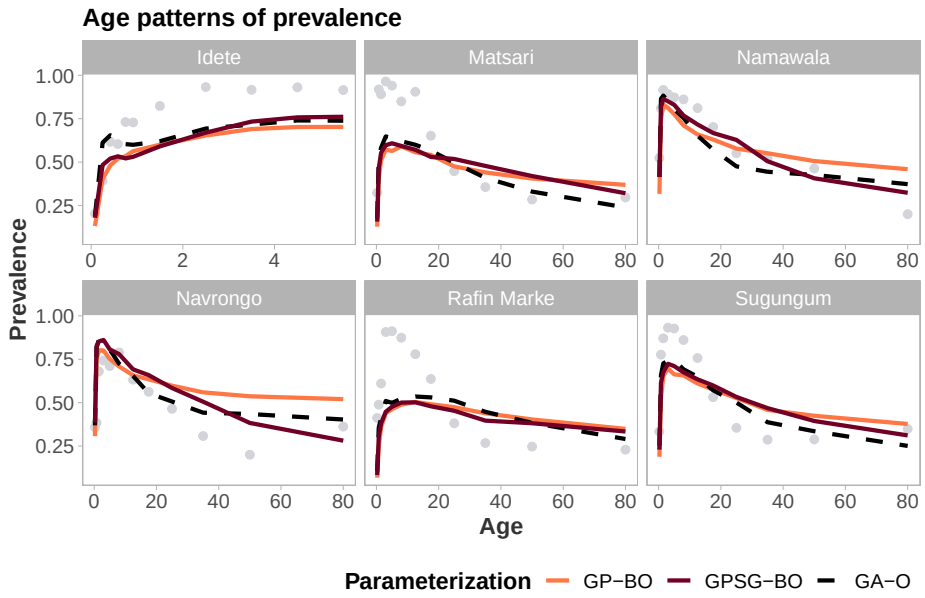

FIGURE A.9: Objective 2: Age pattern of prevalence

Final simulator fit using the parameter sets yielded using GP-BO and GPSG-BO compared to the previous parameterisation (derived using optimisation with a genetic algorithm, GA-O).

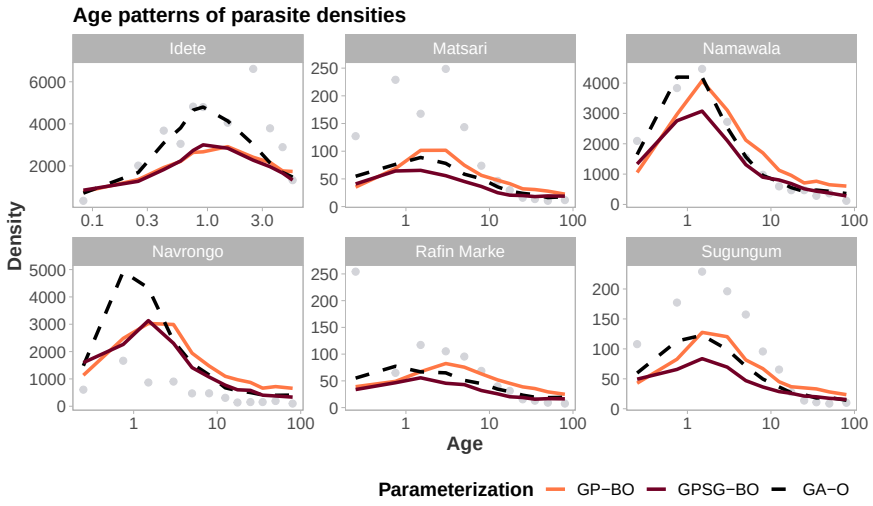

FIGURE A.10: Objective 3: Age pattern of parasite densities (geometric mean)

Final simulator fit using the parameter sets yielded using GP-BO and GPSG-BO compared to the previous parameterisation (derived using optimisation with a genetic algorithm, GA-O).

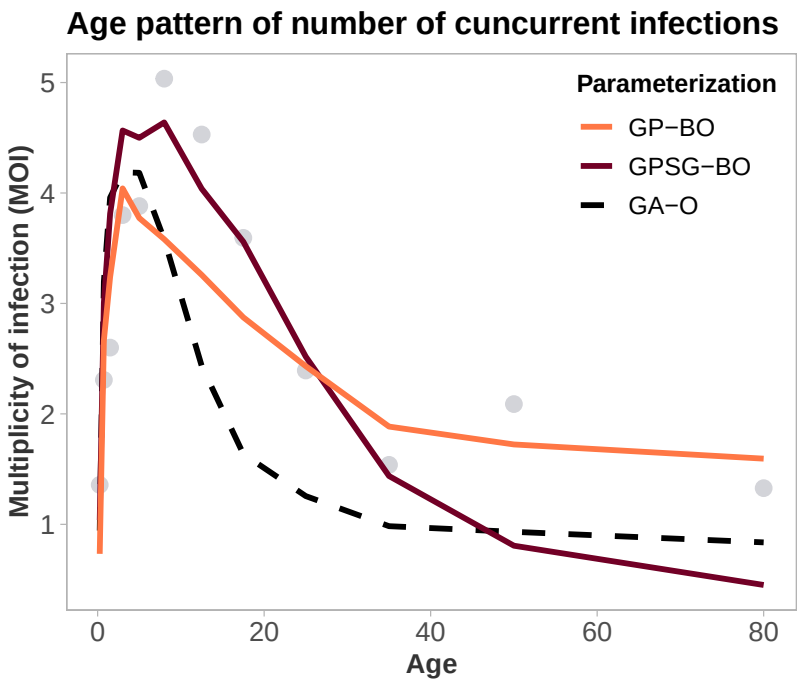

FIGURE A.11: Objective 4: Age pattern of number of concurrent infections

Final simulator fit using the parameter sets yielded using GP-BO and GPSG-BO compared to the previous parameterisation (derived using optimisation with a genetic algorithm, GA-O).

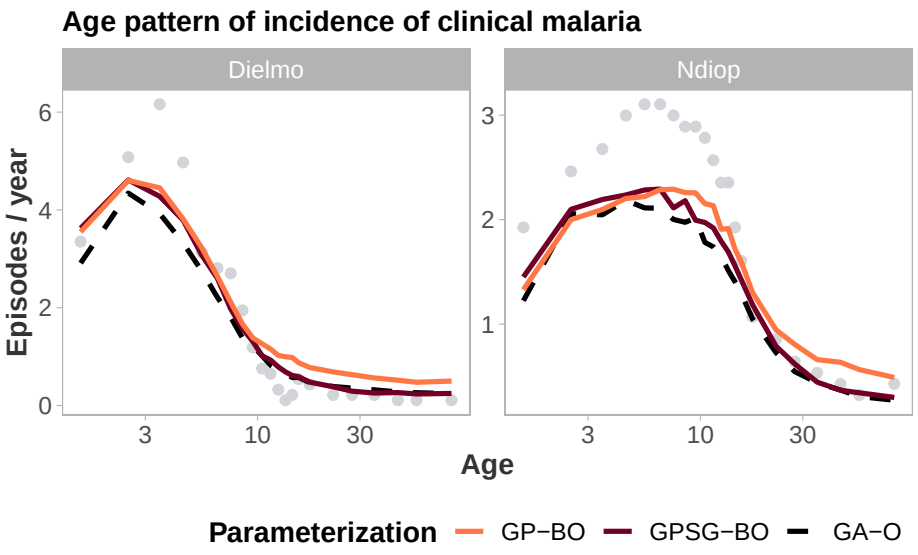

FIGURE A.12: Objective 5: Age pattern of incidence of clinical malaria in Dielmo and Ndiop, Senegal

Final simulator fit using the parameter sets yielded using GP-BO and GPSG-BO compared to the previous parameterisation (derived using optimisation with a genetic algorithm, GA-O).

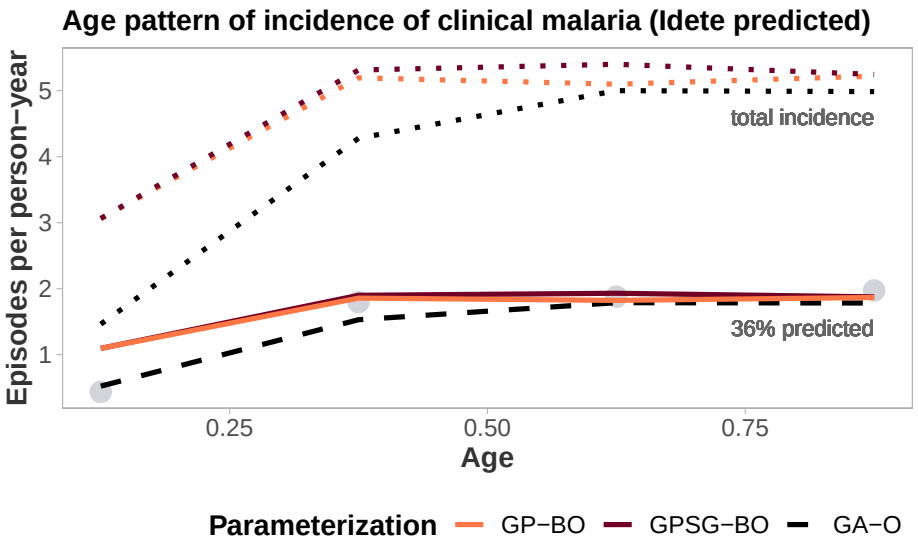

FIGURE A.13: Objective 6: Age pattern of incidence of clinical malaria in Idete, Tanzania

Final simulator fit using the parameter sets yielded using GP-BO and GPSG-BO compared to the previous parameterisation (derived using optimisation with a genetic algorithm, GA-O).

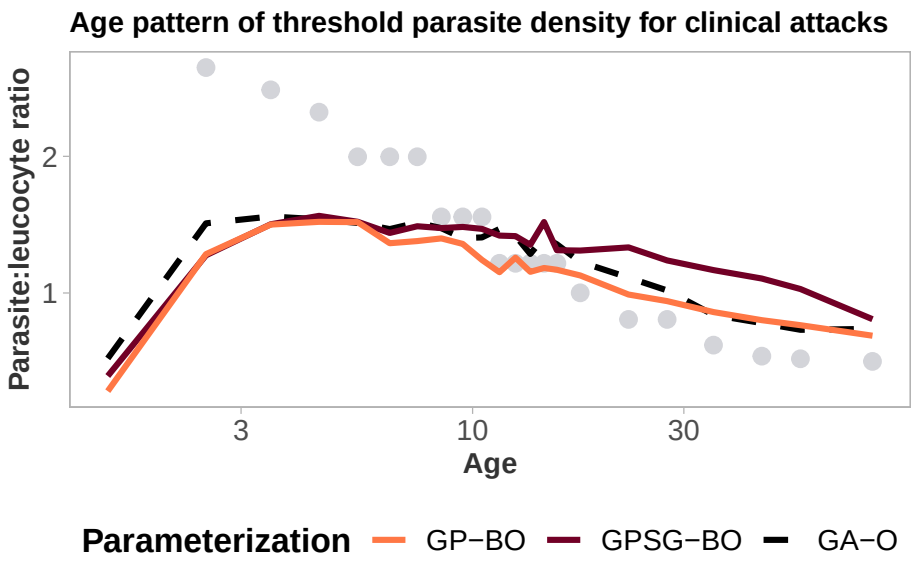

FIGURE A.14: Objective 7: Age pattern of threshold parasite density for clinical attacks

Final simulator fit using the parameter sets yielded using GP-BO and GPSG-BO compared to the previous parameterisation (derived using optimisation with a genetic algorithm, GA-O).

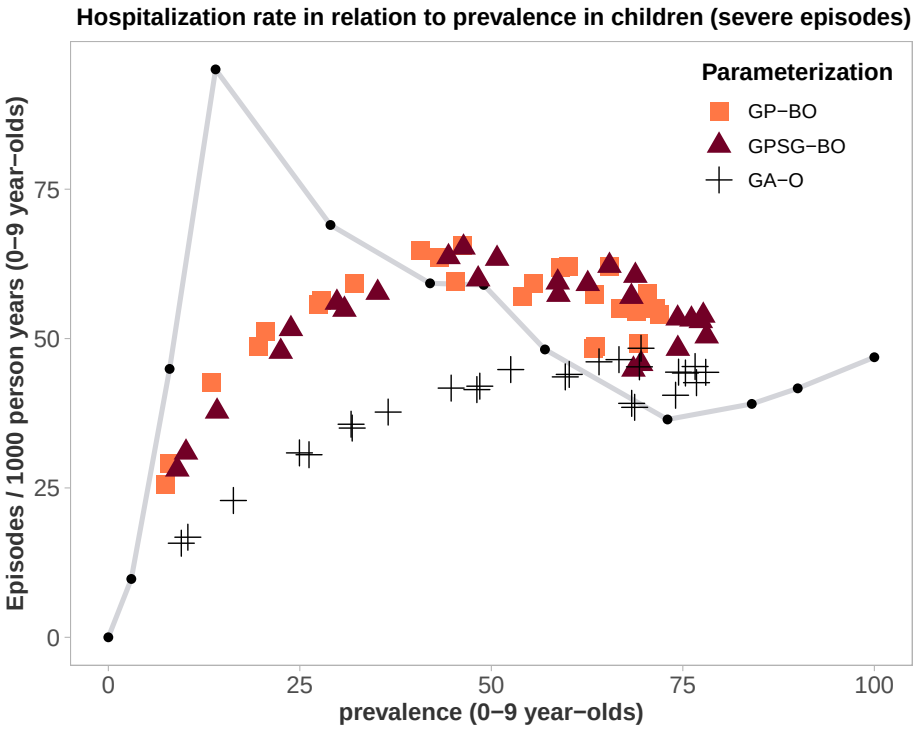

FIGURE A.15: Objective 8: Hospitalisation rate in relation to prevalence in children

Final simulator fit using the parameter sets yielded using GP-BO and GPSG-BO compared to the previous parameterisation (derived using optimisation with a genetic algorithm, GA-O).

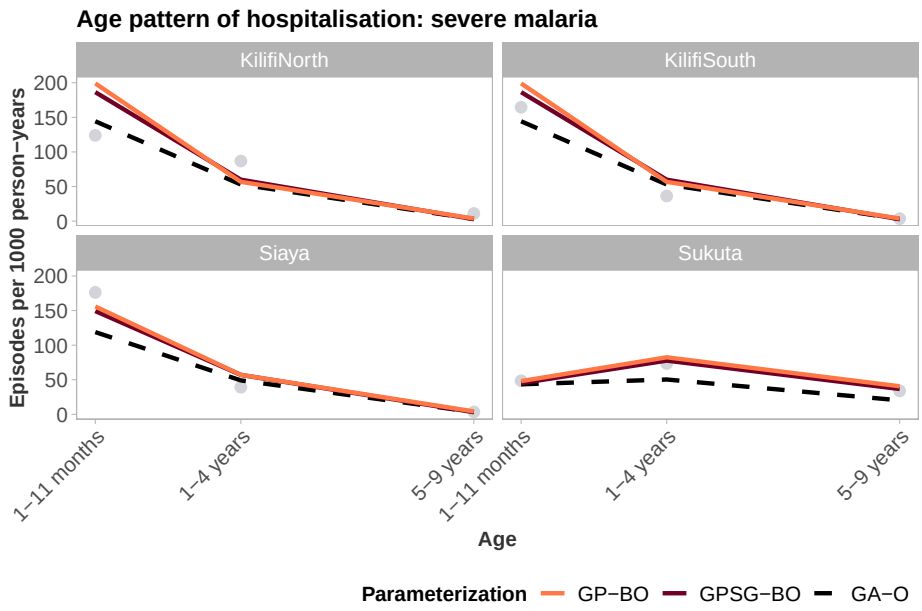

FIGURE A.16: Objective 9. Age pattern of hospitalisation

Final simulator fit using the parameter sets yielded using GP-BO and GPSG-BO compared to the previous parameterisation (derived using optimisation with a genetic algorithm, GA-O).

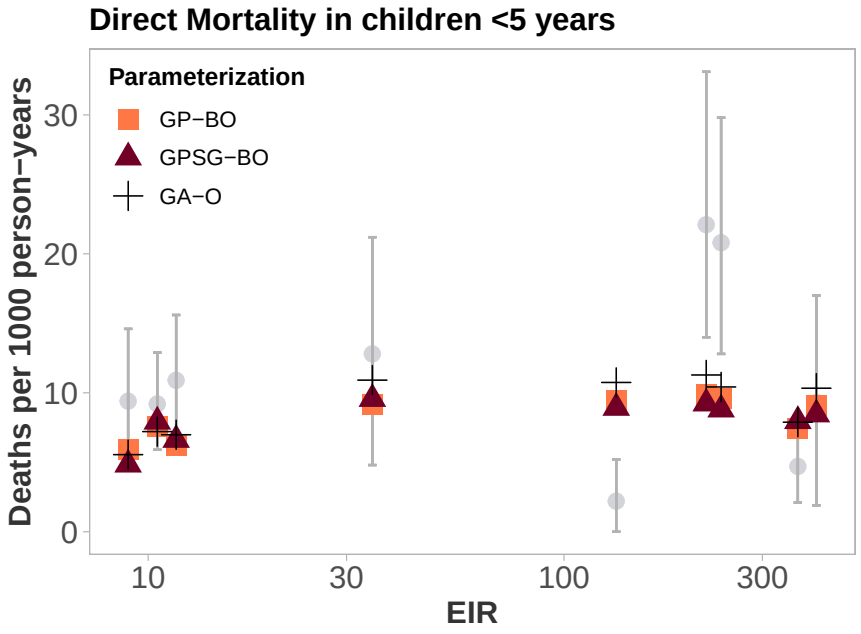

FIGURE A.17: Objective 10: Direct mortality in children <5 years old

Final simulator fit using the parameter sets yielded using GP-BO and GPSG-BO compared to the previous parameterisation (derived using optimisation with a genetic algorithm, GA-O).

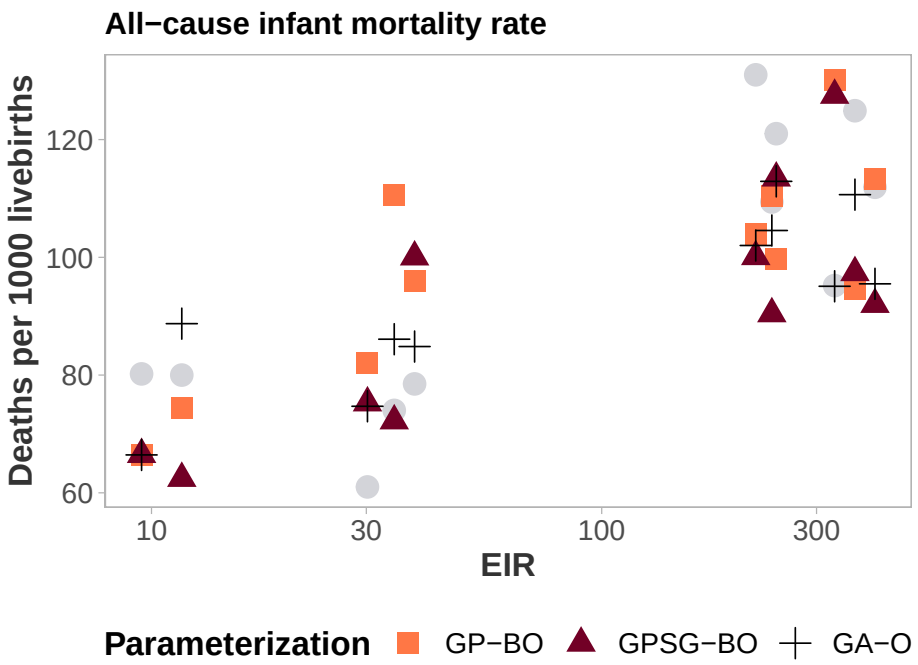

FIGURE A.18: Objective 11: All-cause infant mortality rate

Final simulator fit using the parameter sets yielded using GP-BO and GPSG-BO compared to the previous parameterisation (derived using optimisation with a genetic algorithm, GA-O).

A.6 Validation

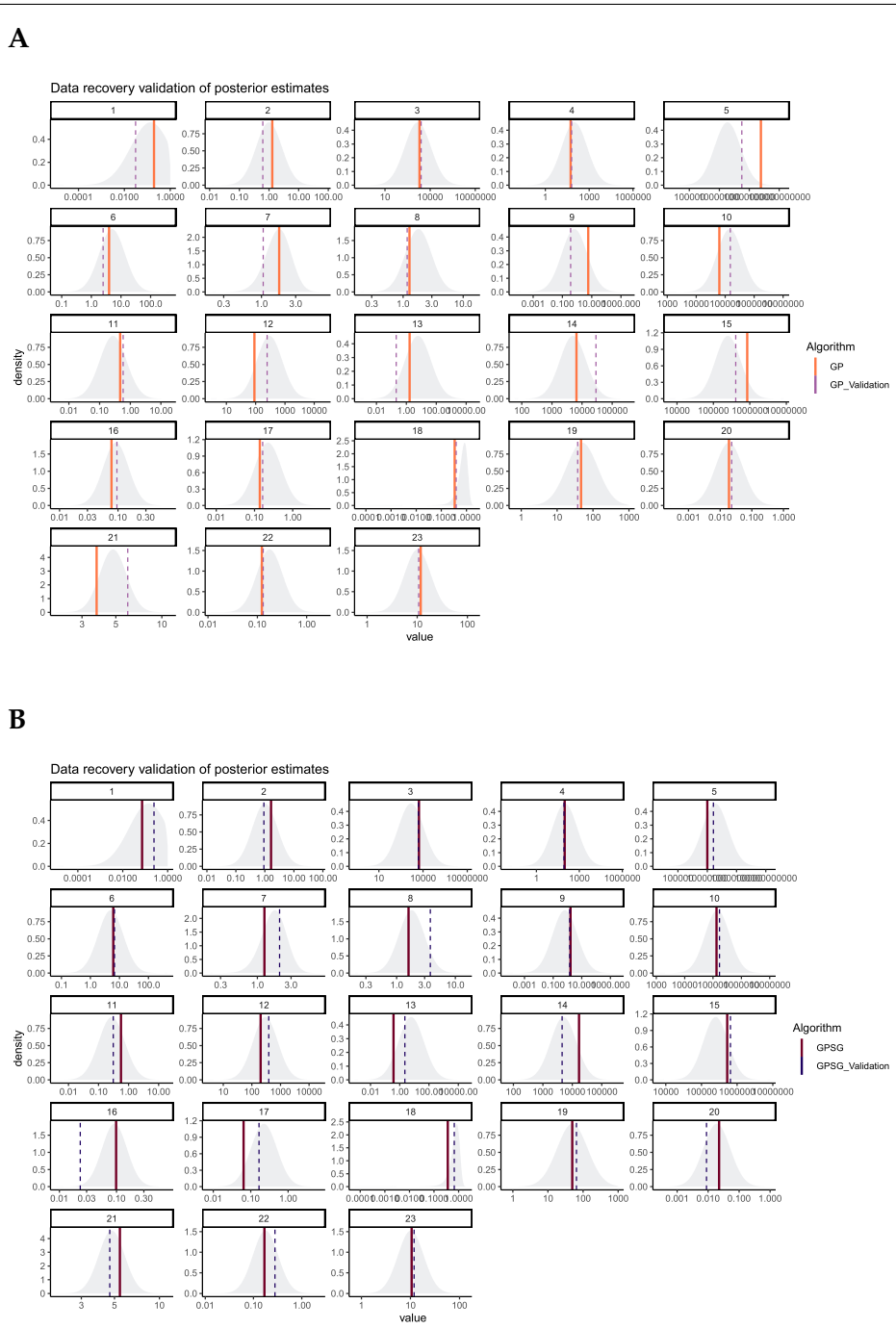

FIGURE A.19: Data recovery validation of posterior estimates

Prior distributions of each parameter and parameter value identified by the optimisation algorithm. The final parameter set was used to generate synthetic field data by simulating each of the 61 scenarios with the respective core parameter sets. The simulation outputs were reformatted to match the original field data, generating a synthetic field data set. The optimisation with both algorithms was repeated using this synthetic field data. The plot shows the best parameter values in each dimension identified at the end of the validation optimisation compared to the values identified in the original optimisation. The grey area shows the prior distribution. **A. GP-BO validation. B. GPSG-BO validation**

A.7 Epidemiological predictions

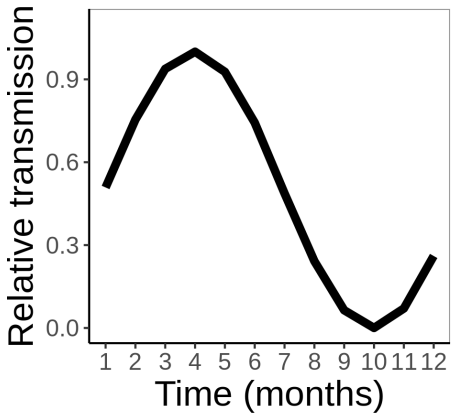

FIGURE A.20: Seasonal pattern assumed for subsequent analyses
The monthly transmission intensity is equivalent to the annual transmission intensity (EIR) scaled by these values and forced to sum to the annual EIR.

