# Computer modelling approaches for improving analysis of anti-malarial clinical trials.

Thesis submitted in accordance with the requirements of the Liverpool School of Tropical Medicine for the degree of Doctor in Philosophy

by

Sam William Jones

August 2019

# Table of Contents

**Acknowledgments.** ..... 3

**Thesis abstract.** ..... 3

**Tables of abbreviations.** ..... 4

**Chapter 1: Background to thesis.** ..... 7

**Chapter 2: Mechanistic Pharmacokinetic/Pharmacodynamic (mPK/PD) modelling to simulate parasite dynamics post-treatment with artemisinin-based combination therapies (ACTs).** ..... 20

**Chapter 3: Improving methods for analysing anti-malarial drug efficacy trials: molecular correction based on length-polymorphic markers msp-1, msp-2 and glurp.** ..... 34

**Chapter 4: Evaluating accuracy of microsatellite markers for classification of recurrent infections during routine monitoring of anti-malarial drug efficacy: A computer modelling approach.** ..... 75

**Chapter 5: Validation of next-generation amplicon sequencing for molecular correction using a computer modelling approach.** ..... 121

**Chapter 6: Optimal treatments for severe malaria and the threat posed by artemisinin resistance.** ..... 154

**Chapter 7: Thesis Conclusions.** ..... 202

**References.** ..... 216

## Acknowledgments.

Firstly, I would like to express my appreciation to my colleagues for their support throughout the lifetime of this thesis, particularly Eva Maria Hodel, Katherine Kay and Raman Sharma for their teaching, advice and encouragement. Most of all, to Ian Hastings, for allowing me to undertake this project and his constant support throughout.

Secondly, to my friends, for your support, company, and good humour.

Finally, to my family, for your unflinching love and support for as long as I can remember.

## Thesis abstract.

Clinical trials of treatments for Plasmodium falciparum are an integral aspect of a continually evolving evidence base that informs public health policy with the aim of reducing malaria morbidity, mortality, preventing the emergence of parasite resistance to drugs and, eventually, permitting elimination of the disease. Despite their importance, obtaining useful information from *in vivo* trials can be hindered through methodological gaps that make it difficult to obtain or analyse results (through an inability to quantify important parameters *in vivo*), cost, required patient numbers or ethical considerations.

This thesis uses a computer modelling approach to address two key research problems relating to *in vivo* trials: Firstly, it quantifies the accuracy failure rate estimates obtained during trials for routine monitoring of artemisinin-based combination therapy (ACT) efficacy in cases of uncomplicated malaria, noting that currently available methods for genotyping patient blood samples are imperfect, and that patients can be infected by new parasite clones (termed reinfection) during the follow-up period which may share (by chance) genetic data with clones present at the time of treatment. Consequently, it is possible for drug failure to be misclassified as a reinfection or vice versa, inducing error in drug failure rate estimates. The true drug failure rate cannot be known *in* vivo so the accuracy of each method is not known. The results presented here show that currently used methods (length-polymorphic markers and microsatellite markers) are under-estimating true drug failure rate and preventing the detection of failing drugs (~10% failure rate). Accuracy of failure rate estimates was greatly improved by using alternative statistical algorithms or through use of novel Amplicon Sequencing techniques for genotyping blood samples.

Secondly, clinical trials of severe malaria generally use reduction in circulating parasite numbers as a clinical endpoint but sequestered - not circulating - parasites are responsible for pathology in severe malaria. A mathematical model was developed to quantify the pathology of severe malaria in an *in silico* patient population based on sequestered parasite numbers. Results from this model then indicated that a simplified treatment regimen was generally non-inferior to the World Health Organization (WHO) recommended regimen, though specific sub-groups of patients may be at increased risk. Model results also indicated that the emergence of resistance to artesunate in parasite early ring-stages would have severe consequences for patient prognosis in cases of severe malaria.

# Tables of abbreviations.

**General abbreviations.**

| | |
|---|---|
| ACT | Artemisinin-based Combination Therapy |
| AmpSeq | Amplicon Sequencing |
| CDC | Centers for Disease Control and Prevention |
| CE | Capillary Electrophoresis |
| DNA | Deoxyribonucleic Acid |
| ETF | Early Treatment Failure |
| GMS | Greater Mekong Subregion |
| He | Expected Heterozygosity |
| HIV | Human Immunodeficiency Virus |
| i.m | Intramuscular |
| i.v | Intravenous |
| iRBC | Infected Red Blood Cell |
| MMV | Malaria for Medicines Venture |
| mPK/PD | Mechanistic Pharmacokinetic/Pharmacodynamic |
| PCR | Polymerase Chain Reaction |
| PD | Pharmacodynamics |
| PK | Pharmacokinetics |
| PK/PD | Pharmacokinetic/Pharmacodynamic |
| PRCC | Partial Rank Correlation Coefficient |
| RCT | Randomized Controlled Trial |
| rfu | Relative Fluorescent Units |
| ROC | Receiver Operator Characteristic |
| SNP | Single Nucleotide Polymorphism |
| TB | Tuberculosis |
| TES | Therapeutic Efficacy Study |
| VPC | Visual Predictive Check |
| WHO | World Health Organization |

**Drug name abbreviations.**

| | |
|---|---|
| AQ | Amodiaquine |
| AR | Artemether |
| AS | Artesunate |
| CQ | Chloroquine |
| DHA | Dihydroartemisinin |
| LF | Lumefantrine |
| MQ | Mefloquine |
| PPQ | Piperaquine |
| PYN | Pyronaridine |
| SP | Sulfadoxine/pyrimethamine |

**Genetic locus abbreviations.**

| | |
|---|---|
| ama1-D3 | Apical Membrane Antigen |
| cpmp | Conserved Plasmodium Membrane Protein |
| cpp | Conserved Plasmodium Protein |
| csp | Circumsporozoite Surface Protein |
| glurp | Glutamate rich protein |
| msp-1 | Merozoite surface protein-1 |
| msp-2 | Merozoite surface protein-2 |
| msp-7 | Merozoite Surface Protein-7 |
| PfHRP2 | Plasmodium Falciparum Histidine Rich Protein 2 |
| pLDH | Plasmodium Lactate Dehydrogenase |

**Abbreviations pertaining to model input or output.**

| | |
|---|---|
| aEIR | Annual Entomological Inoculation Rate |
| $AUC_{PL}$ | Area Under the Pathological Load Curve |
| BW | Patient Bodyweight |
| CL | Clearance |
| CV | Coefficient of Variation |
| FOI | Force of Infection |
| IC50 | Half Maximal Inhibitory Concentration |
| k | Elimination rate |
| Ka | Absorption rate constant |
| MOI | Multiplicity of Infection |
| MPL | Maximum Pathological Load |
| n | Slope of Concentration-Effect Curve |
| *n* | Slope Factor |
| PL | Pathological Load |
| PMR | Parasite Multiplication Rate |
| PRR | Parasite Reduction Ratio |
| Q | Intercompartmental Clearance rate |
| SD | Standard Deviation |
| Vd | Volume of Distribution |
| Vmax | Maximal Parasite Killing constant |
| z | Conversion rate |

# Chapter 1: Background to thesis.

## 1.1 Descriptive epidemiology of falciparum malaria.

*Plasmodium falciparum* is a protozoan parasite that is transmitted between humans by the bite of the female Anopheles mosquito. Following an infectious bite on a human from a mosquito, malaria parasites initially infect human liver cells where they mature, then infecting red blood cells. These infected red blood cells (iRBC) subsequently cause pathology in the host.

In 2017, *P. falciparum* annual mortality was estimated by the World Health Organization (WHO) to be 435,000 [1]. Most of this mortality (approximately 93%) occurred in African countries. Of the 435,000 deaths in 2017, roughly 61% were in children less than 5 years old [1].

Today's mortality figure for falciparum malaria is large, development and deployment of these measures has led to a large reduction of the estimated 1,000,000 annual deaths in 1990 [2]. It is worth noting that the WHO has observed the rate of reduction in mortality slowing over the last several years [1]. From a public health perspective, the "end-game" for falciparum malaria control is elimination of the disease: In 2015, the WHO produced a technical strategy towards this goal [3]. The ultimate aim of the strategy is achieving elimination by 2020 in 10 countries where malaria is currently transmitted, increasing to 35 countries by 2030, while continuing to reduce malaria mortality and incidence worldwide. The public health strategies undertaken by the WHO, national governments, charitable organizations and the wider scientific community in pursuit of this goal are multi-faceted, including but not limited to: Prompt diagnosis and treatment of cases [1, 3], preventative treatment strategies including targeted mass drug administration (MDA) programmes [4-6], vector control strategies [7-9], and development and deployment of novel malaria vaccines [10-12]. In short, the ultimate goal (elimination of the disease) is ambitious, but a diverse toolbox of strategies is available, and great progress has been made.

## 1.2 Drug treatment of falciparum malaria.

Key to the battle against malaria is prompt treatment of infections with artemisinin and partner drugs. An artemisinin derivative (artemether [AR], artesunate [AS], dihydroarteminisin [DHA]) is either given as a monotherapy (parenterally, and only in cases in severe malaria), or as an artemisinin combination therapy (ACT); note that monotherapy for severe malaria will always be followed up by a full course of ACT [1]. ACTs consist of an artemisinin derivative given in combination with a partner drug such as lumefantrine (LF), mefloquine (MQ), amodiaquine (AQ), sulfadoxine/pyrimethamine (SP), piperaquine (PPQ), chlorproguanil/dapsone or pyronaridine (PYN). *P. falciparum* presents in one of two forms: "uncomplicated" or "severe" malaria, and whether treatment is a monotherapy, or an ACT depends on the form of malaria presented by a patient.

In uncomplicated falciparum malaria, the patient has symptoms and detectable parasites in the blood stream but no signs of organ failure or dysfunction. Patients generally have a relatively mild fever, are conscious and capable of taking oral drug regimens; prompt treatment of uncomplicated malaria is associated with low mortality [13]. As such, total clearance of the parasites over the course of follow-up (several weeks) is the primary objective of treatment as this will reduce the risk of onward transmission. ACTs are the front-line drug of choice for uncomplicated malaria across most of the world; the artemisinin component rapidly clears parasites and the partner drug ensures eventual parasite clearance and therapeutic cure [14, 15]. The clinical consequence of this is that parasites rapidly reduce to undetectable levels immediately after ACT treatment initiation but are often not completely cleared during a 3-day course of drugs. Because partner drugs all possess substantial half-lives; infections surviving treatment may only recover to become detectable once partner drug

concentrations have decayed to ineffective concentrations - potentially weeks after treatment, termed “recrudescence”.

Patients with severe malaria present with one, or a combination, of four syndromes: Severe anaemia, respiratory distress, metabolic derangement and cerebral malaria [16, 17]. Patients are treated with parenteral artesunate [1], which rapidly kills parasites, but resolution of pathology lags behind parasite killing; case fatality rates are high even once patients have been admitted to the formal health system (typically between 5 and 12% [13] although studies that are designed to consider small sample sizes report lower mortality of ~2% [18]). The primary objective of treatment for severe malaria is the survival of the patient; complete parasite clearance is a secondary objective [19].

## 1.3 Antimalarial drug resistance.

The pernicious spectre of drug resistance is a key threat to effective treatment of malaria and the elimination effort. “Drug resistance” in the context of malaria is defined by the WHO as delayed parasite clearance following treatment with either a monotherapy or an ACT; note that resistance emerging to partner drugs carries a higher risk of treatment failure than resistance to artemisinin [1]. Anti-malarial resistance is currently “partial”, rather than “full” – i.e., treatment becomes less effective, not ineffective. Resistance (and known mechanisms) of parasite resistance to previous drugs (i.e. chloroquine [CQ]) and current partner drugs (particularly Piperaquine [20]) has been well-documented [21-23] in the Greater Mekong Subregion (GMS), where extensive multi-drug resistance has forced the hand of drug elimination efforts [24, 25] and it has now become a priority to eliminate the disease before *P. falciparum* becomes untreatable with the existing anti-malarial catalogue [1].

In uncomplicated malaria, reduced parasite clearance due to resistance (either to artemisinin or a partner drug) leads to parasites being able to recover i.e. “recrudesce”) and continue causing disease in a patient. The consequences of artemisinin resistance for severe malaria are more immediate: Slower parasite clearance, or the inability to kill parasites before they sequester, can cause patient death. In either case resistance will lead to increased morbidity and mortality from malaria. Concerningly, resistance is self-propagating: When parasites become resistance to current drugs, treatment is less effective and so the resistant strains can spread more easily. Finally, resistance causes economic consequence through the cost of healthcare, loss of workforce productivity, and requiring the development of new drugs [26].

## 1.4 Clinical trials of uncomplicated malaria (Therapeutic Efficacy Studies).

The WHO consider an ACT to be “failing” in a country when failure rates of treatment for uncomplicated malaria within that country reach 10% [3] (i.e., that 10% of patients treated with an ACT fail to fully clear parasites over the course of follow-up). To maintain clinical effectiveness of drugs, the WHO recommends that countries change their first-line drug of choice at that point [27]. It follows that clinical evidence of drug efficacies are required to enact this policy and consequently, drug efficacy estimates are constantly monitored through therapeutic efficacy study / studies (TES), where a group of patients with uncomplicated malaria are recruited, treated, and followed up over several weeks [28].

TES monitor patients for extended periods of time post-treatment to ensure recrudescent infections are detected. Duration of follow up depends on the half-life of the drug being assessed [28, 29] , usually between 4 and 6 weeks (28 to 42 days) [14], sometimes extended to 9 weeks for research purposes. Patients are followed up at regular intervals to check for the presence of malaria parasites. Patients who have detectable parasites during this follow-up period are considered to have a “recurrent infection” or, equivalently, “recurrence”.

A critical problem of long follow-up periods is that in areas of high transmission, new infections (called "reinfections") may be inoculated into patients by mosquitoes. These reinfections must be distinguished from recrudescence to allow accurate estimates of drug efficacy. This is not a trivial problem: Annual entomological inoculation rates (aEIR) of malaria, a measure of malaria exposure in a population, are typically >10 and >100 per patient in areas of moderate to high transmission, respectively. Moderate to high transmission sites are preferred for TES as morbidity from malaria is high, so these trials cover the most at-risk patient populations and patient recruitment is straightforward. Reinfections occurring late during TES follow-up, when partner drug concentrations have reduced from their initial levels, are not indicative of a failing partner drug. An example of the dynamics of malaria parasites in a patient post-treatment are shown graphically in **Figure 1.1** .Consequently, the ability to distinguish between reinfections and initial infections that are not cleared by treatment (termed "recrudescence / recrudescent") is critical to obtain accurate failure rate estimates of ACTs. In theory, if it were possible to perfectly classify recurrent infections as reinfections or recrudescent infections, accurate drug failure rate estimates could be obtained simply. This is possible through sequencing the entire malaria genotype of the patient's initial sample and recurrent sample, but this is (currently) impractical to perform for each patient in a TES as such trials occur frequently and with large numbers of patients.

A variety of methods have been proposed that take a blood sample and genotype of a patient's initial infection and any recurrent infection, subsequently using any difference between genetic profiles to classify a recurrent infection, with a "match" between genotypes indicating a recrudescence, and a "mis-match" a reinfection– this process is called "molecular correction", or, equivalently, "PCR correction". In 2008 the WHO and collaborators reached a consensus of using three length-polymorphic markers to achieve this [14]. Alternative markers are available, including microsatellites (widely used by the Centers for Disease Control and Prevention [CDC]) [30, 31], single nucleotide polymorphisms (SNPs) [32, 33] (not explored in this thesis), and markers obtained from next-generation amplicon sequencing (AmpSeq) [34-36]. Each choice of genetic marker carries their own perils that may induce misclassification of infections – for example, reinfections and initial infections sharing alleles by chance, inability to detect low frequency clones, and errors in correctly reading the base-pair length of a given allele. Length-polymorphic markers are the subject of **chapter 3** of this thesis, microsatellite markers are the subject of **chapter 4** and next-generation amplicon sequencing is the subject of **chapter 5**. A detailed description of each marker and their properties that can result in misclassification of recurrent infections is provided within these chapters, but a summary is provided in **Table 1.1**. Note that the terms "marker" and "locus" are often used interchangeably in the literature. In this thesis, the term "marker" will be used to refer to the choice of methodology i.e. length-polymorphic markers, microsatellite markers and AmpSeq markers, while the term "locus" or "loci" will be used to refer to specific genetic regions, (i.e., *msp-1*, *msp-2*, and *glurp*). Thus, the length-polymorphic marker methodology concerns itself with genotyping the loci *msp-1*, *msp-2* and *glurp*, the microsatellite marker concerns itself with genotyping a variety of microsatellite loci and the AmpSeq marker methodology concerns itself with genotyping a variety of AmpSeq loci.

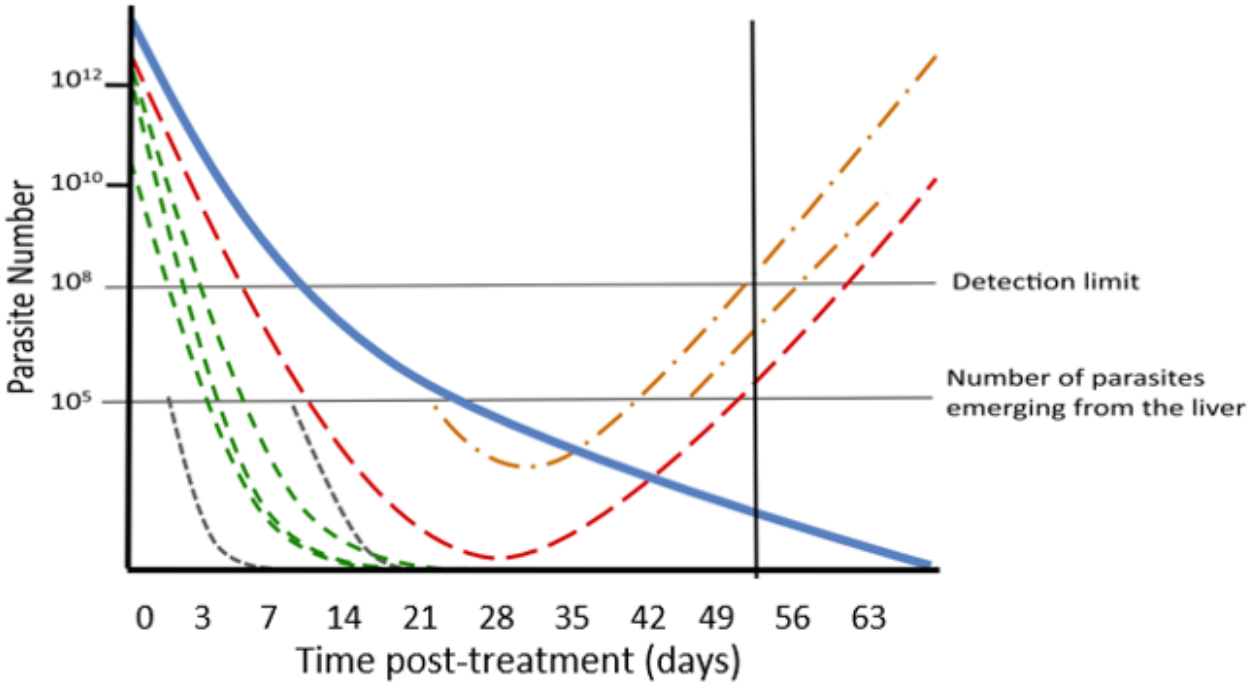



**Figure 1.1** : A graphical representation of the parasite dynamics produced by a mPK/PD model for a hypothetical patient post-treatment. Drug concentration is shown by the solid blue line; it's change over time is determined by the specifics of the model and the patient's PK parameters. Initial clones can be cleared by the drug (green lines) or recrudesce at a later point (red line). New infections (termed "reinfections") may emerge from the liver at a frequency of $10^5$ parasites during follow-up and these can be cleared by the drug (grey dashed lines) or rise in number and eventually become detectable (the solid black line at 53 days denotes the point at which parasites are of a sufficient count to be detected in a blood sample during follow-up). Figure layout adapted from [37].

**Table 1.1 A** brief summary of the types of genotyping for TES explored in this thesis, their proposed advantages, potential disadvantages, and important examples of their use.

| Marker | Notable users | Examples | Advantages | Disadvantages | Chapter of this thesis |
|---|---|---|---|---|---|
| Length-polymorphic markers | WHO consensus methodology | [14, 38] | Well-characterized, generally diverse markers. *Msp-1* and *msp-2* exist in families so family-specific primers can be used to improve resolution of results. Clear guidelines exist for use from WHO. | Alleles below a certain frequency are ignored as noise in the PCR process (to avoid including false alleles that are "noise"). Shorter fragments are amplified preferentially (i.e., longer alleles are harder to detect). Failure to detect alleles in either the initial or recurrent sample means recrudescent infections and reinfections may be misclassified as the other. | 3 |
| Microsatellite markers | CDC | [30, 39, 40] | Lack of immune selection on microsatellite markers. | Alleles below a certain frequency must be ignored in the PCR process (to avoid including false alleles that are "noise"). Microsatellite allele length may be mis-read due to polymerase slippage in the PCR process. | 4 |
| Amplicon Sequencing | Novel methodology still in experimental use / developmental stage | [36, 41] | Able to detect low frequency genetic signals | Novel method without (at the time of writing) clear procedures for best use. Requires use of next-generation technology. | 5 |

*TES: Therapeutic efficacy study, WHO: World Health Organization, msp-1: merozoite surface protein-1, msp-2: merozoite surface protein 2, glurp: glutamate rich protein, PCR: polymerase chain reaction, CDC: Centers for Disease Control and Prevention*

Additionally, there are two inherent sources of error in TES, independent of the choice of marker:

A patient who fails to clear their initial infections may have a reinfection that becomes detectable before the recrudescent clone reaches a detectable level; ethically, that patient must be treated and so is removed (or "censored") from the study before the recrudescence can be observed (**Figure 1.2** (A)).

A patient who fails to clear their initial infection may have that infection persist at a low-lying level, below the limit of detection of light microscopy, such that parasites are never detected during follow-up; the frequency of this event is influenced by the duration of follow-up in the trial, i.e. the longer the follow-up, the less likely it is to occur (**Figure 1.2** (B)).

These issues have led some contributors to suggest the use of molecular correction may be undesirable and non-corrected estimates are better [42, 43],but such an approach would be likely to result in over-estimation of failure rates in the moderate to high transmission areas in which TES are conducted. Thus, a critical issue stands: The true failure rate of the drug cannot be known in vivo (due to errors in correctly classifying recurrence inherent to all approaches) and as such, neither can be accuracy of any given method.

If failure rate estimates are inaccurate there are two critical consequences: Firstly, If current methods under-estimate the true failure rate, we fail to detect resistance to anti-malarial drugs at an appropriate time, leading to increased morbidity, mortality, economic consequence and further spread of resistance. Secondly, If current methods over-estimate the true failure rate, first-line treatments are changed un-necessarily with great economic cost and public health bodies / governments needlessly "waste" an effective treatment.

The potentially dire consequences of inaccurate failure rate estimates lead to the first critical aim of this thesis:

**To develop a methodology that can accurately calculate true failure rates of ACTs in uncomplicated malaria TES, and, against that gold standard figure, compare the accuracy and utility of a range of current and proposed methods to estimate ACT failure rates, and ways in which the accuracy and usage of these methods may be optimized.**

## 1.5 Clinical trials of severe malaria.

The key goal of treatment for severe malaria is patient survival, particularly in the first 48 hours post-treatment (in contrast to uncomplicated malaria where total clearance of parasites is the key objective). Consequently, TES are not conducted for severe malaria and clinical trials of severe malaria generally have a different goal: To explore new treatments or treatment strategies with the aim of improving patient survival rates; notable examples are trials comparing quinine with artesunate that supported eventual policy change replacing quinine with AS as the first line therapy for severe malaria [44, 45].

Parasite clearance rates are a commonly used clinical outcome measure to compare outcomes of antimalarial treatment regimens for severe malaria [46, 47]. However, parasite clearance rates correlate poorly with disease outcome in severe malaria. Large trials comparing intramuscular AR with quinine in African children showed more rapid parasite clearance with AR but no difference in case fatality [48, 49]. With parenteral AS, parasite clearance rates are not different in patients dying from severe malaria compared to survivors (results cited in [50]). There are two potential explanations why parasite clearance is an unsuitable outcome measure in severe malaria: Firstly, parasite clearance rates following treatment for uncomplicated malaria appear to mainly reflect host immunity rather

than drug effectiveness [51-53] so may be a poor metric of overall drug effectiveness. Secondly, parasite clearance rates are measured on circulating parasites [51] whereas non-circulating, sequestered parasites are generally assumed to be responsible for most clinical symptoms, pathology and deaths associated with severe malaria [16]. Given that patient survival is the key clinical outcome, death (and thus, comparison of mortality rates between trial arms) would be the ideal outcome of interest to use to compare the effectiveness of treatments for severe malaria. However, a trial would need to be unfeasibly large to have enough statistical power using death as the outcome of interest, and given that it is seemingly impossible, to obtain accurate estimates of sequestered parasite densities in vivo (though total parasite biomass can be estimated by measuring *P. falciparum* histidine rich protein 2 [PfHRP2] [16, 54]) , parasite clearance rates continue to be used as the outcome of interest despite the noted flaws in this approach.

Consequently, the second critical aim of this thesis is:

**To develop a methodology to quantify the pathology of sequestered parasites in severe malaria and use this methodology to identify rational drug dosing regimens for treatment of severe malaria with AS and quantify the likely impact of AS drug resistance for treating severe**

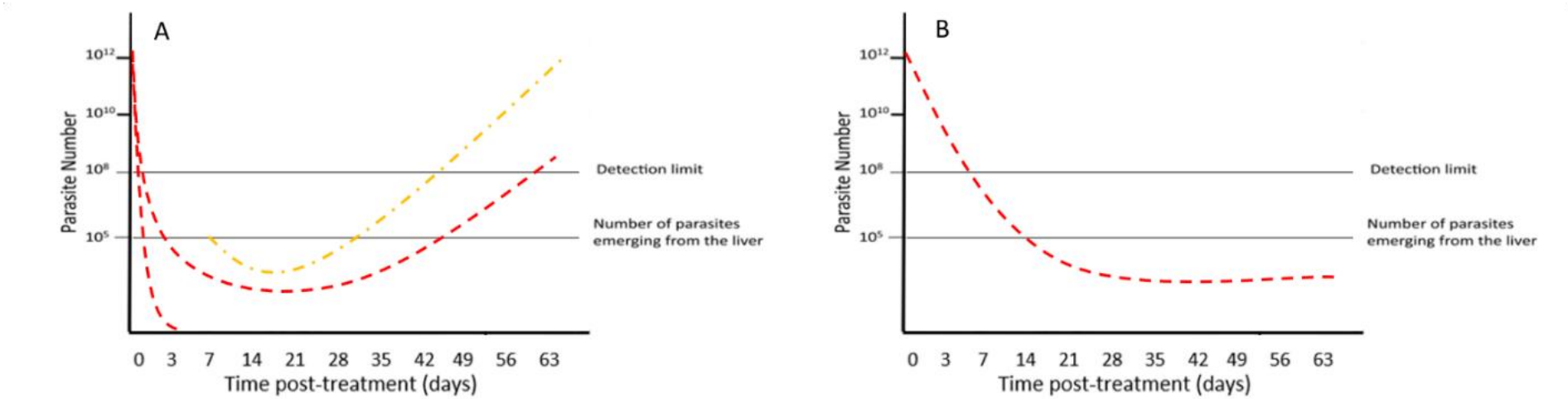



**Figure 1.2** (A) shows a hypothetical patient with two initial infections (dashed red lines) and a reinfection (dashed orange line). An initial infection recrudesces over a long period of time, becoming detectable at day 63, but a reinfecting clone emerging from the liver on day 7 more quickly increases in frequency and becomes detectable on day 45. Consequently, this patient would be removed from follow-up due to the reinfection before the recrudescence can be observed. (B) shows a hypothetical patient with a single initial infection who has no reinfections during follow-up. Treatment fails to clear their initial infection, but the initial infection does not reach detectable levels during follow-up and so parasites are never detected. Figure layout adapted from [37].

## 1.6 Brief summary of mathematical modelling techniques.

The premise of a mathematical model is to use a series of mathematical statements to describe an otherwise non-mathematical system or situation. The first example of modelling applied to malaria was by Ronald Ross in 1910 [55], who used a series of differential equations to describe malaria transmission. Over the last 100 years, particularly following the advent of computer modelling to handle complex calculations, the goals and achievements of models have grown bolder, and there now exists a broad range of applications for models in the context of falciparum malaria, and a rich tapestry of accomplishments [56]. Notable examples include modelling strategies for coping with the emergence of resistance [23], investigating important epidemiological parameters such as the basic reproduction number, endemic equilibrium, and aEIR [57-59], analysing co-morbidity of malaria and Human Immunodeficiency Virus (HIV) or malaria and tuberculosis (TB) [60, 61], the impact of hypothetical or novel malaria vaccines [62, 63], and cost-effectiveness studies of malaria control policies [64].

The overarching goal of this thesis is to interface computational/mathematical modelling methods and anti-malarial drug trials, with the aim of improving patient outcomes either directly (i.e., investigating and providing supporting evidence for novel treatments) or through contributing to the malaria elimination efforts. Essentially, it will focus on simulating anti-malarial trials in a variety of scenarios with key advantages: Parameters that cannot be measured in vivo can be incorporated via sensitivity analysis, errors or flawed methodology can be accounted for (i.e., misclassification of results or inappropriate outcome measures), and the ability to simulate extremely large trials and a variety of treatments without regard to cost or time. Additionally, a modelling methodology can examine matters that would be ethically difficult in vivo (i.e., novel treatment regimens or drugs with unknown efficacy and/or safety) or simply impossible (i.e., simulating resistance to a drug that has never had resistance emerge in vivo, allowing for analysis of the consequences of resistance).

### 1.6.1 History of within-host modelling of malarial parasite dynamics

A wide variety of models have been constructed and deployed to model malaria parasite intra-host dynamics – more plainly, what happens over time to parasites / infected red blood cells (iRBCs) in the host. Broadly speaking, the critical aim of models of intra-host parasite dynamics is to mathematically describe and replicate observed *in vivo* parasite dynamics. Of key historical importance are trials of so-called malaria therapy for neurosyphilis, where patients presenting with neurosyphilis were experimentally treated with *P. falciparum* and *P. vivax* in the early 1900s, under the hypothesis that the fever induced by malaria would alleviate the symptoms of neurosyphilis [65-69]. Data on parasite dynamics that was produced by these trials has been used to inform mathematical models through the data they provided on parameters including infection length, parasite density, and the shape of the parasite growth curve. With this data, mathematical models can be adjusted to "fit"[70] *in vivo* dynamics of *P. vivax* (without real data to validate against, creating a model would be a fairly futile endeavour). In more simple terms, these trials provided the *in vivo* evidence base that models have, historically, sought to mathematically describe. These data were first used to validate a model by Molineaux et al. [71], who noted that previous models (not validated against clinical data) lacked realism – this approach was then used to inform future models by a range of authors [62, 72, 73]. More recently, advanced statistical methods have sought to better describe select parasite dynamics (i.e., duration of infection) from longitudinal genotyping studies [74]

On the purpose of modelling parasite dynamics, I reference Molineaux and Dietz [75], who describe intra-host malaria models as having three purposes: i) explaining observed data with biologically plausible assumptions, ii) predicting the likely outcome of interventions and iii) estimating parameters that cannot be directly observed , noting that these purposes often overlap (as an example, Dietz, Raddatz and Molineaux developed a mathematical model to describe the first wave of *P. falciparum*

parasitaemia following infection, fitting four model parameters that were not directly observable, with the goal of using the model to predict the likely impact of a hypothetical vaccine [76].

Ultimately, available models differ widely in their goals and execution. I will not attempt an exhaustive review here (I describe the approach chosen to model intra-host parasite dynamics in this thesis at length in **chapter 2**), but I would refer readers to [70, 75] and the introduction of [77] for reviews of existing models. I would add that commentators have noted that, in general, models require a large number of parameters, many of which are difficult or impossible to measure *in vivo* and must have their values assumed for the purposes of modelling [70]. Many of these parameters relate to the dynamics of host immunity to malaria, which is still not fully understood [78] - note that observed parasite dynamics reflect the acquisition of human immunity to *P. falciparum* and as such are likely to be highly influenced by a process known as *var* switching, whereby parasite *var* genes (which are responsible for coding Plasmodium falciparum erythrocyte membrane protein 1 (PfEMP1)) switch between ~60 members and allow *P. falciparum* to avoid the host immune response[79, 80]. This switching operates over multiple parasite generations [79] (compared to short term sequestration cycles which operate over 48 hours); and may be a major driver of malaria clones rapidly fluctuating in density. Additionally, there is difficulty in accurately measuring parasite numbers *in vivo* due to iRBCs sequestering as part of the parasite's 48 hour life cycle [16, 81]. Consequently, there is a clear imperative to test the validity of model assumptions and the impact of altering parameters (particularly ones that cannot be measured *in vivo*) on model output and results. Throughout this thesis, I attempt to ensure that the impact of such parameters is investigated. **1.6.2 Pharmacokinetic / Pharmacodynamic modelling.**

Particularly important to this thesis is the concept of Pharmacokinetic (PK) / Pharmacodynamic (PD); PK/PD modelling. Pharmacokinetics (PK) describes the eventual fate of a drug in a patient. In other words, a patient's PK parameters describe the rate at which the drug the drug is absorbed, processed, and eventually eliminated from the body. Pharmacodynamics (PD) describes the effect of a drug on its target (in this case, the effect of drugs on malaria parasites), and specifically the relationship between drug concentration and drug effect. The goal of traditional PK/PD modelling is to investigate an in vivo population of patients and construct a model to estimate PK and PD parameters in that population to a high degree of accuracy. Mechanistic PK/PD (mPK/PD) models are distinct from traditional PK/PD models - they are generally calibrated with parameters obtained from these traditional models (although novel / arbitrary parameters can be used) but use a series of calculations to convert a given set of PK/PD parameters into a quantitative description of the change in parasite clone(s) over time in a patient following treatment (a number of thorough reviews exist in the literature [82-84]). In other words, whereas traditional PK/PD modelling uses drug concentrations and parasite dynamics in a population of patients to describe PK/PD parameters in that population, mPK/PD modelling uses a set of PK/PD parameters to simulate the parasite dynamics post-treatment in a hypothetical patient population.

mPK/PD models for anti-malarial drugs operate in three general steps:

1: Calculate change in drug concentration over time in a patient, given that patient's PK parameters.

2: Calculate the relationship between drug effect and drug concentration, given a parasite clone's PD parameters.

3: Combine steps 1 and 2 to calculate the change in parasite count over time for a given patient following drug treatment.

mPK/PD modelling has been used extensively to address a wide range of research questions in the malaria field; a non-exhaustive summary is displayed in **Table 1.2**.

This thesis will address critical aim 1 by using mPK/PD modelling approaches to simulate parasite dynamics in populations of patients with uncomplicated malaria treated with ACTs and followed up

during the course of a TES. The modelling approach will allow the true fate of each patient's initial infection (and thus, the true failure rate of the population) to be known. Consequently, the accuracy of a range of genotyping methods can be investigated, and their best use optimized.

The results presented in **chapters 3-5** are all generated using a general two-stage process:

1: Simulate intra-host *P. falciparum* dynamics in a population of patients post-treatment using mPK/PD models

2: By allocating genetic data to each parasite clone simulated in step 1, calculate the genetic signals detected from a patient's blood samples during a therapeutic efficacy study (TES) and analyse these to generate drug failure rate estimates (and subsequently analyse the accuracy of these estimates).

The critical difference between each chapter is the genetic markers used. **Chapter 3** investigates the World Health Organization (WHO)/Malaria for Medicines Venture (MMV) recommended length polymorphic markers merozoite surface protein-1 (*msp-1*), merozoite surface protein-2 (*msp-2*) and the glutamate rich protein (*glurp*). **Chapter 4** investigates the use of microsatellite markers, and **chapter 5** investigates the use of next-generation amplicon sequencing (AmpSeq). As such, the second step in this process varies between chapters in both simulation and analysis of data. However, the first step – simulating parasite dynamics over time – occurs using the same core methodology – mPK/PD models of a set of ACTs in each chapter. The detailed methodology behind those mPK/PD models is described fully in **Chapter 2**.

Critical aim 2 will be addressed in **chapter 6** by using a distinct mPK/PD model of artesunate monotherapy in a population of patients. A model will be constructed that accounts for sequestered parasites and develops novel metrics to quantify the burden of disease associated with these parasites. This model will then be used to investigate outcomes with a selection of treatment regimens and compare outcomes between sensitive and resistant parasites.

507 **Table 1.2** Brief summary of published work utilizing mechanistic PK/PD models in a malarial context.

| Authors | Issue | Research question | Practical limitations of in vivo studies | References |
|---|---|---|---|---|
| Hodel et al. | Optimal age/weight-based dosing regimens | Malaria drugs have a therapeutic index of around 2 but must, for practical reasons, be deployed to treat patients in weight or age-bands. Weight often varies >2 fold in these bands. As such, a significant fraction of patients will be under- or over-dosed on a mg/Kg basis. What are the quantitative consequences of this dosing? | Correctly dosing a band of patients whose weight varies >2 fold with a drug with a therapeutic index of around 2 is impossible in vivo. Consequently, heavier patients in the band will be under-dosed and/or lighter patients will be over-dosed. | [85] |
| Simpson et al.; Hodel et al.; Kay, Hodel & Hastings; Dini et al. | Optimal dosage amount and treatment regimes | How does dosage of drug (in mg/Kg) affect therapeutic outcome? | Dose escalation trials are expensive and generally small. Useable data can be extracted post-deployment by meta-analysis from drugs which are given in weight or age bands but must be processed with a modelling approach. | [85-88] |
| Hodel et al.; Challenger et al. | Robustness of drug regimens to poor adherence | Malaria patients often have poor adherence and do not complete a full course of treatment: what is the impact in terms of drug failure rates? | Ethically extremely difficult or impossible to deliberately under-dose patients. | [85, 89] |
| Hastings, Hodel & Kay; Kay & Hastings | Threat posed by the emergence of drug resistance | How will the spread of resistance affect patient cure rates and public health interventions such as mass drug administration? | Impossible to investigate in vivo until resistance has arisen – for many drugs, this has yet to occur. | [15, 90] |
| Kay & Hastings; Stepniewska & White | Possible sources of de novo resistance mutations | Where in the malaria lifecycle are de novo resistance mutations most likely to arise? | Resistance often occurs extremely infrequently, so this is difficult to directly observe in vivo. | [90, 91] |
| Klonis et al. | Inferring drug resistance patterns | What are the physiological patterns of drug resistance? | Sensitivity of parasites to drugs can be measured in vivo but must be placed in a mPK/PD framework to investigate likely impact on therapeutic outcome. | [92] |

| | | | | |
|---|---|---|---|---|
| Jaki et al.;<br>Jones et al. | Simulating data used for clinical trial analysis | Analysis of drug clinical trials show that different methods and assumptions generate different estimates of failure rates: which is the most accurate and/or robust method? | The true failure rate of drugs in real clinical trials is unknown so it is impossible to fully quantify the accuracy of analytical methods and/or whether biases occur. | [37, 93] |
| Jones et al. | Simulating pathology caused by sequestered malaria parasites | Severe malaria pathology is generally caused by sequestered parasites. These cannot be directly observed in patient blood samples. What is their quantitative relationship to level of pathology? | Impossible to directly quantify sequestered parasites in vivo. | [94] |
| Jonhston et al.;<br>Geradin et al. | Simulating the use of drugs in Mass Drug Administration (MDA) campaigns | What is the optimal role of MDA campaigns in malaria control and elimination programmes? | MDA programmes are extremely expensive and difficult to deploy. Comparisons between studies are limited because of the large number of differences between them | [95, 96] |
| Nguyen et al. | Drug deployment policies and selection for resistance | Do differential drug deployment policies (e.g. sequential, rotations, mosaics) have different impacts on the evolution of drug resistance? | It would be virtually impossible to implement different polices for drug deployment in a sufficient number of locations to allow effective comparison of policies | [97] |

508 *MDA: Mass Drug Administration*

# Chapter 2: Mechanistic Pharmacokinetic/Pharmacodynamic (mPK/PD) modelling to simulate parasite dynamics post-treatment with artemisinin-based combination therapies (ACTs).

All computer-based modelling in this thesis was conducted using the programming language R [98]. Due to the time over which this thesis was written, the version of R used varied between 3.1.2 and 3.6.0.

The mPK/PD models used in **Chapter 3 – Chapter 5** are for Dihydroartemisinin-Piperaquine (DHA-PPQ), Artemether-Lumefantrine (AR-LF; commonly referred to in the literature as AL) and Artesunate-Mefloquine (AS-MQ). Other frontline ACTs, i.e. Artesunate-Amodiaquine (AS-AQ), Sulfadoxine/Pyrimethamine (SP) and Artesunate-Pyronaridine (AS-PYN) were not explored for the following reasons. Both the parent form and a metabolite of AQ have antimalarial activity and both are eliminated independently (e.g. [99]); there is currently no robust mPK/PD model [100]. SP exhibits strong synergy between the Sulfadoxine and Pyrimethamine components which has made a robust mPK/PD model difficult to create [101].Finally, PYN is a relatively new drug; thorough AS-PYN safety and efficacy studies are recent [102] and PK parameters are relatively limited and mostly derived from plasma assays (PYN concentrates in red blood cells and so these parameters are not reliable for use in mPK/PD models [103]).

mPK/PD models for DHA-PPQ, AR-LF and AS-MQ, however, have been explored, calibrated and validated in-depth and used by a variety of groups to explore a variety of research questions (see the summary table **Table 1.2** in **Chapter 1**). The model structure I use in this thesis is derived from the work of Katherine Kay (who also published under her maiden name, Winter), Eva Maria Hodel and Ian Hastings [85, 104, 105], and is described in full in [104]. I was provided with a set of code (R language, version 3.1.2) that relates to the publication in reference [104] by Katherine Kay to calculate (for each ACT) drug concentration over time and resulting parasite kill rates over time for any given set of PK/PD parameters (see below). The mechanistic simulations of drug concentration and parasite kill rates with a given set of parameters are described in those papers. I do not detail these methods here due to their length and complexity; rather this chapter details the way I parameterized these models and used them to generate data; expansions to the initial models were made by me (see later) to allow for reinfections to occur during the follow-up period and for patients to have polyclonal infections.

**Chapters 3, 4** and **5** investigate the accuracy of genotyping methods used in clinical trials. To do so (as explained fully in those chapters), they calculate which genetic signals are detectable in conventional finger-prick blood samples by genotyping a selection of loci; as later described this is based on the relative and absolute densities of parasite clones as well as specific characteristics of certain loci that are used as genetic markers (for example, *msp-1* and *msp-2* alleles occur in different families and this affects detectability of the alleles [106]). The purpose of this thesis is to investigate genotyping methods, not to provide a comprehensive study of PK/PD parameters (though see [87, 104, 105] for previous use and detailed discussion of the model used). Parasite dynamics (see **Figure 1.1** in **Chapter** 1) can be generated by other means. Examples include randomly assigning biomass to clones on both the initial day and a day of recurrence using a computer (i.e. [107]), using hazard functions to simulate time to recrudescence or time to reinfection for a range of patients [108], or sophisticated Markov Chain Monte Carlo models [109]. The reason for using a mPK/PD approach in this thesis was to increase realism and lay a foundation for future work – the methodology is adaptable for any group with their own set of PK/PD parameters, genetic data and other relevant information (i.e., multiplicity of infection [MOI], transmission intensity, and allelic distributions of their chosen marker) to investigate the accuracy of their failure rate estimates.

## 2.1 Mechanistic PK model to model drug concentrations over time.

Throughout this thesis, PK parameters for artemisinins and partner drugs remain the same between chapters **3**-**5**. These parameters determine drug concentration over time profiles for each patient in the population. The parameters for any given patient are drawn from a distribution described by the mean value and the coefficient of variation (CV). PK parameters are described in **Table 2.1**; the parameter is assumed to be normally distributed if the CV is < 0.5, and log-normal if the CV is ≥0.5.

There is enormous variation in PK parameters described in the literature (see [85]), though this is not surprising, given that PK studies are conducted in different populations, demographics and locations and different groups use different error models to describe additional variation within their population. For this thesis, I was not trying to replicate any particular population of patients but rather use a mPK/PD model to generate a large, "general" population in silico that could be used to test genotyping methods in trials. To achieve this, I selected mean values from the literature that were generally intermediate within the range of reported values and accompanied them with (generally large) CVs that would achieve a wide range of drug concentration over time profiles for the population.

The mPK/PD model used to generate these drug concentrations (specifically, the PK portion of the model) uses parameters originally derived from in vivo, traditional PK studies (**Table 2.1**). Such studies will typically provide a "visual predictive check" (VPC) of their model – which, simply, is a method of checking that the PK parameters and the error model generated by a PK study can then reproduce the variation in the data that was used to generate the parameters in the first place [110]. Although the goal of the mPK/PD modelling here is to generate a general population for further analysis, not to replicate the VPC of any given study (see above), comparisons to VPCs are useful to check that the range of drug concentrations produced by the mPK/PD model are within reasonable ranges and assure readers that this method of generating parasite dynamics in silico is robust.

**Table 2.1** PK Parameter summary. A summary of the PK parameters used to simulate parasite dynamics post-treatment, adapted from Hodel et al. [85]. The table shows means with coefficient of variation in brackets.

| **Drug** | **Dihydroarteminisin-Piperaquine (2 compartment model)** | | **Artesunate-Mefloquine** | | | **Artemether-Lumefantrine** | | |
|---|---|---|---|---|---|---|---|---|
| | **DHA** | **PPQ** | **AS** | **DHA** | **MQ** | **AR** | **DHA** | **LF** |
| **Vd (L/kg)** | 1.49 (0.48)[85, 104] | 346 (0.93)[111] | 7.1 (0.94)[104] | 1.49(0.48)[104] | 20.8(0.38)[104] | 46.6(0.82)[85] | 15(0.48)[85, 104] | 21(2.63)[85, 104] |
| **$Vd_1$ (L/kg)** | - | 443 (1.70)[111] | - | - | - | - | - | - |
| **ka (/day)** | - | 11.2 (2.17)[111] | 252(1.12)[104] | - | - | 23.98(0.68)[85, 104] | - | - |
| **z (/day)** | - | - | 30.96(0.362)[104] | - | - | 11.97(0.65)[85, 104] | - | - |
| **$Q_1$(L/day/kg)** | - | 69.7(1.01)[111] | - | - | - | - | - | - |
| **k (/day)** | 19.8(0.23)[85, 105] | 0.02*[111, 112] | - | 25.4(0.23)[104] | 0.053(0.63)[104] | - | 44.15(0.23)[85, 104] | 0.16(0.05)[85, 104] |

*PK: Pharmacokinetic, BW: Patient bodyweight, DHA: Dihydroartemisinin, PPQ: Piperaquine, AS: Artesunate, MQ: Mefloquine, AR: Artemether, LF: Lumefantrine, Vd: Volume of Distribution (central compartment for PPQ), $Vd_1$: Volume of Distribution (peripheral compartment), $Q_1$: Intercompartmental clearance (central-peripheral 1), ka: Absorption rate constant, z: Conversion rate of AR/AS into DHA, - : No data / not applicable.*

** elimination rate for PPQ is calculated from clearance (CL) / Vd. CL is not shown here but is 4.5 * $BW^{0.75}$ as in [112]; This means that elimination rate varies with body weight ( a common PK observation) so the value presented here is illustrative and represents a bodyweight of 42kg (the median bodyweight in previous studies[111, 112]). Partner drug IC50 values are not shown here; they vary between and within chapters, see individual chapters for these values. Piperaquine (PPQ) here follows a two-compartment model as described in Kay, Hodel & Hastings[111]. Patient bodyweight (BW) in all simulations was drawn from a uniform distribution between 45-75 kg and is involved in the calculations for PPQ parameters (see[111, 112]). Square brackets are citations in support of the parameter values*

Drug dosing in all models in **chapters 3-5** followed the dosages and timings shown in **Table 2.2**. Dosing values have been chosen for consistency with the recommended dosing regimen published by the WHO in 2015 [19], noting that in practice drug formulations are a fixed weight and so doses are given banded by weight (or age; see references in **Table 1.2** for more detailed discussion of such dosing), but in the mPK/PD models presented in this thesis each patient is given a precise dose according to their body weight.

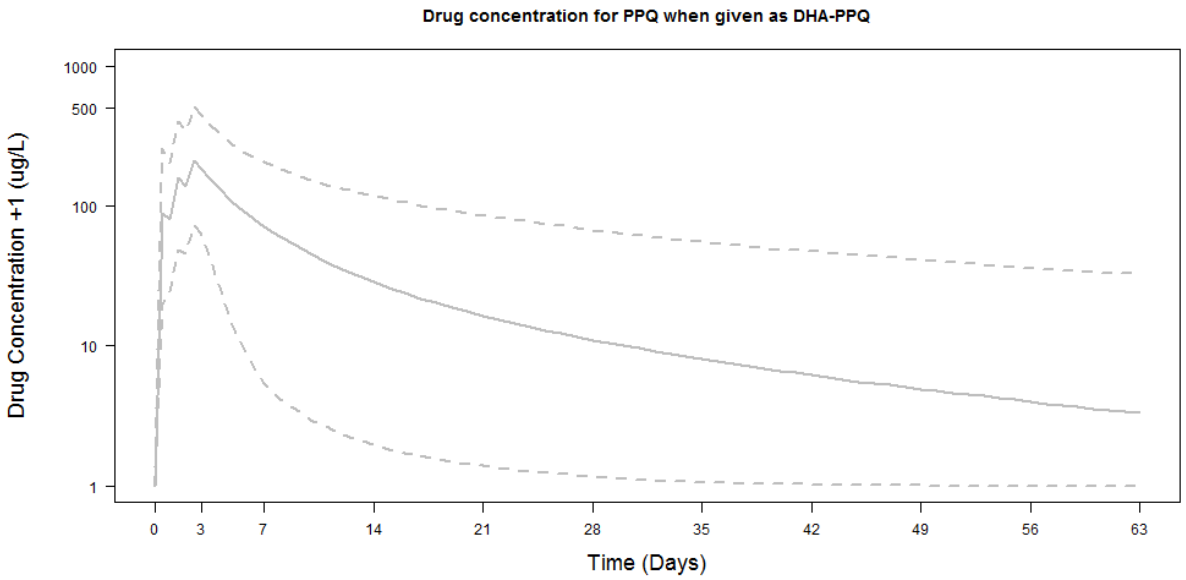
**Table 2.2** Drug dosing of the artemisinin and partner drug components of the ACTs for the mechanistic simulation of DHA-PPQ, AR-LF and AS-MQ

| **Drug** | **DHA-PPQ** | | **AR-LF** | | **AS-MQ** | |
|---|---|---|---|---|---|---|
| | **DHA** | **PPQ** | **AR** | **LF** | **AS** | **MQ** |
| **Dose at 0 days (**mg/kg) | 4 | 18 | 1.7 | 12 | 4 | 8.3 |
| **Dose at 0.5 days (**mg/kg) | | | 1.7 | 12 | | |
| **Dose at 1 days (**mg/kg) | 4 | 18 | 1.7 | 12 | 4 | 8.3 |
| **Dose at 1.5 days (**mg/kg) | | | 1.7 | 12 | | |
| **Dose at 2 days (**mg/kg) | 4 | 18 | 1.7 | 12 | 4 | 8.3 |
| **Dose at 2.5 days (**mg/kg) | | | 1.7 | 12 | | |

*DHA: Di-hydroartemisinin, PPQ: Piperaquine, AR: Artemether, LF: Lumefantrine, AS: Artesunate, MQ: Mefloquine. Dosages listed are mg/kg, e.g., for a 45kg patient, a dose of 180mg of DHA would be given at each interval.*

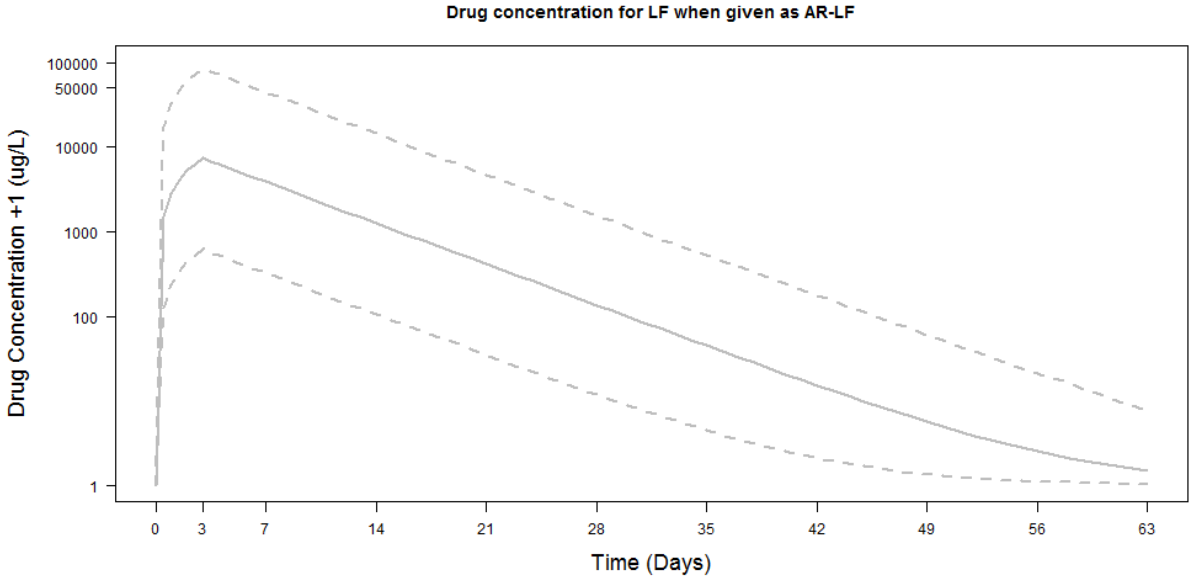
### 2.1.1 Partner drug concentrations over time.

The drug concentration over time profiles produced by the model for PPQ (assuming a two-compartment model for PPQ), LF and MQ, when given as DHA-PPQ, AR-LF and AS-MQ (i.e., only the concentration of the partner drug) produced for 5,000 patients with the parameters shown in **Table 2.1** and the dosing regimen in **Table 2.2** are shown in **Figure 2.1**, **Figure 2.2** and **Figure 2.3** respectively.




609

610 **Figure 2.1** PPQ concentration (in µg/L) over time for a population of 5,000 patients, treated with DHA-
611 PPQ parameterized as in **Table 2.1** with drug dosing as in **Table 2.2**. The solid line is the median
612 population concentration at each day and the dashed lines are the 5% and 95% quantiles. The figure
613 follows patients for 63 days (the maximum length of patient follow-up investigated within this thesis).




614

615 **Figure 2.2** LF concentration (in µg/L) over time for a population of 5,000 patients, treated with AR-LF
616 parameterized as in **Table 2.1** with drug dosing as in **Table 2.2**. The solid line is the median population
617 concentration at each day and the dashed lines are the 5% and 95% quantiles. The figure follows
618 patients for 63 days (the maximum length of patient follow-up investigated within this thesis).

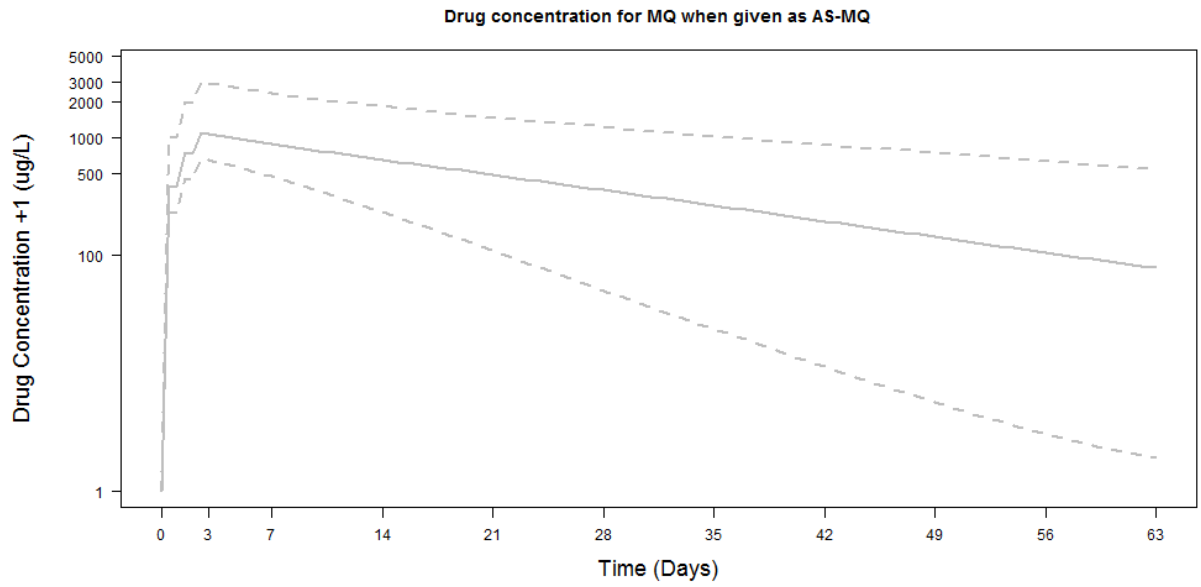



**Figure 2.3** MQ concentration (in µg/L) over time for a population of 5,000 patients, treated with AS-MQ parameterized as in **Table 2.1** with drug dosing as in **Table 2.2**. The solid line is the median population concentration at each day and the dashed lines are the 5% and 95% quantiles. The figure follows patients for 63 days (the maximum length of patient follow-up investigated within this thesis).

In **chapter 3**, a three-compartment calibration of PPQ is also analysed (to reflect the scientific debate over whether PPQ follows a two or three-compartment structure, i.e. [111]). The parameter values chosen were based on the mean values provided in table 2 of Tarning et al. [113] as a starting point. As with the two compartment model, I was not trying to reproduce the exact patient population of that study, but rather create a "general" population. Hence, I do not use their error structure but include sufficiently large CVs across 5,000 patients that the distribution of drug concentration over time curves covers a reasonably shaped parameter space (**Table 2.3).** The key difference between the drug concentrations produced by the two and three-compartment models is that the three-compartment model is slightly more prophylactic (i.e., the drug concentration reduces at a slower rate over time, **Figure 2.4**).

**Table 2.3**: A summary of the PK parameters for PPQ used to generate parasite dynamics post-treatment with DHA-PPQ assuming a three-compartment model for PPQ (as opposed to the two compartment model parameters described in **Table 2.1**).

| **Drug** | **Piperaquine (three compartment model)** |
|---|---|
| **Vd (L)** | 3070 (0.86)[113] |
| **$Vd_1$ (L)** | 4440 (1.21) [113] |
| **$Vd_2$ (L)** | 31400 (0.65) [113] |
| **ka (/day)** | 1.99 (1.08) [51] |
| **$Q_1$(L/hour)** | 427 (1.01)[113] |
| **$Q_2$(L/hour)** | 160 (0.7)[113] |
| **k (/day)** | 0.47* [113] |

Vd: Volume of Distribution (central compartment), $Vd_1$: Volume of Distribution (peripheral compartment 1), $Vd_2$: Volume of Distribution (peripheral compartment 2), ka: Absorption rate constant, $Q_1$: intercompartmental clearance (central-peripheral 1), $Q_2$ : intercompartmental clearance (central-peripheral 2). PK means are derived from [113]; the coefficient of variation (CV) in brackets for each parameter was added by me and is consistent with CV for the two-compartment calibration of PPQ in **Table 2.1**. Citations in support of the parameter values are provided in squared brackets.

* elimination rate for PPQ is calculated from clearance (CL) / Vd; CL (from [113] is 60.2 and a CV of 0.71 is included on this parameter) so the value presented here is illustrative and represents a bodyweight of 42kg.

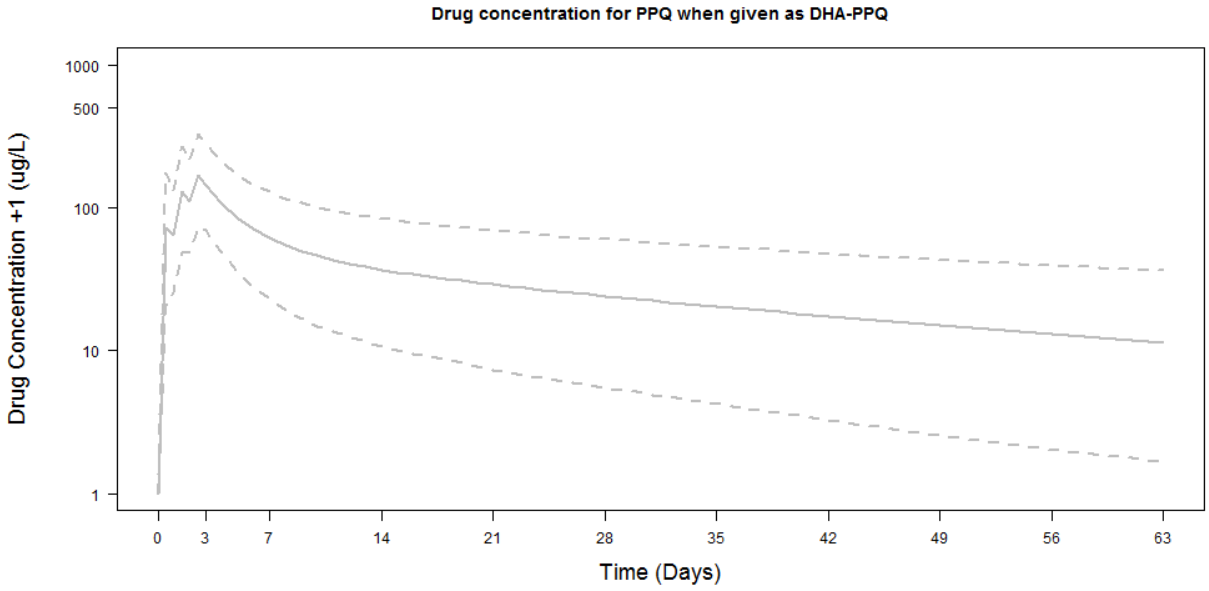



**Figure 2.4** PPQ drug concentration (in µg/L) over time for a population of 5,000 patients, treated with DHA-PPQ under the assumption that PPQ is modelled with a three-compartment model - parameterized as in **Table** 2.3with drug dosing as in **Table 2.2**. The solid line is the median population concentration at each day and the dashed lines are the 5% and 95% quantiles. The figure follows patients for 63 days (the maximum length of trial follow-up investigated within this thesis).

### 2.1.2 Artemisinin concentrations over time.

The drug concentration over time profiles produced by the model for DHA, AR and AS when given as DHA-PPQ, AR-LF and AS-MQ (i.e., only the concentration of the artemisinins) produced for 5,000 patients with the parameters shown in **Table 2.1**and the dosing regimen in **Table 2.2** are shown in **Figure 2.5** , **Figure 2.6** , and **Figure 2.7** respectively. Both AR and AS are metabolized into DHA in these mPK/PD models -see reference [104] for a detailed mechanical description of this process.

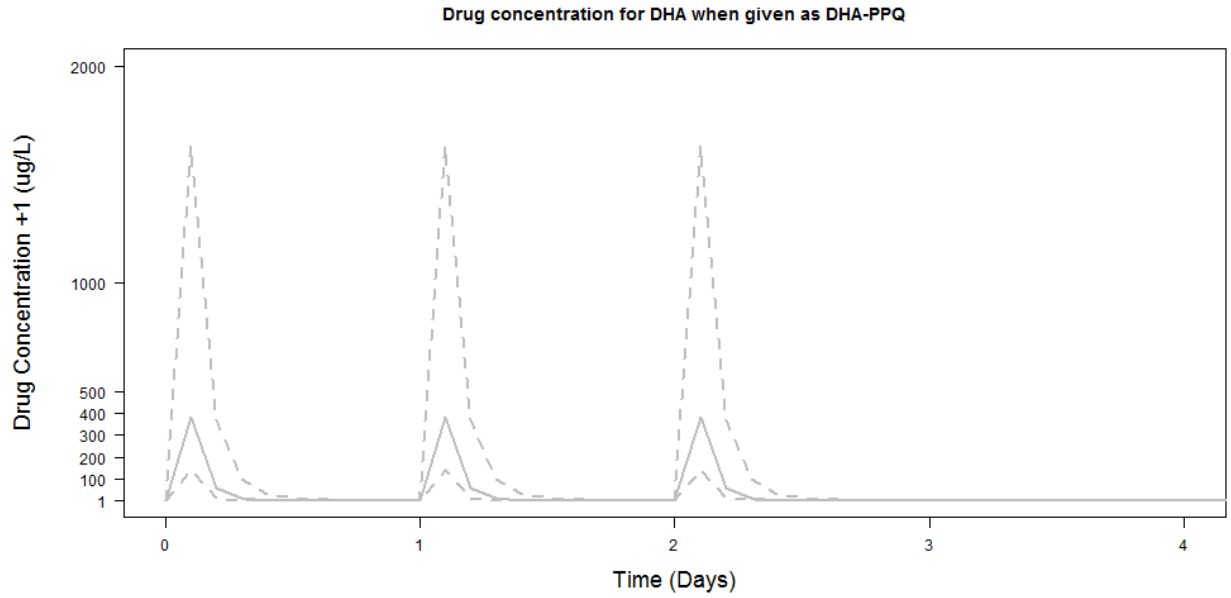



**Figure 2.5** DHA concentration (in µg/L) over time for a population of 5,000 patients, treated with DHA-PPQ parameterized as in **Table 2.1** with drug dosing as in **Table 2.2**. The solid line is the median population concentration at each day and the dashed lines are the 5% and 95% quantiles. The figure follows patients for 4 days (after which all artemisinins have decayed to non-effective and/or zero concentrations).

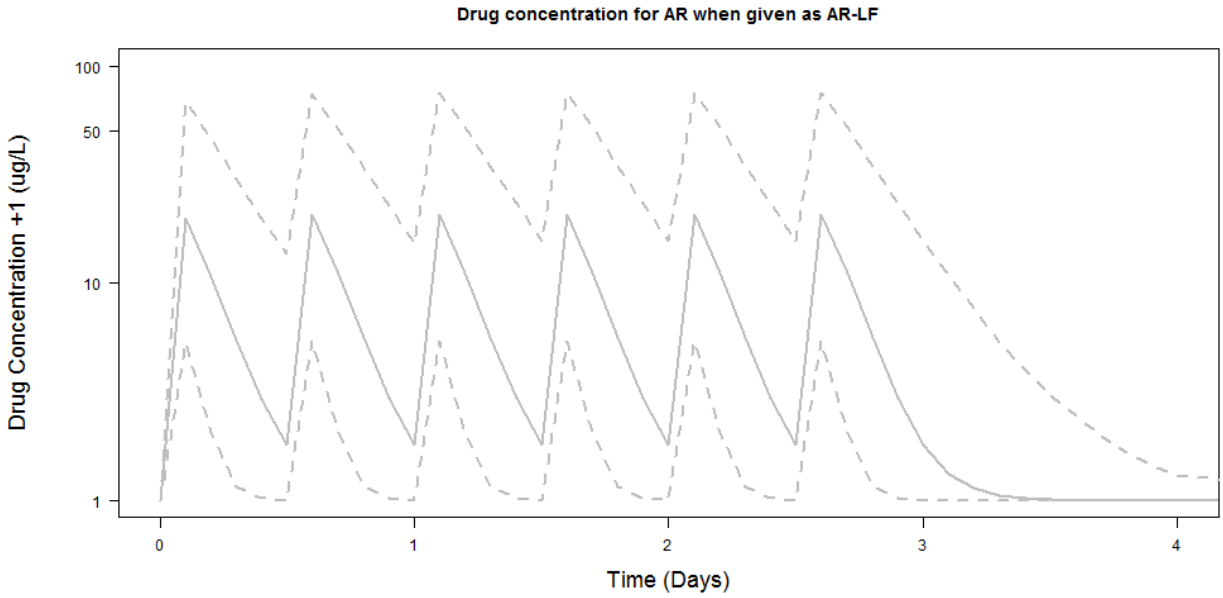



**Figure 2.6** AR concentration (in µg/L) over time for a population of 5,000 patients, treated with AR-LF parameterized as in **Table 2.1** with drug dosing as in **Table 2.2**. The solid line is the median population concentration at each day and the dashed lines are the 5% and 95% quantiles. The figure follows patients for 4 days (after which all artemisinins have decayed to non-effective and/or zero concentrations).

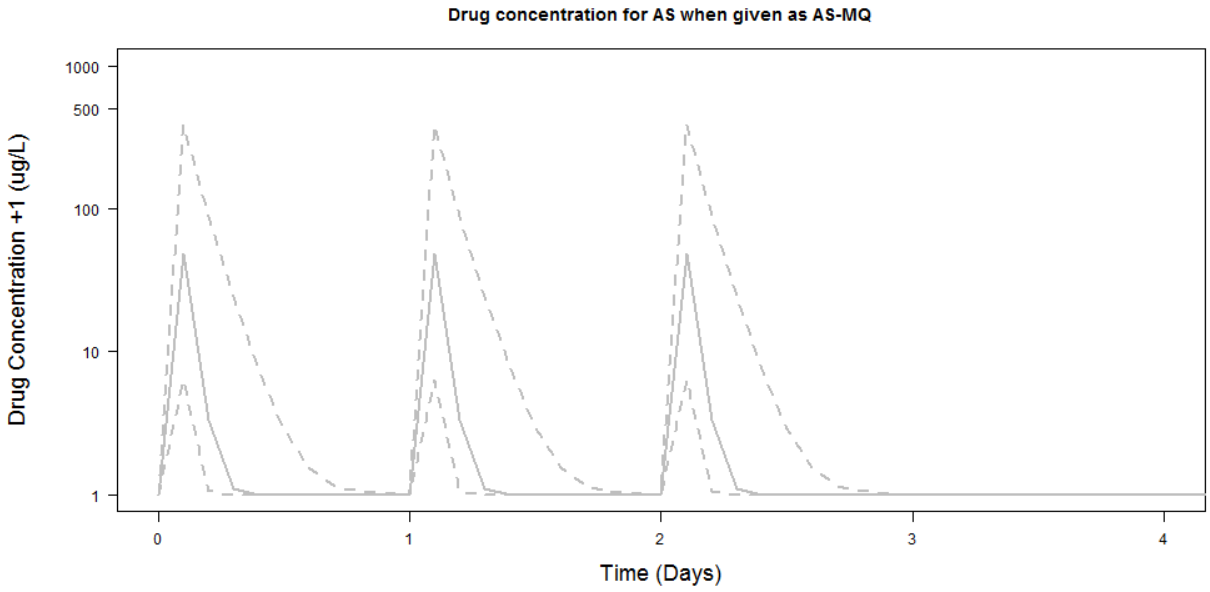



**Figure 2.7** AR concentration (in µg/L) over time for a population of 5,000 patients, treated with AR-LF parameterized as in **Table 2.1** with drug dosing as in **Table 2.2**. The solid line is the median population concentration at each day and the dashed lines are the 5% and 95% quantiles. The figure follows patients for 4 days (after which all artemisinins have decayed to non-effective and/or zero concentrations).

## 2.2 Mechanistic PD model to describe drug effect and relation between drug concentration and drug effect.

The PD parameters determine the rate of drug killing (at a given concentration of drug) for each parasite clone; the mechanistic relationship between these parameters and drug killing is described in [85, 104, 105] (specifically in [104]). The mPK/PD method requires three parameters: The maximal parasite killing constant, Vmax, and the slope factor, n, which remain the same for all drugs described in **chapters 3-5**, and the "half maximal inhibitory concentration" (IC50), which is the concentration of drug at which half-maximal parasite killing occurs. IC50 is identical across all chapters for the artemisinins only, but not for partner drugs, as non-failing / failing ACTs were simulated by altering only the partner drug IC50, given the desire to later investigate the accuracy of classification of recrudescence and reinfection over long follow-up periods. These parameters are shown in **Table 2.4**.

**Table 2.4** A summary of the PD parameters used to generate parasite dynamics in vivo with an mPK/PD model.

| **Drug parameter** | **Di-hydroarteminisin-Piperaquine (2 compartment model)** | | **Artesunate-Mefloquine** | | | **Artemether-Lumefantrine** | | |
|---|---|---|---|---|---|---|---|---|
| | **DHA** | **PPQ** | **AS** | **DHA** | **MQ** | **AR** | **DHA** | **LF** |
| **IC50 (mg/L)** | 0.009 (1.17)[85, 104] | | 0.0016(0.86)[104] | 0.009(1.17)[104] | | 0.0023(0.79)[85, 104] | 0.009(1.17)[85, 104] | |
| **Vmax** | 27.6[85, 104] | 3.45[105] | 27.6[104] | 27.6[104] | 3.45[104] | 27.6[85, 104] | 27.6[85, 104] | 3.45[85, 104] |
| **n** | 4[85, 104, 105] | 6[105] | 4[104, 105] | 4[104, 105] | 5[104] | 4[85, 104, 105] | 4[85, 104, 105] | 4[85, 104, 105] |

*IC50: Half maximal inhibitory concentration, Vmax: Maximal parasite killing constant, n: Slope factor. Half maximal inhibitory concentration (IC50) is shown for all artemisinins but not for partner drugs (partner drug IC50 is described in-chapter for **chapter 3-chapter 5**. Maximal parasite killing constant (Vmax) and the slope factor (n) are shown for all artemisinins and partner drugs. The coefficient of variation (CV) is provided in brackets where appropriate. Citations are provided in square brackets in support of parameter values.*

### 2.2.1 Partner drug IC50.

A key incentive of pursuing accurate drug failure rate estimates (see introduction) is being able to detect when drug failure rates reach ~10% and are classed as failing [24].There is evidence of DHA-PPQ having high estimated failure rates in vivo, and well-documented parameterization for this ACT as it fails. Saunders and colleagues [20], for instance, estimated PPQ IC50 had increased to 23.9ng/ml in recrudescent infections as resistance spread (equivalent to 0.024 mg/L). Consequently, simulating failing DHA-PPQ using in vivo data to calibrate the model was possible. There are field data allowing calibration of the model for non-failing AR-LF and AS-MQ, but failing AR-LF and AS-MQ have not been observed in any known PK/PD studies. Consequently, to model failing AR-LF and AS-MQ I based my parameterization on non-failing PK/PD data and increased the IC50 values to achieve ~10% failure rates.