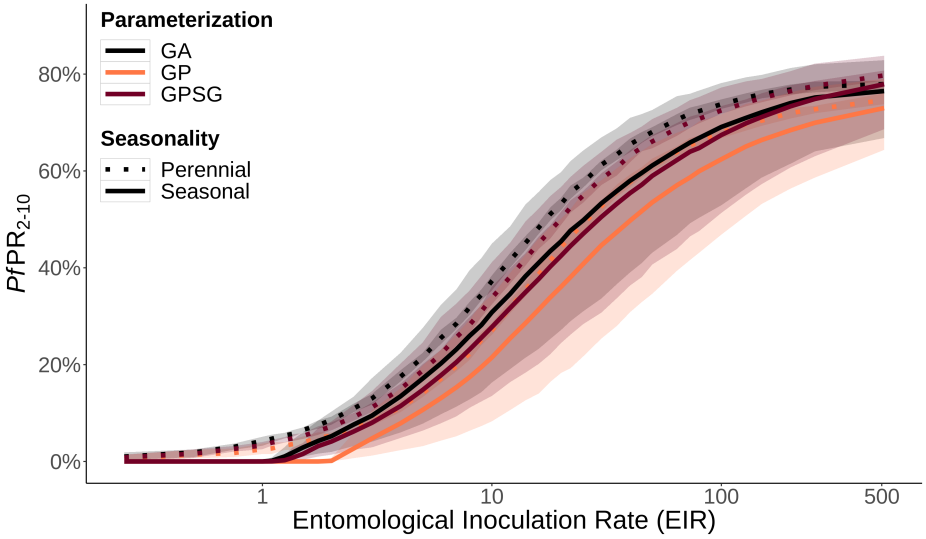

FIGURE A.21: Relationship between EIR and *Pf*PR2-10 under three parameterisations
Solid lines show medians and shaded regions show 95% credible intervals. EIR denotes the entomological inoculation rate.

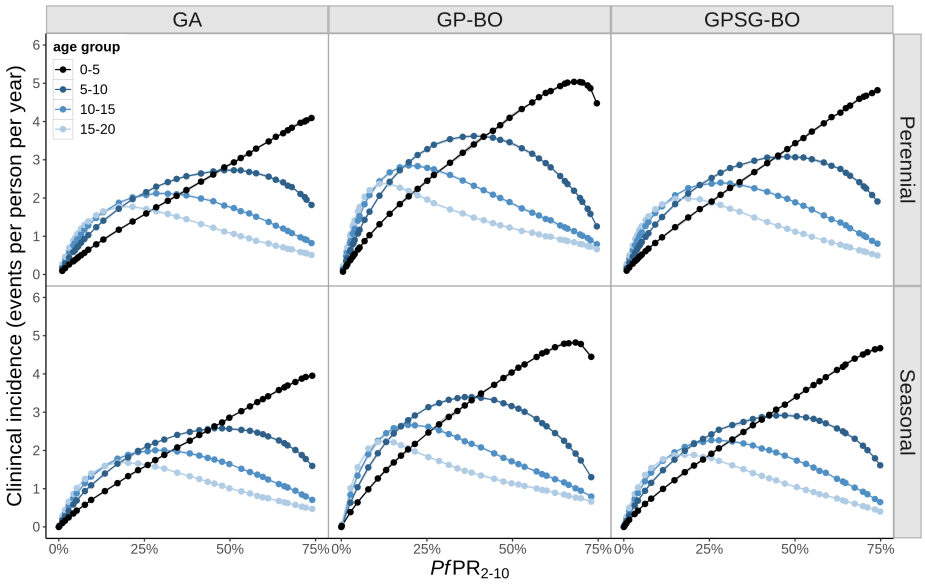

FIGURE A.22: Yearly incidence of clinical (uncomplicated) malaria as a function of $Pf\text{PR}_{2\text{-}10}$ displayed by parameterisation and age group

Clinical incidence is presented in terms of the yearly number of events per person. We assume a probability of effective treatment within 14 days of uncomplicated malaria of 36%.

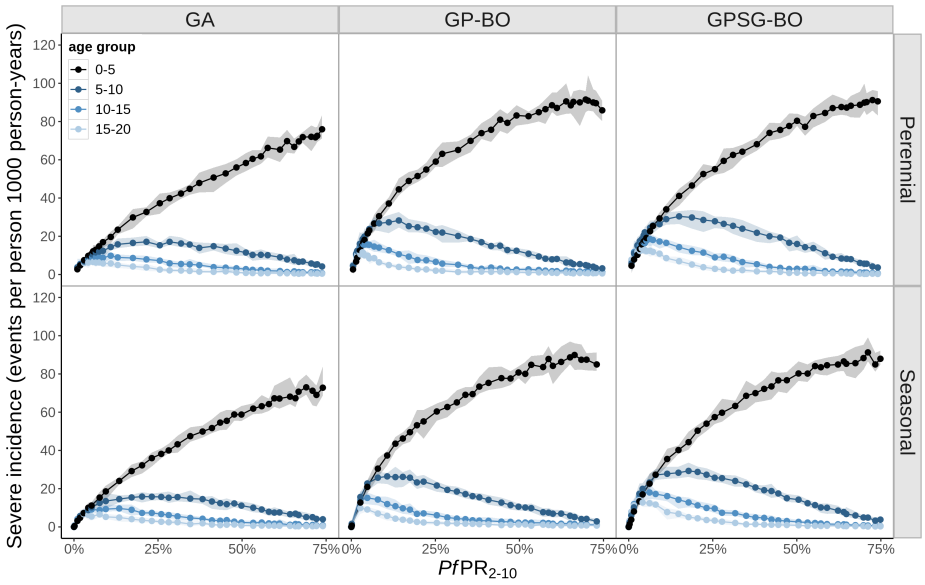

FIGURE A.23: Yearly incidence of total severe malaria

Annual incidence of severe disease as a function of $Pf\text{PR}_{2\text{-}10}$, displayed by parameterisation and age group. Incidence is presented in terms of the yearly number of events in a population of 1000 individuals. It is assumed that 48% of severe malaria cases seek official care at a heath care facility (hospital). We assume a probability of effective treatment within 14 days of uncomplicated malaria of 36%.

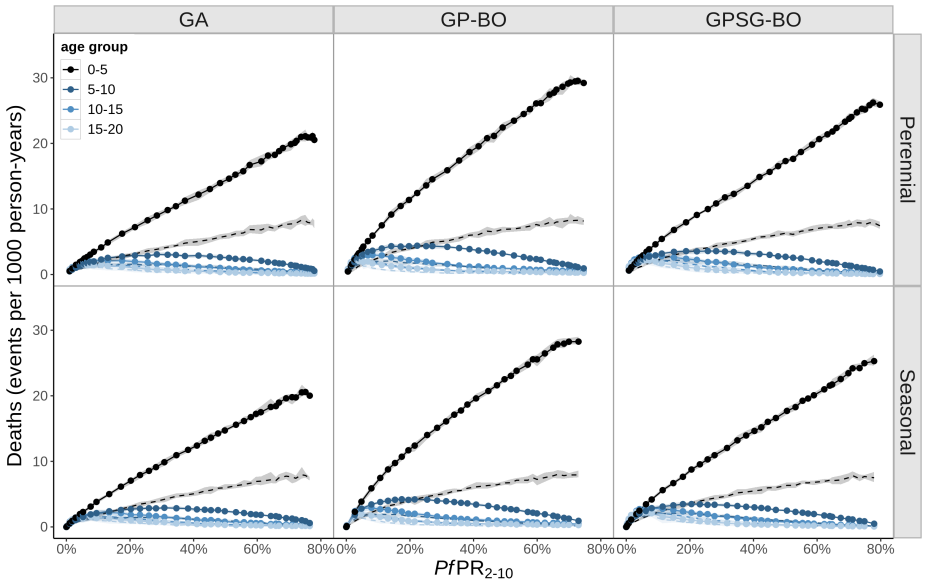

FIGURE A.24: Yearly number of malaria-related deaths

Annual malaria-related deaths as a function of $PfPR_{2\text{-}10}$, displayed by parameterisation and age group. Malaria mortality incidence is presented in terms of the yearly number of deaths in a population of 1000 individuals. For the *OpenMalaria* model both deaths directly attributed to malaria (dotted curve) and all deaths associated with malaria (including both deaths directly attributable to malaria and those associated with comorbidities) are shown (full line).

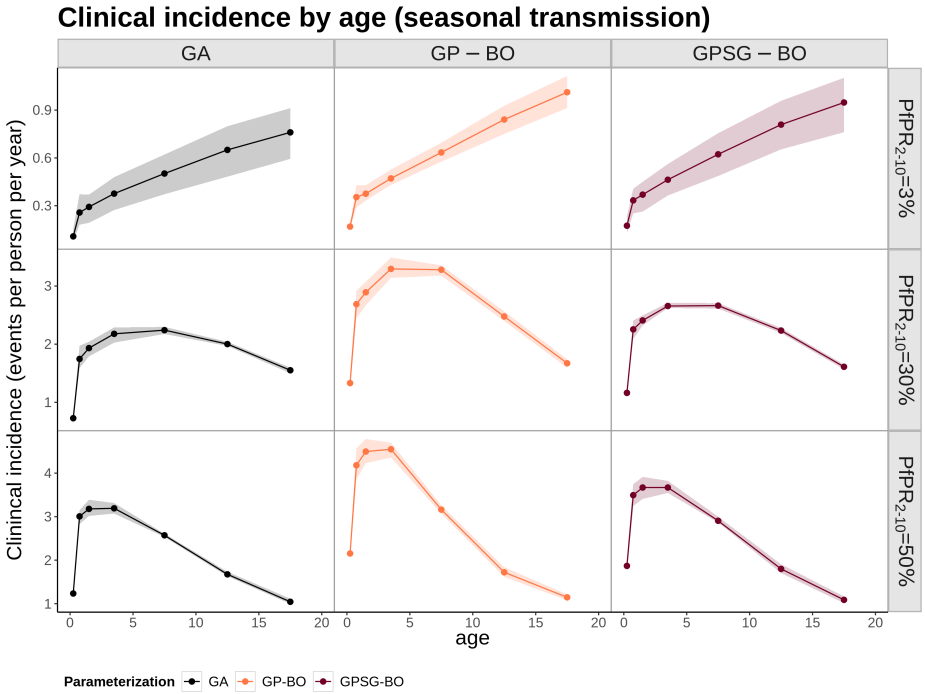

FIGURE A.25: Yearly incidence of clinical malaria (seasonal)

Annual incidence of clinical malaria in a seasonal transmission setting as a function of age, displayed by transmission intensity ($Pf\text{PR}_{2\text{-}10}$) and parameterisation. Clinical incidence is presented in terms of the yearly number of events per person. The $Pf\text{PR}_{2\text{-}10}$ categories include simulated prevalences of 2.5-3.5%, 28-32% , and 47-53% labeled as 3% , 30% , and 50%, respectively.

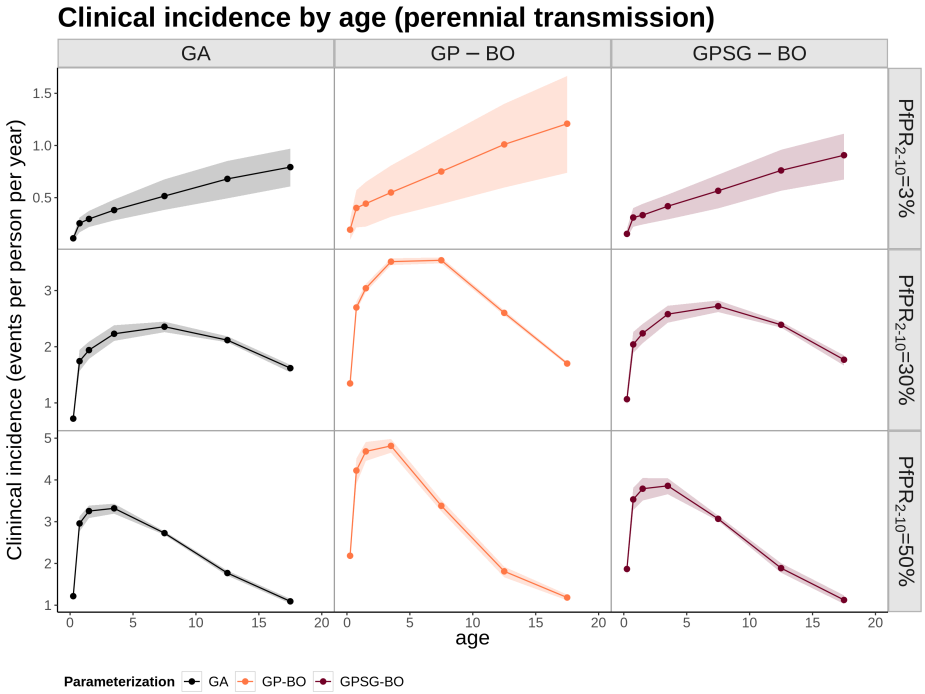

FIGURE A.26: Yearly incidence of clinical malaria (perennial)

Annual incidence of clinical malaria in a perennial transmission setting as a function of age, displayed by transmission intensity ($Pf\text{PR}_{2\text{-}10}$) and parameterisation. Clinical incidence is presented in terms of the yearly number of events per person. The $Pf\text{PR}_{2\text{-}10}$ categories include simulated prevalences of 2.5-3.5%, 28-32% , and 47-53% labeled as 3% , 30% , and 50%, respectively.

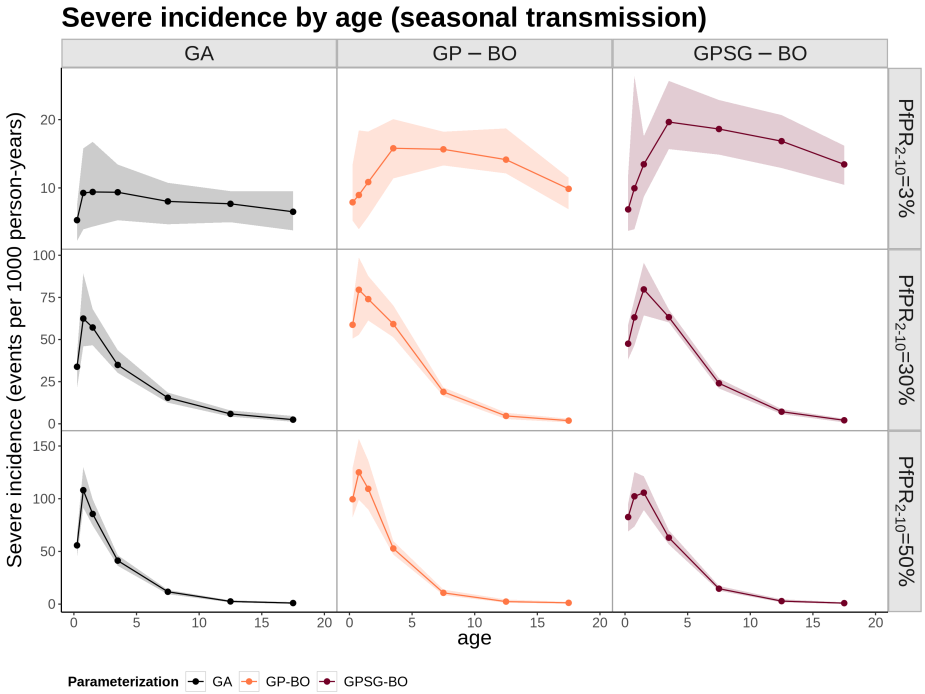

FIGURE A.27: Yearly incidence of total severe malaria (seasonal)

Annual incidence of total severe malaria in a seasonal transmission setting as a function of age, displayed by transmission intensity ($Pf\text{PR}_{2\text{-}10}$) and parameterisation. Incidence is presented in terms of the yearly number of events per 1000 person-years. It is assumed that 48% of severe malaria cases seek official care at a heath care facility (hospital). The $Pf\text{PR}_{2\text{-}10}$ categories include simulated prevalences of 2.5-3.5%, 28-32% , and 47-53% labeled as 3% , 30% , and 50%, respectively.

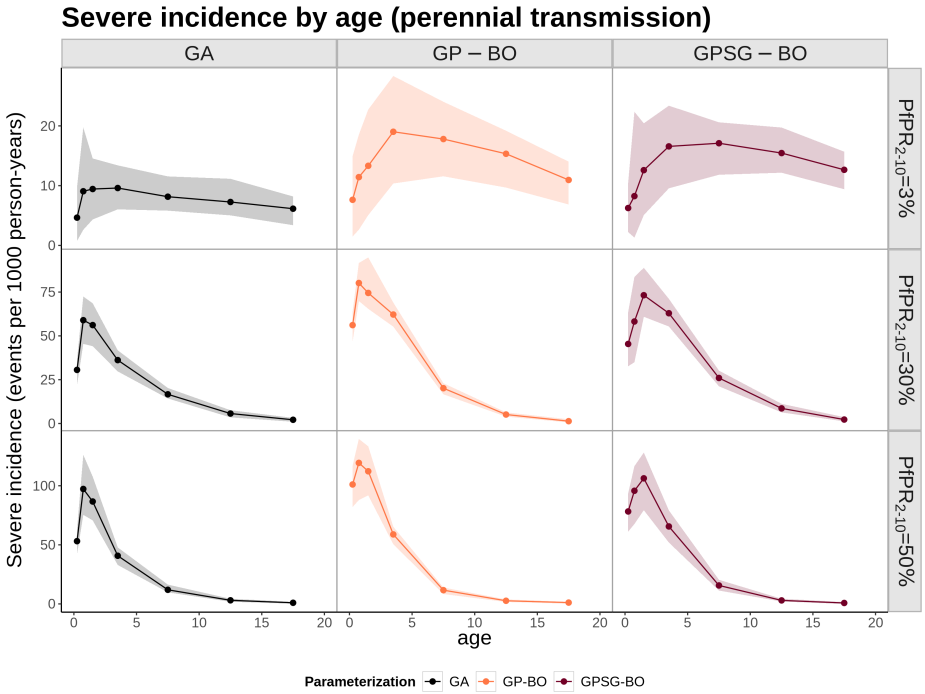

FIGURE A.28: Yearly incidence of total severe malaria (perennial)

Annual incidence of total severe malaria in a perennial transmission setting as a function of age, displayed by transmission intensity ($Pf\text{PR}_{2\text{-}10}$) and parameterisation. Incidence is presented in terms of the yearly number of events per 1000 person-years. It is assumed that 48% of severe malaria cases seek official care at a heath care facility (hospital). The $Pf\text{PR}_{2\text{-}10}$ categories include simulated prevalences of 2.5-3.5%, 28-32% , and 47-53% labeled as 3% , 30% , and 50%, respectively.

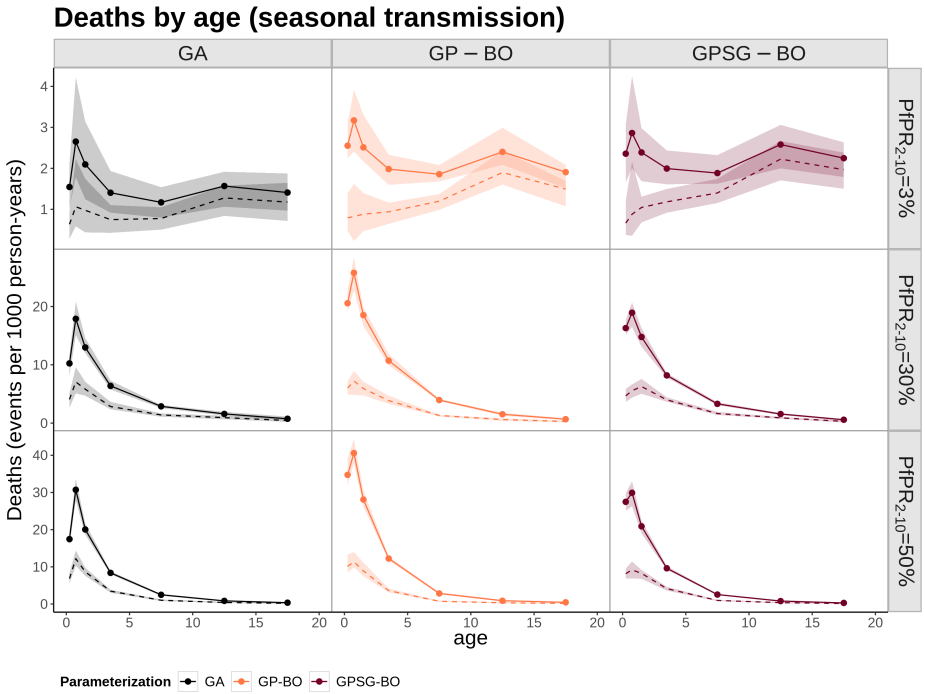

FIGURE A.29: Yearly incidence of malaria-related deaths (seasonal)

Annual incidence of malaria-related deaths in a seasonal transmission setting as a function of age, displayed by transmission intensity ($Pf\text{PR}_{2\text{-}10}$) and parameterisation. Malaria mortality incidence is presented in terms of the yearly number of deaths in a population of 1000 individuals. The dashed estimates represent direct malaria deaths, and the solid all malaria deaths (including those attributable to co-morbidities). The $Pf\text{PR}_{2\text{-}10}$ categories include simulated prevalences of 2.5-3.5%, 28-32% , and 47-53% labeled as 3% , 30% , and 50%, respectively.

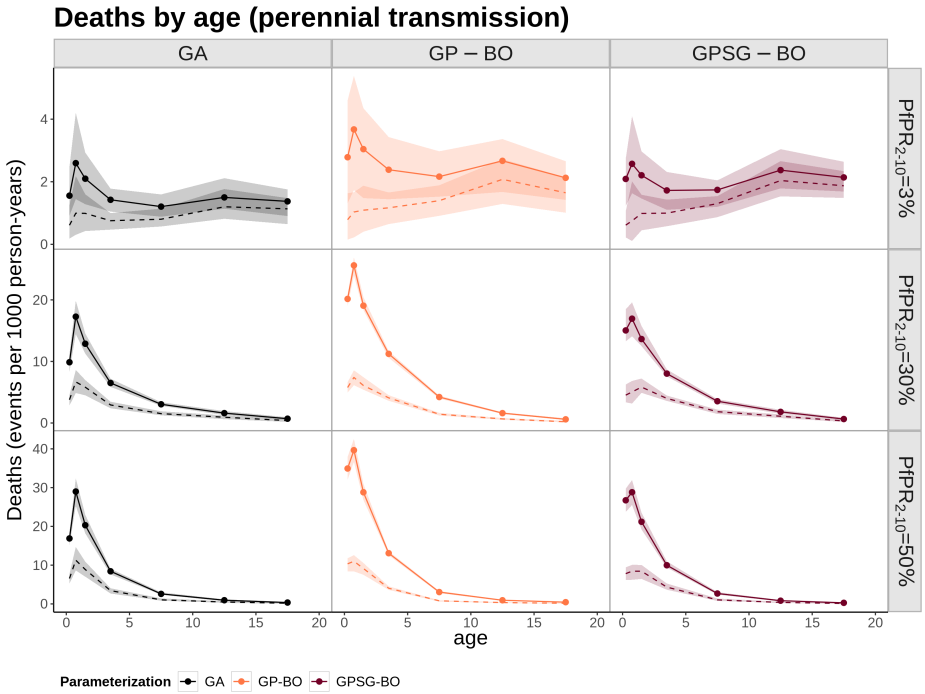

FIGURE A.30: Yearly incidence of malaria-related deaths (perennial)

Annual incidence of malaria-related deaths in a perennial transmission setting as a function of age, displayed by transmission intensity ($Pf\text{PR}_{2\text{-}10}$) and parameterisation. Malaria mortality incidence is presented in terms of the yearly number of deaths in a population of 1000 individuals. The dashed estimates represent direct malaria deaths, and the solid all malaria deaths (including those attributable to co-morbidities). The $Pf\text{PR}_{2\text{-}10}$ categories include simulated prevalences of 2.5-3.5%, 28-32% , and 47-53% labeled as 3% , 30% , and 50%, respectively.