Importantly, partner drug IC50 values had to vary both between and within **chapters 3-5**. Where possible, I wanted to use in vivo data on genetic marker allele distributions and multiplicity of infection (MOI) distributions (see below). This meant that, for example, the MOI distribution for the length polymorphic markers (**chapter 3**) was different to the MOI distribution for microsatellite markers (**chapter 4**), as the data came from different locations.

Furthermore, MOI was varied within **chapters 3-5**. Changing MOI while keeping the IC50 of the partner drug the same will result in different failure rate estimates because a higher MOI means more initial infections and subsequently a higher chance of that patient failing treatment. Because a key purpose of **chapter 4** was to validate a Bayesian algorithm (see details in that chapter), I wanted to keep true failure rates within one percent of each other as different MOI settings were investigated, and this required that IC50 values be changed for each of three MOI distributions in that chapter.

Consequently, the IC50 values used in **chapters 3-5** are not shown here to avoid listing multiple values and creating confusion, but rather are described and discussed within-chapter for **chapters 3-5**. Please note two important points:

Values for failing AR-LF and AS-MQ are not directly derived from in vivo data (see above) and have been created computationally to achieve, for a given model scenario, a particular drug failure rate (~10%).

The partner drug IC50 is altered. The IC50 of the artemisinin component is not. However, the partner drug is not given as a monotherapy in this thesis – the artemisinin component is still killing. Consequently, values of IC50 for a partner drug, for a given true failure rate, will be higher than the value of IC50 needed to produce that same true failure rate if the partner drug were given as a monotherapy.

## 2.3 Number of parasite clones and clone density.

Before the mPK/PD model can be used to track changes in parasite numbers over time post-treatment and produce descriptions of intra-host parasite dynamics, each patient must be "populated" with parasites. The mPK/PD models described here allow for multiple malaria clones to exist simultaneously in a patient, and for infections to emerge in a patient over the course of the treatment that were not present at the time of treatment.

### 2.3.1 Parasite clones at the time of treatment – multiplicity of infection (MOI).

A malaria infection may consist of several genetically-distinct parasite clones and the number of clones in a patient at the time of treatment is termed the multiplicity of infection (MOI) although the equivalent term "complexity of Infection" is also found in the literature. Each patient will have this number drawn from a distribution. The exact distribution **varies both within and between** future chapters of this thesis; for example, simulation of higher transmission areas will be accompanied by higher mean MOI. This thesis endeavours to utilize in vivo distributions of MOI where possible, but these distributions, ideally, must be derived from the same location as the allele distributions of whichever genetic markers are under investigation. This has been possible throughout this thesis and consequently these distributions will be provided within their respective chapters.

Each initial clone was given a starting number of parasites drawn from a uniform (**chapter 3**) or log-uniform (**chapter 4, chapter 5**) distribution. The distribution range was $10^{10}$ to $10^{11}$; note that a wider distribution of $10^{8}$ to $10^{11}$ was additionally explored in **chapter 5** only. For reference, assuming a patient with 4.5L of blood and a WBC count of 8,000/µl of blood, parasitaemia of $10^{10}$ and $10^{11}$ would correspond to densities of 2,222 parasites/µl of blood and 22,222 parasites/µl of blood respectively,

per WHO counting procedure [114]. Previous modelling approaches used $10^{12}$ parasites as the upper limit of parasitaemia; this level of parasitaemia is likely to be lethal or at least exceed the maximum parasite density exclusion criteria in a clinical trial (typically 100,000 parasites /μl); hence $10^{11}$ was chosen as the upper limit for any single clone at the time of treatment. The lower limit of $10^{10}$ would appear to disregard the presence of low density clones. However, the MOI distributions used in each chapter are derived from *in vivo* samples, and the MOI is calculated by genotyping samples from a population with either length-polymorphic markers (**chapter 3** and **chapter 5**) or microsatellite markers (**chapter 4**). Thus, detection of clones (and quantifying MOI) is dependent on these methods – which cannot detect minority clones (this is explored later, but see **3.2.4** and **3.2.5**), and so the lower limit of $10^{10}$ was chosen to ensure that a patient's initial clones reflect the MOI distributions observed *in vivo* and used to parameterize the model. Because of the greater sensitivity of AmpSeq to low density clones, a wider range of $10^{8}$ to $10^{11}$ was explored in **chapter 5** (**5.2.2**) – note that to the best of my knowledge no population MOI estimates obtained with AmpSeq that would be suitable to parameterize these models are available, and so, of necessity, MOI distributions obtained using length-polymorphic markers were used.

### 2.3.2 Reinfection during follow-up – force of infection (FOI).

The number of reinfections that will occur by emerging from the liver in a given patient during the course of follow-up is pre-determined by the "force of infection (FOI)". The number of reinfections per year that emerge from the liver is drawn from a Poisson distribution where the mean of the distribution is the parameter FOI, which is varied to reflect the local intensity of malaria transmission.

FOI is varied between 0 and 16 per year herein, with an FOI of 2 broadly considered to represent a low transmission area, an FOI of 8 broadly considered to represent a medium transmission area and an FOI of 16 broadly considered to represent a high transmission area. Selection of these FOI values were based on the following reasoning (described in the supplemental material of [93], which corresponds to work described in **Chapter 3** of this thesis): "Data from northern Ghana indicates that the average number of new infections per patient per year is 16, and similar estimates can be obtained from efficacy data of effective ACTs (see supplementary material of [90]). Mueller et al. [115] obtain estimates of between 3 and 9 new infections emerging per year with an average of 5.9 in Papa New Guinea. Additional work suggests the FOI in Ghana is highly seasonal with estimates ranging from 44 in the high transmission season to 7 in the low transmission season [115]; but note that any yearly average (such as assumed here) inherently fails to capture the nuances of seasonal transmission. Smith et al [116] explicitly modelled the relationship between entomological inoculation rate (EIR) and FOI. It is technical, but some illustrative data are summarised in their Figure 2: Incidence during a 2 week period at annual EIRs of 36.5 (moderate transmission) and 365 (high transmission) were roughly 0.2 and 0.4 respectively implying annual FOI estimates of 0.2 x26=5.2 and 0.4x26= 10.4 respectively. These may be slight under-estimates because this simple calculation assumed that more than one infection could not become established in a 2 week period but serves as general illustrations of the relationship." The FOI values that are used within **Chapter 3- Chapter 5** to parameterize any given model scenario are described within those chapters.

The day of emergence of a reinfection clone was randomly chosen from any day in the trial follow-up period. When reinfections emerge, they emerge at a frequency of $10^{5}$ parasites [91, 117] (note that all drugs investigated herein are assumed to be inactive against parasite hepatic stages). It is possible in the model for a reinfection to emerge extremely early (i.e., day 1, 2, 3, etc) post-treatment because it was assumed the drugs have no action against parasites while in their liver stages. However, the presence of high drug concentration at these time-points will likely quickly clear these reinfections.

### 2.3.3 Parasite growth rate and density-dependent effects.

The PK/PD model calculates drug concentration over time and parasite kill rate (above), but also requires a the parasite growth rate; this was assumed to be 1.15 for every clone, consistent with previous modelling approaches [86, 105] and equivalent to a parasite multiplication rate (PMR) of 10. The model progresses in daily time-steps to a given follow-up duration (generally 28, 42 or 63 days, but see specific chapters), recalculating the numbers of each parasite clone in a given patient each day until the final day. This is a simple process – each clone increases in number per day according to its growth rate and is killed according to the drug kill rate (known from the PD component of the model) at a given concentration of drug (known from the PK compartment of the model).

The mPK/PD models assumed that if the total parasitaemia (i.e. the sum of parasitaemia of all clones) in a patient at any time, reached $10^{12}$, then density-dependent effects, such as fever, acted to control and stabilise the parasitaemia, effectively setting the growth rate of every clone in that patient to 0 (equivalent to a PMR of 1).

## 2.4 Model output and a note on parasite count versus parasite density.

The models described above calculate the absolute number of parasites of a given clone at any time (i.e., parasitaemia) as opposed to the parasite density measure that may be more familiar to clinicians. Parasite density can be measured in vivo and converted to parasite count; though this conversion requires assumptions to made regarding the white (or red) blood cell count of a patient and is dependent on the ability of the microscopist to identify parasites in a sample [114], and furthermore, these counts may be highly variable [118]. Hence the mPK/PD models do not include blood cell counts, or their variability, and operate in terms of parasite count rather than density.

## 2.5 Modelling gametocytes

The life cycle of *P. falciparum* is complex and contains multiple stages; while the 'parasitaemia' that causes acute disease relates to the numbers of mature asexual parasites in the blood, it is important to note that these merozoites can produce gametocytes that will be ingested by the mosquito vector. These gametocytes will also express genetic loci (be they *msp-1*, *msp-2* and *glurp*, microsatellites, or AmpSeq alleles), meaning that they produce a genetic signal that can be picked up by genotyping. Furthermore, none of the ACTs investigated within this thesis have any impact on malarial gametocytes; following ACT treatment gametocytes can remain detectable in a patient's blood for an average of 55 days [119]. Consequently, there is a concern that gametocytes from initial infections can cause later reinfection to be misclassified as recrudescence as it not possible to distinguish what proportion of the allelic signal comes from gametocytes or asexual parasitaemia, and gametocytes may persist after the initial asexual parasites have been cleared by treatment. I wrote a model to include gametocytes, though they were not eventually included in any chapter of this thesis. A discussion of why I did not eventually include them (a near zero impact on most results) is provided in **Chapter 7**.

Gametocytes can be included in the mPK/PD model presented in this chapter with the following process:

Each clone of malaria in a patient has an initial parasitaemia that reflects the number of asexual parasites in the blood. Each clone will also have some number of gametocytes that share the genetic signal of the earlier asexual form. This was modelled as a certain percentage (say, 10%) of the asexual parasitaemia of that clone. After *x* days, the number of gametocytes will decay according to their half-life ($g_{1/2}$). Between day 0 and day *x* the number of gametocytes will remain constant – this lag period reflects the time taken for drugs to kill asexual parasites – without replenishment (following death of asexual stages), the number of gametocytes can begin to fall.

On any day, the allelic signal from gametocytes and asexual parasites can be summed, and the alleles that would be observed, having accounted for missing minority alleles, can be determined. A large gametocyte signal may mean that alleles from new infections become undetectable minority alleles, and will inflate signal from clones that have fallen below detectable parasitaemia and make it incorrectly appear that they are detectable (a false recrudescence).

The number of gametocytes of a clone present in the blood after *t* time following treatment can thus be expressed by the following equation:

$$G_t = G_0 * e^{\frac{\ln(2)}{g_{\frac{1}{2}}} * t} \qquad \text{Equation 2.1}$$

Where ln(2) / $g_{1/2}$ is the rate of decay of gametocytes when gametocyte half-life is $g_{1/2}$ days and the unit of time is a day.

# Chapter 3: Improving methods for analysing anti-malarial drug efficacy trials: molecular correction based on length-polymorphic markers msp-1, msp-2 and glurp.

A version of this work has been published in Antimicrobial Agents and Chemotherapy : Improving methods for analysing anti-malarial drug efficacy trials: molecular correction based on length-polymorphic markers *msp-1*, *msp-2* and *glurp*. S. Jones, K. Kay, E.M. Hodel, S. Chy, A. Mbituyumuremyi, A. Uwimana, D. Menard, I. Felger, I. Hastings. Antimicrobial Agents and Chemotherapy Jul 2019, AAC.00590-19; DOI: 10.1128/AAC.00590-19.

Chapter-specific acknowledgements: Dr Katherine Kay, Dr Eva Maria Hodel and Dr Ian Hastings provided R code to generate parasite dynamics post-treatment (fully described in **Chapter 2** of this thesis). Professor Ingrid Felger provided in-depth advice on genotyping processes for length-polymorphic markers. Dr Sophy Chy, Dr Aimable Mbituyumuremyi and Dr Aline Uwimana performed additional genotyping on in vivo samples for re-analysis (**3.2.8**). Dr Didier Menard facilitated access to and performed additional genotyping on in vivo samples for re-analysis (**3.2.8**). Dr Jörg Möhrle and Dr Stephan Duparc provided the data shown in **Table 3.4**.

## 3.1 Background.

### 3.1.1 Length-polymorphic markers for genotyping TES.

The current consensus method for distinguishing recrudescence from reinfections is molecular correction or, equivalently, polymerase chain reaction (PCR)-correction. A genetic profile of the malaria infection of each patient is taken just before treatment, with a second profile taken if the patient develops a detectable malaria infection during follow-up (known as "recurrent" parasitaemia). If the profiles 'match' then the patient is considered to have a recrudescent infection if they do not match the patient is considered to have a reinfection. This 'matching' is simple in principle, but in practice has substantial limitations. The main problem is that individual malaria infections may consist of several genetically-distinct clones. Current genotyping techniques struggle to detect minority clones that are present in relatively low numbers and/or which carry alleles that do not amplify well during the genotyping process. These limitations were recognised early in the development of molecular correction methodology [38, 42, 120] and led the World Health Organization (WHO) and Malaria for Medicines Venture (MMV) to co-sponsor a meeting in 2007 to identify a consensus methodology for molecular correction; their findings were published in 2008 [14]. The consensus methodology utilizes three length-polymorphic loci of *P. Falciparum*: Merozoite surface protein 1 (*msp-1*), merozoite surface protein 2 (*msp-2*) and the glutamate-rich protein (*glurp*). Different alleles at all these loci are distinguished by differences in the lengths of their fragments (sequencing the entire allele is possible, but not feasible for the purposes of TES [38]) ; note also that *msp-1* and *msp-2* exist in distinct families, whereas *glurp* does not.

### 3.1.2 Sources of error with length-polymorphic markers.

Concerns surrounding the limitations of this approach have persisted [121, 122]: Previous studies have noted that different algorithms give different results when applied to clinical data (e.g. Table 2 of [122]) and a recent publication quantifying the limitations inherent in PCR detection has led to renewed calls for this methodology to be re-examined [106]. There are three types of misclassification that may occur with length-polymorphic markers:

Recrudescent infections can be misclassified as reinfection if alleles of the recrudescent clone(s) were not detected when genotyping the initial infection.

Recrudescent infections can be misclassified as reinfection if the recurrent infection is mixed (i.e., the recurrence is polyclonal and comprised of both recrudescence and reinfections), and alleles of the recrudescent clone(s) are not detected, but alleles of a reinfecting clone(s) are.

A reinfection could be misclassified as recrudescent if it shares (by chance) alleles with clones present at time of treatment.

Select properties of the three loci used in the length-polymorphic marker methodology and properties inherent in the genotyping process may lead to alleles not being detected when they are truly present (and resulting in a) and b) above). Firstly, the products of PCR are compared using capillary electrophoresis (CE) [106]- note that a variety of genotyping techniques are described by the WHO in their 2008 guidelines [14], but CE has generally superseded comparison using agarose gel based methodology which has been used historically i.e. [123]. This process involves reading the heights of peaks shown in electropherograms that show the relative fluorescent units (rfu) of specific lengths of DNA, which in this context means different alleles (see Figure 1 of [124] for an example of an electropherogram. In this process, small rfu peaks ("stutter peaks") are considered noise generated during the genotyping process and ignored. The size of the peaks at which signals are considered noise varies between laboratories, but for length-polymorphic markers is generally considered to be between 10% and 30% of the majority peak (i.e., the largest signal; for examples see [106, 121], but note the cut-off should be defined in the methodology section of any paper reporting genotyping results with these markers). The obvious consequence of this is that truly present minority alleles (i.e., those of a lower density parasite clone) may be ignored as noise.

Secondly, alleles undergo template competition and so are amplified differently based on their length, with shorter fragments being amplified preferentially in the PCR process [106]. This effect occurs separately for different families of *msp-1* and *msp-2* (i.e., template competition between two alleles of a separate family is negligible). It occurs always with *glurp*, which does not possess allelic families. The consequence of this is that alleles will have varying levels of detectability based on their length, and for *msp-1* and *msp-2*, this is affected by their families (i.e., it is more likely to detect alleles of *msp-1* and *msp-2* in polyclonal infections if clones have different families).

There now exist several proposed sets of rules (referred to hereafter as "algorithms"), including the current WHO/MMV consensus methodology, for interpreting genetic profiles with length-polymorphic markers to classify patients (see definitions in **Table 3.1**). The true failure rate is unknown in vivo, so it has been impossible to identify the level of misclassification that occurs and which algorithm is most accurate; consequently, the molecular correction field is currently in a state of limbo with several alternative methods giving different results, but with no way of knowing which method is most accurate; furthermore, some of these algorithms are newly proposed and have not been used to return failure rate estimates in vivo.

The final source of potential misclassification, c), will occur if a reinfection shares (by chance) a sufficient number (or type) of alleles with the initial sample. The exact number (or type) of alleles that must be shared depended on the molecular correction algorithm chosen (i.e., the no *glurp* algorithm was not affected by sharing an allele at the *glurp* locus, and the allelic family switch algorithm was sensitive to sharing an *msp-1* or *msp-2* family by chance, whereas the other algorithms were not). Of note is that the WHO/MMV algorithm would require matching alleles at all three loci for this misclassification to happen (i.e., it is the algorithm with the highest specificity).

### 3.1.3 Research goals .

Using an mPK/PD approach (**Chapter 2**) to simulate antimalarial therapy provides a gold-standard definition of true response to treatment, something un-obtainable from in vivo TES data, including the true status (reinfection or recrudescence) of all recurrent infections. Simulations of 5,000 patients

were conducted for three ACTs: DHA-PPQ, AR-LF and AS-MQ. Using this simulated data, this chapter had three main objectives:

1: Investigate the accuracy of the range of available algorithms for molecular correction with length polymorphic markers (**Table 3.1**) by comparing the estimated failure rates produced using these methods to the true failure rate known through the mPK/PD model.

2 :Investigate the impact on failure rate estimates (and their accuracy) of varied lengths of follow-up for the three ACTs

3: Re-analyse field data with the range of algorithms available to compare for consistency with simulated results.

**Table 3.1**. Molecular correction algorithms proposed to decide whether a patient presenting with a recurrent malaria infection during follow up is a recrudescence or a reinfection based on the WHO-recommended length-polymorphic markers: msp-1, msp-2 and glurp.

| Algorithm | Reference | Definition | Consequences (identified in the model) |
|---|---|---|---|
| No Correction | | All recurrent infections classified as recrudescence | Grossly over-estimates failure rate at higher FOI |
| WHO/MMV | [28] | Initial and recurrent samples must have shared alleles at all three loci to be classified as recrudescence. | Stringent conditions for recurrences to be classified as recrudescence means that around 50% of true recrudescence are misclassified as reinfections resulting in greatly underestimated failure rates. Most reinfections are correctly classified, so FOI has little impact on estimated failure rate |
| No *glurp* | [106] | As for the WHO/MMV algorithm but based on two loci (i.e. *msp-1* and *msp-2*; *glurp* is omitted as it is prone to genotyping errors). | Largely identical to the WHO/MMV method |
| ≥ 2/3 markers | [106] | As for the WHO/MMV algorithm, but initial and recurrent samples must share alleles at least at two out of three loci to be classified as recrudescence. | Generally intermediate between the no *glurp* and allelic family switchalgorithms |
| Allelic family switch | [106] | Comparison initially based on *msp-1* and *msp-2*. Identical alleles observed at both loci indicate a recrudescence. Absence of shared alleles at both loci indicate a reinfection. If one loci shares alleles and one does not (i.e. the sample is "discordant"), a complete allelic family shift in the non-sharing loci is required to classify a recurrence as a reinfection | Tendency to misclassify reinfections as recrudescences leads to a dependency on FOI and results in large overestimates of failure rates at higher FOI, though produces accurate failure rate estimates at low FOI. |

## 3.2 Methodology.

### 3.2.1 Partner drug choice and IC50 parameterization.

Malaria parasite dynamics were generated using mPK/PD models (**Chapter 2** of this thesis) for three front-line ACTs: Dihydroarteminisin-Piperaquine (DHA-PPQ), Artemether-Lumefantrine (AR-LF) and Artesunate-Mefloquine (AS-MQ).

All PK/PD parameters used in the model described in this chapter are described in **Table 2.1**, **Table 2.3** and **Table 2.4** with the exception of half maximal inhibitory concentration (IC50) of the partner drugs, which are described here and shown in **Table 3.2**:

There is evidence of DHA-PPQ having high estimated failure rates in vivo (**Chapter 2);** a value of 23.9ng/ml in recrudescent infections as resistance spread was equivalent to 0.024 mg/L (rounded here to 0.02mg/L). There are field data allowing calibration of PK/PD parameters for non-failing AR-LF and AS-MQ; 'failing' calibrations of AR-LF and AS-MQ were produced by artificially increasing the mean IC50 values until failure rates reached around 10%. Coefficient of variation (CV) values were taken from the literature for PPQ and non-failing LF and MQ and kept consistent for the failing calibrations of LF and MQ.

**Table 3.2**: Mean values of the half-maximal inhibitory concentration (IC50) for each calibration of three partner drugs used within this chapter.

| Partner Drug | Mean IC50 (mg/L) | Literature Justification |
|---|---|---|
| Failing PPQ | 0.02 (0.3) | [20] |
| Non-Failing LF | 0.032 (1.02) | [87] |
| Failing LF | 4 (1.02) | N/A |
| Non-Failing MQ | 0.027 (0.78) | [87] |
| Failing MQ | 0.37 (0.78) | N/A |

*PPQ: Piperaquine, LF: Lumefantrine, MQ: Mefloquine, IC50: Half-maximal inhibitory concentration. Coefficient of variation (CV) is given in brackets. This table should be considered with **Table 2.1**, **Table 2.3** and **Table 2.4** for a full set of pharmacokinetic/pharmacodynamic parameters for all drugs simulated in this chapter. Note that the IC50 value for a two and three-compartment model calibration of PPQ are the same.*

### 3.2.2 Multiplicity of Infection and Force of Infection, and initial parasite number.

Two MOI distributions were used in this chapter. A "high MOI" representative of the MOI in an area of intense transmission, in this case Tanzania where MOIs of 1-8 were assigned with probabilities 0.036, 0.402, 0.110, 0.110, 0.183, 0.049, 0.061, 0.049 respectively, based on data described in [125]. A "low MOI" distribution was based on data from Papua New Guinea with probabilities of 0.460, 0.370, 0.150 and 0.020 for an MOI of 1-4 respectively [37]; these two distributions were used to check if the accuracy of different algorithms were consistent across different MOIs. Each clone within the MOI (later called "initial clones") had their starting parasitaemia drawn from a uniform distribution spanning from $10^{10}$ to $10^{11}$ asexual parasites per person (see **2.3.1** for discussion and justification of initial parasitaemia).

The FOI values used to calibrate the model in this chapter varied between 0 and 16; all values in that range inclusive were modelled.

### 3.2.3 Genetic Data – allele frequency distributions for msp-1, msp-2 and glurp.

Each clone, whether an initial clone present at treatment or a reinfection that emerged during the follow-period, was assigned a genetic profile based on three loci: *msp-1*, *msp-2*, and *glurp*, using previously established distributions for the frequency of alleles. *Msp-1* and *msp-2* allelic frequency distributions and amplicon sizes were derived from 115 or 108 patients from Tanzania [126]. *Glurp* distributions were drawn from a collection of field samples described in [106]. The length of each allele and its allelic family (for *msp-1* and *msp-2*) was also noted. The distributions used gave *msp-1* expected heterozygosity (He) of 0.915, *msp-2* He of 0.963, *glurp* He of 0.956; [see Supplemental File 1 of [93]] for full data). It was assumed that the genotypes of initial clones were independent of each other and were also independent of the genotypes of reinfections (i.e. it was assumed there is no local genetic structuring of the malaria population). Note that alleles at *msp-1* and *msp-2*, exist in these distributions as members of three or two distinct families, respectively (Families K1, MAD20 and RO33 for *msp-1*, and families 3D7 and Fc27 for *msp-2*).

### 3.2.4 Follow-up length and detection of recurrence.

Multiple lengths of follow-up are permitted in the WHO guidelines [14] and used in practice [127] for TES of DHA-PPQ, AR-LF and AS-MQ. The length of the follow-up period affects drug failure rate estimates in two ways: Firstly, a longer follow-up period will allow more time for recrudescent clones to become detectable (i.e. if a patient had parasites that would recrudesce and become detectable on day 60 and the follow-up period was 28 days, this recrudescence would not be observed). Secondly, a longer follow-up period leads to more reinfections emerging in each patient, some of which may be misclassified as recrudescence and inflate failure rate estimates. Accurate, robust analyses need to balance these two risks through appropriate choice of follow-up duration. WHO guidelines [28] stipulate that patients are checked for recurrent parasitaemia by light microscopy on scheduled days of follow-up. A 28-day follow-up schedule requires patients be examined on days 3, 7, 14, 21 and 28. A 42-day follow-up period uses two additional days i.e. days 35 and 42. A 63-day follow-up period (not recommended in routine surveillance) has scheduled visits as per the 42 days but with 3 extra days i.e. 49, 56 and 63. Novel lengths of follow-up were simulated simply by "ending" the trial on any given day of follow-up, i.e. to investigate a 35-day follow-up length, patients were checked on days 3, 7, 14, 21, 28 and 35.

The parasitaemia of each clone in each patient was tracked and updated each day as described by the mPK/PD model and the PK parameters of the patient and the PD parameters of the clone (**Chapter 2**).

The model checked each day of scheduled follow-up to determine whether a patient had enough parasitaemia that a recurrence would be detectable by light microscopy (a recurrence) – parasitaemia was considered detectable if the total number in a patient was $\geq 10^8$ on that day. Note that variance in the limit of detection by light microscopy exists with respect to the skill of the microscopist [128]; It was assumed this limit was reflective of an "expert" microscopist (corresponding to roughly 20 parasites /µl of blood).

### 3.2.5 Calculating which alleles are observed.

Once the patient parasite dynamics were modelled (as described above), and genetic profiles at the three loci were assigned, the models followed the same process as in vivo trials. Blood samples were taken from each patient immediately prior to treatment (the initial sample), and at pre-determined days during the follow-up period. Samples were screened for the presence of *P. falciparum* by light microscopy to check for recurrent infection. Calculations then took place to determine the genetic signal that would be observed at time of treatment and at any recurrence, using a process to replicate

the technical limitations of acquiring blood samples and genotyping msp-1, msp-2 and glurp as follows:

A "sampling limit" exists; a finite amount of blood is used for genotyping. A parasite clone (and consequently, it's alleles) would not be detected if its density were so low that no parasites are included in the blood sample analysed. Thus, the density and volume of the processed blood sample define the limit of detection. Obviously, this sampling limit differs between methods and laboratories. Here, its likely value is estimated using the genotyping methodology employed for malaria genotyping by the Swiss Tropical and Public Health Institute (personal communication from Ingrid Felger to Sam Jones, April 2016). In this methodology, patient blood is obtained via finger prick. Typically, 3 lots of paper each with a 3mm diameter will have a blood spot placed on them. If this paper is Whatman 3MM paper, as recommended in WHO protocol [14], then each spot will contain 2 µl of blood for a total of 6 µl. The DNA from these spots are extracted in solution to a total volume of 50 µl, of which 5 µl is then taken for PCR and genotyping. Consequently there will be, on average, 0.6 µl of blood presented for PCR. Ideally for a clone to be detected in this process only a single parasite would need to be present per 0.6 µl of blood. Assuming 5L of blood in the human body, there would need to be ($8.3 \times 10^7$) copies of a parasite of a given clone present for that clone to be detected in the genotyping process. It was also necessary to allow for the fact that sub-optimal storage conditions (such as temperature) frequently occurs in the field and will lead to DNA template breakages, and there is periodical absence from the peripheral blood of sequestered parasites. Consequently, the limit of detection will be much higher than 1 parasite per 1µl of blood. It was therefore assumed 10 to 20 parasites per µl would be required to reliably contribute a genetic signal and ensure its detection, corresponding to a total parasitaemia of $5 \times 10^7$ to $10^8$; the upper limit i.e. $10^8$ was selected to ensure reliable detection of that clone and because it is consistent with the microscopy detection limit. The sampling limit was not varied in **chapter 3**, but it's impact was later investigated in simulations of Amplicon sequencing (**chapter 5**) – a lower sampling limit was not found to affect results. More blood being introduced into PCR (for example, if venous blood was taken) would decrease the sampling limit; less blood being introduced would increase the sampling limit.

The magnitude of the genetic signal that will be produced by each malaria allele in the blood sample was proportional to the number of parasites carrying that allele.

An inherent feature of PCR is "template competition" i.e. the relative detectability of alleles at each loci depended on their length, with shorter length alleles being more detectable due to their being better amplified in the PCR process [106] . A linear relationship between allele length and relative detectability was assumed; this was done for simplicity but other relationships, for example log-linear, could also be investigated. The shortest allele in each case was assumed to have a relative detectability of 1 while the longest had a relative detectability of 0.001 i.e. it was assumed the shortest allele generated a thousand times the genetic signal of the longest. In *vitro* experiments have mixed ratios of parasite clones of given lengths, quantifying the ratio at which the longer allele is no longer detectable [106]. Such experiments do not directly quantify the relative detectability – the thousand times detectability assumption was made based on Figure 2 of [106], where a 177 base pair length allele of *msp-1* (K1) was only not detected in a mixture with a 248 base pair length allele (3D7) when the ratio of K1:3D7 was 1:1000. This ratio differed for other families (and for *msp-2* and *glurp* – see [106]) but the largest ratio (1:1000) was translated into a relative detectability assumption for the purpose of these simulations. The sensitivity of the results to this relative detectability was tested by shortening it to 0.1; it is later shown that altering this assumption did not affect the conclusions of the model.

Families within *msp-1* and *msp-2* were assumed to be amplified by separate reactions (i.e. are not multiplexed), so the effect only occurred between alleles within the same families (glurp does not have families so the effect applied to all alleles).

The strength of the genetic signal contributed by an allele in a given blood sample was therefore the product of two factors: The number of parasites carrying the allele times the detectability of the allele. Note that genotyping detects alleles, not parasites. Hence, if two (or more) clones within the infection shared the same allele, the signal for that allele was based on the total number of parasites in the two (or more) clones. The final step was to recognise that, in practice, if one allele makes up a large proportion of the genetic signal, then the smaller signals from 'minority' alleles would be rejected as background "noise" (see **3.1.2**). This threshold at which minority alleles were ignored was assumed to be 25% i.e. that signals from alleles that were less than 25% of the highest allelic signal were rejected as "noise", though other values of this parameter were tested (5% and 30%).

### 3.2.6 Classifying patients according to therapeutic outcome.

Analysis of parasitaemia during patient follow-up and, if required, application of molecular correction algorithms to recurrent infections. Four molecular correction algorithms (and a non-PCR corrected "algorithm") were investigated. The current "WHO/MMV" algorithm [14], a "no glurp" algorithm that only considers msp-1 and msp-2, a "≥ 2/3 markers" algorithm that considers msp-1, msp-2 and glurp but requires matching alleles at only two loci to classify a recrudescence, and an "allelic family switch" algorithm that considers only msp-1 and msp-2 and requires a family shift to classify a recrudescence if the loci are discordant (i.e., one has shared alleles between the initial and recurrent infections and one does not). Full details of these algorithms are presented in **Table 3.1**; they enabled each patient to be classified across four groups as would occur in a real trial i.e.

A patient was classified as an early treatment failure (ETF) if a recurrence occurs on or before day 7; note that all such recurrences are regarded as drug failures and molecular correction is not required. In these simulations, on day 3, if total parasitaemia exceeded $10^8$ but was <25% of the total parasitaemia of the initial sample, the patient continued in the trial per the WHO protocol (consequently, no genotype was taken of the day 3 sample and no classification was made); if parasites were present at >25% of initial parasitaemia, that patient was classified as an early treatment failure, consistent with the WHO procedure [28] . For the purposes of estimating failure rates in this methodology, early treatment failure and recrudescence were not distinguished, as both are indicative of drug failure.

A patient was classified as a drug failure if they had recurrent parasitaemia, and the recurrence was classified as a recrudescence by a PCR-correction algorithm on **Table 3.1**.

A patient was classified as reinfection if they had recurrent parasitaemia, and the recurrence was classified as a reinfection by a PCR-correction algorithm on **Table 3.1**.

A patient was classified as 'Cleared' if no recurrent parasitaemia was detected during follow-up; in these cases, the drug was assumed to have successfully killed all parasites present at time of treatment.

A key objective of this research was to investigate how well the classification algorithms recovered the true status of recurrent infections. Therefore, the following definitions were assigned according to parasitaemia data produced by the mPK/PD model:

True recrudescence was defined as a recurrent infection that contained at least $10^8$ parasites from a clone present at time of treatment (this patient is, by definition, a drug failure). This included patients who have a 'mixed' infection on the day of recurrence i.e. possessed malaria clones that survived treatment plus reinfection clones that were acquired during follow up, providing the former exceed $10^8$; note that all clones contributed to the genetic signal of the recurrence as described above.

True reinfection was defined as a recurrent infection whose blood sample contained only parasites from clone(s) that were reinfection(s) (note that such patients may harbour parasites from original clones if these clones were sub-patent i.e. less than $10^8$ parasites).

It was possible that recrudescent clones may not have reached microscopically detectable levels (i.e. parasite numbers are $<10^8$) on the final day of follow-up; such patients would be classified as "cleared" in vivo as recurrent infection would not be observed and thus, the patient would be considered a treatment success. However, simulated data have confirmed that it is possible for some patients to still harbour parasites below detection level at the end of follow up [129]. This modelling approach classifies these patients as drug failures.

### 3.2.7. Estimating drug failure rates in simulated TES.

The model was run for a cohort of 5,000 patients (although any number can be simulated). This is an unrealistically high number for an in vivo clinical trial but is ideal for the purpose of this research and utilizes the advantages of a modelling approach: A true drug failure rate of 10-12% provided a large number of recurrences (the exact number varied depending on the ACT, FOI, and length of follow up) that can be tested against the various classification algorithms to reduce any uncertainty around results.

The four patient outcomes described above were used to calculate the estimated drug failure rate, $\hat{F}$ in the same manner as outcomes reported in vivo. It was assumed, for simplicity, that no patients were lost to follow-up or removed from the trial for any reason other than recurrent parasitaemia. There were three methods for calculating failure rates which differed in how they processed patients with recurrent parasitaemia that had been classified as reinfections, noting that all patients with recurrent parasitaemia would, in vivo, be re-treated with another antimalarial (for ethical reasons) and removed from the trial. The three methods were: A non-PCR corrected failure rate, a "per protocol" failure rate and a failure rate obtained using survival analysis. The latter two methods are recommended by the WHO to analyse anti-malarial drug trials [14, 28]. Technically, they were calculated as follows using the following nomenclature:

$C_o$ was the number of patients who cleared infection.

$nI_o$, was the number of patients whose recurrent infections were classified as reinfections.

$\hat{F}$ was the estimated drug failure rate.

N was the total number of patients.

(i) The non-PCR corrected failure rate was obtained by considering all patients with recurrent infections as patients who had failed drug treatment. This method did not require distinguishing between reinfections and recrudescent infections. The failure rate $\hat{F}$ could then be estimated as:

$$\widehat{\boldsymbol{F}} = \mathbf{1} - \frac{\boldsymbol{C_o}}{\boldsymbol{N}} \qquad \textbf{Equation 1}$$

(ii) The 'per protocol' method, recommended by WHO [14, 28, 29], simply removed patients who were classified as reinfections from the total number of observations i.e.:

$$\widehat{\boldsymbol{F}} = \mathbf{1} - \frac{\boldsymbol{C_o}}{\boldsymbol{N} - (\boldsymbol{nI_o})} \qquad \textbf{Equation 2}$$

(iii) Survival analysis, as recommended by WHO [14], used the survivor function from a Kaplan Meier plot on the final day of follow-up, right-censoring reinfections.

The Kaplan-Meier estimator (KM) of survivorship at time t was obtained as:

$$\widehat{S}(t) = \prod_{ti \le t} \frac{ni - di}{ni} \qquad \textbf{Equation 3}$$

Where t was a vector of all timepoints i.e. days of follow-up in which an event occurred in the study population, ni was the number of individuals at time ti who remained uninfected, and di was the number of events (drug failures in this case) that occurred at timepoint ti. Plainly, what this method did was calculate the proportion of patients who remained free of recrudescence between consecutive days of follow up, then multiplied all these time periods to obtain the overall probability of 'surviving' recrudescence-free over the whole follow-up period. The advantage was that even those patients who are "censored" (by acquiring a reinfection and leaving the study) will still contribute to the analysis through their inclusion prior to their removal.

The estimator at the final time-point (i.e. the last day of follow-up) was the probability that their treatment was considered a 'success' at the end of the trial. Consequently, it's complement gave the probability that a given individual will fail treatment i.e.

$$\hat{F} = 1 - \hat{S}(t) \qquad \text{Equation 4}$$

The final methodological step was to interrogate the modelled data to determine the "true failure rate" – i.e., the drug failure rate calculated directly from the parasitaemia of each patient (thus, not dependent on genotyped data). For each patient in the simulation, an outcome on the final day of follow-up was determined: If, on the final day, the patient had any parasites from any initial clones (i.e. even a single parasite), the patient was denoted as a drug failure. If no parasites had survived from the initial clones present at treatment, that patient was denoted as a treatment success.

The true failure rate, F, for the patient population was then calculated:

$$F = \frac{f}{N} \qquad \text{Equation 5}$$

Where f was the number of drug failures on the final day of follow-up and N was the total number of patients.

This was the "gold standard" metric and cannot be obtained in vivo. It was compared to the estimated failure rates obtained from modelling the clinical trial and molecular correction process and allowed us to quantify the accuracy of different methods (i.e., their ability to recover the true failure rate).

### 3.2.8 Reanalysis of existing in vivo data with molecular correction algorithms.

Given that it is the WHO-recommended method, failure rate estimates for TES in the published literature have historically been calculated following analysis with the WHO/MMV algorithm. Failure rate estimates can always be calculated using the no correction algorithm (Equation 1) as only the number of recurrences and the total patient number are needed. However, it is not simple to track through historic literature and re-calculate failure rates with the no glurp, ≥ 2/3 markers or allelic family switch methodologies. The reason for this is that as the WHO/MMV algorithm requires a match between the initial and recurrent sample at msp-1, msp-2, and glurp to classify a recrudescence, for efficiency in the field laboratory practice is generally to genotype loci individually until there is a discordance (no matching allele) between samples at a locus, at which point the infection is

considered a recurrence and subsequent loci are not genotyped. In short, much data is technically incomplete in that the recurrent sample is not genotyped at every loci. Collaborators were able to provide me with data-sets with large numbers of "complete" recurrences (i.e., genotyped at all three loci) such that the failure rate estimates with all algorithms could be calculated.

Clinical data was obtained from Rwanda (a relatively high transmission area) across 6 sites between 2013 and 2015, where patients were treated with either AR-LF or DHA-PPQ and genotyped at msp-1, msp-2 and glurp. In patients treated with AR-LF, 137 recurrences were observed, of which 110 could be classified as either a reinfection or a recrudescence (it was not possible to classify 27 patients because they had incomplete genetic data). In patients treated with DHA-PPQ, 48 recurrences were observed, of which 43 could be classified as either a reinfection or a recrudescence (it was not possible to classify 5 patients because they had incomplete genetic data). This data was initially presented internally to the National Malaria Control Programme in Rwanda (a manuscript describing clinical efficacy studies for publication is in preparation).

Clinical data from Cambodia (a relatively low transmission area) was obtained from 6 sites between 2014-2016. Patients were treated with either artesunate plus amodiaquine (AS-AQ), artesunate plus pyronaridine (AS-PYN) or DHA-PPQ, and genotyped at msp-1, msp-2, and glurp. In patients treated with AS-AQ, 12 recurrences were observed, of which 5 could be classified as reinfection or recrudescence (7 patients had incomplete genetic data). In patients treated with AS-PYN, 14 recurrences were observed, of which 12 could be classified as reinfection or recrudescence (2 had incomplete genetic data). In patients treated with DHA-PPQ, 67 recurrences were observed, of which 48 could be classified as reinfection or recrudescence (19 had incomplete genetic data). This data was initially presented internally to the National Malaria Control Programme in Cambodia. A description of the AS-PYN trials has already been published [130].

For all data, the genetic signals (i.e., the msp-1, msp-2 and glurp alleles at the initial sample and any recurrent sample) were re-interpreted using the novel molecular correction algorithms described in **Table 3.1** to investigate how varying the molecular correction algorithm changed the classification (as reinfection or recrudescence) of patients and, consequently, failure rate estimates.

## 3.3 Results.

### 3.3.1 Analysis of failing DHA-PPQ (two compartment model).

**Figure 3.1** shows the failure rates obtained from simulated DHA-PPQ clinical trials using four molecular correction algorithms and the non-corrected algorithm (**Table 3.1**), with a follow-up length of 42 days. Both the true failure rate and the estimated failure rate are presented (calculated using survival analysis) as a function of FOI.

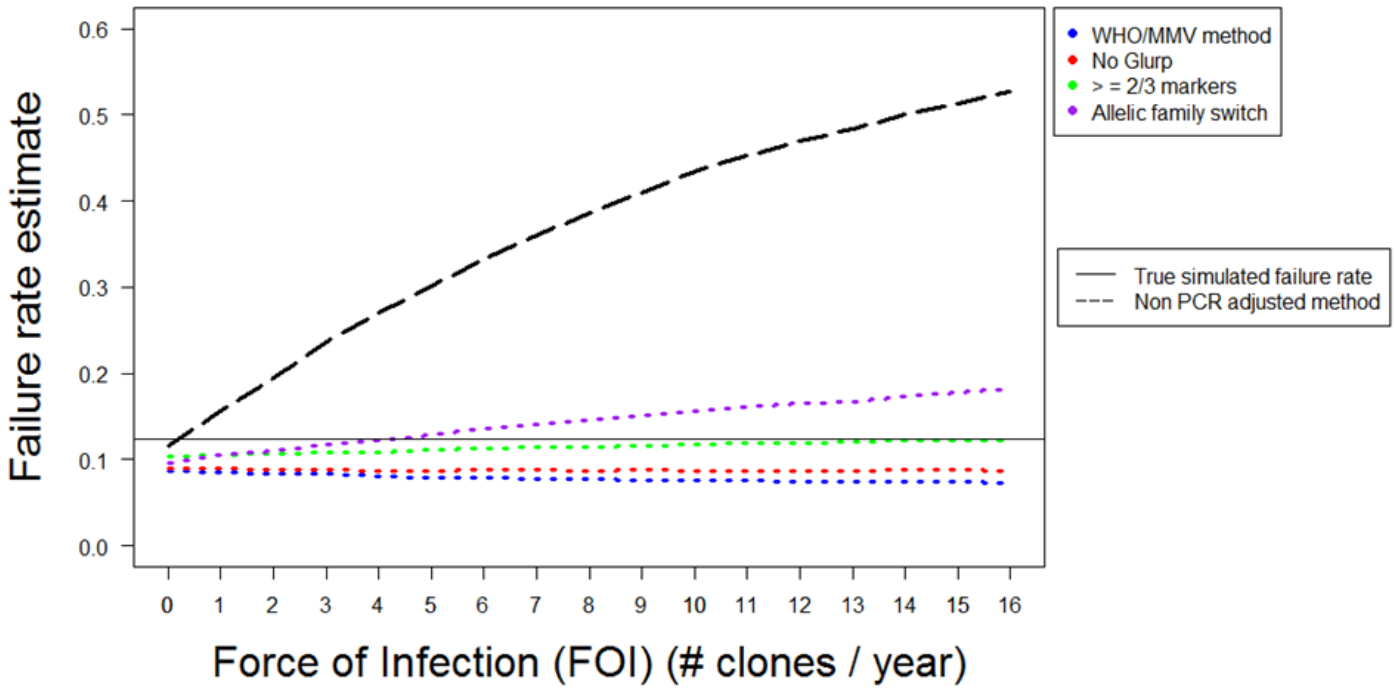



**Figure 3.1** Analysis of simulated trial data for DHA-PPQ with a follow-up period of 42 days. Estimated failure rates are shown for the different algorithms of molecular correction (**Table 3.1**) as a function of Force of Infection (FOI) and are calculated using survival analysis.

The non-corrected algorithm always produced a higher failure rate estimate than any of the four molecular correction algorithms. Failure rate estimates using no correction rose rapidly as FOI increased and at moderate and high levels of transmission estimated failure rates were substantially greater than the true failure rate: At high transmission intensities (FOI of 16) estimated failure rates produced by this algorithm were above 50% - a clear over-estimate of the true failure rate (12%): This pattern occurred because all the additional reinfections that occurred at as FOI increased were misclassified as recrudescence. Conversely, in the absence of any reinfections (when FOI=0), the non-corrected algorithm produced an accurate failure rate estimate by correctly classifying all recurrences as recrudescence (leaving only a slight under-estimate due to patients who had recrudescent parasites at levels of $<10^8$, such that no recurrence occurred during follow-up).

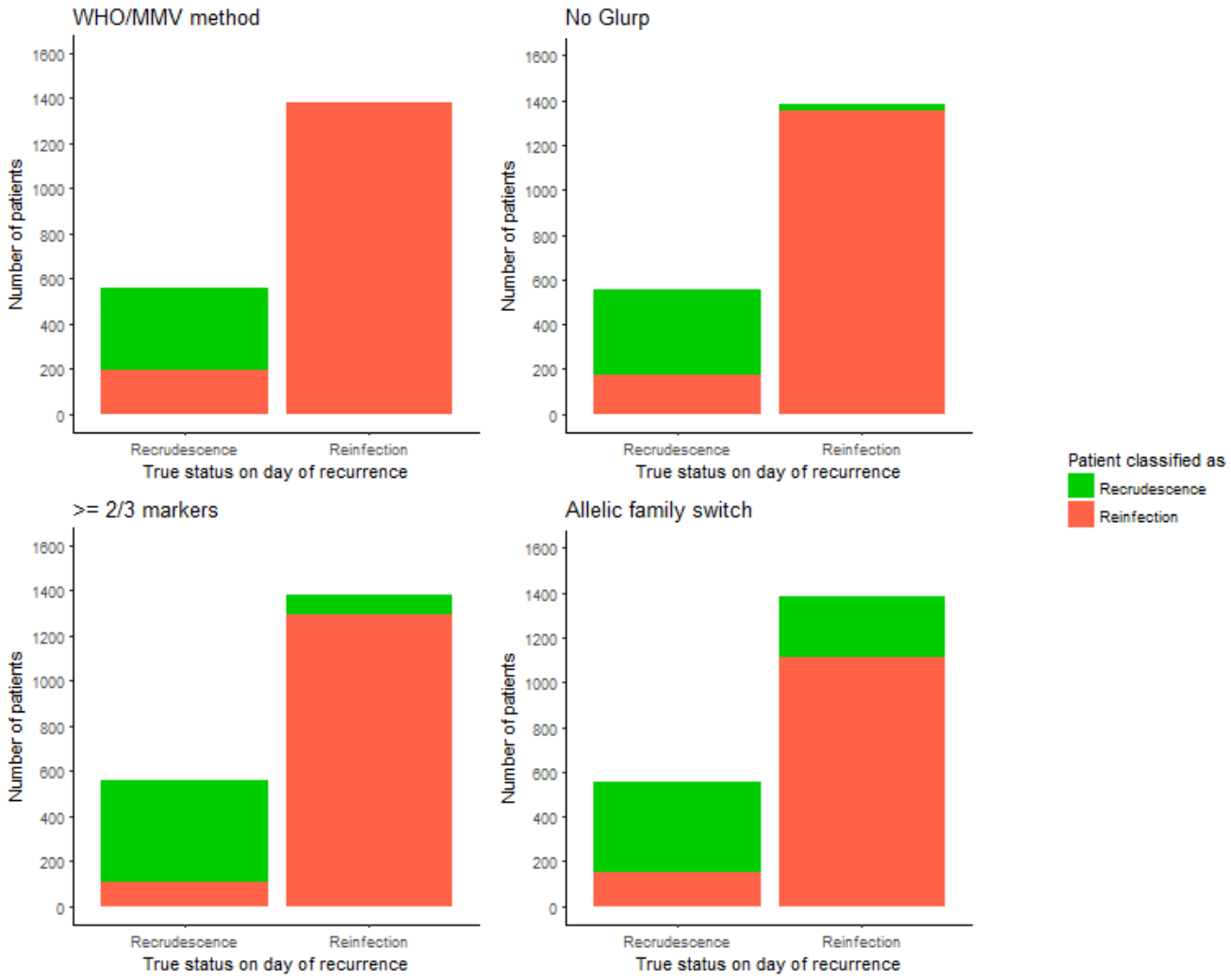



**Figure 3.2** Figure showing the ability of the various molecular correction algorithms to correctly classify patients with recurrent malaria. The data are for DHA-PPQ with a 42-day follow-up obtained with a FOI of 8 (i.e. used to obtain the results shown at FOI=8 in **Figure 3.1**). The X-axis shows the true status of patients on the day of recurrence (i.e. reinfection or a recrudescence) and the colour-coding shows how these patients were classified by each algorithm. The WHO/MMV recommended algorithm correctly classifies nearly all reinfections, but misclassifies around one third of recrudescences. The no glurp algorithm is similar to the WHO/MMV one; it misclassifies only a small number of reinfections, but misclassifies around a third of recrudescences. The ≥2/3 markers algorithm had fewer misclassifications and was also more balanced i.e. misclassified a similar proportion of both reinfections and recrudescences. Finally, the allelic family switch algorithm correctly classifies a large proportion of recrudescences but misclassifies around half of reinfections.

The ability of the four molecular correction algorithms to accurately estimate drug failure rates depended on their ability to correctly classify recrudescences and reinfections. This ability is shown (for an FOI of 8, i.e. a moderate transmission area) in **Figure 3.2**. Each algorithm misclassified some proportion of recrudescences and reinfections. The number of recrudescence misclassified as reinfections was consistent as FOI changed, but the number of reinfections misclassified as recrudescence increased as FOI increased – results shown in **Figure 3.3** (note that while results for all parameterizations DHA-PPQ and, later, AR-LF and AS-MQ are not shown, the proportion of misclassification occurring was extremely robust between drugs). General trends were extremely clear:

The WHO/MMV algorithm consistently under-estimated failure rates at all transmission intensities as shown in **Figure 3.1** . The algorithm frequently failed to detect drug failures i.e. it misclassified around 40% of recrudescent infections as reinfections (**Figure 3.2**). These misclassifications occurred because of failure to detect recrudescent alleles in either the initial or recurrent blood sample – this algorithm was so stringent (requiring matching alleles at all three loci) that even missing a single allele could result in misclassification. As FOI increased, the estimated failure rate did not change to any meaningful extent because the algorithm correctly classified nearly all reinfections (**Figure 3.2**).

The no glurp algorithm produced slightly higher estimated failure rates than the WHO/MMV algorithm across all FOI settings (**Figure 3.1** ). This occurred because recrudescences were slightly less likely to be misclassified as reinfections while reinfections were slightly more likely to be misclassified as recrudescences than under the WHO/MMV algorithm (**Figure 3.2**). At low FOI, this difference was small; the high allelic diversity of *msp-1* and *msp-2* meant misclassification of reinfections as recrudescences was rare. The difference between the no glurp algorithm and the WHO/MMV algorithm increased as FOI increased, but, like the WHO/MMV algorithm, the no glurp algorithm always under-estimated the true failure rate.

The ≥ 2/3 markers algorithm produced higher estimated failure rates than the no glurp algorithm across all FOI levels. This occurred because this algorithm reduced the chance of a recrudescence being misclassified as reinfection (due to failure to detect recrudescent alleles) and increased the chance of a reinfections being misclassified as a recrudescence (**Figure 3.2**). Both effects occurred because only needing matching alleles between samples at 2/3 loci gave the algorithm some tolerance to un-detectable alleles.

The allelic family switch algorithm produced higher estimated failure rates than the ≥ 2/3 markers algorithms at all but the lowest FOI (0-2) settings (**Figure 3.1** ). A complete family switch in *msp-1* or *msp-2* in a discordant sample (**Table 3.1**) would be sufficient to classify a recrudescence; this led to a similar number of recrudescence being correctly classified as the ≥ 2/3 markers algorithm, but this algorithm misclassified the largest number of reinfections as recrudescence out of all the molecular correction algorithms – the family switch could still occur (by chance); the difference in numbers misclassified between the no glurp algorithm and the allelic family switch algorithm is the result of this misclassification by chance.

**Figure 3.2** shows the misclassification of recurrent infections (recrudescence classified as reinfection and vice versa) for an FOI of 8. **Figure 3.3** shows the same plot for an FOI of 2 and 16. These plots show that the number of recrudescence misclassified as reinfection is stable as FOI increased for all algorithms. Furthermore, it shows that increased FOI had nearly no impact on the number of reinfections being misclassified for the WHO/MMV algorithm (which correctly classified all reinfections), and a very minor impact for the no glurp algorithm. For the ≥ 2/3 markers and allelic family switch algorithm, this figure demonstrates that increased FOI led to greatly increased numbers of reinfections being misclassified as recrudescence. The proportion of reinfections misclassified was stable as FOI increased, but the greater total number of misclassifications produced the increased failure rates seen with these algorithms in **Figure 3.1** .

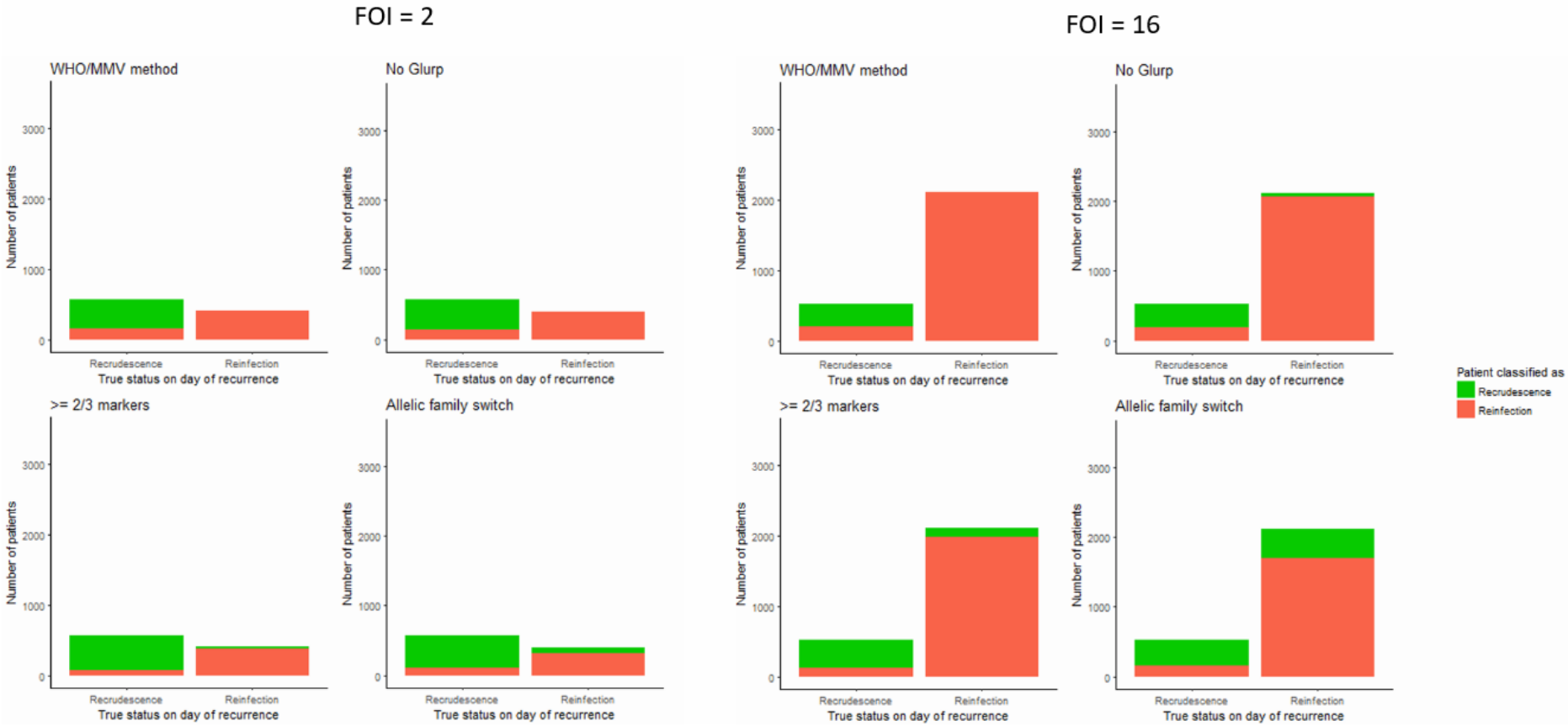
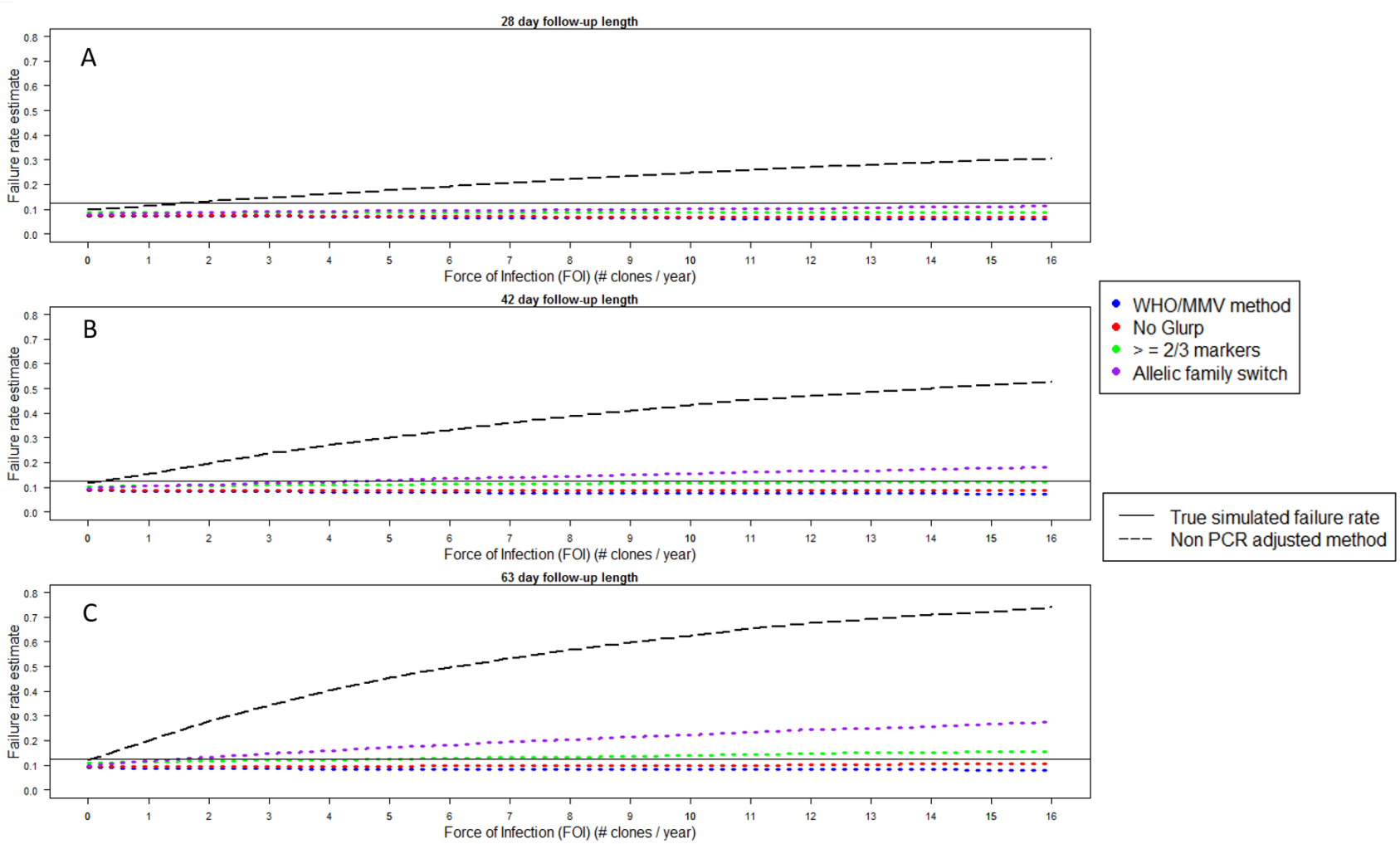



**Figure 3.3** Figure showing the ability of the various molecular correction algorithms to correctly classify patients with recurrent malaria. The data are for DHA-PPQ with a 42-day follow-up obtained with FOIs of 2 and 16 (see **Figure 3.1** for an FOI of 8), showing how misclassification by each algorithm alters as FOI changes. The X-axis shows the true status of patients on the day of recurrence (i.e. reinfection or a recrudescence) and the colour-coding shows how these patients were classified by each algorithm.

Alternate durations of follow-up length were simulated for DHA-PPQ and their impact on estimated failure rates are shown in **Figure 3.4** for 28, 42 and 63 days of follow-up. Longer durations of follow-up led to larger estimated failure rates for all algorithms. This occurred because longer follow-up (i) allowed more time for recrudescences to become detectable, (ii) allowed more reinfections to emerge, some of which were misclassified as recrudescences (**Figure 3.3**).

Under-estimation of the true failure rate occurred with all algorithms when a 28-day follow-up period was chosen. With a 42-day follow-up period, the allelic family switch algorithm produced the most accurate failure rate estimate with an FOI of <7, and the ≥ 2/3 markers algorithm produced the most accurate failure rate estimate with FOI ≥7. As length of follow-up increased to 63 days, the ≥ 2/3 markers algorithm tended to slightly over-estimate the failure rate. This effect was more apparent as FOI increased. These patterns emerged because only a small number of initial clones recrudesced after 42 days. **Figure 3.5** shows the proportion of recurrent infections on each day of the follow-up period that were truly recrudescent or reinfections. On days 49, 56 and 63, the number of recurrent infections that were truly recrudescent was small. Almost all recurrent infections on these days were reinfections and consequently, inclusion of these three extra days of follow-up inflated the estimated failure rate due to misclassification of these reinfections as recrudescences (as alleles were shared by chance between these reinfections and the initial blood sample). However, the increased failure rate of a 42 day follow-up compared to a 28 day follow-up (due to both detection of true recrudescence and misclassification of extra reinfections) meant that a 42 day follow-up period analysed with either the ≥2/3 markers or allelic family switch algorithm produced more accurate failure rate estimates than the WHO/MMV algorithm.




1407

1408

1409 **Figure 3.4** Analysis of simulated trial data for DHA-PPQ showing the impact of changing follow-up period with follow-up lengths of (A) 28 days, (b) 42 days
1410 (as in **Figure 3.1** ), and (C) 63 days. Estimated failure rates are shown the different molecular correction algorithms (**Table 3.1**) as a function of FOI and
1411 calculated using survival analysis.

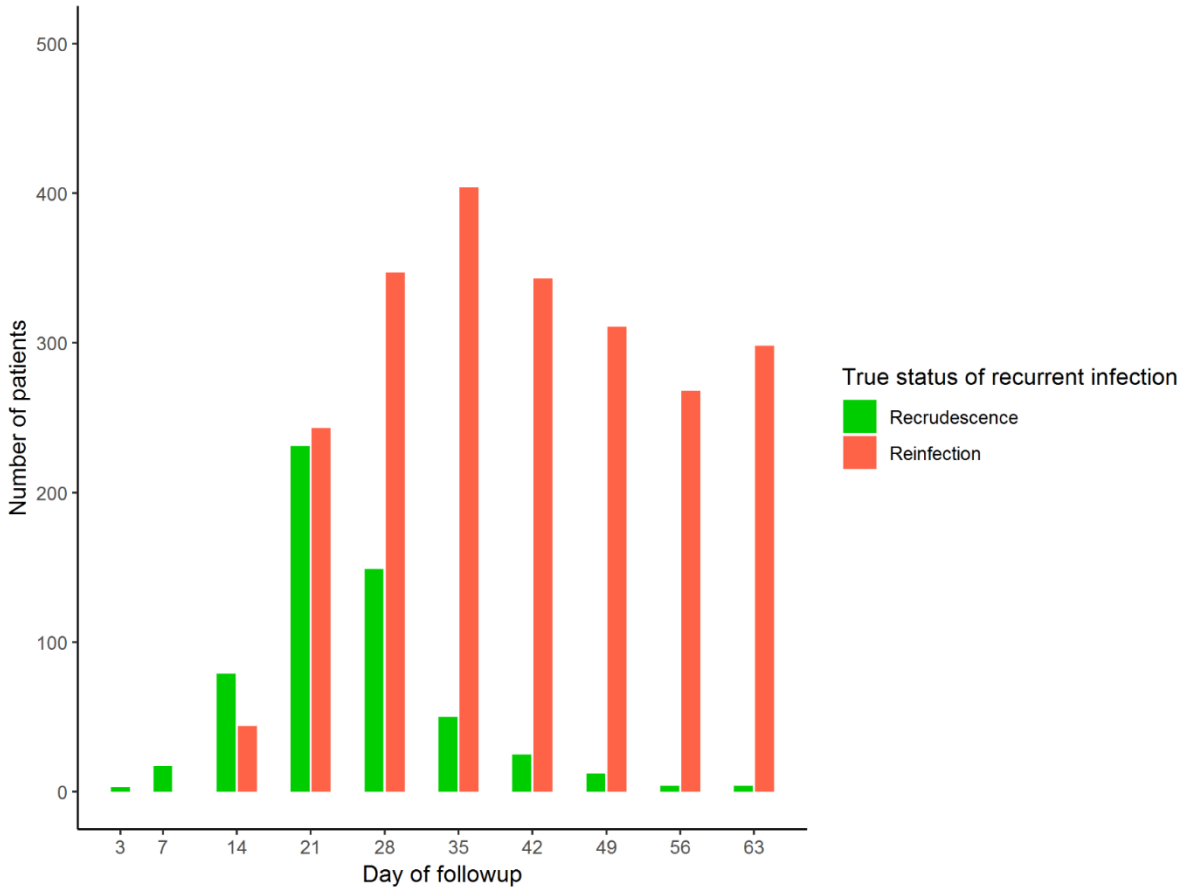



**Figure 3.5** The true status of recurrent infections on each day of follow-up for a simulated trial of DHA-PPQ with a true failure rate of 12% and an FOI of 8. The total height of the bars indicates the number of recurrent infections detected on that day of follow-up, and the color-coding shows the number of those recurrent infections that were truly recrudescent or reinfections.

### 3.3.3 Analysis of failing DHA-PPQ (three compartment model)

The principal difference between the parasite dynamics generated for DHA-PPQ under the assumptions of a two-compartment for PPQ and a three-compartment model for PPQ is that the three-compartment model is slightly more prophylactic and has a greater total area under the drug kill curve (see **2.1.1**); consequently, true failure rate is slightly lower, and a smaller number of reinfections become patent (i.e., survive to become detectable). However, failure rate estimates obtained using each algorithm are not significantly different between the two compartment and three compartment models; results (the relative performance of molecular correction algorithms) are qualitatively the same with both model calibrations. I am not attempting to comment, here, on whether DHA-PPQ is best represented by a two or three compartment model or its exact parameterization, but rather confirm and stress the consistency of the molecular correction algorithms across both parameterizations, suggesting that, regardless of the number of PPQ PK compartments included, conclusions regarding the accuracy of these molecular correction algorithms to estimate treatment failure rates are robust.

Parasite dynamics for each patient using a three-compartment model calibration for DHA-PPQ (rather than the two-compartment calibration in **Table 2.4**), i.e. **Table 2.3**. Results are shown in **Figure 3.6** , the qualitative patterns were the same as for the two-compartment model, i.e., that WHO/MMV algorithm produced the lowest failure rate estimate, then no glurp, then ≥ 2/3 markers, then allelic

family switch (at most FOI, ≥ 2/3 markers produced a slightly higher failure rate estimate at 0-2 FOI). True failure rate was slightly lower for the three compartment model (10% vs 12%). The relative failure rate estimates of the algorithms and the no-correction approach were the same – i.e., that WHO/MMV algorithm produced the lowest failure rate estimate, followed by no glurp, ≥ 2/3 markers, and the allelic family switch algorithm. Failure rate estimates are lower across all algorithms than with the shorter-prophylaxis two-compartment model, and a 63-day follow-up appears to be the most suitable under this calibration; the ≥2/3 markers algorithm produced an accurate failure rate estimate at all but the lowest FOI levels with this follow-up length). Crucially, the key message is the same: The WHO/MMV algorithm under-estimates true failure rate and other algorithms can produce more accurate failure rate estimates. Perhaps the most interesting difference between the two DHA-PPQ compartment assumptions is that they suggested, given use of the same molecular correction algorithm, different optimal length of follow-up.

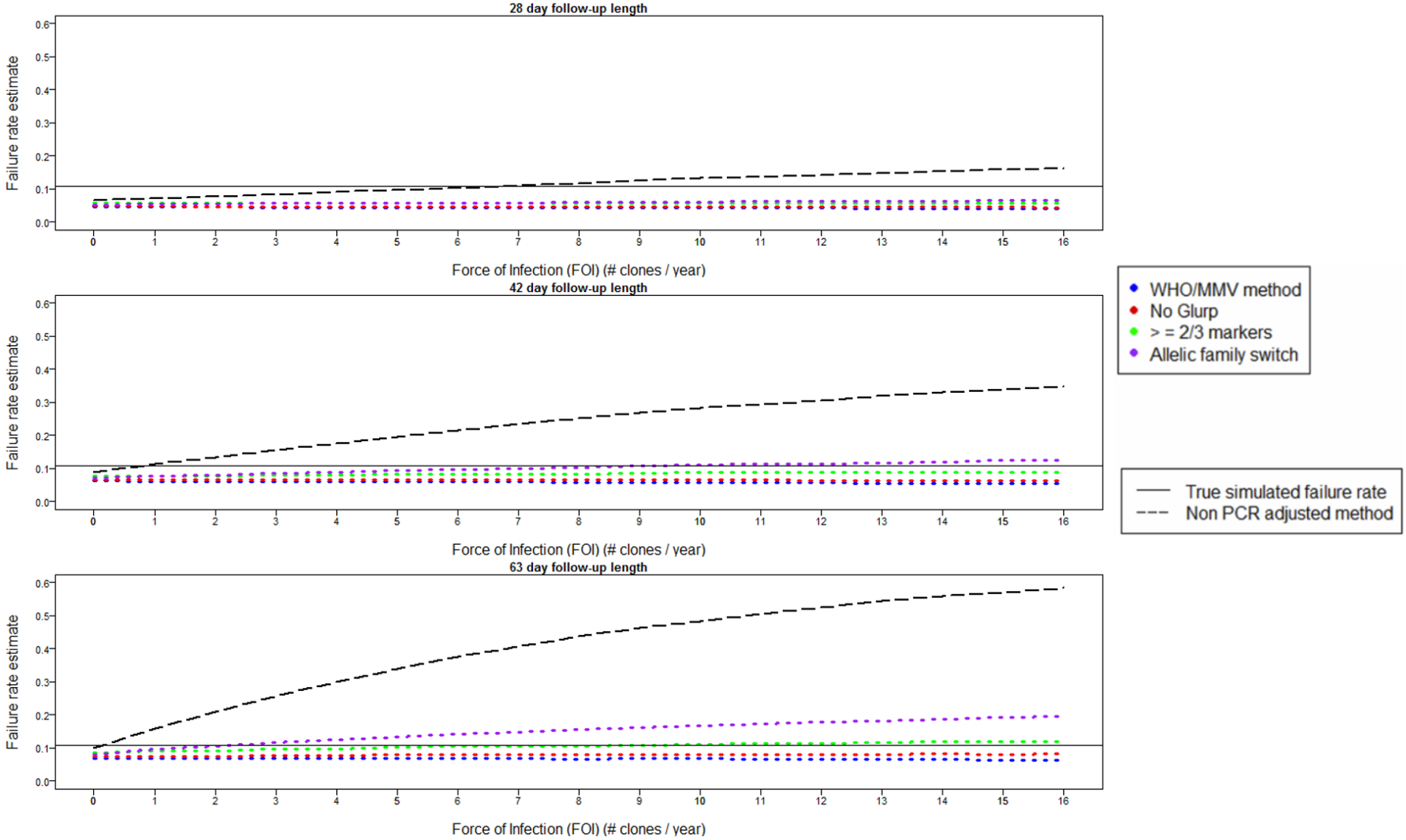



**Figure 3.6** Analysis of simulated trial data for DHA-PPQ using a three compartment model with follow-up lengths of (A) 28 days, (b) 42 days and (C) 63 days. Estimated failure rates are shown for the different algorithms of molecular correction as a function of FOI and calculated using survival analysis.

### 3.3.4 Analysis of failing AR-LF.

Failure rate estimates for failing AR-LF for 21-day and 28-day follow-up lengths are presented in **Figure 3.7** . The true failure rate of AR-LF in these simulations was 0.918 (9%). The same pattern was observed as for DHA-PPQ: The non-PCR corrected algorithm over-estimated the failure rate at any FOI higher than 1, and severely overestimated failure rates at high FOI; the WHO/MMV algorithm and the no glurp algorithm slightly under-estimated the failure rate across all levels of FOI. Use of a 21-day follow-up period led to both the allelic family switch algorithm and the ≥ 2/3 markers algorithm under-estimating the failure rate, only at a high FOI of 13 did the allelic family switch algorithm accurately recover the true failure rate. Use of a 28-day follow-up period produced more accurate failure rate estimates: The ≥ 2/3 markers algorithm accurately recovered the true failure rate between an FOI of 5-16, with both the ≥ 2/3 markers algorithm and the allelic family switch algorithm under-estimating the failure rate slightly at lower FOI. These results combined with the true classifications of recurrent infections as recrudescence and reinfections (**Figure 3.8** ) suggested a 28-day follow-up period led to more accurate failure rate estimates.