A.8 Log prior distributions and posterior estimates

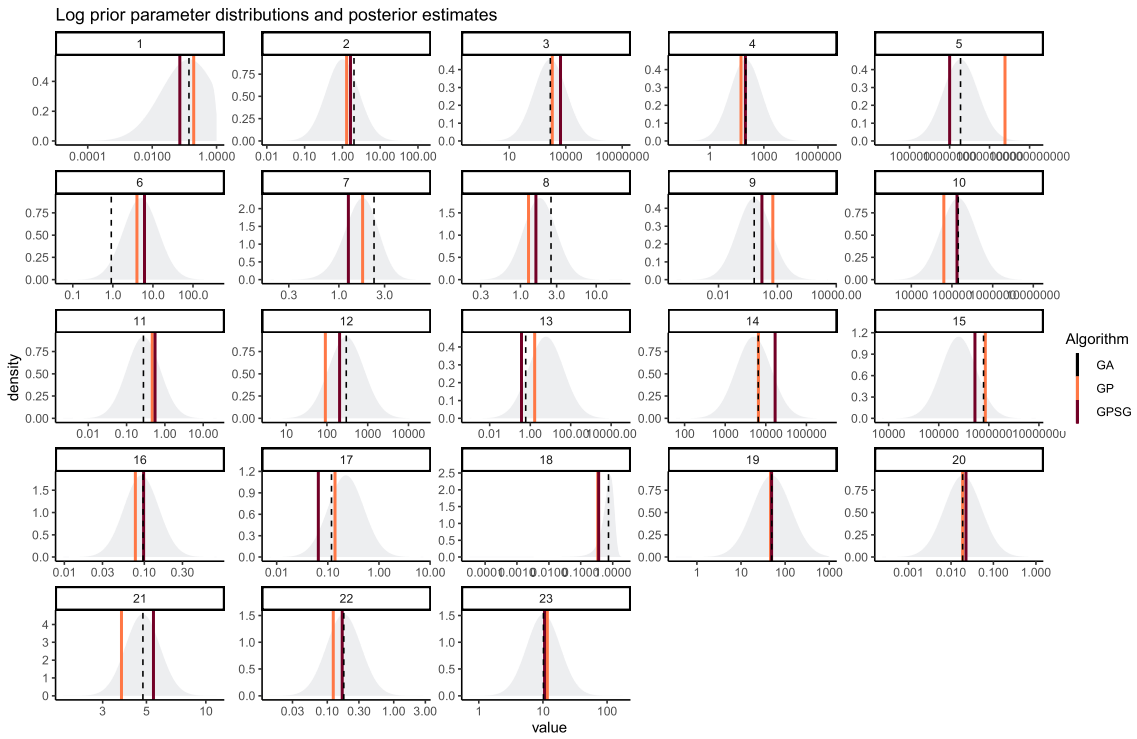

FIGURE A.31: Log prior distributions and final posterior estimates

Prior distributions of each parameter and final parameter values identified by each optimisation algorithm (GP-BO and GPSG-BO) and compared to the current parameterisation (derived using a genetic algorithm, GA).

A.9 Random forest: Ranger importance

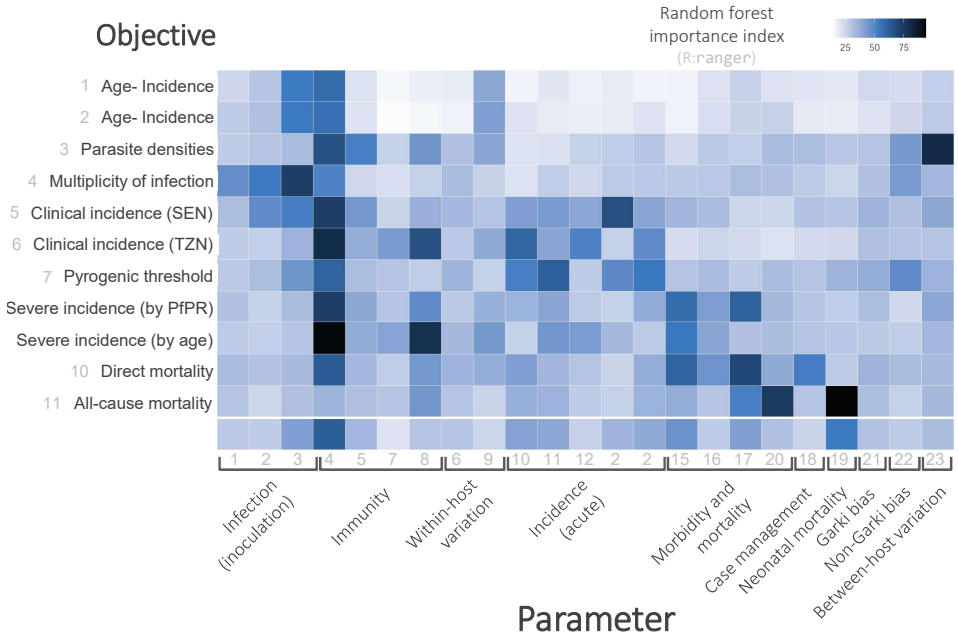

FIGURE A.32: Random forest importance

Estimated parameter importance indices for all parameters and objectives. The indices were calculated using the *ranger* random forest package in R.

B Supplement to: Calibrating infctious disease models to real-world data: Context matters

B.1 Contextual covariates

B.1.1 Population structure and monitoring

Since all incidence and prevalence data was given by age group, study population structures were disregarded. Monitored age groups were set in accordance with age groups reported in the main reference studies. Where incidence age bands did not match prevalence age bands, we simulated monitoring at the smallest intervals, such that all required intervals could be derived by summation.

B.1.2 Transmission intensity and seasonality.

Annual transmission for each site at the according time was extracted from the literature. Seasonality patterns (relative monthly EIR) were derived from site- specific transmission data using EIR values and the following hierarchy of information: 1) Primary reference; 2) Extended literature for the same site; 3) Extended literature for the same region. Where monthly EIR was not available, relative human biting rates or rainfall data with a one-month time lag (`github.com/SwissTPH/openmalaria/wiki/`) were used to derive relative monthly transmission. Where no information on rainfall or EIR seasonality was given in primary or secondary references, historical rainfall data openly accessible through the World Bank Group Climate Change Knowledge Portal (`climateknowledgeportal.worldbank.org/`) was used. Assuming that even if the study year was dryer or wetter than average, rain patterns and relative monthly rainfall should remain relatively constant between years, we extracted average rainfall data for 1961-1990 or 1992-2015. Locations were matched by longitude and latitude. Where data was available for more than one year, the mean monthly value was used in the parameterisation.

B.1.3 Health system

Care seeking behaviour in *OpenMalaria* simulations is defined through the probability that an individual seeks official health care, *pSeekOfficialCareUncomplicated* (and receives curative

treatment). In the simulations, this value must reflect the probability that an individual receives curative treatment, which combines the probability of care in the health care system, as well as from active case detection in the study. Health care system quality of care commonly derived from effective care rate estimates (E_{14}, the probability that an individual receives effective treatment within 14 days of infection). None of the studies included provided estimates of effective care rates. For studies conducted after 2000, we propose using country level ACT coverage data as published by MAP (`https://map.ox.ac.uk/`) to approximate case management levels. Downstream factors reducing effective coverage compared to ACT coverage include adherence to drug regime and cure rate (adequate drug formulation, absence of parasite drug resistance and parasite clearance). In the absence of information to the contrary, we assume that these rates are low, but it is possible that effective care is truly slightly lower (<5%) than in our parameterisation. For studies carried out prior to 2000, no data was available in the MAP database. In order to estimate case management levels, we used the same approach as in the original parameterisation of OpenMalaria as published by Tediosi et al. (*271*). For studies conducted prior to 2000 (and thus prior to malaria awareness campaigns in the considered countries) we assumed low care seeking probabilities for uncomplicated malaria in the formal health sector (0.04), for self-treatment (0.01) and seeking outpatient care in the case of treatment failure (0.04) unless indicated otherwise (*63*). These values originally were derived by triangulating model predictions for clinical episodes with health systems data from Manhica, Mozambique(*63*). In order to simulate the study conditions under which incidence was recorded, active case detection was added at given and 100% coverage or at a coverage equivalent to the follow-up rate provided in the study.

B.1.4 Diagnostics and treatment

Case detection was simulated as with thick blood smear testing and case definition thresholds. Where no threshold was given, we used the sensitivity of the given measure as threshold, 100 parasites per μl for thick blood smears (*272*). For studies considering any patent infection, in routine case management we set the test threshold to 0 because all suspected cases that present will be treated. Rather than explicitly modelling different drugs and their properties, we modelled generic drug treatment and accounted adjusted the effective care value according to the drug (quinine (QN), chloroquine (CQ) sulfadoxine-pyrimethamine (SP), artemetherlumefantrine (AL), amodaquinine (AQ) or artemisinin combination therapy (ACT)). We assume negligible resistance to treatment unless otherwise specified in the literature.

B.1.5 Interventions

For each setting, we consider its history with regards to insecticide-treated net (ITN), distribution, long-lasting insecticide treated net (LLIN) use, indoor residual spraying (IRS), and intermittent preventive treatment in pregnancy (IPTp). We used the following hierarchy of information: 1) Site-specific information from primary references; 2) Region-specific information from LINK malaria profiles (`www.linkmalaria.org/`); 3) Country-specific survey-based information (Demographic and Health Survey (DHS) or Malaria Indicator Survey (MIS));

4) Country-specific imputed information provided by the Malaria Atlas Project (MAP; `https://map.ox.ac.uk`). Prior to 2000 we assumed no interventions unless indicated otherwise.

B.2 Excluded records

TABLE B.1: Excluded records

ID	Study	Location	Country	Year(s)	Exclusion reason
x01	Bloland et al. (*273*)	Asembo Bay	Kenya	1992	PfPR statistically estimated (MAP)
x02	Bougouma et al. (*274*)	Saponé	Burkina Faso	2007	PfPR statistically estimated (MAP)
x03	Coulibaly et al. (*275*)	Bandiagara	Mali	1999	age-specific PfPR extrapolated for all but one age group
x04	Diallo et al. (*276*)	Dakar (central)	Senegal	1996-1997	PfPR statistically estimated (MAP)
x05	Diallo et al. (*277*)	S. Dakar	Senegal	1994	Unstable transmission? Imported infections
x06	Fillol et al. (*278*)	Niakhar	Senegal	2003	PfPR statistically estimated (MAP)
x07	Greenwood et al. (*279*)	Farafenni	The Gambia	1981-1982	PfPR statistically estimated (MAP)
x08	Guinovart et al. (*280*)	Manhiça	Mozambique	2003-2005	PfPR statistically estimated (MAP)
x09	Loha et al. (*281*)	Chano Mille	Ethiopia	2009-2011	PfPR statistically estimated (MAP)
x10	Molez et al. (*282*)	Barkedji	Senegal	1994-1995	PfPR statistically estimated (MAP)
x11	Nebie et al. (*283*)	Baloghin	Burkina Faso	2003	PfPR statistically estimated (MAP)
x12	Owusu-Agyei et al. (*284*)	Kintampo	Ghana	2004	PfPR statistically estimated (MAP)
x13	Rogier et al. (*285*)	Dielmo	Senegal	1990	PfPR statistically estimated (MAP)
x14	Trape et al. (*286*)	Lizolo	Republic of Congo	1983-1984	PfPR statistically estimated (MAP)
x15	Velema et al. (*287*)	Pahou	Benin	1989	PfPR statistically estimated (MAP)

B.3 Included records

TABLE B.2: Overview of references to data and seasonality tables for included studies

sID	Study	Location	Country	Year(s)	Data record	Seasonality table
s01	Ba et al. (*143*)	Ndiop	Senegal	1993-1994	B.3	B.4
s02	Bonnet et al. (*145*)	Koundou	Cameroon	1997-1998	B.5	B.6
s03		Ebolakounou	Cameroon	1997-1998	B.7	B.8
s04	Dicko et al. (*148*)	Douneguebougou	Mali	1999-2000	B.9	B.10
s05		Sotuba	Mali	1999-2000	B.11	B.12
s06	Henry et al. (*151*)	Katiola "R0"	Côte d'Ivoire	1997-1998	B.13	B.14
s07		Korhogo "R1"	Côte d'Ivoire	1997-1998	B.15	B.16
s08		Korhogo "R2"	Côte d'Ivoire	1997-1998	B.17	B.18
s09	Lusingu et al. (*155*)	Mgome	Tanzania	2001	B.19	B.20
s10		Ubiri	Tanzania	2001	B.21	B.22
s11		Magamba	Tanzania	2001	B.23	B.24
s12	Mwangi et al. (*159–161*)	Ngerenya	Kenya	2001	B.25	B.26
s13		Chonyi	Kenya	2001	B.27	B.28
s14	Saute et al. (*167*, *168*)	Manhiça	Mozambique	1996-1999	B.29	B.30
s15	Schellenberg et al. (*172*)	Ifakara	Tanzania	2000-2001	B.31	B.32
s16	Thomposon et al. (*175*)	Matola	Mozambique	1992-1995	B.33	B.34
s17	Trape et al. (*177*)	Dielmo	Senegal	2007-2008	B.35	B.36

B.3.1 Ba et al. (2000): Ndiop, Senegal

Data

TABLE B.3: *Pf*PR-incidence records from Ba et al. (*143*)

Data extracted from Battle et al. (*140*). LAR = Lower bound of the incidence age range, UAR = Upper bound of the incidence age range, INC 1000 PYO^{-1} = incidence (number of events) per 1,000 person-years observed.

Study	Year	LAR	UAR	N	INC 1000 PYO^{-1}	*Pf*PR
Ba 2000	1993	0	1	26	1231	0.226
Ba 2000	1993	1	2	15	1867	0.155
Ba 2000	1993	2	4	40	2600	0.229
Ba 2000	1993	5	9	61	2574	0.287
Ba 2000	1993	10	14	34	1647	0.41
Ba 2000	1993	15	19	41	780	0.537
Ba 2000	1993	20	39	96	583	0.344
Ba 2000	1993	40	59	31	484	0.251
Ba 2000	1993	60	85	7	857	0.287

Seasonality

TABLE B.4: Seasonality profile for Ndiop, Senegal

Inferred from Fontenille et al. (*144*)

Month	Monthly EIR	Relative transmission
Jul 93	1.283	0.0203
Aug 93	9.626	0.152
Sep 93	38.289	0.605
Oct 93	14.118	0.223
Nov 93	0	0
Dec 93	0	0
Jan 94	0	0
Feb 94	0	0
Mar 94	0	0
Apr 94	0	0
May 94	0	0
Jun 94	0	0
Jul 94	0	0
Aug 94	7.273	0.425
Sep 94	8.342	0.488
Oct 94	1.497	0.088
Nov 94	0	0

B.3.2 Bonnet et al. (2002): Koundou, Cameroon

Data

TABLE B.5: *Pf*PR-incidence records from (*145*)

Data extracted from Battle et al. (*140*). LAR = Lower bound of the incidence age range, UAR = Upper bound of the incidence age range, INC 1000 PYO^{-1} = incidence (number of events) per 1,000 person-years observed.

Study	Year	LAR	UAR	N INC	INC 1000 PYO^{-1}	PR parasitaemia	PR carrier
Bonnet 2002 Koundou	1997	0	1	15	400	NA	NA
Bonnet 2002 Koundou	1997	2	5	23	1000	NA	NA
Bonnet 2002 Koundou	1997	0	5	NA	NA	0.264	0.667
Bonnet 2002 Koundou	1997	6	10	29	610	0.202	0.61
Bonnet 2002 Koundou	1997	11	15	20	240	0.159	0.695
Bonnet 2002 Koundou	1997	16	85	110	160	0.067	0.395

Seasonality

TABLE B.6: Seasonality profile for Koundou, Cameroon

Inferred from Bonnet et al. (*145*)

Month	Monthly EIR	Relative transmission
Jun 97	22.15	0.295
Jul 97	33.72	0.449
Aug 97	12.78	0.170
Sep 97	0.37	0.005
Oct 97	0.1	0.001
Nov 97	5.34	0.071
Dec 97	0.65	0.009
Jan 98	0.1	0.001
Feb 98	0.37	0.003
Mar 98	11.12	0.103
Apr 98	0.37	0.003
May 98	96.27	0.889
Jun 98	0	0

B.3.3 Bonnet et al. (2002): Ebolakounou, Cameroon

Data

TABLE B.7: *Pf*PR-incidence records from (*145*)

Data extracted from Battle et al. (*140*). LAR = Lower bound of the incidence age range, UAR = Upper bound of the incidence age range, INC 1000 PYO^{-1} = incidence (number of events) per 1,000 person-years observed.

Study	Year	LAR	UAR	N INC	INC 1000 PYO^{-1}	PR parasitaemia	PR carrier
Bonnet 2002 Ebolakounou	1997	0	1	38	182	NA	NA
Bonnet 2002 Ebolakounou	1997	2	5	79	99	NA	NA
Bonnet 2002 Ebolakounou	1997	0	5	NA	NA	0.212	0.591
Bonnet 2002 Ebolakounou	1997	6	10	69	80	0.17	0.662
Bonnet 2002 Ebolakounou	1997	11	15	53	45	0.124	0.586
Bonnet 2002 Ebolakounou	1997	16	85	288	13	0.053	0.343

Seasonality

TABLE B.8: Seasonality profile for Ebolakounou, Cameroon

Inferred from Bonnet et al. (*145*)

Month	Monthly EIR	Relative transmission
Jun 97	2.24	1
Jul 97	0	0
Aug 97	0	0
Sep 97	0	0
Oct 97	0	0
Nov 97	0	0
Dec 97	0	0
Jan 98	0	0
Feb 98	0	0
Mar 98	0	0
Apr 98	0	0
May 98	15.48	1
Jun 98	0	0

B.3.4 Dicko et al. (2007): Douneguebougou, Mali

Data

TABLE B.9: *Pf*PR-incidence records from (*148*)

Data extracted from Battle et al. (*140*). LAR = Lower bound of the incidence age range, UAR = Upper bound of the incidence age range, INC 1000 PYO^{-1} = incidence (number of events) per 1,000 person-years observed.

Study	Year	LAR	UAR	N	INC 1000 PYO^{-1}	*Pf*PR
Dicko 2007 Douneguebougou	1999	0	5	53	2000	0.395
Dicko 2007 Douneguebougou	1999	6	10	48	1395.83	0.677
Dicko 2007 Douneguebougou	1999	11	15	52	1153.85	0.664
Dicko 2007 Douneguebougou	1999	16	20	43	674.42	0.461
Dicko 2007 Douneguebougou	2000	0.25	5	53	1962.26	0.383
Dicko 2007 Douneguebougou	2000	6	10	48	1634.62	0.608
Dicko 2007 Douneguebougou	2000	11	15	52	1000	0.524
Dicko 2007 Douneguebougou	2000	16	20	43	720.93	0.397

Seasonality

TABLE B.10: Seasonality profile for Douneguebougou, Mali

Inferred from Dicko et al. (*148*)

Month	Monthly EIR	Relative transmission
Jun 99	0	0
Jul 99	4.34	0.026
Aug 99	25.32	0.151
Sep 99	53.81	0.322
Oct 99	58.89	0.352
Nov 99	20.83	0.125
Dec 99	4.05	0.024
Jun 00	0.08	0.001
Jul 00	0.54	0.004
Aug 00	22.07	0.161
Sep 00	58.4	0.425
Oct 00	42.56	0.310
Nov 00	11.41	0.083
Dec 00	2.26	0.016

B.3.5 Dicko et al. (2007): Sotuba, Mali

Data

TABLE B.11: *Pf*PR-incidence records from (*148*)

Data extracted from Battle et al. (*140*). LAR = Lower bound of the incidence age range, UAR = Upper bound of the incidence age range, INC 1000 PYO^{-1} = incidence (number of events) per 1,000 person-years observed.

Study	Year	LAR	UAR	N	INC 1000 PYO^{-1}	*Pf*PR
Dicko 2007 Sotuba	1999	0	5	58	1637.93	0.145
Dicko 2007 Sotuba	1999	6	10	49	1979.59	0.171
Dicko 2007 Sotuba	1999	11	15	48	2104.17	0.166
Dicko 2007 Sotuba	1999	16	20	46	1869.57	0.120
Dicko 2007 Sotuba	2000	0.25	5	51	745.1	0.040
Dicko 2007 Sotuba	2000	6	10	53	1264.15	0.074
Dicko 2007 Sotuba	2000	11	15	49	918.37	0.048
Dicko 2007 Sotuba	2000	16	20	46	1021.74	0.081

Seasonality

TABLE B.12: Seasonality profile for Sotuba, Mali

Inferred from Dicko et al. (*148*)

Month	monthly EIR	Relative transmission
Jun 99	0	0
Jul 99	1.7	0.139
Aug 99	1.23	0.100
Sep 99	1.67	0.136
Oct 99	5.54	0.452
Nov 99	2.12	0.173
Dec 99	0	0
Jun 00	0	0
Jul 00	0.09	0.025
Aug 00	1.41	0.387
Sep 00	1.91	0.525
Oct 00	0.23	0.063
Nov 00	0	0
Dec 00	0	0

B.3.6 Henry et al. (2003): Katiola R0, Côte d'Ivoire

Data

TABLE B.13: *Pf*PR-incidence records from (*151*)

Data extracted from Battle et al. (*140*). LAR = Lower bound of the incidence age range, UAR = Upper bound of the incidence age range, INC 1000 PYO^{-1} = incidence (number of events) per 1,000 person-years observed.

Study	Year	LAR	UAR	N	INC 1000 PYO^{-1}	*Pf*PR	INC diag threshold
Henry 2003 Katiola R0	1997	0	2	1116	4800	0.54506	2500
Henry 2003 Katiola R0	1997	2	5	1687	2300	0.9103	2500
Henry 2003 Katiola R0	1997	5	10	2870	500	0.87	1000
Henry 2003 Katiola R0	1997	10	20	2667	600	0.753773	500
Henry 2003 Katiola R0	1997	20	40	2968	300	0.572175	500
Henry 2003 Katiola R0	1997	40	85	2997	200	0.62	500

Seasonality

TABLE B.14: Seasonality profile for Katiola "R0", Côte d'Ivoire

Inferred from Henry et al. (*151*)

Month	Bites per 1000 man-nights	Relative transmission
Dec 96	311.3208	0.073
Jan 97	NA	NA
Feb 97	400.9434	0.095
Mar 97	28.30189	0.007
Apr 97	731.1321	0.172798
May 97	NA	NA
Jun 97	608.4906	0.144
Jul 97	613.2075	0.145
Aug 97	311.3208	0.074
Sep 97	NA	NA
Oct 97	613.2075	0.145
Nov 97	613.2075	0.145

B.3.7 Henry et al. (2003): Korhogo R1, Côte d'Ivoire

Data

TABLE B.15: *Pf*PR-incidence records from (*151*)

Data extracted from Battle et al. (*140*). LAR = Lower bound of the incidence age range, UAR = Upper bound of the incidence age range, INC 1000 PYO^{-1} = incidence (number of events) per 1,000 person-years observed.

Study	Year	LAR	UAR	N	INC 1000 PYO^{-1}	*Pf*PR	INC diag threshold
Henry 2003 Korhogo R1	1997	0	2	1065	3000	0.526264	2500
Henry 2003 Korhogo R1	1997	2	5	1356	1100	0.87891	2500
Henry 2003 Korhogo R1	1997	5	10	2516	800	0.84	1000
Henry 2003 Korhogo R1	1997	10	20	3479	200	0.802404	500
Henry 2003 Korhogo R1	1997	20	40	2868	100	0.609089	500
Henry 2003 Korhogo R1	1997	40	85	3257	100	0.66	500

Seasonality

TABLE B.16: Seasonality profile for Korhogo "R1", Côte d'Ivoire

Inferred from Henry et al. (*151*)

Month	Bites per 1000 man-nights	Relative transmission
Dec 96	161.017	0.065
Jan 97	59.322	0.024
Feb 97	NA	NA
Mar 97	NA	NA
Apr 97	144.068	0.058
May 97	NA	NA
Jun 97	609.208	0.246
Jul 97	899.377	0.363
Aug 97	302.691	0.122
Sep 97	NA	NA
Oct 97	NA	NA
Nov 97	298.605	0.121

B.3.8 Henry et al. (2003): Korhogo R2, Côte d'Ivoire

Data

TABLE B.17: *Pf*PR-incidence records from (*151*)

Data extracted from Battle et al. (*140*). LAR = Lower bound of the incidence age range, UAR = Upper bound of the incidence age range, INC 1000 PYO^{-1} = incidence (number of events) per 1,000 person-years observed.

Study	Year	LAR	UAR	N	INC 1000 PYO^{-1}	*Pf*PR	INC diag threshold
Henry 2003 Korhogo R2	1997	0	2	970	3000	0.494939196	2500
Henry 2003 Korhogo R2	1997	2	5	1132	3000	0.826593872	2500
Henry 2003 Korhogo R2	1997	5	10	2113	1200	0.79	1000
Henry 2003 Korhogo R2	1997	10	20	3026	300	0.717300301	500
Henry 2003 Korhogo R2	1997	20	40	2703	100	0.544488813	500
Henry 2003 Korhogo R2	1997	40	85	4028	100	0.59	500

Seasonality

TABLE B.18: Seasonality profile for Korhogo "R2", Côte d'Ivoire

Inferred from Henry et al. (*151*)

Month	Bites per 1000 man-nights	Relative transmission
Dez 96	747.604	0.201
Jan 97	67.093	0.018
Feb 97	NA	NA
Mär 97	57.508	0.015
Apr 97	31.949	0.009
Mai 97	NA	NA
Jun 97	904.153	0.243
Jul 97	699.681	0.188
Aug 97	207.668	0.056
Sep 97	NA	NA
Okt 97	405.751	0.109
Nov 97	603.834	0.162

B.3.9 Lusingu et al. (2004): Mgome, Tanzania

Data

TABLE B.19: *Pf*PR-incidence records from (*155*)

Data extracted from Battle et al. (*140*). LAR = Lower bound of the incidence age range, UAR = Upper bound of the incidence age range, INC 1000 PYO^{-1} = incidence (number of events) per 1,000 person-years observed.

Study	Year	LAR	UAR	N	INC 1000 PYO^{-1}	*Pf*PR
Lusingu 2004 Mgome	2001	0	1	14	3858.46	0.594948
Lusingu 2004 Mgome	2001	1	2	13	3812.63	0.7959
Lusingu 2004 Mgome	2001	2	3	13	3143.58	0.8289
Lusingu 2004 Mgome	2001	3	4	13	632.38	0.9216
Lusingu 2004 Mgome	2001	4	5	13	1228.11	0.9688
Lusingu 2004 Mgome	2001	5	9	60	412.42	0.9135
Lusingu 2004 Mgome	2001	10	14	53	384.93	0.7299
Lusingu 2004 Mgome	2001	15	19	45	311.61	0.5588

Seasonality

TABLE B.20: Seasonality profile for Mgome, Tanzania

Inferred from Bødker et al. (*156*). Relative rainfall plus one month lag.

Month	Relative transmission
Oct 95	0.000
Nov 95	0.000
Dec 95	0.000
Jan 96	0.000
Feb 96	0.029
Mar 96	0.079
Apr 96	0.066
May 96	0.308
Jun 96	0.395
Jul 96	0.027
Aug 96	0.032
Sep 96	0.013
Oct 96	0.012
Nov 96	0.032
Dec 96	0.008

B.3.10 Lusingu et al. (2004): Ubiri, Tanzania

Data

TABLE B.21: *Pf*PR-incidence records from (*155*)

Data extracted from Battle et al. (*140*). LAR = Lower bound of the incidence age range, UAR = Upper bound of the incidence age range, INC 1000 PYO^{-1} = incidence (number of events) per 1,000 person-years observed.

Study	Year	LAR	UAR	N	INC 1000 PYO^{-1}	*Pf*PR
Lusingu 2004 Ubiri	2001	0	1	14	0	0.0873
Lusingu 2004 Ubiri	2001	1	2	13	219.96	0.2942
Lusingu 2004 Ubiri	2001	2	3	13	714.87	0.4042
Lusingu 2004 Ubiri	2001	3	4	13	192.46	0.2302
Lusingu 2004 Ubiri	2001	4	5	13	568.23	0.2089
Lusingu 2004 Ubiri	2001	5	9	59	119.15	0.2399
Lusingu 2004 Ubiri	2001	10	14	52	0	0.2385
Lusingu 2004 Ubiri	2001	15	19	44	109.98	0.3745

Seasonality

TABLE B.22: Seasonality profile for Ubiri, Tanzania

Inferred from Bødker et al. (*156*). Relative rainfall plus one month lag.

Month	Relative transmission
Oct 95	0.008
Nov 95	0.002
Dec 95	0.000
Jan 96	0.014
Feb 96	0.021
Mar 96	0.048
Apr 96	0.100
May 96	0.281
Jun 96	0.311
Jul 96	0.042
Aug 96	0.026
Sep 96	0.013
Oct 96	0.008
Nov 96	0.050
Dec 96	0.076

B.3.11 Lusingu et al. (2004): Magamba, Tanzania

Data

TABLE B.23: *Pf*PR-incidence records from (*155*)

Data extracted from Battle et al. (*140*). LAR = Lower bound of the incidence age range, UAR = Upper bound of the incidence age range, INC 1000 PYO^{-1} = incidence (number of events) per 1,000 person-years observed.

Study	Year	LAR	UAR	N	INC 1000 PYO^{-1}	*Pf*PR
Lusingu 2004 Magamba	2001	0	1	14	0	0.0885
Lusingu 2004 Magamba	2001	1	2	13	0	0.1055
Lusingu 2004 Magamba	2001	2	3	13	0	0.0366
Lusingu 2004 Magamba	2001	3	4	13	0	0.0259
Lusingu 2004 Magamba	2001	4	5	13	0	0.1153
Lusingu 2004 Magamba	2001	5	9	60	100.82	0.0858
Lusingu 2004 Magamba	2001	10	14	53	0	0.1144
Lusingu 2004 Magamba	2001	15	19	45	0	0.0277

Seasonality

TABLE B.24: Seasonality profile for Magamba, Tanzania

Inferred from Bødker et al. (*156*). Relative rainfall plus one month lag.

Month	Relative transmission
Oct 95	0.000
Nov 95	0.000
Dec 95	0.000
Jan 96	0.000
Feb 96	0.000
Mar 96	0.152
Apr 96	0.184
May 96	0.374
Jun 96	0.201
Jul 96	0.000
Aug 96	0.000
Sep 96	0.000
Oct 96	0.000
Nov 96	0.000
Dec 96	0.089

B.3.12 Mwangi et al. (2003, 2005): Ngerenya, Kenya

Data

TABLE B.25: *Pf*PR-incidence records from (*159*) and (*161*)

Data extracted from Battle et al. (*140*). LAR = Lower bound of the incidence age range, UAR = Upper bound of the incidence age range, INC 1000 PYO^{-1} = incidence (number of events) per 1,000 person-years observed.

Study	Year	LAR	UAR	N	INC 1000 PYO^{-1}	*Pf*PR	INC diag threshold	sensitivity	specificity
Mwangi 2005 Ngerenya	1999	0	1	NA	862	0.1023	1	1	0.95
Mwangi 2005 Ngerenya	1999	1	2	NA	1408	0.1616	2500 p ul-1	0.95	0.89
Mwangi 2005 Ngerenya	1999	2	3	NA	1574	0.3489	2500 p ul-1	0.95	0.89
Mwangi 2005 Ngerenya	1999	3	4	NA	1462	0.357	2500 p ul-1	0.95	0.89
Mwangi 2005 Ngerenya	1999	4	5	NA	1541	0.3653	2500 p ul-1	0.95	0.89
Mwangi 2005 Ngerenya	1999	5	6	NA	1513	0.3907	2500 p ul-1	0.95	0.89
Mwangi 2005 Ngerenya	1999	6	7	NA	1658	0.4118	2500 p ul-1	0.95	0.88
Mwangi 2005 Ngerenya	1999	7	8	NA	1073	0.4158	2500 p ul-1	0.95	0.88
Mwangi 2005 Ngerenya	1999	8	9	NA	924	0.3857	2500 p ul-1	0.95	0.88
Mwangi 2005 Ngerenya	1999	9	10	NA	765	0.3342	2500 p ul-1	0.95	0.88
Mwangi 2005 Ngerenya	1999	10	11	NA	501	0.3085	2500 p ul-1	0.9	0.93
Mwangi 2005 Ngerenya	1999	11	14	NA	425	0.224096	2500 p ul-1	0.9	0.93
Mwangi 2005 Ngerenya	1999	15	19	NA	377	0.231	1	1	0.87
Mwangi 2005 Ngerenya	1999	20	39	NA	106	0.216493	1	1	0.87
Mwangi 2005 Ngerenya	1999	40	59	NA	47	0.182502	1	1	0.87
Mwangi 2005 Ngerenya	1999	60	85	NA	48	0.0511	1	1	0.87

Seasonality

TABLE B.26: Seasonality profile for Ngerenya, Kenya

Inferred from Mbogo et al. (*162*).

Month	Relative transmission
Jan	0.286
Feb	0.000
Mar	0.000
Apr	0.000
May	0.000
Jun	0.000
Jul	0.508
Aug	0.206
Sep	0.000
Oct	0.000
Nov	0.000
Dec	0.000

B.3.13 Mwangi et al. (2003, 2005): Chonyi, Kenya

Data

TABLE B.27: *Pf*PR-incidence records from (*159*) and (*161*)

Data extracted from Battle et al. (*140*). LAR = Lower bound of the incidence age range, UAR = Upper bound of the incidence age range, INC 1000 PYO^{-1} = incidence (number of events) per 1,000 person-years observed.

Study	Year	LAR	UAR	N	INC 1000 PYO^{-1}	*Pf*PR	INC diag threshold	sensitivity	specificity
Mwangi 2005 Chonyi	1999	0	1	NA	1340	0.1962	1	1	0.95
Mwangi 2005 Chonyi	1999	1	2	NA	1411	0.3878	2500 p ul-1	0.95	0.89
Mwangi 2005 Chonyi	1999	2	3	NA	1503	0.5837	2500 p ul-1	0.95	0.89
Mwangi 2005 Chonyi	1999	3	4	NA	1337	0.5704	2500 p ul-1	0.95	0.89
Mwangi 2005 Chonyi	1999	4	5	NA	982	0.5744	2500 p ul-1	0.95	0.89
Mwangi 2005 Chonyi	1999	5	6	NA	460	0.5913	2500 p ul-1	0.95	0.89
Mwangi 2005 Chonyi	1999	6	7	NA	510	0.6123	2500 p ul-1	0.95	0.88
Mwangi 2005 Chonyi	1999	7	8	NA	477	0.6292	2500 p ul-1	0.95	0.88
Mwangi 2005 Chonyi	1999	8	9	NA	372	0.629	2500 p ul-1	0.95	0.88
Mwangi 2005 Chonyi	1999	9	10	NA	304	0.6128	2500 p ul-1	0.95	0.88
Mwangi 2005 Chonyi	1999	10	11	NA	276	0.5859	2500 p ul-1	0.9	0.93
Mwangi 2005 Chonyi	1999	11	14	NA	279	0.4605	2500 p ul-1	0.9	0.93
Mwangi 2005 Chonyi	1999	15	19	NA	234	0.372	1	1	0.87
Mwangi 2005 Chonyi	1999	20	39	NA	99	0.196048	1	1	0.87
Mwangi 2005 Chonyi	1999	40	59	NA	29	0.182502	1	1	0.87
Mwangi 2005 Chonyi	1999	60	85	NA	48	0.2688	1	1	0.87

Seasonality

TABLE B.28: Seasonality profile for Chonyi, Kenya

Inferred from Mbogo et al. (*162*).

Month	Relative transmission
Jan	0.286
Feb	0.000
Mar	0.000
Apr	0.000
May	0.000
Jun	0.000
Jul	0.508
Aug	0.206
Sep	0.000
Oct	0.000
Nov	0.000
Dec	0.000

B.3.14 Saute et al. (2003): Manhiça, Mozmbique

Data

TABLE B.29: *Pf*PR-incidence records from (*168*) and (*167*)

Data extracted from Battle et al. (*140*). LAR = Lower bound of the incidence age range, UAR = Upper bound of the incidence age range, INC 1000 PYO^{-1} = incidence (number of events) per 1,000 person-years observed.

Study	Year	LAR	UAR	N	INC 1000 PYO^{-1}	*Pf*PR
Saute 2003	1999	0	1	235	505.53	0.1117
Saute 2004	1999	1	2	125	188.9	0.2216
Saute 2005	1999	2	3	125	322.06	0.242
Saute 2006	1999	3	4	125	257.03	0.3062
Saute 2007	1999	4	5	125	157.94	0.2302
Saute 2008	1999	5	10	245	203.82	0.259967

Seasonality

TABLE B.30: Seasonality profile for Manhiça, Mozmbique

Inferred from Mendis et al. (*176*).

month	EIR	Relative transmisison
Nov 94	0	NA
Dec 94	0.75	NA
Jan 95	1.03	0.08
Feb 95	0.5	0.04
Mar 95	3.73	0.29
Apr 95	0	0.00
May 95	1.67	0.13
Jun 95	2.28	0.18
Jul 95	0	0.00
Aug 95	0.67	0.05
Sep 95	0	0.00
Oct 95	0.42	0.03
Nov 95	1.39	0.11
Dec 95	1.07	0.08
Jan 96	2.05	NA
Feb 96	4.3	NA
Mar 96	0	NA
Apr 96	0	NA

B.3.15 Schellenberg et al. (2003): Ifakara, Tanzania

Data

TABLE B.31: *Pf*PR-incidence records from (*172*)

Data extracted from Battle et al. (*140*). LAR = Lower bound of the incidence age range, UAR = Upper bound of the incidence age range, INC 1000 PYO^{-1} = incidence (number of events) per 1,000 person-years observed.

Study	Year	LAR	UAR	N	INC 1000 PYO^{-1}	*Pf*PR
Schellenberg 2002	2000	0	1	191	129.32	0.043
Schellenberg 2003	2000	1	2	115	174.44	0.121
Schellenberg 2004	2000	2	3	158	102.06	0.218
Schellenberg 2005	2000	3	4	107	138.67	0.25
Schellenberg 2006	2000	4	5	47	212.56	0.176
Schellenberg 2007	2000	0	5	618	143.09	0.1655

Seasonality

TABLE B.32: Seasonality profile for Ifakara, Tanzania

Inferred from Drakeley et al. (*173*).

Month	Relative transmission
Mar	0.319
Apr	0.311
May	0.016
Jun	0.047
Jul	0.004
Aug	0.047
Sep	0.008
Oct	0.000
Nov	0.023
Dec	0.140
Jan	0.031
Feb	0.054

B.3.16 Thompson et al. (1997): Matola, Mozmbique

Data

TABLE B.33: *Pf*PR-incidence records from (*175*)

Data extracted from Battle et al. (*140*). LAR = Lower bound of the incidence age range, UAR = Upper bound of the incidence age range, INC 1000 PYO^{-1} = incidence (number of events) per 1,000 person-years observed.

Study	Year	LAR	UAR	N	PYO	INC 1000 PYO^{-1}	*Pf*PR
Thompson 1997	1994	0	1	NA	70.5882	340	0.33
Thompson 1998	1994	2	4	NA	77.7778	630	0.378
Thompson 1999	1994	5	9	NA	111.5385	260	0.382
Thompson 2000	1994	10	14	NA	100	80	0.362
Thompson 2001	1994	15	19	NA	85.7143	70	0.307
Thompson 2002	1994	20	39	NA	250	40	0.3
Thompson 2003	1994	40	85	NA	100	40	0.154

Seasonality

TABLE B.34: Seasonality profile for Matola, Mozambique

From Mendis et al. (*176*).

month	EIR	Relative transmisison
Nov 94	0	NA
Dec 94	0.75	NA
Jan 95	1.03	0.08
Feb 95	0.5	0.04
Mar 95	3.73	0.29
Apr 95	0	0.00
May 95	1.67	0.13
Jun 95	2.28	0.18
Jul 95	0	0.00
Aug 95	0.67	0.05
Sep 95	0	0.00
Oct 95	0.42	0.03
Nov 95	1.39	0.11
Dec 95	1.07	0.08
Jan 96	2.05	NA
Feb 96	4.3	NA
Mar 96	0	NA
Apr 96	0	NA

B.3.17 Trape et al. (2011): Dielmo, Senegal

Data

TABLE B.35: *Pf*PR-incidence records from (*177*)

Data extracted from Battle et al. (*140*). LAR = Lower bound of the incidence age range, UAR = Upper bound of the incidence age range, INC 1000 PYO^{-1} = incidence (number of events) per 1,000 person-years observed.

Study	Year	LAR	UAR	N	PYO	INC 1000 PYO^{-1}	*Pf*PR
Trape 2011	2007	0	4	NA	104.4493	1072.29	0.1
Trape 2011	2007	5	9	NA	93.5233	1390.03	0.16
Trape 2011	2007	10	14	NA	81.5589	662.1	0.39
Trape 2011	2007	15	29	NA	113.1233	291.72	0.31
Trape 2011	2007	30	44	NA	64.7534	324.31	0.09
Trape 2011	2007	45	85	NA	89.7616	89.12	0.11
Trape 2011	2008	0	4	NA	146.1041	20.53	0.02
Trape 2011	2008	5	9	NA	139.2411	107.73	0.04
Trape 2011	2008	10	14	NA	114.2575	87.52	0.14
Trape 2011	2008	15	29	NA	190.7726	41.93	0.1
Trape 2011	2008	30	44	NA	92.663	32.38	0.04
Trape 2011	2008	45	85	NA	136.9041	7.3	0
Trape 2011	2010	0	4	NA	20.4329	636.23	0
Trape 2011	2010	5	9	NA	18.0164	721.56	0.05
Trape 2011	2010	10	14	NA	16.1863	803.15	0.18
Trape 2011	2010	15	29	NA	26.4301	529.7	0.1
Trape 2011	2010	30	44	NA	16.7342	537.82	0.05
Trape 2011	2010	45	85	NA	22.4548	178.14	0.04

Seasonality

TABLE B.36: Seasonality profile for Dielmo, Senegal

Inferred from Trape et al. (*177*).

Month	Relative transmission
Jan 07	0.064
Feb 07	0.075
Mar 07	0.039
Apr 07	0.059
May 07	0.270
Jun 07	0.057
Jul 07	0.139
Aug 07	0.129
Sep 07	0.081
Oct 07	0.030
Nov 07	0.020
Dec 07	0.037

C Supplement to: Insights into data needs for calibration and implications for predictions

C.1 Convergence

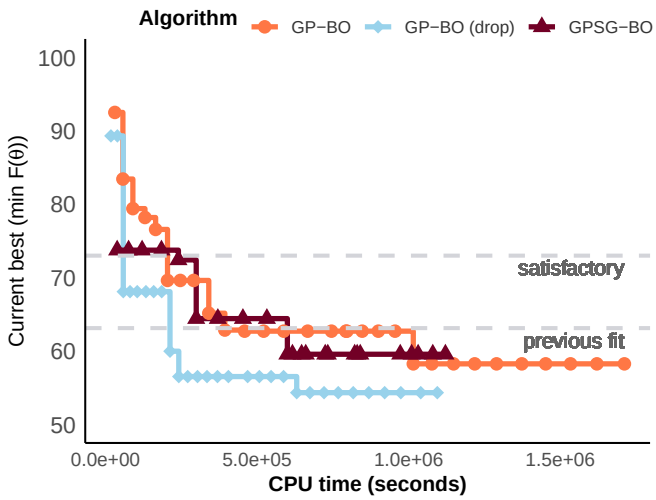

FIGURE C.1: Convergence plot

Weighted sum of loss functions over 11 objectives associated with the current best fit parameter set by CPU time in seconds. Satisfactory fit of *OpenMalaria* refers to a weighted sum of loss functions value of 73.2 (as defined by Smith 2012 (*64*)). The previous best fit for *OpenMalaria* was achieved by the genetic algorithm had a loss function value of 63.7. GP-BO_{drop} reaches its best fit of 54.3 in iteration 20 in $6.3e^5$ CPU seconds (~7 days) compared to 58.2 for GP-BO in iteration 21 in $1.02e^6$ CPU seconds (~12 days) and 59.6 for GPSG-BO in iteration 10 in $6.00e^5$ CPU seconds (~7 days). GP-BO_{drop} = Gaussian process emulator Bayesian optimisation excluding contentious severe disease data point, GP-BO = Gaussian process emulator Bayesian optimisation, GPSG-BO = Gaussian process stacked optimisation emulator Bayesian optimisation.

C.2 Final simulator (*OpenMalaria*) fit including GP-BO$_{\text{drop}}$

C.2.1 Objective 1: Age pattern of prevalence in Matsari, Nigeria during the intervention

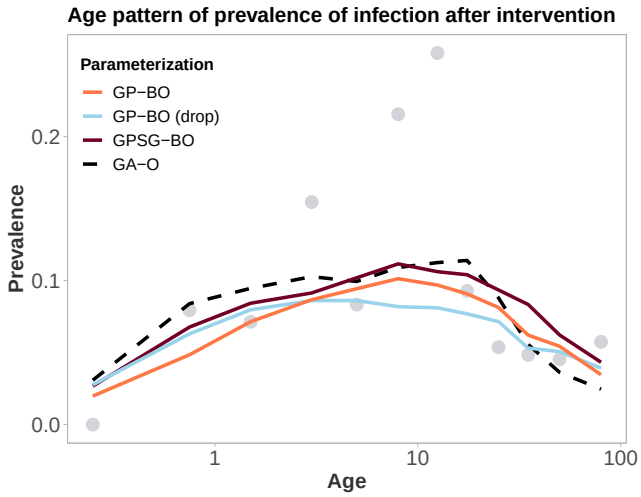

FIGURE C.2: Objective 1: Age pattern of prevalence in Matsari, Nigeria during the intervention

Final simulator fit using the parameter sets yielded using GP-BO$_{\text{drop}}$, GP-BO and GPSG-BO compared to the previous parameteriation (derived using optimization with a genetic algorithm, GA-O).

C.2.2 Objective 2: Age pattern of prevalence

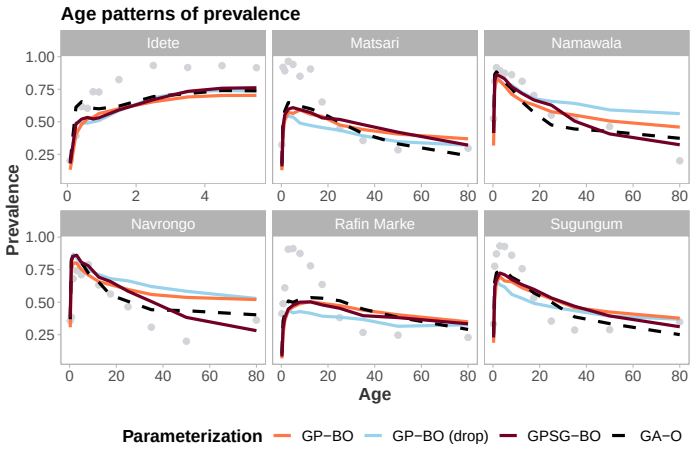

FIGURE C.3: Objective 2: Age pattern of prevalence

Final simulator fit using the parameter sets yielded using GP-BO$_{drop}$, GP-BO and GPSG-BO compared to the previous parameterization (derived using optimization with a genetic algorithm, GA-O).

C.2.3 Objective 3: Age pattern of parasite densities

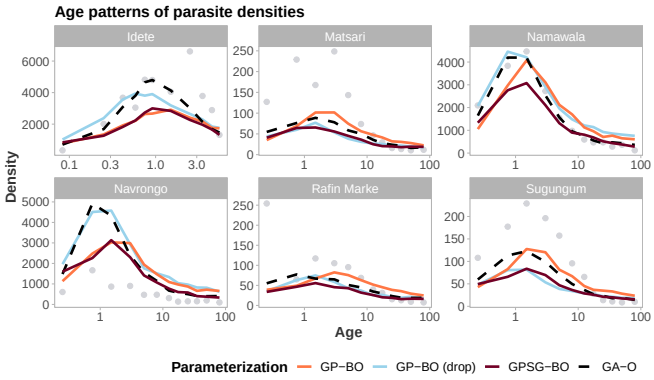

FIGURE C.4: Objective 3: Age pattern of parasite densities (geometric mean)

Final simulator fit using the parameter sets yielded using GP-BO$_{drop}$, GP-BO and GPSG-BO compared to the previous parameterization (derived using optimization with a genetic algorithm, GA-O).