### 3.3.5 Analysis of failing AS-MQ.

Failure rate estimates for failing AS-MQ for a 42, 49 and 63-day follow-up length are presented in **Figure 3.9**. The true failure rate of AS-MQ in these simulations was 0.1032(10%). With a 42-day follow-up period (**Figure 3.9**(A)), the ≥ 2/3 markers algorithm under-estimated the true simulated failure rate at all FOI settings – the allelic family switch and ≥ 2/3 markers algorithm were close in value up to an FOI of 9-10. As with DHA-PPQ and AR-LF, the WHO/MMV and no glurp algorithms under-estimated the failure rate consistently and using no PCR correction generated a large over-estimate of the true failure rate. A novel follow-up length of 49 days was simulated (**Figure 3.9**(B)) under which the ≥ 2/3 markers algorithm produced a more accurate failure rate estimate than a 42-day follow-up at all FOI levels. With a 63-day follow-up period (**Figure 3.9** (C)), the allelic family switch algorithm over-estimated the true failure rate from an FOI of 4 and upwards. The ≥ 2/3 markers algorithm over-estimated from an FOI of 8 and up, but only by a small amount. AS-MQ is more prophylactic than DHA-PPQ and AR-LF: Given the same period of follow-up, fewer reinfections became patent, and recrudescences occurred later in the follow-up period (**Figure 3.10** ). As such, it was unsurprising that a longer period of follow-up led to more accurate failure rate estimates. Using the ≥ 2/3 markers algorithm and assuming an FOI of <8, a 63-day follow-up period resulted in a more accurate estimate than the 42 and 49-day follow-up lengths, but the differences in estimates between 49 and 63 days were small and the operational, logistical advantages of a 49-day trial over a 63-day trial are likely to be substantial. Furthermore, with an FOI of ≥8, a shorter follow-up (49 days) produced a more accurate failure rate estimate with the ≥ 2/3 markers algorithm – a 63 day follow-up period over-estimated the true failure rate slightly with higher transmission intensity using this algorithm.

### 3.3.6 Analysis of non-failing AR-LF and non-failing AS-MQ.

The simulations were run for the non-failing (i.e. effective drug) PK/PD calibrations for AR-LF (**Figure 3.11**) and AS-MQ (**Figure 3.12**), which had true failure rates of 0.0046 (0.5%) and 0.0208 (2%) respectively. This was to investigate whether the new algorithms could incorrectly identify effective drugs as failing. Crucially, the under-estimate associated with of the ≥ 2/3 markers algorithm was so small in terms of absolute value that the use of the algorithm can be recommended without concern for over-estimating the failure rate of effective drugs i.e. there was no danger in the model of an effective drug being misclassified as failing. The large under-estimate of failure rate that was observed using the WHO/MMV algorithm with failing AR-LF and AS-MQ (**Figure 3.7** ; **Figure 3.9**) was not observed with non-failing drugs as the number of recurrences and the true failure rate were small.

However, these results highlighted the dangers of not using a molecular correction: The non-PCR-corrected algorithm generated estimated failure rates >10% in areas of high FOI when using long durations of follow-up. Detailed investigations of the proportion of infections misclassified (i.e. **Figure 3.2** (for DHA-PPQ)), or the times at which true recrudescence / reinfection occurred (i.e. **Figure 3.5**, **Figure 3.8** , **Figure 3.10** ) were not conducted for non-failing AR-LF and AS-MQ due to the small number of recurrences that occur with non-failing drugs and how close failure rate estimates with all molecular correction algorithms are in absolute value to the true failure rate.

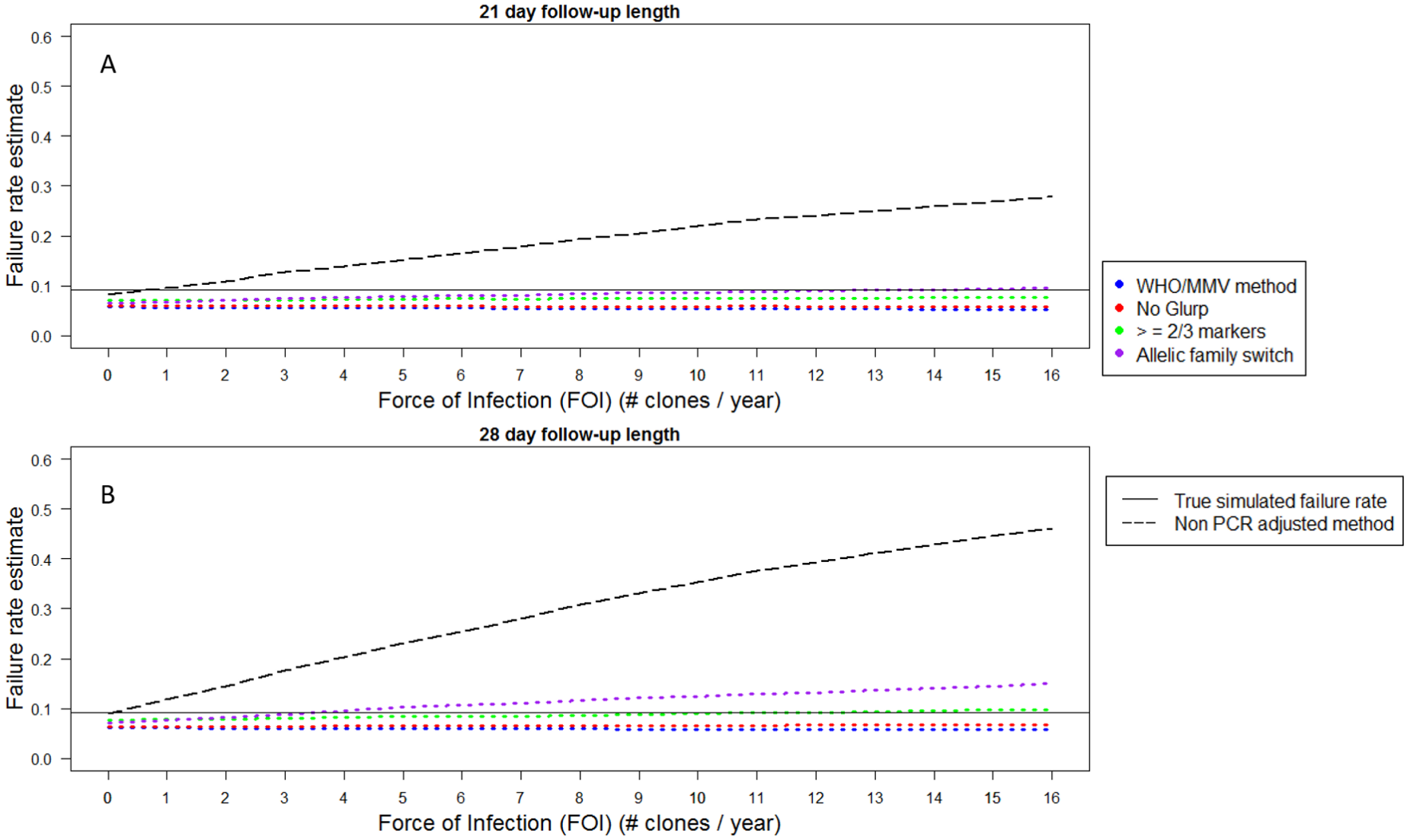



**Figure 3.7** Analysis of simulated trial data for failing AR-LF with follow-up lengths of 21 days (A) and 28 days (B). Estimated failure rates are shown for the different algorithms of molecular correction as a function of FOI and calculated using survival analysis.

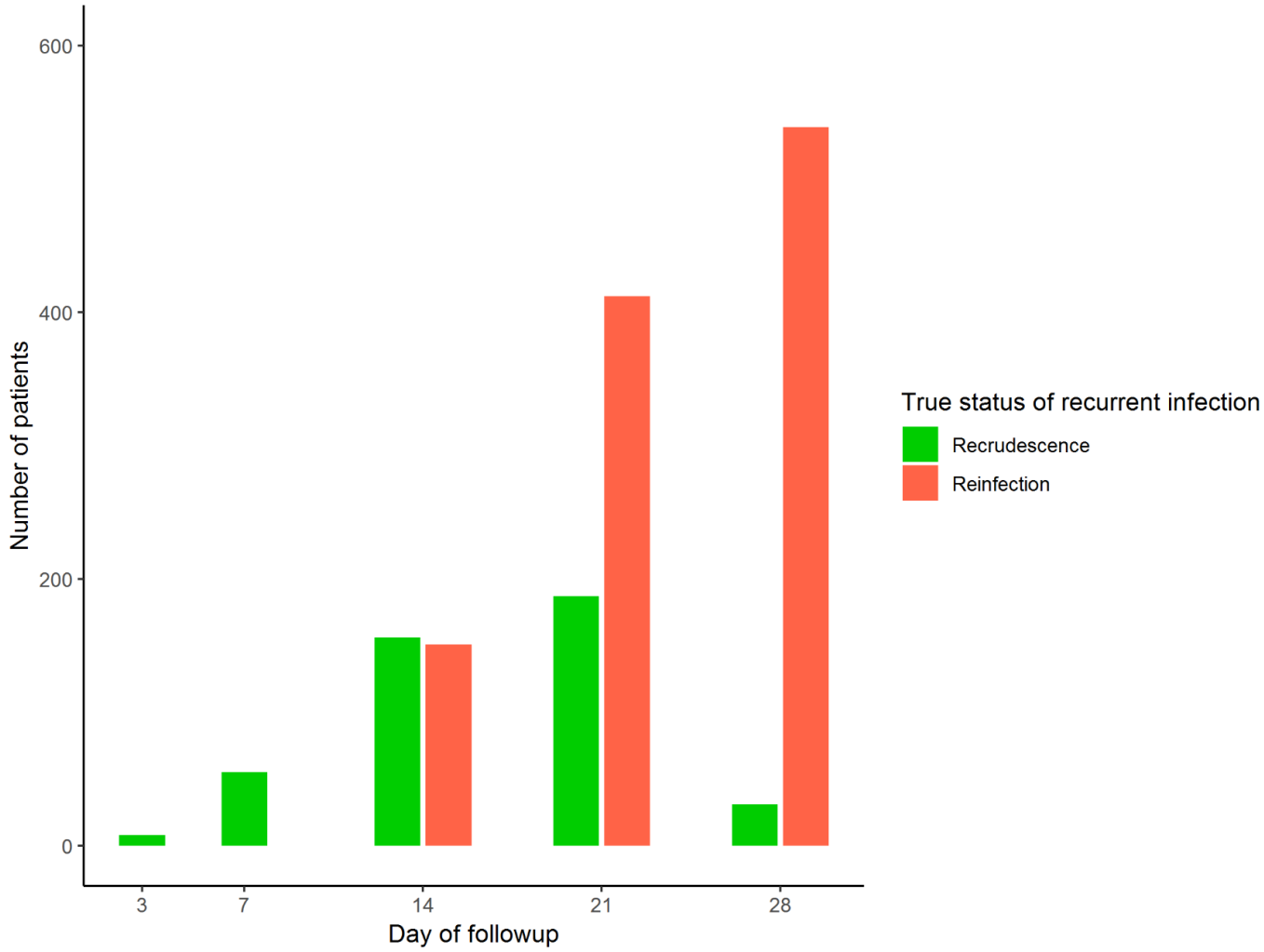



**Figure 3.8** The true status of recurrent infections on each day of follow-up for a simulated trial of AR-LF with a true simulated failure rate of 9% and an FOI of 8. The total height of the bars indicates the number of recurrent infections detected on that day of follow-up, and the color-coding shows the number of those recurrent infections that were truly recrudescent or reinfections.

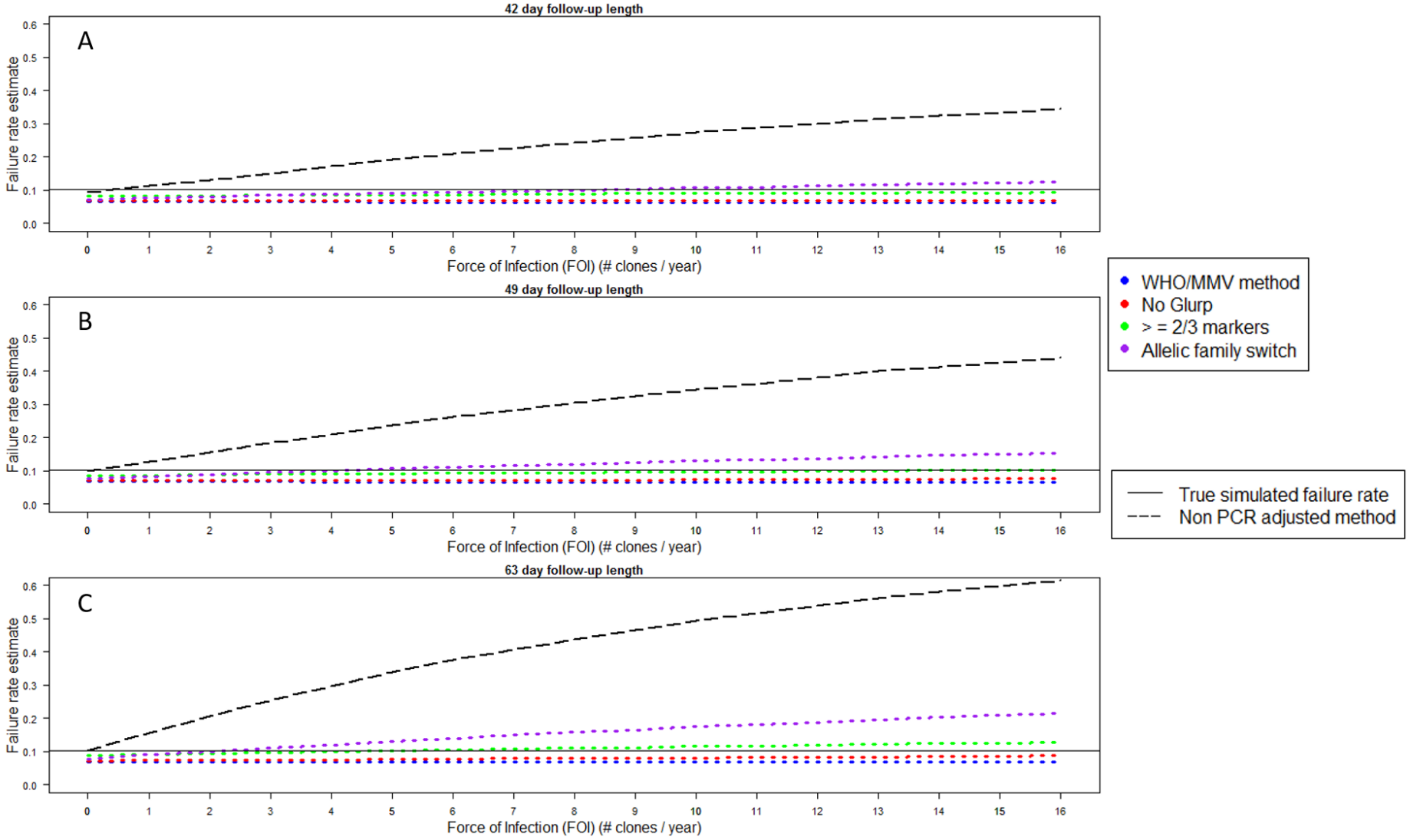



**Figure 3.9** Analysis of simulated trial data for failing AS-MQ with follow-up lengths of 42 days (A), 49 days (B) and 63 days (C). Estimated failure rates are shown for the different algorithms of molecular correction as a function of FOI and calculated using survival analysis.

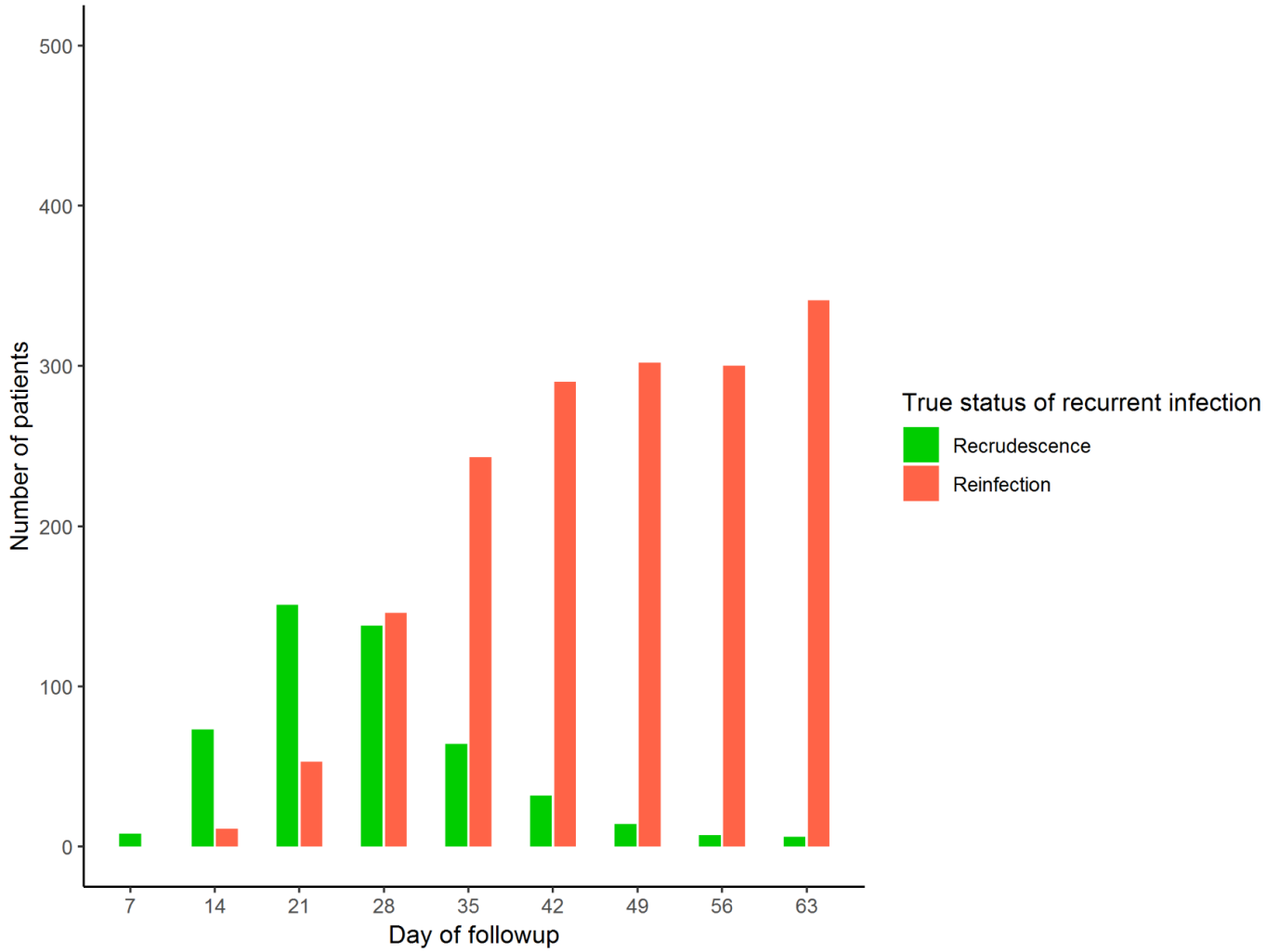



**Figure 3.10** The true status of recurrent infections on each day of follow-up for a simulated trial of AS-MQ with a true simulated failure rate of 10% and an FOI of 8. The total height of the bars indicates the number of recurrent infections detected on that day of follow-up, and the color-coding shows the number of those recurrent infectoins that were truly recrudescent or reinfections.

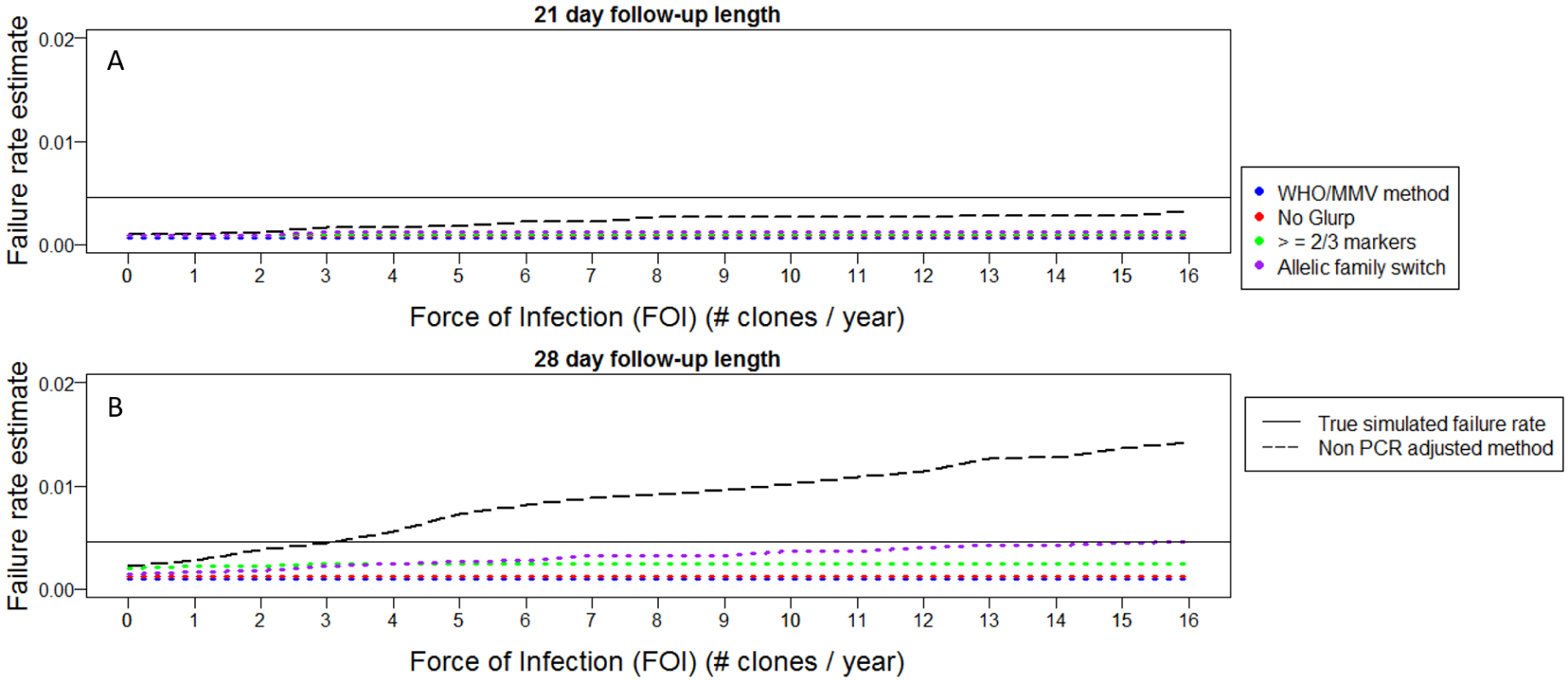



**Figure 3.11** Analysis of simulated trial data for effective AR-LF with follow-up lengths of 21 days (A) and 28 days (B). Estimated failure rates are shown for the different algorithms of molecular correction as a function of FOI and calculated using survival analysis.

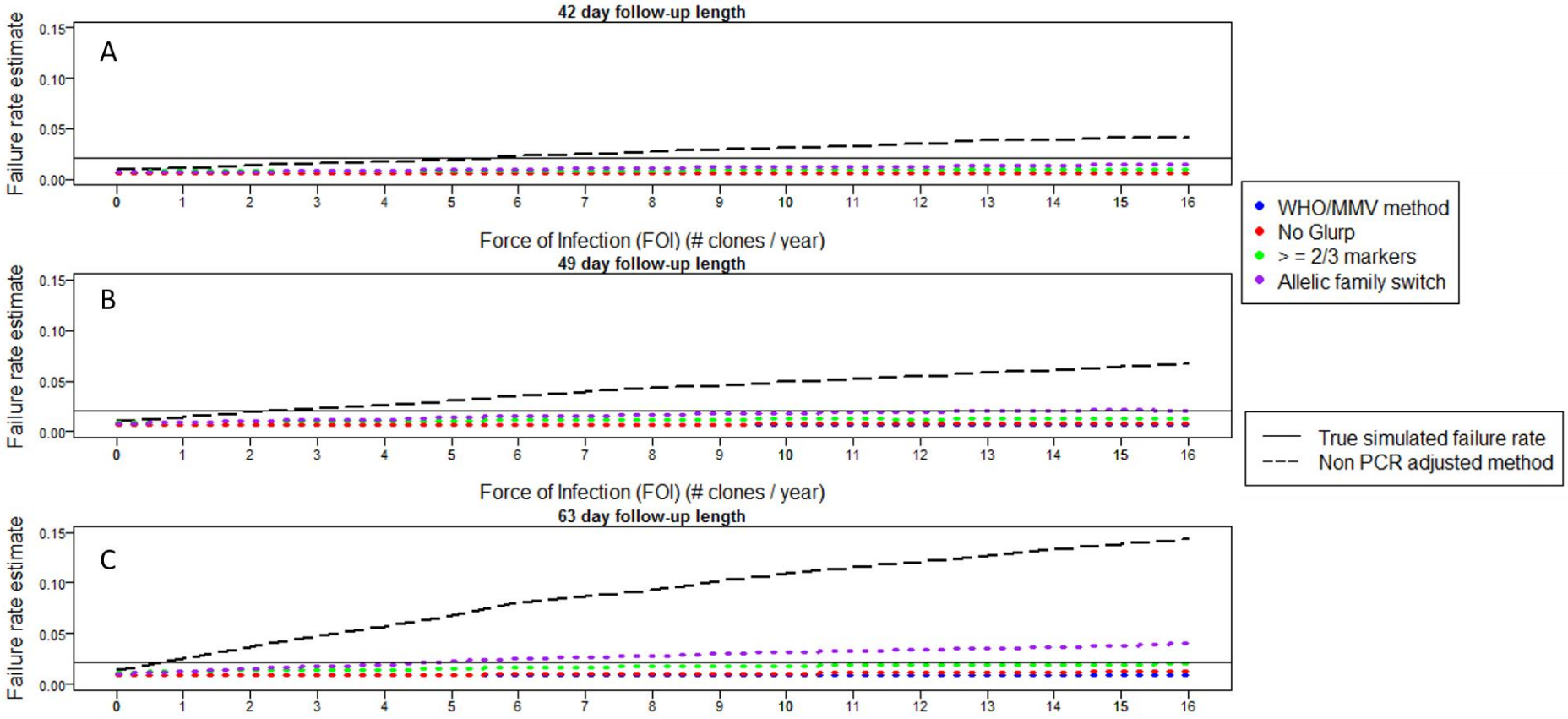



**Figure 3.12** Analysis of simulated trial data for effective AS-MQ with follow-up lengths of 42 days (A), 49 days (B) and 63 days (C). Estimated failure rates are shown for the different algorithms of molecular correction as a function of FOI and calculated using survival analysis.

### 3.3.7 Comparison of failure rate estimates for "per protocol" method and survival analysis

WHO guidelines [28] recommend two methods for statistical analysis of molecular-corrected data: Survival analysis and per-protocol analysis. The results presented above for DHA-PPQ, AR-LF, and AS-MQ are failure rate estimates obtained using survival analysis. The same models were analysed to obtain failure rate estimates calculated using per-protocol method (**Figure 3.13**, **Figure 3.14**, **Figure 3.15**). Comparison of these results showed that the per protocol method generates slightly higher estimated failure rates than survival analysis. The differences were dependant on the FOI level and duration of follow-up – the more reinfections that become patent over the course of follow-up (as occurred with higher FOI and longer follow-up), the greater this difference. With a 63-day follow-up and an FOI of 16 the failure rate estimate obtained for DHA-PPQ with the per-protocol method was nearly 30%, compared to the estimate with survival analysis of 15%. The reason is a "denominator effect". The per-protocol analysis simply removes all patients identified with reinfections from the analysis. Take the example where 20 of 200 patients are drug failures, giving a true underlying failure rate of 20/(20+180)=10%. If, for example, 50 of the 180 cured patients had reinfections and were removed from the analysis then the estimated per-protocol failure rate would rise to 20/(20+130)= 13% and if 100 of the cured patients had reinfections then failure rate would further increase to 20/(20+80)=20%. This example is somewhat artificial because reinfections will also occur in the recrudescence group and if they occur first, a later recrudescence could be masked, but it does serve to illustrate this denominator effect. It is important to appreciate that use of the per-protocol method with the newly proposed ≥ 2/3 markers algorithm (which generally produced more accurate failure rate estimates with appropriate follow-up length) will result in an over-estimate of failure rate. A detailed discussion of statistical analysis of malaria drug trials can be found elsewhere [131] but I emphasise that reporting the failure rate estimate obtained through survival analysis is essential for researchers wishing to utilize the ≥ 2/3 markers algorithm or the allelic family switch algorithm.

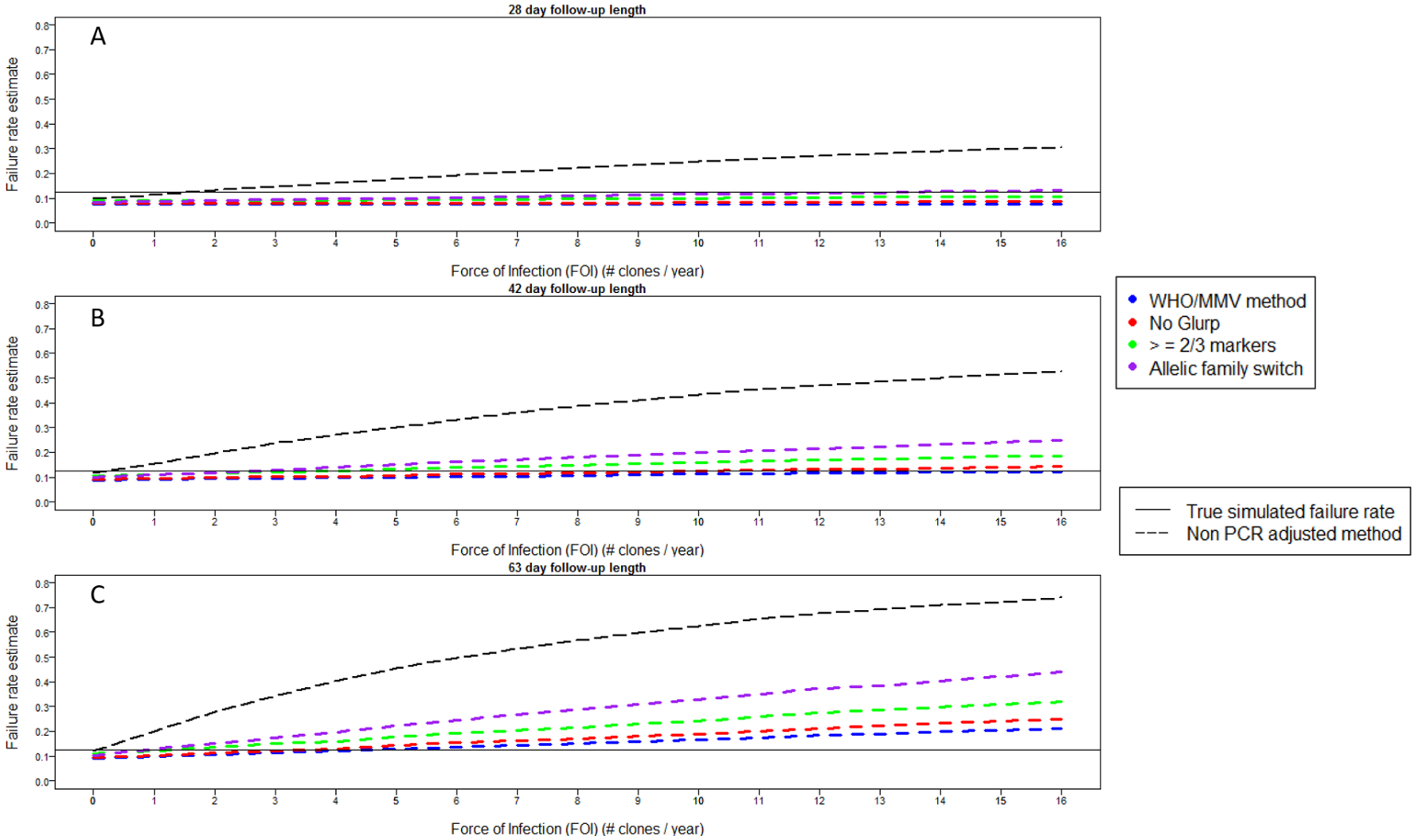
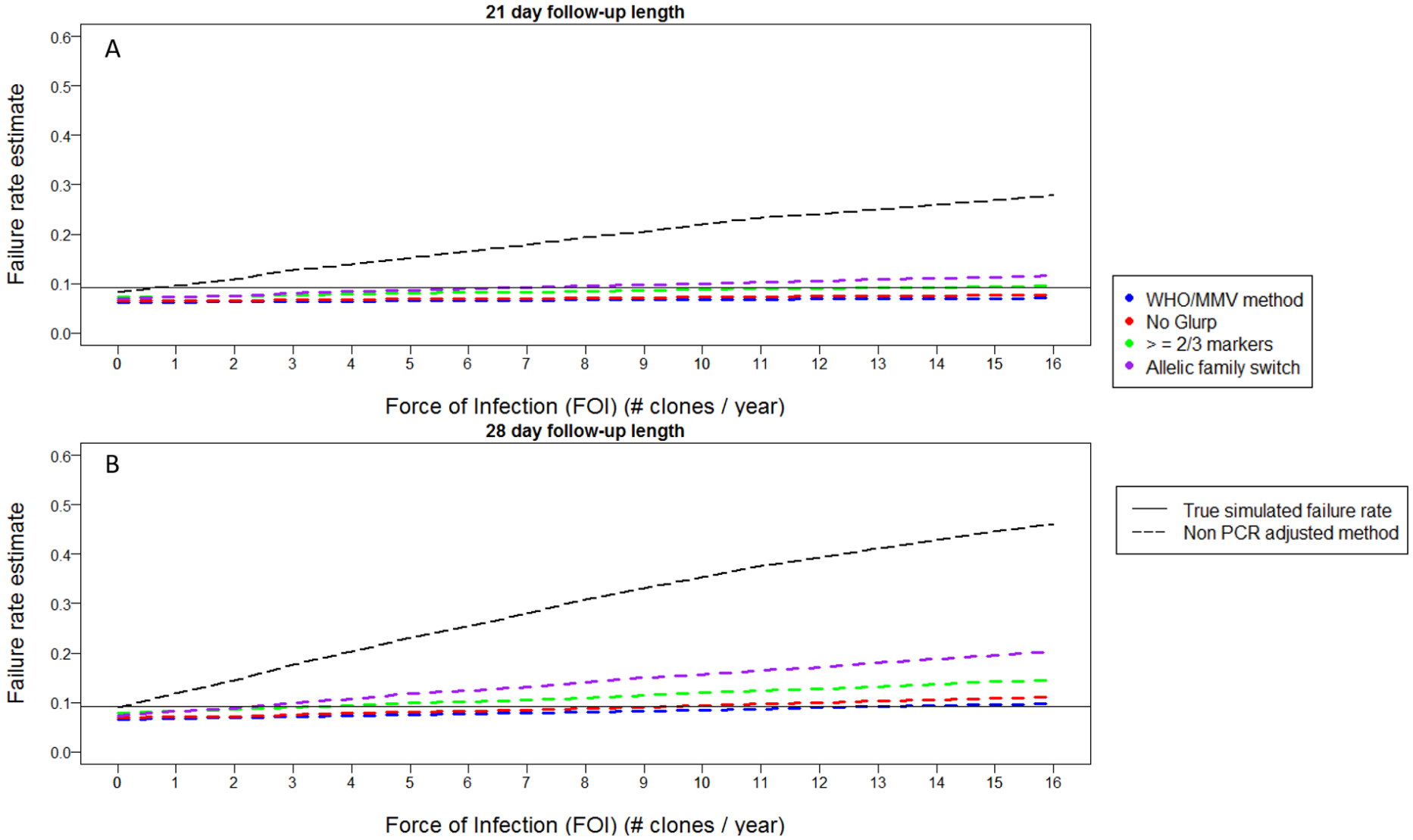



**Figure 3.13**: Analysis of simulated trial data for DHA-PPQ with follow-up lengths of 28 days (A), 42 days (B) and 63 days (C). Estimated failure rates are shown for the different algorithms of molecular correction as a function of FOI and calculated using the per protocol method.




1562

1563 **Figure 3.14**: Analysis of simulated trial data for failing AR-LF with follow-up lengths of 21 days (A) and 28 days (B). Estimated failure rates are shown for the
1564 different algorithms of molecular correction as a function of FOI and calculated using the per protocol method.

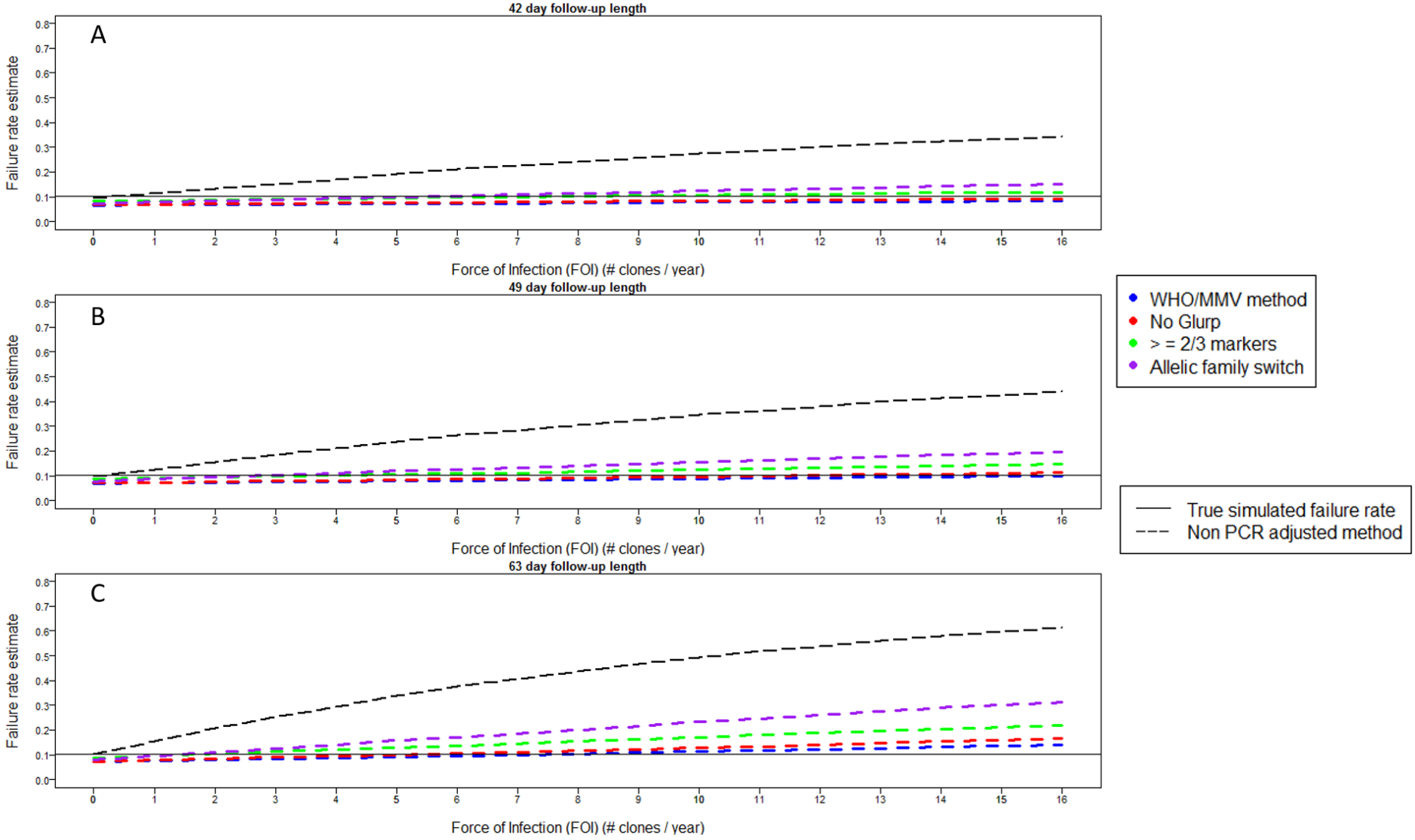



**Figure 3.15**: Analysis of simulated trial data for failing AS-MQ with follow-up lengths of 42 days (A), 49 days (B) and 63 days (C). Estimated failure rates are shown for the different algorithms of molecular correction as a function of FOI and calculated using the per protocol method.

### 3.3.8 Sensitivity analysis of model parameters.

Sensitivity analysis was conducted on model parameters: MOI, relative detectability of alleles and the minority allele detection threshold. Given the consistency of results for DHA-PPQ, AR-LF and AS-MQ, sensitivity analysis was conducted modelling DHA-PPQ under the assumption of a two-compartment model for PPQ.

The results presented above all assumed MOI at time of treatment is representative of high transmission i.e. using Tanzanian data. This was assumed because high MOI makes detection of recrudescent alleles more difficult (due to the issues with detection of minority alleles) so represents a "worst case" scenario. Readers will notice that results presented earlier in this chapter used this high MOI assumption even at low transmission intensities (quantified by FOI). This may not be a true reflection of epidemiology in vivo (lower MOI would generally be assumed in a lower transmission area), but a high MOI was used for the following reasons:

Keeping the same MOI across all transmission intensities allowed a direct comparison of molecular correction algorithms.

This assumption of high MOI at treatment is conservative (i.e. "worst case" scenarios) for low transmission areas because the results show that there is little operational difference between the algorithms even if initial MOI is high; it is therefore a robust conclusion that algorithm choice is not important in these areas because if MOI at treatment is lower, then there will be even less difference between the algorithms (this is later illustrated by Cambodian field data in **Table 3.3**).

High MOI at time of treatment can occur even in low transmission areas if people immigrate from areas of higher transmission or have acquired sufficient protective immunity that several clones may co-circulate asymptomatically before the patient falls ill. More plausibly, this scenario may arise in areas of seasonally intense transmission where MOI at time of treatment is high, but trials are conducted during the low-transmission season to reduce the impact of reinfections.

I investigated the impact of a reduced MOI to ensure that model findings would be consistent in different MOI settings. Analysis of simulated data for DHA-PPQ with a 42-day follow-up and a low MOI setting (the distribution obtained from PNG; see **3.3.2**) is shown in **Figure 3.16**. First note that the true failure rate was slightly lower than that obtained in a high MOI setting (**Figure 3.4** ) because patients harboured fewer clones at time of treatment which made their infection easier to clear. Reducing the MOI to reflect a low-transmission setting reduced the difference between algorithms. Overall, the results were consistent with those obtained from a high MOI setting i.e. the allelic family switch algorithm produced an accurate failure rate estimate at an FOI of 4 and below, and the ≥ 2/3 markers algorithm produced the most accurate failure rate estimate at all higher FOI.

The relative detectability of the longest allele to the shortest allele was altered from 0.001:1 to 0.1:1. The results are shown in **Figure 3.17**. Failure rate estimates obtained using this altered relative detectability are nearly identical to those obtained with the relative detectability of 0.001:1 used elsewhere in this chapter (i.e. **Figure 3.4** ).

The threshold at which minority genotyping signals are discounted as "noise" and disregarded was varied from 0.25 to 0.05 (the value used in the results described above is 0.25, as described in **3.2.5**). Analysis of simulated data for DHA-PPQ with a 42-day follow-up under these conditions is shown in **Figure 3.18**. The failure rate estimate produced by each algorithm increased as the threshold decreased. At the lower threshold of 0.05 the no glurp algorithm (rather than the ≥ 2/3 markers algorithm) produced the most accurate failure rate estimate from an FOI of 6 and higher. A minority detection threshold of 0.05 is unrealistic because large amounts of experimental/laboratory noise would be included in the signal, so this threshold could not be used in practice. The threshold was changed to 0.2 (a more realistic value) in **Figure 3.19**. Under this assumption the ≥ 2/3 markers

algorithm produced the most accurate failure rate estimate, robust across all FOI levels, the same as when the minority detection threshold is set to 0.25.

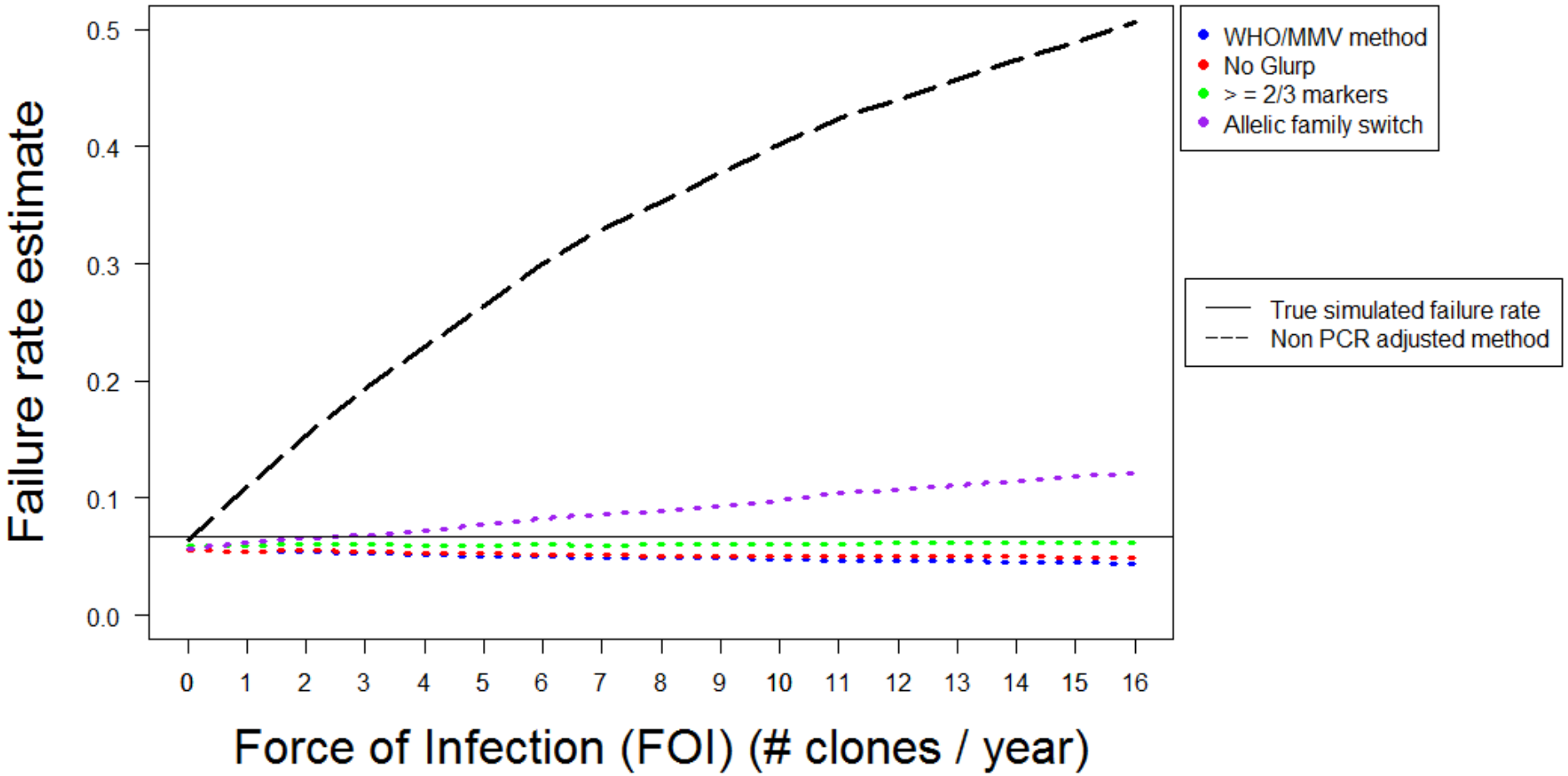



**Figure 3.16**: Analysis of simulated trial data for DHA-PPQ with a follow-up period of 42 days in a low MOI setting. Estimated failure rates are shown for the different algorithms of molecular correction as a function of FOI and calculated using survival analysis.

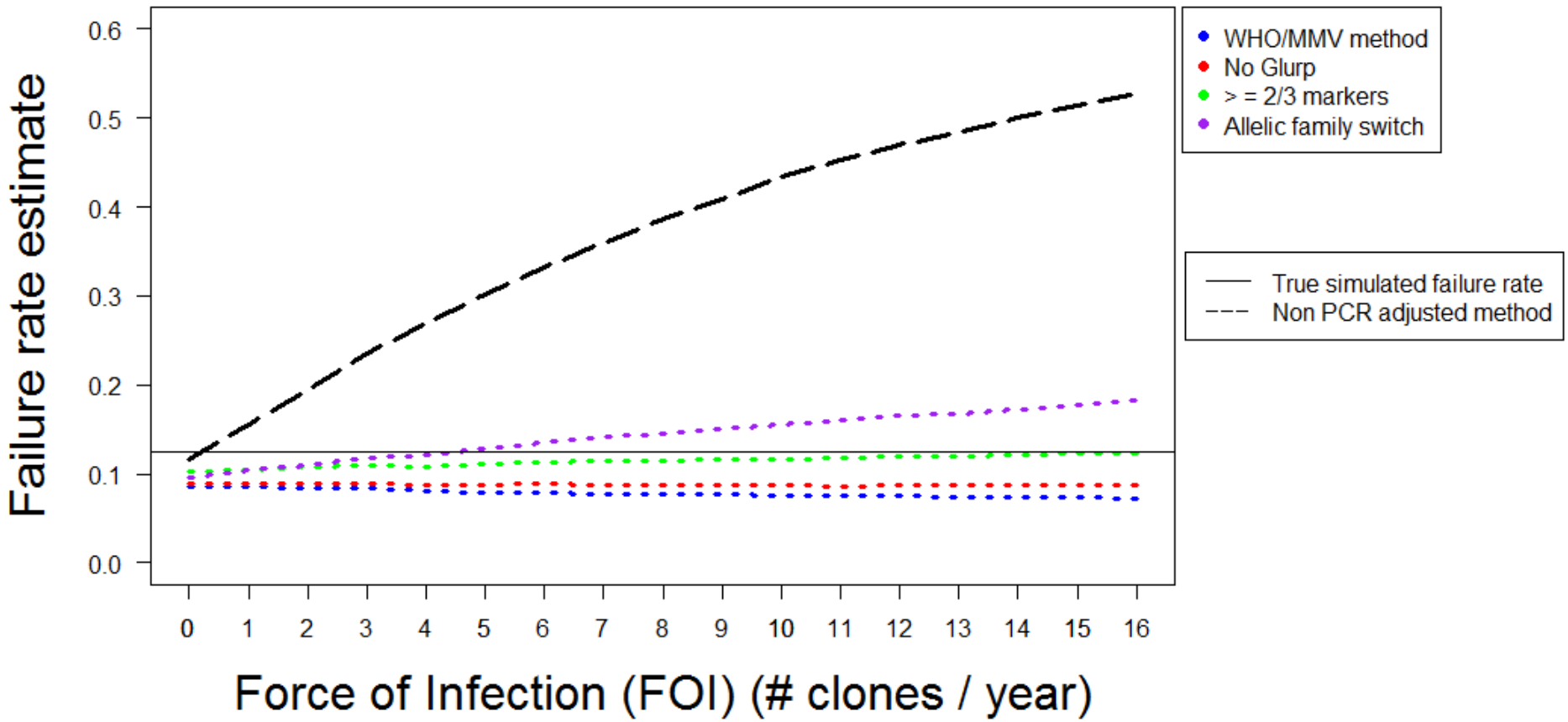



**Figure 3.17**: Analysis of simulated trial data for DHA-PPQ with a follow-up period of 42 days with the relative detectability of the longest allele to the shortest allele set to be 0.1:1. Estimated failure rates are shown for the different algorithms of molecular correction as a function of FOI and calculated using survival analysis.

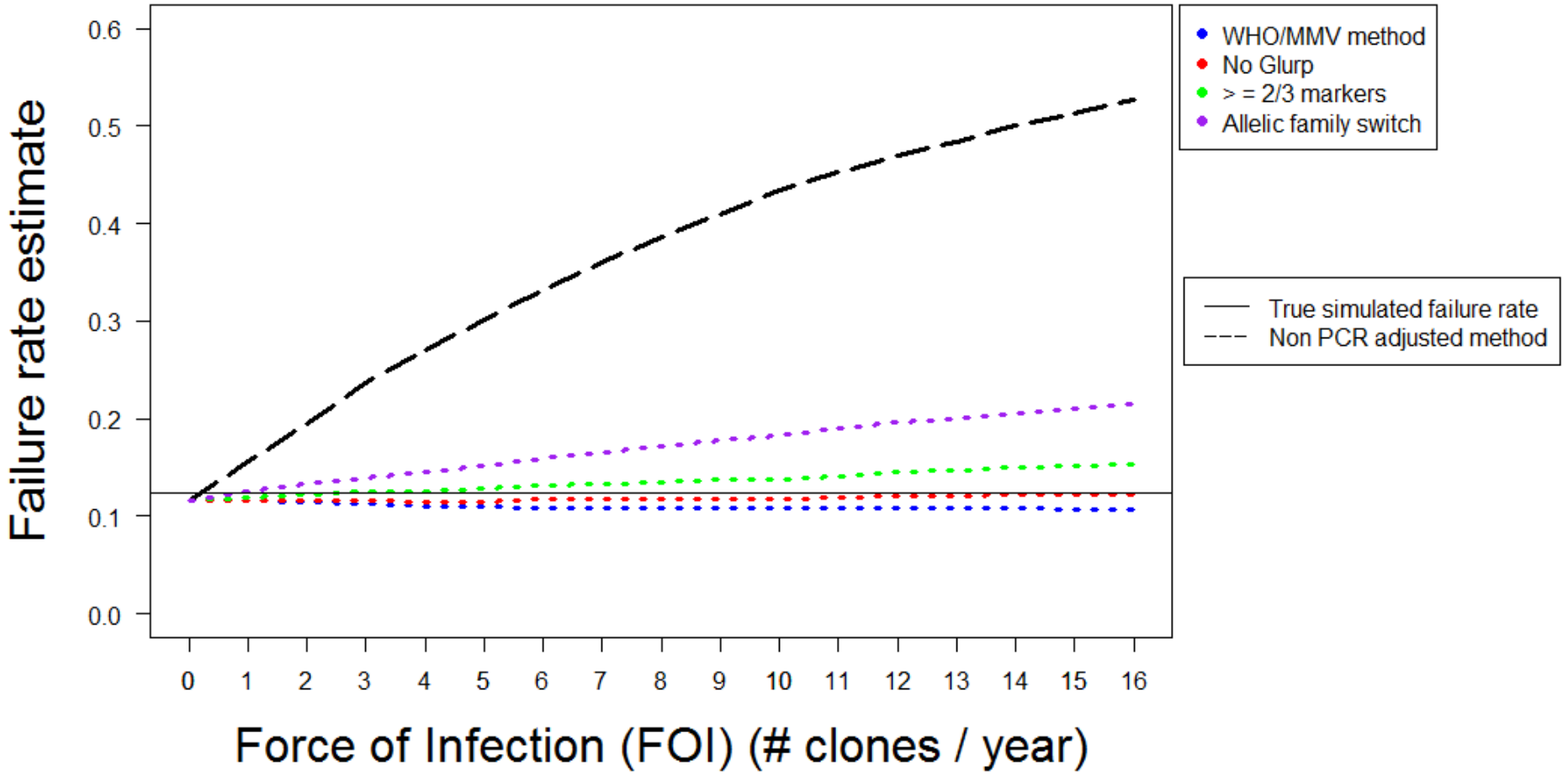



**Figure 3.18**: Analysis of simulated trial data for DHA-PPQ with a follow-up period of 42 days and a minority allele detection threshold of 0.05. Estimated failure rates are shown for the different algorithms of molecular correction as a function of FOI and calculated using survival analysis.

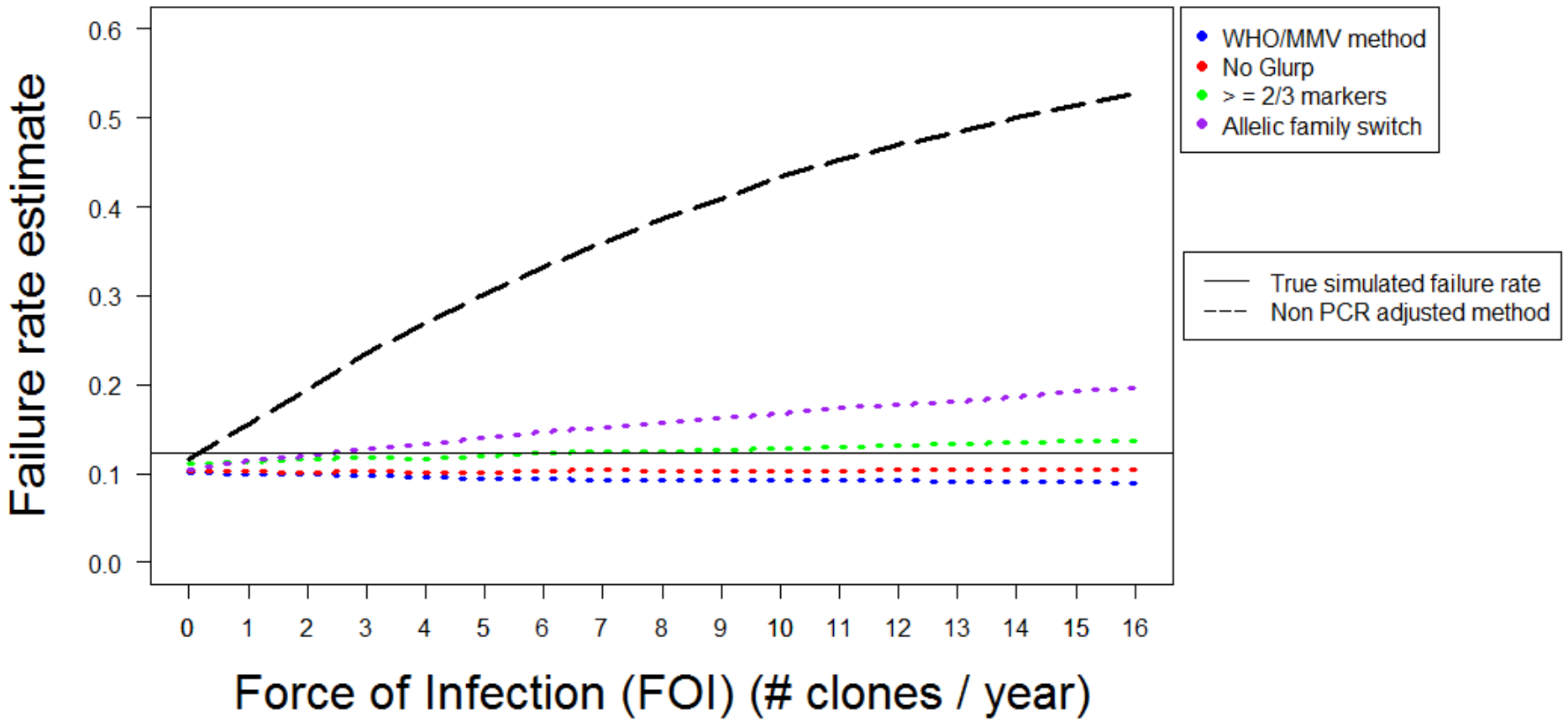



**Figure 3.19**: Analysis of simulated trial data for DHA-PPQ with a follow-up period of 42 days and a minority allele detection threshold of 0.2. Estimated failure rates are shown for the different algorithms of molecular correction as a function of FOI and calculated using survival analysis.

### 3.3.9 Re-analysis of clinical data .

Clinical data from Rwanda (a relatively high transmission area) were re-analysed using the proposed molecular correction algorithms (**Table 3.3**), and were highly consistent with simulated results i.e. the WHO/MMV algorithm produced the lowest estimated failure rate, followed by no glurp, then the ≥ 2/3 markers algorithm, then the allelic family switch algorithm. The pattern was quantitively consistent: The WHO/MMV algorithm estimated failure rates to be around half that obtained by the ≥ 2/3 markers algorithm. Results are similarly consistent with re-analysis of a trial from low transmission settings in Cambodia (**Table 3.3**). The impact of algorithm choice was not so large in Cambodia because FOI was low: 62 of the recurrences had matching alleles at all 3 loci so were presumably drug failures and would have been classified as such by all four algorithms. There were only 3 potential reinfections (all following DHA-PPQ treatment): 1 had no shared alleles at any locus so was classified as a reinfection under all four algorithms, but the other two patients shared alleles at both msp-1 and msp-2 and were only classified as reinfections under the WHO/MMV algorithm because no common alleles were noted at glurp. In contrast, the other algorithms all classified both patients as being drug failures. In summary, as in the high transmission data, the WHO/MMV algorithm had a higher tendency to classify recurrences as reinfections compared to the other algorithms. Note also that, consistent with **Figure 3.4** , the choice of algorithm makes little operational difference at low FOI: using the WHO/MMV algorithm identified 62 drug failures and three reinfections, while the other algorithms give 64 drug failures and one reinfections, a negligible increase in number of drug failures.

**Table 3.3**. Molecular correction with multiple algorithms from re-analysis of clinical trial data from Rwanda (a high transmission area) and Cambodia (a low transmission area)

| Country | Drug* | Classification of recurrent infection | WHO/MMV | No glurp | ≥ 2/3 markers | Allelic family switch |
|---|---|---|---|---|---|---|
| Rwanda | AR-LF | recrudescence | 17 | 27 | 36 | 59 |
| | | reinfections | 93 | 83 | 73 | 51 |
| | DHA-PPQ | recrudescence | 3 | 6 | 8 | 18 |
| | | reinfections | 40 | 37 | 35 | 25 |
| Cambodia | AS-AQ | recrudescence | 5 | 5 | 5 | 5 |
| | | reinfections | 0 | 0 | 0 | 0 |
| | DHA-PPQ | recrudescence | 45 | 47 | 47 | 47 |
| | | reinfections | 3 | 1 | 1 | 1 |
| | AS-PYN | recrudescence | 12 | 12 | 12 | 12 |
| | | reinfections | 0 | 0 | 0 | 0 |

*AR-LF: Artemether-Lumefantrine, DHA-PPQ: Di-hydroartemisinin-Piperaquine, AS-AQ: Artesunate-Amodiaquine, AS-PYN: Artesunate-Pyronaridine. Full details of study sites and methodology are provided in **3.2.8**.*

Finally, a review of clinical trials that reported failure rates based on no correction and the WHO/MMV algorithm is provided in **Table 3.4**. The magnitude of differences in failure rate estimates were similar to those noted in the results from simulations where the non-corrected algorithm and the WHO/MMV algorithm produced the highest and lowest failure rate estimates respectively.

**Table 3.4**: The need for molecular correction: a comparison of estimated drug failure rates obtained without correction vs with molecular correction performed according to the current WHO/MMV recommended algorithm.

| Drug tested | Uncorrected vs corrected failure rates | Country/yr | Ref |
|---|---|---|---|
| AR-LF<br>AS-AQ | 54% vs 10%<br>42% vs 10% | Burkina Faso, 2014 | [132] |
| AS-AQ | 17% vs 6% | Congo, 2013 | [133] |
| AR-LF | 22% vs 0% | Tanzania, 2014 | [134] |
| AR-LF | 13% vs 0% | Benin, 2016 | [135] |
| AR-LF | 9% vs 2% | Mozambique, 2015 | [136] |
| AR-LF | 2% vs 1% | India 2015 | [137] |
| AR-LF<br>AS-AQ | 16% vs 1%<br>22% vs 5% | Congo 2012 | [138] |

*AR-LF: Artemether-Lumefantrine, AS-AQ: Artesunate-Amodiaquine. Failure rate was calculated as 1 minus the 28-day adequate clinical and parasitological response reported in the studies (data collated and provided by Drs Jörg Möhrle and Stephan Duparc).*

## 3.4 Discussion

The key message presented here is that none of the proposed algorithms using msp-1, msp-2 and glurp correctly classified all recurrent infections (**Figure 3.2**; **Figure 3.3**) nor is it likely that such an algorithm exists due to the limitations of the PCR correction process [106]. The ability of each algorithm to accurately recover the true failure rate was dependent on the transmission intensity (quantified in these models by FOI) due to the differing propensity of each algorithm to misclassify reinfections as recrudescence (which occurred when alleles are shared by chance or a clone that later recrudesces was not observed in the initial sample). The 2-fold under-estimation of true failure rates that occurred at all FOI levels using the current consensus methodology of the WHO/MMV algorithm is a cause for considerable concern. This under-estimate occurred because this algorithm was extremely stringent – it did not misclassify any reinfections as recrudescence (**Figure 3.2**; **Figure 3.3**) – but did misclassify some recrudescences as reinfections when a clone that later recrudesced wasn't detected in the initial sample (due to the issues inherent in the PCR methodology with detecting minority alleles and longer alleles). These issues are shared between algorithms; however, no glurp, ≥2/3 markers and the allelic family switch algorithm are all less stringent and misclassified some reinfections as recrudescence, which increased failure rate estimates and accounted – to some extent – for the under-estimation of failure rates. The ≥ 2/3 markers algorithm correctly classifies the largest proportion of recrudescence; Requiring a match at only 2/3 markers allows for tolerance when minority alleles at a single loci are not detectable in the molecular correction process.

Key to identifying a methodology that gives consistently accurate estimated failure rates is to minimise and balance errors that arise from molecular correction, which are, in turn, influenced by factors including FOI, duration of follow-up, and sensitivity of the PCR protocols. Despite these concerns, these results show that operationally-important increases in accuracy of estimated failure rates for anti-malarial efficacy trials are achievable with alternate genotyping algorithms. It is undesirable to recommend different molecular correction algorithms for different ACTs and transmission intensity levels (as this would be likely to cause confusion), hence the approach of investigating multiple ACTs and varying transmission intensity through FOI to assess if a single algorithm may be identified that gives robust and accurate estimates. Based on the results presented here, it appeared that the ≥ 2/3 markers algorithm was the most robust in areas of moderate to high transmission, and provided estimated failure rates close (typically within 2 percentage units) to the true failure rate (**Figure 3.4 ; Figure 3.7 Figure 3.9**).

The other factor that can affect estimates of drug efficacy, given that molecular correction is imperfect, is the duration of follow-up. Recommended duration has gradually increased over the last 20 to 30 years, with the objective of capturing all (or at least the majority) of recrudescences. However, the objective of clinical trials is not to capture every recrudescence, but to obtain accurate and robust estimates of efficacy. **Figure 3.5** shows, for DHA-PPQ, that in areas of moderate to high FOI, the penalty for detecting the last few recrudescences by extending the follow-up period was the inclusion of a much larger number of reinfections. These reinfections inflate the estimated failure rate due to the propensity of molecular correction algorithms to misclassify some reinfections as recrudescence (as seen in **Figure 3.2**). It is obviously preferable to have the shortest follow-up possible while retaining accuracy of failure rate estimates; based on the results shown in **Figure 3.5**, and analogous plots for failing AR-LF (**Figure 3.8** )and AS-MQ (**Figure 3.10** ), using the ≥ 2/3 markers algorithm provided accurate estimates using a follow-up of 28 days for AR-LF, 42 days for DHA-PPQ and 49 days for AS-MQ, all roughly in line with current WHO recommendations [14, 28]. Importantly, the accuracy of the estimates with this algorithm appeared to be relatively robust to changes in transmission intensity, quantified in these models by FOI (the WHO/MMV and no glurp algorithms were also robust to changes in FOI, but had an under-estimate of failure rate associated with them). Note that a different DHA-PPQ parameterization (assuming a three compartment model for PPQ**; Figure 3.6** ) favoured a longer follow-up period more in line with MQ which also has longer prophylaxis post-treatment. The trends across all drugs modelled are clear: it is highly likely that use of the current WHO/MMV algorithm will generate substantial (near two-fold) underestimates of failure rates and that switching to an alternative correction algorithm should be considered as matter of urgency.

Technical problems with molecular correction approaches exist (identified and explained in, for example, [106, 121]) which gives rise to the temptation to simply ignore molecular correction and just use uncorrected data. The results presented here strongly suggest that appropriate use of molecular correction is essential. Trials conducted in areas of moderate to high transmission intensity, which are the areas where most malaria morbidity and mortality occur, analysed without molecular correction will lead to severe over-estimates of the true failure rate. This assertion is supported by clinical data (**Table 3.4**), which clearly shows that large discrepancies may arise in the absence of molecular correction. Ignoring molecular correction (i.e. non-PCR corrected algorithm in **Figure 3.6 Figure 3.7** and **Figure 3.9**) only produced accurate estimates of failure rates when FOI is very low (a fact generally acknowledged in the literature [14, 28]). However, caution must be taken even when using no-correction in low transmission areas. Malaria transmission is highly focal and even if an area is, on average, very low transmission, it is plausible that most patients will be recruited from foci of high transmission where FOI may well be sufficient to invalidate estimates based on no-correction.

Confidence in this modelling approach was assured given the past success of pharmacological modelling to correctly reflect and predict clinical data e.g. [15, 87, 104, 105, 139, 140], and the consistency of the simulated results with in vivo Rwandan and Cambodian data-sets (**Table 3.3**). This model may not reflect the in vivo PK parameters of these trials (see discussion of mPK/PD modelling methods in **2.1**), however, the purpose of re-analysis of these data were to investigate the change in failure rates from us of proposed algorithms on in vivo data – analysis of trial results with these algorithms has not previously taken place. This re-analysis is not dependent on model parameter space (nor vice versa), and all algorithms require the same data (the msp-1, msp-2 and glurp alleles (and families for the former two)); consequently, this re-analysis showing similar trends to modelled results is encouraging.

There is concern in the literature that reinfections may share alleles with the initial infection purely by chance and that subsequent misclassification of reinfections as recrudescence would lead to over-estimation of failure rates [121]. This could arise in areas of high transmission [120] as increased MOI leads to more alleles in the initial sample; these can later be shared with a reinfection purely by chance. It could also occur in low-transmission areas where genetic diversity is lower and there is more chance of a match by chance. Importantly, large-scale over-estimation is not observed (e.g. the low

impact of FOI on estimated failure rate using the ≥ 2/3 markers algorithm in **Figure 3.6 Figure 3.7** and **Figure 3.9**) with increased transmission intensity with either a high MOI (**Figure 3.6**) or a low MOI **(Figure 3.16)**, suggesting these fears are unlikely to have a large impact in practice.

In conclusion, this modelling approach of the length-polymorphic markers *msp-1*, *msp-2* and *glurp*, and re-analysis of clinical data both suggest that more accurate and easily implemented algorithms are available to analyse clinical data and the field should consider implementing these methods. Which algorithm will perform best will depend on factors in the patient population/area – the results here demonstrate this explicitly for transmission intensity (FOI) and follow-up length. The four algorithms investigated here are not mutually exclusive and are based on the same data. A firm recommendation is that initial and recurrent samples should be genotyped at all three loci: when using the current WHO/MMV algorithm, there is no need to genotype after a mismatch has occurred at one locus (**see 3.2.8**), so genotyping is often incomplete. These complete data would allow results obtained from all four algorithms be presented; this maintains consistency with previous analyses based on the WHO/MMV algorithm while also providing results that are likely to provide a substantially more robust estimate of malaria drug clinical failure rates.

# Chapter 4: Evaluating accuracy of microsatellite markers for classification of recurrent infections during routine monitoring of anti-malarial drug efficacy: A computer modelling approach.

A version of this work has been submitted for publication (26/07/2019): Evaluating accuracy of microsatellite markers for classification of recurrent infections during routine monitoring of anti-malarial drug efficacy: A computer modelling approach. Authors: Sam Jones, Mateusz Plucinski, Katherine Kay, Eva Maria Hodel, Ian Hastings.

Chapter-specific acknowledgements: Dr Katherine Kay, Dr Eva Maria Hodel and Dr Ian Hastings provided R code to generate parasite dynamics post-treatment (fully described in **Chapter 2** of this thesis). Dr Mateusz Plucinski provided extensive advice on genotyping microsatellite markers, provided access to R code to run a Bayesian analysis algorithm (**4.2.7**) and provided code to generate plots for distributions of posterior probability of recrudescence, receiver operator characteristic curves and contour plots. Additionally, Dr Plucinski provided access to sets of microsatellite allele frequency data. Dr Simon Wagstaff and Mr Andrew Bennett provided access to the high-performance computing required to analyse simulated data-sets using a Bayesian algorithm.

## 4.1 Introduction.

### 4.1.1 Microsatellite markers for genotyping Therapeutic Efficacy Studies (TES).

In the "molecular correction" process to genotype the results of therapeutic efficacy studies (TES) and distinguish between recrudescence and reinfections, microsatellite markers are an alternative method to the length-polymorphic markers msp-1, msp-2, and glurp – the WHO consensus methodology [14] that was explored in **Chapter 3** of this thesis. Microsatellites are (generally small) segments of repeated genetic motifs; they have been extensively explored for this purpose [40, 141, 142], with one key proposed advantage being the lack of immune selection on microsatellite loci [39] – note, however, that this is a historic reason for the development of microsatellite markers for TES, and whether or not markers are under immune selection should not have any bearing on the simulated results in this thesis for the purposes of calculating efficacy estimates as markers being under immune selection, whilst having consequences for population genetic structure [39] should not affect allelic frequency distributions in the 4-6 week period of a TES

In this methodology, researchers genotype microsatellite loci in both initial and recurrent infections and count the number of matching loci in each patient i.e. the number of loci at which at least a single allele is shared between the initial and recurrent infection. They then define a certain number of matches to be indicative of recrudescence. In addition to their use in TES, microsatellites have also been commonly used to assess treatment failure in returning travellers in non-endemic areas [143-145].