C.2.4 Objective 4: Age pattern of number of concurrent infections

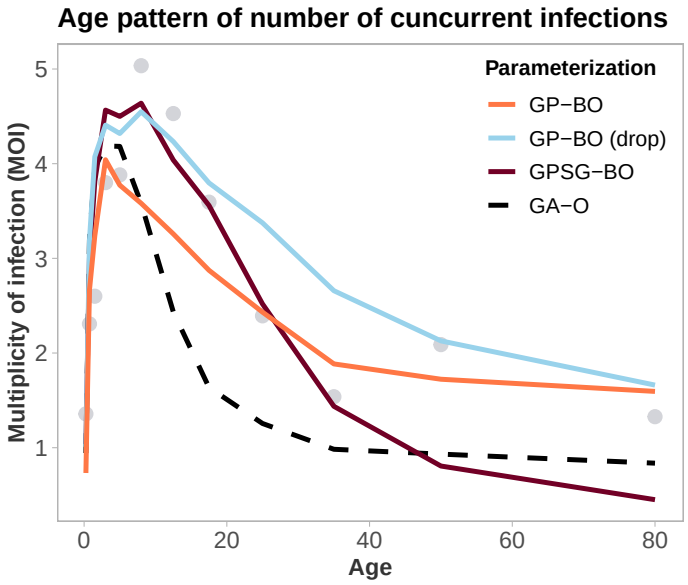

FIGURE C.5: Objective 4: Age pattern of number of concurrent infections

Final simulator fit using the parameter sets yielded using GP-BO_{drop}, GP-BO and GPSG-BO compared to the previous parameterization (derived using optimization with a genetic algorithm, GA-O).

C.2.5 Objective 5: Age pattern of incidence of clinical malaria in Dielmo and Ndiop, Senegal

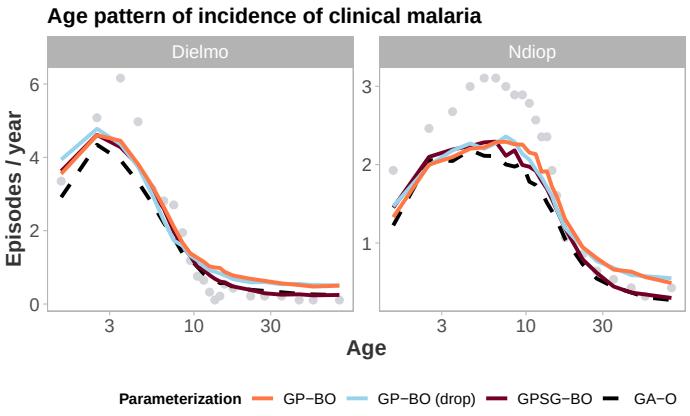

FIGURE C.6: Objective 5: Age pattern of incidence of clinical malaria in Dielmo and Ndiop, Senegal

Final simulator fit using the parameter sets yielded using GP-BO_{drop}, GP-BO and GPSG-BO compared to the previous parameterization (derived using optimization with a genetic algorithm, GA-O).

C.2.6 Objective 6: Age pattern of incidence of clinical malaria in Idete, Tanzania

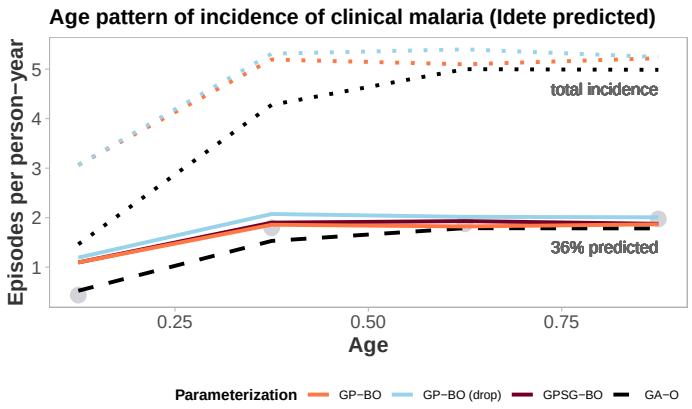

FIGURE C.7: Objective 6: Age pattern of incidence of clinical malaria in Idete, Tanzania

Final simulator fit using the parameter sets yielded using GP-BO$_{drop}$, GP-BO and GPSG-BO compared to the previous parameterization (derived using optimization with a genetic algorithm, GA-O).

C.2.7 Objective 7: Age pattern of threshold parasite density for clinical attacks

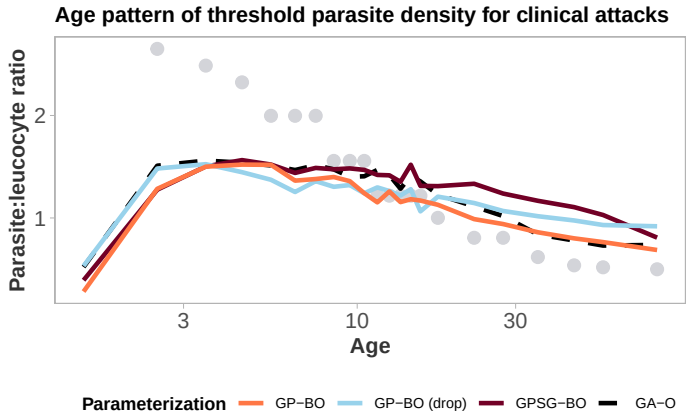

FIGURE C.8: Objective 7: Age pattern of threshold parasite density for clinical attacks

Final simulator fit using the parameter sets yielded using GP-BO$_{drop}$, GP-BO and GPSG-BO compared to the previous parameterization (derived using optimization with a genetic algorithm, GA-O).

C.2.8 Objective 8: Hospitalisation rate in relation to prevalence in children

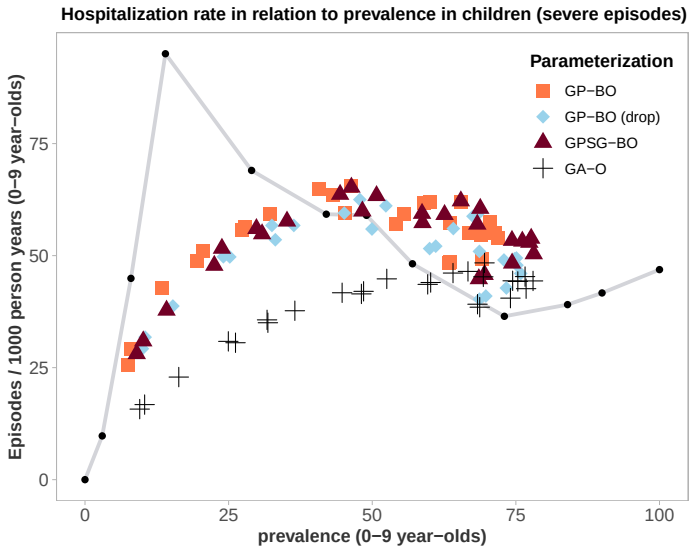

FIGURE C.9: Objective 8: Hospitalization rate in relation to prevalence in children

Final simulator fit using the parameter sets yielded using GP-BO$_{\text{drop}}$, GP-BO and GPSG-BO compared to the previous parameterization (derived using optimization with a genetic algorithm, GA-O).

C.2.9 Objective 9: Age pattern of hospitalisation

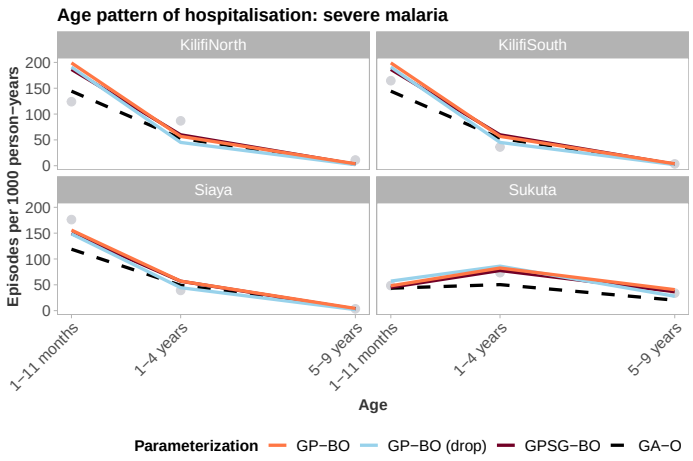

FIGURE C.10: Objective 9: Age pattern of hospitalization

Final simulator fit using the parameter sets yielded using GP-BO$_{\text{drop}}$, GP-BO and GPSG-BO compared to the previous parameterization (derived using optimization with a genetic algorithm, GA-O).

C.2.10 Objective 10: Direct mortality in children <5 years old

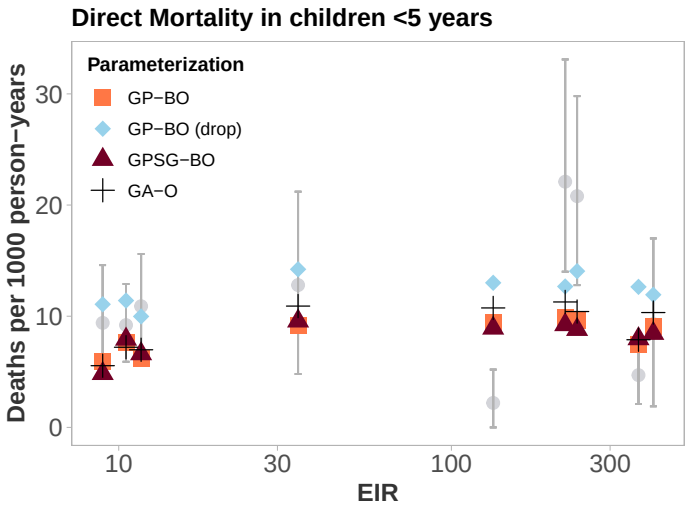

FIGURE C.11: Objective 10: Direct mortality in children <5 years old

Final simulator fit using the parameter sets yielded using GP-BO$_{drop}$, GP-BO and GPSG-BO compared to the previous parameterization (derived using optimization with a genetic algorithm, GA-O).

C.2.11 Objective 11: All-cause infant mortality rate

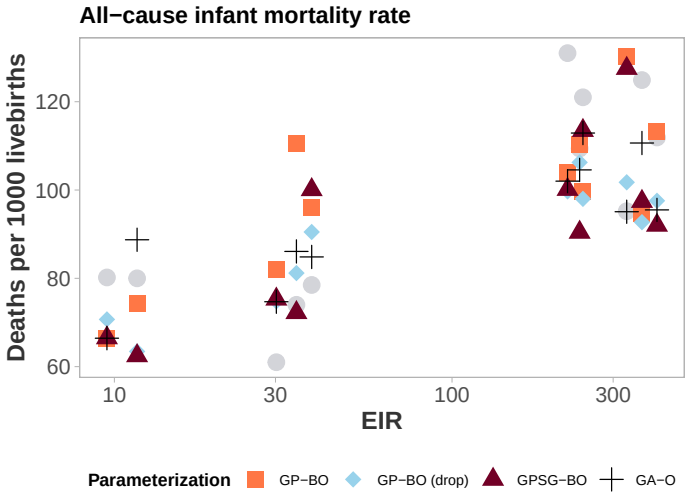

FIGURE C.12: Objective 11: All-cause infant mortality rate

Final simulator fit using the parameter sets yielded using GP-BO$_{drop}$, GP-BO and GPSG-BO compared to the previous parameterization (derived using optimization with a genetic algorithm, GA-O).

D Supplement to: Modelling reactive case detection strategies for interrupting transmisson of *Plasmodium faciparum* malaria

D.1 E_{14} to E_5 conversion

E_5 (OM input, in %)	E_{14}, adjusted for treatment failure	Reported E_{14}, in %
4.0	11.2	13.9
6.5 (REF)	17.6	21.8
8.0	21.1	26.0
12.0	29.7	36.7
16.0	37.3	46.1
20.0	44.1	54.6
24.0	50.2	62.1
28.0	55.7	68.9

Conversion from E_5 OM input to E_{14} Effective care levels can be described by the following equation, derived by fitting a 4^{th} degree polynomial regression.

$$f(x) = -3.831e-06x^4 + 6.143e-04x^3 - 4.971e-02x^2 + 2.982x + 7.723e-02$$

where x is the *OpenMalaria* input (pSeekOfficialCareUncomplicated). We found that a 4^{th} degree polynomial regression provided a sufficiently good fit (adjusted R^2 = 0.9995) whilst keeping the degree low compared to polynomial regressions of degrees 1 to 9. Note that due to the maximum values of the data the regression was fitted to, it is only accurate up to approximately 80% E_{14}

D.2 Population attributable risk

As calculated in the EpiR package, reference: Statistics for Epidemiology by N. P. Jewell (*288*, *289*).

Suppose the following 2 x 2 contingency table:

	O+	O-	
E+	a	b	(a+b)
E-	c	d	(c+d)
	(a+c)	(b+d)	n

where O is the outcome, (here: interruption of transmission) and E is the exposure (here: RCD).

The population attributable risk is given by

$$\widehat{AR} = \frac{\frac{a+c}{n} - \frac{c}{c+d}}{\frac{a+c}{n}}$$
$$= \frac{ad - bc}{(a+c)(c+d)}$$

D.3 EIR and case management to prevalence $PfPR_{0\text{-}99}$ (mean)

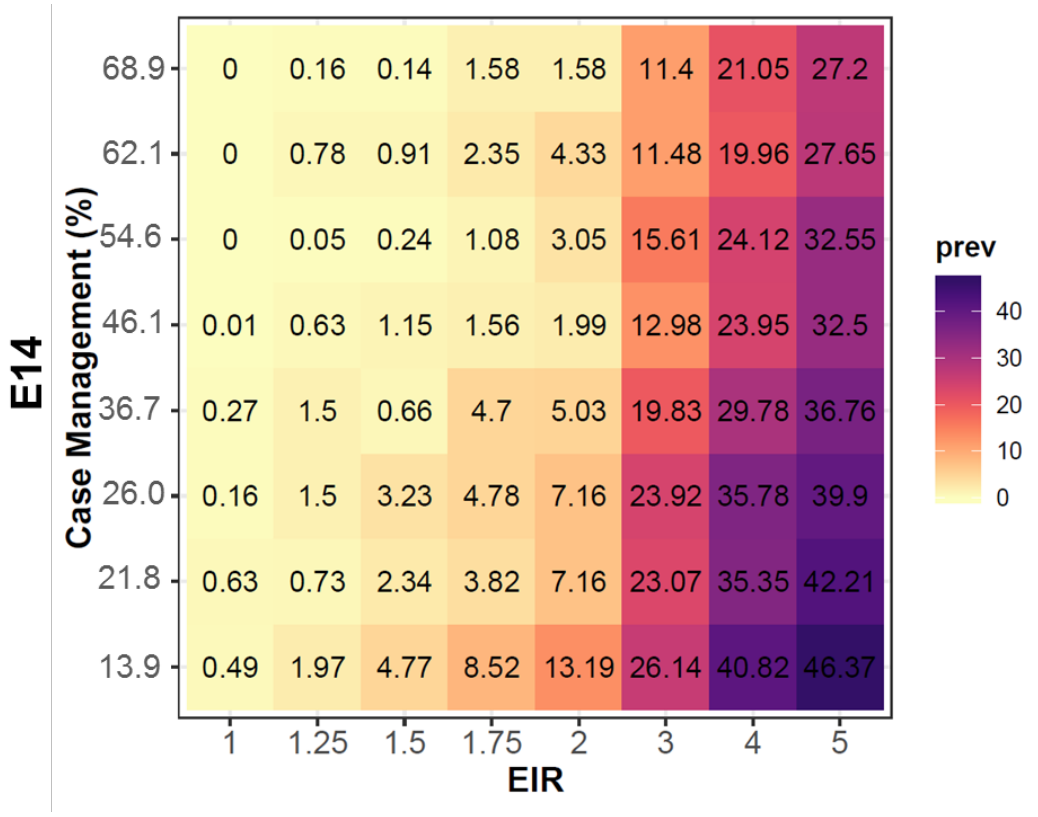

FIGURE D.1: EIR and case management to prevalence $PfPR_{0\text{-}99}$ (mean)

Mean all-age prevalence across five simulations with a given transmission intensity (EIR) and case management level. Case management level is indicated as the probability that a clinical case of malaria receives effective treatment within 14 days (E_{14}).

D.4 Prevalence

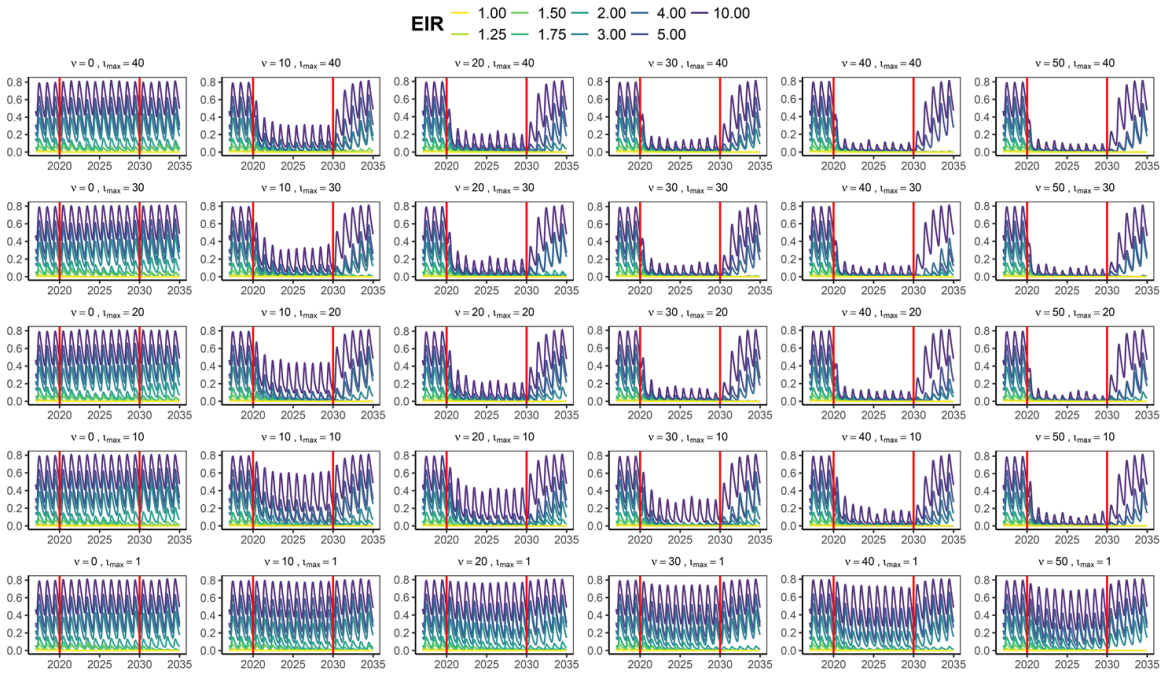

FIGURE D.2: Simulated all-age prevalence for different EIRs and implementations of RCD

Each subplot shows the simulated all-age prevalence over time in one simulation with a given implementation of RCD. The simulated period spans from 2017 to 2035 with RCD being implemented between 2020 and 2030. ι_{max} and ν detail the implementation strategy of RCD, where ν is the number of individuals in the search radius and ι_{max} is the maximum number of index cases that can be followed up per 5-day time step. The different colours denote different transmission intensities. Simulations were carried out at reference case management level.

Bibliography

(*1*) Organization, W. H. et al. Infections and infectious diseases: a manual for nurses and midwives in the WHO European Region, tech. rep., Copenhagen: WHO Regional Office for Europe, 2001.

(*2*) Egger, M., Johnson, L., Althaus, C., Schöni, A., Salanti, G., Low, N., and Norris, S. L. (2017). Developing WHO Guidelines: Time to Formally Include Evidence from Mathematical Modelling Studies. *F1000Research 6*.

(*3*) Smith, T., Ross, A., Maire, N., Rogier, C., Trape, J.-F., and Molineaux, L. (2006). An Epidemiologic Model of the Incidence of Acute Illness in Plasmodium Falciparum Malaria. *The American Journal of Tropical Medicine and Hygiene 75*, 56–62.

(*4*) Griffin, J. T., Hollingsworth, T. D., Okell, L. C., Churcher, T. S., White, M., Hinsley, W., Bousema, T., Drakeley, C. J., Ferguson, N. M., and Basáñez, M.-G. (2010). Reducing Plasmodium Falciparum Malaria Transmission in Africa: A Model-Based Evaluation of Intervention Strategies. *PLOS Med 7*, e1000324.

(*5*) Eckhoff, P. A. (2011). A Malaria Transmission-Directed Model of Mosquito Life Cycle and Ecology. *Malaria journal 10*, 1–17.

(*6*) Cohen, T., Colijn, C., Finklea, B., Wright, A., Zignol, M., Pym, A., and Murray, M. (2008). Are Survey-Based Estimates of the Burden of Drug Resistant TB Too Low? Insight from a Simulation Study. *PLOS ONE 3*, e2363.

(*7*) Halloran, M. E. et al. (2008). Modeling Targeted Layered Containment of an Influenza Pandemic in the United States. *Proceedings of the National Academy of Sciences of the United States of America 105*, 4639–44.

(*8*) Karl, S., Halder, N., Kelso, J. K., Ritchie, S. A., and Milne, G. J. (2014). A Spatial Simulation Model for Dengue Virus Infection in Urban Areas. *BMC Infectious Diseases 14*, 447.

(*9*) Smith, D. L., Battle, K. E., Hay, S. I., Barker, C. M., Scott, T. W., and McKenzie, F. E. (2012). Ross, Macdonald, and a Theory for the Dynamics and Control of Mosquito-Transmitted Pathogens. *PLOS Pathogens 8*, e1002588.

(*10*) Willem, L., Verelst, F., Bilcke, J., Hens, N., and Beutels, P. (2017). Lessons from a Decade of Individual-Based Models for Infectious Disease Transmission: A Systematic Review (2006-2015). *BMC Infectious Diseases 17*, 612.

(*11*) Penny, M. A. et al. (2016). Public Health Impact and Cost-Effectiveness of the RTS,S/AS01 Malaria Vaccine: A Systematic Comparison of Predictions from Four Mathematical Models. *The Lancet 387*, 367–375.

(*12*) Runge, M., Snow, R. W., Molteni, F., Thawer, S., Mohamed, A., Mandike, R., Giorgi, E., Macharia, P. M., Smith, T. A., and Lengeler, C. (2020). Simulating the Council-Specific Impact of Anti-Malaria Interventions: A Tool to Support Malaria Strategic Planning in Tanzania. *PLOS ONE 15*, e0228469.

(*13*) Gomes, M. F., y Piontti, A. P., Rossi, L., Chao, D., Longini, I., Halloran, M. E., and Vespignani, A. (2014). Assessing the International Spreading Risk Associated with the 2014 West African Ebola Outbreak. *PLOS Currents 6*.

(*14*) Chinazzi, M., Davis, J. T., Ajelli, M., Gioannini, C., Litvinova, M., Merler, S., y Piontti, A. P., Mu, K., Rossi, L., and Sun, K. (2020). The Effect of Travel Restrictions on the Spread of the 2019 Novel Coronavirus (COVID-19) Outbreak. *Science 368*, 395–400.

(*15*) Prem, K., Liu, Y., Russell, T. W., Kucharski, A. J., Eggo, R. M., Davies, N., Flasche, S., Clifford, S., Pearson, C. A., and Munday, J. D. (2020). The Effect of Control Strategies to Reduce Social Mixing on Outcomes of the COVID-19 Epidemic in Wuhan, China: A Modelling Study. *The Lancet Public Health 5*, e261–e270.

(*16*) Gilbert, M., Pullano, G., Pinotti, F., Valdano, E., Poletto, C., Boëlle, P.-Y., d'Ortenzio, E., Yazdanpanah, Y., Eholie, S. P., and Altmann, M. (2020). Preparedness and Vulnerability of African Countries against Importations of COVID-19: A Modelling Study. *The Lancet 395*, 871–877.

(*17*) Ferguson, N., Laydon, D., Nedjati-Gilani, G., Imai, N., Ainslie, K., Baguelin, M., Bhatia, S., Boonyasiri, A., Cucunubá, Z., Cuomo-Dannenburg, G., et al. (2020). Report 9: Impact of Non-Pharmaceutical Interventions (NPIs) to Reduce COVID19 Mortality and Healthcare Demand. *Imperial College London 10*, 491–497.

(*18*) O'Grady, C. The UK backed off on herd immunity. To beat COVID-19, we'll ultimately need it *National Geographic* [Online], Published March 20, 2020, `https://www.nationalgeographic.com/science/2020/03/uk-backed-off-on-herd-immunity-to-beat-coronavirus-we-need-it/`.

(*19*) Sarah Boseley New Data, New Policy: Why UK's Coronavirus Strategy Changed *The Guardian* [Online], Published March 16, 2020, `https://www.theguardian.com/world/2020/mar/16/new-data-new-policy-why-uks-coronavirus-strategy-has-changed`.

(*20*) Smith, T. A., Chitnis, N., Penny, M., and Tanner, M. (2017). Malaria Modeling in the Era of Eradication. *Cold Spring Harbor Perspectives in Medicine 7*, a025460.

(*21*) World Health Organization, *World Malaria Report 2017*; World Health Organizaion: Geneva, 2017.

(*22*) World Health Organizaion, *World Malaria Report 2019*; World Health Organization: Geneva, 2019.

(*23*) Snow, R. W. (2015). Global Malaria Eradication and the Importance of Plasmodium Falciparum Epidemiology in Africa. *BMC Medicine 13*, 1–3.

(*24*) Mo, A. X., and McGugan, G. (2018). Understanding the Liver-Stage Biology of Malaria Parasites: Insights to Enable and Accelerate the Development of a Highly Efficacious Vaccine. *The American Journal of Tropical Medicine and Hygiene 99*, 827–832.

(*25*) Sullivan, D., and Krishna, S., *Malaria: Drugs, Disease and Post-Genomic Biology*; Springer Science & Business Media: 2006; Vol. 295.

(*26*) Reece, J. B., Urry, L. A., Cain, M. L., Wasserman, S. A., Minorsky, P. V., and Jackson, R. B., *Campbell Biology*; Pearson Boston: 2014.

(*27*) Votýpka J., Modrý D., Oborník M., Šlapeta J., and Lukeš J. In *Handbook of the Protists*; Springer International Publishing AG: 2017.

(*28*) Kurth, F., Develoux, M., Mechain, M., Malvy, D., Clerinx, J., Antinori, S., Gjørup, I. E., Gascon, J., Mørch, K., Nicastri, E., et al. (2017). Severe Malaria in Europe: An 8-Year Multi-Centre Observational Study. *Malaria Journal 16*, 1–11.

(*29*) Conroy, A. L., Datta, D., and John, C. C. (2019). What Causes Severe Malaria and Its Complications in Children? Lessons Learned over the Past 15 Years. *BMC Medicine 17*, 1–4.

(*30*) Doolan, D. L., Dobaño, C., and Baird, J. K. (2009). Acquired Immunity to Malaria. *Clinical Microbiology Reviews 22*, 13–36.

(*31*) Maire, N., Smith, T., Ross, A., Owusu-Agyei, S., Dietz, K., and Molineaux, L. (2006). A Model for Natural Immunity to Asexual Blood Stages of Plasmodium Falciparum Malaria in Endemic Areas. *The American Journal of Tropical Medicine and Hygiene 75*, 19–31.

(*32*) Ariey, F., Gay, F., and Ménard, R., *Malaria Control and Elimination*; Springer: 2019.

(*33*) Bhatt, S., Cameron, E., Flaxman, S. R., Weiss, D. J., Smith, D. L., and Gething, P. W. (2017). Improved Prediction Accuracy for Disease Risk Mapping Using Gaussian Process Stacked Generalization. *Journal of The Royal Society Interface 14*, 20170520.

(*34*) World Health Organization, *Global Technical Strategy for Malaria 2016-2030*; World Health Organization: Geneva, 2015.

(*35*) Moonen, B., Cohen, J. M., Snow, R. W., Slutsker, L., Drakeley, C., Smith, D. L., Abeyasinghe, R. R., Rodriguez, M. H., Maharaj, R., and Tanner, M. (2010). Operational Strategies to Achieve and Maintain Malaria Elimination. *The Lancet 376*, 1592–1603.

(*36*) World Health Organization Training Module on Malaria Elimination: Malaria Elimination Guide for Participants, tech. rep., 2016.

(*37*) World Health Organization, *A Framework for Malaria Elimination*; World Health Organizaion: Geneva, 2017.

(*38*) Perera, R., Caldera, A., and Wickremasinghe, A. R. (2020). Reactive Case Detection (RACD) and Foci Investigation Strategies in Malaria Control and Elimination: A Review. *Malaria Journal 19*, 1–11.

(*39*) Sturrock, H. J., Hsiang, M. S., Cohen, J. M., Smith, D. L., Greenhouse, B., Bousema, T., and Gosling, R. D. (2013). Targeting Asymptomatic Malaria Infections: Active Surveillance in Control and Elimination. *PLOS Medicine 10*, e1001467.

(*40*) Hustedt, J., Canavati, S. E., Rang, C., Ashton, R. A., Khim, N., Berne, L., Kim, S., Sovannaroth, S., Ly, P., and Ménard, D. (2016). Reactive Case-Detection of Malaria in Pailin Province, Western Cambodia: Lessons from a Year-Long Evaluation in a Pre-Elimination Setting. *Malaria Journal 15*, 1–10.

(*41*) Gueye, C. S., Sanders, K. C., Galappaththy, G. N., Rundi, C., Tobgay, T., Sovannaroth, S., Gao, Q., Surya, A., Thakur, G. D., and Baquilod, M. (2013). Active Case Detection for Malaria Elimination: A Survey among Asia Pacific Countries. *Malaria Journal 12*, 1–9.

(*42*) Gerardin, J., Bever, C. A., Bridenbecker, D., Hamainza, B., Silumbe, K., Miller, J. M., Eisele, T. P., Eckhoff, P. A., and Wenger, E. A. (2017). Effectiveness of Reactive Case Detection for Malaria Elimination in Three Archetypical Transmission Settings: A Modelling Study. *Malaria Journal 16*, 1–17.

(*43*) Ross, R., *The Prevention of Malaria*; John Murray: 1911.

(*44*) Mandal, S., Sarkar, R. R., and Sinha, S. (2011). Mathematical Models of Malaria-a Review. *Malaria Journal 10*, 1–19.

(*45*) Macdonald, G. (1950). The Analysis of Infection Rates in Diseases in Which Super Infection Occurs. *Tropical Diseases Bulletin 47*, 907–915.

(*46*) Macdonald, G. (1952). The Analysis of the Sporozoite Rate. *Tropical Diseases Bulletin 49*.

(*47*) MacDonald, G., Cuellar, C. B., and Foll, C. V. (1968). The Dynamics of Malaria. *Bulletin of the World Health Organization 38*, 743.

(*48*) Macdonald, G. (1957). The Epidemiology and Control of Malaria. *The Epidemiology and Control of Malaria.*

(*49*) Chitnis, N., Hyman, J. M., and Cushing, J. M. (2008). Determining Important Parameters in the Spread of Malaria through the Sensitivity Analysis of a Mathematical Model. *Bulletin of Mathematical Biology 70*, 1272–1296.

(*50*) Chitnis, N., Pemberton-Ross, P., Yukich, J., Hamainza, B., Miller, J., Reiker, T., Eisele, T. P., and Smith, T. A. (2019). Theory of Reactive Interventions in the Elimination and Control of Malaria. *Malaria Journal 18*, 1–13.

(*51*) DeAngelis, D. L., and Grimm, V. (2014). Individual-Based Models in Ecology after Four Decades. *F1000prime reports 6*.

(*52*) Smith, T, Maire, N, Ross, A, Penny, M, Chitnis, N, Schapira, A, Studer, A, Genton, B, Lengeler, C, and Tediosi, F. (2008). Towards a Comprehensive Simulation Model of Malaria Epidemiology and Control. *Parasitology 135*, 1507–1516.

(*53*) Cameron, E., Battle, K. E., Bhatt, S., Weiss, D. J., Bisanzio, D., Mappin, B., Dalrymple, U., Hay, S. I., Smith, D. L., Griffin, J. T., Wenger, E. A., Eckhoff, P. A., Smith, T. A., Penny, M. A., and Gething, P. W. (2015). Defining the Relationship between Infection Prevalence and Clinical Incidence of Plasmodium Falciparum Malaria. *Nature Communications 6*, 1–10.

(*54*) Chitnis, N., Hardy, D., and Smith, T. (2012). A Periodically-Forced Mathematical Model for the Seasonal Dynamics of Malaria in Mosquitoes. *Bulletin of Mathematical Biology 74*, 1098–124.

(*55*) Eckhoff, P. A. (2012). Malaria Parasite Diversity and Transmission Intensity Affect Development of Parasitological Immunity in a Mathematical Model. *Malaria Journal 11*, 1–14.

(*56*) Slater, H. C., Walker, P. G., Bousema, T., Okell, L. C., and Ghani, A. C. (2014). The Potential Impact of Adding Ivermectin to a Mass Treatment Intervention to Reduce Malaria Transmission: A Modelling Study. *J Infect Dis 210*, 1972–80.

(*57*) Winskill, P., Walker, P. G., Griffin, J. T., and Ghani, A. C. (2017). Modelling the Cost-Effectiveness of Introducing the RTS,S Malaria Vaccine Relative to Scaling up Other Malaria Interventions in Sub-Saharan Africa. *BMJ Global Health 2*, e000090.

(*58*) Nguyen, T. D., Olliaro, P., Dondorp, A. M., Baird, J. K., Lam, H. M., Farrar, J., Thwaites, G. E., White, N. J., and Boni, M. F. (2015). Optimum Population-Level Use of Artemisinin Combination Therapies: A Modelling Study. *Lancet Glob Health 3*, e758–66.

(*59*) Brady, O. J. et al. (2017). Role of Mass Drug Administration in Elimination of Plasmodium Falciparum Malaria: A Consensus Modelling Study. *The Lancet Global Health 5*, e680–e687.

(*60*) Okell, L, Slater, H., Ghani, A., Pemberton-Ross, P, Smith, T. A., and Chitnis, N. In *Malaria Policy Advisory Committee Meeting*, 2015, pp 16–18.

(*61*) Penny, M. A., Maire, N., Bever, C. A., Pemberton-Ross, P., Briët, O. J., Smith, D. L., Gething, P. W., and Smith, T. A. (2015). Distribution of Malaria Exposure in Endemic Countries in Africa Considering Country Levels of Effective Treatment. *Malaria Journal 14*, 384.

(*62*) World Health Organization (2016). Malaria Vaccine: WHO Position Paper–January 2016. *Weekly Epidemiological Record= Relevé épidémiologique hebdomadaire 91*, 33–52.

(*63*) Smith, T., Killeen, G. F., Maire, N., Ross, A., Molineaux, L., Tediosi, F., Hutton, G., Utzinger, J., Dietz, K., and Tanner, M. (2006). Mathematical Modeling of the Impact of Malaria Vaccines on the Clinical Epidemiology and Natural History of Plasmodium Falciparum Malaria: Overview. *The American Journal of Tropical Medicine and Hygiene 75*, 1–10.

(*64*) Smith, T., Ross, A., Maire, N., Chitnis, N., Studer, A., Hardy, D., Brooks, A., Penny, M., and Tanner, M. (2012). Ensemble Modeling of the Likely Public Health Impact of a Pre-Erythrocytic Malaria Vaccine. *PLOS Medicine 9*, e1001157.

(*65*) Molineaux, L., Gramiccia, G., and Organization, W. H., *The Garki Project: Research on the Epidemiology and Control of Malaria in the Sudan Savanna of West Africa*; World Health Organization: Geneva, 1980.

(*66*) Forrester, A., Sobester, A., and Keane, A., *Engineering Design via Surrogate Modelling: A Practical Guide*; John Wiley & Sons: 2008.

(*67*) Vernon, I., Jackson, S. E., and Cumming, J. A. (2019). Known Boundary Emulation of Complex Computer Models. *SIAM/ASA Journal on Uncertainty Quantification 7*, 838–876.

(*68*) Asher, M. J., Croke, B. F., Jakeman, A. J., and Peeters, L. J. (2015). A Review of Surrogate Models and Their Application to Groundwater Modeling. *Water Resources Research 51*, 5957–5973.

(*69*) Hussain, M. F., Barton, R. R., and Joshi, S. B. (2002). Metamodeling: Radial Basis Functions, versus Polynomials. *European Journal of Operational Research 138*, 142–154.

(*70*) Stone, N. Gaussian Process Emulators for Uncertainty Analysis in Groundwater Flow, Ph.D. Thesis, University of Nottingham, 2011.

(*71*) Kennedy, M. C., and O'Hagan, A. (2001). Bayesian Calibration of Computer Models. *Journal of the Royal Statistical Society: Series B (Statistical Methodology) 63*, 425–464.

(*72*) Yoon, H., Jun, S.-C., Hyun, Y., Bae, G.-O., and Lee, K.-K. (2011). A Comparative Study of Artificial Neural Networks and Support Vector Machines for Predicting Groundwater Levels in a Coastal Aquifer. *Journal of Hydrology 396*, 128–138.

(*73*) Kourakos, G., and Mantoglou, A. (2009). Pumping Optimization of Coastal Aquifers Based on Evolutionary Algorithms and Surrogate Modular Neural Network Models. *Advances in Water Resources 32*, 507–521.

(*74*) Breiman, L. (2001). Random Forests. *Machine Learning 45*, 5–32.

(*75*) Hogue, A. Lecture 12: Support Vector Machines and Singular Value Decomposition MIT, Math 18.337, Computer Science 6.338, SMA 5505, `http://courses.csail.mit.edu/18.337/2004/book/Lectures_2004.pdf`.

(*76*) Bourinet, J.-M. (2016). Rare-Event Probability Estimation with Adaptive Support Vector Regression Surrogates. *Reliability Engineering & System Safety 150*, 210–221.

(*77*) Vapnik, V., Golowich, S. E., and Smola, A. J. In *Advances in Neural Information Processing Systems*, 1997, pp 281–287.

(*78*) Chang, E. Y. In *Foundations of Large-Scale Multimedia Information Management and Retrieval*; Springer: 2011, pp 213–230.

(*79*) Zeng, Z.-Q., Yu, H.-B., Xu, H.-R., Xie, Y.-Q., and Gao, J. In *2008 3rd International Conference on Intelligent System and Knowledge Engineering*, IEEE: 2008; Vol. 1, pp 997–1001.

(*80*) Forrester, A. I., and Keane, A. J. (2009). Recent Advances in Surrogate-Based Optimization. *Progress in Aerospace Sciences 45*, 50–79.

(*81*) Lippmann, R. (1987). An Introduction to Computing with Neural Nets. *IEEE Assp Magazine 4*, 4–22.

(*82*) Lapedes, A., and Farber, R. In *Evolution, Learning and Cognition*; World Scientific: 1988, pp 331–346.

(*83*) Nelson, M. M., and Illingworth, W. T. (1991). A Practical Guide to Neural Nets.

(*84*) Perkins, T. A., Jr, R. C. R., España, G., ten Bosch, Q. A., Verma, A., Liebman, K. A., Paz-Soldan, V. A., Elder, J. P., Morrison, A. C., Stoddard, S. T., Kitron, U., Vazquez-Prokopec, G. M., Scott, T. W., and Smith, D. L. (2019). An Agent-Based Model of Dengue Virus Transmission Shows How Uncertainty about Breakthrough Infections Influences Vaccination Impact Projections. *PLOS Computational Biology 15*, e1006710.

(*85*) Bellman, R. E., *Dynamic Programming*, Sixth; Princeton University Press: 1957.

(*86*) Craig, A. Astronomers Count the Stars BBC News, `http://news.bbc.co.uk/2/hi/science/nature/3085885.stm`.

(*87*) Goldberg, D. E. (1989). Genetic Algorithms in Search. *Optimization and Machine Learning*.

(*88*) Oliveto, P. S., Paixão, T., Pérez Heredia, J., Sudholt, D., and Trubenová, B. In *Proceedings of the Genetic and Evolutionary Computation Conference 2016*, 2016, pp 1163–1170.

(*89*) Eckhoff, P. (2012). P. Falciparum Infection Durations and Infectiousness Are Shaped by Antigenic Variation and Innate and Adaptive Host Immunity in a Mathematical Model. *PLOS ONE 7*, e44950.

(*90*) Eckhoff, P. (2013). Mathematical Models of Within-Host and Transmission Dynamics to Determine Effects of Malaria Interventions in a Variety of Transmission Settings. *The American Journal of Tropical Medicine and Hygiene 88*, 817–827.

(*91*) Griffin, J. T., Ferguson, N. M., and Ghani, A. C. (2014). Estimates of the Changing Age-Burden of Plasmodium Falciparum Malaria Disease in Sub-Saharan Africa. *Nature Communications 5*, 1–10.

(*92*) Hazelbag, C. M., Dushoff, J., Dominic, E. M., Mthombothi, Z. E., and Delva, W. (2020). Calibration of Individual-Based Models to Epidemiological Data: A Systematic Review. *PLOS Computational Biology 16*, e1007893.

(*93*) Fer, I., Kelly, R., Moorcroft, P. R., Richardson, A. D., Cowdery, E. M., and Dietze, M. C. (2018). Linking Big Models to Big Data: Efficient Ecosystem Model Calibration through Bayesian Model Emulation. *Biogeosciences (Online) 15*, 5801–5830.

(*94*) Mockus, J. In *Bayesian Approach to Global Optimization: Theory and Applications*, Mockus, J., Ed.; Mathematics and Its Applications; Springer Netherlands: Dordrecht, 1989, pp 125–156.

(*95*) Chong, A., and Menberg, K. (2018). Guidelines for the Bayesian Calibration of Building Energy Models. *Energy and Buildings 174*, 527–547.

(*96*) Gramacy, R. B., Bingham, D., Holloway, J. P., Grosskopf, M. J., Kuranz, C. C., Rutter, E., Trantham, M., and Drake, R. P. (2015). Calibrating a Large Computer Experiment Simulating Radiative Shock Hydrodynamics. *The Annals of Applied Statistics 9*, 1141–1168.

(*97*) Snoek, J., Larochelle, H., and Adams, R. P. (2012). Practical Bayesian Optimization of Machine Learning Algorithms. *Advances in Neural Information Processing Systems 25*, 2951–2959.

(*98*) Snoek, J., Rippel, O., Swersky, K., Kiros, R., Satish, N., Sundaram, N., Patwary, M., Prabhat, M., and Adams, R. In *International Conference on Machine Learning*, 2015, pp 2171–2180.

(*99*) Wolpert, D. H. (1992). Stacked Generalization. *Neural Networks 5*, 241–259.

(*100*) Sobol, I. M. (1993). Sensitivity Analysis for Nonlinear Mathematical Models. *Mathematical Modelling and Computational Experiments 1*, 407–414.

(*101*) Benkeser, D., Ju, C., Lendle, S., and van der Laan, M. (2018). Online Cross-validation-based Ensemble Learning. *Statistics in Medicine 37*, 249–260.

(*102*) Breiman, L. (1996). Stacked Regressions. *Machine Learning 24*, 49–64.

(*103*) Sill, J., Takacs, G., Mackey, L., and Lin, D. Feature-Weighted Linear Stacking (2009), `https://arxiv.org/abs/0911.0460`.

(*104*) Van der Laan, M. J., Polley, E. C., and Hubbard, A. E. (2007). Super Learner. *Statistical Applications in Genetics and Molecular Biology 6*.

(*105*) Srinivas, N., Krause, A., Kakade, S. M., and Seeger, M. Gaussian Process Optimization in the Bandit Setting: No Regret and Experimental Design (2009), `http://arxiv.org/abs/0912.3995`.

(*106*) Baker, R. E., Peña, J.-M., Jayamohan, J., and Jérusalem, A. (2018). Mechanistic Models versus Machine Learning, a Fight Worth Fighting for the Biological Community? *Biology Letters 14*, 20170660.

(*107*) Moriconi, R., Deisenroth, M. P., and Kumar, K. S. (2020). High-Dimensional Bayesian Optimization Using Low-Dimensional Feature Spaces. *Machine Learning 109*, 1925–1943.

(*108*) Zhou, D., Li, L., and Gu, Q. In *International Conference on Machine Learning*, PMLR: 2020, pp 11492–11502.

(*109*) Marler, R. T., and Arora, J. S. (2010). The Weighted Sum Method for Multi-Objective Optimization: New Insights. *Structural and Multidisciplinary Optimization 41*, 853–862.

(*110*) Binois, M., Gramacy, R. B., and Ludkovski, M. (2018). Practical Heteroscedastic Gaussian Process Modeling for Large Simulation Experiments. *Journal of Computational and Graphical Statistics 27*, 808–821.

(*111*) Hadji, A., and Szábo, B. Can We Trust Bayesian Uncertainty Quantification from Gaussian Process Priors with Squared Exponential Covariance Kernel? (2019), `https://arxiv.org/abs/1904.01383`.

(*112*) Foresee, F. D., and Hagan, M. T. In *Proceedings of International Conference on Neural Networks (ICNN'97)*, IEEE: 1997; Vol. 3, pp 1930–1935.

(*113*) MacKay, D. J. (1992). Bayesian Interpolation. *Neural Computation 4*, 415–447.

(*114*) Rodriguez, P., and Gianola, D R Package 'brnn' Version 0.6 (2016), `https://cran.r-project.org/web/packages/brnn/brnn.pdf`.

(*115*) Hastie, T., and Qian, J. Glmnet Vignette (2016), `https://hastie.su.domains/Papers/Glmnet_Vignette.pdf`.

(*116*) Liaw, A., Wiener, M., Breiman, L, and Cutler, A R Package: 'randomForest' Version 4.6-14 (2015), `https://cran.r-project.org/web/packages/randomForest/randomForest.pdf`.

(*117*) Reiker, T., Chitnis, N., and Smith, T. (2019). Modelling Reactive Case Detection Strategies for Interrupting Transmission of Plasmodium Falciparum Malaria. *Malaria Journal 18*, 259.

(*118*) Cauwet, M.-L., Couprie, C., Dehos, J., Luc, P., Rapin, J., Riviere, M., Teytaud, F., Teytaud, O., and Usunier, N. In *International Conference on Machine Learning*, PMLR: 2020, pp 1338–1348.

(*119*) Kucherenko, S., Albrecht, D., and Saltelli, A. Exploring Multi-Dimensional Spaces: A Comparison of Latin Hypercube and Quasi Monte Carlo Sampling Techniques (2015), `https://arxiv.org/abs/1505.02350`.

(*120*) Auer, P. (2002). Using Confidence Bounds for Exploitation-Exploration Trade-offs. *Journal of Machine Learning Research 3*, 397–422.

(*121*) Brochu, E., Cora, V. M., and De Freitas, N. A Tutorial on Bayesian Optimization of Expensive Cost Functions, with Application to Active User Modeling and Hierarchical Reinforcement Learning (2010), `https://arxiv.org/abs/1012.2599`.

(*122*) Bischl, B., Lang, M., Kotthoff, L., Schiffner, J., Richter, J., Studerus, E., Casalicchio, G., and Jones, Z. M. (2016). Mlr: Machine Learning in R. *The Journal of Machine Learning Research 17*, 5938–5942.

(*123*) Ripley, B., Venables, W., and Ripley, M. B. R Package 'nnet' Version 7.3-12 (2016), `https://cran.r-project.org/web/packages/nnet/nnet.pdf`.

(*124*) Hofner, B., Mayr, A., Robinzonov, N., and Schmid, M. (2014). Model-Based Boosting in R: A Hands-on Tutorial Using the R Package Mboost. *Computational Statistics 29*, 3–35.

(*125*) Ishwaran, H., Kogalur, U. B., and Kogalur, M. U. B. R Package 'randomForestSRC' Version 2.9.3 (2020), https://cran.r-project.org/web/packages/randomForestSRC/randomForestSRC.pdf.

(*126*) Wright, M. N., and Ziegler, A. Ranger: A Fast Implementation of Random Forests for High Dimensional Data in C++ and R (2015), https://doi.org/10.18637\%2Fjss.v077.i01.

(*127*) Meinshausen, N. R Package 'nodeHarvest' https://cran.r-project.org/web/packages/nodeHarvest.

(*128*) Hastie, T., Tibshirani, R., Leisch, F, Hornik, K, and Ripley, B. R package 'mda' Version 0.4-10 (2017), https://cran.r-project.org/web/packages/mda/mda.pdf.

(*129*) Jansen, M. J. (1999). Analysis of Variance Designs for Model Output. *Computer Physics Communications 117*, 35–43.

(*130*) Saltelli, A., Annoni, P., Azzini, I., Campolongo, F., Ratto, M., and Tarantola, S. (2010). Variance Based Sensitivity Analysis of Model Output. Design and Estimator for the Total Sensitivity Index. *Computer Physics Communications 181*, 259–270.

(*131*) Iooss, B., Da Veiga, S., Janon, A., and Pujol, G. R package 'sensitivity' Version 1.24.0 (2021), https://cran.r-project.org/web/packages/sensitivity/sensitivity.pdf.