Typically, microsatellite data are analysed by applying a mathematically simple match counting algorithm which uses an arbitrary threshold for the number of loci that have common alleles between the initial and day of failure samples. In these algorithms, if the two samples have matching alleles at, or above, the threshold number of loci, they are classified as recrudescence, and otherwise, reinfections. Typically, classification of an infection as a recrudescence requires a match at most, if not all, sampled loci [40, 98, 146], though the specific microsatellites used, the total number of microsatellites genotyped, and the number of matches at which a recrudescence is classified is variable in the literature. For example, Hwang et al. [98] used 8 loci and defined a match at 7 or more loci to be a recrudescence. Greenhouse et al. [40] investigated 6 loci, and subsequently used 4 to analyse samples, with a match at every locus being required to classify a recurrence as a

recrudescence. Mwangi et al. [146] used 5 loci and considered a match at 5 to be a recrudescence, 0 to be a reinfection, and intermediary values to be mixed infections.

### 4.1.2 Sources of error with microsatellite markers.

As with length-polymorphic markers, there are three potential mechanisms by which misclassification of a recurrent infection can occur using microsatellite markers:

a) Recrudescent infections can be misclassified as reinfection if alleles of the recrudescent clone(s) were not detected when genotyping the initial infection.

b) Recrudescent infections can be misclassified as reinfection if the recurrent infection is mixed (i.e., the recurrence is polyclonal and comprised of both recrudescence and reinfections), and alleles of the recrudescent clone(s) are not detected, but alleles of a reinfecting clone(s) are.

c) A reinfection could be misclassified as recrudescent if it shares (by chance) alleles with clones present at time of treatment.

The types of misclassification a) and b) may occur due to imperfect detectability of microsatellite alleles. Alleles at each microsatellite loci are defined purely by their length in base pairs – they are not characterized by families (as the length-polymorphic loci msp-1 and msp-2 are) and so are comparable to the length-polymorphic marker glurp in this respect. The process of characterizing microsatellite alleles in a blood sample is broadly similar to that undertaken for length-polymorphic markers: The sample is induced into PCR, amplified, and examined through electropherograms produced after electrophoresis (i.e., Figure 1 of [39]). Low frequency peaks are discounted in the electropherogram as they are considered to be noise induced during the PCR process with the consequence that truly present minority alleles may be ignored [40]. The peak height (relative to the highest peak) at which minority peaks are ignored will vary between laboratories, software and operators but is generally considered to be between 20 and 30%.

Additionally, it is possible to “mis-read” the length of microsatellite alleles. That is to say, if a given allele was 100 base pairs in length, that allele might instead be observed as 98, 99, 101, or 102 base pairs (depending on the potential range of mis-reading). This has obvious consequences as a researcher may falsely observe shared alleles between the initial and recurrent sample or fail to observe shared alleles that are truly there. In other words, mis-reading allele length could cause reinfections to be misclassified as recrudescence or vice-versa. Amplifying microsatellites in PCR has the potential to induce mutations in the genetic material (i.e., errors in reading base-pair length) through so-called “replication slippage” of *Taq* DNA polymerase, [147]. Evidence shows that the choice of sequencing method (or machine) will induce variation in the allele length observed [148]. Finally, external factors such as laboratory temperature have been shown to induce variation in observed allelic length [149].

As with length polymorphic markers, it is theoretically possible for a third type of misclassification to occur, i.e:

c) A reinfection can be misclassified as a recrudescence if sufficient alleles are shared, by chance, between a reinfecting clone and clones of the initial infection. The probability of this occurring will depend on the MOI, transmission intensity, level of genetic diversity, and when using the match counting method of classifying recrudescence described above, is also dependant on the number of loci at which an allele is shared between the initial and recurrent sample that is required to classify a recrudescence.

### 4.1.3 Bayesian methodology for microsatellite markers.

A recent publication [150] presented a statistical method based on Bayesian probability to analyse microsatellite data to calculate drug failure rates. This method generates the posterior probability that a recurrent infection is a recrudescence and has subsequently been used to analyse TES data [30, 31, 151]. The problems inherent in accurately genotyping a blood sample using microsatellite markers listed above mean that a simple method of counting matching microsatellites between samples may never be able to reliably classify a patient as reinfection or recrudescence. Bayesian analyses constitute a better, more flexible approach capable of dealing with these uncertainties. The full advantages of a Bayesian approach, and a detailed description of the Bayesian algorithm later used to analyse simulated data-sets in this thesis chapter are explained in exhaustive detail elsewhere [150].

### 4.1.4 Research goals.

Using an mPK/PD approach (**Chapter 2**) to simulate antimalarial therapy provides a gold-standard definition of true response to treatment, something un-obtainable from field trial data, including the true status (reinfection or recrudescence) of all recurrent infections. Simulations of 10,000 patients were conducted for two ACTs: AR-LF and AS-MQ. Using this simulated data, this chapter had three main objectives:

Evaluate the accuracy of failure rate estimates generated using microsatellite data in conjunction with a match counting algorithm (as is currently typical), and determine whether the stringent requirement of a match between the initial and recurrent sample at all or most loci to classify a recrudescence is able to recover accurate failure rate estimates.

Assess the advantages of a previously published Bayesian analysis methodology [150], both in its ability to recover the true failure rate and the diagnostic ability to distinguish recrudescence from reinfections.

Determine whether the methodologies based on microsatellite loci are robust across drugs with different post-treatment prophylactic profiles (i.e., partner drugs with varying half-lives) which determine when reinfections start to occur, across different transmission intensities (which determine rates of reinfection in TES) and in regions with differing levels of genetic diversity at microsatellite loci.

## 4.2 Methodology.

### 4.2.1 Trial scenarios: Partner drug IC50, Multiplicity of Infection (MOI) and Force of Infection (FOI).

10,000 patients treated with AR-LF and AS-MQ were simulated under a selection of "scenarios" with pre-defined parameters multiplicity of Infection (MOI; **2.3.1**), force of infection (FOI; **2.3.2** ), level of genetic diversity and partner drug half maximal inhibitory concentrations (IC50; **2.2.1**) to represent in vivo scenarios using microsatellite markers for molecular correction. This was possible due to the wealth of data (MOI and allele frequency distributions) provided to me by Mateusz Plucinski of the Centers for Disease Control and Prevention (CDC). This approach differs to that undertaken in **Chapter 3** for the length-polymorphic markers, where there were two available MOI settings, FOI was varied arbitrarily and only a single set of allele frequency distributions were available. The goal of modelling set scenarios was to increase realism (i.e., a low transmission area is generally more likely to have lower MOI, FOI and reduced genetic diversity compared to a medium or high transmission area) to make the model as robust as possible for investigating the accuracy of the Bayesian approach (see research goal 2 in **4.1.4**).

Initially, 6 scenarios were modelled for two different drugs (AR-LF and AS-MQ), for a total of 12 different scenarios.

Non-failing drug in a low transmission setting

Non-failing drug in a medium transmission setting

Non-failing drug in a high transmission setting

Failing drug in a low transmission setting

Failing drug in a medium transmission setting

Failing drug in a high transmission setting.

"Transmission setting" was defined by the FOI, the MOI, and the frequency distribution of microsatellite alleles. Whether the drug is "non-failing" or "failing" was determined by modifying the half maximal inhibitory concentration (IC50) parameter of the drug in the PK/PD model.

In this chapter, parasite dynamics post-treatment were simulated with AR-LF and AS-MQ (see **Chapter 2** for mPK/PD methodology and PK parameterization) in order to model a drug with a relatively short post-treatment prophylactic period (AR-LF) and a drug with a relatively long post-treatment prophylactic period (AS-MQ). DHA-PPQ was not modelled here due to computational limits imposed by running the Bayesian algorithm, which are fully described later in this chapter (**4.2.9**), due to the length of follow-up needed (42 days) and the number of recurrences that would occur (DHA-PPQ is not as prophylactic as AS-MQ and so more reinfections will occur; AS-MQ simulations already take a long computation time, (see section **4.2.9**). This is without considering the debate in the field about whether DHA-PPQ should is represented by a 2 or 3-compartment model. To simulate both calibrations, as parameterized in **Chapter 2**, would take an extremely long time. The goal of this chapter was to investigate the accuracy of failure rate estimates using microsatellite markers – while simulating both a drug with short post-treatment prophylaxis and long post-treatment prophylaxis was important to ensure the validity of the model, the goal was not to simulate every available Artemisinin-based combination therapy (ACT; **2.1.1**) and so in the context of this thesis, it was prudent to simulate AR-LF and AS-MQ in this chapter.

Three MOI distributions were modelled: Low, medium and high MOI. These distribution were data-sets obtained in vivo from Angola [31, 151] and are represented graphically in **Figure 4.1** . All low transmission scenarios used an FOI of 2, all medium transmission scenarios used an FOI of 8 and all high transmission scenarios used an FOI of 16. Partner drug IC50s were altered such that true failure rates were ~2% for non-failing drugs and ~10% for failing drugs (the true failure rate is directly observable from the mPK/PD model and is described mathematically in **3.2.7**. A ~2% failure rate in "non-failing" drugs was allowed because even very good drugs occasional fail due to factors such as sub-optimal patient adherence (e.g. vomiting up one of the doses). True failure rate increases as multiplicity of infection (MOI) increases, which meant altering the IC50 between scenarios for failing drugs. For non-failing drugs, it wasn't necessary to alter the IC50 as MOI changed to keep true failure rates close to 2%. The IC50 values for the partner drugs used for each scenario are shown **Table 4.1**.

Three additional scenarios were modelled for AR-LF only. These scenarios utilized an "extremely low genetic diversity" distribution of alleles in order to investigate the impact on failure rate estimates of extremely low genetic diversity, which could occur naturally in low transmission areas or even in high transmission areas due to effects such as a genetic bottleneck. Microsatellite allelic distributions were not available in vivo for an area of extremely low diversity so these had to be created arbitrarily by me (**4.2.2.2**). The three scenarios were then low, medium and high transmission with the MOI distribution, FOI values and IC50 values for LF described above.

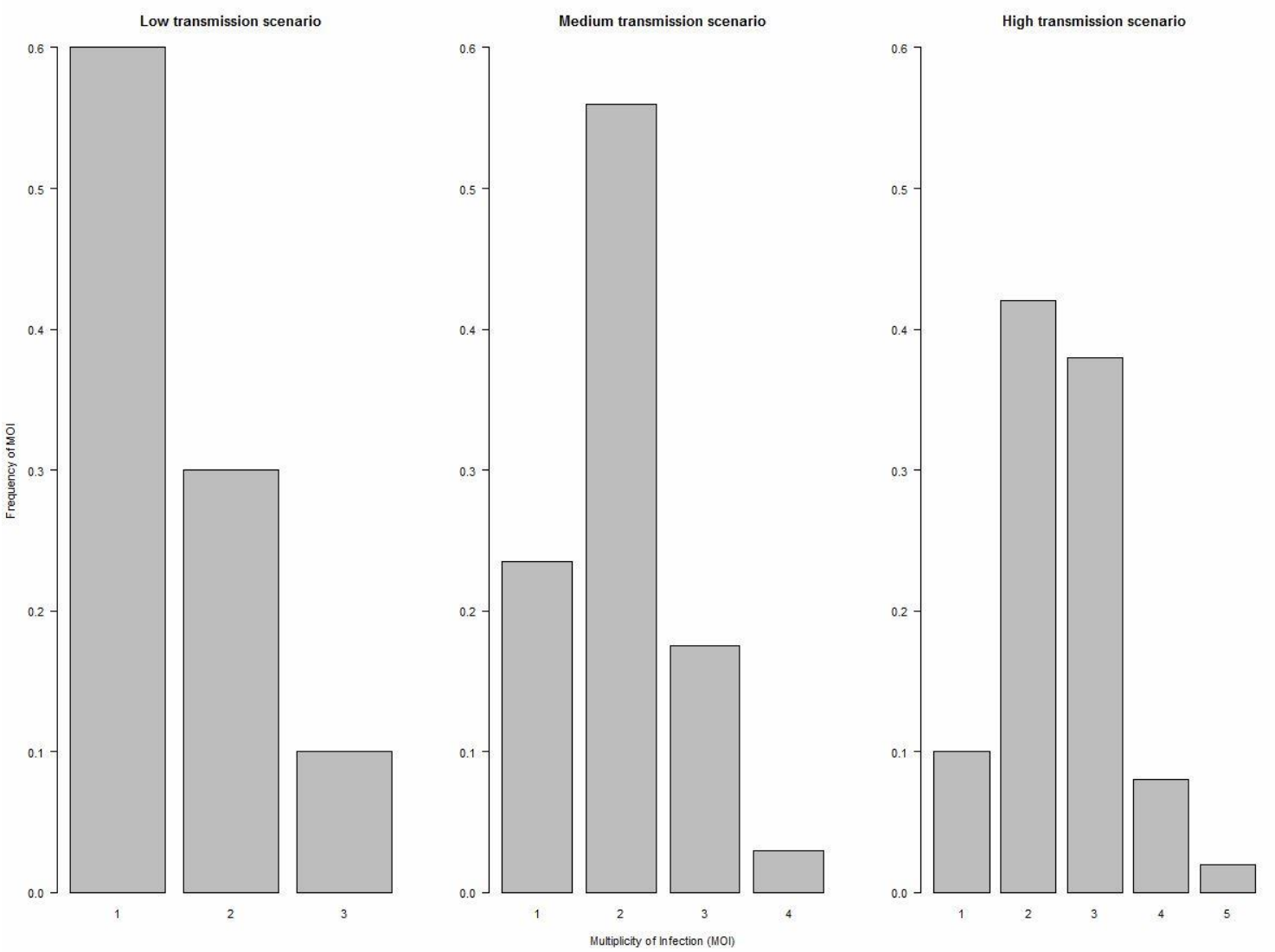



**Figure 4.1** Multiplicity of infection (MOI) distributions for three different transmission intensity scenarios, based on data from Angola [31, 151]. The MOI for each patient in the simulation was sampled from one of these distributions, depending on which scenario was being simulated.

**Table 4.1** Drug concentration at which half maximal inhibitory concentration (IC50) occurs for each of six model scenarios in these simulations for Artemether-Lumefantrine (AR-LF) and Artesunate-Mefloquine (AS-MQ). Corresponding true failure rates for each modelled scenario are provided.

| | | Artemether-Lumefantrine (AR-LF) | | Artesunate-Mefloquine (AS-MQ) | |
|---|---|---|---|---|---|
| Transmission | Drug Failure | IC50 (μg/L) | True Failure | IC50 (μg/L) | True Failure |
| Low | Non-fail | 500 | 0.0149 | 27 | 0.006 |
| Mid | | 500 | 0.0167 | 27 | 0.0097 |
| High | | 500 | 0.0204 | 27 | 0.0119 |
| Low | Fail | 220000 | 0.0965 | 450 | 0.0978 |
| Mid | | 20000 | 0.0997 | 400 | 0.106 |
| High | | 10000 | 0.103 | 370 | 0.1058 |

*AR: Artemether, LF: Lumefantrine, AS: Artesunate, MQ: Mefloquine, IC50: Half-maximal inhibitory concentration.*

### 4.2.2 Genetic Data – allele frequency distributions for seven microsatellite loci.

#### 4.2.2.1 Low, medium and high genetic diversity allelic frequency distributions.

Genotypes were assigned to every clone (both initial and reinfections) at seven microsatellite loci: 313, 383, TA1, polya, PfPK2, 2490 and TA109; alleles at each loci were defined by their length (base pairs). The sources of data were studies in Angola, described in [31, 151]. Relative frequencies and length of alleles used to parameterize these simulations are shown in Table 4.2 for a low genetic diversity setting, Table 4.3 for a medium genetic diversity setting and Table 4.4 for a high genetic diversity setting.

#### 4.2.2.2 Extremely low genetic diversity allelic frequency distributions.

I simulated an area of extremely low genetic diversity (for AR-LF). In the absence of in vivo distribution for such an area, I took the number of unique alleles for each locus in the low transmission setting and divided that number by 4 (rounded up). This resulted in:

- 313: 4 alleles
- 383: 4 alleles
- TA1: 3 alleles
- POLYA: 4 alleles
- PFPK2: 5 alleles
- 2490: 2 alleles
- TA109: 4 alleles

It was then assumed these remaining alleles were those with the highest frequency in the low transmission setting. They were then assigned equal frequency, such that (for example) 313 has 4 alleles, each with 25% frequency. The relative frequency of alleles and their length for this extremely low genetic diversity setting is shown in Table 4.5.

This distribution is obviously arbitrary but is likely to be conservative (single or two extremely frequent alleles would reduce diversity even further) and is useful to illustrate the point that low genetic diversity leads to higher failure rate estimates as the rate of false positives from reinfections being misclassified as recrudescence increases and examine to what extent this will threaten the accuracy of failure rate estimates.

**Table 4.2** Frequency distributions for alleles at seven microsatellite loci for a low transmission setting.

| Frequency distributions for low transmission setting | | | | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| MS Number | 313 | MS Number | 383 | MS Number | TA1 | MS Number | POLYA | MS Number | PFPK2 | MS Number | 2490 | MS Number | TA109 |
| 13 | | 16 | | 11 | | 13 | | 19 | | 5 | | 13 | |
| Length | Frequency | Length | Frequency | Length | Frequency | Length | Frequency | Length | Frequency | Length | Frequency | Length | Frequency |
| 220 | 0.111111 | 124 | 0.243243 | 178 | 0.2 | 155 | 0.181818 | 162 | 0.137255 | 82 | 0.571429 | 163 | 0.230769 |
| 222 | 0.111111 | 140 | 0.216216 | 163 | 0.171429 | 152 | 0.151515 | 168 | 0.117647 | 79 | 0.25 | 160 | 0.192308 |
| 244 | 0.111111 | 104 | 0.081081 | 166 | 0.114286 | 164 | 0.121212 | 171 | 0.117647 | 73 | 0.107143 | 148 | 0.134615 |
| 262 | 0.111111 | 86 | 0.054054 | 169 | 0.114286 | 167 | 0.121212 | 177 | 0.098039 | 85 | 0.035714 | 175 | 0.134615 |
| 224 | 0.074074 | 122 | 0.054054 | 175 | 0.114286 | 158 | 0.090909 | 174 | 0.078431 | 88 | 0.035714 | 172 | 0.096154 |
| 226 | 0.074074 | 144 | 0.054054 | 172 | 0.085714 | 143 | 0.060606 | 165 | 0.058824 | | | 151 | 0.057692 |
| 232 | 0.074074 | 146 | 0.054054 | 73 | 0.057143 | 170 | 0.060606 | 180 | 0.058824 | | | 169 | 0.038462 |
| 238 | 0.074074 | 88 | 0.027027 | 181 | 0.057143 | 173 | 0.060606 | 150 | 0.039216 | | | 166 | 0.019231 |
| 240 | 0.074074 | 136 | 0.027027 | 160 | 0.028571 | 104 | 0.030303 | 156 | 0.039216 | | | 178 | 0.019231 |
| 250 | 0.074074 | 138 | 0.027027 | 184 | 0.028571 | 113 | 0.030303 | 159 | 0.039216 | | | 181 | 0.019231 |
| 218 | 0.037037 | 148 | 0.027027 | 202 | 0.028571 | 161 | 0.030303 | 186 | 0.039216 | | | 184 | 0.019231 |
| 230 | 0.037037 | 150 | 0.027027 | | | 176 | 0.030303 | 195 | 0.039216 | | | 187 | 0.019231 |
| 246 | 0.037037 | 152 | 0.027027 | | | 179 | 0.030303 | 138 | 0.019608 | | | 196 | 0.019231 |
| | | 162 | 0.027027 | | | | | 141 | 0.019608 | | | | |
| | | 164 | 0.027027 | | | | | 183 | 0.019608 | | | | |
| | | 170 | 0.027027 | | | | | 189 | 0.019608 | | | | |
| | | | | | | | | 192 | 0.019608 | | | | |
| | | | | | | | | 198 | 0.019608 | | | | |
| | | | | | | | | 201 | 0.019608 | | | | |

**Table 4.3** Frequency distributions for alleles at seven microsatellite loci for a medium genetic diversity setting.

| Frequency distributions for medium transmission setting | | | | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| MS Number | 313 | MS Number | 383 | MS Number | TA1 | MS Number | POLYA | MS Number | PFPK2 | MS Number | 2490 | MS Number | TA109 |
| 17 | | 18 | | 10 | | 13 | | 13 | | 3 | | 15 | |
| Length | Frequency | Length | Frequency | Length | Frequency | Length | Frequency | Length | Frequency | Length | Frequency | Length | Frequency |
| 219 | 0.117647 | 124 | 0.225 | 160 | 0.228571 | 155 | 0.205128 | 165 | 0.2 | 82 | 0.666667 | 163 | 0.220339 |
| 245 | 0.117647 | 140 | 0.125 | 166 | 0.171429 | 164 | 0.179487 | 162 | 0.133333 | 73 | 0.166667 | 160 | 0.152542 |
| 259 | 0.117647 | 86 | 0.075 | 169 | 0.114286 | 158 | 0.102564 | 168 | 0.133333 | 79 | 0.166667 | 175 | 0.152542 |
| 217 | 0.088235 | 138 | 0.075 | 172 | 0.114286 | 149 | 0.076923 | 159 | 0.088889 | | | 172 | 0.101695 |
| 225 | 0.088235 | 144 | 0.075 | 175 | 0.114286 | 152 | 0.076923 | 171 | 0.088889 | | | 148 | 0.067797 |
| 233 | 0.088235 | 102 | 0.05 | 178 | 0.114286 | 170 | 0.076923 | 177 | 0.088889 | | | 151 | 0.067797 |
| 235 | 0.058824 | 122 | 0.05 | 181 | 0.057143 | 176 | 0.076923 | 174 | 0.066667 | | | 166 | 0.067797 |
| 237 | 0.058824 | 136 | 0.05 | 73 | 0.028571 | 161 | 0.051282 | 156 | 0.044444 | | | 154 | 0.033898 |
| 211 | 0.029412 | 146 | 0.05 | 163 | 0.028571 | 167 | 0.051282 | 180 | 0.044444 | | | 181 | 0.033898 |
| 221 | 0.029412 | 88 | 0.025 | 187 | 0.028571 | 137 | 0.025641 | 186 | 0.044444 | | | 169 | 0.016949 |
| 231 | 0.029412 | 100 | 0.025 | | | 143 | 0.025641 | 183 | 0.022222 | | | 178 | 0.016949 |
| 239 | 0.029412 | 108 | 0.025 | | | 182 | 0.025641 | 189 | 0.022222 | | | 202 | 0.016949 |
| 241 | 0.029412 | 128 | 0.025 | | | 188 | 0.025641 | 192 | 0.022222 | | | 205 | 0.016949 |
| 247 | 0.029412 | 150 | 0.025 | | | | | | | | | 208 | 0.016949 |
| 249 | 0.029412 | 152 | 0.025 | | | | | | | | | 211 | 0.016949 |
| 257 | 0.029412 | 154 | 0.025 | | | | | | | | | | |
| 275 | 0.029412 | 156 | 0.025 | | | | | | | | | | |
| | | 170 | 0.025 | | | | | | | | | | |

**Table 4.4** Frequency distributions for alleles at seven microsatellite loci for a high genetic diversity setting.

| Frequency distributions for high transmission setting | | | | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| MS | 313 | MS | 383 | MS | TA1 | MS | POLYA | MS | PFPK2 | MS | 2490 | MS | TA109 |
| Number | 18 | Number | 23 | Number | 14 | Number | 15 | Number | 16 | Number | 4 | Number | 16 |
| Length | Frequency | Length | Frequency | Length | Frequency | Length | Frequency | Length | Frequency | Length | Frequency | Length | Frequency |
| 217 | 0.145455 | 138 | 0.142857 | 160 | 0.180328 | 155 | 0.196721 | 162 | 0.2375 | 82 | 0.627451 | 172 | 0.19 |
| 225 | 0.127273 | 140 | 0.142857 | 166 | 0.131148 | 152 | 0.163934 | 168 | 0.1625 | 79 | 0.313725 | 160 | 0.18 |
| 233 | 0.127273 | 124 | 0.12987 | 172 | 0.114754 | 167 | 0.114754 | 171 | 0.125 | 85 | 0.039216 | 163 | 0.15 |
| 221 | 0.109091 | 102 | 0.090909 | 178 | 0.114754 | 149 | 0.081967 | 159 | 0.1 | 73 | 0.019608 | 175 | 0.12 |
| 229 | 0.072727 | 136 | 0.090909 | 163 | 0.098361 | 179 | 0.081967 | 165 | 0.0875 | | | 151 | 0.07 |
| 243 | 0.072727 | 84 | 0.051948 | 169 | 0.098361 | 173 | 0.065574 | 174 | 0.05 | | | 184 | 0.07 |
| 239 | 0.054545 | 122 | 0.038961 | 184 | 0.065574 | 143 | 0.04918 | 177 | 0.0375 | | | 148 | 0.06 |
| 241 | 0.054545 | 142 | 0.038961 | 139 | 0.04918 | 158 | 0.04918 | 180 | 0.0375 | | | 166 | 0.04 |
| 223 | 0.036364 | 86 | 0.025974 | 175 | 0.04918 | 161 | 0.04918 | 183 | 0.0375 | | | 178 | 0.04 |
| 231 | 0.036364 | 98 | 0.025974 | 181 | 0.032787 | 140 | 0.032787 | 153 | 0.025 | | | 187 | 0.02 |
| 251 | 0.036364 | 100 | 0.025974 | 136 | 0.016393 | 164 | 0.032787 | 186 | 0.025 | | | 145 | 0.01 |
| 211 | 0.018182 | 130 | 0.025974 | 142 | 0.016393 | 170 | 0.032787 | 189 | 0.025 | | | 154 | 0.01 |
| 215 | 0.018182 | 164 | 0.025974 | 187 | 0.016393 | 119 | 0.016393 | 135 | 0.0125 | | | 157 | 0.01 |
| 235 | 0.018182 | 172 | 0.025974 | 193 | 0.016393 | 137 | 0.016393 | 138 | 0.0125 | | | 169 | 0.01 |
| 237 | 0.018182 | 88 | 0.012987 | | | 182 | 0.016393 | 144 | 0.0125 | | | 181 | 0.01 |
| 245 | 0.018182 | 90 | 0.012987 | | | | | 147 | 0.0125 | | | 196 | 0.01 |
| 255 | 0.018182 | 104 | 0.012987 | | | | | | | | | | |
| 261 | 0.018182 | 106 | 0.012987 | | | | | | | | | | |
| | | 144 | 0.012987 | | | | | | | | | | |
| | | 148 | 0.012987 | | | | | | | | | | |
| | | 150 | 0.012987 | | | | | | | | | | |
| | | 152 | 0.012987 | | | | | | | | | | |
| | | 162 | 0.012987 | | | | | | | | | | |

1995 **Table 4.5** Frequency distributions for alleles at seven microsatellite loci for an extremely low genetic diversity setting.

| Frequency distributions for very low genetic diversity setting | | | | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| MS | 313 | MS | 383 | MS | TA1 | MS | POLYA | MS | PFPK2 | MS | 2490 | MS | TA109 |
| Number | 4 | Number | 4 | Number | 3 | Number | 4 | Number | 5 | Number | 2 | Number | 4 |
| Length | Frequency | Length | Frequency | Length | Frequency | Length | Frequency | Length | Frequency | Length | Frequency | Length | Frequency |
| 220 | 0.25 | 124 | 0.25 | 178 | 0.33 | 155 | 0.25 | 162 | 0.2 | 82 | 0.5 | 163 | 0.25 |
| 222 | 0.25 | 140 | 0.25 | 163 | 0.33 | 152 | 0.25 | 168 | 0.2 | 79 | 0.5 | 160 | 0.25 |
| 244 | 0.25 | 104 | 0.25 | 166 | 0.33 | 164 | 0.25 | 171 | 0.2 | | | 148 | 0.25 |
| 262 | 0.25 | 86 | 0.25 | | | 167 | 0.25 | 177 | 0.2 | | | 175 | 0.25 |
| | | | | | | | | 174 | 0.2 | | | | |

1996

### 4.2.3 Follow-up length and detection of recurrence.

In this chapter, a 28 day follow-up period was used for models of AR-LF and a 42 day follow-up period was used for models of AS-MQ, as permitted by WHO guidelines [14]. Unlike in **Chapter 3**, novel lengths of follow-up were not explored for microsatellite markers. A 28-day follow-up schedule required patients be examined on days 3, 7, 14, 21 and 28. A 42-day follow-up period used two additional days i.e. days 35 and 42. The parasitaemia of each clone in each patient was tracked and updated each day as described by the mPK/PD model and the PK parameters of the patient and the PD parameters of the clone (**Chapter 2**).

The model checked each day of scheduled follow-up to determine whether a patient had enough parasitaemia that a recurrence would be detectable by light microscopy (a recurrence) – parasitaemia was considered detectable if the total number in a patient was $\geq 10^8$ on that day. Note that variance in the limit of detection by light microscopy exists with respect to the skill of the microscopist [18]; it was assumed this limit was reflective of an "expert" microscopist (corresponding to roughly 20 parasites /µl of blood). In short, follow-up length and detection of recurrence proceeded in an identical manner to as described for length polymorphic markers in **Chapter 3**.

### 4.2.4 Calculating which alleles are observed.

The genotype of the initial malaria infection of each patient was taken on the day of treatment. This genotype signal is a composite of all the clone(s) present in the initial infection and is determined by the technical accuracy and sensitivity of genotyping.

On all days of follow-up except day 3, a recurrence was identified if the sum parasitaemia of all clones in a patient exceeded $10^8$ which was assumed to be the minimum parasitaemia at which detection by light microscopy was possible [152]. This corresponded to a parasite density of roughly 20 parasites/µl of blood. If total parasitaemia was less than $10^8$ then recurrent parasites would not be observed by microscopy (and thus, the patient would not be genotyped on that day). On day 3, if total parasitaemia exceeded $10^8$ but was <25% of the total parasitaemia on the initial sample, the patient continued in the trial; if parasites were present at >25% of initial parasitaemia, that patient was classed as an early treatment failure, per WHO procedure [28]. Note that for subsequent calculations and analysis, an early treatment failure is considered to be identical to a recrudescence. Calculations then occurred using a three-step process to replicate the technical limitations of acquiring a profile of microsatellite alleles from a blood sample

Firstly, a "sampling" limit was included: A finite volume of blood is available for genotyping. A parasite clone would not be detected if its density were so low that no parasites were included in the blood sample analysed. Thus, the density and volume of the processed blood sample defined the limit of detection. It was assumed this limit was $10^8$ (i.e., no clone present in less than $10^8$ parasites would be detected). This limit is identical to that used in **Chapter 3**, see **3.2.5** for calculations and justification.

Secondly, the "majority" allele for each microsatellite locus is the allele with the highest parasitaemia (if multiple clones share alleles at a locus, the allelic signal for that loci is the sum of parasitaemia of the clones. It was assumed that for an allele to be detected, the parasitaemia of that allele must be ≥25% of the parasitaemia of the majority allele.

Finally, the chance that the length of each allele may be mis-read due to genotyping errors was included (see **4.1.2**). The chance of an error of +/- length x was assumed to be described by the geometric distribution $0.8 * (0.2)^x$ ; this distribution has been described and validated in [150].

The output of these simulations was, for each patient, the microsatellite alleles (quantified by their length in base-pairs for each loci) at each of the seven loci, observed in the initial sample, and at any

recurrent infection in that patient. This is exactly the data recorded in standard TES using microsatellites (and is the input used for the Bayesian algorithm in vivo as in [31, 150]).

### 4.2.5 Classifying patients: True failures, high and low density recrudescence.

A key advantage of simulating patients with microsatellite markers using an mPK/PD model is the ability to interrogate the Bayesian algorithm; i.e., investigate diagnostic ability and determine in which circumstances it would misclassify recurrences.

The term "true failure" is consistent with the mathematical description given in **Chapter 3** (**3.2.7**) but is re-summarised here for ease of reading: It was determined whether each patient was a "true failure" based on parasitaemia: A patient was a true failure if, on the final day of follow-up (day 28 for AR-LF, day 42 for AS-MQ), they still harboured any parasites from any initial clone. The true failure rate is the frequency of these patients across the entire population. The model tracked patients over the full length of follow-up, thus this "true failure" classification captured patients who would in vivo, have been removed earlier in the TES with a recurrent infection classified as a reinfection (and whose recrudescent clones would not then be observed).

For this chapter, true failures were separated into 'high' and 'low' density recrudescence. A high density recrudescence was defined as occurring when three conditions were met: (i) the sum parasitaemia of all recrudescent clones on the day of recurrence is >25% the sum parasitaemia of all clones on the day of recurrence (i.e., if there is a mixed infection of new and recrudescent clones on the day of recurrence, recrudescent clones must be >25% of the total infection) and (ii), the sum parasitaemia of all recrudescent clones on the day of recurrence must have been >25% of the total parasitaemia of all clones in the initial sample (i.e., clones that later recrudesce constitute at least 25% of the initial infection) . (iii) the total number of recrudescing clones on the day of recurrence must be $\geq 10^8$ (to be consistent with the sampling limit defined above). If any one of these conditions is not met then the failure is defined as "low density". In this manner, the true classification of each recurrence as a reinfection, high density recrudescence or low density recrudescence was defined.

### 4.2.6 Estimating drug failure rates: Match counting algorithm.

A match counting algorithm compared the number of microsatellite loci that have at least a single allele shared between the initial and recurrent sample (termed a "matching" loci). Typically, use of microsatellite markers in vivo requires a high number of matching loci to classify an infection as recrudescent (either all loci, or permitting a single locus not to match, i.e.: [40, 98, 146]). Herein, with the 7 loci modelled, threshold number of matching loci required to classify a recrudescence was varied in order to determine the impact of this choice of threshold on failure rate estimates. This is a counting algorithm where a recurrent infection is defined as a recrudescence when the number of matching loci is greater than or equal to a specified threshold. Six threshold values were analysed for this method: 2, 3, 4, 5, 6 and 7 matching loci (e.g. if a recurrent infection had 3 matching loci with the initial infection, that recurrence was classified as a recrudescence with a threshold of 2 or 3 loci, but as a reinfection with the other thresholds. When all recurrent samples had been classified as a recrudescence or reinfection based on this algorithm, failure rates were estimated using survival analysis as described in **3.2.7**.

### 4.2.7 Estimating drug failure rates: Bayesian analysis method.

The Bayesian analysis method described in [150] was used to interpret simulated results and obtain posterior probabilities of recrudescence for each patient. R code was provided to me by Mateusz

Plucinski (the original author of the Bayesian algorithm) to run this Bayesian algorithm; computational expansions (**4.2.9**) were made by me to allow the algorithm to run for 10,000 patients across all the scenarios modelled.

The Bayesian algorithm is extremely complex and is fully described in [150]. In brief, the Bayesian algorithm uses a Markov chain Monte Carlo approach to sample from the posterior probability of recrudescence for each sample, with the ratio of likelihoods of a reinfection versus a recrudescence derived from the frequencies of the observed alleles. The algorithm jointly estimates several key parameters, such as the error rate, and accounts for missing data by sampling hidden alleles. The Bayesian algorithm was used to define a recurrence as being a recrudescence when posterior probability of recrudescence in that patient exceeded a value p, where p lies between 0 and 1. I investigated the impact on failure rate estimates of the value of p by varying it between 0.1 and 1 inclusive. The situation in which p = 0 indicated a recrudescence was not investigated, as that would be reflective of assuming every recurrence was a recrudescence (i.e., a non-molecular correction approach). When all recurrent samples had been classified as a recrudescence or reinfection based on this algorithm, failure rates were estimated using survival analysis as described in **3.2.7**.

### 4.2.8 Assessment of algorithm accuracy.

Both the match-counting algorithm and Bayesian analysis classified a recurrent infection as either reinfection or recrudescence depending on the choice of threshold (for the match counting algorithm) or posterior probability *p* (for the Bayesian analysis). These classifications were then used to generate failure rate estimates for the simulated TES using survival analysis. The failure estimates for both methods were then compared with the true failure rate to assess their accuracy.

The distribution of the posterior probability of recrudescence calculated using the Bayesian algorithm was plotted for each scenario, with recurrences stratified into their true status: low-density recrudescence, high-density recrudescence or reinfection. Receiver operator characteristic (ROC) curves were constructed using the posterior probability at which an infection would be classified as a recrudescence (from 0 to 1). The area under the ROC curve (AUC) was used to quantify the diagnostic ability of the method [153]. An AUC of >0.8 is considered to be a "good" test and an AUC of >0.9 is considered to be an "excellent" test.

The ability of the Bayesian algorithm to detect low-density recrudescence was evaluated by calculating the posterior probability of recrudescence estimated by the Bayesian algorithm for each recurrent infection (**4.2.7**) and categorizing each infection as reinfection, low-density recrudescence or high-density recrudescence (**4.2.5**).

### 4.2.9 Computational considerations.

All modelling and subsequent analysis was conducted using the statistical programming language R (version 3.5.1) [98]. The mPK/PD simulations, assigning genetic signals to initial and recurrent infections, obtaining failure rate estimates using the match counting method and all analysis (i.e., generating ROC curves, producing plots, etc) takes an inconsequential amount of computational time and should not be a concern to anyone looking to replicate this methodology.

The Bayesian analysis method is extremely computationally intense. The exact amount of time required to analyse a given trial depends on the number of patients in the trial, the number of recurrences, and the average number of microsatellites observed at both initial and recurrent samples. Naturally, this means that higher transmission scenarios took longer to analyse than lower transmission scenarios and failing drug scenarios took longer than non-failing drugs.

Simulations were run on an Ubuntu 18.04 Virtual Machine with 104 dedicated processor cores and 128GB of RAM, hosted on a 4 socket Ubuntu 18.04 Dell PowerEdge R940 server (768TB DDR4-2666 RAM, 4x Intel Xeon Gold 6154) running KVM. This machine was made available by the scientific computing department at the Liverpool School of Tropical Medicine (in particular, Simon Wagstaff and Andrew Bennett). I parallelized the simulations and analysis using the R packages doParallel and foreach to simulate 10,000 patients as 100 trials of 100 patients each. Run-time per scenarios varied between 1 day (low transmission scenario, non-failing AR-LF) and 6 days (high transmission scenario, failing AS-MQ). Most of this run-time is allocated to the Bayesian algorithm (generating the parasitaemia data and assigning microsatellites took around an hour for all scenarios). Consequently, running these simulations would not have been feasible without a multi-core computing and parallelization approach. For groups seeking to use the Bayesian method in vivo to analyse smaller trials a parallelization approach should still be taken to lower run-times, i.e., separating a trial of 300 patients into three 100 patient trials for analysis should be feasible on most current personal computers.

## 4.3 Results.

### 4.3.1 Analysis of AR-LF.

#### 4.3.1.1 Failure rate estimates and comparison to true failure rate.

The match counting algorithm was sensitive to transmission intensity; no threshold value of matching loci at which a recurrence was classified as recrudescent was able to accurately estimate true failure rate across all transmission scenarios for either failing (**Figure 4.2** ) or non-failing (**Figure 4.3** ) AR-LF. Failure rate estimates declined as the threshold increased. Failure rate estimates increased as transmission increased, presumably due to the greater number of reinfections, some of which were misclassified as recrudescence; this effect was greater at low thresholds when the probability of such misclassification was greater. A threshold of 4 produced estimates close to the true failure rate for all non-failing AR-LF scenarios. For failing AR-LF scenarios, a threshold of 3 produced the closest estimate to true failure in the low transmission scenario, and a threshold of 4 produced the closest estimate in the high transmission scenario, with the medium transmission scenario intermediate between the two. However, using a threshold of 3 in a high transmission scenario over-estimated failure rate (estimated failure rate of 0.18 compared to a true failure rate of 0.1). A threshold of 4 gave an estimate of 0.08 relative to a 0.0997 true failure rate for the failing, medium transmission scenario and an estimate of 0.077 relative to a true failure rate of 0.0965 for the failing, low transmission scenario. A threshold of 7 (requiring all 7 loci to be matching) resulted in extremely large under-estimates of failure rates for failing AR-LF: 0.005 relative to true failure rate of 0.0965 in the low transmission scenario, 0.008 relative to true failure of 0.0997 in the medium transmission scenario and 0.006 relative to true failure rate of 0.1 in the high transmission scenario.

In contrast to the match-counting method, the Bayesian algorithm recovered true failure rate to a high degree of accuracy across all transmission settings and for both calibrations of true drug failure rate (**Figure 4.2** and **Figure 4.3** ). Values of the posterior probability $p$ used to distinguish recrudescence from reinfection between 0.1 and 0.9 produced good, consistent failure rates estimates with only a slight decline as $p$ increased; using $p = 1$ to classify a recrudescence resulted in a substantial decrease in failure rate estimates from using $p \geq 0.9$. For all non-failing and failing drug scenarios, treating all infections with $p \geq 0.1$ as recrudescence generated a failure rate estimate within 0.01 (1%) of the true failure rate.

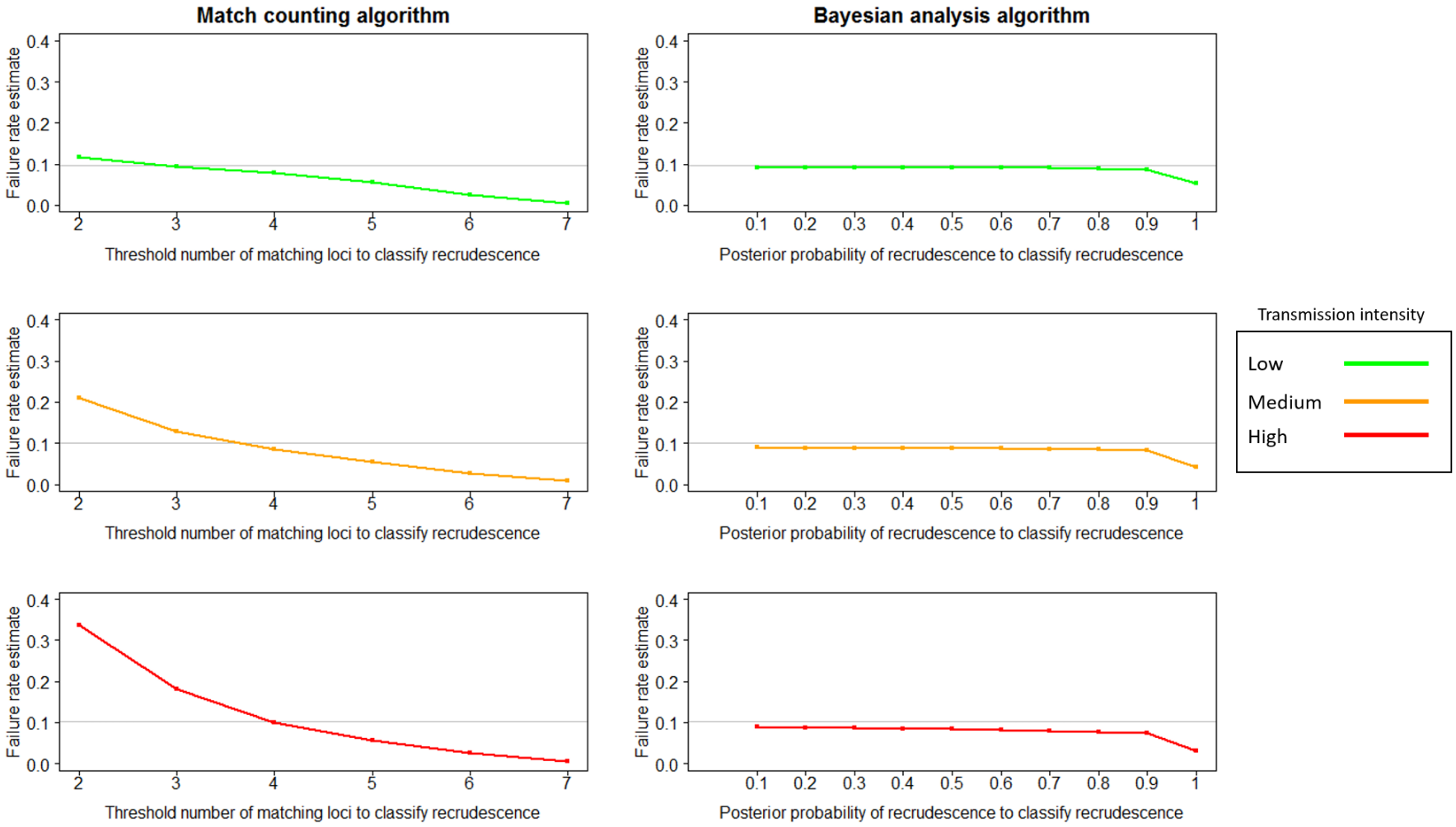



**Figure 4.2** Failure rate estimates obtained using the match counting algorithm and the Bayesian analysis algorithm for failing AR-LF under low, medium and high transmission scenarios. The true failure rate is denoted in each plot by the horizontal grey line. For the match counting algorithm, the threshold for the number of matching loci at which a recurrence is classified as a recrudescence varies between 2 and 7. For the Bayesian analysis, the cut-off for posterior probability at which a recurrence is classified as a recrudescence varies between ≥0.1 and ≥0.9.

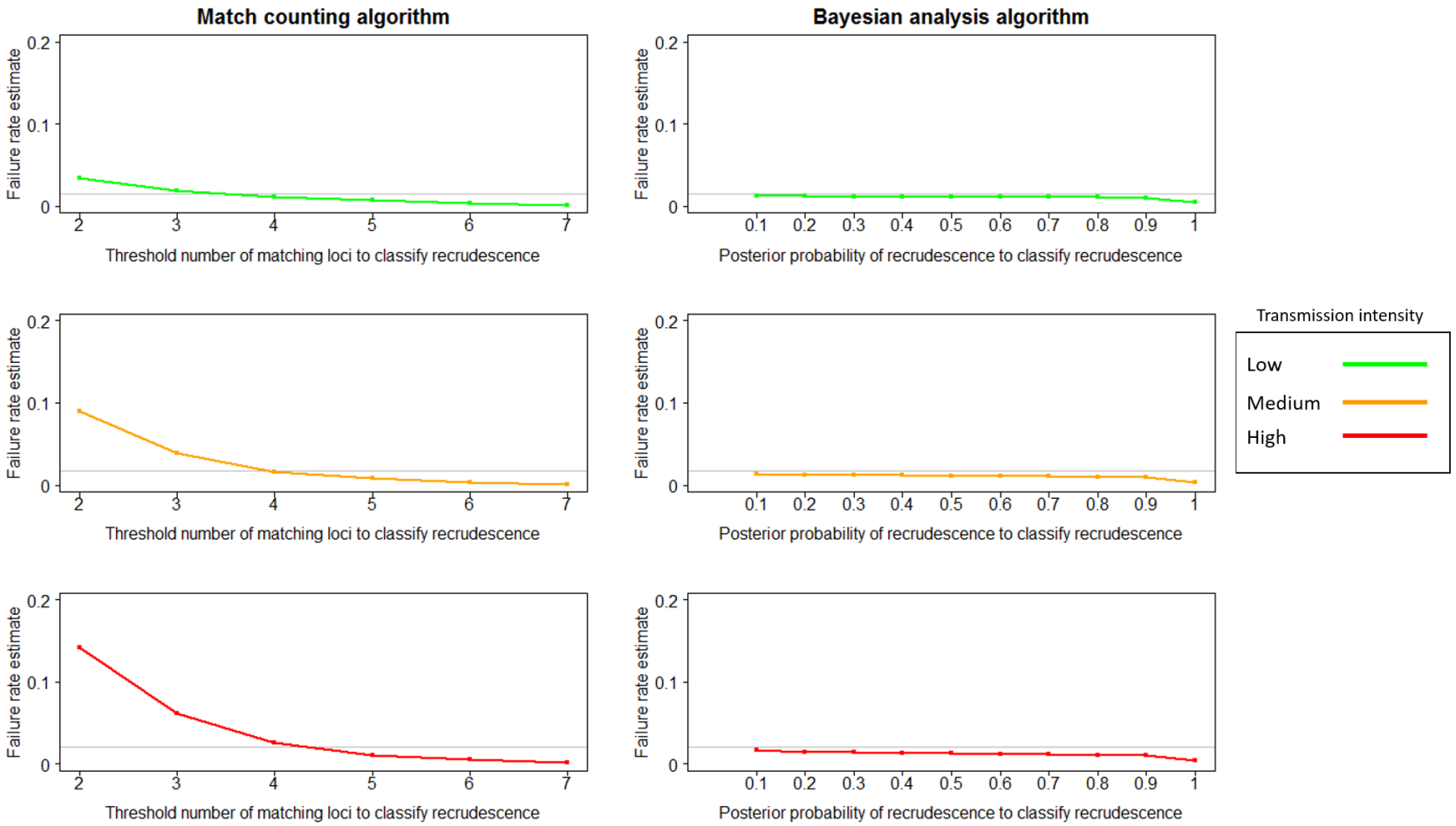



**Figure 4.3** Failure rate estimates obtained using the match counting algorithm and the Bayesian analysis algorithm for non-failing AR-LF under low, medium and high transmission scenarios. The true failure rate is denoted in each plot by the horizontal grey line. For the match counting algorithm, the threshold for the number of matches at which a recurrence is classified as a recrudescence varies between 2 and 7. For the Bayesian analysis, the cut-off for posterior probability at which a recurrence is classified as a recrudescence varies between ≥0.1 and ≥0.9.

#### 4.3.1.2 Receiver Operator Characteristic (ROC) curves for the Bayesian algorithm.

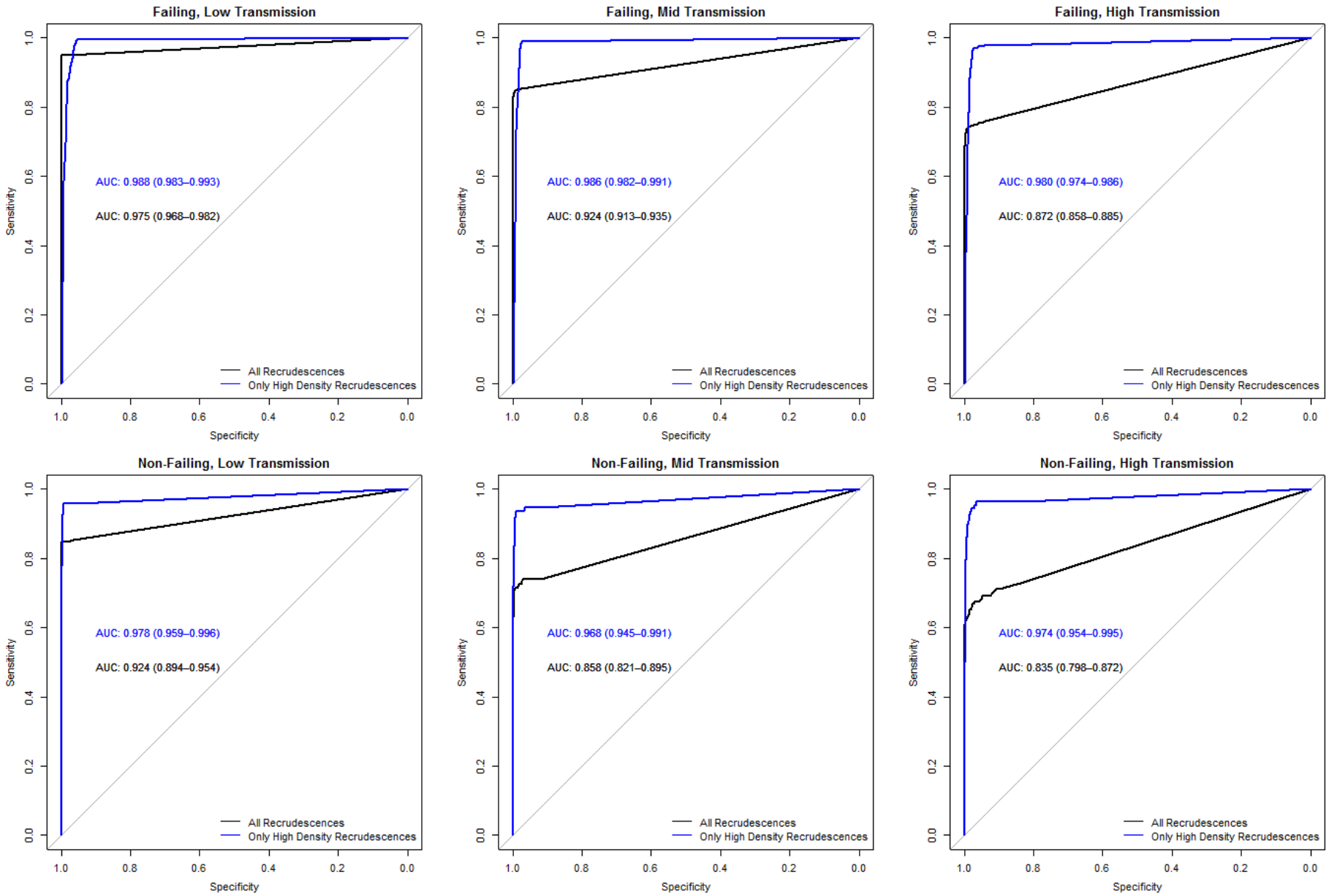
**Figure 4.4** shows the analysis of the specificity and sensitivity of the Bayesian algorithm in its ability to identify low and high density recrudescences. The general trend was that the AUC of the ROC curve decreased as transmission intensity increased with values of 0.872 and 0.835 in the failing and non-failing high transmission scenarios respectively – these correspond to a "good" diagnostic test. AUC was higher for any given transmission scenario in failing AR-LF than non-failing AR-LF. When the ROC curve was calculated for only high-density recrudescence AUC increased to ≥0.968 in all scenarios – an "excellent" diagnostic test.

#### 4.3.1.3 Distribution of posterior probability of recrudescence.

**Figure 4.5** shows the distribution of the posterior probabilities of recrudescence for all recurrences, stratified according to the true classification of their recurrence: Reinfection, low-density recrudescence, or high-density recrudescence. The distributions were nearly binary in every scenario: Nearly all posterior probabilities in the patient population were <0.1 or≥0.9. Some trends here were intuitive (note different scales on the Y axes): i.e., larger number of reinfections occurred as transmission intensity increased and larger number of recrudescences occurred in scenarios in which failing drugs were administered. The small number of patients whose infections had estimated probabilities of recrudescence between (but not including) 0.1 and 0.9 was reflected in the minor changes in failure rate estimates as p changed in **Figure 4.2** and **Figure 4.3** .

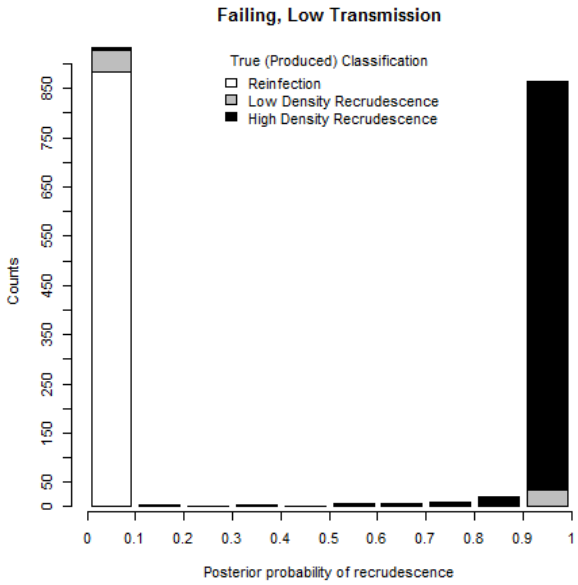
Most patients whose recurrence had $p$ <0.1 were reinfections. Given that ≥0.1 was the choice of p that produces the most accurate failure rate estimate , the cause of the (slight) under-estimate of failure rate was due to the proportion of patients with infections at $p$ <0.1 who had, in reality, recrudescent infections. For failing drugs, ~5% of recurrent infections were recrudescent infections with $p$ <0.1 at all transmission intensities. For non-failing drugs, ~2.5% of recurrent infections were recrudescent infections with $p$ <0.1 at all transmission intensities. Notably most of these were low density recrudescence; only 0.03%-0.05% of recurrent infections with $p$ <0.1 were high-density recrudescences for failing drug scenarios, and 0.02%-0.06% of recurrent infections with $p$ <0.1 were high-density recrudescences for non-failing drug scenarios. There were a small number of recurrent infections with $p$ ≥0.1 which were truly reinfections but in all scenarios this number was small relative to the number of recrudescent infections where $p$ <0.1. Consequently, the under-estimation due to truly recrudescent infections having $p$ <0.1 was greater than the over-estimation due to reinfections having $p$ ≥0.1; thus these reinfections with $p$ ≥0.1 were not leading to an over-estimation of failure rate.

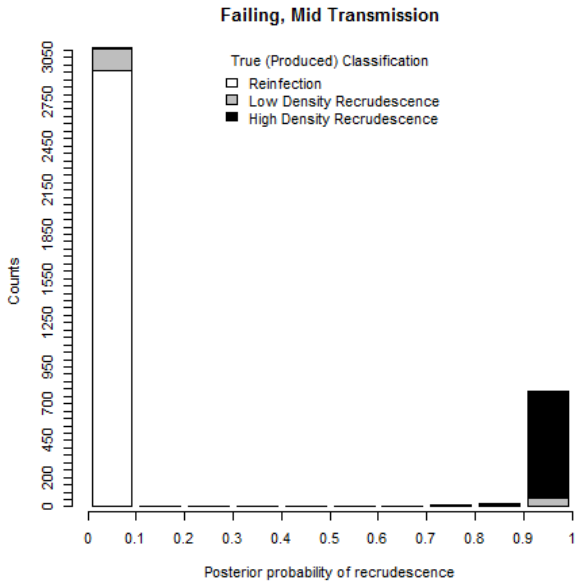
**Figure 4.2 Figure 4.3** and **Figure 4.5** show that over-estimation of failure rate due to misclassification of reinfection as recrudescence did not significantly affect the Bayesian algorithm due to its high specificity (an assertion that is further supported by ROC analysis in **Figure 4.4**); nearly all reinfections had a posterior probability of recrudescence of <0.1. A slight-under-estimate of failure rate occurred with all values of $p$ ≥0.1 to ≥0.9 inclusive to classify a recrudescence, due to the algorithm assigning posterior probabilities of <0.1 to a small proportion of infections with low density recrudescence.

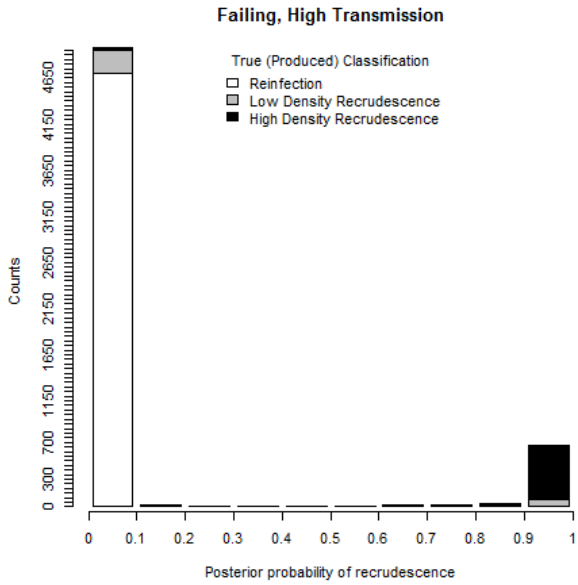

Failing, Low Transmission
AUC: 0.988 (0.983–0.993)
AUC: 0.975 (0.968–0.982)
Failing, Mid Transmission
AUC: 0.986 (0.982–0.991)
AUC: 0.924 (0.913–0.935)
Failing, High Transmission
AUC: 0.980 (0.974–0.986)
AUC: 0.872 (0.858–0.885)
Non-Failing, Low Transmission
AUC: 0.978 (0.959–0.996)
AUC: 0.924 (0.894–0.954)
Non-Failing, Mid Transmission
AUC: 0.968 (0.945–0.991)
AUC: 0.858 (0.821–0.895)
Non-Failing, High Transmission
AUC: 0.974 (0.954–0.995)
AUC: 0.835 (0.798–0.872)
All Recrudescences
Only High Density Recrudescences
Sensitivity
Specificity
1.0
0.8
0.6
0.4
0.2
0.0

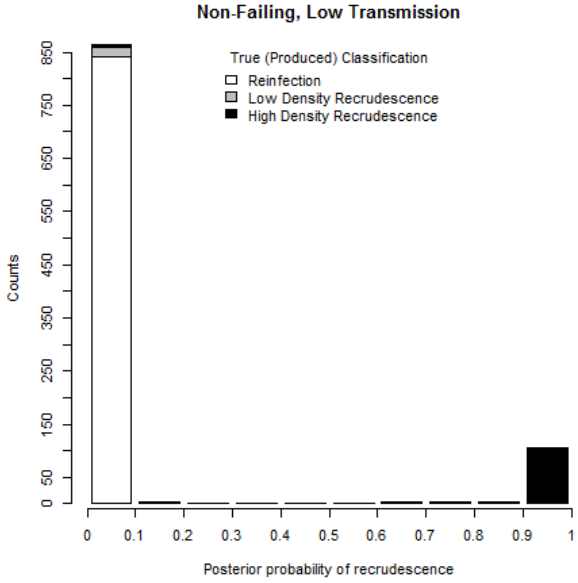
2228

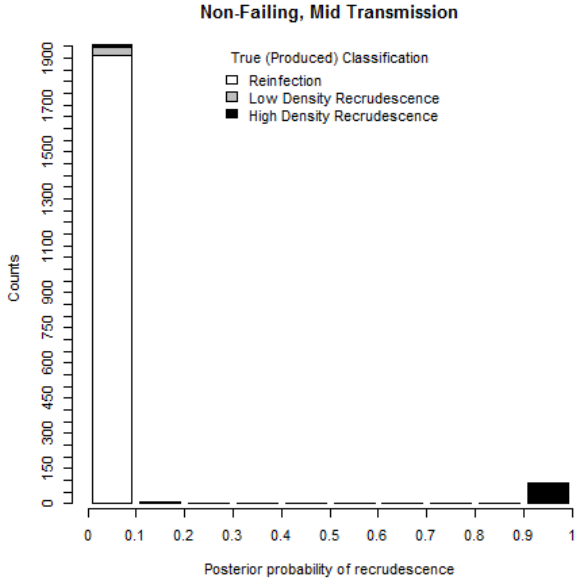
**Figure 4.4** Receiver operator characteristic (ROC) curves showing diagnostic ability of the Bayesian analysis method for 3 scenarios of transmission intensity for non-failing and failing artemether-lumefantrine (AR-LF). ROC curves and area under the roc curve (AUC) are shown for all recrudescence and for high density recrudescence. A high density recrudescence was defined as explained in **4.2.5.**

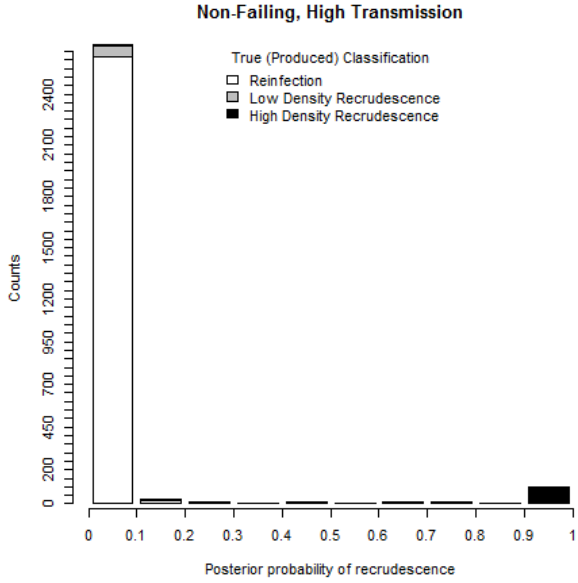

Failing, Low Transmission
True (Produced) Classification
Reinfection
Low Density Recrudescence
High Density Recrudescence
Counts
0 50 150 250 350 450 550 650 750 850
0 0.1 0.2 0.3 0.4 0.5 0.6 0.7 0.8 0.9 1
Posterior probability of recrudescence


Failing, Mid Transmission
True (Produced) Classification
Reinfection
Low Density Recrudescence
High Density Recrudescence
Counts
0 200 450 700 950 1250 1550 1850 2150 2450 2750 3050
0 0.1 0.2 0.3 0.4 0.5 0.6 0.7 0.8 0.9 1
Posterior probability of recrudescence


Failing, High Transmission
True (Produced) Classification
Reinfection
Low Density Recrudescence
High Density Recrudescence
Counts
0 300 700 1150 1650 2150 2650 3150 3650 4150 4650
0 0.1 0.2 0.3 0.4 0.5 0.6 0.7 0.8 0.9 1
Posterior probability of recrudescence


Non-Failing, Low Transmission
True (Produced) Classification
Reinfection
Low Density Recrudescence
High Density Recrudescence
Counts
0 50 150 250 350 450 550 650 750 850
0 0.1 0.2 0.3 0.4 0.5 0.6 0.7 0.8 0.9 1
Posterior probability of recrudescence


Non-Failing, Mid Transmission
True (Produced) Classification
Reinfection
Low Density Recrudescence
High Density Recrudescence
Counts
0 150 300 450 600 750 900 1100 1300 1500 1700 1900
0 0.1 0.2 0.3 0.4 0.5 0.6 0.7 0.8 0.9 1
Posterior probability of recrudescence


Non-Failing, High Transmission
True (Produced) Classification
Reinfection
Low Density Recrudescence
High Density Recrudescence
Counts
0 200 450 700 950 1200 1500 1800 2100 2400
0 0.1 0.2 0.3 0.4 0.5 0.6 0.7 0.8 0.9 1
Posterior probability of recrudescence

2232

**Figure 4.5** Distribution of the posterior probabilities of recrudescence estimated by the Bayesian algorithm for 3 scenarios of transmission intensity for non-failing and failing artemether-lumefantrine (AR-LF). A high density recrudescence was defined as explained in **4.2.5**

#### 4.3.1.4 Determinants of posterior probability of recrudescence.

**Figure 4.6** is a contour plot showing the estimated posterior probabilities of recrudescence estimated by the Bayesian algorithm as a function of the densities of the recrudescent clone(s) in the recurrent and initial sample. There was a clear trend of the posterior probability of recrudescence increasing as both densities increase, reinforcing the result illustrated in **Figure 4.5**: the density of recrudescent clones was an important determinant of the posterior probability of recrudescence returned for a given patient. The slight under-estimate of failure rates that occurred using the Bayesian algorithm (**4.3.1.1**) were due almost entirely to the finite sensitivity of genotyping causing some low-density clones to be missed during genotyping.

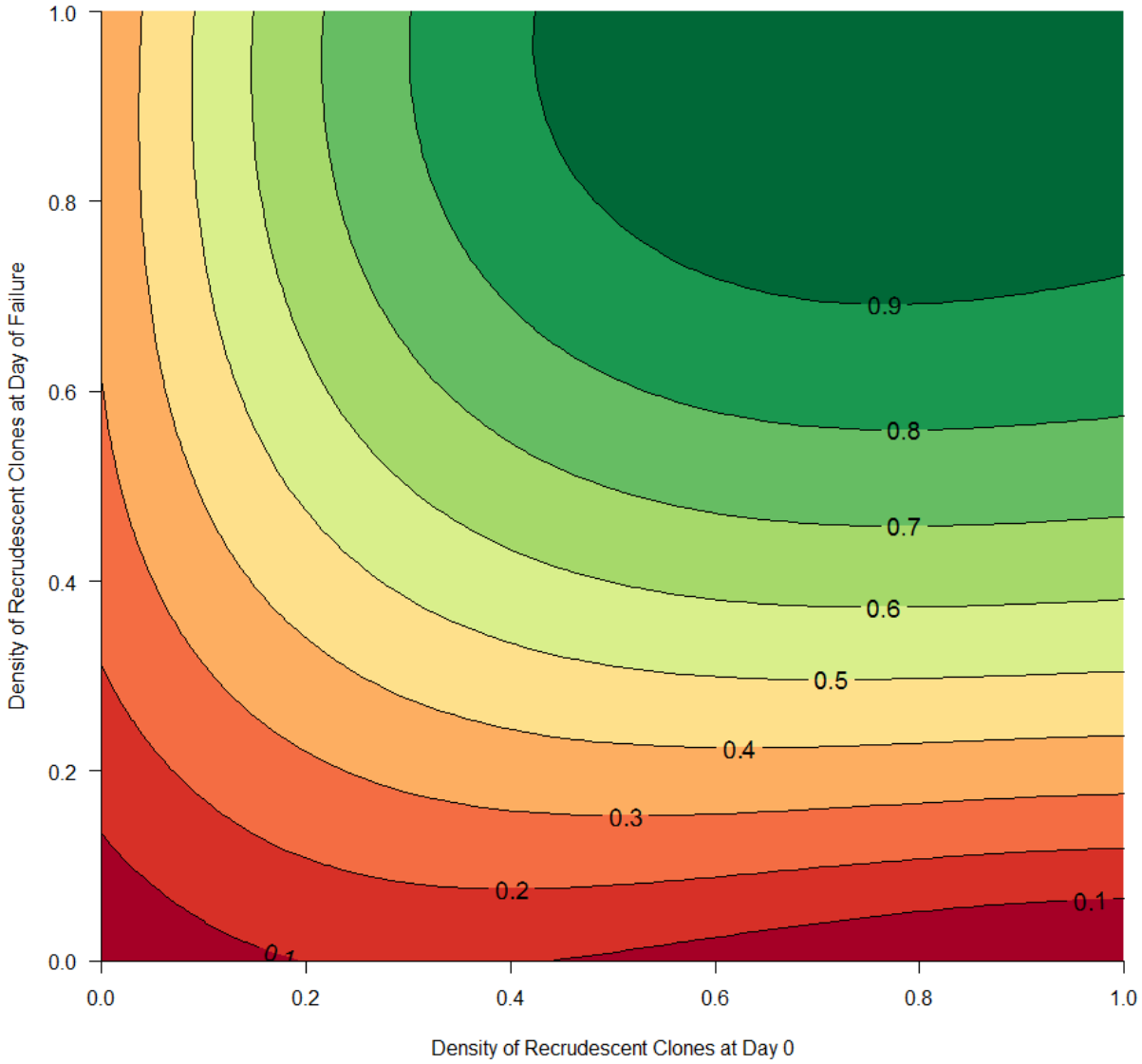



**Figure 4.6** Contour plot of the posterior probability of recrudescence estimated by Bayesian algorithm as a function of the density of recrudescent clones (i.e., the proportion of the recrudescent clones in the total infection biomass) in the initial sample and the recurrent sample. This plot is the combined data of all 6 scenarios modelled for artemether-lumefantrine (AR-LF). Each contour line indicates the posterior probability of recrudescence and the area between the lines the number of recurrent infections in the population with those posterior probabilities.

#### 4.3.1.5 The impact of patients with undetectable parasitaemia during follow-up.

It is possible, both in vivo and in silico, for patients to bear infections (either initial infections or reinfections) that do not fully clear during follow-up but also never become detectable (see Figure **1.1**). More specifically, some patients possessed low, but non-detectable parasitaemia on the final day of follow-up, but never had observable recurrence during follow-up. This occurred because the patient had a total parasitaemia of $<10^8$, throughout follow-up so parasites were unobservable through microscopy (unless venous blood was taken, which is not current practice); thus, no genotype was taken.

A patient with an initial infection that does not clear is, by definition, a true failure. However, in the absence of detectable parasites, the patient cannot be classified as a recrudescence or a reinfection (correctly or incorrectly) and is classified as a treatment success on the final day of follow-up. Because one of the purposes of this research is determining the accuracy of the Bayesian algorithm, it was necessary to quantify the number of these patients exactly.

The number of patients who had undetectable parasitaemia throughout follow-up (i.e. a total $<10^8$ parasites, either reinfection or initial infection) was calculated, as was the proportion of these patients who had undetectable recrudescent infections (as opposed to undetectable reinfections). These results are shown in **Table 4.6** . The proportion of patients with non-detectable parasitaemia increased as transmission intensity increased, but the proportion of patients with non-detectable parasitaemia who are true failures was consistently extremely low. The chance of a patient having an initial clone that does not clear, does not increase to detectable levels, and that patient not having reinfections that lead to parasites become detectable is extremely low. In other words there is a negligible (nearly zero) under-estimation of failure rate from patients who are true failures but who never have an observed recurrence.

**Table 4.6** The proportion of patients who have undetectable parasites during follow-up (due to having total parasitaemia $<10^8$ , such that parasites are not observed by light microscopy and no genotype is taken), and the proportion of patients with undetectable parasites and a recrudescent clone(s) (i.e. the proportion of patients who are true failures but do not have a recurrence).

| Scenario | Proportion of patients with non-detectable parasitaemia on the final day of the follow-up | Proportion of patients with non-detectable parasitaemia during follow-up who are also true failures |
|---|---|---|
| Low transmission, non-failing AR-LF | 0.0377 | 0.0006 |
| Medium transmission, non-failing AR-LF | 0.1462 | 0.001 |
| High transmission, non-failing AR-LF | 0.2269 | 0.0017 |
| Low transmission, failing AR-LF | 0.0313 | 0 |
| Medium transmission, failing AR-LF | 0.0746 | 0.0004 |
| High transmission, failing AR-LF | 0.1023 | 0.0004 |

*AR: Artemether, LF: Lumefantrine*

### 4.3.2 Analysis of AS-MQ.

AS-MQ treatment was simulated and analysed in the same manner as for AR-LF, i.e., the ability to recover the true failure rate of the match counting algorithm and Bayesian analysis across a variety of cut-off points, the diagnostic ability in the form of receiver operator characteristic (ROC) curves and analysis of the distribution of the posterior probability of recrudescence in the Bayesian analysis.