(*132*) Bernoulli, D. (1766). Essai d'une Nouvelle Analyse de La Mortalité Cause Par La Petite Vérole et Des Avantages de l'inoculation Pour La Prévenir. Histoire de l'académie Royale Des Sciences Avec Les Mémoires de Mathématique et de Physique Tirés Des Registres de Cette Académie. Paris 1766 (Année 1760). *History of Actuarial Science 8*, 1766.

(*133*) Maude, R. J., Lubell, Y., Socheat, D., Yeung, S., Saralamba, S., Pongtavornpinyo, W., Cooper, B. S., Dondorp, A. M., White, N. J., and White, L. J. (2010). The Role of Mathematical Modelling in Guiding the Science and Economics of Malaria Elimination. *International Health 2*, 239–246.

(*134*) Yan, P., and Chowell, G., *Quantitative Methods for Investigating Infectious Disease Outbreaks*; Springer: 2019; Vol. 70.

(*135*) Den Boon, S., Jit, M., Brisson, M., Medley, G., Beutels, P., White, R., Flasche, S., Hollingsworth, T. D., Garske, T., and Pitzer, V. E. (2019). Guidelines for Multi-Model Comparisons of the Impact of Infectious Disease Interventions. *BMC Medicine 17*, 1–13.

(*136*) Reiker, T., Golumbeanu, M., Shattock, A., Burgert, L., Smith, T. A., Filippi, S., Cameron, E., and Penny, M. A. (2021). Emulator-Based Bayesian Optimization for Efficient Multi-Objective Calibration of an Individual-Based Model of Malaria. *Nature Communications 12*, 1–11.

(*137*) Galactionova, K., Tediosi, F., De Savigny, D., Smith, T., and Tanner, M. (2015). Effective Coverage and Systems Effectiveness for Malaria Case Management in Sub-Saharan African Countries. *PLOS ONE 10*, e0127818.

(*138*) Cibulskis, R. E., Aregawi, M., Williams, R., Otten, M., and Dye, C. (2011). Worldwide Incidence of Malaria in 2009: Estimates, Time Trends, and a Critique of Methods. *PLOS Medicine 8*, e1001142.

(*139*) Weiss, D. J., Lucas, T. C., Nguyen, M., Nandi, A. K., Bisanzio, D., Battle, K. E., Cameron, E., Twohig, K. A., Pfeffer, D. A., and Rozier, J. A. (2019). Mapping the Global Prevalence,

Incidence, and Mortality of Plasmodium Falciparum, 2000–17: A Spatial and Temporal Modelling Study. *The Lancet 394*, 322–331.

(*140*) Battle, K. E., Guerra, C. A., Golding, N., Duda, K. A., Cameron, E., Howes, R. E., Elyazar, I. R., Baird, J. K., Reiner, R. C., and Gething, P. W. (2015). Global Database of Matched Plasmodium Falciparum and P. Vivax Incidence and Prevalence Records from 1985–2013. *Scientific Data 2*, 1–12.

(*141*) Battle, K. E., Cameron, E., Guerra, C. A., Golding, N., Duda, K. A., Howes, R. E., Elyazar, I. R., Price, R. N., Baird, J. K., and Reiner, R. C. (2015). Defining the Relationship between Plasmodium Vivax Parasite Rate and Clinical Disease. *Malaria Journal 14*, 1–14.

(*142*) Patil, A. P., Okiro, E. A., Gething, P. W., Guerra, C. A., Sharma, S. K., Snow, R. W., and Hay, S. I. (2009). Defining the Relationship between Plasmodium Falciparum Parasite Rate and Clinical Disease: Statistical Models for Disease Burden Estimation. *Malaria Journal 8*, 1–11.

(*143*) Ba, F. F. (2000). Le Paludisme En Zone Mésoendémique: Relations Entre La Transmission, l'infection et La Morbidité Palustre à Ndiop (Sénégal).

(*144*) Fontenille, D., Lochouarn, L., Diagne, N., Sokhna, C., Lemasson, J.-J., Diatta, M., Konate, L., Faye, F., Rogier, C., and Trape, J.-F. (1997). High Annual and Seasonal Variations in Malaria Transmission by Anophelines and Vector Species Composition in Dielmo, a Holoendemic Area in Senegal. *The American Journal of Tropical Medicine and Hygiene 56*, 247–253.

(*145*) Bonnet, S., l Paul, R., Gouagna, C., Safeukui, I., Meunier, J.-Y., Gounoue, R., and Boudin, C. (2002). Level and Dynamics of Malaria Transmission and Morbidity in an Equatorial Area of South Cameroon. *Tropical Medicine & International Health 7*, 249–256.

(*146*) Meunier, J.-Y., Safeukui, I., Fontenille, D., and Boudin, C. (1999). Etude de La Transmission Du Paludisme Dans Une Future Zone d'essai Vaccinal En Forêt Équatoriale Du Sud Cameroun. *Bulletin de la Societe de Pathologie Exotique 92*, 309–312.

(*147*) Lemnge, M., Alifrangis, M., Kafuye, M. Y., Gesase, S., Minja, D., Massaga, J. J., Rønn, A. M., and Bygbjerg, I. C. (2006). High Reinfection Rate and Treatment Failures in Children Treated with Amodiaquine for Falciparum Malaria in Muheza Villages, Northeastern Tanzania. *The American Journal of Tropical Medicine and Hygiene 75*, 188–193.

(*148*) Dicko, A., Sagara, I., Diemert, D., Sogoba, M., Niambele, M. B., Dao, A., Dolo, G., Yalcouye, D., Diallo, D. A., and Saul, A. (2007). Year-to-Year Variation in the Age-Specific Incidence of Clinical Malaria in Two Potential Vaccine Testing Sites in Mali with Different Levels of Malaria Transmission Intensity. *The American Journal of Tropical Medicine and Hygiene 77*, 1028–1033.

(*149*) de Radigues, X., Diallo, K. I., Diallo, M., Ngwakum, P. A., Maiga, H., Djimdé, A., Sacko, M., Doumbo, O., and Guthmann, J.-P. (2006). Efficacy of Chloroquine and Sulfadoxine/Pyrimethamine for the Treatment of Uncomplicated Falciparum Malaria in Koumantou, Mali. *Transactions of the Royal Society of Tropical Medicine and Hygiene 100*, 1013–1018.

(*150*) Djimde, A. A., Fofana, B., Sagara, I., Sidibe, B., Toure, S., Dembele, D., Dama, S., Ouologuem, D., Dicko, A., and Doumbo, O. K. (2008). Efficacy, Safety, and Selection of Molecular Markers of Drug Resistance by Two ACTs in Mali. *The American Journal of Tropical Medicine and Hygiene 78*, 455–461.

(*151*) Henry, M.-C., Rogier, C., Nzeyimana, I., Assi, S. B., Dossou-Yovo, J., Audibert, M., Mathonnat, J., Keundjian, A., Akodo, E., and Teuscher, T. (2003). Inland Valley Rice Production Systems and Malaria Infection and Disease in the Savannah of Cote d'Ivoire. *Tropical Medicine & International Health 8*, 449–458.

(*152*) Zogo, B., Soma, D. D., Tchiekoi, B. N., Somé, A., Alou, L. P. A., Koffi, A. A., Fournet, F., Dahounto, A., Coulibaly, B., and Kandé, S. (2019). Anopheles Bionomics, Insecticide Resistance Mechanisms, and Malaria Transmission in the Korhogo Area, Northern Côte d'Ivoire: A Pre-Intervention Study. *Parasite 26*.

(*153*) Henry, M.-C., Koné, M., Guillet, P., and Mouchet, J. (1998). Resistance to Chloroquine and Malaria Control in the Ivory Coast. *Cahiers d'Études et de Recherches Francophones/Santé 8*, 287–291.

(*154*) Henry, M.-C., van Leeuwen, W. D., Watson, P., Jansen, A., Jacobs, K., Agricola, K., Nahounou, N., Dossou, J., and Eggelte, T. E. (1994). In Vivo Sensitivity of Plasmodium Falciparum to Chloroquine in Rural Areas of Côte d'Ivoire. *Acta Tropica 58*, 275–281.

(*155*) Lusingu, J. P., Vestergaard, L. S., Mmbando, B. P., Drakeley, C. J., Jones, C., Akida, J., Savaeli, Z. X., Kitua, A. Y., Lemnge, M. M., and Theander, T. G. (2004). Malaria Morbidity and Immunity among Residents of Villages with Different Plasmodium Falciparum Transmission Intensity in North-Eastern Tanzania. *Malaria Journal 3*, 1–11.

(*156*) Bødker, R., Akida, J., Shayo, D., Kisinza, W., Msangeni, H. A., Pedersen, E. M., and Lindsay, S. W. (2003). Relationship between Altitude and Intensity of Malaria Transmission in the Usambara Mountains, Tanzania. *Journal of Medical Entomology 40*, 706–717.

(*157*) Alifrangis, M., Enosse, S., Khalil, I. F., Tarimo, D. S., Lemnge, M. M., Thompson, R., Bygbjerg, I. C., and Rønn, A. M. (2003). Prediction of Plasmodium Falciparum Resistance to Sulfadoxine/Pyrimethamine in Vivo by Mutations in the Dihydrofolate Reductase and Dihydropteroate Synthetase Genes: A Comparative Study between Sites of Differing Endemicity. *The American Journal of Tropical Medicine and Hygiene 69*, 601–606.

(*158*) Jelinek, T., Jelinek, T., Rønn, A. M., Lemnge, M. M., Curtis, J., Mhina, J., Duraisingh, M. T., Bygbjerg, I. C., and Warhurst, D. C. (1998). Polymorphisms in the Dihydrofolate Reductase (DHFR) and Dihydropteroate Synthetase (DHPS) Genes of Plasmodium Falciparum and in Vivo Resistance to Sulphadoxine/Pyrimethamine in Isolates from Tanzania. *Tropical Medicine & International Health 3*, 605–609.

(*159*) Mwangi, T. W., Ross, A., Snow, R. W., and Marsh, K. (2005). Case Definitions of Clinical Malaria under Different Transmission Conditions in Kilifi District, Kenya. *The Journal of Infectious Diseases 191*, 1932–1939.

(*160*) Mwangi, T. W. Clinical Epidemiology of Malaria under Differing Levels of Transmission, Ph.D. Thesis, The Open University, 2004.

(*161*) Mwangi, T. W., Ross, A., Marsh, K., and Snow, R. W. (2003). The Effects of Untreated Bednets on Malaria Infection and Morbidity on the Kenyan Coast. *Transactions of the Royal Society of Tropical Medicine and Hygiene 97*, 369–372.

(*162*) Mbogo, C. N., Snow, R. W., Khamala, C. P., Kabiru, E. W., Ouma, J. H., Githure, J. I., Marsh, K., and Beier, J. C. (1995). Relationships between Plasmodium Falciparum Transmission by Vector Populations and the Incidence of Severe Disease at Nine Sites on the Kenyan Coast. *The American Journal of Tropical Medicine and Hygiene 52*, 201–206.

(*163*) Mbogo, C. M., Mwangangi, J. M., Nzovu, J., Gu, W., Yan, G., Gunter, J. T., Swalm, C., Keating, J., Regens, J. L., and Shililu, J. I. (2003). Spatial and Temporal Heterogeneity of Anopheles Mosquitoes and Plasmodium Falciparum Transmission along the Kenyan Coast. *The American Journal of Tropical Medicine and Hygiene 68*, 734–742.

(*164*) O'Meara, W. P., Mwangi, T. W., Williams, T. N., McKenzie, F. E., Snow, R. W., and Marsh, K. (2008). Relationship between Exposure, Clinical Malaria, and Age in an Area of Changing Transmission Intensity. *The American Journal of Tropical Medicine and Hygiene 79*, 185–191.

(*165*) Snow, R. W., McCabe, E., Mbogo, C. N. M., Molyneux, C. S., Some, E. S., Mung'Ala, V. O., and Nevill, C. G. (1999). The Effect of Delivery Mechanisms on the Uptake of Bed Net Re-Impregnation in Kilifi District, Kenya. *Health Policy and Planning 14*, 18–25.

(*166*) Snow, R. W., Peshu, N., Forster, D., Mwenesi, H., and Marsh, K. (1992). The Role of Shops in the Treatment and Prevention of Childhood Malaria on the Coast of Kenya. *Transactions of the Royal Society of Tropical Medicine and Hygiene 86*, 237–239.

(*167*) Saúte, F., Aponte, J., Almeda, J., Ascaso, C., Abellana, R., Vaz, N., Dgedge, M., and Alonso, P. (2003). Malaria in Southern Mozambique: Malariometric Indicators and Malaria Casedefinition in Manhiça District. *Transactions of the Royal Society of Tropical Medicine and Hygiene 97*, 661–666.

(*168*) Saúte, F., Aponte, J., Ahmeda, J., Ascaso, C., Vaz, N., Dgedge, M., and Alonso, P. (2003). Malaria in Southern Mozambique: Incidence of Clinical Malaria in Children Living in a Rural Community in Manhica District. *Transactions of the Royal Society of Tropical Medicine and Hygiene 97*, 655–660.

(*169*) Aranda, C., Aponte, J. J., Saute, F., Casimiro, S., Pinto, J., Sousa, C., Rosario, V. D., Petrarca, V., Dgedge, M., and Alonso, P. (2005). Entomological Characteristics of Malaria Transmission in Manhica, a Rural Area in Southern Mozambique. *Journal of Medical Entomology 42*, 180–186.

(*170*) Yamba, E. I., Tompkins, A. M., Fink, A. H., Ermert, V., Amelie, M. D., Amekudzi, L. K., and Briët, O. J. (2020). Monthly Entomological Inoculation Rate Data for Studying the Seasoanality of Malaria Transmission in Africa. *Data 5*, 31.

(*171*) Mayor, A. G., Gómez-Olivé, X., Aponte, J. J., Casimiro, S., Mabunda, S., Dgedge, M., Barreto, A., and Alonso, P. L. (2001). Prevalence of the K76T Mutation in the Putative Plasmodium Falciparum Chloroquine Resistance Transporter (Pfcrt) Gene and Its Relation to Chloroquine Resistance in Mozambique. *The Journal of Infectious Diseases 183*, 1413–1416.

(*172*) Schellenberg, D. M., Aponte, J. J., Kahigwa, E. A., Mshinda, H., Tanner, M., Menendez, C., and Alonso, P. L. (2003). The Incidence of Clinical Malaria Detected by Active Case Detection in Children in Ifakara, Southern Tanzania. *Transactions of the Royal Society of Tropical Medicine and Hygiene 97*, 647–654.

(*173*) Drakeley, C., Schellenberg, D., Kihonda, J., Sousa, C. A., Arez, A. P., Lopes, D., Lines, J., Mshinda, H., Lengeler, C., and Schellenberg, J. A. (2003). An Estimation of the Entomological Inoculation Rate for Ifakara: A Semi-urban Area in a Region of Intense Malaria Transmission in Tanzania. *Tropical Medicine & International Health 8*, 767–774.

(*174*) Schellenberg, D., Menendez, C., Aponte, J., Guinovart, C., Mshinda, H., Tanner, M., and Alonso, P. (2004). The Changing Epidemiology of Malaria in Ifakara Town, Southern Tanzania. *Tropical Medicine & International Health 9*, 68–76.

(*175*) Thompson, R., Begtrup, K., Cuamba, N., Dgedge, M., Mendis, C., Gamage-Mendis, A., Enosse, S. M., Barreto, J., Sinden, R. E., and Hogh, B. (1997). The Matola Malaria Project: A Temporal and Spatial Study of Malaria Transmission and Disease in a Suburban Area of Maputo, Mozambique. *The American Journal of Tropical Medicine and Hygiene 57*, 550–559.

(*176*) Mendis, C., Jacobsen, J. L., Gamage-Mendis, A., Bule, E., Dgedge, M., Thompson, R., Cuamba, N., Barreto, J., Begtrup, K., and Sinden, R. E. (2000). Anopheles Arabiensis and An. Funestus Are Equally Important Vectors of Malaria in Matola Coastal Suburb of Maputo, Southern Mozambique. *Medical and Veterinary Entomology 14*, 171–180.

(*177*) Trape, J.-F., Tall, A., Diagne, N., Ndiath, O., Ly, A. B., Faye, J., Dieye-Ba, F., Roucher, C., Bouganali, C., and Badiane, A. (2011). Malaria Morbidity and Pyrethroid Resistance after the Introduction of Insecticide-Treated Bednets and Artemisinin-Based Combination Therapies: A Longitudinal Study. *The Lancet Infectious Diseases 11*, 925–932.

(*178*) Kasasa, S., Asoala, V., Gosoniu, L., Anto, F., Adjuik, M., Tindana, C., Smith, T., Owusu-Agyei, S., and Vounatsou, P. (2013). Spatio-Temporal Malaria Transmission Patterns in Navrongo Demographic Surveillance Site, Northern Ghana. *Malaria Journal 12*, 1–10.

(*179*) Gosoniu, L., Vounatsou, P., Tami, A., Nathan, R., Grundmann, H., and Lengeler, C. (2008). Spatial Effects of Mosquito Bednets on Child Mortality. *BMC Public Health 8*, 1–9.

(*180*) Rumisha, S. F., Smith, T., Abdulla, S., Masanja, H., and Vounatsou, P. (2014). Modelling Heterogeneity in Malaria Transmission Using Large Sparse Spatio-Temporal Entomological Data. *Global Health Action 7*, 22682.

(*181*) Amek, N. O., Van Eijk, A., Lindblade, K. A., Hamel, M., Bayoh, N., Gimnig, J., Laserson, K. F., Slutsker, L., Smith, T., and Vounatsou, P. (2018). Infant and Child Mortality in Relation to Malaria Transmission in KEMRI/CDC HDSS, Western Kenya: Validation of Verbal Autopsy. *Malaria Journal 17*, 1–11.

(*182*) Walski, T. M. (2000). Model Calibration Data: The Good, the Bad, and the Useless. *Journal – American Water Works Association 92*, 94–99.

(*183*) Smith, T., Maire, N., Dietz, K., Killeen, G. F., Vounatsou, P., Molineaux, L., and Tanner, M. (2006). Relationship between the Entomologic Inoculation Rate and the Force of

Infection for Plasmodium Falciparum Malaria. *The American Journal of Tropical Medicine and Hygiene 75*, 11–18.

(*184*) Gerardin, J., Ouédraogo, A. L., McCarthy, K. A., Eckhoff, P. A., and Wenger, E. A. (2015). Characterization of the Infectious Reservoir of Malaria with an Agent-Based Model Calibrated to Age-Stratified Parasite Densities and Infectiousness. *Malaria Journal 14*, 1–13.

(*185*) Collins, W. E., and Jeffery, G. M. (1999). A Retrospective Examination of Sporozoite- and Trophozoite-Induced Infections with Plasmodium Falciparum: Development of Parasitologic and Clinical Immunity during Primary Infection. *The American Journal of Tropical Medicine and Hygiene 61*, 4–19.

(*186*) Felger, I., Maire, M., Bretscher, M. T., Falk, N., Tiaden, A., Sama, W., Beck, H.-P., Owusu-Agyei, S., and Smith, T. A. (2012). The Dynamics of Natural Plasmodium Falciparum Infections. *PLOS ONE 7*, e45542.

(*187*) Marsh, K. (2006). PREFACE. *The American Journal of Tropical Medicine and Hygiene 75*, i–i.

(*188*) Brooks, A., Briët, O. J., Hardy, D., Steketee, R., and Smith, T. A. (2012). Simulated Impact of RTS, S/AS01 Vaccination Programs in the Context of Changing Malaria Transmission. *PLOS ONE 7*, e32587.

(*189*) Maire, N., Shillcutt, S. D., Walker, D. G., Tediosi, F., and Smith, T. A. (2011). Cost-Effectiveness of the Introduction of a Pre-Erythrocytic Malaria Vaccine into the Expanded Program on Immunization in Sub-Saharan Africa: Analysis of Uncertainties Using a Stochastic Individual-Based Simulation Model of Plasmodium Falciparum Malaria. *Value in Health 14*, 1028–1038.

(*190*) Nunes, J. K., Cárdenas, V., Loucq, C., Maire, N., Smith, T., Shaffer, C., Måseide, K., and Brooks, A. (2013). Modeling the Public Health Impact of Malaria Vaccines for Developers and Policymakers. *BMC Infectious Diseases 13*, 1–15.

(*191*) Penny, M. A., Galactionova, K., Tarantino, M., Tanner, M., and Smith, T. A. (2015). The Public Health Impact of Malaria Vaccine RTS, S in Malaria Endemic Africa: Country-Specific Predictions Using 18 Month Follow-up Phase III Data and Simulation Models. *BMC Medicine 13*, 1–20.

(*192*) Tediosi, F., Maire, N., Penny, M., Studer, A., and Smith, T. A. (2009). Simulation of the Cost-Effectiveness of Malaria Vaccines. *Malaria Journal 8*, 1–17.

(*193*) Crowell, V., Hardy, D., Briët, O., Chitnis, N., Maire, N., and Smith, T. (2012). Can We Depend on Case Management to Prevent Re-Establishment of P. Falciparum Malaria, after Local Interruption of Transmission? *Epidemics 4*, 1–8.

(*194*) Ross, A., Maire, N., Sicuri, E., Smith, T., and Conteh, L. (2011). Determinants of the Cost-Effectiveness of Intermittent Preventive Treatment for Malaria in Infants and Children. *PLOS ONE 6*, e18391.

(*195*) Briët, O. J., and Chitnis, N. (2013). Effects of Changing Mosquito Host Searching Behaviour on the Cost Effectiveness of a Mass Distribution of Long-Lasting, Insecticidal Nets: A Modelling Study. *Malaria Journal 12*, 1–11.

(*196*) Briët, O. J., Penny, M. A., Hardy, D., Awolola, T. S., Van Bortel, W., Corbel, V., Dabiré, R. K., Etang, J., Koudou, B. G., and Tungu, P. K. (2013). Effects of Pyrethroid Resistance

on the Cost Effectiveness of a Mass Distribution of Long-Lasting Insecticidal Nets: A Modelling Study. *Malaria Journal 12*, 1–12.

(*197*) Briët, O. J., Gething, P. W., Maire, N., Tarantino, M., and Hay, S. I. (2012). Estimated Malaria Epidemiologically Effective Lifetime of Mass LLIN Distributions Depending on Transmission in African Countries. *Report for African Leaders Malaria Alliance*.

(*198*) Briët, O. J., Hardy, D., and Smith, T. A. (2012). Importance of Factors Determining the Effective Lifetime of a Mass, Long-Lasting, Insecticidal Net Distribution: A Sensitivity Analysis. *Malaria Journal 11*, 1–27.

(*199*) Briët, O. J., and Penny, M. A. (2013). Repeated Mass Distributions and Continuous Distribution of Long-Lasting Insecticidal Nets: Modelling Sustainability of Health Benefits from Mosquito Nets, Depending on Case Management. *Malaria Journal 12*, 1–19.

(*200*) Chitnis, N., Schapira, A., Smith, T., and Steketee, R. (2010). Comparing the Effectiveness of Malaria Vector-Control Interventions through a Mathematical Model. *The American Journal of Tropical Medicine and Hygiene 83*, 230–240.

(*201*) Hamilton, M., Mahiane, G., Werst, E., Sanders, R., Briët, O., Smith, T., Cibulskis, R., Cameron, E., Bhatt, S., and Weiss, D. J. (2017). Spectrum-Malaria: A User-Friendly Projection Tool for Health Impact Assessment and Strategic Planning by Malaria Control Programmes in Sub-Saharan Africa. *Malaria Journal 16*, 1–15.

(*202*) Korenromp, E., Hamilton, M., Sanders, R., Mahiané, G., Briët, O. J., Smith, T., Winfrey, W., Walker, N., and Stover, J. (2017). Impact of Malaria Interventions on Child Mortality in Endemic African Settings: Comparison and Alignment between LiST and Spectrum-Malaria Model. *BMC Public Health 17*, 29–42.

(*203*) Korenromp, E., Mahiané, G., Hamilton, M., Pretorius, C., Cibulskis, R., Lauer, J., Smith, T. A., and Briët, O. J. (2016). Malaria Intervention Scale-up in Africa: Effectiveness Predictions for Health Programme Planning Tools, Based on Dynamic Transmission Modelling. *Malaria Journal 15*, 1–14.

(*204*) Korenromp, E. L., Williams, B. G., Gouws, E., Dye, C., and Snow, R. W. (2003). Measurement of Trends in Childhood Malaria Mortality in Africa: An Assessment of Progress toward Targets Based on Verbal Autopsy. *The Lancet Infectious Diseases 3*, 349–358.