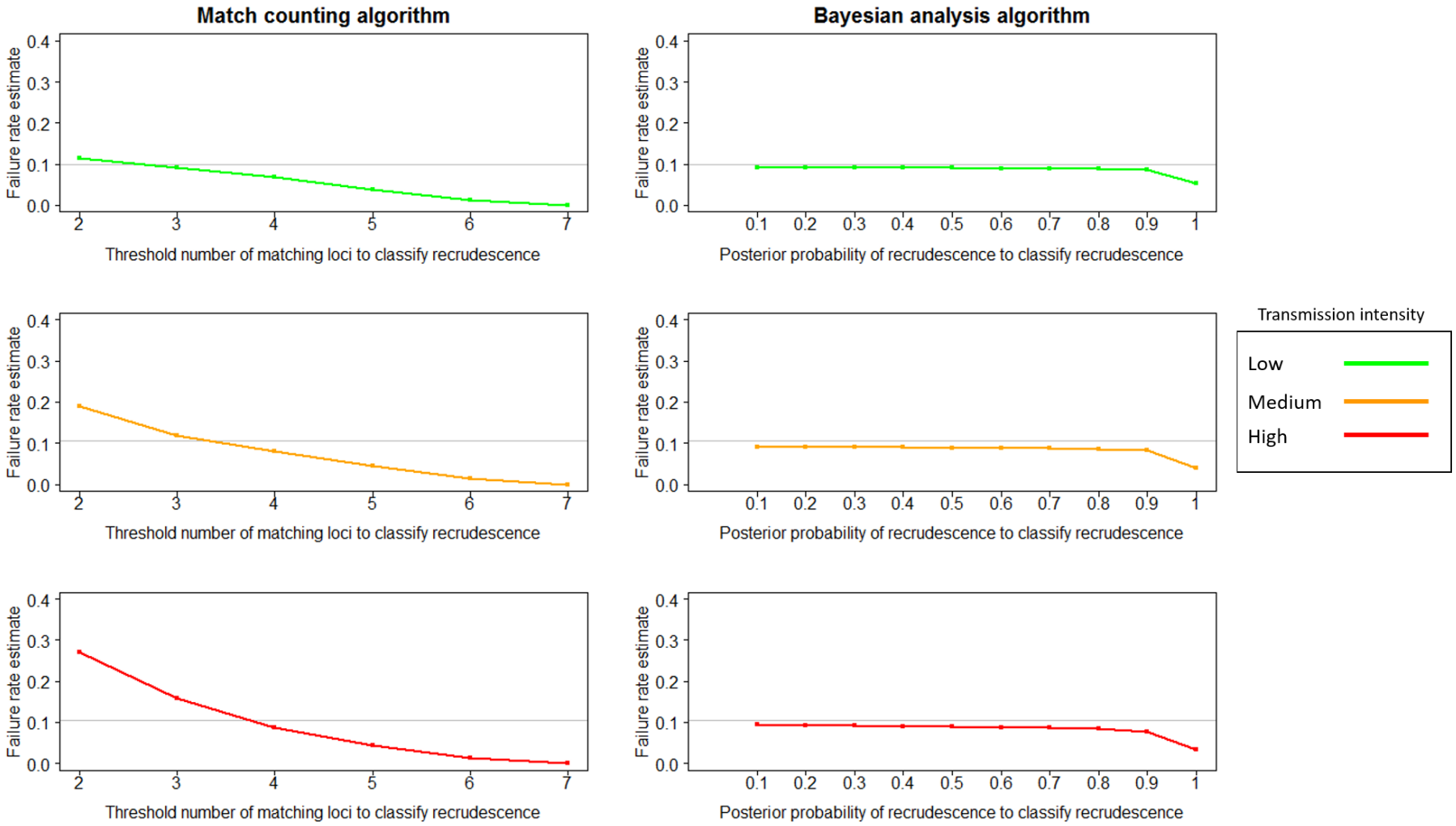
Results were very consistent with those of AR-LF: The match counting algorithm for classifying recurrences as reinfection or recrudescence could not consistently provide accurate failure rate estimates across a variety of scenarios and often resulted in extreme over or under-estimates of true failure rate, depending on the choice of threshold. The Bayesian analysis method generated failure rate estimates to a high degree of accuracy across all scenarios, although there was an under-estimate of 1.6 percentage units in the high transmission, failing drug scenario. As with AR-LF, using $p \geq 0.1$ to classify an infection as a recrudescence provided the most accurate failure rate estimate for AS-MQ in every scenario.

#### 4.3.2.1 Failure rate estimates and comparison to true failure rate.

Six thresholds (2 to 7 matching loci between initial and recurrent sample inclusive) were analysed and the results presented in **Figure 4.7 Figure 4.8** for non-failing AS-MQ. The trends were similar to those seen for AR-LF**:** using a low threshold over-estimated failure rates, using a higher threshold under-estimated the true failure rate, and no threshold consistently recovered an accurate estimate across all failing drug scenarios – while a threshold of 3 produces a relatively accurate failure rate in a low transmission, failing drug scenario, the same threshold will over-estimate true failure rate in medium and high transmission scenarios. Note, though, for non-failing drugs, a threshold of 3 recovered the true failure rate in all transmission scenarios – this did not occur with AR-LF and likely occurs with AS-MQ because the number of reinfections that occur is lower due to the superior post-treatment prophylactic properties of MQ compared to LF. The over-estimate of true failure rate that occurred in medium and high transmission scenarios with a threshold of 2 or 3 was much lower with AS-MQ for both failing and non-failing drugs than with AR-LF – again, this occurred because the prophylactic effects of AS-MQ meant there were fewer reinfections occurring that could be misclassified as recrudescence (and the chance of misclassifying a reinfection is, obviously, greater at lower thresholds).

The Bayesian analysis method for AS-MQ was analysed using 9 break-points of $p$ from 0.1 to 1 (i.e., the posterior probability above which a patient's recurrence was classified as recrudescence) across all scenarios. The Bayesian algorithm under-estimates true failure rate slightly using all values of $p \geq 0.1$ and thus classifying recurrent infections as recrudescence at $p \geq 0.1$ is the value of $p$ that will result in the most accurate failure rate estimate. This under-estimate is not large: with a $p$ of 0.1, the low transmission scenario has a failure rate estimate of 0.914 relative to a true failure of 0.0978, the medium transmission scenario has a failure rate estimate of 0.915 relative to a true failure of 0.106 and the high transmission scenario has a failure rate estimate of 0.915 relative to a true failure of 0.1058. In short, using $p$ of $\geq 0.1$ to classify recrudescence, the Bayesian estimate is within 0.016 (1.6%) of the true failure rate in all failing drug scenarios. Failure rate estimates were highly consistent across transmission intensities (as was observed with AR-LF), highlighting the robustness of the Bayesian analysis method.

For all non-failing drug scenarios, the Bayesian estimate is even closer to the true failure rate; for example $p \geq 0.1$ produces a failure rate estimate of 0.0061 relative to a true failure rate of 0.0119 for the high transmission scenario, so while this is an under-estimate the absolute difference is so small that, practically, the method recovers the true failure rate with extreme accuracy for the non-failing drug parameterization (as it does for AR-LF).




2330

2331 **Figure 4.7** Failure rate estimates obtained using either the match counting algorithm or the Bayesian analysis algorithm for failing AS-MQ in low, medium and
2332 high transmission scenarios. The true failure rate is denoted in each plot by the horizontal grey line. For the match counting algorithm, the threshold for the
2333 number of matching loci with which a recurrence is classified as a recrudescence varies between 2 and 7. For the Bayesian analysis, the cut-off for posterior
2334 probability at which a recurrence is classified as a recrudescence varies between ≥0.1 and ≥0.9. This plot is analogous to **Figure 4.2** for failing AR-LF.

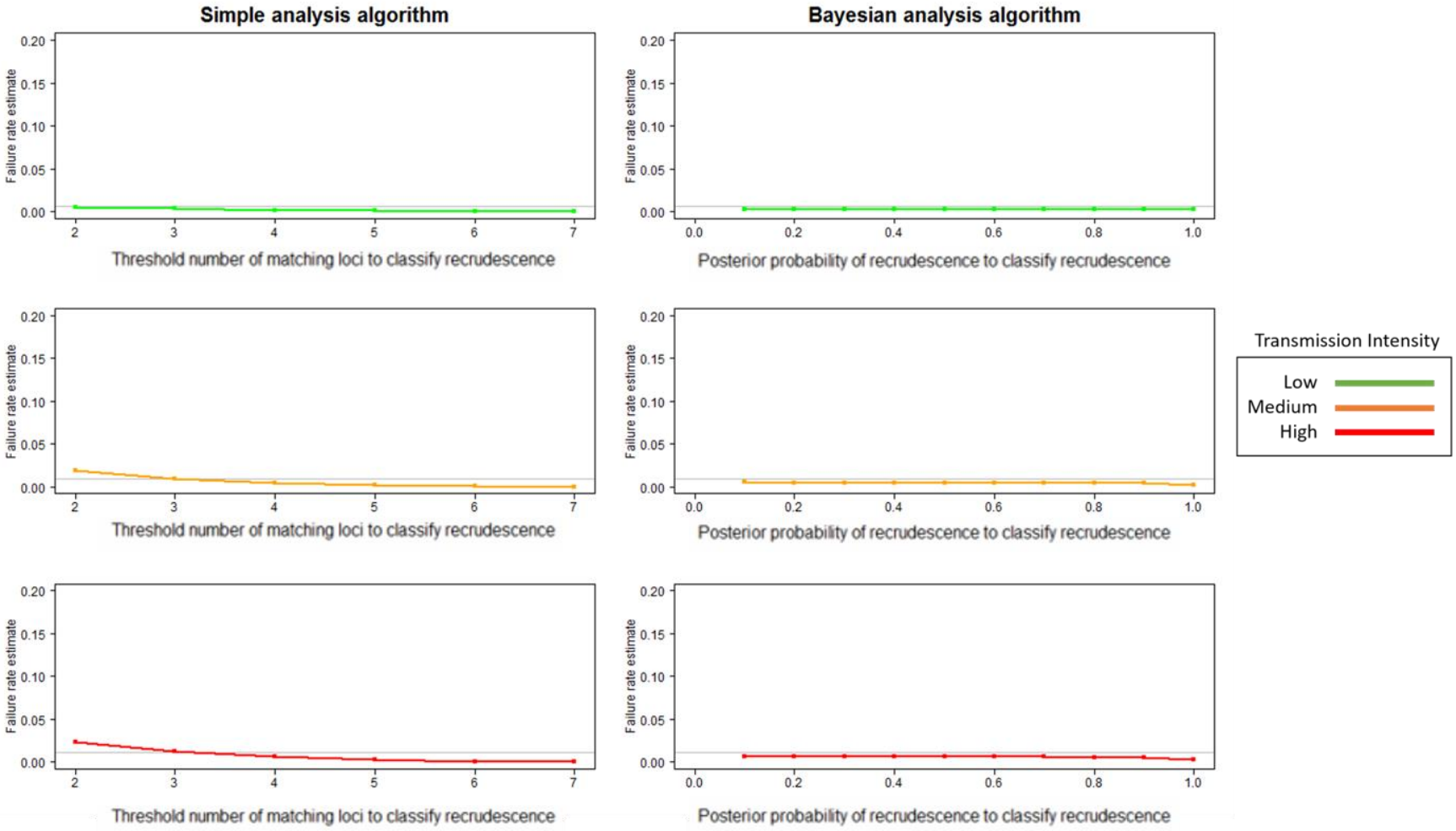



**Figure 4.8** Failure rate estimates obtained using either the match counting algorithm or the Bayesian analysis algorithm for non-failing AS-MQ in low, medium and high transmission scenarios. The true failure rate is denoted in each plot by the horizontal grey line. For the match counting algorithm, the threshold for the number of matching loci with which a recurrence is classified as a recrudescence varies between 2 and 7. For the Bayesian analysis, the cut-off for posterior probability at which a recurrence is classified as a recrudescence varies between ≥0.1 and ≥0.9. This plot is analogous to **Figure 4.3** for non-failing AR-LF.

#### 4.3.2.2 Receiver Operator Characteristic (ROC) curves for the Bayesian algorithm.

ROC curves for the Bayesian algorithm are shown in Figure 4.9, where the ROC curve was constructed using the posterior probability at which an infection would be classified as a recrudescence (from 0 to 1). An area under the curve (AUC) of ≥0.8 is considered a "good diagnostic test" while an AUC of ≥0.9 is considered an "excellent diagnostic test".

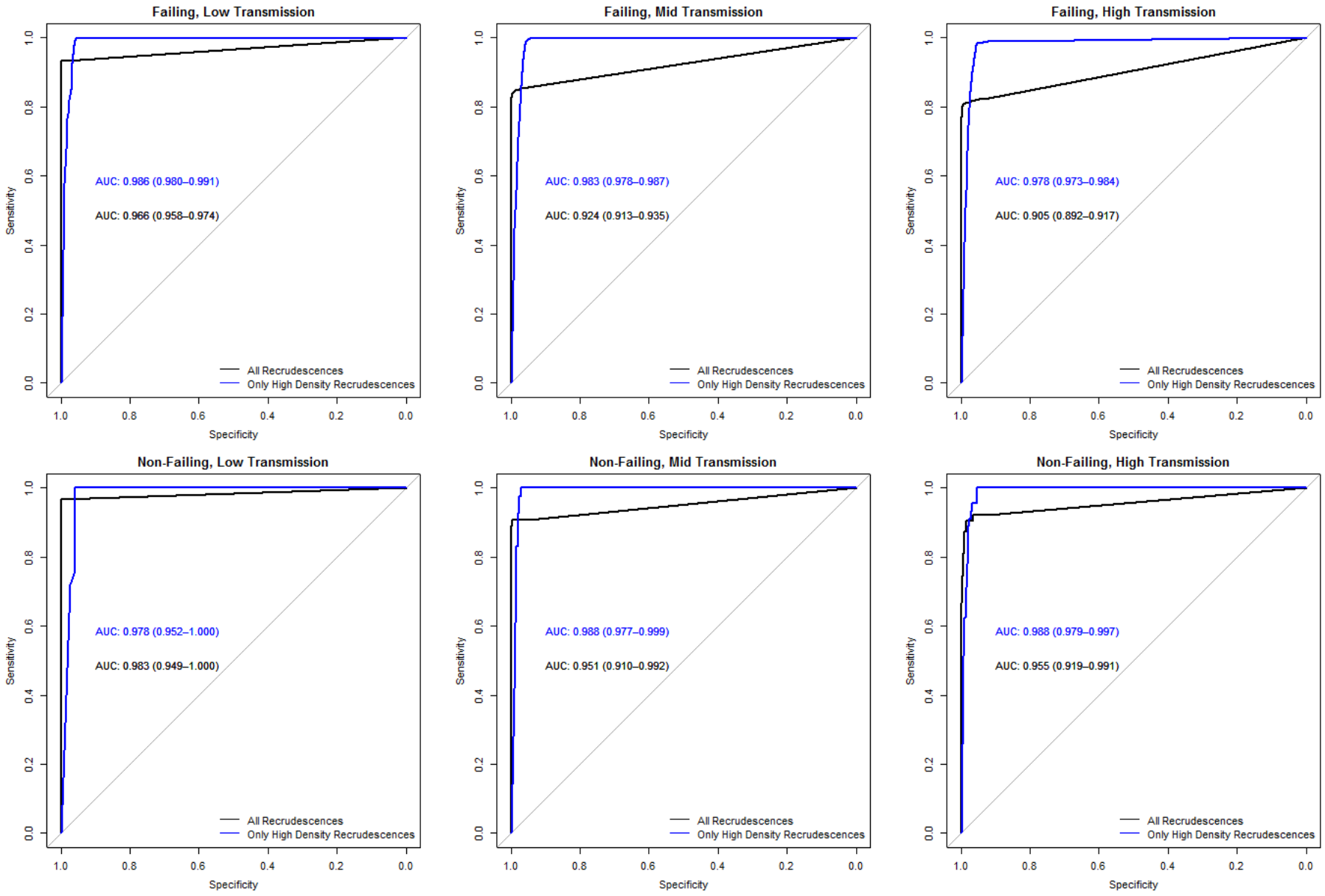
The ROC curves show the Bayesian method to be an excellent diagnostic test for both non-failing and failing AS-MQ in all transmission intensity scenarios. There was a notable trend in the failing drug scenarios that AUC decreases as transmission intensity increases (AUC is "only" 0.905 in the failing drug, high transmission scenario). Similar to the results for AR-LF, the method had a higher diagnostic ability when considering only high density recrudescence (where AUC was very close to 1 in all scenarios).

AUC was greater in all scenarios for AS-MQ than AR-LF, and this difference was more pronounced in the non-failing drug scenarios than the failing drug scenarios. This difference appeared to arise from differences in sensitivity rather than specificity and was negligible when considering only high-density recrudescence. Consequently, these results appear to show that the Bayesian method had a higher sensitivity when analysing AS-MQ than AR-LF – the cause of this is most likely the increased prophylactic effect of AS-MQ resulting in fewer reinfections becoming patent; thus, recurrences are less likely to be reinfections, which could mask low density recrudescence (i.e., the recrudescing clone was not able to increase in frequency to a detectable parasitaemia before the reinfection became detectable and the patient was removed from follow-up for further treatment).

#### 4.3.2.3 Distribution of posterior probability of recrudescence.

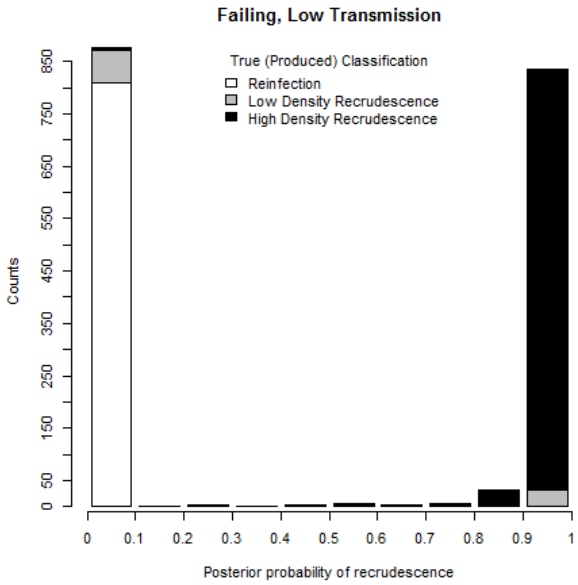
**Figure 4.10** shows the distribution of the posterior probabilities of recrudescence for all recurrent infections, stratified according to the true classification of their recurrence: Reinfection, low density recrudescence, or high density recrudescence. As with AR-LF, distributions are nearly binary in every scenario: Nearly all posterior probabilities in the patient population are <0.1 or ≥0.9.

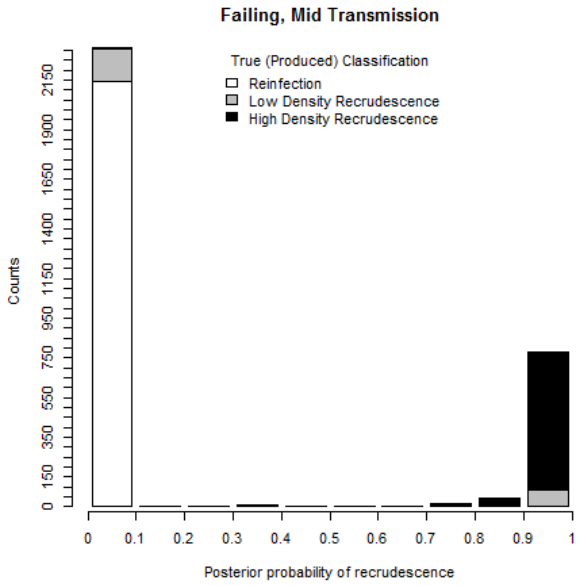
Results were markedly similar to those of AR-LF (**Figure 4.5**); the under-estimate of failure rate at $p$ ≥0.1 was due to the small proportion of infections with $p$ ≥0.1 that are a low-density recrudescence. The number of infections that are low density recrudescence and $p$ ≥0.1 is extremely low (and is 0 for all the non-failing drug scenarios).

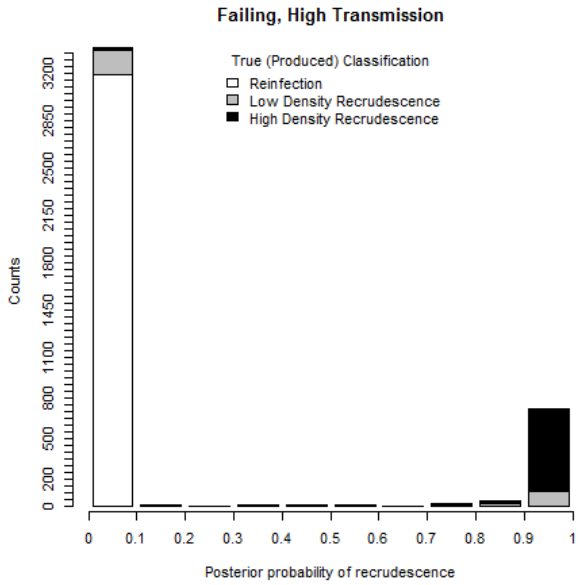

Failing, Low Transmission
AUC: 0.986 (0.980–0.991)
AUC: 0.966 (0.958–0.974)
Failing, Mid Transmission
AUC: 0.983 (0.978–0.987)
AUC: 0.924 (0.913–0.935)
Failing, High Transmission
AUC: 0.978 (0.973–0.984)
AUC: 0.905 (0.892–0.917)
Non-Failing, Low Transmission
AUC: 0.978 (0.952–1.000)
AUC: 0.983 (0.949–1.000)
Non-Failing, Mid Transmission
AUC: 0.988 (0.977–0.999)
AUC: 0.951 (0.910–0.992)
Non-Failing, High Transmission
AUC: 0.988 (0.979–0.997)
AUC: 0.955 (0.919–0.991)
All Recrudescences
Only High Density Recrudescences
Sensitivity
Specificity
1.0
0.8
0.6
0.4
0.2
0.0

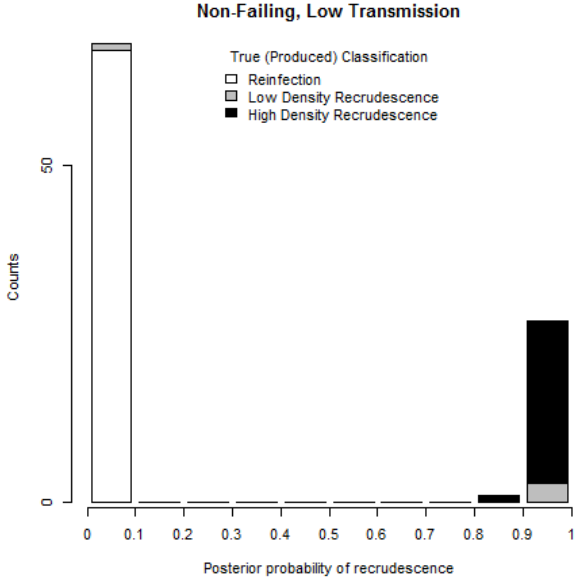
2379

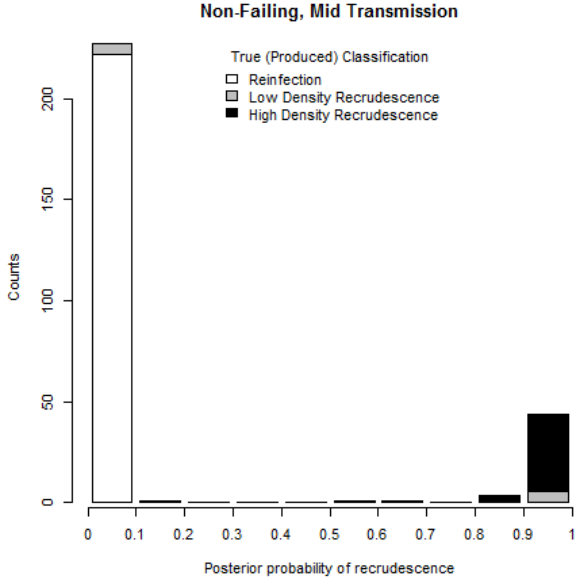
**Figure 4.9** : Receiver operator characteristic (ROC) curves showing diagnostic ability for the Bayesian analysis method for 6 model scenarios for artesunate-mefloquine (AS-MQ). ROC curves and area under the roc curve (AUC) are shown for all recrudescence and for high density recrudescence, where a high density recrudescence is defined in **4.2.5**. This plot is complementary to the plot for artemether-lumefantrine (AR-LF) shown in **Figure 4.4**

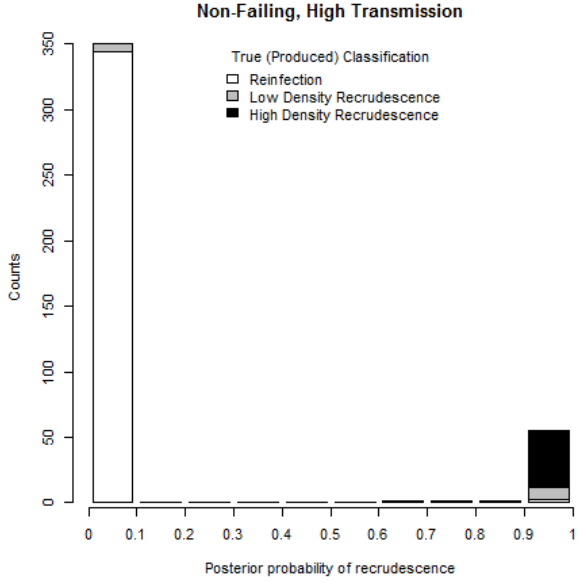

Failing, Low Transmission
True (Produced) Classification
Reinfection
Low Density Recrudescence
High Density Recrudescence
Counts
0 50 150 250 350 450 550 650 750 850
0 0.1 0.2 0.3 0.4 0.5 0.6 0.7 0.8 0.9 1
Posterior probability of recrudescence


Failing, Mid Transmission
True (Produced) Classification
Reinfection
Low Density Recrudescence
High Density Recrudescence
Counts
0 150 350 550 750 950 1150 1400 1650 1900 2150
0 0.1 0.2 0.3 0.4 0.5 0.6 0.7 0.8 0.9 1
Posterior probability of recrudescence


Failing, High Transmission
True (Produced) Classification
Reinfection
Low Density Recrudescence
High Density Recrudescence
Counts
0 200 500 800 1100 1450 1800 2150 2500 2850 3200
0 0.1 0.2 0.3 0.4 0.5 0.6 0.7 0.8 0.9 1
Posterior probability of recrudescence


Non-Failing, Low Transmission
True (Produced) Classification
Reinfection
Low Density Recrudescence
High Density Recrudescence
Counts
0 50
0 0.1 0.2 0.3 0.4 0.5 0.6 0.7 0.8 0.9 1
Posterior probability of recrudescence


Non-Failing, Mid Transmission
True (Produced) Classification
Reinfection
Low Density Recrudescence
High Density Recrudescence
Counts
0 50 100 150 200
0 0.1 0.2 0.3 0.4 0.5 0.6 0.7 0.8 0.9 1
Posterior probability of recrudescence


Non-Failing, High Transmission
True (Produced) Classification
Reinfection
Low Density Recrudescence
High Density Recrudescence
Counts
0 50 100 150 200 250 300 350
0 0.1 0.2 0.3 0.4 0.5 0.6 0.7 0.8 0.9 1
Posterior probability of recrudescence

2383

**Figure 4.10** Distribution of the posterior probability of recrudescence estimated the by Bayesian algorithm across 6 different scenarios for artesunate-mefloquine (AS-MQ). A high density recrudescence was defined as explained in **4.2.5**. This plot is complementary to the plot for artemether-lumefantrine (AR-LF) shown in **Figure 4.5**.

### 4.3.3 Analysis of AR-LF in extremely low genetic diversity scenarios.

Failing AR-LF and non-failing AR-LF in an area of extremely low genetic diversity (**4.2.2.2)** was simulated under scenarios of low, medium and high transmission. The model parameters MOI, FOI and PK/PD parameters (including IC50) are the same as for other simulations of failing AR-LF (i.e., as described in **4.2.1**).

Note that the parasite dynamics of each clone are identical to those in the results for AR-LF in the main text, i.e., the parasitaemia over time of clone 1 in patient 1 for the failing, high transmission scenario (main text) is the same as the failing, high transmission scenario presented here; however, the microsatellite alleles assigned to each clone differs, so the classification of each recurrent infection may also differ.

#### 4.3.3.1 Failure rate estimates and comparison to true failure rate.

True failure rate for failing and non-failing AR-LF at low, medium and high transmission scenarios, all calibrated with allelic distributions of extremely low genetic diversity were identical to the true failure rates of AR-LF when calibrated as described in **4.2.1 (Figure 4.2 )**. Failure rate estimates obtained using the match counting algorithm and the Bayesian analysis algorithm are shown for failing AR-LF and non-failing AR-LF in

**Figure** 4.11 and **Figure 4.12** respectively.

Six thresholds for the match counting algorithm (2 to 7 matches between initial and recurrent sample inclusive) were analysed. Compared to AR-LF with higher levels of genetic diversity (**4.3.1**), there was a clear trend of greatly increased failure rate estimates, particularly as transmission intensity increases. For example, a threshold of 3 matching loci for the failing, high transmission scenario gave a failure rate estimate of 35% with extremely low genetic diversity compared to 18% with the allelic distribution of a "high" transmission area i.e. **Figure 4.2** – this is an over-estimate of the true failure rate by 25%; an area of high transmission but very low genetic diversity could represent a genetic bottleneck in vivo. The results in

**Figure** 4.11 and **Figure 4.12** should be compared with those in **Figure 4.2** and **Figure 4.3** ; they clearly show that a reduction in genetic diversity will lead to increased failure rate estimates using a match counting algorithm and a threshold of 3 or 4 (i.e., the thresholds that produced the failure rate estimates closest to the true failure rate in areas of higher genetic diversity) as reinfections are misclassified as recrudescence. Failure rate estimates at a threshold value of at least 6 matching loci will still under-estimate the failure rate in all scenarios because, even in a very low diversity setting, the mathematical chance of a reinfection matching an initial clone at 6 or 7 loci is small.

The Bayesian analysis algorithm was then applied to the extremely low genetic diversity AR-LF simulations with nine values of $p$ at which to classify a recurrence as a recrudescence ranging from ≥0.1 to 1 across all patients, as previously described.

Failure rate estimates obtained using the Bayesian analysis algorithm are higher across the board with extremely low genetic diversity compared to the higher genetic diversity scenarios; though note that using $p \geq 0.1$ to classify a recrudescence will still produce the best failure rate estimate (and it remained highly accurate here, as it was with higher diversity). Note, though, that while using this $p$ as the threshold at which to classify a recrudescence would under-estimate true failure (slightly, see **Figure 4.2** and **Figure 4.3** with higher diversity, here there was a slight over-estimate in the high transmission intensity scenarios for both failing and non-failing AR-LF.

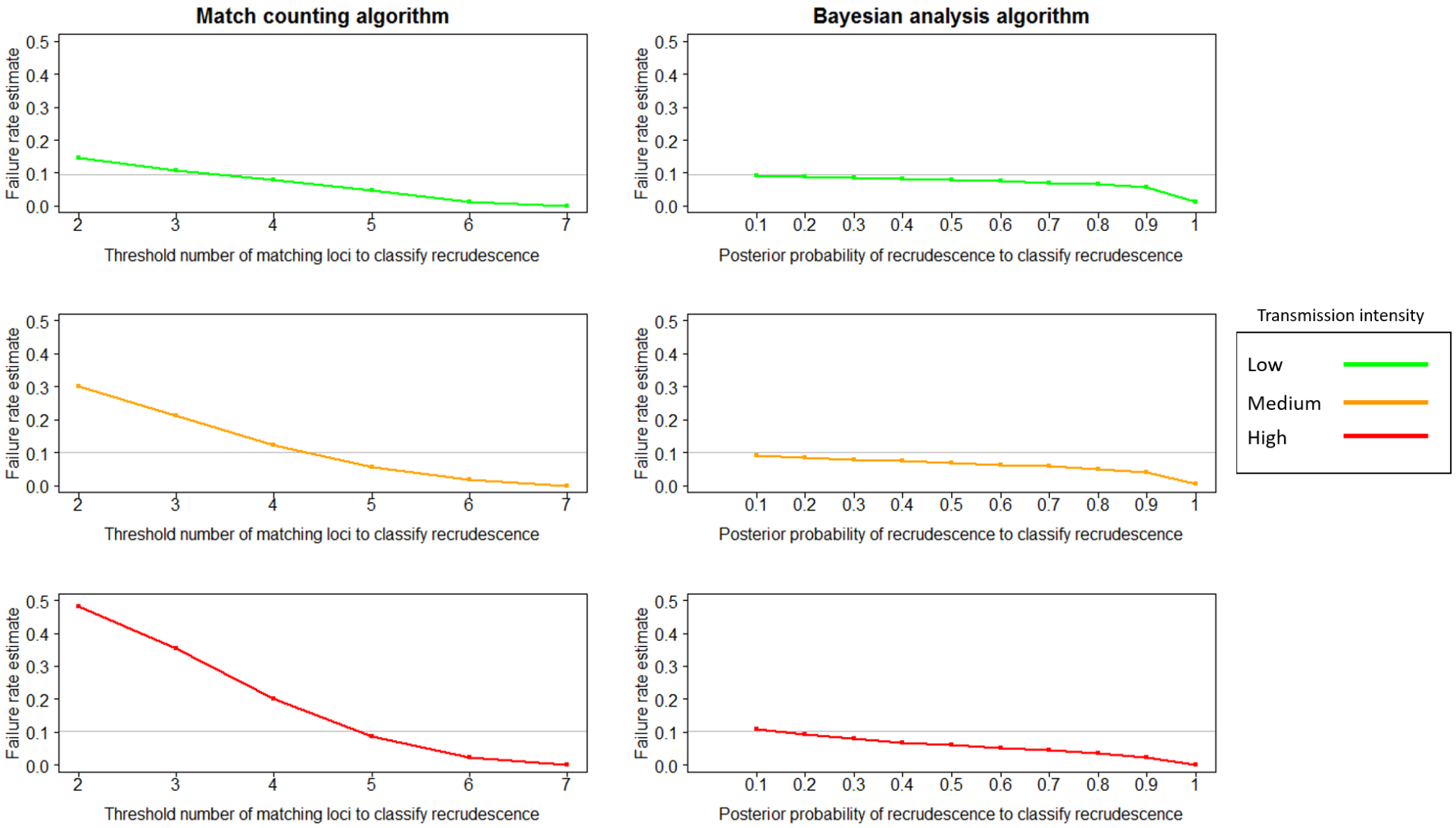
The most notable trend is that the decrease in failure rate estimates obtained from the Bayesian method was considerably sharper in a very low diversity setting – more so with higher transmission

intensity, such that using high *p* values as the threshold at which to classify a recurrence as a recrudescence resulted in larger under-estimates of true failure rate in a very low diversity setting than in a higher diversity setting.

#### 4.3.3.2 Receiver Operator Characteristic (ROC) curves for the Bayesian algorithm.

ROC curves for the Bayesian algorithm for AR-LF with very low genetic diversity are shown in **Figure 4.13** where the ROC curve was constructed using the posterior probability at which an infection would be classified as a recrudescence (from 0 to 1). An area under the curve (AUC) of ≥0.8 was considered a "good diagnostic test" while an AUC of ≥0.9 was considered an "excellent diagnostic test".

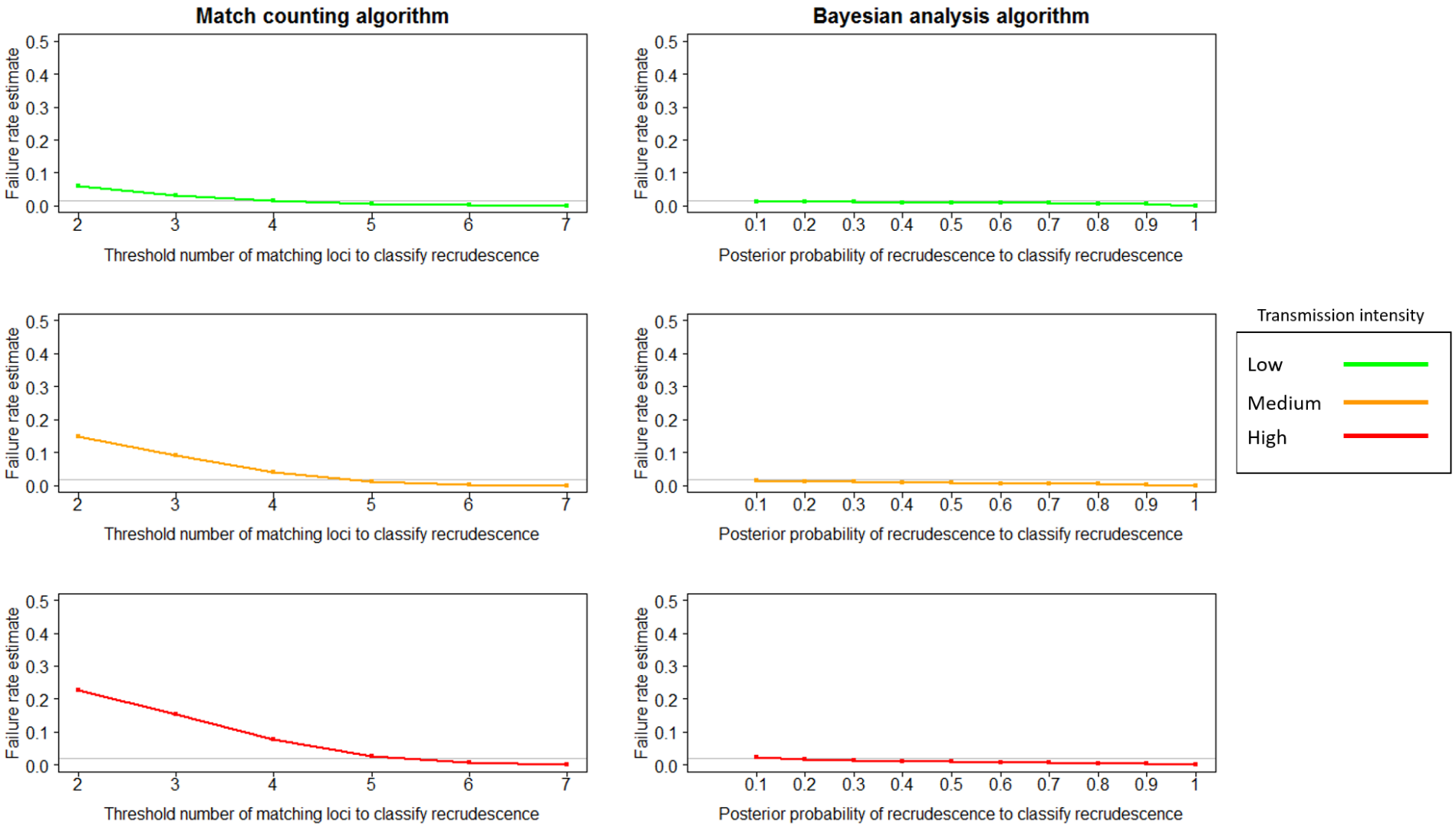
AUC is similar between the extremely low genetic diversity scenarios and the higher diversity scenarios; the trends shown in **Figure 4.13** follow the same key trends as those in **Figure 4.4** i.e., that AUC decreases as transmission intensity increases, driven by a decrease in sensitivity, and that the diagnostic ability for "high density recrudescence" is notably higher than that for all recrudescence.


Match counting algorithm
Bayesian analysis algorithm
Failure rate estimate
0.0
0.1
0.2
0.3
0.4
0.5
2
3
4
5
6
7
0.1
0.2
0.3
0.4
0.5
0.6
0.7
0.8
0.9
1
Threshold number of matching loci to classify recrudescence
Posterior probability of recrudescence to classify recrudescence
Transmission intensity
Low
Medium
High

2461

2462

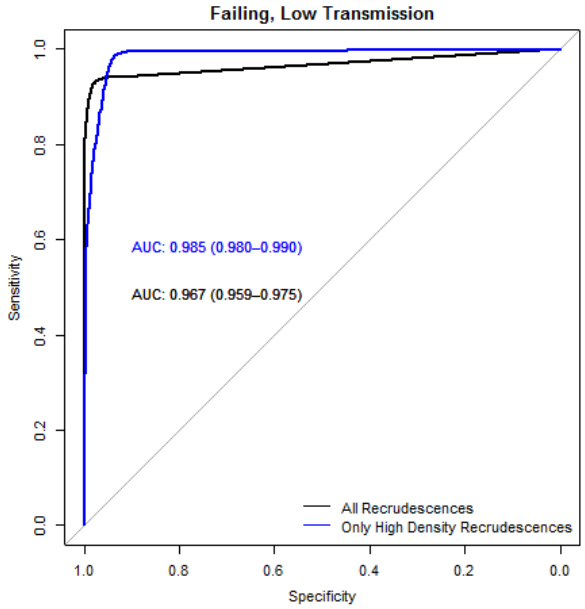
**Figure 4.11** Failure rate estimates obtained using either the match counting algorithm or the Bayesian analysis algorithm for failing AR-LF in an extremely low genetic diversity setting and with low, medium and high transmission intensity (MOI and FOI). The true failure rate is denoted in each plot by the horizontal grey line. For the match counting algorithm, the threshold for the number of matching loci with which a recurrence is classified as a recrudescence varies between 2 and 7. For the Bayesian analysis, the cut-off for posterior probability at which a recurrence is classified as a recrudescence varies between ≥0.1 and ≥0.9. This plot is analogous to **Figure 4.2** .

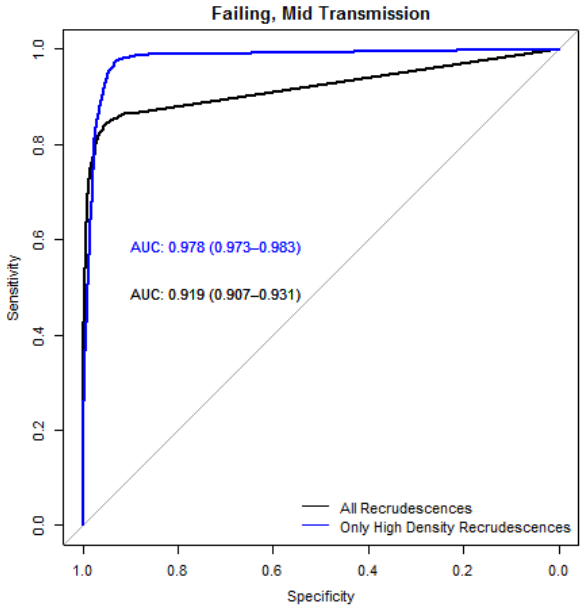

Match counting algorithm
Bayesian analysis algorithm
Failure rate estimate
0.0
0.1
0.2
0.3
0.4
0.5
2
3
4
5
6
7
0.1
0.2
0.3
0.4
0.5
0.6
0.7
0.8
0.9
1
Threshold number of matching loci to classify recrudescence
Posterior probability of recrudescence to classify recrudescence
Transmission intensity
Low
Medium
High

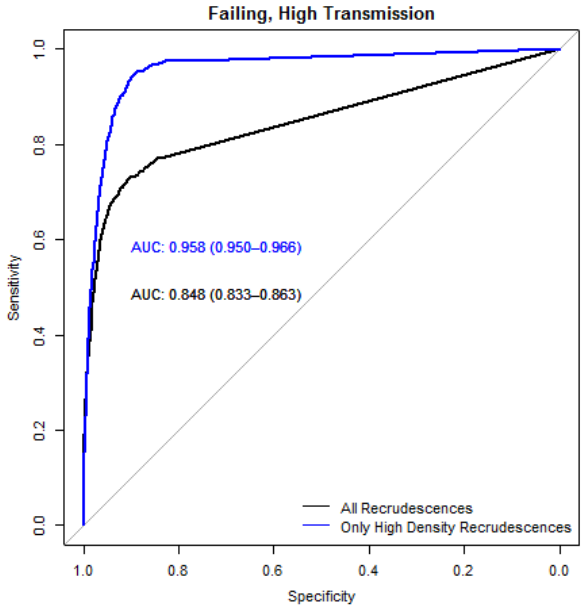
2469

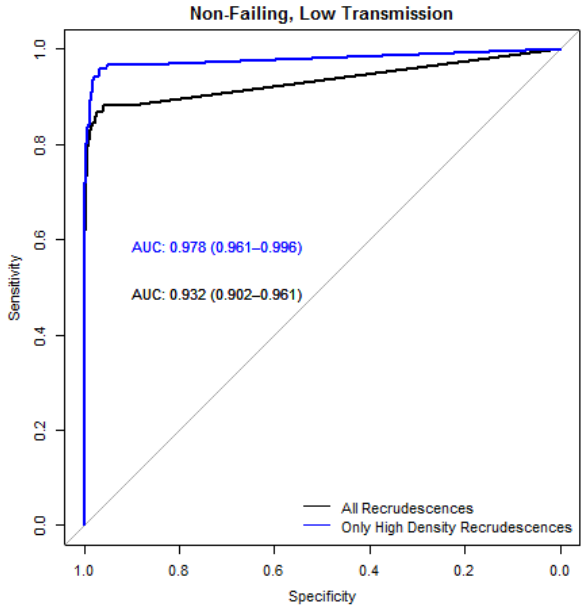
2470

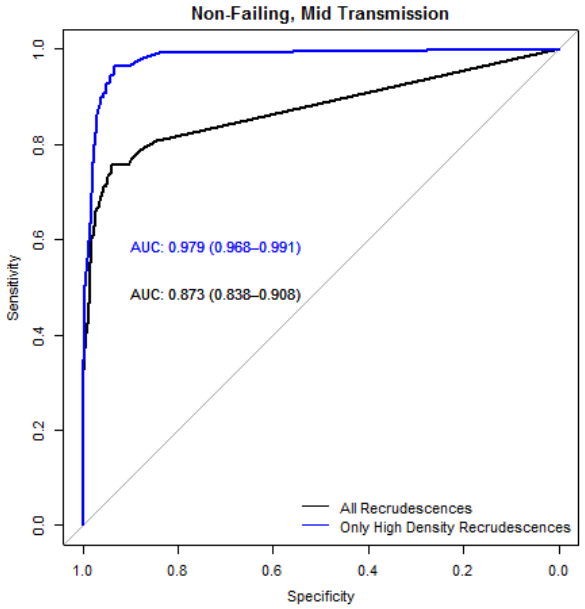
**Figure 4.12** Failure rate estimates obtained using either the match counting algorithm or the Bayesian analysis algorithm for failing AR-LF in an extremely low genetic diversity setting and with low, medium and high transmission intensity (MOI and FOI). The true failure rate is denoted in each plot by the horizontal grey line. For the match counting algorithm, the threshold for the number of matching loci with which a recurrence is classified as a recrudescence varies between 2 and 7. For the Bayesian analysis, the cut-off for posterior probability at which a recurrence is classified as a recrudescence varies between ≥0.1 and ≥0.9. This plot is analogous to **Figure 4.3** .

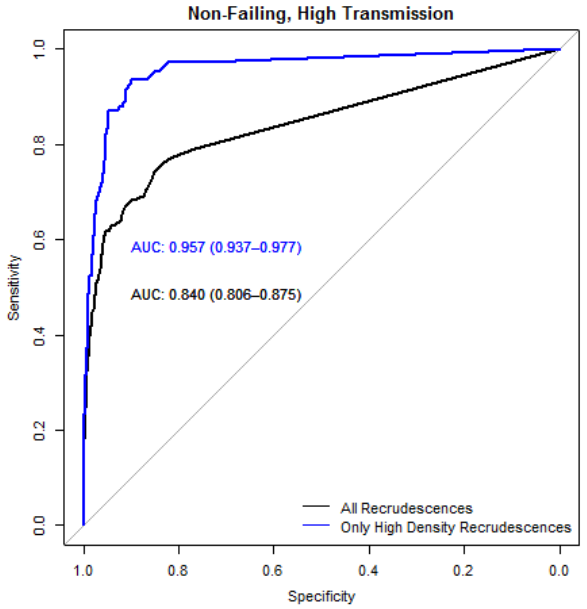

Failing, Low Transmission
Sensitivity
Specificity
AUC: 0.985 (0.980–0.990)
AUC: 0.967 (0.959–0.975)
All Recrudescences
Only High Density Recrudescences


Failing, Mid Transmission
Sensitivity
Specificity
AUC: 0.978 (0.973–0.983)
AUC: 0.919 (0.907–0.931)
All Recrudescences
Only High Density Recrudescences


Failing, High Transmission
Sensitivity
Specificity
AUC: 0.958 (0.950–0.966)
AUC: 0.848 (0.833–0.863)
All Recrudescences
Only High Density Recrudescences


Non-Failing, Low Transmission
Sensitivity
Specificity
AUC: 0.978 (0.961–0.996)
AUC: 0.932 (0.902–0.961)
All Recrudescences
Only High Density Recrudescences


Non-Failing, Mid Transmission
Sensitivity
Specificity
AUC: 0.979 (0.968–0.991)
AUC: 0.873 (0.838–0.908)
All Recrudescences
Only High Density Recrudescences


Non-Failing, High Transmission
Sensitivity
Specificity
AUC: 0.957 (0.937–0.977)
AUC: 0.840 (0.806–0.875)
All Recrudescences
Only High Density Recrudescences

2476

**Figure 4.13** Receiver operator characteristic (ROC) curves showing diagnostic ability for the Bayesian analysis method for 6 model scenarios for artemether-lumefantrine (AR-LF) in a very low genetic diversity setting. ROC curves and area under the roc curve (AUC) are shown for all recrudescence and for high density recrudescence, where a high density recrudescence was defined as explained in **4.2.5.**

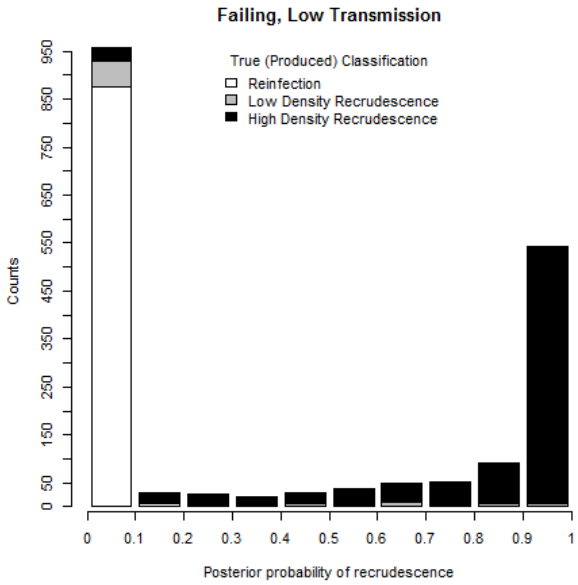
#### 4.3.2.3 Distribution of posterior probability of recrudescence.

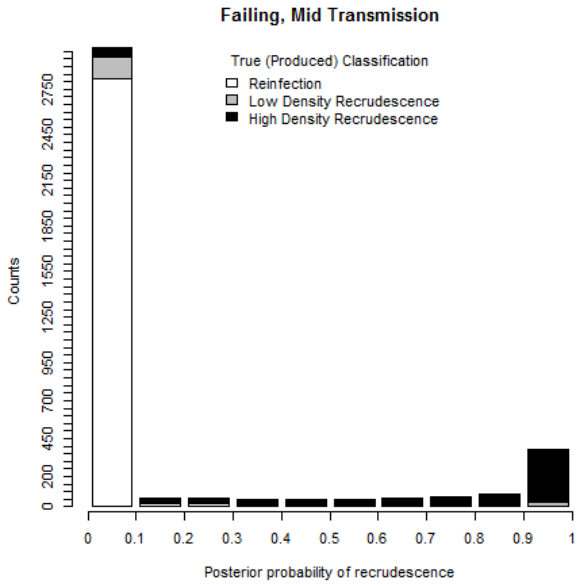
**Figure 4.14** shows the distribution of the posterior probabilities of recrudescence for all recurrences, stratified according to the true classification of their recurrence: Reinfection, low density recrudescence, or high density recrudescence.

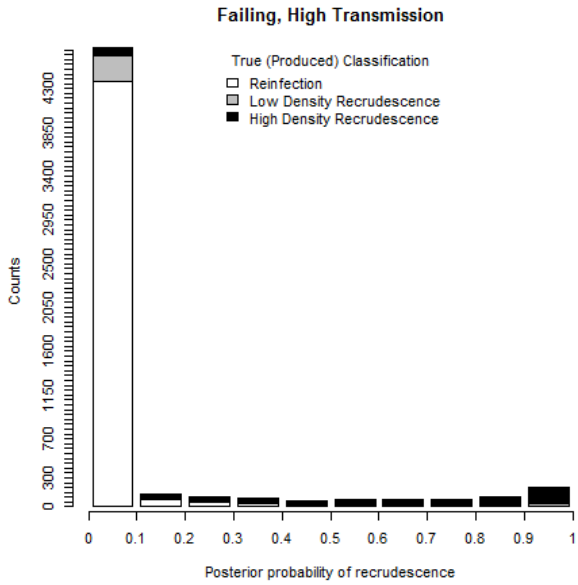
A notable, important difference between the extremely low diversity scenarios and the higher diversity scenarios was that posterior probabilities under the assumption of extremely low genetic diversity were less binary than those for higher diversities (i.e., **Figure 4.5**), where only a small number of patients had $p$ between 0.1 and 0.9 (non-inclusive). By contrast, in the extremely low diversity genetic scenarios, recurrences with a $p$ of <0.1 and ≥0.9 were the largest groups, but larger numbers of recurrences possessed interim values. This distribution explains why the decrease in failure rate estimate shown in

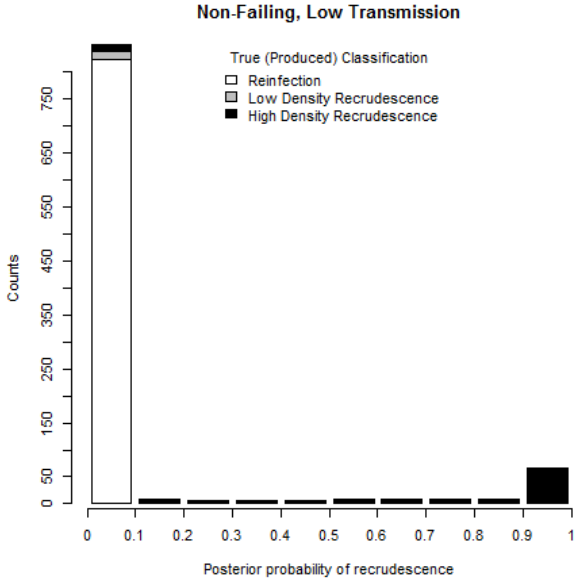
**Figure** 4.11 and **Figure 4.12** as $p$ increased was much sharper than occurred in the higher genetic diversity scenarios (**Figure 4.2** and **Figure 4.3** ) – each unit increase in p moved a greater proportion of recurrences from being classified as a reinfection to being classified as a recrudescence.

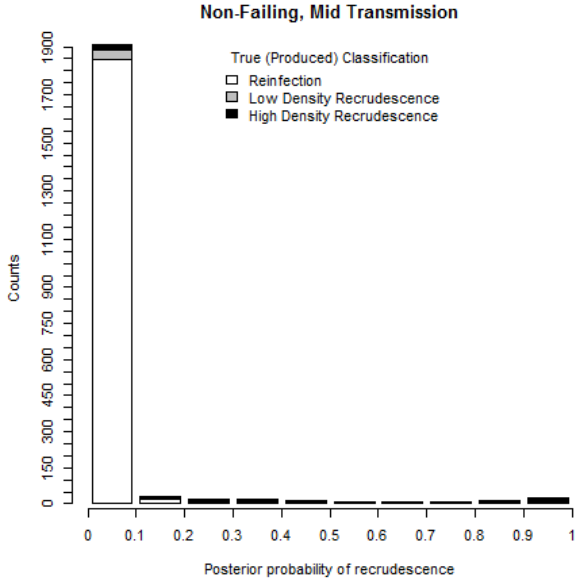
There were a larger number of recurrences that were high density recrudescence with $p$ values of <0.1 in the extremely low genetic diversity scenarios compared to the higher genetic diversity scenarios. Specifically, in the failing drug scenarios: 27, 56 and 81 recurrences for low, medium and high transmission respectively, relative to 5, 9 and 21 recurrences with higher genetic diversity. In the non-failing drug scenarios these figures are 11, 18 and 22 patients relative to 5, 7 and 6 for higher genetic diversity. The distribution of low-density recrudescence also changed – there are comparable numbers of recurrences with $p < 0.1$ in the extremely low genetic diversity scenarios to the higher genetic diversity scenarios, but a reduction in the number of recurrences which have $p \geq 0.9$ and low density recrudescence – the posterior probabilities of these recurrences have shifted to the interim posterior probabilities between 0.1 and 0.9. Both effects occurred because of the increased probability of shared alleles between initial and recurrent samples due to reduced genetic diversity – this can also be observed by the reduced specificity shown in the ROC curves (**Figure 4.13**). Ultimately the increased failure rate observed when classifying a recrudescence as a given value of $p$, but specifically $p \geq 0.1$ was because of the larger numbers of recurrences with $p \geq 0.1$ who are truly reinfections when compared to the higher genetic diversity scenarios. There was a modest increase in the numbers of recurrences that were high density recrudescence which had $p < 0.1$, which would drive an under-estimate of true failure rate, but this was not enough to fully offset the impact of reinfections having increased values of $p$.

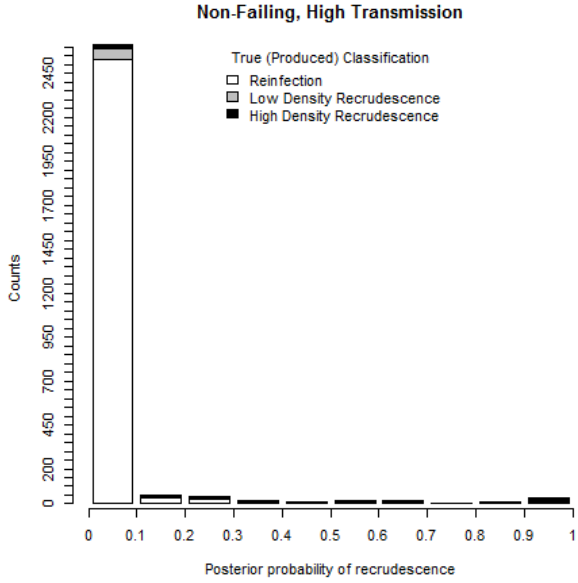

Failing, Low Transmission
True (Produced) Classification
Reinfection
Low Density Recrudescence
High Density Recrudescence
Counts
0 50 150 250 350 450 550 650 750 850 950
0 0.1 0.2 0.3 0.4 0.5 0.6 0.7 0.8 0.9 1
Posterior probability of recrudescence


Failing, Mid Transmission
True (Produced) Classification
Reinfection
Low Density Recrudescence
High Density Recrudescence
Counts
0 200 450 700 950 1250 1550 1850 2150 2450 2750
0 0.1 0.2 0.3 0.4 0.5 0.6 0.7 0.8 0.9 1
Posterior probability of recrudescence


Failing, High Transmission
True (Produced) Classification
Reinfection
Low Density Recrudescence
High Density Recrudescence
Counts
0 300 700 1150 1600 2050 2500 2950 3400 3850 4300
0 0.1 0.2 0.3 0.4 0.5 0.6 0.7 0.8 0.9 1
Posterior probability of recrudescence


Non-Failing, Low Transmission
True (Produced) Classification
Reinfection
Low Density Recrudescence
High Density Recrudescence
Counts
0 50 150 250 350 450 550 650 750
0 0.1 0.2 0.3 0.4 0.5 0.6 0.7 0.8 0.9 1
Posterior probability of recrudescence


Non-Failing, Mid Transmission
True (Produced) Classification
Reinfection
Low Density Recrudescence
High Density Recrudescence
Counts
0 150 300 450 600 750 900 1100 1300 1500 1700 1900
0 0.1 0.2 0.3 0.4 0.5 0.6 0.7 0.8 0.9 1
Posterior probability of recrudescence


Non-Failing, High Transmission
True (Produced) Classification
Reinfection
Low Density Recrudescence
High Density Recrudescence
Counts
0 200 450 700 950 1200 1450 1700 1950 2200 2450
0 0.1 0.2 0.3 0.4 0.5 0.6 0.7 0.8 0.9 1
Posterior probability of recrudescence

2520

**Figure 4.14** Distribution of the posterior probability of recrudescence estimated the by Bayesian analysis algorithm for 6 model scenarios for artemether-lumefantrine (AR-LF) in an extremely low genetic diversity setting. A high density recrudescence was defined as explained in **4.2.**

## 4.4 Discussion.

### 4.4.1 Evaluation of the match counting algorithm.

Despite its wide use, match counting of microsatellites for distinguishing recrudescence from reinfection does not appear to be a robust approach because the estimated drug failure rate is highly dependent on the threshold used to define a recrudescence. The same clone of malaria will have, by definition, the same genotype between the initial and recurrent sample. However, the observed genotype (described by the microsatellite alleles) may differ due to issues inherent in the genotyping method (failure to detect minority alleles or errors in measuring base-pair length of alleles) – accounting for this difference is the purpose of including a degree of flexibility in the molecular correction process i.e., varying thresholds.

The high thresholds generally used to classify a recurrence as a recrudescence in vivo (either most, or all, of the available loci must match to define a recrudescence, see **4.1**) are likely resulting in substantial under-estimate of failure rate. For the in silico failing AR-LF results presented here, failure rate estimates with a threshold of 2 ranged between 15% in a low transmission scenario to 50% in a high transmission scenario, relative to true failure rates of ~10% (**Figure 4.2** ). However, a threshold of 7 provided estimates that ranged between 0.5% and 0.6% relative to true failure rates of ~10% . For non-failing AR-LF (**Figure 4.3** ) failure rate estimates with a threshold of 2 ranged from 7% in a low transmission scenario to 24% in a high transmission scenario, relative to true failure rates of ~2%. In other words, the potential bias induced by choice of a break-point for the match counting algorithm could result in either rejecting an efficacious drug or continuing to use a failing drug and this is further complicated by the sensitivity of the break-point to transmission intensity; it is difficult to recommend, for policy use, a specific number of microsatellites to type or a specific number of matching loci at which to classify a recurrence as recrudescent, because these specifics would have to vary across areas of different transmission intensities. The same issues are present in using the match counting algorithm for AS-MQ (**Figure 4.7** and **Figure 4.8)**.

The results presented here strongly suggest that stringent thresholds (i.e., requiring all or most loci to have matching alleles) will under -estimate failure rate (i.e., over-estimate efficacy). With the seven microsatellites used in these simulations, failure rate estimates produced by the match counting algorithm varied with both the choice of threshold and the transmission intensity but in all scenarios (assuming “normal” levels of genetic diversity) a threshold of 5 matching loci under-estimated failure rate; either 3 or 4 produced the closest estimate (**Figure 4.2** , **Figure 4.3** , **Figure 4.7** and **Figure 4.8**). Note that the threshold producing the most accurate estimate increased from 3 to 4 as transmission increased from low to high – this is because in higher transmission areas there was a greater number of reinfections, and consequently the impact of reinfections being incorrectly classified as recrudescence due to sharing alleles by chance was increased. A threshold of 2 would lead to large over-estimates of failure rate. The reason that stringent thresholds under-estimated failure rate is two-fold:

Firstly, low-density recrudescence can be overlooked in patients who have a polyclonal initial or recurrent infection. This is best observed in the figures showing the distribution of the posterior probabilities of recrudescence; **Figure 4.5** for AR-LF and **Figure 4.10** for AS-MQ. The mechanistic cause of missing low density recrudescence is due to the inability to detect minority alleles in the PCR process, as described in **4.1.2**. Secondly, mis-reading the length of microsatellite alleles (also described in **4.1.2**) both occurs in reality and was included in the model. Mis-reading allelic length by one or more base pairs can result in the conditions for the stringent matching threshold not being met.

### 4.4.2 Evaluation of the Bayesian analysis algorithm.

Application of the Bayesian analysis algorithm produced accurate and stable estimates of failure rate in all transmission scenarios for both failing and non-failing drugs with use of a posterior probability *p* of ≥0.1. This result is consistent for analysis of AR-LF, AS-MQ and even for AR-LF in an extremely low genetic diversity scenarios. Existing in vivo studies that have utilized the Bayesian analysis algorithm used in this research have considered a *p* of ≥0.5 to be indicative of recrudescence [31, 150]. Applying the Bayesian analysis algorithm to simulated data where the true failure rate is known permits calculation of it's sensitivity and specificity (**Figure 4.4** and **Figure 4.9**). This is not possible in vivo, and the results obtained show that the algorithm is highly specific, producing posterior probabilities ≥0.1 for reinfections at a very low frequency (**Figure 4.5** and **Figure 4.10** ). Because of this, and because the results produced by the algorithm are very binary (a result that also occurs when the algorithm has been applied to in vivo data elsewhere, see Figure 1 of [150]), a small increase in accuracy of failure rate estimates produced by this algorithm may be realised by using the lower threshold of *p* ≥0.1.

The slight under-estimate of true failure rate that occurs in all scenarios (apart from the extremely low genetic diversity scenarios) when using the Bayesian analysis algorithm occurs primarily because the algorithm is unable to accurately identify low density recrudescence (**Figure 4.5** and **Figure 4.10** ); this occurred because of the inherent inability to detect minority alleles during the genotyping process of microsatellite markers; nonetheless, the number of recurrences that were, in truth, low density recrudescence was low; most recurrences were either high density recrudescence or reinfection. The diagnostic ability of the Bayesian analysis algorithm for high density recrudescence was extremely high (**Figure 4.4** and **Figure 4.9**) and consequently the algorithm is able to estimate failure rates to a high degree of accuracy.

### 4.4.3 Microsatellite markers in areas of extremely low genetic diversity.

Extremely low diversity has been a key concern in the field due to the increased threat of misclassifying reinfection as recrudescence, and the stringent thresholds that have been used with the match counting algorithm were likely chosen, historically, to help mitigate the frequency of these misclassifications. Extremely low genetic diversity was simulated in low, medium and high transmission scenarios here for AR-LF. A low transmission, extremely low diversity scenario is simply a plausible situation that is likely to occur, while medium and high transmission scenarios with extremely low genetic diversity may reflect phenomena such as genetic bottlenecks.

Under the assumption of extremely low genetic diversity, the match counting algorithm becomes critically compromised. The stringent thresholds (5, 6 and 7 matching loci) still result in under-estimates of the true failure rate, while requiring matches at 3 or 4 loci over-estimates failure rates for failing AR-LF (**Figure 4.12** ). As with areas of higher genetic diversity, failure rate estimates obtained with a given threshold of matches increase as transmission intensity increases. The match counting algorithm is able to estimate an accurate failure rate for non-failing AR-LF at a threshold of 5, 6 or 7 matching loci (**Figure 4.12** )

Crucially, the Bayesian algorithm was able to recover accurate failure rate estimates for failing AR-LF even under the assumption of extremely low genetic diversity (**Figure 4.12** ). The results here show that, while failure rate estimates using the Bayesian analysis algorithm will increase slightly with extremely low genetic diversity, the algorithm is still highly specific (**Figure 4.13**) and sensitive to high-density recrudescence (the majority of recrudescence). This is likely to be a result of using seven microsatellite loci – use of fewer loci would be more compromised by extremely low genetic diversity.

### 4.4.4 Policy recommendations for use of microsatellite markers.

The main practical drawback of the Bayesian algorithm is the need to run a Bayesian analysis. The methodology is published and available [150] but application requires some experience in programming and Bayesian statistics. The analysis is computationally expensive (see **[SI]**) and may be difficult to run on an average personal computer (fully described in **4.2.9**). However, this should not be allowed to be an impediment, given the importance of accurate malaria drug trials, and one solution to this would be for a central body to offer such analyses as a service, or to support application of the algorithm through an internet-based application.

Nonetheless, match counting algorithms will likely continue to be used due to their considerable ease and inertia in the field. When a match counting algorithm for interpreting microsatellite data is used, the results presented here strongly suggest that failure rates obtained with multiple thresholds points are reported, (for example Plucinski et al. reported failure rate estimates based on thresholds of matching at all loci and matching at all except a single locus [150]; their table 2). This reflects the difficulty (or the impossibility) of identifying a robust number of matching loci which to consider a recurrence as a recrudescence, a priori. Additionally, the results presented here suggest that stringent thresholds (requiring all or a very high proportion of loci to be matching) should generally be avoided due to the under-estimate of true failure rate induced by these stringent thresholds.

In the results presented here, a threshold of 4 matching loci to classify a recurrent infection as recrudescent appeared to be a reasonable approach in non-failing drug scenarios (**Figure 4.3** and **Figure 4.8**). In non-failing drug scenarios, most recurrences were likely to be reinfection and 4 appeared to be a sufficient threshold to prevent over-estimation of failure rates due to misclassifying reinfections as recrudescence (such as would be observed using 2 or 3 matching loci as a threshold, particularly for AR-LF). Consequently, perhaps a feasible approach for using microsatellites in TES would be to use the match counting algorithm initially, assess the failure rate estimates produced with a range of thresholds and pass any result that indicates a drug failure rate of higher than 5% through a Bayesian algorithm for re-analysis. In the results presented here, the failure rate estimates produced when varying the threshold number of matching loci are sensitive to transmission intensity, but even in a high transmission intensity, a threshold of 4 would not mistakenly indicate that a failing drug was non-failing (**Figure 4.3** and **Figure 4.8**).

# Chapter 5: Validation of next-generation amplicon sequencing for molecular correction using a computer modelling approach.

Chapter-specific acknowledgements: Dr Katherine Kay, Dr Eva Maria Hodel and Dr Ian Hastings provided R code to generate parasite dynamics post-treatment (fully described in **Chapter 2** of this thesis). Professor Ingrid Felger, Dr Anita Lerch and Maria Gruenberg provided data-sets detailing amplicon allele frequency distributions.

## 5.1 Introduction.

### 5.1.1 Next-generation amplicon sequencing for molecular correction.

Sequencing of length-polymorphic markers (msp-1, msp-2 and glurp; **Chapter 3**) and microsatellite markers (**Chapter 4**) for molecular correction follow the same general laboratory methods, i.e. that blood samples are passed through Polymerase Chain Reaction (PCR), and the products of PCR are sequenced using gel or capillary electrophoresis so the frequency of alleles (identified by their length in base pairs) can be determined [14].

Amplicon deep-sequencing (AmpSeq) is able to quantify, in greater detail, the genetic information present in a given blood sample. This technique has seen notable use in a malaria context: Genotyping infections for the purpose of describing and tracking specific genes (i.e., for resistance) within populations [154-156], and for evaluating efficacy of the novel RTS,S/AS01 vaccine [157]. Using AmpSeq as a method for the molecular correction process in therapeutic efficacy studies (TES) is now being investigated [36, 41, 158]. Amplicon deep-sequencing amplifies a target region of Deoxyribonucleic acid (DNA), creating a large amount of genetic material that can then be sequenced using next generation sequencing (NGS; alternatively known as deep sequencing). The region amplified is the amplicon locus and NGS is able to identify individual variants of that locus through their distinctive genetic sequences. The large amount of genetic material generated through amplification allows the genetic sequence to be read an extremely large number of times, with the consequence that amplicon variants can be fully sequenced and identified with a high degree of accuracy, and the number of "reads" of a specific amplicon variant denotes its frequency in the blood sample. In principle, any short region of DNA can be used as an amplicon locus (for example, merozoite surface protein-1 (*msp-1*) and merozoite surface protein-2 (*msp-2*), typically used in the length-polymorphic marker methodology for genotyping malaria infections [**Chapter 3**] have been characterized using an AmpSeq approach [159]), but it is hypothesized that select loci with high numbers of single nucleotide polymorphisms (SNPs) are the best candidates for effective genotyping of malaria infections [36].

Note that some existing AmpSeq literature has a tendency to refer individual gene variants as "haplotypes" (see [36, 41, 160]), while other AmpSeq literature uses the term haplotype in the more traditional sense i.e. a group of alleles that are inherited together i.e. [157] ; in this thesis the term haplotype is avoided and individual gene variants will be referred to as "alleles" to avoid confusion and for consistency of comparisons with other approaches (i.e., traditional use of length-polymorphic markers and microsatellite markers [**Chapter 3, Chapter 4**]).

A critical proposed advantage of amplicon sequencing in the context of molecular correction to distinguish between recrudescence and reinfection in anti-malarial TES is its ability to detect low frequency alleles, and because the allele is fully sequenced there is no risk of introducing errors by mis-reading allelic length, as can happen with microsatellite markers [122]. Consequently, low-frequency alleles appear to be considerably more detectable using amplicon deep-sequencing than when using traditional methods [36, 158]. Deep-sequencing techniques have been experimentally shown to identify more alleles in a selection of samples than the length-polymorphic markers *msp-1* and *msp-2* [159]. Additionally, minority clones were consistently detected in mixtures using the

amplicon conserved Plasmodium membrane protein (*Cpmp*) at a frequency of >1% of the majority clone, but for the amplicon circumsporozoite surface protein (*csp*) this was accomplished with a frequency of >0.7% of the majority clone [36] –so while the exact ability to detect minority clones does vary between amplicons, it has been experimentally shown that there is consistently a much higher detectability of minority alleles compared to similar experiments involving mixtures of clones and using traditional genotyping of *msp-1*, *msp-2* and *glurp* [36, 106].

### 5.1.2 Sources of misclassification of recurrent infections when genotyping with AmpSeq.

There were three manners of misclassification of recurrent infections with length-polymorphic and microsatellite markers – these may also occur using AmpSeq.

a) Recrudescent infections can be misclassified as reinfection if alleles of the recrudescent clone(s) were not detected when genotyping the initial infection.

b) Recrudescent infections can be misclassified as reinfection if the recurrent infection is mixed (i.e., the recurrence is polyclonal and comprised of both recrudescence and reinfections), and alleles of the recrudescent clone(s) are not detected, but alleles of a reinfecting clone(s) are.

c) A reinfection could be misclassified as recrudescent if it shares (by chance) alleles with clones present at time of treatment.

Misclassification a) and b) above both require an allele that is truly present in a blood sample to not be detected. How could this occur using an AmpSeq methodology? Low frequency false alleles (artefacts) can be generated in the amplicon deep-sequencing process – this is a similar concept to minority "noise" peaks generated in traditional genotyping and is explained in detail elsewhere [36]. Consequently, there must be a user-defined cut-off point for allelic frequency (generally termed in the AmpSeq literature "number of reads") below which an allele would be considered to be an artefact, with the consequence that truly present low frequency alleles will also be considered noise. These false alleles have been observed experimentally at frequencies of up to 0.01% [36] and this cut-off point has been utilized in a longitudinal infection dynamics study [160]. However, for TES, a 1% cut-off is generally used for two purposes: Firstly, malaria gametocytes may persist at low levels, and use of an extremely minor cut-off risks misclassifying gametocyte genetic signals as a recrudescent infection. Secondly, to further mitigate the impact of artefacts (they have not been observed at frequencies of 1% of the majority allele, so a 1% cut-off is more conservative). Development of AmpSeq for genotyping is a very recent development and a limited number of labs have explored the implications of cut-off points, but see [36, 41, 160] for details.

### 5.1.3 Research Goals.

AmpSeq is purported to be a highly accurate means of genotyping infections during TES [41]. However, the true failure rate of a drug cannot be known in vivo and so the quantitative accuracy of AmpSeq used in molecular correction remains unknown. Using a mechanistic pharmacokinetic/pharmacodynamic (mPK/PD) approach (**Chapter 2**), 5,000 patients were modelled in silico following treatment of an uncomplicated malaria infection with either Dihydroartemisinin-Piperaquine (DHA-PPQ) or Artemether-Lumefantrine (AR-LF) in a range of different levels of transmission intensity. These ACTs were chosen to investigate both a drug with a relatively long post-treatment prophylactic period (DHA-PPQ) and one with a relatively short post-treatment prophylactic period (AR-LF). Parasite clones were allocated genetic data at AmpSeq loci. The simulated patient populations were analysed with two goals:

1. Quantify the accuracy of failure rate estimates obtained from a variety of simulated patient populations when infections are genotyped using AmpSeq markers by comparing the estimated failure rates to the true failure rate, which is known from the mPK/PD model.
2. Quantify how accuracy of failure rate estimates changed when fewer AmpSeq markers are genotyped (increasing ease and decreasing cost) and identify the fewest number of AmpSeq markers that should be genotyped while maintaining the accuracy of failure rate estimates.

## 5.2 Methodology.

Parasite dynamics post-treatment with DHA-PPQ and AR-LF were simulated for 5,000 patients using a mPK/PD approach (see **Chapter 2** for full description of the mechanistic simulations for these drugs). DHA-PPQ and AR-LF were the drugs modelled for the purpose of investigating failure rate estimates produced by AmpSeq. All PK/PD parameters (with the exception of partner drug IC50, below) are given in **Table 2.1** and **Table 2.4**.

### 5.2.1 Partner Drug choice and IC50.

Failing DHA-PPQ and AR-LF were simulated. The IC50 values used for the partner drugs in this chapter were identical to those used for failing PPQ and LF in **Chapter 3** and are shown in **Table 5.1**. As previously described in **Chapter 2 and Chapter 3**, failing LF was simulated by arbitrarily increasing its IC50 until simulated true failure was ~10%. Only a two-compartmental model of PPQ was used in this chapter, a three-compartmental model (i.e., as described in **Table 2.3**) was not considered for the reasons provided in **2.1.1.**

**Table 5.1** Mean values of the half-maximal inhibitory concentration (IC50) for each calibration of two partner drugs (PPQ and LF) used within this chapter.

| Partner Drug | Mean IC50 (mg/L) | Literature Justification |
|---|---|---|
| Failing PPQ | 0.02 (0.3) | [20] |
| Failing LF | 10 (1.02) | N/A |

PPQ: Piperaquine, LF: Lumefantrine, IC50: Half-maximal inhibitory concentration. Coefficient of variation (CV) is given in brackets. This table should be considered with **Table 2.1** and **Table 2.4** for a full set of PK/PD parameters for DHA-PPQ and AR-LF.

### 5.2.2 Multiplicity of Infection (MOI), Force of Infection (FOI), and initial parasite number.

In this chapter, two Multiplicity of Infection (MOI) distributions were modelled, from which the number of clones in a given patient's initial infection were drawn. A "high MOI" was representative of the MOI in an area of intense transmission, in this case Tanzania where MOIs of 1-8 were assigned with probabilities 0.036, 0.402, 0.110, 0.110, 0.183, 0.049, 0.061, 0.049 respectively [14]. A "low MOI" distribution was based on data from Papua New Guinea with probabilities of 0.460, 0.370, 0.150 and 0.020 for an MOI of 1-4 respectively [15]; these two distributions were subsequently used to check if the accuracy of different algorithms were consistent across different MOIs.

These MOI distributions are identical to those used for simulation of length-polymorphic markers described in **2.3.1**. Note that the MOI distributions were derived, in the first instance, from in vivo data using length-polymorphic markers, not AmpSeq. Given that AmpSeq provides a higher detectability of minority clones [158, 160], MOI estimates from a given population with AmpSeq should be higher than MOI estimates with length-polymorphic markers. However, use of AmpSeq for

this purpose is extremely novel, and useable MOI distributions obtained with are limited. To the best of my knowledge, the only available MOI distributions obtained using AmpSeq are across multiple countries or study sites [41, 157] and would not be appropriate to use the source for a population MOI distribution for this mPK/PD approach. The purpose of simulating different MOIs was to investigate whether qualitative conclusions regarding the accuracy of failure rate estimates obtained using AmpSeq were consistent in different MOI (i.e., different endemicity) settings and so use of a low and high MOI distribution obtained using length-polymorphic markers was appropriate, even though MOI estimates from those same populations obtained with AmpSeq would result in different distributions (presumably a higher number of clones on average).

The Force of Infection (FOI) values used to calibrate the model in this chapter were 0, 2, 8 and 16, broadly representing an area with no, low, medium and high ongoing transmission, respectively. Intermediate values were not modelled as previous work on length-polymorphic markers (**Chapter 3**) showed the relationship between failure rate estimates and FOI to be relatively linear, so an approach of modelling four specific values to represent scenarios of transmission intensity (similar to the approach taken in the microsatellite work, **Chapter 4**) was chosen to reduce computational time compared to investigating the whole range of FOI values.

In short, 8 different scenarios were simulated for treatment with each of DHA-PPQ and AR-LF: Two MOI distributions with four FOI values each. Sensitivity analysis was later conducted by varying select model parameters and re-simulating these scenarios.

Each clone within the MOI (later called "initial clones") had their starting parasitaemia drawn from a log-uniform distribution. This log-uniform distribution was varied between two ranges within this chapter. Firstly, a log-uniform distribution spanning from $10^{10}$ to $10^{11}$ asexual parasites per person was used (**2.3.1**). Previous modelling approaches [15] used $10^{12}$ parasites as the upper limit of parasitaemia because this level of parasitaemia is likely to be lethal or at least be a parasite density sufficiently high that such patients would not be enrolled in a clinical trial; hence $10^{11}$ was used as the upper limit for any single clone at the time of treatment (see **2.3.1** for full discussion).

Secondly, a wider log-uniform distribution of $10^{8}$ to $10^{11}$ parasites was investigated. This wider distribution was investigated only for AmpSeq (and not for length-polymorphic markers (**Chapter 3**) or microsatellite markers (**Chapter 4**)) due to the theorized advantage of AmpSeq being able to detect minority alleles down to 1% of the majority allele. For other markers, where this minority detection threshold is ~25% (though see additional values investigated for length-polymorphic markers in **3.3.8**), varying this initial parasite distribution would have little impact on results compared to the large number of alleles that would be missed due to the minority detection threshold (i.e., the increased number of alleles that would be missed by creating a wider distribution would be negligible). However, because of the higher detectability of minority alleles afforded by AmpSeq, it was theoretically possible that the lower range of initial parasitaemia could impact results and it was important to check that it does not.

### 5.2.3 Genetic Data of Malaria Clones.

Genetic information was assigned to each parasite clone in a patient at five AmpSeq loci: conserved plasmodium protein (*cpp*), conserved plasmodium membrane protein (*cpmp*), circumsporozoite surface protein (*csp*), apical membrane antigen (ama1-D3) and merozoite surface protein 7 (*msp-7*).

Details of the identification, sequencing and additional information for these loci can be found in [36, 41, 160]. For the simulations herein, I was provided with frequency distributions for alleles at each of these loci by Maria Gruenberg, Anita Lerch and Ingrid Felger. The data were drawn from a worldwide mix of samples, fully described across [36, 41, 147, 160] . Each allele was uniquely by genetic variation in their sequence (note that this is as opposed to length-polymorphic markers and microsatellite

markers which are identified by their length (and, for msp-1 and msp-2, family). The expected heterozygosity (He) of each loci in these distributions were: 0.975 for cpp, 0.982 for cpmp, 0.949 for csp, 0.966 for ama1-D3 and 0.899 for msp-7.

To investigate the accuracy of AmpSeq molecular correction when genotyping fewer loci (a strategy designed to reduce the complexity and cost of conducting AmpSeq), the total number of loci investigated was 5 (the maximum), 4 and 3. To simulate genotyping of 4 total loci, the least diverse locus (msp-7) was excluded, and to simulate genotyping 3 total loci, the second least diverse locus (csp) was also excluded – this is a consistent approach with laboratory experiments that used 4 loci (excluding msp-7) for genotyping, and only using csp to adjudicate in the event that any of the other three loci failed to amplify [41].

**Table 5.2** Allele frequency distribution for 5 AmpSeq loci simulated: Ama1-D3, cpmp, cpp, csp, msp-7.

| Ama1-D3 | | cpmp | | cpp | | csp | | msp-7 | |
|---|---|---|---|---|---|---|---|---|---|
| Allele name | Frequency | Allele name | Frequency | Allele name | Frequency | Allele name | Frequency | Allele name | Frequency |
| ama1-D3-1 | 4.39 | cpmp-1 | 0.83 | cpp-1 | 0.83 | csp-1 | 1.89 | msp7-1 | 23.33 |
| ama1-D3-10 | 0.88 | cpmp-10 | 0.83 | cpp-10 | 7.44 | csp-10 | 0.94 | msp7-10 | 1.11 |
| ama1-D3-11 | 6.14 | cpmp-11 | 2.50 | cpp-11 | 0.83 | csp-12 | 5.66 | msp7-11 | 1.11 |
| ama1-D3-13 | 1.75 | cpmp-117 | 0.83 | cpp-12 | 0.83 | csp-13 | 10.38 | msp7-12 | 2.22 |
| ama1-D3-14 | 1.75 | cpmp-118 | 0.83 | cpp-13 | 3.31 | csp-14 | 0.94 | msp7-14 | 2.22 |
| ama1-D3-15 | 2.63 | cpmp-119 | 0.83 | cpp-14 | 1.65 | csp-16 | 2.83 | msp7-15 | 3.33 |
| ama1-D3-16 | 7.89 | cpmp-12 | 3.33 | cpp-15 | 1.65 | csp-17 | 3.77 | msp7-16 | 1.11 |
| ama1-D3-2 | 0.88 | cpmp-123 | 0.83 | cpp-16 | 0.83 | csp-18 | 1.89 | msp7-17 | 5.56 |
| ama1-D3-20 | 0.88 | cpmp-127 | 0.83 | cpp-17 | 2.48 | csp-19 | 1.89 | msp7-19 | 1.11 |
| ama1-D3-21 | 0.88 | cpmp-128 | 0.83 | cpp-18 | 0.83 | csp-2 | 8.49 | msp7-2 | 3.33 |
| ama1-D3-22 | 2.63 | cpmp-129 | 0.83 | cpp-19 | 0.83 | csp-20 | 2.83 | msp7-20 | 1.11 |
| ama1-D3-23 | 4.39 | cpmp-13 | 2.50 | cpp-2 | 1.65 | csp-21 | 1.89 | msp7-21 | 1.11 |
| ama1-D3-24 | 4.39 | cpmp-130 | 0.83 | cpp-20 | 0.83 | csp-24 | 1.89 | msp7-25 | 1.11 |
| ama1-D3-25 | 0.88 | cpmp-14 | 1.67 | cpp-23 | 0.83 | csp-25 | 0.94 | msp7-26 | 1.11 |
| ama1-D3-26 | 2.63 | cpmp-15 | 0.83 | cpp-24 | 1.65 | csp-29 | 5.66 | msp7-28 | 1.11 |
| ama1-D3-27 | 0.88 | cpmp-16 | 0.83 | cpp-26 | 4.13 | csp-3 | 3.77 | msp7-29 | 1.11 |
| ama1-D3-28 | 2.63 | cpmp-17 | 1.67 | cpp-27 | 0.83 | csp-30 | 0.94 | msp7-3 | 10.00 |
| ama1-D3-29 | 1.75 | cpmp-174 | 0.83 | cpp-28 | 1.65 | csp-31 | 1.89 | msp7-30 | 3.33 |
| ama1-D3-3 | 0.88 | cpmp-18 | 1.67 | cpp-29 | 2.48 | csp-32 | 0.94 | msp7-31 | 6.67 |
| ama1-D3-31 | 1.75 | cpmp-2 | 4.17 | cpp-30 | 1.65 | csp-33 | 0.94 | msp7-33 | 1.11 |
| ama1-D3-32 | 4.39 | cpmp-21 | 0.83 | cpp-31 | 3.31 | csp-34 | 0.94 | msp7-35 | 1.11 |
| ama1-D3-33 | 0.88 | cpmp-24 | 1.67 | cpp-32 | 2.48 | csp-35 | 2.83 | msp7-36 | 1.11 |
| ama1-D3-35 | 3.51 | cpmp-25 | 0.83 | cpp-33 | 0.83 | csp-37 | 2.83 | msp7-37 | 1.11 |
| ama1-D3-36 | 0.88 | cpmp-26 | 0.83 | cpp-36 | 0.83 | csp-38 | 0.94 | msp7-38 | 1.11 |
| ama1-D3-37 | 1.75 | cpmp-27 | 1.67 | cpp-37 | 2.48 | csp-39 | 0.94 | msp7-39 | 1.11 |
| ama1-D3-38 | 0.88 | cpmp-28 | 2.50 | cpp-38 | 0.83 | csp-4 | 8.49 | msp7-4 | 14.44 |
| ama1-D3-39 | 0.88 | cpmp-3 | 0.83 | cpp-39 | 0.83 | csp-40 | 1.89 | msp7-45 | 1.11 |
| ama1-D3-4 | 4.39 | cpmp-30 | 3.33 | cpp-4 | 0.83 | csp-44 | 0.94 | msp7-5 | 2.22 |
| ama1-D3-40 | 0.88 | cpmp-31 | 1.67 | cpp-40 | 0.83 | csp-45 | 0.94 | msp7-6 | 1.11 |
| ama1-D3-42 | 4.39 | cpmp-32 | 0.83 | cpp-41 | 1.65 | csp-5 | 5.66 | msp7-7 | 2.22 |
| ama1-D3-43 | 0.88 | cpmp-33 | 0.83 | cpp-42 | 0.83 | csp-6 | 0.94 | msp7-9 | 1.11 |
| ama1-D3-44 | 0.88 | cpmp-34 | 1.67 | cpp-43 | 2.48 | csp-7 | 3.77 | | |
| ama1-D3-45 | 0.88 | cpmp-35 | 0.83 | cpp-44 | 0.83 | csp-8 | 6.60 | | |
| ama1-D3-46 | 1.75 | cpmp-36 | 0.83 | cpp-45 | 0.83 | csp-9 | 1.89 | | |

| | |
|---|---|
| ama1-D3-48 | 0.88 |
| ama1-D3-49 | 0.88 |
| ama1-D3-5 | 1.75 |
| ama1-D3-50 | 0.88 |
| ama1-D3-51 | 0.88 |
| ama1-D3-52 | 0.88 |
| ama1-D3-54 | 0.88 |
| ama1-D3-55 | 1.75 |
| ama1-D3-59 | 0.88 |
| ama1-D3-60 | 0.88 |
| ama1-D3-61 | 0.88 |
| ama1-D3-62 | 0.88 |
| ama1-D3-64 | 0.88 |
| ama1-D3-65 | 0.88 |
| ama1-D3-66 | 0.88 |
| ama1-D3-8 | 5.26 |
| ama1-D3-9 | 1.75 |

| | |
|---|---|
| cpmp-4 | 0.83 |
| cpmp-41 | 0.83 |
| cpmp-43 | 0.83 |
| cpmp-44 | 1.67 |
| cpmp-45 | 0.83 |
| cpmp-46 | 1.67 |
| cpmp-48 | 0.83 |
| cpmp-49 | 0.83 |
| cpmp-5 | 0.83 |
| cpmp-50 | 1.67 |
| cpmp-51 | 0.83 |
| cpmp-52 | 1.67 |
| cpmp-53 | 1.67 |
| cpmp-54 | 0.83 |
| cpmp-55 | 0.83 |
| cpmp-56 | 0.83 |
| cpmp-58 | 0.83 |
| cpmp-59 | 2.50 |
| cpmp-6 | 3.33 |
| cpmp-60 | 2.50 |
| cpmp-62 | 1.67 |
| cpmp-63 | 0.83 |
| cpmp-66 | 0.83 |
| cpmp-7 | 3.33 |
| cpmp-70 | 0.83 |
| cpmp-71 | 0.83 |
| cpmp-72 | 0.83 |
| cpmp-73 | 0.83 |
| cpmp-75 | 0.83 |
| cpmp-76 | 0.83 |
| cpmp-77 | 0.83 |
| cpmp-79 | 0.83 |
| cpmp-80 | 1.67 |
| cpmp-81 | 0.83 |
| cpmp-83 | 0.83 |
| cpmp-84 | 0.83 |
| cpmp-85 | 0.83 |
| cpmp-86 | 1.67 |
| cpmp-87 | 0.83 |

| | |
|---|---|
| cpp-48 | 2.48 |
| cpp-49 | 0.83 |
| cpp-50 | 1.65 |
| cpp-51 | 2.48 |
| cpp-52 | 0.83 |
| cpp-53 | 0.83 |
| cpp-54 | 0.83 |
| cpp-56 | 0.83 |
| cpp-57 | 1.65 |
| cpp-59 | 0.83 |
| cpp-6 | 1.65 |
| cpp-60 | 3.31 |
| cpp-62 | 0.83 |
| cpp-64 | 0.83 |
| cpp-66 | 0.83 |
| cpp-68 | 0.83 |
| cpp-69 | 0.83 |
| cpp-7 | 3.31 |
| cpp-70 | 1.65 |
| cpp-71 | 1.65 |
| cpp-78 | 1.65 |
| cpp-8 | 3.31 |
| cpp-80 | 1.65 |
| cpp-84 | 0.83 |
| cpp-85 | 0.83 |
| cpp-9 | 3.31 |
| cpp-93 | 0.83 |
| cpp-96 | 0.83 |
| cpp-99 | 0.83 |

| | |
|---|---|
| cpmp-9 | 1.67 |
| cpmp-95 | 1.67 |
| cpmp-99 | 1.67 |

*Ama1-D3: apical membrane antigen, cpmp: conserved plasmodium membrane protein, cpp: conserved plasmodium protein, csp: circumsporozoite surface protein, msp-7: merozoite surface protein 7. Data is presented in tabulated form despite it's length to improve clarity, noting the high number of low-frequency alleles at loci Ama1-D3, cpmp and cpp.*

### 5.2.4 Follow-up length and detection of recurrence.

In this chapter, a 28 day follow-up period was used for models of AR-LF and a 42 day follow-up period was used for models of DHA-PPQ, as permitted by WHO guidelines [14]. Unlike in **Chapter 3**, novel lengths of follow-up were not explored for microsatellite markers. A 28-day follow-up schedule required patients be examined on days 3, 7, 14, 21 and 28. A 42-day follow-up period used two additional days i.e. days 35 and 42. The parasitaemia of each clone in each patient was tracked and updated each day as described by the mPK/PD model and the PK parameters of the patient and the PD parameters of the clone (**Chapter 2**).

The model checked each day of scheduled follow-up to determine whether a patient had enough parasitaemia that a recurrence would be detectable by light microscopy (a recurrence) – recurrent parasitaemia was considered detectable if the total number in a patient was $\geq 10^8$ on that day. Note that variance in the limit of detection by light microscopy exists with respect to the skill of the microscopist [18]; it was assumed this limit was reflective of an "expert" microscopist (corresponding to roughly 20 parasites /µl of blood). In short, follow-up length and detection of recurrence proceeded in an identical manner to as described for length-polymorphic markers and microsatellite markers (**3.2.4**; **4.2.3**).

### 5.2.5 Calculating which alleles are observed.

The genotype of the initial malaria infection of each patient was calculated on the day of treatment. This genotype signal is a composite of all the clone(s) present in the initial infection and is determined by the technical accuracy and sensitivity of genotyping. This is broadly similar to the method described in **3.2.5** for length-polymorphic markers, though note differences in the minority allele detection threshold, and that there is no differential detectability of alleles based on length or allelic families for AmpSeq markers.

On all days of follow-up except day 3, a recurrence was identified if the sum parasitaemia of all clones in a patient exceeded $10^8$ which was assumed to be the minimum parasitaemia at which detection by light microscopy was possible [152]. This corresponded to a parasite density of roughly 20 parasites/µl of blood. If total parasitaemia was less than $10^8$ then recurrent parasites would not be observed by microscopy (and thus, the patient would not be genotyped on that day). On day 3, if total parasitaemia exceeded $10^8$ but was <25% of the total parasitaemia on the initial sample, the patient continued in the trial; if parasites were present at >25% of initial parasitaemia, that patient was classed as an early treatment failure, per WHO procedure [28]. Note that for subsequent calculations and analysis, an early treatment failure is considered to be identical to a recrudescence. Calculations then occurred using a three-step process to replicate the technical limitations of acquiring a profile of microsatellite alleles from a blood sample

Firstly, a "sampling" limit was included: A finite volume of blood is available for genotyping. A parasite clone would not be detected if its density were so low that no parasites were included in the blood sample analysed. Thus, the density and volume of the processed blood sample defined the limit of detection. Two assumptions were made for this value.

Firstly, it was assumed this limit was $10^8$ (i.e., no clone present in less than $10^8$ parasites would be detected). This limit is identical to that used in **Chapter 3**, see **3.2.5** for calculations and justification.

Secondly, the limit was varied to the lower limit of $10^7$. The sampling limit was based on in vivo considerations on whether parasites would enter genotyping using finger-prick blood samples stored on sample paper (**3.2.5**) but does still include simplifications and assumptions. This was deemed not to be an issue for investigation of length-polymorphic and microsatellite markers as the large minority detection thresholds would far overweight the impact of small changes in the sampling limit, but

under the increased resolution of the AmpSeq methodology, the assumed value of sampling limit was theorized to be more important, and so the impact of varying it was investigated here.

A "minority allele detection threshold" was included in a similar principle to that included for length-polymorphic markers and microsatellite markers. The "majority allele" in a sample is the most frequent allele. The threshold at which minority alleles were ignored as artefacts or gametocytes in the AmpSeq process was varied herein between 0%, 1% and 2%, noting that 1% is reflective of the value used experimentally and 0% would mean perfect detection of all alleles, providing their frequency met the sampling limit. Note that this modelling methodology does not attempt to include false alleles generated as artefacts or genetic signals originating from gametocytes, and 0% is modelled simply to explore the difference in failure rate estimates between this hypothetical perfect detection limit and other values.

#### 5.2.6 Classifying patients: True failures and failure rate estimates with a match counting algorithm.

The term "true failure" is consistent with the mathematical description given in **3.2.7** but is re-summarised here for ease of reading: It was determined whether each patient was a "true failure" based on parasitaemia: A patient was a true failure if, on the final day of follow-up (day 28 for AR-LF, day 42 for DHA-PPQ), they still harboured any parasites from any initial clone. The true failure rate is the frequency of these patients across the entire population. The model tracked patients over the full length of follow-up, thus this "true failure" classification captured patients who would in vivo, have been removed earlier in the TES with a recurrent infection classified as a reinfection (and whose recrudescent clones would not then be observed).

To calculate estimated failure rates using AmpSeq, a "match-counting" algorithm was employed in a similar manner to the match-counting algorithm described for microsatellite markers in **4.2.6.** This is a counting algorithm where a recurrent infection is defined as a recrudescence when the number of AmpSeq loci which share at least one allele between the initial and recurrent sample (i.e., a "matching" loci or a "match") is greater than or equal to a specified threshold. As five AmpSeq loci were simulated herein, the threshold was varied between ≥1 and =5 matching loci. For simulations that used a lower total number of AmpSeq loci (i.e., 3 and 4), the threshold was varied between ≥1 and =3 and ≥1 and =4 respectively. Under each threshold, a failure rate estimate was then calculated for the patient population using survival analysis as described in **3.2.7.**

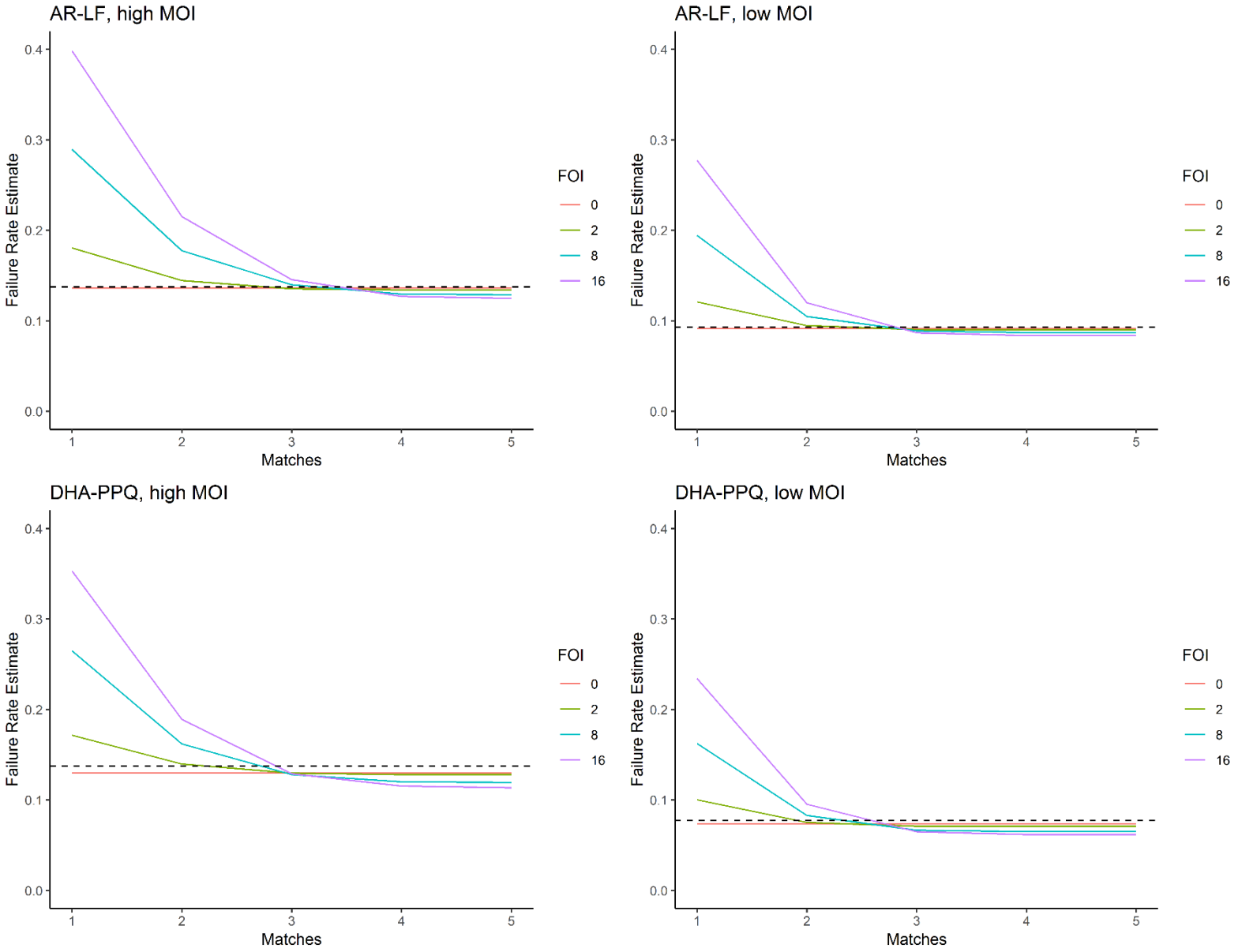
## 5.3 Results.

### 5.3.1 Failure rate estimates using AmpSeq for DHA-PPQ and AR-LF: 5 AmpSeq loci.

Failure rate estimates were obtained under the "baseline" scenario for DHA-PPQ and AR-LF as described above (i.e., minority allele detection threshold of 1%, a "sampling limit" of $10^8$ parasites (**5.2.5**) and initial parasite number drawn from a log-uniform distribution between $10^{10}$ and $10^{11}$ (**5.2.2**). The true failure rate for DHA-PPQ was 0.138 and 0.0774 in areas of high and low MOI respectively. The true failure rate for AR-LF was 0.1376 and 0.0928 in areas of high and low MOI respectively. Note that, as would be expected, the true failure rate is higher in the high MOI area because the IC50 value of the partner drug was not changed between MOI settings (see **5.2.1**), and so an area of higher MOI (and thus, on average a larger number of initial clones per patients) gives a greater chance of a given patient having a recrudescence during follow-up.

Patients were classified as a recrudescence if the number of AmpSeq loci with shared alleles between the initial and recurrent samples ("matches") was equal to or greater than a given threshold, which varied between 1 and 5 (**5.2.5**). Failure rate estimates were then calculated using survival analysis, per

WHO procedure [28], and are shown in **Figure 5.1**. Several trends were clear for both drugs, both MOI settings and all FOI values: the failure rate estimates obtained using ≥3, ≥4 and =5 matching loci to classify a recrudescence were very close in value and very close to the true failure rate (in all instances, classifying a recrudescence at ≥3 matches produced the failure rate estimate closest to the true failure rate). Classifying a recrudescence at ≥1 or ≥2 matches lead to over-estimation of true failure rate with any non-zero value of FOI. When FOI = 0, failure rate estimates did not change as the required number of matches changed (there were no reinfections, so every recurrence was a recrudescence). For both DHA-PPQ and AR-LF at both high and low MOI, failure rate was slightly under-estimated using ≥4 or ≥5 matches to classify a recrudescence; this under-estimate was higher with higher MOI. In summary: Classifying a recurrence as a recrudescence at ≥ 3 matches produced accurate failure rate estimates for both DHA-PPQ and AR-LF in all MOI and FOI settings modelled.


AR-LF, high MOI
AR-LF, low MOI
DHA-PPQ, high MOI
DHA-PPQ, low MOI
Failure Rate Estimate
Matches
FOI
0
2
8
16
0.0
0.1
0.2
0.3
0.4
1
2
3
4
5

2952

**Figure 5.1** Failure rate estimates obtained using AmpSeq for Di-hydroartemisinin-Piperaquine (DHA-PPQ) and Artemether-Lumefantrine (AR-LF) in low and high Multiplicity of Infection (MOI) settings with a range of Force of Infection (FOI) values. Failure rate estimates were obtained using a given number of matching loci between the initial and recurrent sample to classify a recrudescence (x axis). The true failure rate is marked by the horizontal dashed black line.

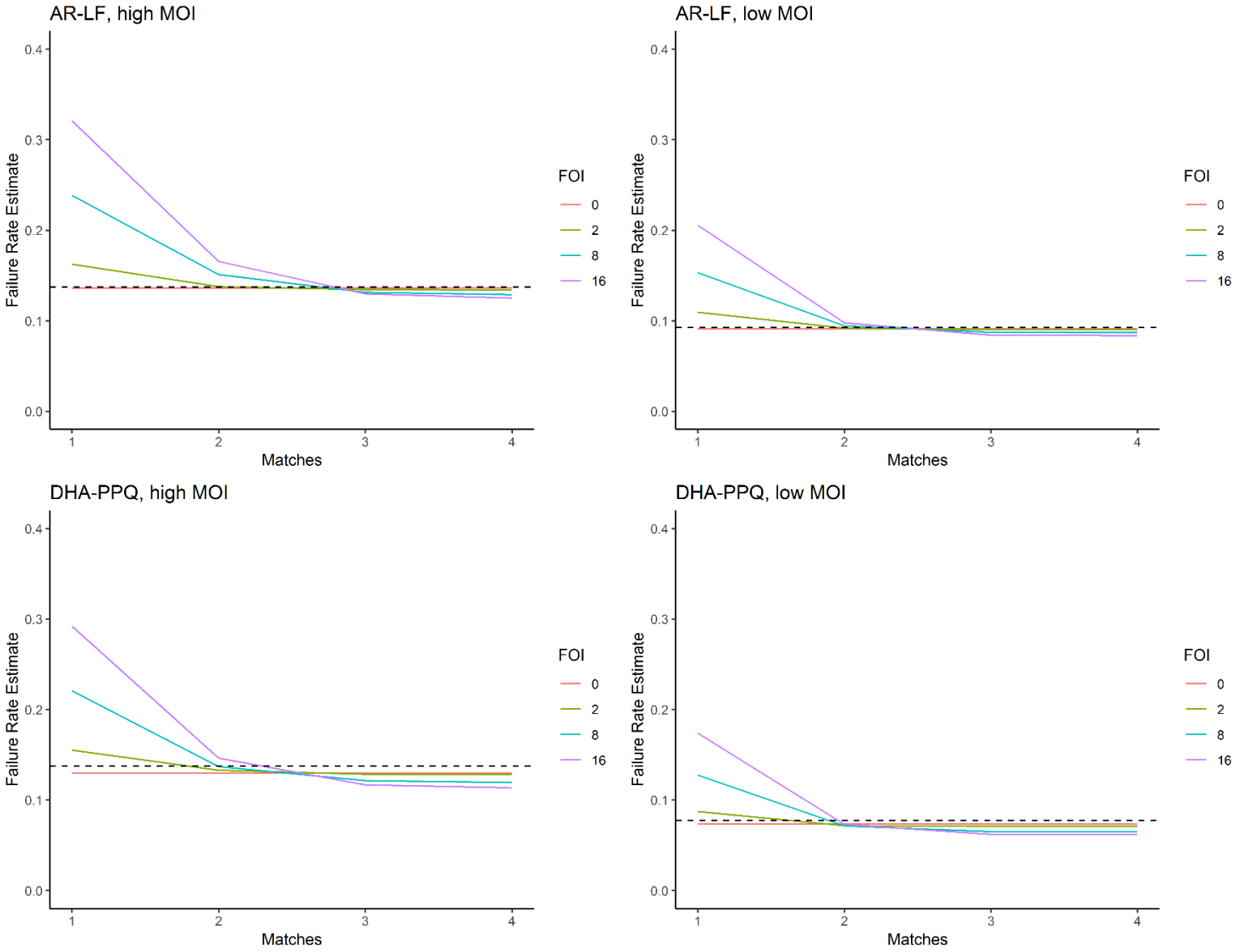
### 5.3.2 Failure rate estimates using AmpSeq for DHA-PPQ and AR-LF: 3 and 4 AmpSeq loci.

The simulations used to generate **Figure 5.1** were repeated, but the total number of AmpSeq loci genotyped was reduced to 4, then again to 3, excluding *msp-7* first then excluding *csp* (see **5.2.3 and 5.2.5).** Results are displayed in **Figure 5.2** and **Figure 5.3** respectively. True failure rates were identical to those in **Figure 5.1** and described above – the same simulated patients are analysed, just with fewer total loci genotyped.

For both drugs and MOI settings, failure rate estimates with an FOI of 0 (i.e., representing an area with zero ongoing transmission) did not change as the number of loci genotyped was reduced. For all other FOI values, failure rate estimates fell as the total number of AmpSeq loci genotyped fell, provided that the threshold number of matches required to classify a recurrent infection as a recrudescence remained the same. Note that, when requiring a match at ≥4 loci to classify a recrudescence, the fall in the failure rate estimate going from 5 to 4 total loci was negligible. However, the reduction in failure rate estimate using when using ≥1 match to classify a recrudescence was significant. In other words, the decrease in failure rate estimates that occurred as the total number of AmpSeq loci genotyped fell increased in significance at a lower number of matches required to classify a recrudescence. However, classifying a recrudescence at ≥1 or ≥2 loci still over-estimated true failure rate at all except the lowest (0-2) FOI. To improve clarity, these results have also been tabulated in **Table 5.3** for DHA-PPQ and **Table 5.4** for AR-LF.

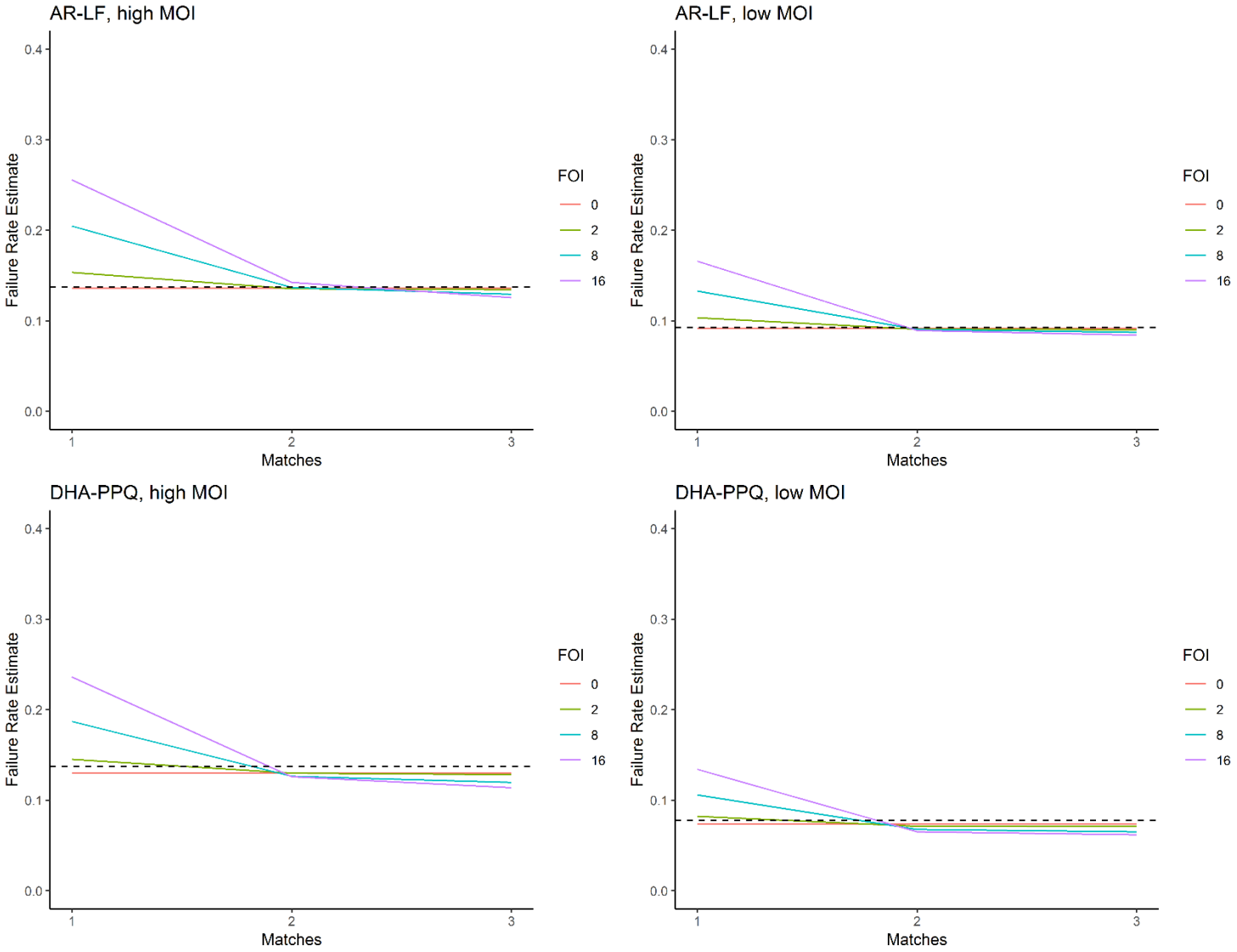
Genotyping 3 or 4 rather than 5 total AmpSeq loci still produced highly accurate failure rate estimates in all MOI and FOI settings (within 1% of true failure rate when recrudescence is classified at an appropriate number of matching loci; **Table 5.3**, **Table 5.4**). When genotyping 4 loci, classifying a recurrence as a recrudescence with ≥2 matches produced accurate failure rate estimates in low MOI areas for both DHA-PPQ and AR-LF, at all values of FOI (though there was a slight over-estimate of failure rate for AR-LF with an FOI of 16). This was not true when using 4 total loci in high MOI areas, as classifying a recrudescence at ≥2 matches caused over-estimation of failure rates at higher FOI, more substantially for AR-LF than DHA-PPQ (Figure 5.2). In the high MOI simulations when genotyping 4 total loci, classifying a recrudescence at ≥3 matches produced generally accurate failure rate estimates. The difference in estimates between classifying a recrudescence at ≥2 or ≥3 matches was only notable at FOI 16 (3% for DHA-PPQ, 4% for AR-LF). The difference between classifying a recrudescence at ≥3 or ≥4 matches was never more than 1%, but as both ≥3 and ≥4 matches always slightly under-estimated the true failure rate, classifying a recrudescence at ≥3 matches was always more accurate. It was dependant on FOI whether ≥2 or ≥3 matches were more accurate; at higher FOIs ≥2 matches tended to over-estimate. ≥3 matches would slightly under-estimate at all FOIs.

When genotyping 3 total loci, classifying a recurrence as a recrudescence with ≥2 matches produced accurate failure rate estimates for both drugs in both low and high MOI areas and at all FOI values (Figure 5.3). Notably, the changes in failure rate estimates as FOI increased were very small (<1% as FOI changed from 0 to 16) in all simulations when genotyping 3 total loci and classifying a recrudescence at ≥2 matches, so genotyping a smaller number of loci appeared to still be robust against changes in transmission intensity.


AR-LF, high MOI
AR-LF, low MOI
DHA-PPQ, high MOI
DHA-PPQ, low MOI
Failure Rate Estimate
Matches
FOI
0
2
8
16
0.0
0.1
0.2
0.3
0.4
1
2
3
4

2995

**Figure 5.2** Failure rate estimates obtained using 4 Amplicon Sequencing (AmpSeq) loci for Di-hydroartemisinin-Piperaquine (DHA-PPQ) and Artemether-Lumefantrine (AR-LF) in low and high Multiplicity of Infection (MOI) settings with a range of Force of Infection (FOI) values. Failure rate estimates were obtained using a given number of matching loci between the initial and recurrent sample to classify a recrudescence (x axis). The true failure rate is marked by the horizontal dashed black line.


AR-LF, high MOI
AR-LF, low MOI
DHA-PPQ, high MOI
DHA-PPQ, low MOI
Failure Rate Estimate
Matches
FOI
0
2
8
16
0.0
0.1
0.2
0.3
0.4
1
2
3

3001

**Figure 5.3** Failure rate estimates obtained using 3 Amplicon Sequencing (AmpSeq) loci for Di-hydroartemisinin-Piperaquine (DHA-PPQ) and Artemether-Lumefantrine (AR-LF) in low and high Multiplicity of Infection (MOI) settings with a range of Force of Infection (FOI) values. Failure rate estimates were obtained using a given number of matching loci between the initial and recurrent sample to classify a recrudescence (x axis). The true failure rate is marked by the horizontal dashed black line.

**Table 5.3** Failure rate estimates obtained for modelled scenarios of DHA-PPQ when genotyping 5, 4 and 3 AmpSeq loci, to two significant figures.

| Drug | DHA-PPQ | | | | | | | | | | | | | | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| MOI | High | | | | | | | | | | | | Low | | | | | | | | | | | |
| FOI | 0 | | | 2 | | | 8 | | | 16 | | | 0 | | | 2 | | | 8 | | | 16 | | |
| AmpSeq loci genotyped | 5 | 4 | 3 | 5 | 4 | 3 | 5 | 4 | 3 | 5 | 4 | 3 | 5 | 4 | 3 | 5 | 4 | 3 | 5 | 4 | 3 | 5 | 4 | 3 |
| Failure rate estimate: 1 match | 0.13 | 0.13 | 0.13 | 0.17 | 0.16 | 0.15 | 0.27 | 0.22 | 0.19 | 0.35 | 0.29 | 0.24 | 0.07 | 0.07 | 0.07 | 0.10 | 0.09 | 0.08 | 0.16 | 0.13 | 0.11 | 0.23 | 0.17 | 0.13 |
| Failure rate estimate: 2 match | 0.13 | 0.13 | 0.13 | 0.14 | 0.13 | 0.13 | 0.16 | 0.14 | 0.13 | 0.19 | 0.15 | 0.13 | 0.07 | 0.07 | 0.07 | 0.08 | 0.07 | 0.07 | 0.08 | 0.07 | 0.07 | 0.10 | 0.07 | 0.07 |
| Failure rate estimate: 3 match | 0.13 | 0.13 | 0.13 | 0.13 | 0.13 | 0.13 | 0.13 | 0.12 | 0.12 | 0.13 | 0.12 | 0.11 | 0.07 | 0.07 | 0.07 | 0.07 | 0.07 | 0.07 | 0.07 | 0.07 | 0.07 | 0.06 | 0.06 | 0.06 |
| Failure rate estimate: 4 match | 0.13 | 0.13 | | 0.13 | 0.13 | | 0.12 | 0.12 | | 0.12 | 0.11 | | 0.07 | 0.07 | | 0.07 | 0.07 | | 0.07 | 0.07 | | 0.06 | 0.06 | |
| Failure rate estimate: 5 match | 0.13 | | | 0.13 | | | 0.12 | | | 0.11 | | | 0.07 | | | 0.07 | | | 0.07 | | | 0.06 | | |

*AmpSeq: Amplicon Sequencing, DHA-PPQ: Di-hydroartemisinin-Piperaquine, MOI: Multiplicity of Infection, FOI: Force of Infection, "Match" refers to the number of AmpSeq loci at which at least one allele must be shared between the initial and recurrent blood samples for the recurrence to be classified as a recrudescence.*

**Table 5.4** Failure rate estimates obtained for modelled scenarios of AR-LF when genotyping 5, 4 and 3 AmpSeq loci, to two significant figures.

| Drug | AR-LF | | | | | | | | | | | | | | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| MOI | High | | | | | | | | | | | | Low | | | | | | | | | | | |
| FOI | 0 | | | 2 | | | 8 | | | 16 | | | 0 | | | 2 | | | 8 | | | 16 | | |
| Total AmpSeq loci genotyped | 5 | 4 | 3 | 5 | 4 | 3 | 5 | 4 | 3 | 5 | 4 | 3 | 5 | 4 | 3 | 5 | 4 | 3 | 5 | 4 | 3 | 5 | 4 | 3 |
| Failure rate estimate: 1 match | 0.14 | 0.14 | 0.14 | 0.18 | 0.16 | 0.15 | 0.29 | 0.24 | 0.20 | 0.40 | 0.32 | 0.26 | 0.09 | 0.09 | 0.09 | 0.12 | 0.11 | 0.10 | 0.19 | 0.15 | 0.13 | 0.28 | 0.21 | 0.17 |
| Failure rate estimate: 2 match | 0.14 | 0.14 | 0.14 | 0.14 | 0.14 | 0.14 | 0.18 | 0.15 | 0.14 | 0.22 | 0.17 | 0.14 | 0.09 | 0.09 | 0.09 | 0.09 | 0.09 | 0.09 | 0.10 | 0.09 | 0.09 | 0.12 | 0.10 | 0.09 |
| Failure rate estimate: 3 match | 0.14 | 0.14 | 0.14 | 0.14 | 0.13 | 0.13 | 0.14 | 0.13 | 0.13 | 0.15 | 0.13 | 0.13 | 0.09 | 0.09 | 0.09 | 0.09 | 0.09 | 0.09 | 0.09 | 0.09 | 0.09 | 0.09 | 0.08 | 0.08 |
| Failure rate estimate: 4 match | 0.14 | 0.14 | | 0.13 | 0.13 | | 0.13 | 0.13 | | 0.13 | 0.13 | | 0.09 | 0.09 | | 0.09 | 0.09 | | 0.09 | 0.09 | | 0.08 | 0.08 | |
| Failure rate estimate: 5 match | 0.14 | | | 0.13 | | | 0.13 | | | 0.12 | | | 0.09 | | | 0.09 | | | 0.09 | | | 0.08 | | |

*AmpSeq: Amplicon Sequencing, AR-LF: Artemether-Lumefantrine, MOI: Multiplicity of Infection, FOI: Force of Infection, "Match" refers to the number of AmpSeq loci at which at least one allele must be shared between the initial and recurrent blood samples for the recurrence to be classified as a recrudescence.*

### 5.3.3 Failure rate estimates using AmpSeq for DHA-PPQ and AR-LF: Sensitivity analysis of model parameters.

Three important model assumptions were varied to assess their impact on results: The minority detection threshold (**5.2.5**), the sampling limit (**5.2.5**), and the lower limit of the initial parasite number distribution (**5.2.2**).

#### 5.3.3.1 Sensitivity analysis: Minority detection threshold.

The minority detection threshold was 1% in the baseline model (i.e., alleles were detected as long as they were at least 1% the frequency of the most frequent (majority) allele). The rationale behind this threshold is described in **5.1**, but importantly it is user-defined on genotyping software when AmpSeq markers are genotyped in vivo (i.e., [36]) and thus assessing the impact of this parameter on the findings of the models within this chapter was important. It was varied up to 2% to assess the impact of a user-defined higher cut-off point, and down to 0% to assess the difference between 1% and 2% from a hypothetical perfect detection, noting that 0% is unfeasible in vivo because that would permit inclusion of false alleles created by genotyping artefacts or gametocytes. The true failure rates of DHA-PPQ and AR-LF in each MOI setting were identical to those in the baseline model (**Figure 5.1**, and described in **5.3.1**).

Failure rate estimates for DHA-PPQ and AR-LF with low and high MOI and a range of FOI values are shown in **Figure 5.4** for a minority allele detection threshold of 0% and **Figure 5.5** for a minority allele detection threshold of 0.02%. When compared to a minority allele detection threshold of 1% (**Figure 5.1**), a threshold of 0% resulted in slightly higher failure rate estimates and a threshold of 2% resulted in slightly lower failure rate estimates. In both cases, the difference was extremely negligible and failure rate estimates obtained using all minority allele detection thresholds were close in value (less than 0.1% difference at every FOI value modelled).

#### 5.3.3.2 Sensitivity analysis: Sampling limit.

The sampling limit (the parasitaemia of a clone required for that clone to be included on a blood sample) was $10^8$ in the baseline model (**5.2.5**). This limit was based on calculations derived from real processes but given the high ability to detect low frequency alleles using AmpSeq markers, it was necessary to check this assumption was not biasing results. The sampling limit was thus varied by reducing it to $10^7$ (i.e., so lower frequency clones would be able to be included in the blood sample). The true failure rates of DHA-PPQ and AR-LF in each MOI setting are identical to those in the baseline model (**Figure 5.1**). The minority allele detection threshold was 1%, as for the baseline model. As in the baseline model, 5 loci were genotyped.

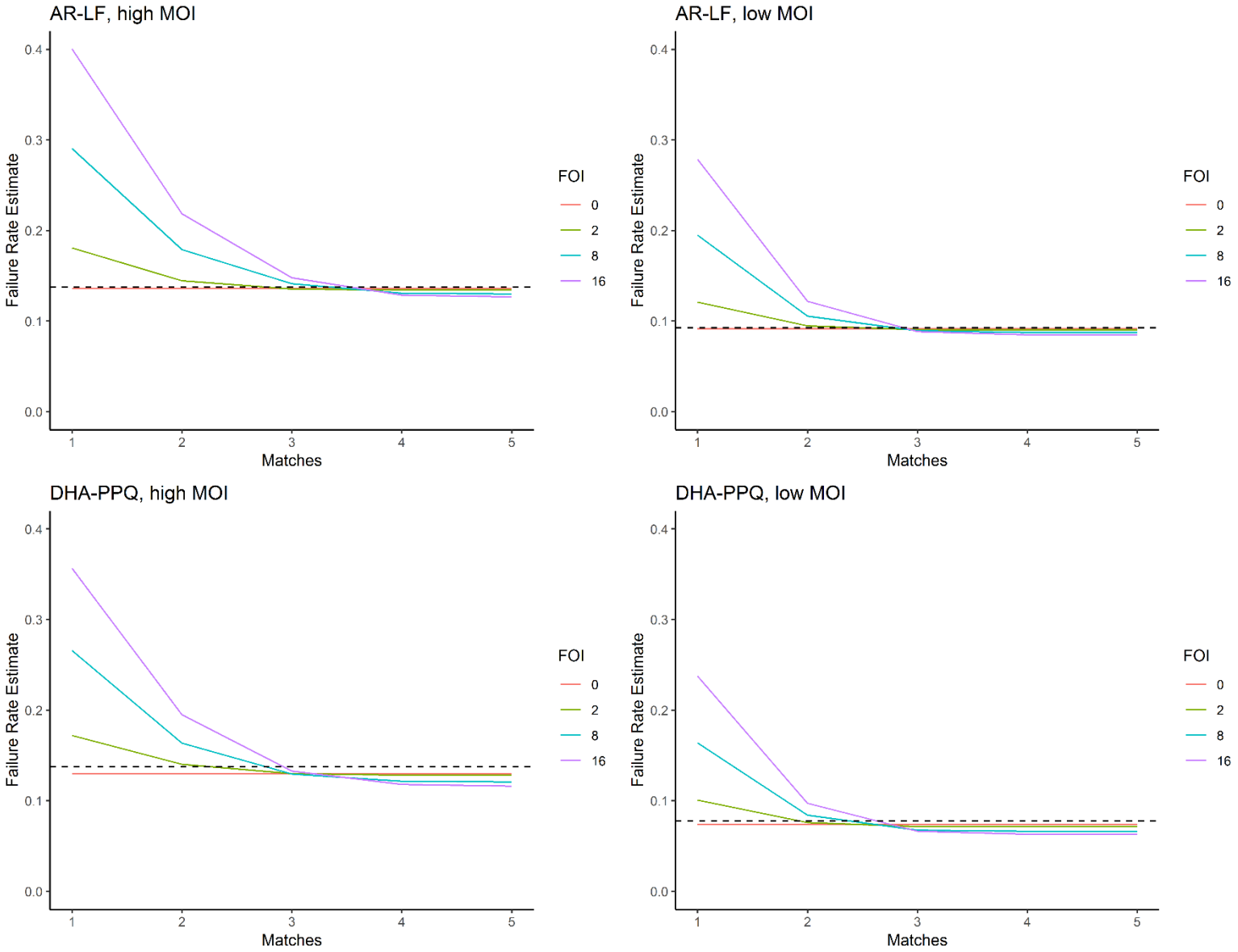
Failure rate estimates under the assumption of lower sampling limit are shown for DHA-PPQ and AR-LF in **Figure 5.6**. Results were qualitatively extremely similar to the baseline model (**Figure 5.1**). There was an extremely minor increase in failure rate estimates at higher FOI (8 and 16) of ~0.02% when using lower number of matches (≥1 or ≥2) to classify a recrudescence. In short, results were functionally identical to the baseline assumption of a sampling limit of $10^8$, so the assumed value of the sampling limit did not appear to affect the failure rate estimates obtained using AmpSeq markers.

#### 5.3.3.3 Sensitivity analysis: Lower limit of Initial parasite numbers.

The lower limit of the log-uniform distribution was varied from $10^{10}$ (the baseline lower limit) to $10^8$ (the upper limit remains as $10^{11}$, see **5.2.2**). This was done to investigate if the accuracy of failure rate

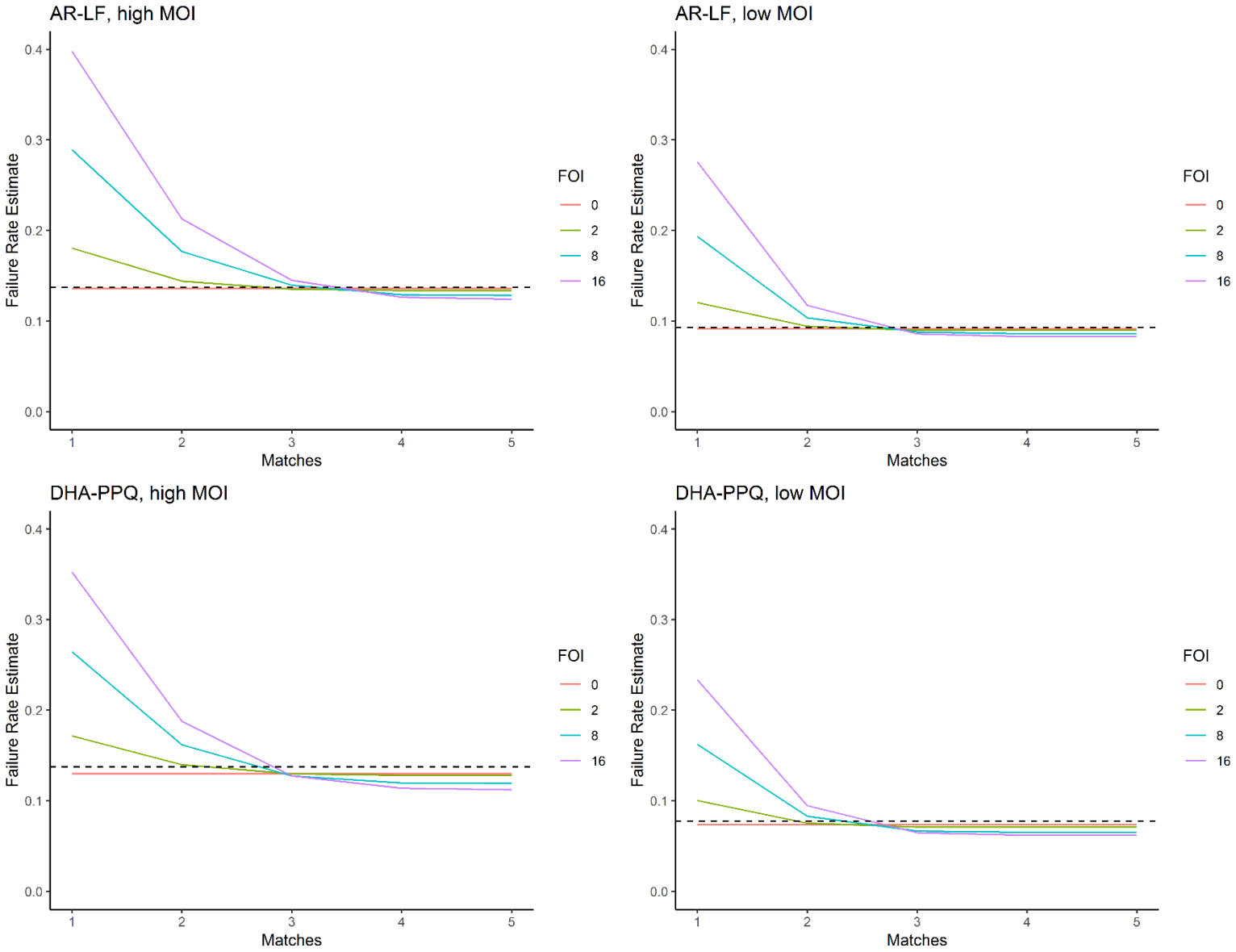
estimates generated using AmpSeq markers were affected by assuming a wider range of initial parasitaemia across clones (this will lead to an average increase in the number of low frequency alleles). The true failure rate changed as the range of this distribution changed: 0.1358 and 0.0758 for DHA-PPQ in high and low MOI settings respectively and 0.1342 and 0.092 for AR-LF in high and low MOI settings respectively. These true failure rates are slightly lower than the baseline scenario (DHA-PPQ true failure of 0.138 and 0.0774 in areas of high and low MOI respectively; AR-LF true failure of 0.1376 and 0.0928 in areas of high and low MOI respectively), due to initial clones with a lower starting parasitaemia being less likely to recrudesce (though the absolute change in true failure rate was negligible, so this effect did not appear to be large). The minority allele detection threshold was 1%, as for the baseline model. The sampling limit was $10^{8,}$ as for the baseline model.

Failure rate estimates using a wider initial parasite number distribution are shown for DHA-PPQ and AR-LF in Figure 5.7. Failure rate estimates were slightly lower in all cases with a wider distribution, though this should be considered relative to the slightly lower true failure rate. The relative distance of each estimate from the true failure rate and thus, the qualitative conclusions, were identical to the baseline model i.e. classifying a recrudescence at ≥3 matching loci accurately recovered the true failure rate for both drugs, both MOI settings and all FOI values.


AR-LF, high MOI
AR-LF, low MOI
DHA-PPQ, high MOI
DHA-PPQ, low MOI
Failure Rate Estimate
Matches
FOI
0
2
8
16
0.0
0.1
0.2
0.3
0.4
1
2
3
4
5

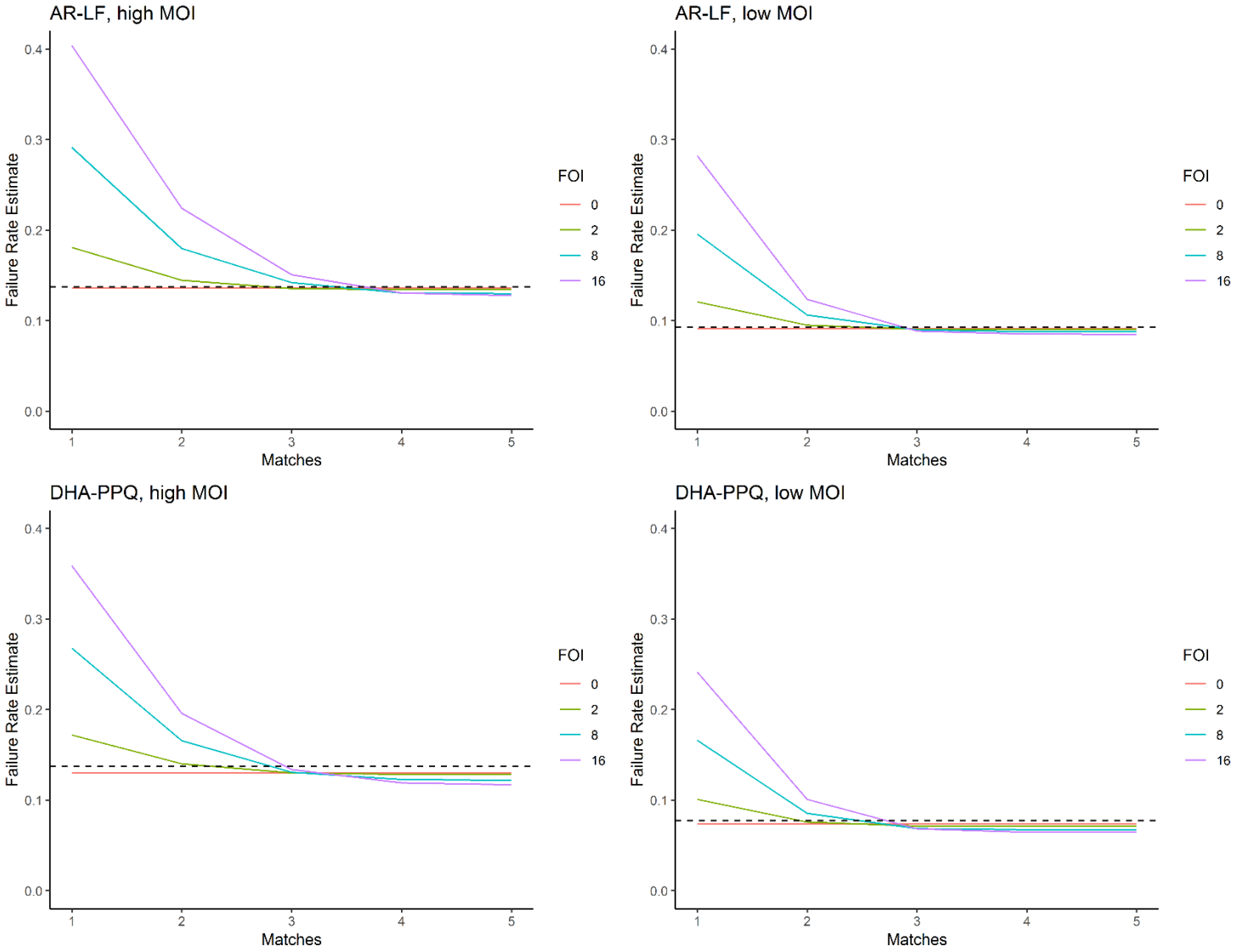
3101

**Figure 5.4** Failure rate estimates obtained using 5 Amplicon Sequencing (AmpSeq) loci for Di-hydroartemisinin-Piperaquine (DHA-PPQ) and Artemether-Lumefantrine (AR-LF) in low and high Multiplicity of Infection (MOI) settings with a range of Force of Infection (FOI) values, using a minority allele detection threshold of 0%. Failure rate estimates were obtained using a given number of matching loci between the initial and recurrent sample to classify a recrudescence. The true failure rate is marked by the horizontal dashed black line.


AR-LF, high MOI
AR-LF, low MOI
DHA-PPQ, high MOI
DHA-PPQ, low MOI
Failure Rate Estimate
Matches
FOI
0
2
8
16
0.0
0.1
0.2
0.3
0.4
1
2
3
4
5

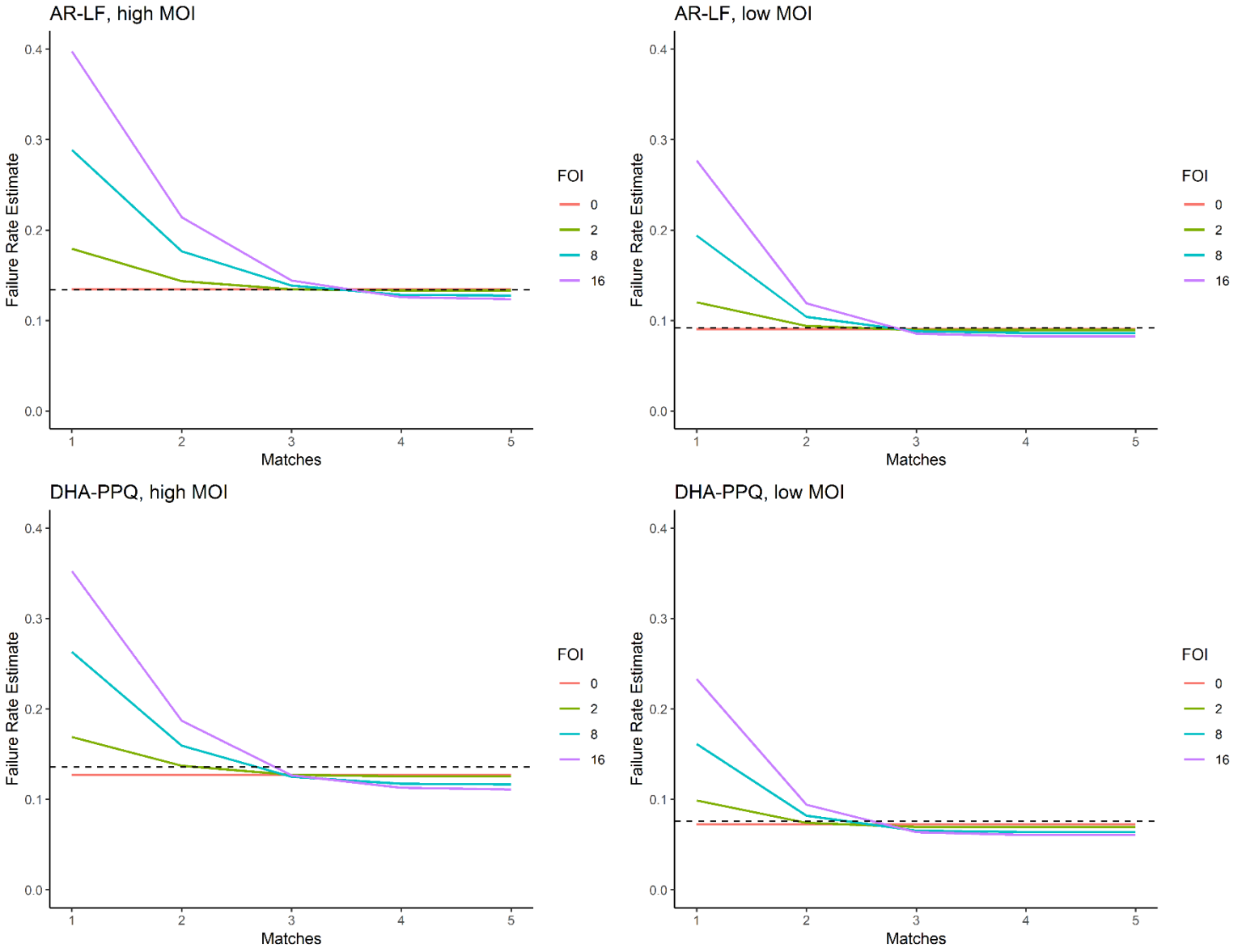
3106

**Figure 5.5** Failure rate estimates obtained using 5 Amplicon Sequencing (AmpSeq) loci for Di-hydroartemisinin-Piperaquine (DHA-PPQ) and Artemether-Lumefantrine (AR-LF) in low and high Multiplicity of Infection (MOI) settings with a range of Force of Infection (FOI) values, using a minority allele detection threshold of 2%. Failure rate estimates were obtained using a given number of matching loci between the initial and recurrent sample to classify a recrudescence. The true failure rate is marked by the horizontal dashed black line.


AR-LF, high MOI
AR-LF, low MOI
DHA-PPQ, high MOI
DHA-PPQ, low MOI
Failure Rate Estimate
Matches
FOI
0
2
8
16
0.0
0.1
0.2
0.3
0.4
1
2
3
4
5

3112

**Figure 5.6** Failure rate estimates obtained using 5 Amplicon Sequencing (AmpSeq) loci for Di-hydroartemisinin-Piperaquine (DHA-PPQ) and Artemether-Lumefantrine (AR-LF) in low and high Multiplicity of Infection (MOI) settings with a range of Force of Infection (FOI) values, using a sampling limit of $10^7$. Failure rate estimates were obtained using a given number of matching loci between the initial and recurrent sample to classify a recrudescence. The true failure rate is marked by the horizontal dashed black line.


AR-LF, high MOI
AR-LF, low MOI
DHA-PPQ, high MOI
DHA-PPQ, low MOI
Failure Rate Estimate
Matches
FOI
0
2
8
16
0.0
0.1
0.2
0.3
0.4
1
2
3
4
5

3117

**Figure 5.7** Failure rate estimates obtained using 5 Amplicon Sequencing (AmpSeq) loci for Di-hydroartemisinin-Piperaquine (DHA-PPQ) and Artemether-Lumefantrine (AR-LF) in low and high Multiplicity of Infection (MOI) settings with a range of Force of Infection (FOI) values, using a wider initial parasite number range of $10^8$ to $10^{11}$. Failure rate estimates were obtained using a given number of matching loci between the initial and recurrent sample to classify a recrudescence. The true failure rate is marked by the horizontal dashed black line.

## 5.4 Discussion.

I have presented an in silico approach designed to quantify the accuracy of failure rate estimates obtained using novel AmpSeq methodology. The proposed advantages of AmpSeq for molecular genotyping of have been explored in vivo and in vitro [36, 41, 158-160]. In particular, existing research has identified suitable AmpSeq loci for genotyping [36, 41, 157], quantified greater detectability of minority alleles relative to traditional genotyping methods [36, 159], provided improved estimates of MOI [160] and developed and compared analysis tools and identified appropriate exclusion thresholds for minority alleles [36, 41, 158]. However, while the true failure rate of patient populations remains unknown in vivo, the accuracy of the failure rate estimates and best practice of using AmpSeq (i.e., how many loci should be genotyped, what should be the number of matching alleles between the initial and recurrent samples at which a recurrence is classified as a recrudescence) have not been possible to quantify. The modelling approach presented here bridges this gap and validates the accuracy of using AmpSeq for molecular correction.

### 5.4.1 Accuracy of failure rate estimates.

Modelled results with an FOI of 0 (i.e., representing an area of zero ongoing transmission) show the degree to which AmpSeq under-estimated the true failure rate by failing to observe minority alleles. In an area of FOI=0, there are only two reasons a recrudescence would not be detected: Firstly, if the recrudescing clone(s) did not reach detectable levels by the time the follow-up period ends. Secondly, if a recrudescing clone (and thus, it's alleles) was not detected in either blood sample. Because AmpSeq is capable of detecting minority clones at a frequency of 1% of the majority clone, failure rate estimates obtained when an FOI value of 0 is assumed were extremely close to true failure rates, with negligible under-estimates (note that with an FOI of 0, the number of matches at which a recurrence was classified as a recrudescence had no impact on the failure rate estimate).

As FOI was increased, failure rate estimates varied as the number of matching loci required to classify a recurrent infection as a recrudescence was varied. This occurred because introducing reinfections a) meant there were more total clones and thus a higher chance for recrudescent clones to be below the minority detection threshold in the recurrent sample (i.e., some truly recrudescent alleles may not be observed and b) more alleles are, on average, in the recurrent blood sample and so the chance of misclassifying a reinfection as a recrudescence increases (note this chance will be close to 0 when a match is required at all loci genotyped). Unsurprisingly, the higher the FOI, the greater the difference in failure rate estimates as the matching loci threshold was changed (**Figure 5.1**). The key operational question is thus: What matching loci threshold produces the failure rate estimate closest to the true failure rate, and is this finding consistent for multiple drugs and in multiple MOI and FOI settings?

When genotyping five total AmpSeq loci, classifying a recrudescence with three or more matching loci produced the most accurate failure rate estimates. With more stringent thresholds (≥4 or =5 matches), there was an under-estimate of the true failure rate in every instance. Notably the under-estimate was higher as FOI increased; the increased number of alleles meant there was a greater chance of an allele not being detected in either the initial or the recurrent sample, and with a threshold of 5, only a single allele would have to be missed in this way to misclassify a recrudescence as a reinfection. On the other hand, classifying a recrudescence with ≥1 or ≥2 matching loci would clearly be inappropriate and lead to large over-estimates of failure rate with any FOI value above 0.