(*205*) Stuckey, E. M., Stevenson, J., Galactionova, K., Baidjoe, A. Y., Bousema, T., Odongo, W., Kariuki, S., Drakeley, C., Smith, T. A., and Cox, J. (2014). Modeling the Cost Effectiveness of Malaria Control Interventions in the Highlands of Western Kenya. *PLOS ONE 9*, e107700.

(*206*) Stuckey, E. M., Miller, J. M., Littrell, M., Chitnis, N., and Steketee, R. (2016). Operational Strategies of Anti-Malarial Drug Campaigns for Malaria Elimination in Zambia's Southern Province: A Simulation Study. *Malaria Journal 15*, 1–14.

(*207*) Stuckey, E. M., Stevenson, J. C., Cooke, M. K., Owaga, C., Marube, E., Oando, G., Hardy, D., Drakeley, C., Smith, T. A., and Cox, J. (2012). Simulation of Malaria Epidemiology and Control in the Highlands of Western Kenya. *Malaria Journal 11*, 1–14.

(*208*) Stuckey, E. M., Smith, T. A., and Chitnis, N. (2013). Estimating Malaria Transmission through Mathematical Models. *Trends in Parasitology 29*, 477–482.

(*209*) Stuckey, E. M., Smith, T., and Chitnis, N. (2014). Seasonally Dependent Relationships between Indicators of Malaria Transmission and Disease Provided by Mathematical Model Simulations. *PLOS Computational Biology 10*, e1003812.

(*210*) Camponovo, F., Bever, C. A., Galactionova, K., Smith, T., and Penny, M. A. (2017). Incidence and Admission Rates for Severe Malaria and Their Impact on Mortality in Africa. *Malaria Journal 16*, 1–12.

(*211*) Killeen, G. F., Ross, A., and Smith, T. (2006). Infectiousness of Malaria-Endemic Human Populations to Vectors. *The American Journal of Tropical Medicine and Hygiene 75*, 38–45.

(*212*) Ross, A., Killeen, G., and Smith, T. (2006). Relationships between Host Infectivity to Mosquitoes and Asexual Parasite Density in Plasmodium Falciparum. *The American Journal of Tropical Medicine and Hygiene 75*, 32–37.

(*213*) Carneiro, I. A., Smith, T., Lusingu, J. P., Malima, R., Utzinger, J., and Drakeley, C. J. (2006). Modeling the Relationship between the Population Prevalence of Plasmodium Falciparum Malaria and Anemia. *The American Journal of Tropical Medicine and Hygiene 75*, 82–89.

(*214*) Ross, A., and Smith, T. (2006). The Effect of Malaria Transmission Intensity on Neonatal Mortality in Endemic Areas. *The American Journal of Tropical Medicine and Hygiene 75*, 74–81.

(*215*) Ross, A., Maire, N., Molineaux, L., and Smith, T. (2006). An Epidemiologic Model of Severe Morbidity and Mortality Caused by Plasmodium Falciparum. *The American Journal of Tropical Medicine and Hygiene 75*, 63–73.

(*216*) Reyburn H., Drakely C., Cerneiro I., Jones C., Cox J., Bruce, J., Riley E., B. Greenwood, and C., W. The Epidemiology of Severe Malaria Due to Plasmodium Falciparum at Different Transmisison Intensities in NE Tanzania, tech. rep., 2004.

(*217*) Marsh, K, and Snow, R. (1999). Malaria Transmission and Morbidity. *Parassitologia 41*, 241.

(*218*) Griffin, J. T., Hollingsworth, T. D., Reyburn, H., Drakeley, C. J., Riley, E. M., and Ghani, A. C. (2015). Gradual Acquisition of Immunity to Severe Malaria with Increasing Exposure. *Proceedings of the Royal Society B: Biological Sciences 282*, 20142657.

(*219*) Bousema, T., Okell, L., Felger, I., and Drakeley, C. (2014). Asymptomatic Malaria Infections: Detectability, Transmissibility and Public Health Relevance. *Nature Reviews Microbiology 12*, 833–840.

(*220*) World Health Organization, *World Malaria Report 2015*; World Health Organization: Geneva, 2016.

(*221*) Van Eijk, A. M., Ramanathapuram, L., Sutton, P. L., Kanagaraj, D., Priya, G. S. L., Ravishankaran, S., Asokan, A., Tandel, N., Patel, A., and Desai, N. (2016). What Is the Value of Reactive Case Detection in Malaria Control? A Case-Study in India and a Systematic Review. *Malaria Journal 15*, 1–13.

(*222*) Larsen, D. A., Ngwenya-Kangombe, T., Cheelo, S., Hamainza, B., Miller, J., Winters, A., and Bridges, D. J. (2017). Location, Location, Location: Environmental Factors Better

Predict Malaria-Positive Individuals during Reactive Case Detection than Index Case Demographics in Southern Province, Zambia. *Malaria Journal 16*, 1–9.

(*223*) Zhou, S.-S., Zhang, S.-S., Zhang, L., Rietveld, A. E., Ramsay, A. R., Zachariah, R., Bissell, K., Van den Bergh, R., Xia, Z.-G., and Zhou, X.-N. (2015). China's 1-3-7 Surveillance and Response Strategy for Malaria Elimination: Is Case Reporting, Investigation and Foci Response Happening According to Plan? *Infectious Diseases of Poverty 4*, 1–9.

(*224*) Cao, J., Sturrock, H. J., Cotter, C., Zhou, S., Zhou, H., Liu, Y., Tang, L., Gosling, R. D., Feachem, R. G., and Gao, Q. (2014). Communicating and Monitoring Surveillance and Response Activities for Malaria Elimination: China's "1-3-7" Strategy. *PLOS Medicine 11*, e1001642.

(*225*) Cotter, C., Sudathip, P., Herdiana, H., Cao, Y., Liu, Y., Luo, A., Ranasinghe, N., Bennett, A., Cao, J., and Gosling, R. D. (2017). Piloting a Programme Tool to Evaluate Malaria Case Investigation and Reactive Case Detection Activities: Results from 3 Settings in the Asia Pacific. *Malaria Journal 16*, 1–13.

(*226*) Murray, C. K., Gasser, R. A., Magill, A. J., and Miller, R. S. (2008). Update on Rapid Diagnostic Testing for Malaria. *Clinical Microbiology Reviews 21*, 97–110.

(*227*) Yukich, J., Bennett, A., Yukich, R., Stuck, L., Hamainza, B., Silumbe, K., Smith, T., Chitnis, N., Steketee, R. W., and Finn, T. (2017). Estimation of Malaria Parasite Reservoir Coverage Using Reactive Case Detection and Active Community Fever Screening from Census Data with Rapid Diagnostic Tests in Southern Zambia: A Re-Sampling Approach. *Malaria Journal 16*, 1–13.

(*228*) Okumu, F. O., Govella, N. J., Moore, S. J., Chitnis, N., and Killeen, G. F. (2010). Potential Benefits, Limitations and Target Product-Profiles of Odor-Baited Mosquito Traps for Malaria Control in Africa. *PLOS ONE 5*, e11573.

(*229*) Hogan, A. B., Winskill, P., Verity, R., Griffin, J. T., and Ghani, A. C. (2018). Modelling Population-Level Impact to Inform Target Product Profiles for Childhood Malaria Vaccines. *BMC Medicine 16*, 1–11.

(*230*) Penny, M. A., Camponovo, F., Chitnis, N., Smith, T. A., and Tanner, M. (2020). Future Use-Cases of Vaccines in Malaria Control and Elimination. *Parasite Epidemiology and Control 10*, e00145.

(*231*) Runge, M., Molteni, F., Mandike, R., Snow, R. W., Lengeler, C., Mohamed, A., and Pothin, E. (2020). Applied Mathematical Modelling to Inform National Malaria Policies, Strategies and Operations in Tanzania. *Malaria Journal 19*, 1–10.

(*232*) Rebholz-Schuhmann, D., Oellrich, A., and Hoehndorf, R. (2012). Text-Mining Solutions for Biomedical Research: Enabling Integrative Biology. *Nature Reviews Genetics 13*, 829–839.

(*233*) Rzhetsky, A., Seringhaus, M., and Gerstein, M. (2008). Seeking a New Biology through Text Mining. *Cell 134*, 9–13.

(*234*) Golumbeanu, M., Yang, G., Camponovo, F., Stuckey, E. M., Hamon, N., Mondy, M., Rees, S., Chitnis, N., Cameron, E., and Penny, M. A. (2021). Combining Machine Learning and Mathematical Models of Disease Dynamics to Guide Development of Novel Disease Interventions. *medRxiv*.

(*235*) Matott, L. S., Babendreier, J. E., and Purucker, S. T. (2009). Evaluating Uncertainty in Integrated Environmental Models: A Review of Concepts and Tools. *Water Resources Research 45*.

(*236*) Japkowicz, N. In *AAAI Workshop on Evaluation Methods for Machine Learning*, 2006, pp 6–11.

(*237*) Faes, M., Broggi, M., Beer, M., and Moens, D. In *Proceedings of the Joint ICVRAM ISUMA UNCERTAINTIES Conference*, 2018; Vol. 3.

(*238*) Kandasamy, K., Vysyaraju, K. R., Neiswanger, W., Paria, B., Collins, C. R., Schneider, J., Poczos, B., and Xing, E. P. Tuning Hyperparameters without Grad Students: Scalable and Robust Bayesian Optimisation with Dragonfly (2019), `https://arxiv.org/abs/1903.06694`.

(*239*) Ghaderi, A., Morovati, V., and Dargazany, R. A Bayesian Surrogate Constitutive Model to Estimate Failure Probability of Rubber-like Materials (2020), `https://arxiv.org/abs/2010.13241`.

(*240*) Li, J., Li, J., and Xiu, D. (2011). An Efficient Surrogate-Based Method for Computing Rare Failure Probability. *Journal of Computational Physics 230*, 8683–8697.

(*241*) Li, J., and Xiu, D. (2010). Evaluation of Failure Probability via Surrogate Models. *Journal of Computational Physics 229*, 8966–8980.

(*242*) Peherstorfer, B., Kramer, B., and Willcox, K. (2017). Combining Multiple Surrogate Models to Accelerate Failure Probability Estimation with Expensive High-Fidelity Models. *Journal of Computational Physics 341*, 61–75.

(*243*) Xiao, N.-C., Zhan, H., and Yuan, K. (2020). A New Reliability Method for Small Failure Probability Problems by Combining the Adaptive Importance Sampling and Surrogate Models. *Computer Methods in Applied Mechanics and Engineering 372*, 113336.

(*244*) Contreras, M. T., Gironás, J., and Escauriaza, C. (2020). Forecasting Flood Hazards in Real Time: A Surrogate Model for Hydrometeorological Events in an Andean Watershed. *Natural Hazards and Earth System Sciences 20*, 3261–3277.

(*245*) Gaspar, B., Teixeira, A. P., and Soares, C. G. (2014). Assessment of the Efficiency of Kriging Surrogate Models for Structural Reliability Analysis. *Probabilistic Engineering Mechanics 37*, 24–34.

(*246*) Collins, W. E., Jeffery, G. M., and Roberts, J. M. (2004). A Retrospective Examination of Reinfection of Humans with Plasmodium Vivax. *The American Journal of Tropical Medicine and Hygiene 70*, 642–644.

(*247*) Molineaux, L., Träuble, M., Collins, W. E., Jeffery, G. M., and Dietz, K. (2002). Malaria Therapy Reinoculation Data Suggest Individual Variation of an Innate Immune Response and Independent Acquisition of Antiparasitic and Antitoxic Immunities. *Transactions of the Royal Society of Tropical Medicine and Hygiene 96*, 205–209.

(*248*) World Health Organization Malaria Eradication: Benefits, Future Scenarios and Feasibility, tech. rep., Geneva, 2020.

(*249*) Drosten, C., and Henning, K. (16) Wir brauchen Abkuerzungen bei der Impfstoffzulassung Podcast: Das Coronavirus-Update von NDR Info, Published 18.03.2020 (2020),

https://www.ndr.de/nachrichten/info/16-Wir-brauchen-Abkuerzungen-bei-der-Impfstoffzulassung,audio655164.html.

(*250*) Greenwood, B., *Elimination of Malaria: Halfway There*; Oxford University Press: 2017.

(*251*) Port, G., Boreham, P., and Bryan, J. H. (1980). The Relationship of Host Size to Feeding by Mosquitoes of the Anopheles Gambiae Giles Complex (Diptera: Culicidae). *Bulletin of Entomological Research 70*, 133–144.

(*252*) Owusu-Agyei, S., Smith, T, Beck, H.-P., Amenga-Etego, L, and Felger, I (2002). Molecular Epidemiology of Plasmodium Falciparum Infections among Asymptomatic Inhabitants of a Holoendemic Malarious Area in Northern Ghana. *Tropical Medicine & International Health 7*, 421–428.

(*253*) Smith, T, Charlwood, J., Kihonda, J, Mwankusye, S, Billingsley, P, Meuwissen, J, Lyimo, E, Takken, W, Teuscher, T, and Tanner, M. (1993). Absence of Seasonal Variation in Malaria Parasitaemia in an Area of Intense Seasonal Transmission. *Acta Tropica 54*, 55–72.

(*254*) Kitua, A. Y., Smith, T, Alonso, P., Masanja, H, Urassa, H, Menendez, C, Kimario, J, and Tanner, M (1996). Plasmodium Falciparum Malaria in the First Year of Life in an Area of Intense and Perennial Transmission. *Tropical Medicine & International Health 1*, 475–484.

(*255*) Earle, W. C., and Perez, M. (1932). Enumeration of Parasites in the Blood of Malarial Patients. *Journal of Laboratory and Clinical Medicine 17*.

(*256*) Trape, J.-F., and Rogier, C. (1996). Combating Malaria Morbidity and Mortality by Reducing Transmission. *Parasitology Today 12*, 236–240.

(*257*) Trape, J.-F., Rogier, C., Konate, L., Diagne, N., Bouganali, H., Canque, B., Legros, F., Badji, A., Ndiaye, G., and Ndiaye, P. (1994). The Dielmo Project: A Longitudinal Study of Natural Malaria Infection and the Mechanisms of Protective Immunity in a Community Living in a Holoendemic Area of Senegal. *The American Journal of Tropical Medicine and Hygiene 51*, 123–137.

(*258*) Charlwood, J., Smith, T, Lyimo, E, Kitua, A., Masanja, H, Booth, M, Alonso, P., and Tanner, M (1998). Incidence of Plasmodium Falciparum Infection in Infants in Relation to Exposure to Sporozoite-Infected Anophelines. *The American Journal of Tropical Medicine and Hygiene 59*, 243–251.

(*259*) Snow, R. W., Omumbo, J. A., Lowe, B., Molyneux, C. S., Obiero, J.-O., Palmer, A., Weber, M. W., Pinder, M., Nahlen, B., and Obonyo, C. (1997). Relation between Severe Malaria Morbidity in Children and Level of Plasmodium Falciparum Transmission in Africa. *The Lancet 349*, 1650–1654.

(*260*) Rogier, C., Commenges, D., and Trape, J.-F. (1996). Evidence for an Age-Dependent Pyrogenic Threshold of Plasmodium Falciparum Parasitemia in Highly Endemic Populations. *The American Journal of Tropical Medicine and Hygiene 54*, 613–619.

(*261*) Vounatsou, P, Smith, T, Kitua, A., Alonso, P., and Tanner, M (2000). Apparent Tolerance of Plasmodium Falciparum in Infants in a Highly Endemic Area. *Parasitology 120*, 1–9.

(*262*) Beier, J. C., Killeen, G. F., and Githure, J. I. (1999). Entomologic Inoculation Rates and Plasmodium Falciparum Malaria Prevalence in Africa. *The American Journal of Tropical Medicine and Hygiene 61*, 109–113.

(*263*) Barnish, G, Maude, G., Bockarie, M., Eggelte, T., Greenwood, B., and Ceesay, S (1993). Malaria in a Rural Area of Sierra Leone. I. Initial Results. *Annals of Tropical Medicine & Parasitology 87*, 125–136.

(*264*) INDEPTH Network, and International Development Research Centre, *Population and Health in Developing Countries: Population, Health and Survival at INDEPTH Sites*; IDRC: 2002; Vol. 1.

(*265*) Spencer, H. C., Kaseje, D. C., Mosley, W. H., Sempebwa, E. K., Huong, A. Y., and Roberts, J. M. (1987). Impact on Mortality and Fertility of a Community-Based Malaria Control Programme in Saradidi, Kenya. *Annals of Tropical Medicine & Parasitology 81*, 36–45.

(*266*) D'Alessandro, U., Olaleye, B, Langerock, P, Aikins, M., Thomson, M., Cham, M., Greenwood, B., McGuire, W, Bennett, S, and Cham, B. (1995). Mortality and Morbidity from Malaria in Gambian Children after Introduction of an Impregnated Bednet Programme. *The Lancet 345*, 479–483.

(*267*) Duboz, P, Vaugelade, J, and Debouverie, M Mortalité Dans l'enfance Dans La Région de Niangoloko, tech. rep., Ouagadougou, Burkina Faso: ORSTOM, 1989.

(*268*) Schellenberg, J. A., Abdulla, S., Minja, H., Nathan, R., Mukasa, O., Marchant, T., Mponda, H., Kikumbih, N., Lyimo, E., and Manchester, T. (1999). KINET: A Social Marketing Programme of Treated Nets and Net Treatment for Malaria Control in Tanzania, with Evaluation of Child Health and Long-Term Survival. *Transactions of the Royal Society of Tropical Medicine and Hygiene 93*, 225–231.

(*269*) Premji, Z., Ndayanga, P, Shiff, C, Minjas, J, Lubega, P, and MacLeod, J (1997). Community Based Studies on Childhood Mortality in a Malaria Holoendemic Area on the Tanzanian Coast. *Acta Tropica 63*, 101–109.

(*270*) Trape, J.-F., Pison, G., Preziosi, M.-P., Enel, C., du Loû, A. D., Delaunay, V., Samb, B., Lagarde, E., Molez, J.-F., and Simondon, F. (1998). Impact of Chloroquine Resistance on Malaria Mortality. *Comptes Rendus de l'Académie des Sciences-Series III-Sciences de la Vie 321*, 689–697.

(*271*) Tediosi, F., Maire, N., Smith, T., Hutton, G., Utzinger, J., Ross, A., and Tanner, M. (2006). An Approach to Model the Costs and Effects of Case Management of Plasmodium Falciparum Malaria in Sub-Saharan Africa. *The American Journal of Tropical Medicine and Hygiene 75*, 90–103.

(*272*) Wongsrichanalai, C., Barcus, M. J., Muth, S., Sutamihardja, A., and Wernsdorfer, W. H. (2007). A Review of Malaria Diagnostic Tools: Microscopy and Rapid Diagnostic Test (RDT). *The American Journal of Tropical Medicine and Hygiene 77*, 119–127.

(*273*) Bloland, P. B., Boriga, D. A., Ruebush, T. K., McCormick, J. B., Roberts, J. M., Oloo, A. J., Hawley, W., Lal, A., Nahlen, B., and Campbell, C. C. (1999). Longitudinal Cohort Study of the Epidemiology of Malaria Infections in an Area of Intense Malaria Transmission II. Descriptive Epidemiology of Malaria Infection and Disease among Children. *The American Journal of Tropical Medicine and Hygiene 60*, 641–648.

(*274*) Bougouma, E. C., Tiono, A. B., Ouédraogo, A., Soulama, I., Diarra, A., Yaro, J.-B., Ouédraogo, E., Sanon, S., Konaté, A. T., and Nébié, I. (2012). Haemoglobin Variants and

Plasmodium Falciparum Malaria in Children under Five Years of Age Living in a High and Seasonal Malaria Transmission Area of Burkina Faso. *Malaria Journal 11*, 1–10.

(*275*) Coulibaly, D., Diallo, D. A., Thera, M. A., Dicko, A., Guindo, A. B., Koné, A. K., Cissoko, Y., Coulibaly, S., Djimdé, A., and Lyke, K. (2002). Impact of Preseason Treatment on Incidence of Falciparum Malaria and Parasite Density at a Site for Testing Malaria Vaccines in Bandiagara, Mali. *The American Journal of Tropical Medicine and Hygiene 67*, 604–610.

(*276*) Diallo, S., Konate, L., Ndir, O., Dieng, T., Dieng, Y., Bah, I. B., Faye, O., and Gaye, O. (2000). Malaria in the Central Health District of Dakar (Senegal). Entomological, Parasitological and Clinical Data. *Cahiers d'Études et de Recherches Francophones/Santé 10*, 221–229.

(*277*) Diallo, S., Ndir, O., Faye, O., Diop, B. M., Dieng, Y., Bah, I. B., Dieng, T., Gaye, O., and Konate, L. (1998). Malaria in the Southern Sanitary District of Dakar (Senegal). 1. Parasitemia and Malarial Attacks. *Bulletin de la Societe de Pathologie Exotique 91*, 208–213.

(*278*) Fillol, F., Cournil, A., Boulanger, D., Cissé, B., Sokhna, C., Targett, G., Trape, J.-F., Simondon, F., Greenwood, B., and Simondon, K. B. (2009). Influence of Wasting and Stunting at the Onset of the Rainy Season on Subsequent Malaria Morbidity among Rural Preschool Children in Senegal. *The American Journal of Tropical Medicine and Hygiene 80*, 202–208.

(*279*) Greenwood, B. M., Bradley, A. K., Greenwood, A. M., Byass, P., Jammeh, K., Marsh, K., Tulloch, S., Oldfield, F. S. J., and Hayes, R. (1987). Mortality and Morbidity from Malaria among Children in a Rural Area of The Gambia, West Africa. *Transactions of the Royal Society of Tropical Medicine and Hygiene 81*, 478–486.

(*280*) Guinovart, C., Bassat, Q., Sigaúque, B., Aide, P., Sacarlal, J., Nhampossa, T., Bardají, A., Nhacolo, A., Macete, E., and Mandomando, I. (2008). Malaria in Rural Mozambique. Part I: Children Attending the Outpatient Clinic. *Malaria Journal 7*, 1–9.

(*281*) Loha, E., Lunde, T. M., and Lindtjørn, B. (2012). Effect of Bednets and Indoor Residual Spraying on Spatio-Temporal Clustering of Malaria in a Village in South Ethiopia: A Longitudinal Study. *PLOS ONE 7*, e47354.

(*282*) Molez, J. F., Diop, A., Gaye, O., Lemasson, J. J., and Fontenille, D. (2006). Malaria Morbidity in Barkedji, Village of Ferlo, in Senegal Sahelian Area. *Bulletin de la Societe de Pathologie Exotique 99*, 187–190.

(*283*) Nebie, I., Diarra, A., Ouedraogo, A., Soulama, I., Bougouma, E. C., Tiono, A. B., Konate, A. T., Chilengi, R., Theisen, M., and Dodoo, D. (2008). Humoral Responses to Plasmodium Falciparum Blood-Stage Antigens and Association with Incidence of Clinical Malaria in Children Living in an Area of Seasonal Malaria Transmission in Burkina Faso, West Africa. *Infection and Immunity 76*, 759–766.

(*284*) Owusu-Agyei, S., Asante, K. P., Adjuik, M., Adjei, G., Awini, E., Adams, M., Newton, S., Dosoo, D., Dery, D., and Agyeman-Budu, A. (2009). Epidemiology of Malaria in the Forest-Savanna Transitional Zone of Ghana. *Malaria Journal 8*, 1–10.

(*285*) Rogier, C., and Trape, J.-f (1995). Etude de l'acquisition de La Prémunition En Zones d'holo-et de Méso-Endémie Palustre et Dielmo et a Ndiop (Sénégal): Résultats Préliminaires, 1990-1994. *Médecine Tropicale 55*, 71–76.

(*286*) Trape, J.-F., Zoulani, A., and Quinet, M. C. (1987). Assessment of the Incidence and Prevalence of Clinical Malaria in Semi-Immune Children Exposed to Intense and Perennial Transmission. *American Journal of Epidemiology 126*, 193–201.

(*287*) Velema, J. P., Alihonou, E. M., Chippaux, J.-P., van Boxel, Y., Gbedji, E., and Adegbini, R. (1991). Malaria Morbidity and Mortality in Children under Three Years of Age on the Coast of Benin, West Africa. *Transactions of the Royal Society of Tropical Medicine and Hygiene 85*, 430–435.

(*288*) Jewell, N. P., *Statistics for Epidemiology*; CRC Press: 2003.

(*289*) Stevenson, M., Stevenson, M. M., and BiasedUrn, I. R Package 'epiR' Version 2.0-17 (2020), `https://cran.ism.ac.jp/web/packages/epiR/epiR.pdf`.