*In vitro* AmpSeq experiments have excluded the loci *msp-7* based on its low diversity relative to other AmpSeq loci, and only used *csp* as a back-up locus if *ama1-D3*, *cpmp* or *cpp* failed to amplify [41]. The impact of genotyping a smaller number of total loci was investigated *in silico* here, under the theory that genotyping fewer loci is operationally easier and more economical. The modelling approach was able to quantify the difference in the accuracy of failure rate estimates that arose from genotyping a

lower number of loci. Comparison of failure rate estimates obtained here (**Figure 5.1**, **Figure 5.2**, **Figure 5.3**) showed that accurate failure rate estimates could be obtained by genotyping only the three most diverse loci. Notably, when genotyping three loci instead of five (**Figure 5.3**), the most accurate failure rate estimates will be obtained when classifying a recurrence as recrudescence when ≥2 matches occurred between the initial and recurrent sample. When using 5 loci (**Figure 5.1**), classifying a recrudescence with ≥2 matches over-estimated the true failure rate; removing the lower diversity loci reduced the chance of a reinfection sharing alleles with an initial clone (and so being misclassified as a recrudescence at lower thresholds). The difference in failure rate estimates obtained through classifying a recrudescence at ≥2 or 3 matching loci with 3 loci genotyped was extremely small in these simulations; researchers may wish, in vivo, to classify a recrudescence when 3/3 loci match as this will be more conservative with regards to potential misclassification of reinfection, particularly if lower diversity AmpSeq loci than the ones presented here must be used. In short, genotyping 3 total amplicon loci appeared to be sufficient for the purposes of producing accurate failure rate estimates in these simulated TES. It is important to note, however, that genotyping a greater number of total loci is likely to be useful in practice as it allows results to be interpreted even if samples fail to amplify at any given locus (see [41]).

Note that the loci used in these simulations (**Table 5.2**) were relatively diverse. Lower diversity distributions of AmpSeq loci were not currently available (as this is a relatively new methodology, populations with low genetic diversity have not been genotyped using AmpSeq to date). The impact of lower genetic diversity would be increased failure rate estimates as the frequency of reinfections being misclassified as recrudescence increases due to them sharing alleles purely by chance. In such areas, it may be necessary to consider a) genotyping a larger number of total loci and b) to use higher (more stringent) thresholds of matching loci to classify a recurrence as a recrudescence. Use of AmpSeq in TES should involve obtaining MOI estimates of the initial samples and quantifying the level of genetic diversity of those samples, such that informed decisions around the total number of markers to analyse and the threshold chosen can be made. Fortunately, obtaining accurate MOI and diversity estimates is possible with AmpSeq due to their high resolution (**5.1**) [160]. If and when lower diversity data-sets become available, this modelling work can be repeated to quantify the accuracy of failure rate estimates in such areas.

Importantly, the qualitative findings of the model were robust as key model parameters were altered. A lower sampling limit did not affect results, and the qualitative findings when changing the initial parasite number distribution were identical (given the low true failure rate). The difference between a 0% minority detection threshold (i.e., hypothetical perfect detection) and a 1% (baseline) and 2% threshold were negligible. This indicates that a 1-2% cut-off can be safely included to prevent the inclusion of false alleles in the analysis without any notable decrease in accuracy from misclassifying minority recrudescent clones.

### 5.4.2 Policy implications and future work.

AmpSeq for molecular correction recovers the true failure rate to an extremely high degree of accuracy. Based on this modelling work, my strong recommendation is that the field should consider seeking to implement its wide-spread use as the new gold standard for genotyping in TES. Crucially, provided an appropriate number of sufficiently diverse loci are genotyped, failure rate estimates are robust for drugs with either a long or short post-treatment prophylactic effect and across a range of transmission intensities. Highly diverse loci are desirable for genotyping and the modelling work here shows that increases in accuracy were produced by removing lower diversity loci from analysis. Consequently, efforts should be made to identify, in vivo, diverse loci for use in each sentinel site where TES are conducted by using large data-sets such as those curated by the Malaria Genomic

Epidemiology Network (MalariaGEN) [147]. A full discussion of a strategy for identifying AmpSeq loci suitable for genotyping can be found in [36].

Traditional molecular correction with either length-polymorphic markers or microsatellite markers has been conducted using either gel-based or capillary electrophoresis (CE) of PCR products. The 2008 WHO guidelines contained protocols for both gel-based electrophoresis and CE [14, 161] but CE offered higher sensitivity and ability to discriminate between alleles with minimal size differences; CE is now widely used and has generally phased out gel-based electrophoresis for molecular correction [106, 161]. The adoption of AmpSeq will require use of next-generation sequencing platforms such as 454 GS Junior (Roche), Ion Torrent PGM (Life Technologies) or MiSeq (Illumina); see [36, 162] for a comparison of these platforms and note that existing literature using AmpSeq in a malaria context has preferred the MiSeq platform [36, 41, 160]. The economic cost of deploying these machines to sub-Sarahan Africa and South East Asia for use in a malaria context has not been fully quantified but is likely to be significant and will necessarily include training, reagent supply and maintenance. Having one in every sentinel site is likely to be extremely unfeasible, particularly in the short term. Prompt deployment of AmpSeq for analysis of malaria TES should focus on equipping a central site – one per country or even regionally if necessary with the technology required for the methodology. This economic factor appears to be the largest obstacle for AmpSeq as a molecular correction methodology, but should be balanced against the long-term economic benefits of accurate failure rate estimates (warding against drug resistance).

# Chapter 6: Optimal treatments for severe malaria and the threat posed by artemisinin resistance.

A version of this work has been published in The Journal of Infectious Diseases: Sam Jones, Eva Maria Hodel, Raman Sharma, Katherine Kay, Ian M Hastings, Optimal Treatments for Severe Malaria and the Threat Posed by Artemisinin Resistance, The Journal of Infectious Diseases, Volume 219, Issue 8, 15 April 2019, Pages 1243–1253, https://doi.org/10.1093/infdis/jiy649

Chapter-specific acknowledgements: Dr Eva Maria Hodel provided R code with which to simulate the duration of artesunate killing of parasites from a given set of pharmacokinetic parameters. Dr Raman Sharma provided R code with which to run a Partial Rank Correlation Coefficient analysis on outcome metrics and specific covariates of interest. Dr Katherine Kay provided R code with which to calculate correlation between select pharmacokinetic parameters.

## 6. 1 Background.

### 6.1.1 Severe malaria.

*Plasmodium falciparum* is the malaria species responsible for the largest number of deaths worldwide [1] and presents clinically in two forms. Patients with "uncomplicated" malaria have a relatively mild fever, are conscious and capable of taking oral drug regimens; prompt treatment of uncomplicated malaria is associated with low mortality [13]. Patients with "severe" malaria present with one, or a combination, of four syndromes: Severe anaemia, respiratory distress, metabolic derangement and cerebral malaria [16, 17]. Patients are treated with parenteral artesunate, which rapidly kills parasites, but resolution of pathology lags behind parasite killing; case fatality rates are high even once patients have been admitted to the formal health system (typically between 5 and 12% [13] although these have been falling to ~2% [18]).

A key factor responsible for severe malaria is the binding of parasitized erythrocytes (subsequently called infected red blood cells, iRBCs) to microvascular endothelium, a process known as sequestration. iRBC sequestration induces pathology through three main causes: (i) impairing blood flow to organs through direct physical blockage of the capillaries [163], (ii) indirect blockage via host defence mechanisms such as inflammation [16, 164] and (iii) physical damage to microvascular endothelium and the blood/brain barrier [165]. High case fatality rates occur, even if the drug kills parasites within sequestered iRBCs, because the molecules responsible for sequestration (for example, *P. falciparum* erythrocyte membrane protein 1 (PfEMP1) [166]) are still present on iRBC surfaces and it takes a significant amount of time for these ligands to decline sufficiently for the sequestered iRBC to detach and/or for the pathology associated with sequestration to resolve[167, 168].

### 6.1.2 Parasite clearance rates as an endpoint for severe malaria clinical trials.

Parasite clearance rates are a commonly used clinical outcome measure to compare efficacy of antimalarial treatment regimens. However, parasite clearance rates correlate poorly with disease outcome in severe malaria. Large trials comparing intramuscular artemether with quinine in African children showed more rapid parasite clearance with artemether but no difference in case fatality [48, 49]. With parenteral artesunate, parasite clearance rates are not different in patients dying from severe malaria compared to survivors (results cited in [50]). There are two potential explanations why parasite clearance is an unsuitable outcome measure in severe malaria: Firstly, parasite clearance rates following treatment for uncomplicated malaria appear to mainly reflect host immunity rather than drug effectiveness [52, 53, 129] so may be a poor metric of overall drug effectiveness. Secondly,

parasite clearance rates are measured on circulating parasites [129] whereas non-circulating, sequestered parasites are responsible for most clinical symptoms, pathology and deaths associated with severe malaria [16].

There is considerable interest in conducting clinical trials for severe malaria to test new treatments and treatment regimens with the goal of improving patient survival (the key clinical aim of treating severe malaria), and a clinical outcome is required to quantify the differences between trial arms. Mortality would be a good clinical outcome to use, but a trial would have to be unfeasibly large to have sufficient statistical power for this to be acceptable. It is extremely difficult (if not impossible) to measure sequestered parasite load in vivo, and so, despite the noted flaws in using parasite clearance rates (or half-life) as the clinical outcome in severe malaria trials as listed above, they continue to be widely used.

### 6.1.3 Research Goals.

1: To develop a mechanistic pharmacokinetic(PK)/pharmacodynamic(PD); (mPK/PD) model that describes the dynamics of sequestered parasites (and their pathology) in severe malaria in the first 48 hours following treatment with artesunate.

2: To use this model to compare the differences between existing and proposed novel drug regimens on the pathology of severe malaria.

3: To use this model to investigate the likely consequences of artemisinin resistance on the pathology of severe malaria.

## 6.2 Methodology.

A computer-based mPK/PD model was utilized to track changes in the number of sequestered iRBCs following drug administration. This model was for AS monotherapy and was implemented in the statistical programming software R [169] version 3.4.1. *P. falciparum* parasites undergo a 48-hour developmental cycle in human erythrocytes with two main implications for pathology and treatment. Firstly, parasites initially circulate freely in blood vessels but sequester (i.e. bind to capillaries) at mature stages of their intra-erythrocytic cycle. Secondly, parasites differ in their sensitivity to drugs over the course of this 48-hour cycle.

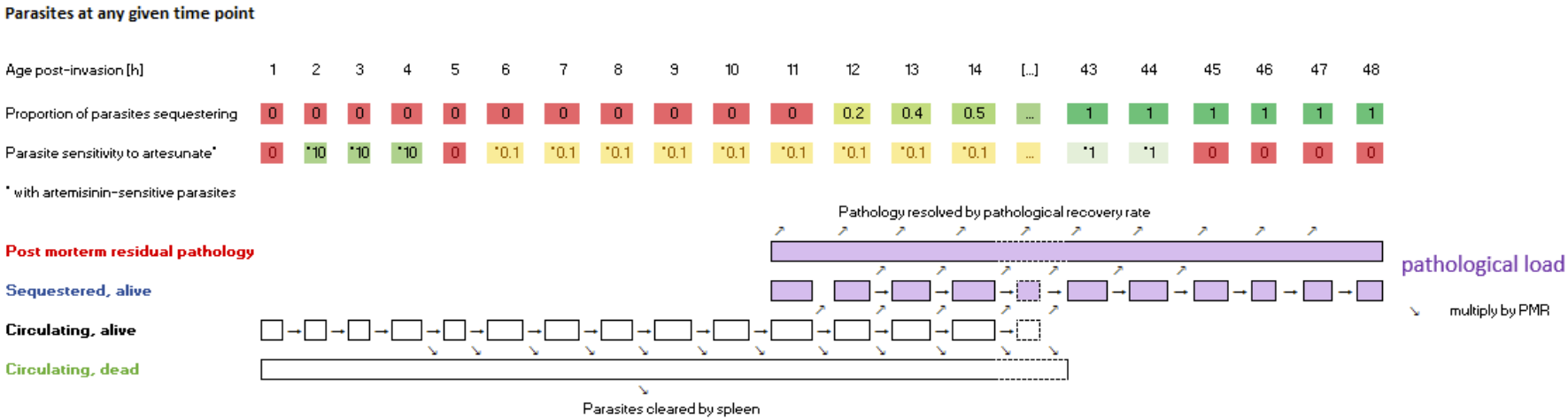
As previously described [15], the parasite population was separated into within patient 48 ‘age-bins’ that each represented a one-hour long development stage in the parasite’s 48-hour life-cycle within human erythrocytes. Parasites within age-bins had differing propensities to sequester and had varying degrees of drug sensitivity. The model tracked the number of iRBCs in each of four classes at any time post-treatment depending on whether the parasites are alive or dead, and whether the iRBC is circulating or sequestered: Alive & circulating, alive & sequestered, dead & circulating, and dead & sequestered (see **Figure 6.1** for illustration). Note that iRBCs classed as “dead & sequestered” are those iRBCs whose parasites have died while sequestered and are either: (i) still sequestered and causing pathology or (ii) have ruptured/detached from the capillary but are still associated with continued, lingering pathology.

I initially wrote the model to simulate a single patient with a user-defined duration and rate of artesunate killing. In other words, the methodology to analyse a single patient with a given duration of AS killing was developed by me. As it was felt at a relatively early stage that this research project was better served by simulating a large population of patients with explicit variation in AS killing duration (via variation in PK/PD parameters), Eva Maria Hodel provided me with R code to calculate artesunate killing from sampled PK parameters. I sampled durations of killing using this code with a

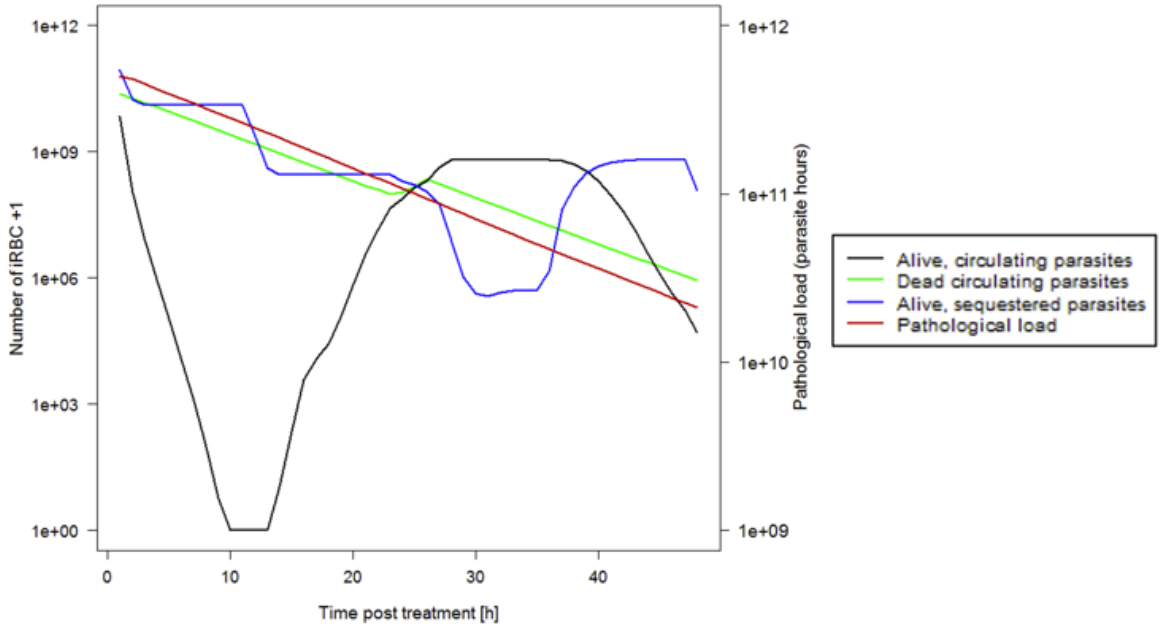
variety of PK parameters from different sources (**6.2.3**). Katherine Kay contributed code to calculate correlation between Volume of Distribution (VD) and Clearance (CL); described in **6.2.3**.

A cohort of 10,000 patients were then simulated who had parasitological, pharmacological, and patient-specific parameters drawn from the distributions given in Table **6.1** . Individual patient profiles allowed individual PK/PD variation to be incorporated to generate individual patient post-treatment parasite clearance dynamics. Each patient was simulated three times under different scenarios: Once for drug sensitive parasites treated by the standard WHO regimen (2.4mg/kg artesunate twice a day in the first 24h), once for sensitive parasites treated with the simplified regimen (4mg/kg artesunate once a day, as proposed by Kremsner et. al [170]), and once for artemisinin resistant parasites treated by the standard WHO regimen. This allowed us to compare the two dosing regimens ("standard" versus "simplified") and the impact of resistance ("sensitive" versus "resistant"), in each patient. Follow-up time was 48 hours after drug administration; this reflected a whole parasite life-cycle within an iRBC but, more importantly, covers the period post-treatment where a patient is most likely to die [171, 172].

Panel A




Panel B




3360

3361

**Figure 6.1** A schematic of how the model tracks parasitaemia and pathology post-treatment.

**Panel (A)** shows how the simulation tracks parasitaemia and pathology. The parasite population is separated into 48 hourly 'age-bins' corresponding to their developmental age within their 48hour intra-erythrocytic cycle. A certain proportion of parasites in each age-bin will be sequestered, with 0% of parasites sequestering in age-bins 1 to 11 and ~100% sequestering in age-bins 14-48 (the proportions given in the figure are illustrative). Parasites in age-bin 48 rupture to produce new 'daughter' parasites that enter age-bin 1; the number of daughter parasites that successfully invade new erythrocytes is the parasite multiplication rate PMR. The simulation runs in one-hour time steps and, if drug is present, it kills parasites according to their drug sensitivity which is given in the second row of boxes as a proportion of basal kill rate (Supplementary information). Parasites that survive drug action are moved forward one age-bin (unless they are in age-bin 48 in which case they rupture to produce daughter parasites as described above). Parasites killed by drug in the time-step have two fates depending on their status. Those killed in circulating stages enter a pool of "dead circulating parasites" and will eventually be removed by splenic or other host clearance mechanisms. Those parasites that are killed while sequestered are removed from the simulation but their pathology does not instantly disappear with their death, so their "post-mortem pathology" (second term of equation 1) is tracked – this pathology resolves at the user-defined "pathological recovery rate".

**Panel B** shows how this methodology is used to simulate treatment of one exemplar individual. The number of alive circulating plus dead circulating parasites can be tracked over time post treatment. These two classes can be directly observed (but not distinguished) in human blood samples and their rate of clearances, usually known as "parasite clearance rate" is often used as a proxy of clinical outcome; this enables verification of simulations (in terms of the ability of simulations to recover these clinical observations). Live sequestered parasites are added to the lingering effects of sequestered parasites killed in earlier stages (i.e. those contributing to "post-mortem pathology") to obtain the pathological load L(t) at any time point post-treatment (Equation 1). The dynamics of L(t) following treatment are used to calculate key pathology metrics i.e. area under the pathology curve ($AUC_{PL}$) and the maximum parasite load (MPL).

**Table 6.1** Parameter values used in the simulations. (Not including volume of distribution (Vd) / clearance (Cl) that determine artesunate killing duration, see 6.2.3 for discussion of these parameters).

| Parameter | Unit | Abbreviation | Range | Format | Distribution | Justification |
|---|---|---|---|---|---|---|
| Initial parasite number | | $P_0$ | $10^x$, where $(x \in \mathbb{R} \mid 10 < x < 12)$ | Double | Uniform | [173, 174] |
| Mean of initial age bin distribution | [h] | Mean | x + 0.5, where $(x \in \mathbb{N} \mid 0 \leq x \leq 47)$ | Integer | Triangular with mode = 10 | [174, 175] |
| Standard deviation of initial age bin distribution | [h] | SD | x, where $(x \in \mathbb{N} \mid 2 \leq x \leq 4)$ | Integer | Uniform | [173] |
| Parasite multiplication rate | | PMR | x, where $(x \in \mathbb{N} \mid 1 \leq x \leq 10)$ | Integer | Triangular with mode = 1 | [173, 175] |
| Pathological recovery rate half-life | $[h^{-1}]$ | r = ln(2)/x | x, where $(x \in \mathbb{N} \mid 4 \leq x \leq 12)$ | Integer | Uniform | |
| Splenic clearance rate half-life | $[h^{-1}]$ | u = ln(2)/x | x, where mean = 2.7 and CV = 0.3 | Double | Normal | [176, 177] |
| Half-maximum inhibitory concentration of AS | [mg/L] | $IC50_{AS}$ | x, where mean = 0.0016 and CV = 0.86 | Double | Log-normal | [87] |
| Half-maximum inhibitory concentrations of DHA | [mg/L] | $IC50_{DHA}$ | x, where mean = 0.009 and CV = 1.17 | Double | Log-normal | [87] |
| Maximal rate of drug killing | $[h^{-1}]$ | $V_{max}$ | x, where mean = 1.78 and CV = 0.1 | Double | Normal | [87, 140] |
| Slope factor | | n | x, where mean = 4 and CV = 0.3 | Double | Normal | [87] |

*AS: artesunate; CV: coefficient of variation; DHA: dihydroartemisinin; i.m: intramuscular*

### 6.2.1 Modelling parasite age-bins.

Note that it was explicitly assume only a single clone is responsible for sequestration-based pathology, consistent with existing research [178, 179]. Modelling a mixed infection with this model, should a reader desire, is straightforward: Two or more clones can be simultaneously tracked in the same individual by running one simulation for each clone. It is assumed that pathology is additive so total pathology would be the sum of the individual clonal pathologies.

Developmental age-bins were enumerated chronologically: Bin 1 represents an infected red blood cell (iRBC) in the first hour following parasite invasion, and bin 48 is the final age-bin. Parasite development ends after bin 48 with the rupturing of the iRBC to release merozoites that re-invade RBCs and re-enter age-bin 1. The intra-host model tracks the number of parasites in each of the 48 age-bins at each time point post-treatment. Specifically, the number of alive iRBCs (both circulating and sequestered) at each time-point in each bin was tracked and used to calculate the number of alive iRBCs in the next age bin at the next time-point as:

$$N_{t+1}^{b+1} = N_t^b (1 - D_t^b) \qquad \text{Equation 6.1}$$

where $N_{t+1}^{b+1}$ represents the number of alive iRBCs in age-bin b+1 at time t+1 post-treatment, which depended on the iRBCs in the current age-bin and time period $N_t^b$ and the proportion which survived drug treatment in the bin at time $(1 - D_t^b)$ where $D_t^b$ is the drug-specific killing rate for that age bin, i.e. the proportion of parasites killed in age bin b at time t (see below). The only exception is for age-bin 1 which reflected parasites released from iRBCs at age-bin 48 i.e.

$$N_{t+1}^{1} = N_t^{48} (1 - D_t^{48}) PMR \qquad \text{Equation 6.2}$$

where PMR is the parasite multiplication rate, i.e. the number of merozoites released from a schizont that successful infect a new erythrocyte (Table **6.1** ).

### 6.2.2 Modelling parasite sequestration.

Sequestration was incorporated by following, for convenience, the assumption of Saralamba et al. [173], i.e. that parasites begin to leave circulation and sequester in/after age-bin 11, with a half-life of z=3 hours, such that 50% of parasites are sequestered when reaching age-bin 14. Consequently, the proportion of parasites in age-bin, b that remain in the circulation, $P_c$, is

$$P_C = 2^{\frac{11-b}{z}} \quad \text{(for } 12 \leq b \leq 48\text{)} \qquad \text{Equation 6.3}$$

and

$$P_C = 0 \quad \text{(for } 1 \leq b \leq 11\text{)} \qquad \text{Equation 6.4}$$

The number of alive iRBCs ( Equation 6.1 & Equation 6.2) in each stage can then be multiplied by the proportion of iRBCs circulating in each stage (Equation 6.3 & Equation 6.4) to calculate the number of parasites in the alive circulating and alive sequestered compartments (this is done before incorporating drug killing in that stage).

### 6.2.3 Modelling artesunate killing and stage specificity of drug killing

The duration of artesunate killing was determined as follows: First the kill rate-over-time profiles for artesunate and its active metabolite dihydroartemisinin (DHA) were calculated using standard equations [105, 180] based on the dose and an individual's PK/PD parameters **(**Table **6.1** ). The time at which the half-maximal rate of drug killing ($V_{max}/2$) occurs for both artesunate and DHA was used to estimate the duration of killing by both; duration of artesunate killing was set to whichever was longer. Artesunate is eliminated so rapidly that it decays from concentrations generating maximal killing rates to physiologically negligible concentrations very rapidly; this means that it can be regarded as being either present (and killing at maximum rate) or absent [15].

The mPK/PD model used one-hour time steps. Thus, a patient with "extra" killing over an hour would have that "extra" killing added to the next hour, such that a patient with 2.2 hours killing would have 2 hours of killing at $V_{max}$ and would have $0.2*V_{max}$ killing in the third hour. Any values of artesunate duration below 1 hour or above 12 hours (for a single dose) were deemed unrealistic and resampled.

Recall that a key objective of this research was to model the likely extent of pathology caused by sequestered iRBCs in patients treated with two alternative regimens of AS treatment i.e. The "Standard" WHO recommended regimen [181] of 2.4 mg/kg i.m., given twice in the first day and daily thereafter, i.e. at 0, 12, 24, and 48 hours (referred to by Kremsner et al. [170] as the "five dose regimen" - an additional dose is given at 72 hours), and the "Simplified" regimen as proposed by Kremsner et al. [170] (and referred to by them as the "three dose regimen") consisting of larger doses of 4 mg/kg i.m. given once each day i.e. at 0, 24 and 48 hours. Consequently, the PK parameters provided by Kremsner et al. [170] were initially utilized (**Table 6.2** ) to generate artemisinin killing durations for the patient population, but found the distribution to be an atypical shape to that which would usually be expected (**Figure 6.2**) – this is likely because it was not possible to model the correlation between the PK parameters Volume of Distribution (Vd) and Clearance (CL) using the parameters provided by Kremsner et al. [170]. Notably, however, use of their PK parameters and this distribution did later allow for the recovery of Kremsner's clinical observations (Figure 5 of [170]; see results).

3470 **Table 6.2** Pharmacokinetic (PK) parameters drawn from Kremsner et al. [170]

| Parameter | Unit | Abbreviation | Range |
|---|---|---|---|
| Volume of distribution i.m. AS [170] | [L] | $V_{AS,i.m}$ | x, where mean = 21.1 and CV = 0.97 |
| Volume of distribution i.m. DHA [170] | [L] | $V_{DHA,i.m}$ | x, where mean = 25.3 and CV = 0.81 |
| Clearance i.m. AS [170] | [L/h] | $CL_{AS,i.m}$ | x, where mean = 33.3 and CV = 0.81 |
| Clearance i.m. DHA [170] | [L/h] | $CL_{DHA,i.m}$ | x, where mean = 8.5 and CV = 0.87 |

3471

3472

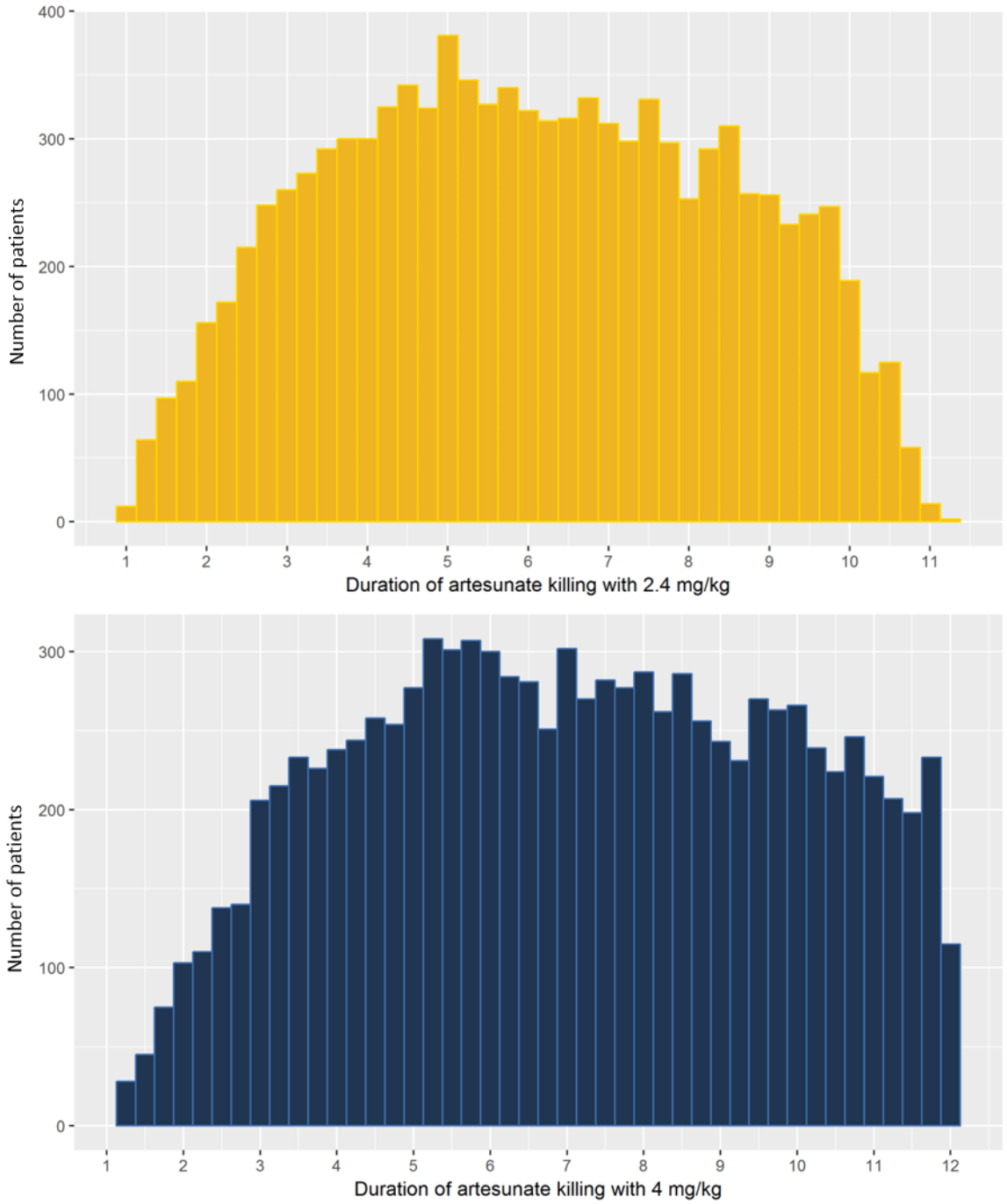



**Figure 6.2** Distribution of artemisinin killing duration obtained when using the pharmacokinetic (PK) parameters reported in Kremsner et. al [170]

As mentioned above, the atypical distribution of artemisinin killing durations obtained using the parameters in **Table 6.2** may be due to correlation between Vd and CL, but this was impossible to calculate using the parameters given in [170]. Thus, the PK parameters from Hendriksen et al. [182] were utilized, who provided the random effects that allowed for the incorporation parameter correlation (these calculations were performed for this research by Katherine Kay). Specifically, population estimates for the mean of Vd and CL were drawn from the fixed effects in Table 2 of [182] and the associated CV of these parameters from the random effects (ETA), where CV was calculated assuming they used a proportional or exponential error structure: CV = sqrt(exp(ETA)-1). The

correlation term "ηCL/F ~ ηV/F" listed under random effects was incorporated and it was assumed that Vd and CL were correlated in subsequent simulations. The final PK values used in these simulations to generate results are thus summarized in **Table 6.3.**

The resulting distribution using Hendriksen et al. [182] parameters (**Figure 6.3**) was appropriately shaped.

**Table 6.3** Pharmacokinetic (PK) parameters drawn from Hendriksen et al. [182]

| Parameter | Unit | Abbreviation | Range |
|---|---|---|---|
| Volume of distribution i.m. AS [182] | [L] | $V_{AS,i.m}$ | x, where mean = 28.2 and CV = 0.98 |
| Volume of distribution i.m. DHA [182] | [L] | $V_{DHA,i.m}$ | x, where mean = 13.5 and CV = 0.98 |
| Clearance i.m. AS (Hendriksen) | [L/h] | $CL_{AS,i.m}$ | x, where mean = 45.8 and CV = 0.71 |
| Clearance i.m. DHA (Hendriksen) | [L/h] | $CL_{DHA,i.m}$ | x, where mean = 22.4 and CV = 0.59 |
| Correlation of random effects on V and CL | | ηCL/F ARS ~ ηV/F ARS | X, where mean = 0.497 |

*AS: Artesunate, DHA: Di-hydroartemisinin*

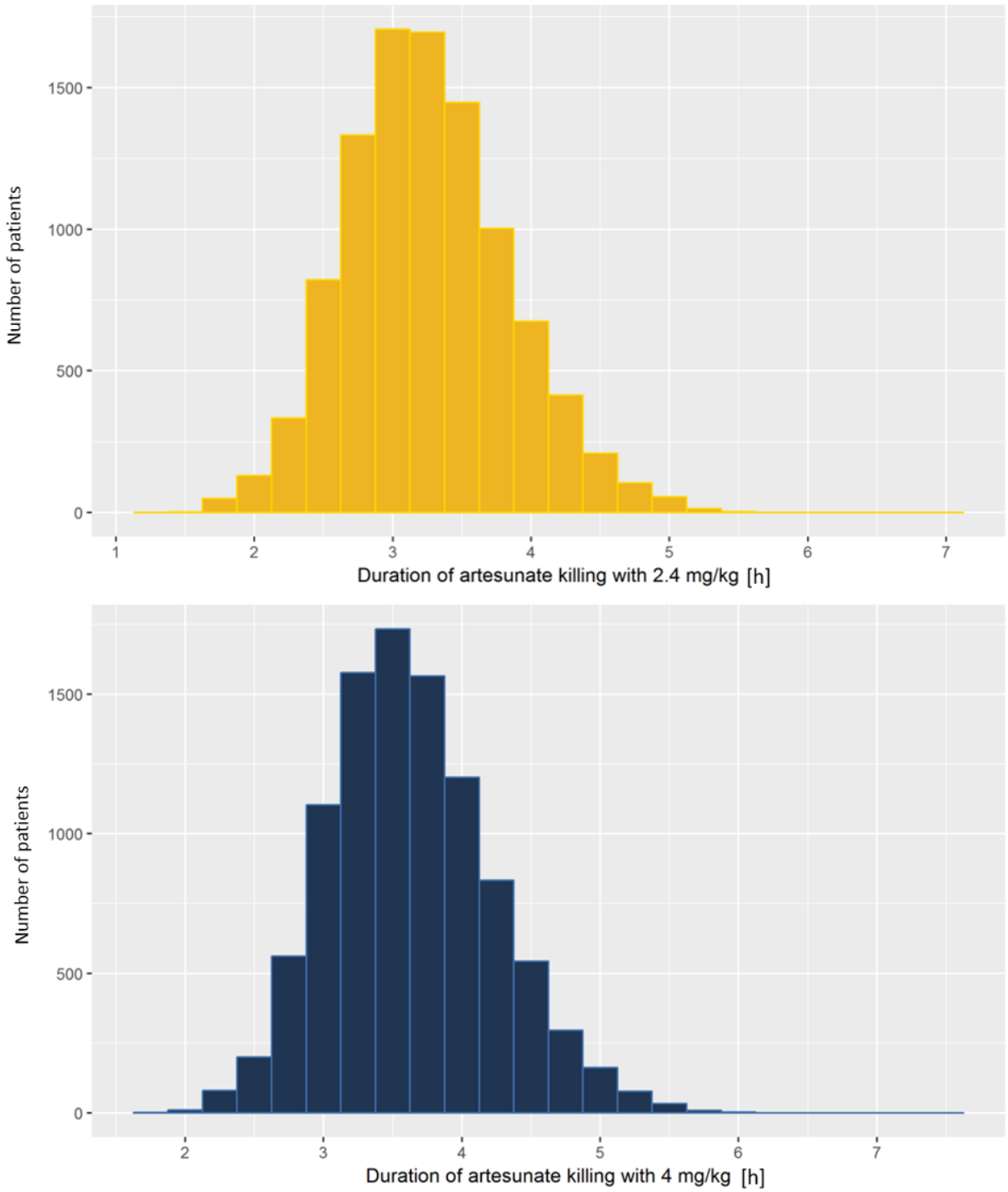



**Figure 6.3** Distribution of Artesunate killing duration. Data for 10,000 patients following treatment with a single dose of artesunate of either 2.4mg/kg (top panel) or 4mg/ml (bottom); note the duration includes that of the active metabolite dihydroartemisinin (DHA). This distribution was obtained using parameters from Hendriksen et al. [182].

Most malaria mortality occurs in children under 5 years old (and pregnant women) in areas of intense falciparum transmission in sub-Saharan Africa [183]. However, severe malaria does occur in adults in areas of low transmission, where patients have low levels of acquired immunity. The PK parameters from Hendriksen et al. used in these simulations were obtained from children [182], as were the PK parameters from Kremsner et al. [170]. To check the robustness of the model, it was adapted to replicate treatment in adults with the longer durations of artesunate killing that would be expected.

A distribution of 'adult' durations of artesunate killing were produced by increasing the duration obtained from each set of paediatric PK values by 50%, based on the assumption that:

Data from Zaloumis et al. [184] can be used to calculate artesunate clearance as: $Cl = \frac{Dose * F}{AUC}$

Using their DHA exposure figure of 2,077 h *ng/ml and assuming that a 60kg adult has an exposure of 2800 h *ng/ml, noting that F and dose are equal across groups, the ratio of CL of children: adults can be calculated as 1.35. This is not as large as a 50% increase between adults and children (as was assumed). However, combined with subsequent results of the simulations presented here, the larger duration serves to illustrate that there is little difference in the ratios of outcome metrics between children and adults, and so it was used.

This produced the distribution shown in **Figure 6.4.** Typical values of this distribution are around 4 to 6 hours artesunate killing, consistent with expectations in adults.

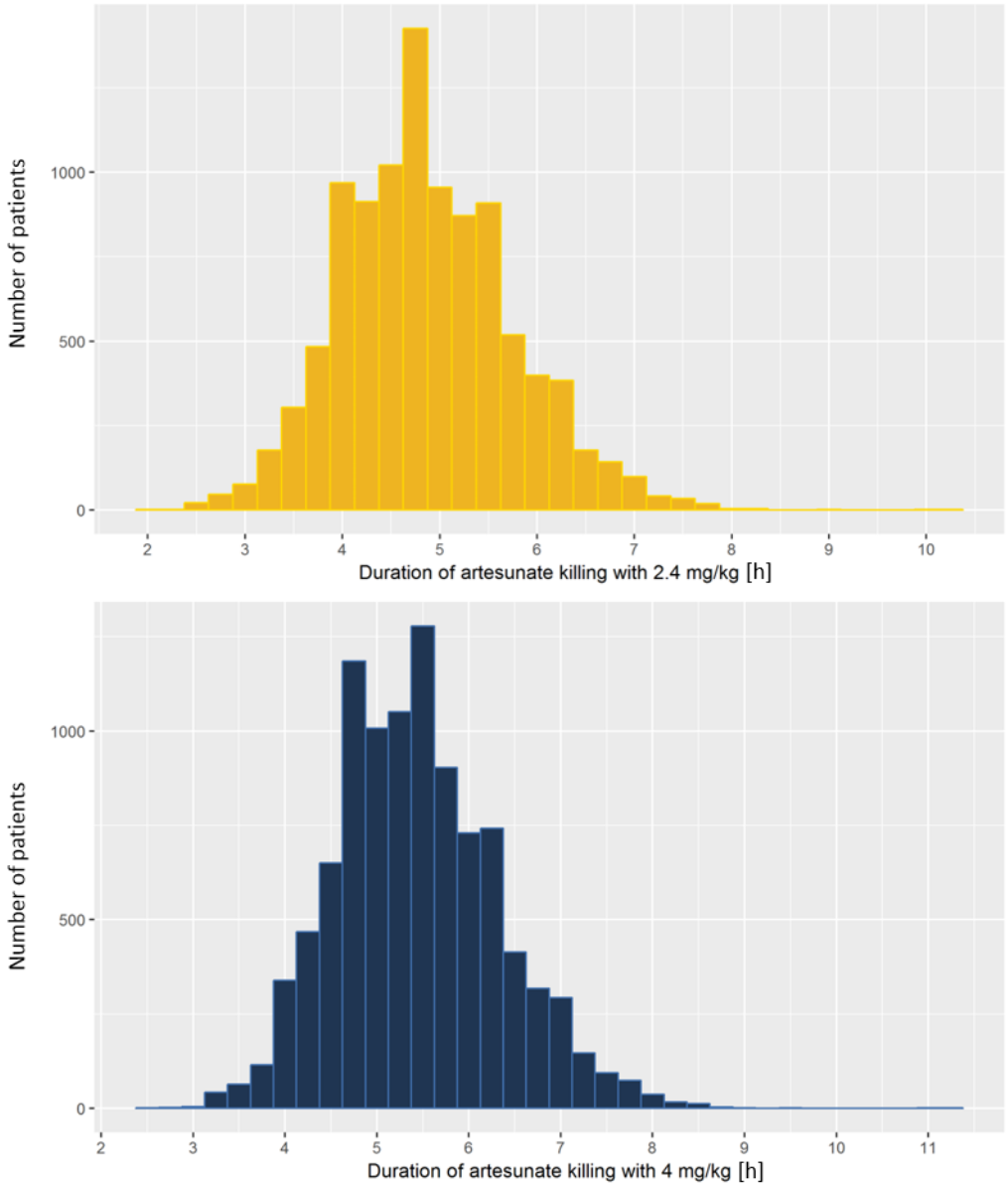



**Figure 6.4** Distribution of artemisinin killing duration obtained when using the pharmacokinetic (PK) parameters reported in Hendriksen et al. [182], and increased by 50% to better represent expected distributions for adult patients.

Stage-specific drug sensitivity was incorporated into the model as previously described [15] using their "hyper sensitive" profile to include findings from the rings-stage assay (described by [92]) which revealed parasites to be very sensitive to artemisinin just after their invasion of erythrocytes. Thus, a "sensitive" profile of parasites was calibrated by assuming a "baseline" sensitivity of parasites to artesunate in parasite age-bins 18 to 44 inclusive (Equation 6.7), hyper-sensitivity in age-bins 2 to 4 inclusive (Equation 6.5) which were ten times more sensitive than baseline, and reduced-sensitivity stages between age-bins 6 and 17 inclusive which were ten times less sensitive than baseline (Equation 6.6). Drug killing in other age-bins is 0 (Equation 6.8). The factor $D^{x\text{-}y}$ denotes drug killing between age-bins x to y inclusive and is defined as

$$D^{2\text{-}4} = 1 - e^{-d*V\ \max*10} \qquad \text{Equation 6.5}$$

$D^{6\text{-}17} = 1 - e^{-d * V_{max} * 0.1}$ Equation 6.6

$D^{18\text{-}44} = 1 - e^{-d * V_{max}}$ Equation 6.7

$D^{1} = D^{5} = D^{45-48} = 0$ Equation 6.8

where d is the duration of the time-step (i.e. one hour in this case). The mean value of $V_{max}$ is 1.78 as estimated previously [185] (**Table 6.1** ).

Recent reports of artemisinin "resistance" suggest resistance is restricted to the early "hyper-sensitive" stages [186] so the impact of this resistance was investigated by setting the artesunate sensitivity of parasites in these age-bins (i.e. 2-4) to zero, while killing in other age-bins is unchanged. This "resistant" profile was therefore described as

$D^{6\text{-}17} = 1 - e^{-d * V_{max} * 0.1}$ Equation 6.9

$D^{18\text{-}44} = 1 - e^{-d * V_{max}}$ Equation 6.10

$D^{1-5} = D^{45-48} = 0$ Equation 6.11

The mean age bin must be grouped as a categorical variable for the purposes of later carrying out Partial Rank Correlation Coefficient (PRCC) analysis; each category must be given an ordinal rank. Age-bins have different sensitivity to artesunate (Equation 6.7-Equation 6.11) and so were ranked according to the amount of drug killing that occurred in that age-bin, the category with lowest killing being the lowest rank and the category with highest killing being the highest rank. The categories are shown in **Table 6.4** for sensitive parasites and **Table 6.5** for resistant parasites.

Note that, for sensitive parasites, while age bins 45-1 and 5 have the same (lack of) killing, 45-1 is given a lower rank as it is a longer, more continuous section without killing, so it felt prudent to separate 45-1 and 5 into different categories.

**Table 6.4** Grouping of the mean age-bin parameter for sensitive parasites into ordinal categories according to the amount of killing that occurs in that age-bin, suitable for partial rank correlation coefficient (PRCC) analysis

| Age bin | Rank | Killing in age bins |
|---|---|---|
| 45-1 | 1 | None (i.e. $0*V_{max}$) |
| 5 | 2 | None (i.e. $0*V_{max}$) |
| 6-17 | 3 | $0.1*V_{max}$ |
| 18-44 | 4 | $1*V_{max}$ |
| 2-4 | 5 | $10*V_{max}$ |

**Table 6.5** Grouping of the mean age-bin parameter for resistant parasites into ordinal categories according to the amount of killing that occurs in that age-bin, suitable for partial rank correlation coefficient (PRCC) analysis

| Age bin | Rank | Killing in age bins |
|---|---|---|
| 45-5 | 1 | None (i.e. $0*V_{max}$) |
| 6-17 | 2 | $0.1*V_{max}$ |
| 18-44 | 3 | $1*V_{max}$ |

The standard deviation (SD) of the mean age-bin was constrained between 2 and 4. This results in reasonably narrow age-bin distributions and reflects existing belief and observations that severe malaria infections are synchronized [185, 187]. Mathemtically, 95% of the population fall with +/- 1.96 SD of the mean. A maximum of SD=4 therefore means that 95% of parasites fall within a 16 hour develolment perod which is one third of the 48 hour intraerythrocyte developmental cycle This is consistent with clinical observations that some clones are genetically unobservable one day (i.e, sequestered) and observable (i.e, circulating) the next day (see [187] for a complete discussion of the synchronization of parasites and note that in this model, 100% of parasites are circulating hours 1-11 then gradually sequester, see [173]).

In vivo, fever brought on by host immunity is thought to play a role in killing parasites in cases of severe malaria. Various existing models of malaria account for this effect by preventing parasite numbers reaching unfeasibly high levels [71, 188]. This regulation was induced into the model presented here by assuming "fever" and/or other immune mechanisms act to slow the multiplication rate of the parasites. This is the "parasite multiplication rate" (PMR) which was varied from 1 to 10 using a triangular distribution with mode =1 (**Table 6.1** ). Obviously, a value of 1 indicates the parasitaemia is being controlled and held constant by host mechanisms but it was important to investigate situations where parasitaemia is not completely regulated (hence PMR can go up to 10). This assumption is justified in the publications cited in Table **6.1** . Additionally, note that PMR is not a parameter that is later correlated with ratios of outcomes for either comparison of regimens or resistant / sensitive parasites.

### 6.2.4 Modelling Pathological load and pathological recovery rate.

Severity of the malaria infection is determined by a novel metric generated by this model: 'pathological load', i.e. the number of sequestered iRBCs (containing either living or dead parasites) physically restricting blood flow and/or eliciting patient's immune and/or inflammatory response that may also contribute to pathology [16, 189]. It is unlikely that the iRBC immediately ruptures on death of the parasites (which would reduce physical blockage of the capillary) or that the immune/inflammatory responses immediately disappear when the parasite dies, so it was assumed that pathology persists for a period after the death of the sequestered parasites. This effect was captured by defining a 'pathological recovery rate', r, which is the rate at which the pathology caused by sequestered iRBCs disappears with time following the death of the parasite. As will be discussed later, there are no clinical estimates of this 'recovery rate' so a strategy was undertaken to quantify the impact of dosing regimen and artemisinin resistance across a range of values of recovery rate to test whether results of the model were dependent on assumed values for recovery rate (it will be later shown that they were not). The 'recovery rate' r in the simulations was varied by altering its half-life (Table **6.1** ), which is the time it takes pathology caused by dead sequestered parasites to reduce by half. It was assumed that parasite death, with consequent rupturing of the iRBC or reduction of binding ligands (allowing iRBCs to detach from blood vessel walls), was essential to allow the start of

pathological recovery, hence sequestered iRBCs with living parasites were not subject to the pathological recovery rate. The pathological load L(t) at any time t post-treatment was quantified as the sum of the current number of sequestered iRBCs with living parasites α(t) and the lingering pathological effects of once-sequestered iRBC whose parasites were killed in the current or previous time periods, β(i), i.e.

$$L(t) = \alpha(t) + \sum_{i=1}^{t} \beta(i) e^{-(t-i)r} \qquad \text{Equation 6.12}$$

Two metrics were used to analyse treatment regimens and resistance: (i) Maximum pathological load (MPL), the maximum value of L(t) occurring during a defined time period post-treatment, and (ii) the area under the pathological load curve ($AUC_{PL}$) during a defined time period post-treatment, i.e. the total pathology in that period. For example, the $AUC_{PL}$ in the period 0 to 24 hours post-treatment is:

$$AUC_{PL} = \sum_{t=1}^{24} L(t) \qquad \text{Equation 6.13}$$

### 6.2.6 Measuring parasite clearance: Parasite reduction ratios.

The model was used to track the number of circulating and sequestered iRBCs, containing both dead and alive parasites (**Figure 6.1** ). The number of circulating parasites were used to determine the rate at which the observed (i.e. circulating) number of iRBC declined post-treatment: These metrics are often measured in clinical trials, including those of Kremsner et al. [170, 190] and allowed for comparison of the simulation against clinical data. The number of sequestered iRBCs were tracked until they are cleared from the host.

The number of circulating iRBCs containing either living or dead parasites at time t is represented as Q(t) which can be calculated as

$$Q(t) = x(t) + \sum_{i=1}^{t} Y(i) e^{-(t-i)c} \qquad \text{Equation 6.12}$$

where x(t) is the number of living parasites in circulating iRBC at time t, Y(i) is the number of parasites killed while in circulating iRBCs during time period i post treatment, and c was the rate of 'splenic' clearance of circulating iRBCs containing dead parasites (by the spleen and other possible host mechanisms such as "pitting"). The clearance half-live, u, and its equivalent rate, c, are interconverted using the formula

$$u = \frac{ln(2)}{c} \qquad \text{Equation 6.13}$$

The parasite reduction ratio (PRR) is simply the ratio of iRBC circulating at time of treatment to the number circulating at a given time of follow-up Q(t). PRR is usually measured over 48 hours ($PRR_{48}$) but the model was also later checked against the clinical observations of Kremsner et al. [170] who measured PRR over 24 hours ($PRR_{24}$).

### 6.2.7 Sensitivity analysis.

Partial rank correlation coefficient (PRCC) using Spearman's Rho was conducted to establish the strength of the relationship between model parameters and dependent variables (i.e. the pathology metrics $AUC_{PL}$ and MPL). Raman Sharma provided me with R code with which to perform the PRCC analysis described here.

Most parameters are quantitative so can enter the PRCC without modification. The exception is mean age-bin which, although numeric, has a 'circular' scale, age-bin 1 being adjacent to age-bin 48, due to parasites from ruptured iRBCs (at hour 48) reinvading to restart the asexual lifecycle. The mean age-bin variable was therefore split into either 5 or 3 ordinal classes (depending on whether parasites were hyper-sensitive or resistant to artemisinin) as described in **Table 6.4** and **Table 6.5**

The following parameters were included in the PRCC analysis:

Duration of artesunate killing post-treatment; this captures all the PK/PD parameters in Table 1 except maximal artesunate kill rate

Maximal rate of artesunate killing ($V_{max}$)

Initial mean age-bin as a categorical variable (see above)

Variation of initial age-bin distribution (measured as the standard deviation (SD) around the mean).

Initial parasite number

Parasite multiplication rate (PMR)

Half-life of the 'pathological recovery rate' (r)

The splenic clearance rate was not included in the analysis as it has no impact on sequestered iRBC based pathology.

## 6.3 Results.

### 6.3.1 Consistency of model outputs with existing field data.

The mPK/PD model calculated parasite reduction ratios (PRR) from circulating parasite numbers to allow for direct comparison of the model to published clinical data. The clinical endpoint of the trials by Kremsner and colleagues was the proportion of patients in each arm whose PRR at 24 hours ($PRR_{24}$) was >99% [170], reported as 79% and 78% for the five-dose standard and the three-dose simplified regimen, respectively. When calibrated with PK parameters from Kremsner's study [170], the modelled results here were consistent with these clinical observations, i.e. the model predicted 78% and 74% for the standard and simplified regimen with hyper-sensitive parasites, respectively (**Table 6.6**). However, the results presented in later in this chapter are calibrated using PK parameters from Hendriksen et al. [182] (**6.2.3**), with which lower values of 70% and 62% of patients with $PRR_{24}$>99% were observed for the standard and simplified i.m regimens, respectively.

Hendriksen et al. [182] did not report the percentage of patients with $PRR_{24}$ > 99% in their study, so it was not possible to simultaneously compare the findings of the model presented here with the findings of Kremsner et al. [170] and Hendriksen et al. [182]. However, Hendriksen et al. [182] reported the population geometric mean of the fractional reduction in parasite counts at 24 hours as 96% (94-98%, 95% CI) following treatment with the standard regimen. The population geometric mean

obtained for the reduction in parasite counts at 24 hours (i.e. $PRR_{24}$) in these simulation using parameters from Hendriksen et al. [182] was >99%.

The general accepted value for $PRR_{48}$ following artemisinin treatment is $10^{-4}$ [191] which is very close to the value obtained here: For the standard regimen, using the artesunate duration derived from Hendriksen's PK parameters a mean $PRR_{48}$ of $5.18^{-5}$ was obtained (**Table 6.6**).

**Table 6.6** Clinical outcome (Parasite Reduction Ratio, PRR) observed in simulations using three different pharmacokinetic (PK) parameterizations. (from Kremsner et al. [170], Hendriksen et al. [182], and Hendriksen et al. [182] with increased duration).

| Parameter choice | Kremsner et. al [170] | | Hendriksen et. al [182] | | Hendriksen et. al [182] (50% increase in artesunate duration) | |
|---|---|---|---|---|---|---|
| Regimen | Standard | Simplified | Standard | Simplified | Standard | Simplified |
| Artesunate duration figure | **Figure 6.2** | | **Figure 6.3** | | **Figure 6.4** | |
| % of patients with >99% PRR at 24h | 78% | 74% | 70% | 62% | 77% | 72% |
| Population geometric mean PRR at 48h | $5.56^{-6}$ | $2.33^{-5}$ | $5.18^{-5}$ | 0.0009 | $5.21^{-6}$ | $5.28^{-5}$ |

*PRR: Parasite Reduction Ratio*

### 6.3.2 Comparison of pathological load metrics.

The model calculated pathological load and returns two outcome metrics: $AUC_{PL}$ and MPL. **Figure 6.5** shows the values of these metrics for 3 model scenarios: Patients with sensitive parasites treated with the standard WHO regimen, a comparison of the ratios of $AUC_{PL}$ and MPL for treatment with simplified regimen v standard regimen, and the impact or artemisinin resistance on outcomes following treatment with standard WHO regimen.

Ratios of outcome metrics are calculated as simplified regimens scaled by standard regimen and as resistant parasites scaled by sensitive parasites. High metrics are deleterious, thus ratios of >1 indicate worse prognosis associated with the simplified or resistant parasites. These ratios quantify the impact e.g. a ratio of 5 for resistant vs sensitive parasites indicates pathological metrics are 5 times higher when treating resistant parasites. Four time periods post-treatment were investigated: 0-12h, 0-24h, 12-24h and 24-48h.

#### 6.3.2.1 Standard regimen treatment of artemisinin-sensitive parasites ("baseline scenario").

Treatment of drug-sensitive parasites with the standard regimen was simulated, and the key drivers of pathology were identified by calculating which parameters were most correlated with absolute values of $AUC_{PL}$ and MPL (**Figure 6.6**, **Table 6.7**). The most highly correlated parameter for both metrics was the initial parasite number: Large positive PRCCs (between 0.88 and 0.98) were observed with associated p values <=0.001 at all time-periods. The half-life of the recovery rate r had PRCC of 0.46 for $AUC_{PL}$ and 0.34 for MPL in the 24-48h time-period (p values <=0.001), but PRCC of <0.3 in earlier time periods. All other parameters had PRCC values of <0.3, indicating that outcome metrics were not highly correlated as per accepted statistical criteria [192]. All other model parameters had negligible correlation. The most likely explanation is that such a large proportion of parasites are killed by artesunate that small differences in number killed are negligible compared to the initial parasite number and pathological recovery rate.

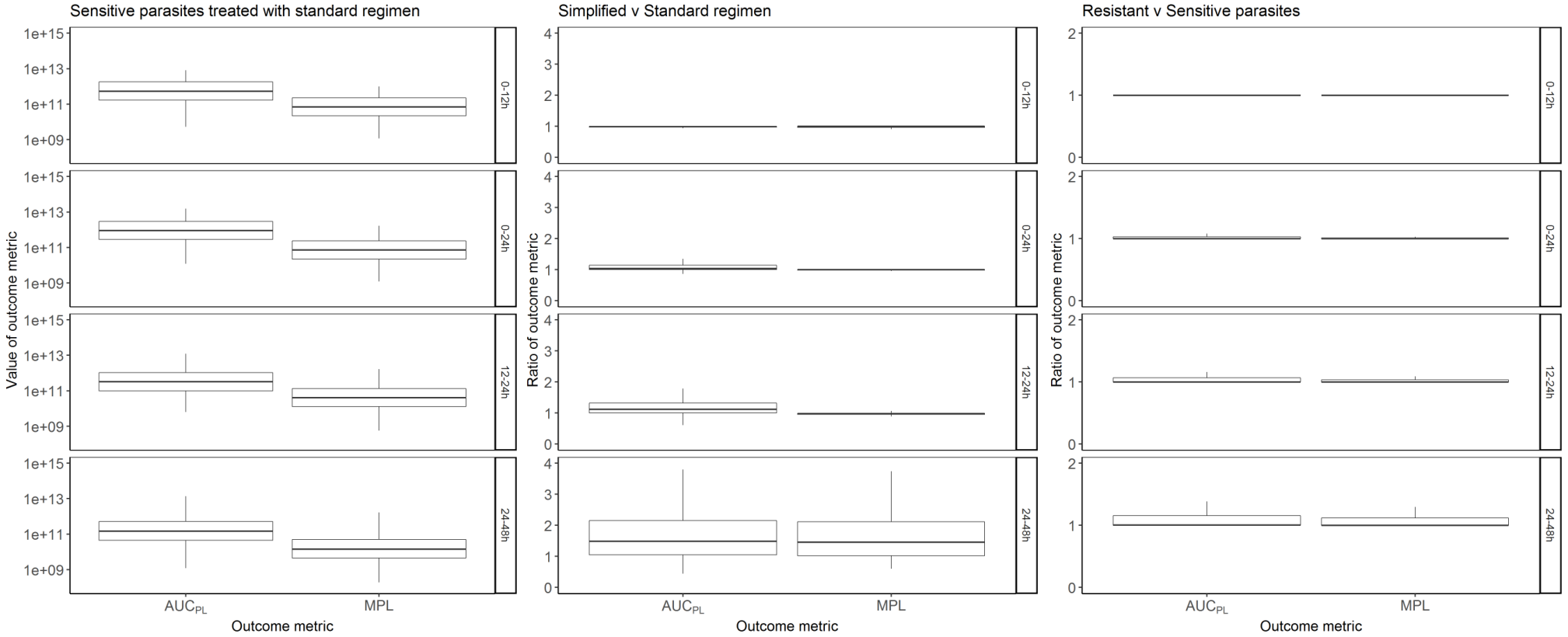
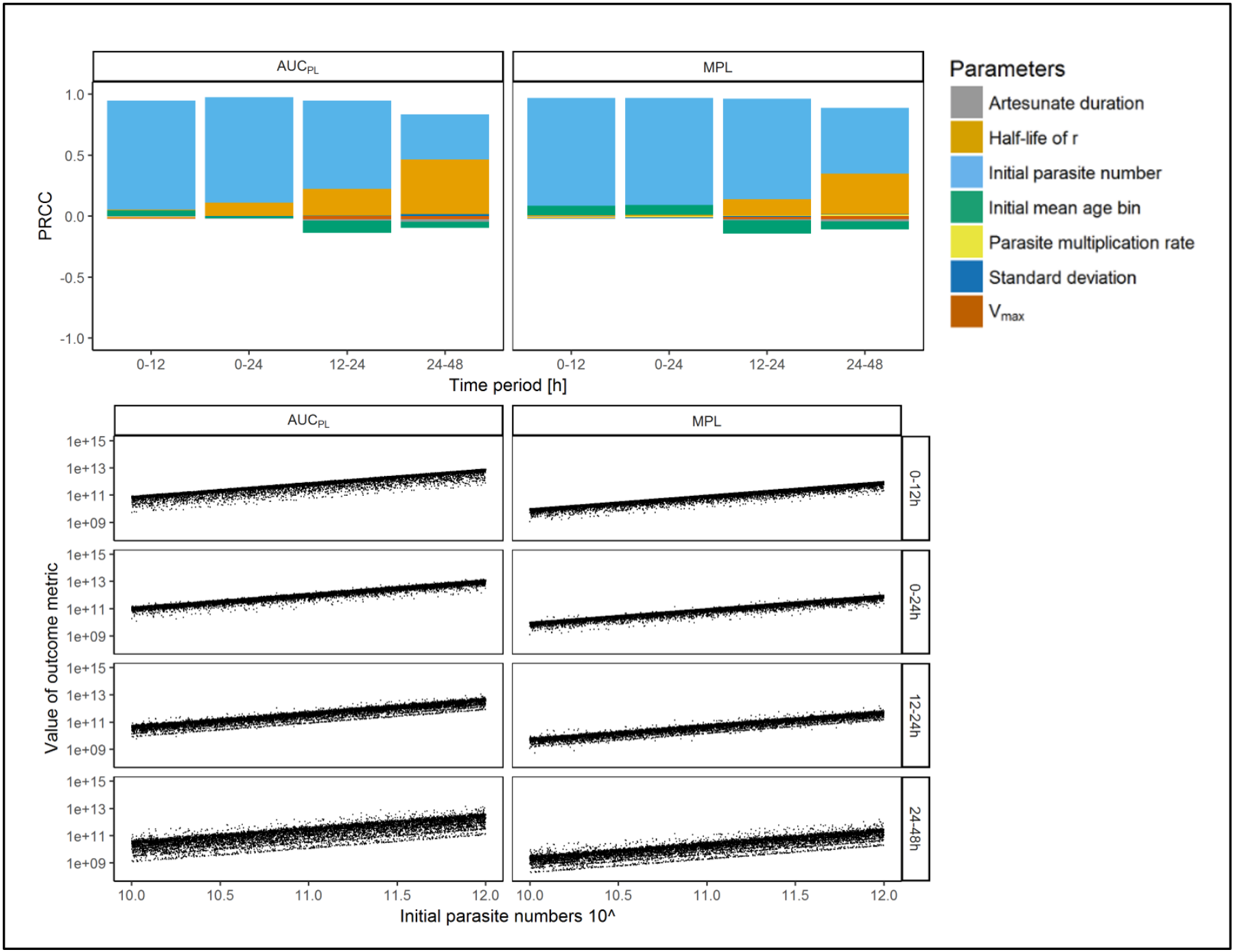



**Figure 6.5** Values or ratios of area under the pathological load curve ($AUC_{PL}$) and maximum pathological load (MPL) obtained for each of 3 model scenarios across four time periods post-treatment: 0-12h, 0-24h, 12-24h and 24-48h. The left-hand plot shows absolute values of $AUC_{PL}$ and (MPL in the "baseline scenario" when artemisinin sensitive parasites are treated with the standard regimen. The centre plot shows a comparison of the simplified v standard regimen (ratios >1 show the standard regimen is superior). The right-hand plot shows a comparison of the standard regimen when used to treat resistant v sensitive parasites (ratios >1 show that sensitive parasites produce better outcomes).


AUC_PL
MPL
PRCC
1.0
0.5
0.0
-0.5
-1.0
0-12
0-24
12-24
24-48
Time period [h]
Parameters
Artesunate duration
Half-life of r
Initial parasite number
Initial mean age bin
Parasite multiplication rate
Standard deviation
V_max
0-12h
0-24h
12-24h
24-48h
Value of outcome metric
1e+15
1e+13
1e+11
1e+09
10.0
10.5
11.0
11.5
12.0
Initial parasite numbers 10^

3730

**Figure 6.6** Analysis of the baseline scenario. The impact of underlying factors on the standard World Health Organization (WHO) regimen used to treat patients with artemisinin-sensitive parasites. Top panel: Partial rank correlation coefficients (PRCC) using Spearman's Rho of model parameters on values of area under the pathological load curve ($AUC_{PL}$) and maximum pathological load (MPL) obtained from a population. Lower panel: Values of $AUC_{PL}$ and MPL are plotted against the most highly correlated parameter, i.e. initial parasite number, for four time periods post-treatment.

**Table 6.7** Partial rank correlation coefficient (PRCC) values with corresponding p values (brackets) for values of $AUC_{PL}$ and MPL for a patient population simulated with sensitive parasites and treated with the standard regimen, using seven key model parameters.

| Outcome Metric | Time period | Parameter | | | | | | |
|---|---|---|---|---|---|---|---|---|
| | | Initial parasite number | Initial mean age-bin | Standard deviation | PMR | $V_{max}$ | Half-life of r | Artesunate duration |
| $AUC_{PL}$ | 0-12h | 0.95 (<0.001) | 0.05 (<0.001) | 0.001 (0.71) | -0.012 (0.003) | -0.02 (<0.001) | 0.05 (<0.001) | -0.004 (0.27) |
| | 0-24h | 0.97 (<0.001) | -0.01 (<0.001) | -0.001 (0.77) | -0.003 (0.386) | -0.02 (<0.001) | 0.11 (<0.001) | -0.02 (<0.001) |
| | 12-24h | 0.95 (<0.001) | -0.14 (<0.001) | 0.002 (0.52) | 0.003 (0.34) | -0.02 (<0.001) | 0.22 (<0.001) | -0.04(<0.001) |
| | 24-48h | 0.83 (<0.001) | -0.09 (<0.001) | 0.015 (<0.001) | 0.015 (<0.001) | -0.02 (<0.001) | 0.46 (<0.001) | -0.04 (<0.001) |
| MPL | 0-12h | 0.97 (<0.001) | 0.08 (<0.001) | -0.02 (<0.001) | -0.013 (<0.001) | -0.02 (<0.001) | 0.008 (0.03) | -0.005 (0.17) |
| | 0-24h | 0.97 (<0.001) | 0.09 (<0.001) | -0.02 (<0.001) | -0.007 (0.06) | -0.02 (<0.001) | 0.01 (0.01) | -0.01 (0.01) |
| | 12-24h | 0.096 (<0.001) | -0.14 (<0.001) | -0.005 (0.21) | 0.003 (0.47) | -0.03 (<0.001) | 0.14 (<0.001) | -0.03 (<0.001) |
| | 24-48h | 0.89 (<0.001) | -0.1 (<0.001) | 0.02 (<0.001) | 0.01 (0.001) | -0.02 (<0.001) | 0.35 (<0.001) | -0.04 (<0.001) |

*PRCC: Partial Rank Correlation Coefficient, $AUC_{PL}$: Area under the pathological load curve, MPL: Maximum value of pathological load, PMR: Parasite multiplication rate, $V_{max}$ : Maximal rate of artesunate killing, r: pathological load recovery rate.*

### 6.3.2.2 Comparison of simplified and standard regimen.

Alternative treatment regimens on artemisinin-sensitive parasites were evaluated using the model. These results are presented as ratios of $AUC_{PL}$ and MPL. The simplified regimen had a slightly higher median ratio in 0-24h of 1.03; MPL was 1. At 24-48h, higher medians of 1.49 and 1.45 for $AUC_{PL}$ and MPL respectively were observed (**Figure 6.5**; **Table 6.8**). There was negligible difference in outcome metrics between assumption of paediatric artesunate killing duration and adult artesunate killing duration (**Table 6.8**)

**Table 6.8** Median ratios of the area under the pathological load curve ($AUC_{PL}$) and maximum pathological load (MPL) across the patient population for 4 time-periods and for comparison of standard and simplified regimen with sensitive parasites and for sensitive and resistant parasites treated with standard regimen. Parameters are from Hendriksen et al. [182] (top; assuming paediatric artesunate duration) and from Hendriksen et al. with 50% increase in artesunate killing duration (bottom, inside square brackets and assuming adult artesunate killing duration). A ratio of <1 indicates lower metric with the simplified regimen or resistant parasites respectively:

| Median ratios for standard v simplified regimen with sensitive parasites | | | | |
|---|---|---|---|---|
| Time (h) | 0-12 | 0-24 | 12-24 | 24-48 |
| $AUC_{PL}$ | 0.99<br>[0.99] | 1.03<br>[1.007] | 1.12<br>[1.03] | 1.49<br>[1.18] |
| MPL | 1<br>[1] | 1<br>[1] | 0.98<br>[0.98] | 1.45<br>[1.17] |
| Median ratios for sensitive v resistant parasites, treated with standard regimen | | | | |
| Time (h) | 0-12 | 0-24 | 12-24 | 24-48 |
| $AUC_{PL}$ | 1.000003<br>[1.000003] | 1.000203<br>[1.000161] | 1.000615<br>[1.000505] | 1.006516<br>[1.00208] |
| MPL | 1<br>[1] | 1<br>[1] | 1.000092<br>[1.000081] | 1.002546<br>[1.001053] |

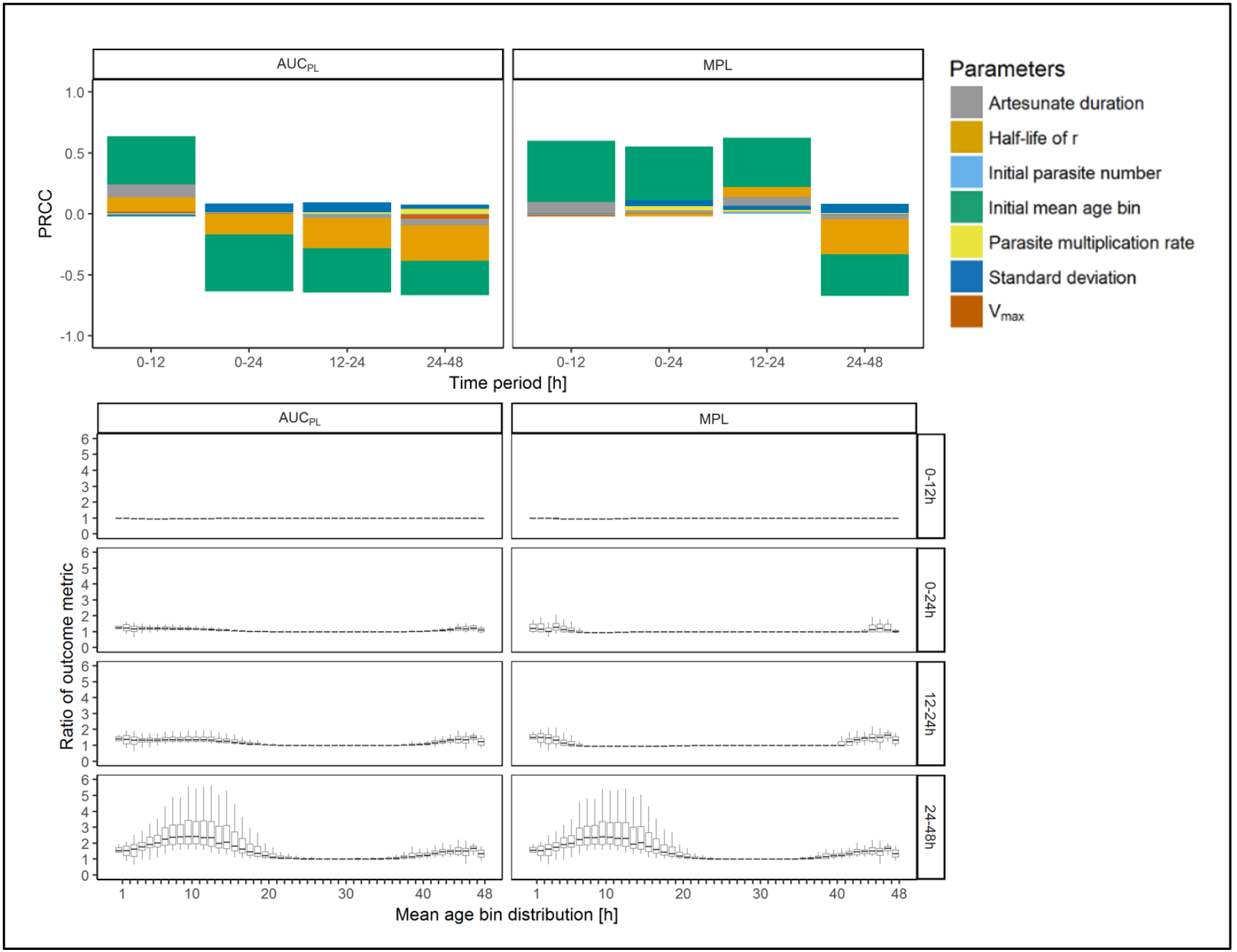
*$AUC_{PL}$: Area under the pathological load curve, MPL: maximum pathological load*

Parameter analysis with PRCC (**Figure 6.7**, **Table 6.9**) revealed that patients whose initial infections were in either very late or very early initial mean age-bins will have worse outcomes with the simplified regimen. This occurred because parasites in these stages are largely insensitive to artesunate at first treatment, and the simplified regimen lacks the second dose, 12 hours later, of the standard regimen that would effectively target these parasites that had matured into more artemisinin sensitive age-bins.

The half-life of the recovery rate r had a moderate correlation with outputs in the 12-24h and 24-48h periods indicating that assumption of slower recovery made the simplified regimen perform relatively better (Figure 6.8). This parameter does not appear to affect the validity of results; for complete

discussion see **6.3.3**. No other parameters have notable correlation with sequestration-based pathology when comparing regimens. This is probably because they "cancel out" as explained above e.g. initial parasite numbers is the same within patients thus cancels when comparing the impact of different regimens within the same patient

This analysis was repeated to compare regimens when treating resistant (as opposed to drug-sensitive) parasites. Results were extremely similar to those for comparison of regimens with drug-sensitive parasites (shown in **Figure 6.7**) and are displayed in **Figure 6.9** and **Table 6.10**.


AUC$_{PL}$
MPL
PRCC
1.0
0.5
0.0
-0.5
-1.0
0-12
0-24
12-24
24-48
Time period [h]
Parameters
Artesunate duration
Half-life of r
Initial parasite number
Initial mean age bin
Parasite multiplication rate
Standard deviation
$V_{max}$
0-12h
0-24h
12-24h
24-48h
Ratio of outcome metric
1
10
20
30
40
48
Mean age bin distribution [h]

3796

**Figure 6.7** Evaluation of alternative drug treatment regimens. Comparison of the simplified v. World Health Organization (WHO) standard regimen for treatment of artemisinin-sensitive parasites; ratios of >1 indicate the simplified regimen produces worse outcome metrics. Top panel: Partial rank correlation coefficients (PRCC) using Spearman's Rho of model parameters on the ratios of area under the pathological load curve ($AUC_{PL}$) and maximum pathological load (MPL). Lower panel: Ratios of $AUC_{PL}$ and MPL are plotted against the most highly correlated parameter (initial mean age bin), for four time periods post-treatment.

**Table 6.9** PRCC values with corresponding p values (brackets) for **ratios** of area under the pathological load curve ($AUC_{PL}$) and maximum pathological load (MPL) for the simplified v standard regimen for a patient population simulated with sensitive parasites using seven key model parameters. The ratio is calculated as $\frac{Outcome\ metric\ of\ simplified\ regimen}{Outcome\ metric\ of\ standard\ regimen}$ such that higher ratios (and thus, positive correlation) indicate better performance of the standard regimen.

| Outcome Metric | Time period | Parameter | | | | | | |
|---|---|---|---|---|---|---|---|---|
| | | Initial parasite number | Initial mean age-bin | Standard deviation | PMR | $V_{max}$ | Half-life of r | Artesunate duration |
| $AUC_{PL}$ | 0-12h | 0.007 (0.07) | 0.64 (<0.001) | -0.02 (<0.001) | -0.006 (0.1) | 0.02 (<0.001) | 0.13 (<0.001) | 0.24 (<0.001) |
| | 0-24h | -0.003 (0.41) | -0.63 (<0.001) | 0.09 (<0.001) | 0.013 (<0.001) | 0.007 (0.06) | -0.16 (<0.001) | 0.009 (0.02) |
| | 12-24h | -0.003 (0.35) | -0.64 (<0.001) | 0.1 (<0.001) | 0.012 (0.002) | -0.006 (0.1) | -0.28 (<0.001) | -0.03 (<0.001) |
| | 24-48h | -0.005 (0.24) | -0.66 (<0.001) | -0.08 (<0.001) | 0.04 (<0.001) | -0.04 (<0.001) | -0.38 (<0.001) | -0.09 (<0.001) |
| MPL | 0-12h | -0.009 (0.02) | 0.59 (<0.001) | 0.003 (0.4) | -0.00005 (0.89) | -0.02 (<0.001) | -0.003 (0.52) | 0.1 (<0.001) |
| | 0-24h | 0.002 (0.58) | 0.55 (<0.001) | 0.11 (<0.001) | 0.06 (<0.001) | 0.002 (0.52) | -0.02 (<0.001) | 0.03 (<0.001) |
| | 12-24h | 0.01 (0.002) | 0.62 (<0.001) | 0.06 (<0.001) | 0.03 (<0.001) | 0.01 (<0.001) | 0.22 (<0.001) | 0.14 (<0.001) |
| | 24-48h | -0.006 (0.15) | -0.67 (<0.001) | 0.08 (<0.001) | 0.007 (0.08) | -0.008 (0.04) | -0.33 (<0.001) | -0.04 (<0.001) |

*PRCC: Partial Rank Correlation Coefficient, $AUC_{PL}$: Area under the pathological load curve, MPL: Maximum value of pathological load, PMR: Parasite multiplication rate, $V_{max}$: Maximal rate of artesunate killing, r: pathological load recovery rate.*

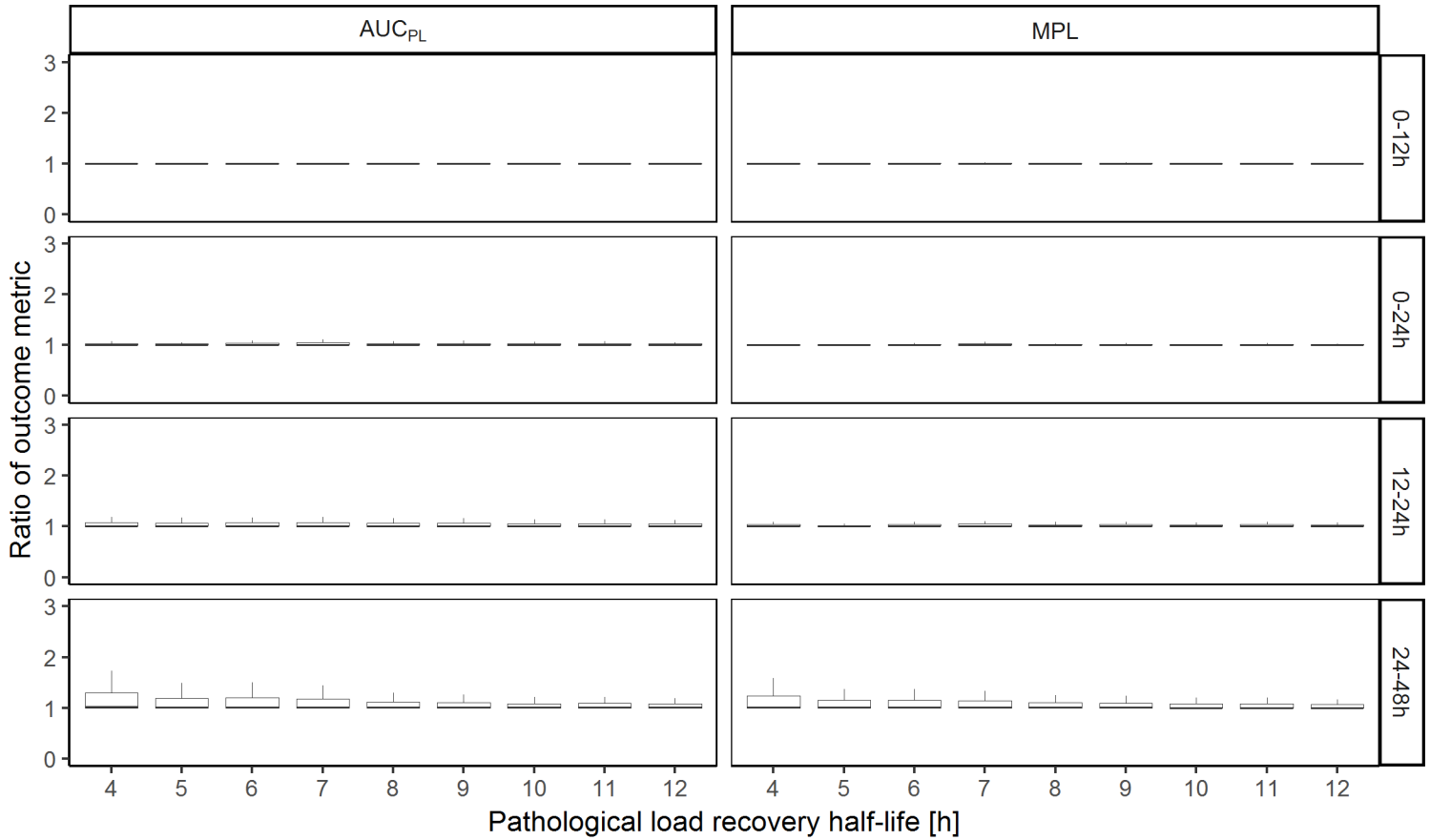



**Figure 6.8** Ratios of area under the pathological load curve ($AUC_{PL}$) and maximum pathological load (MPL) for sensitive v resistant parasites treated with standard regimen, plotted according to half-life of the pathological load recovery rate r across four time periods post-treatment: 0-12h, 0-24h, 12-24h and 24-48h.

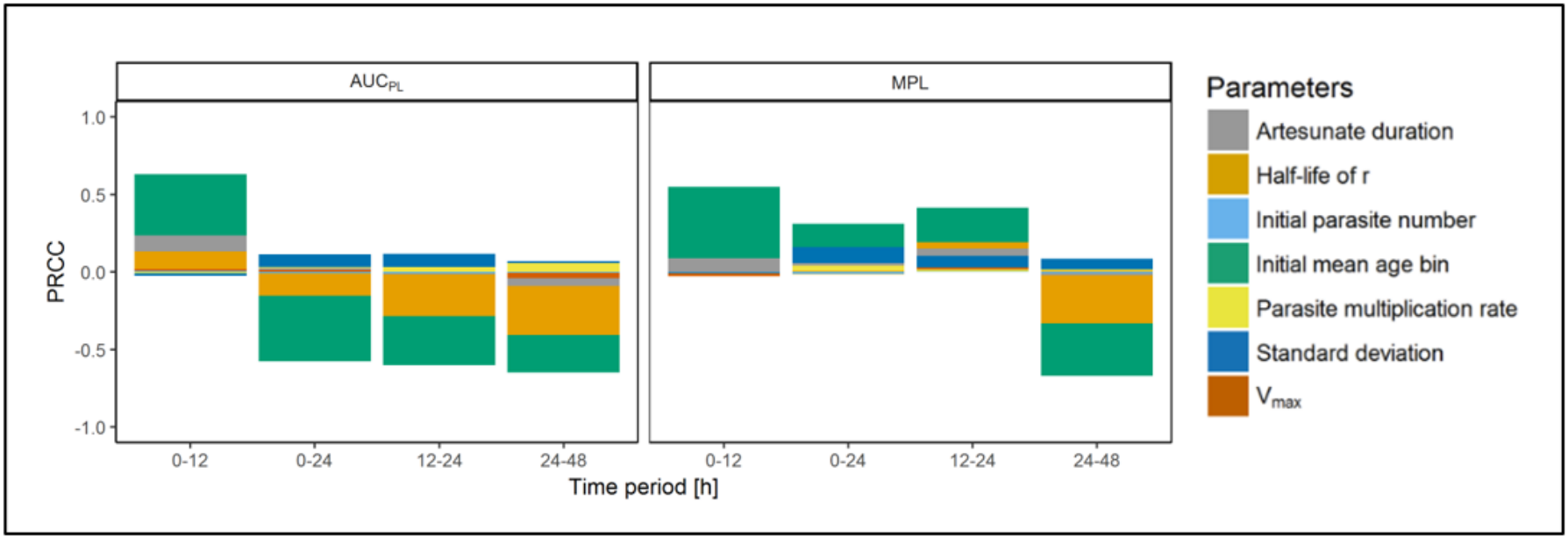



**Figure 6.9** Partial rank correlation coefficient (PRCC) analysis using Spearman's Rho of model parameters on the ratios of area under the pathological load curve ($AUC_{PL}$) and maximum pathological load (MPL) for the standard v the simplified regimen when parasites are resistant. A PRCC of over 0.3 (+ or -) indicates that the parameter has notable correlation with the ratio.

**Table 6.10** Partial rank correlation coefficient (PRCC) values with corresponding p values (brackets) for ratios of area under the pathological load curve ($AUC_{PL}$) and maximum pathological load (MPL) for the simplified v standard regimen for a patient population simulated with resistant parasites using seven key model parameters. The ratio is calculated as $\frac{Outcome\ metric\ of\ simplified\ regimen}{Outcome\ metric\ of\ standard\ regimen}$ such that higher ratios (and thus, positive correlation) indicate better performance of the standard regimen.

| Outcome Metric | Time period | Parameter | | | | | | |
|---|---|---|---|---|---|---|---|---|
| | | Initial parasite number | Initial mean age-bin | Standard deviation | PMR | $V_{max}$ | Half-life of r | Artesunate duration |
| $AUC_{PL}$ | 0-12h | 0.007 (0.05) | 0.63 (0) | -0.02 (<0.001) | -0.07 (0.08) | 0.019 (<0.001) | 0.13 (<0.001) | 0.23 (0) |
| | 0-24h | -0.007 (0.08) | -0.57 (0) | 0.11 (<0.001) | 0.03 (<0.001) | 0.01 (<0.001) | -0.15 (2.22) | 0.02 (<0.001) |
| | 12-24h | -0.008 (0.04) | -0.6 (0) | 0.11 (<0.001) | 0.03 s(<0.001) | $-3.45^{-5}$ (0.99) | -0.28 (0) | -0.01 (<0.001) |
| | 24-48h | -0.006 (0.11) | -0.64 (0) | 0.07 (<0.001) | 0.05 (<0.001) | -0.03 (<0.001) | -0.4 (0) | -0.09 (<0.001) |
| MPL | 0-12h | -0.002 (0.48) | 0.54 (0) | -0.008 (0.03) | -0.002 (0.47) | -0.02 (<0.001) | 0.001 (0.73) | 0.08 (<0.001) |
| | 0-24h | -0.01 (0.002) | 0.31 (0) | 0.16 (0) | 0.04 (<0.001) | 0.003 (0.44) | -0.01 (<0.001) | 0.05 (<0.001) |
| | 12-24h | 0.002 (0.47) | 0.41 (0) | 0.1 (<0.001) | 0.015 (<0.001) | 0.027 (<0.001) | 0.19 (0) | 0.15 (0) |
| | 24-48h | -0.007 (0.06) | -0.66 (0) | 0.085 (<0.001) | 0.017 (<0.001) | 0.003 (0.41) | -0.33 (0) | -0.01 (<0.001) |

*PRCC: Partial Rank Correlation Coefficient, $AUC_{PL}$: Area under the pathological load curve, MPL: Maximum value of pathological load, PMR: Parasite multiplication rate, $V_{max}$ : Maximal rate of artesunate killing, r: pathological load recovery rate.*

### 6.3.2.3 The impact of artemisinin resistance on treatment by the standard regimen.

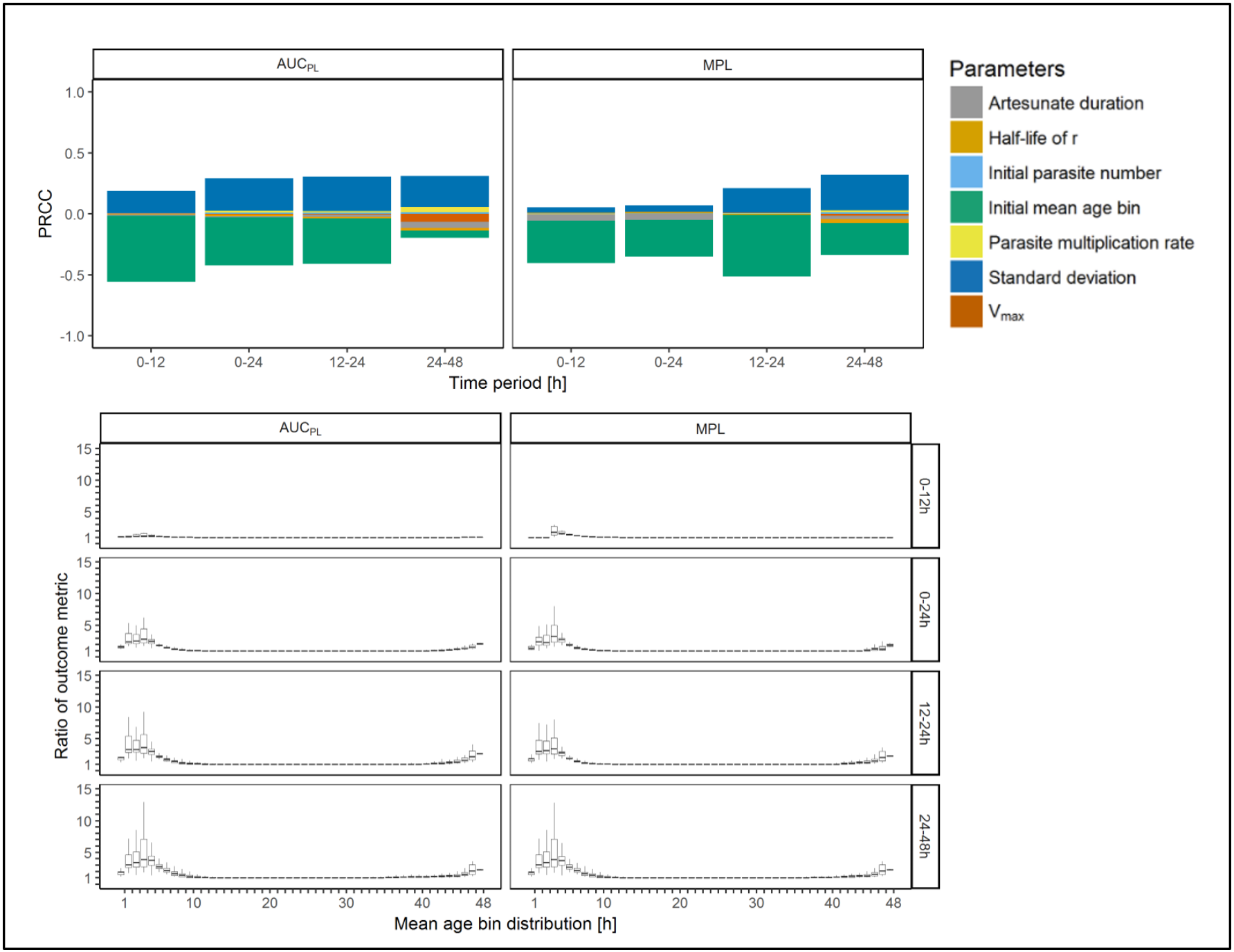
Unsurprisingly ratios of $AUC_{PL}$ and MPL when comparing resistant and sensitive parasites are never less than 1 (**Figure 6.5**) i.e. under no circumstance did patients have a better outcome when parasites are resistant. Differences in median values (**Figure 6.5**, **Table 6.8**) were extremely small (in other words, while better outcomes were never observed with resistant parasites, the impact of resistance across the entire 10,000 patient population seemed negligible.

PRRC analysis (**Figure 6.10**,**6.3.3.1**) was conducted to investigate whether this small difference obscured the presence of a vulnerable sub-group of patients. This appeared to be the case: Patients whose infections are clustered in the early age-bins at time of treatment had pathological outcomes which were significantly worse in the presence of resistance (**Figure 6.10**).

In these early age bins, ratios for $AUC_{PL}$ and MPL are as high as 5 in the 0-24h period (comparisons based on the upper quartile value). This occurs because artesunate presence post-treatment largely coincides with parasites in age-bins insensitive to artesunate through resistance, rendering the initial dose nearly or completely ineffective.

SD of the initial mean age-bin had a positive correlation with the ratio (indicating that resistant parasites had worse outcomes as SD increased). This occurred because higher SD "nudged" parts of the age-bin distribution into (or out of) resistant age-bins (i.e. the contiguous bin 45-48 and 1-5 where killing is absent). PRCC analysis showed no other parameter had a PRCC value of >0.01, suggesting the initial mean age-bin (and, to a lesser extent, it's SD) are the sole determinants of whether a patient's outcome will be worse in the presence of resistance.

There was negligible difference in outcome metrics when comparing sensitive and resistant parasites between assumption of paediatric artesunate killing duration and adult artesunate killing duration (**Table 6.8**)


AUC$_{PL}$
MPL
PRCC
1.0
0.5
0.0
-0.5
-1.0
0-12
0-24
12-24
24-48
Time period [h]
Parameters
Artesunate duration
Half-life of r
Initial parasite number
Initial mean age bin
Parasite multiplication rate
Standard deviation
$V_{max}$
AUC$_{PL}$
MPL
0-12h
0-24h
12-24h
24-48h
Ratio of outcome metric
15
10
5
1
1
10
20
30
40
48
Mean age bin distribution [h]

3870

**Figure 6.10** Analysis of the impact of artemisinin-resistance. The effectiveness of the World Health Organization (WHO) standard regimen used to treat resistant v sensitive parasites; ratios of >1 indicate that resistant parasites have worse outcome metrics. Top panel: Partial rank correlation coefficients (PRCC) using Spearman's Rho of model parameters on the ratios of area under the pathology curve ($AUC_{PL}$) and maximum pathological load (MPL) Lower panel: Ratios of $AUC_{PL}$ and MPL are plotted against the most highly correlated parameter (mean age bin), for four time periods post-treatment.

**Table 6.11** Partial rank correlation coefficient (PRCC) values with corresponding p values (brackets) for **ratios** of area under the pathological load curve ($AUC_{PL}$) and maximum pathological load (MPL) for resistant v sensitive parasites for a patient population simulated treated with the standard regimen, using seven key model parameters. The ratio is calculated as $\frac{Outcome\ metric\ of\ resistant\ parasites}{Outcome\ metric\ of\ sensitive\ parasites}$ such that higher ratios (and thus, positive correlation) indicate better outcomes with sensitive parasites

| Outcome Metric | Time period | Parameter | | | | | | |
|---|---|---|---|---|---|---|---|---|
| | | Initial parasite number | Initial mean age-bin | Standard deviation | PMR | $V_{max}$ | Half-life of r | Artesunate duration |
| $AUC_{PL}$ | 0-12h | 0.003 (0.51) | -0.55 (<0.001) | 0.18 (<0.001) | 0.004 (0.32) | 0.009 (0.03) | -0.005 (0.18) | -0.01 (<0.001) |
| | 0-24h | 0.009 (0.03) | -0.42 (<0.001) | 0.29 (<0.001) | 0.025 (<0.001) | -0.005 (0.15) | -0.02 (<0.001) | -0.03 (<0.001) |
| | 12-24h | 0.01 (0.02) | -0.4 (<0.001) | 0.3 (<0.001) | 0.24 (<0.001) | -0.006 (0.15) | -0.04 (<0.001) | -0.02 (<0.001) |
| | 24-48h | 0.01 (0.001) | -0.19 (<0.001) | 0.31 (<0.001) | 0.06 (<0.001) | -0.07 (<0.001) | -0.13 (<0.001) | -0.11 (<0.001) |
| MPL | 0-12h | 0.003 (0.36) | -0.4 (<0.001) | 0.05 (<0.001) | -0.0007 (0.87) | -0.001 (0.71) | 0.01 (0.02) | -0.05 (<0.001) |
| | 0-24h | 0.002 (0.52) | -0.35 (<0.001) | 0.06 (<0.001) | 0.02 (<0.001) | -0.0001 (0.97) | 0.02 (<0.001) | -0.05 (<0.001) |
| | 12-24h | 0.004 (0.3) | -0.51 (<0.001) | 0.21 (<0.001) | 0.007 (<0.001) | 0.01 (0.02) | -0.01 (0.04) | -0.0002 (0.96) |
| | 24-48h | 0.01 (0.002) | -0.33 (<0.001) | 0.32 (<0.001) | 0.03 (<0.001) | -0.01 (<0.001) | -0.1 (<0.001) | -0.05 (<0.001) |

*PRCC: Partial Rank Correlation Coefficient, $AUC_{PL}$: Area under the pathological load curve, MPL: Maximum value of pathological load, PMR: Parasite multiplication rate, $V_{max}$: Maximal rate of artesunate killing, r: pathological load recovery rate*

### 6.3.3 Impact of the recovery rate r (and its half-life).

The key unknown parameter in these simulations is the pathological recovery rate r. It's value was varied by altering its half-life, i.e. the amount of time it took pathology from sequestered parasites to reduce by half, following the death of the parasite. There are no clinical estimates of this parameter, so it was arbitrarily varied between 4 and 12 hours in these simulations. Most death occurs within 24 hours of admission, so a 12-hour half-life as the upper limit was selected because, if pathology lingered longer, then presumably more deaths would be expected in the post-24hour period. Four hours was taken as the minimum because under this assumption 50% of the pathology disappears by 4 hours, 75% by 8 hours, 87.5% by 12 hours, and this seemed rapid. Future researchers can change these values as they see fit or these simulations can be re-run to their specifications. The key operational question is to what extent assumptions of the value of r alter the results on treatment outcome presented by the model. There is negligible impact of this parameter on two key comparisons i.e. the impact of alternative regimens and the impact of resistance – this is described and justified in extensive detail below.

#### 6.3.3.1 Impact of pathological recovery rate on the baseline scenario i.e. treatment of artemisinin-sensitive parasites with the standard, WHO-recommended regimen.

The impact of half-life on the values of $AUC_{PL}$ and MPL (for patients with sensitive parasites treated under the standard regimen) is shown in **Figure 6.11**; a shorter half-life (indicating a faster rate of clearance) resulted in lower outcome metrics. This is unsurprising – a faster resolution of pathology from dead sequestered parasites would result in better outcomes. Analysis of the baseline was conducted purely to establish basic dynamics of treatment and pathology and to check the results were consistent with expectations. This is clearly the case i.e. that increased recovery rate half-life r causes increased pathology.

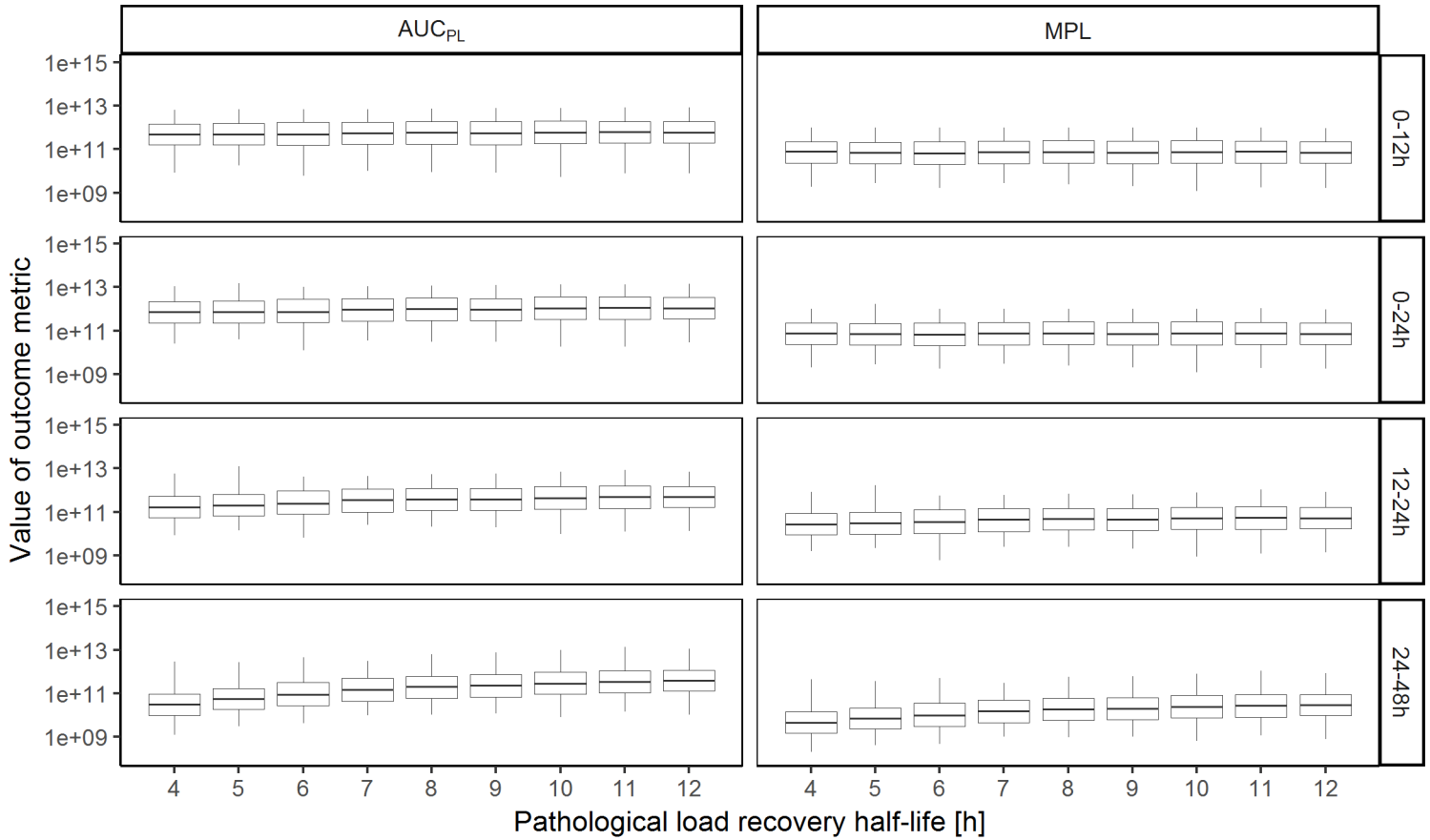



**Figure 6.11** Population values of area under the pathological load curve ($AUC_{PL}$) and maximum pathological load (MPL) for patients with sensitive parasites treated with the standard regimen, plotted according to half-life of the pathological load recovery rate r across four time periods post-treatment: 0-12h, 0-24h, 12-24h and 24-48h.

### 6.3.3.2 Impact of pathological recovery rate on the impact of artemisinin resistance.

The assumed value of pathological recovery rate had no effect on estimates of the impact of artemisinin resistance. This can be seen in **Figure 6.10**, **6.3.3.1** and additionally in **Figure 6.12**.

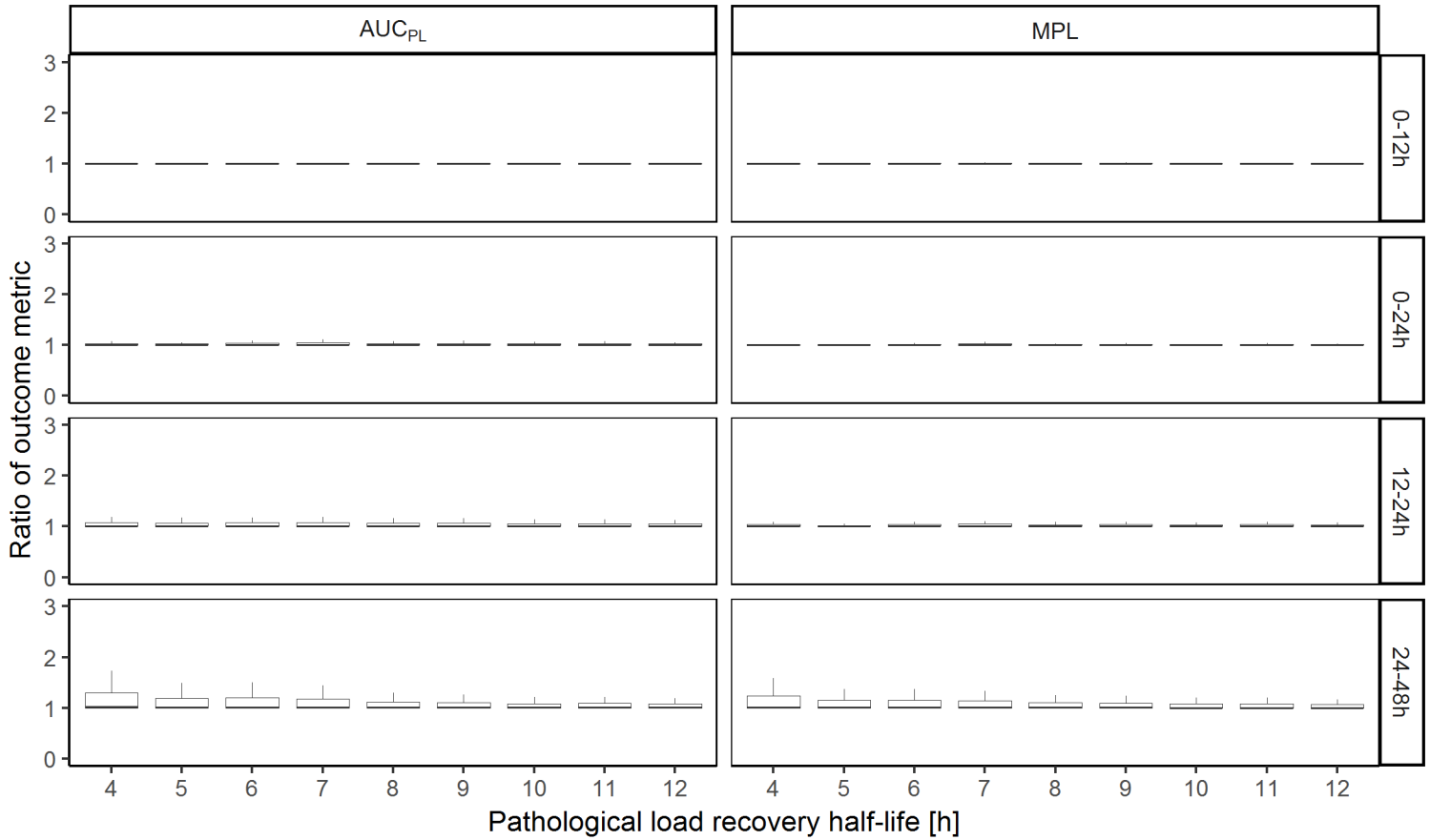



**Figure 6.12** Ratios of area under the pathological load curve ($AUC_{PL}$) and maximum pathological load (MPL) for sensitive v resistant parasites treated with standard regimen, plotted according to half-life of the pathological load recovery rate r across four time periods post-treatment: 0-12h, 0-24h, 12-24h and 24-48h.

### 6.3.3.3 Impact of pathological recovery rate on the relative performance of alternative regimens.

The assumed value of pathological recovery rate had a negligible impact when comparing the relative performance of simplified versus standard regimen in the 0-12, 12-24 and 0-24h time periods (**Figure 6.7**, **Table 6.9**). However, at 24-48h there was notable negative correlation between $AUC_{PL}$ and MPL and half-life of r meaning that the simplified regimen performs increasingly worse during this period as pathological recovery rate increases (i.e. half-life decreases) – this is shown in **Figure 6.13** .

The most plausible reason for the pattern can be explained by recalling that pathological load is the sum of two contributions: the pathology caused by alive sequestered parasites (first term of Equation 6.12) and the lingering pathology due to previously-killed parasites (the second term of Equation 6.12). The pathological recovery rate only affects the second term and so alters the relative size of the two contributions to the overall pathological load. This effect will be explained intuitively and illustrated by dynamics of treatment in an exemplar patient shown on **Figure 6.14**, (note that this is an illustrative patient out of the 10,000 sampled). Consider the first term of Equation 6.12 i.e. number of alive sequestered parasites. The two doses of artemisinin in the first day means that the WHO-standard regimen kills more parasites (for reasons more fully explained in [111]); so the alive, sequestered component is always higher in the simplified than standard treatment (**Figure 6.14** Panel A v B and Panels C vs D). However, the lingering post-mortem pathology from previously killed parasites (the second term in Equation 6.12) obscures this difference in live parasites. In the first 24 hours post treatment, the lingering pathology of killed parasites is large, more than an order of magnitude higher than live sequestered burden (**Figure 6.14**), and there is negligible difference between the regimens. After 24 hours the lingering pathology is not such a dominant component, the effect of live sequestered parasites starts to become noticeable. Increased values of pathological recovery rate (lower half-life, r) means the lingering pathology becomes a smaller contribution to pathological load and the difference between the regimens becomes more apparent; hence after around 24 hours, the value of pathological recovery does start to have an impact on the comparison between the simplified and standard regimen. To make this algebra clearer, recall that comparisons are being conducted with ratios of outcome metrics rather than their absolute values:

$$Ratio\ of\ outcome\ metric = \frac{alive\ sequestered\ parasites\ (simplified\ regimen) + lingering\ pathology\ (simplified\ regimen)}{alive\ sequestered\ parasites\ (standard\ regimen) + lingering\ pathology\ (standard\ regimen)}$$

(see also Equation 6.12)

Lingering pathology is extremely large in the first 24 hours due to the large amount of killed parasites after initial treatment and makes up an overwhelming proportion of the pathological load. There is negligible difference between the regimens in terms of the lingering pathology (which is primarily governed by the recovery rate, r), and so because the lingering pathology component of the equation is by far the largest, the ratio of the outcome metrics will be very close to 1. However, the magnitude of lingering pathology values falls over time post treatment (due to it being resolved by the rate of recovery, r) and is lower in the 24-48h period than in 0-24h. These lower values are less able to obscure the difference in effectiveness between the regimens in terms of the number of alive sequestered parasites, so the ratios increase and become more variable (**Figure 6.13** ). This can be observed in **Figure 6.14**: Compare panels A and B to panels C and D – with the faster resolution of pathology afforded by a shorter half-life (A, B), the alive sequestered parasites constitute the largest proportion of pathological load by hour 48. With a longer half-life (C, D), the lingering pathology is still the largest component. The faster pathology is resolved, the smaller the magnitude of lingering pathology and the larger the value of the ratio.

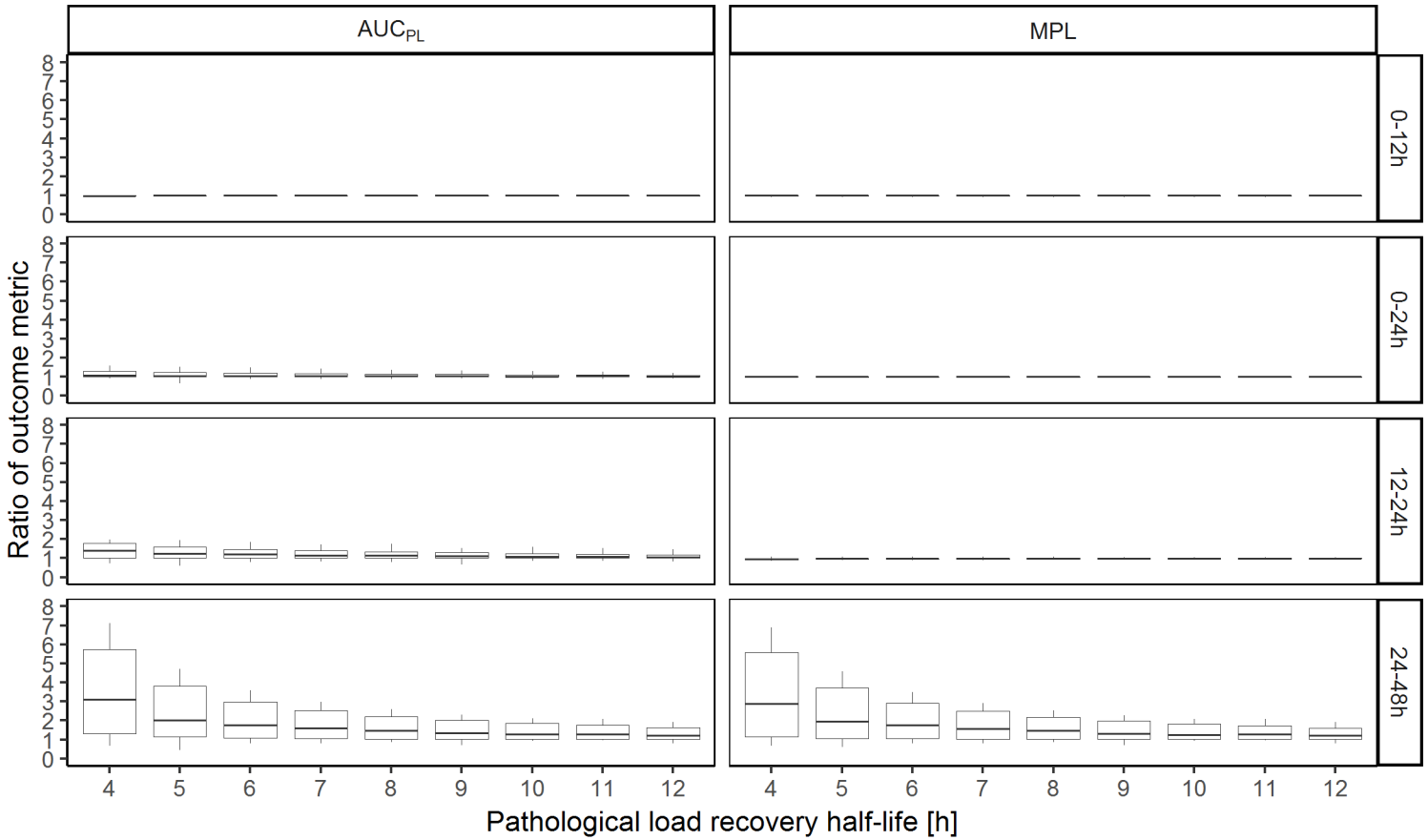
Using illustrative values, at 24h when lingering pathology is high:

$$ratio = \frac{1e8\ (alive\ sequestered) + 1e10\ (lingering\ pathology)}{1e7\ (alive\ sequestered) + \ 1e10\ (lingering\ pathology)} = 1.009$$

At 48h, when lingering pathology is low:

$$ratio = \frac{1e8\ (alive\ sequestered) + 1e8\ (lingering\ pathology)}{1e7\ (alive\ sequestered) + \ 1e8\ (lingering\ pathology)} = 1.81$$

In summary, pathological recovery rate has no impact in the first 24 hours but has an important impact on the relative performance of the regimens (quantified as ratios of $AUC_{PL}$) in the later, 24-48h post-treatment period (see **Figure 6.7 and Table 6.9**) (note that it is not important for MPL as the value of MPL in the 24-48h period typically occurs at 24 hours, this is reflected in the PRCC analysis showing no correlation between MPL at 24-48h and the half-life of r). This analysis of different drug regimens to treat severe malaria is robust to assumptions of the value of recovery rate for the following two reasons. Firstly, in the critical first 24 hour period following treatment (the period focused on by Kremsner et al. [170]), PRCC results indicate that the half-life parameter has no impact on the ratio of $AUC_{PL}$ and MPL of the regimens. This is further evidenced by **Figure 6.13 and Figure 6.14 –** even with an extremely short half-life, the pathological load did not differ sufficiently in the first 24 hours to make any difference between the regimens when r changes. Secondly, pathological load is much higher in the 0-24h period than the 24-48h period, so the lack of impact of r in the 0-24h period is critical ($AUC_{PL}$ and MPL at 24-48h are generally 20-30% of the values at 0-24h, though this will vary depending on the value of patient parameters, including the half-life of the pathological recovery rate) This reflects the belief that pathology/death in the 24-48 hour period post-treatment is not described solely by pathological metrics in that period: rather it is the cumulated damage since treatment was initiated that is important and that includes the metrics in the 0-24h period. Finally, it is clear from the full set of results presented here that while an extremely low half-life is associated with better outcome from the standard regimen in all time periods except 0-12h, the standard regimen produces lower $AUC_{PL}$ across the entire range of parameter values (MPL is equivalent at 0-24h and 12-24h), so no matter the parameterization of the half-life of r, the standard regimen still produces lower outcome metrics.




4022

4023 **Figure 6.13** Ratios of area under the pathological load curve (AUC$_{PL}$) and maximum pathological load (MPL) for standard v simplified regimen with sensitive
4024 parasites, plotted according to half-life of the pathological load recovery rate r across four time periods post-treatment: 0-12h, 0-24h, 12-24h and 24-48h.

4025

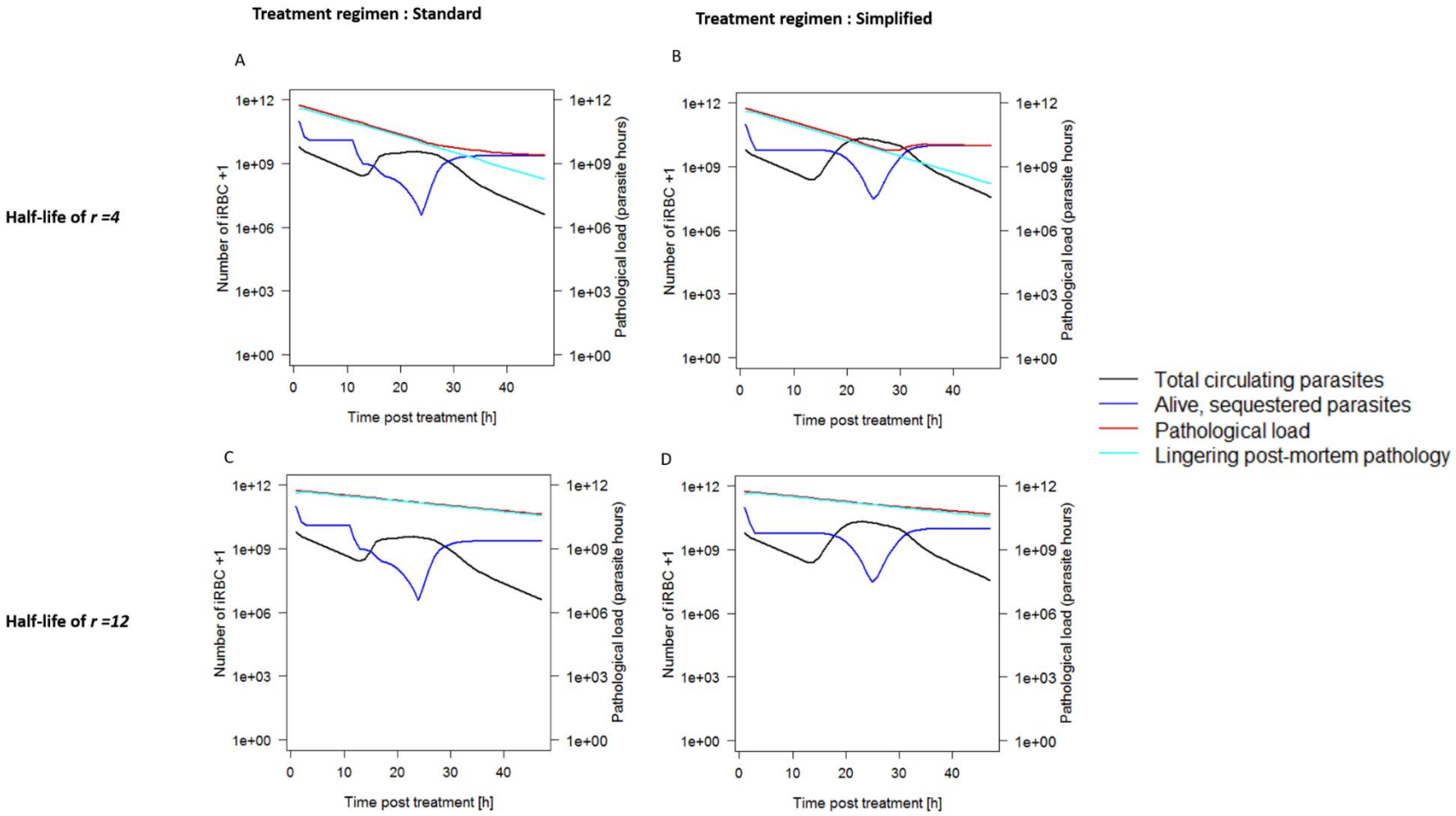



**Figure 6.14** Plots to illustrate the proportion of total pathological load contributed by alive sequestered parasites and by lingering post-mortem pathology, and the differences in these contributions when the half-life of the recovery rate r is varied. With a lower half-life (faster recovery), alive sequestered parasites constitute a larger proportion of pathological load and so differences between standard and simplified regimens with respect to the number of alive, sequestered parasites become more apparent with the assumption of faster recover.

## 6.4 Discussion.

### 6.4.1 Model validity.

An mPK/PD modelling methodology capable of investigating the treatment of severe malaria was developed. Kremsner et al. [47] recognised the clinical necessity of this, and noted that "for the first time, we [i.e. Kremsner et al.] are assessing artesunate using similar pharmacokinetic and dynamic approaches". Parasite clearance is likely to be a poor measure of regimen effectiveness (and, by extension, clinical outcome) in severe malaria where pathology is due to sequestered parasites. The effects of alternative regimens and the impact of drug resistance can only be investigated by traditional clinical outcomes using large scale clinical trials, so pharmacological modelling of the type proposed here is essential to help generate the evidence base for rational treatment design. This modelling was highly flexible and, of necessity, reflected limitations in understanding of pathology, for example, how rapidly pathology is resolved following parasite death and whether pathology depends on maximal sequestered load (measured as MPL) or on total exposure (measured as $AUC_{PL}$). An interesting, highly important result is that the key quantitative assumption made in the analysis, the rate of resolution of pathology (measured as the half-life of r), had little effect on the conclusions when comparing alternative regimens or the impact of resistance (**6.3.3.3**) implying that the pathological model is a robust to assumptions made in this comparative investigation Importantly, while circulating parasite loads do not reflect the pathology of severe malaria they are currently the regular endpoint of choice in severe malaria trials, including those undertaken by Kremsner et al. [170, 190]; this model was able to reproduce the clinical outcomes reported in [170, 182] (when appropriately parameterized), and recover expected $PRR_{48}$, so confidence that it is reflective of in vivo scenarios is present.

#### 6.4.1.1 Model metrics and alternatives.

To the best of my knowledge, there is no biological proxy, or direct clinical measurement, for the rate of detachment of iRBCs containing dead parasites in vivo, and it is unknown whether all sequestered iRBCs cause the same pathological burden irrespective of how long they have been sequestered. Therefore two simple metrics of pathology were used to enable comparison of the impact of different drug regimens and resistance. Note that this is not attempting to be definitive description of severe malaria pathology but present a flexible methodology that allows users to construct more sophisticated models of pathology should they so wish. A simple exponential decline of pathology was assumed, quantified as the pathological recovery rate r and that it applies to all sequestered iRBCs containing dead parasites, but this could be expanded in a number of ways, most likely:

Pathology may be affected by duration of binding e.g. a iRBC that has been sequestered for three hours may have caused less damage than one sequestered for 30 hours; it would be possible to construct a pathology index that reflects this (although this would entail constructing an index to quantify this relationship between duration and pathology).

The reduction in pathology may not be linear. For example, it may decline very slowly for the first few hours following parasite death, then very rapidly thereafter (or vice versa).

It was assumed pathology was proportional to the number of sequestered parasites and their lingering pathology. Readers may prefer to set a pathological trigger (e.g. $>10^9$ sequestered parasites) and use time above that limit as a metric for pathology.

In short there is great flexibility to model pathology, but there is little precedent on how to properly construct or calibrate a definitive model. Hence, this simple methodology was chosen to minimise complexity and arbitrary choice of parameters and to avoid "cherry picking" metrics to support results.

Severe malaria is the result of a myriad symptoms and patients often present with different levels of cerebral malaria, severe anaemia, metabolic derangement and respiratory distress. It was assumed that reducing sequestered parasitaemia is the key therapeutic objective and that the MPL and $AUC_{PL}$ are realistic metrics to compare treatment regimens of severe malaria. These metrics do not have to reflect pathology exactly, but are sufficiently accurate proxies that they can serve as comparisons for the clinical efficacy of treatment regimens. The simplifications in model calibrations primarily reflect the currently poor understanding of the pathology of severe malaria [81], however, one advantage of this methodology is that it is highly flexible and transparent. Users may easily change the calibration to reflect their beliefs of the underlying pathology and test how treatment regimens impact prognosis. For example, respiratory distress syndrome is associated with higher levels of circulating parasitaemia and the clinical priority may be to reduce circulating parasitaemia [16], in which case more detailed analysis of apparent PRR could be conducted [193].

The results of this research focused on pathology in the first 24 and 48 hour periods. Pathological effects are known to last much longer and, for example, neurological sequalae can be observed months after treatment[194]. This long-lasting pathology may not be due to parasites present at the time of observations but may instead be a long-lasting consequence(s) of the parasitaemia and pathology that occurred within the first 48 hours as the pathological load has typically fallen by 4 orders of magnitudes by 48 hours. Therefore, these results should apply equally well to these longer lasting pathological effects as to the more immediate metrics of death within 24 or 48 hours.

Little difference was observed between ratios of $AUC_{PL}$ and MPL when comparing regimens or resistance (i.e., the ratios of $AUC_{PL}$ are close to the ratios of MPL). This occurs because, for each time period, the highest point of pathological load is almost always the earliest hour (i.e., highest pathological load is at 0 hours in 0-12h, 12 hours in 12-24h, so on). Consequently, the MPL for each patient in a time period typically occurs at the very beginning of that time period. It follows that the largest contributions to $AUC_{PL}$ also occur at the very beginning of the time period; **Figure 6.1** of an exemplar patient shows that pathological load follows a linear decline over time.

The similarity between ratios of $AUC_{PI}$ and MPL was thus un-concerning, it follows that a patient with a higher MPL will have a higher $AUC_{PI}$ and any parameters that affect one will affect the other equally (i.e., initial parasite number, mean age-bin).

### 6.4.2 Standard v Simplified regimen.

Kremsner and colleagues [170, 190] concluded that their simplified regimen was non-inferior to the standard WHO regimen and possessed operational advantages due to less frequent drug administration[170, 190]. This work was influential and initiated a wider debate about the best drug regimen(s) to treat severe malaria [47, 50, 195] to which the results presented herein can contribute. Comparison of the 0-24h and 12-24h period was used to compare the effects of the initial, larger dose of the simplified regimen against the additional dose at 12h with the standard regimen. The standard regimen produced slightly lower median $AUC_{PL}$ within the first 24 hours post-treatment (**Figure 6.5**, **Table 6.8**). This difference was greater in the 24-48h period, but the majority of pathological load occurred within the first 24 hours as artesunate rapidly kills parasites– $AUC_{PL}$ in the 24-48h period is, on average, between 20-30% that of $AUC_{PL}$ in the 0-24h period (data not shown). The first 24 hours are critical for patient survival[171], so outcome metrics at 24-48h may have little relevance in choosing between regimens. However, the simplified regimen performed much worse in the sub-group of patients with very late or very early initial mean age-bins. Based on these results, care should be taken if recommending use of the simplified regimen but there is an important rider to this. Kremsner et al. never claimed this simplified regimen would be superior, but argued that any inferiority, if it exists, would be within acceptable margins. I will not comment on this, and leave it to clinically qualified personnel to judge whether 50% in some subgroups is within an acceptable margin

of inferiority, especially given the inability to directly link the pathological outcomes investigated here with the likelihood of mortality.

An important note is that any value of SD can be modelled. The key change as the distribution widens is that the mean age-bin parameter becomes less important. This is intuitive as the importance of this parameter derives from the fact that, with a narrow distribution, patients with certain initial mean age-bins have treatment falling in less/non-sensitive bins. The technical problem with this distribution is that there is no "mean age bin" so it is not possible to analyse its impact nor incorporate this distribution into formal sensitivity analyses. Under the circumstances of uniformly distributed parasite age-bins, it is fair to claim, as Kremsner et al. did, that the simplified regimen is "non-inferior". This occurs because when parasites are uniformly distributed there are no "at risk" patients with infections clustered in early mean age-bins at the time of treatment, and all patients will experience high levels of killing from the initial artesunate dose.

### 6.4.3 Impact of artemisinin resistance.

The impact of artemisinin resistance on treatment of severe malaria was assessed, i.e. the extent to which resistance increased MPL and $AUC_{PL}$. Resistance prevents drug killing in age-bins 2-4 (these bins are otherwise hyper-sensitive) resulting in no killing for a contiguous 8 hour period in resistant parasites (i.e. age-bins 45 to 5). Results show the initial mean age-bin and it's SD are the only parameters that distinguish outcomes between sensitive and resistant parasites (**Figure 6.10**) . It has been argued previously [15] that artemisinin resistance would have a negligible impact on eventual cure rates in uncomplicated malaria (provided there was no resistance to partner drugs) but artemisinin resistance clearly poses a much larger threat to treatment of severe malaria than it does to uncomplicated malaria. Although differences between sensitive and resistant parasites across the entire population are minor (**Figure 6.5**, **Table 6.8**), there is an extremely vulnerable sub-group of patients whose infections at the time of treatment are clustered in very late or very early age-bins (i.e., where parasites are resistant in the model; **Figure 6.10**).

# Chapter 7: Thesis Conclusions.

This thesis has used a computer modelling approach to achieve two critical objectives, one focusing on uncomplicated falciparum malaria and one focusing on severe falciparum malaria:

**1: To develop a methodology that can accurately calculate true failure rates of Artemisinin-based combination therapies (ACTs) in uncomplicated malaria therapeutic efficacy studies (TES), and, against that gold standard figure, compare the accuracy and utility of a range of current and proposed methods to estimate ACT failure rates, and ways in which the accuracy and usage of these methods may be optimized.**

**2**: **To develop a methodology to quantify the pathology of sequestered parasites in severe malaria and use this methodology to identify rational drug dosing regimens for treatment of severe malaria with artesunate (AS) and quantify the likely impact of AS drug resistance for treating severe malaria.**

## 7.1 A Computer modelling approach to improving failure rate estimates from TES through simulation of molecular correction.

Three methodologies for molecular correction in therapeutic efficacy studies (TES) were explored within this thesis. Each are topical: The length polymorphic marker methodology using merozoite surface protein-1 (msp-1), merozoite surface protein-2 (msp-2) and the glutamate rich protein (glurp) is the presently recommended methodology by the World Health Organization (WHO) [14], but the accuracy of this approach has long been under scrutiny [121]. Microsatellite markers are used in TES by the Centers for Disease Control and Prevention (CDC), and novel methods have suggested that simple statistical analysis may be severely under-estimating drug failure rates [31, 150]. Finally, use of next-generation amplicon sequencing (AmpSeq) for molecular correction is still in exploratory stages, and a small but promising selection of research has been undertaken [36, 41, 160].

The computer modelling approach undertaken in this thesis had, broadly, the same goal for all three methods: Determine the accuracy of failure rate estimates produced using each methodology through comparison to a true failure rate calculated using mechanistic pharmacokinetic/pharmacodynamic (mPK/PD) modelling. (**Chapter 2**). The simulation of each methodology has, necessarily, had different requirements and different goals. The objectives, core methodology and key findings of modelling TES using length-polymorphic markers (**Chapter 3**), microsatellite markers (**Chapter 4**) and AmpSeq (**Chapter 5**) are briefly summarized below:

### 7.1.1 Length-polymorphic Markers (Chapter 3).

A selection of algorithms are available to classify recurrent infections as either recrudescence or reinfection using length-polymorphic markers (**Table 3.1**). The accuracy of each algorithm in a variety of settings describing transmission intensity was modelled. The impact on failure rate accuracy of altered follow-up length of TES was also investigated.

The currently recommended WHO/MMV algorithm was extremely conservative in classifying recurrent infections as recrudescent and misclassified a large proportion of recrudescence as reinfection. Consequently, modelling work suggested that use of this algorithm is likely to under-estimate the failure rate of failing drugs (DHA-PPQ, AR-LF and AS-MQ were modelling with ~10% true failure rates). Failure rates for non-failing drugs were also under-estimated, but the absolute difference between the true and estimated failure rate was small and so the operational consequences are minor.

Novel algorithms each misclassified some proportion of recrudescence and reinfection, and which was most accurate was determined by transmission intensity (and thus, the number of reinfections). As a general trend, a novel "allelic family switch" algorithm was the most accurate in areas of lower transmission intensity and a "≥2/3 markers" algorithm was most accurate in areas of higher transmission intensity. Follow-up lengths of 28 days for AR-LF and 42 days for DHA-PPQ and AS-MQ (49 days would also be appropriate for AS-MQ) were appropriate to recover accurate failure rate estimates with the algorithms described above; this was consistent with current WHO recommendations of 28 days for AR-LF and 42 days for DHA-PPQ and AS-MQ. Sensitivity analysis of model parameters showed that conclusions were robust to model assumptions, and re-analysis of existing clinical data with the new algorithms showed qualitatively the same trends observed with the modelling approach.

### 7.1.2 Microsatellite Markers (Chapter 4).

Microsatellite markers are not recommended for use in molecular correction by the WHO but are used in practice by the CDC and other groups (examples:[40, 150, 151], **Chapter 4**). The use of microsatellite markers to classify recurrent infections was modelled here. The accuracy of failure rate estimates for AR-LF and AS-MQ calculated with both a simplified "match counting algorithm"- where a recurrent infection was classified as recrudescent when a threshold number of microsatellite loci had matching alleles between the initial and recurrent sample- and a novel Bayesian algorithm [150] were investigated. A detailed analysis of Bayesian algorithm was then conducted to determine how robust it was in a variety of transmission intensity settings and how well it could identify low density recrudescent infections.

Failure rate estimates obtained using the match counting algorithm consistently under-estimated the true failure rate of failing AR-LF and AS-MQ when the threshold number of matching loci was all or most microsatellite loci genotyped. Lower thresholds increased the failure rate estimate, but no threshold was able to consistently recover the true failure rate as failure rate estimates varied widely as transmission intensity was changed. Low thresholds lead to large over-estimation of true failure rate in high transmission settings.

A novel Bayesian algorithm was able to recover the true failure rate to a high degree of accuracy in all scenarios. The Bayesian algorithm was highly specific (negligible misclassification of recrudescence as reinfection) and classifying every recurrence with a posterior probability of recrudescence, p, ≥0.1 as a recrudescence was the most accurate approach. A slight under-estimate of failure rate still occurred; some low-density recrudescences were not correctly identified by the algorithm. The Bayesian algorithm was able to accurately estimate failure rate even in a simulation area of extremely low genetic diversity.

### 7.1.3 Amplicon Sequencing (Chapter 5).

Use of AmpSeq for molecular correction is a subject of recent interest; they have not (to my knowledge) been formally used to obtain failure rate estimates in a TES, and research is currently focusing on developing protocols for their use [36, 41, 158]. The purpose of modelling work was to determine the accuracy of failure rate estimates using AmpSeq markers and answer two important operational questions: (i) How many (and which AmpSeq loci must be genotyped to obtain accurate failure rate estimates in a variety of scenarios of transmission intensity? (ii) can a simple matching algorithm return accurate and robust failure rate estimates?

AmpSeq methods were able to estimate the true failure rate with a high degree of accuracy in multiple MOI settings and levels of transmission intensity. It was sufficient to base classification of recurrent

infections as recrudescence or reinfection on genotyping 3 highly diverse AmpSeq loci, and the AmpSeq methodology was capable of estimating failure rates to be within 1% of the true failure rate at all levels of transmission intensity investigated in the model.

### 7.1.4 Approaches for obtaining accurate failure rate estimates in TES.

The two molecular correction methods that are currently most widely used (i.e., the WHO/MMV algorithm for length-polymorphic markers and the match counting algorithm with a stringent threshold for microsatellite markers) will result in large under-estimates of failure rate of failing drugs. Under-estimation of true failure rate with these approaches has previously been inferred from field data and in vitro studies for both length-polymorphic markers [106] and microsatellite markers [31, 150]. However, the true failure rate cannot be known in vivo, and such comparisons have been based on the simple fact that, for length-polymorphic markers, alternative algorithms produce different failure rate estimates, and for microsatellite markers that failure rate estimates differ as the threshold changes for a match-counting algorithm and that a Bayesian algorithm produced different estimates to match-counting algorithms. The modelling work presented here has been able to quantify these differences and reveals the magnitude of under-estimation with current methods.

The WHO/MMV algorithm and the microsatellite match-counting algorithm share an important trait: They have both been extremely conservative (stringent) when classifying recurrent infections as recrudescent. The WHO/MMV algorithm requires shared alleles between the initial and recurrent sample at all three loci to classify a recrudescence [14], and the microsatellite match-counting algorithm has historically requires matching alleles at "all" or "all but one" loci, when the number of loci has varied but has generally been 6 or 7 [31, 40, 98, 142, 151].

There is a clear problem with this approach of classifying recurrent infections: length-polymorphic markers and microsatellite markers methodologies will both fail to detect low-frequency alleles during the PCR process (**3.1.2**; **4.1.2**) The practical consequence of this is that, for both methodologies, minority alleles <20-30% of the majority allele will not be detected (see individual chapters for full discussion). The modelling approach undertaken in this thesis quantified the impact on failure rate estimates of the inability to detect these minority alleles.

Before discussing the impact of modelling less stringent algorithms, an important note must be made: It was possible for a given allele(s) (of any type of marker) of a clone to be detected while others were not. For all three methodologies, alleles could be shared between clones and the sum frequency of a given allele in a sample was determined by the number of clones with that allele in the sample. In other words, alleles on a given clone may exist in different relative frequencies because some were present on multiple clones while others were not. This could result in one allele on a clone being the majority allele in the sample while the others were minority alleles. For length-polymorphic markers, alleles were variable in their detectability according to their relative lengths (this is family-specific for *msp-1* and *msp-2*). Consequently, a situation could arise where an allele at a given locus of a clone was not observed but alleles at another locus on that same clone were, due to the length (and frequency) of other alleles at that locus present in the sample. Finally, it was possible for microsatellite alleles at a given locus to not be correctly observed through errors in reading microsatellite length (**4.1.2**)

As such, less stringent methodology may allow recrudescent clones to be accurately identified in circumstances where an allele(s) of that clone was not observed whereas a stringent methodology requiring an allele to be shared at all loci between the initial and recurrent samples would misclassify that sample as a reinfection.

It is almost inevitable that less stringent algorithms misclassify reinfections as recrudescence at a greater frequency than stringent algorithms. Intuitively, this seems extremely undesirable, but the results herein suggest that in practical terms "two wrongs can make a right" – less stringent algorithms

allowed for opposing sources of misclassifications to cancel out, given that low density recrudescence are difficult to detect with length polymorphic or microsatellite markers (resulting in the under-estimate of true failure rate using stringent algorithms described above). This may seem extremely unscientific but the modelled results for length-polymorphic markers suggest that novel algorithms are relatively robust across a variety of transmission scenarios and produce reasonably accurate failure rate estimates (**Chapter 3**). It is not presently possible to increase detection of low density recrudescence with these markers (this is a goal that is achieved using AmpSeq markers instead), but using different algorithms, as shown in **Chapter 3**, can improve the accuracy of failure rates by a) correctly identifying more recrudescent infections and b) misclassifying some proportion of reinfections as recrudescence, cancelling out under-estimation of true failure rate due to low density recrudescence.

A less stringent approach for microsatellite markers using a traditional match counting algorithm appeared to be unsuitable. The modelled results in **Chapter 4** showed that no threshold number of matching loci produced accurate failure rate estimates that were robust across a variety of transmission intensity scenarios. Lower thresholds induced over-estimates of true failure rates in high transmission scenarios, and stringent thresholds consistently under-estimated true failure rate (as described above).

However, use of a Bayesian analysis methodology produced extremely accurate failure rate estimates. The mPK/PD methodology used here allowed for a detailed analysis of results using this algorithm. Current usage of the algorithm in vivo has classified recurrent infections as recrudescence when the posterior probability of recrudescence, p, was ≥0.5 [31, 150]; modelling results indicate that the algorithm has extremely high specificity and classifying recurrences with p ≥0.1 will produce slightly more accurate failure rate estimates (generally an extremely small under-estimate of the true failure rate. However, the as the algorithm calculated p as either <0.1 or ≥0.9 for most recurrences, the difference between using p ≥0.1 and p ≥0.5 to classify a recrudescence will be small in practice.

Modelling of AmpSeq markers showed an extremely high ability to recover the true failure rate in all scenarios modelled. It was sufficient to genotype three highly diverse loci and classify a recrudescence with matching alleles between the initial and recurrent sample at ≥2 loci. Note that the length-polymorphic markers and microsatellite markers (with the exception of the extremely low diversity setting) were similarly diverse (as quantified by expected heterozygosity; He), so it is the greatly increased detectability of low-density clones that enables this method to provide a marked improvement over the status quo.

A direct comparison of all the failure rate estimates obtained from the different methodologies investigated in this thesis is difficult because different drugs were investigated for each chapter (see earlier chapters for reasoning), and where possible the models were calibrated with MOI distributions obtained using the markers of interest (i.e., the MOI distributions used in **chapter 3** and **chapter 5** were obtained from field data using length-polymorphic markers and the MOI distributions used in **chapter 4** were obtained using microsatellite markers. However, failing AR-LF was modelled for all three methods and a comparison of AR-LF failure rate estimates obtained during a 28 day follow-up assuming a "high" MOI (the MOI from Tanzania for length-polymorphic markers and AmpSeq markers and the MOI from the "high transmission" scenario for microsatellites) is possible for illustrative purposes, noting that the true failure rate of AR-LF differed between chapters. Failure rate estimates at different transmission intensity levels (quantified by FOI) can be shown for FOI 0, 2, 8 and 16 for length-polymorphic markers and AmpSeq markers. For microsatellite markers the "high transmission scenario" was calibrated only with an FOI of 16. Consequently, length-polymorphic marker results and AmpSeq results are directly comparable as the same MOI distribution and FOI values were used for the "low", "medium" and "high" transmission scenarios. However, microsatellite markers are calibrated differently, so comparison of microsatellite results with other markers requires caution. Comparison of estimates is shown in Figure 7.1 for length-polymorphic markers and AmpSeq and

Figure 7.2 for microsatellite markers, noting that the "best" estimate is presented (i.e., the Bayesian algorithm classifying a recurrence as recrudescence with a posterior probability of recrudescence, p, of ≥0.1 for microsatellite markers and for AmpSeq, genotyping 3 loci and classifying a recurrence as a recrudescence when either 2 or 3 loci match between samples; all length-polymorphic marker algorithms are shown). These values of failure rate estimates are all calculated using survival analysis and correspond to those shown in **Figure 3.7** for length-polymorphic markers, **Figure 4.2** for microsatellite markers and **Figure 5.1** for AmpSeq.

Modelling work evaluating the three methodologies available to recover accurate failure rate estimates for failing drugs, is summarized in Table **7.1** according to the order they are discussed in this thesis: The length-polymorphic markers using novel algorithms, microsatellite markers using a Bayesian algorithm, and finally AmpSeq markers with failure rates estimated using a match counting algorithm. These methods are distinct from those currently being employed in the field for TES; the modelling work presented in this thesis, combined with existing literature [31, 41, 106, 122] and re-analysis of TES data (**Table 3.3**) strongly suggest that drug failure rates are being under-estimated in vivo with current practice for length-polymorphic markers and microsatellites.

If the field were to adopt any of these new methods, it should be expected to observe failing drugs in areas they were previously thought to be efficacious, particularly for DHA-PPQ where drug failure has already been observed (using length-polymorphic marker methodology) in vivo [20]. Given their high level of accuracy and the lack of need for a complicated Bayesian analysis, AmpSeq markers appears to be the most desirable methodology, but it is important to note that currently used methods (length-polymorphic markers, microsatellites) can achieve more accurate failure rate estimates with the described methodological tweaks.

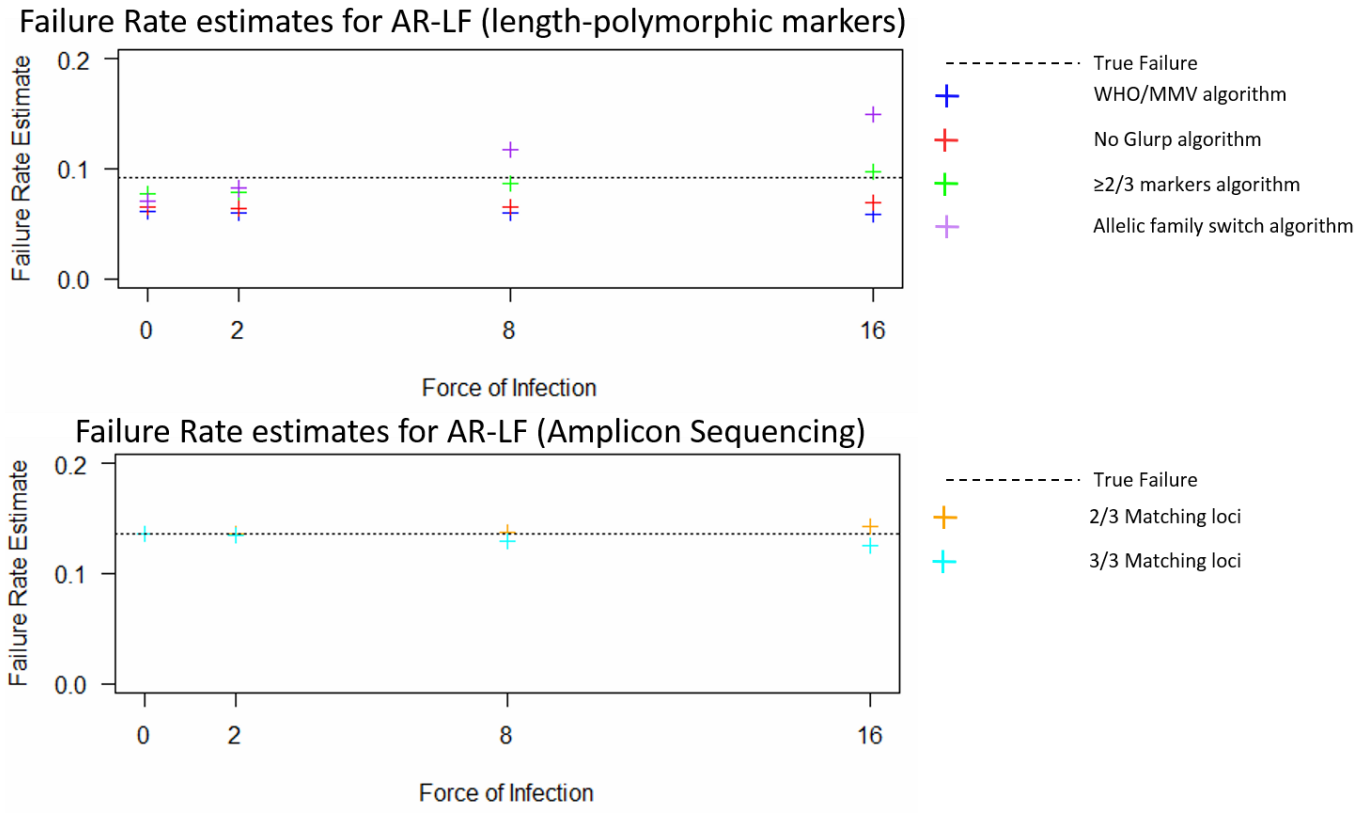



**Figure 7.1** Comparison of failure rate estimates obtained using length-polymorphic marker algorithms and AmpSeq (with 3 loci genotyped and classifying a recurrence as recrudescence at either 2/3 matching loci or 3/3 matching loci), with a high Multiplicity of Infection (MOI) distribution as described in **2.3.1** and with Force of Infection (FOI) set to either 0, 2, 8 or 16

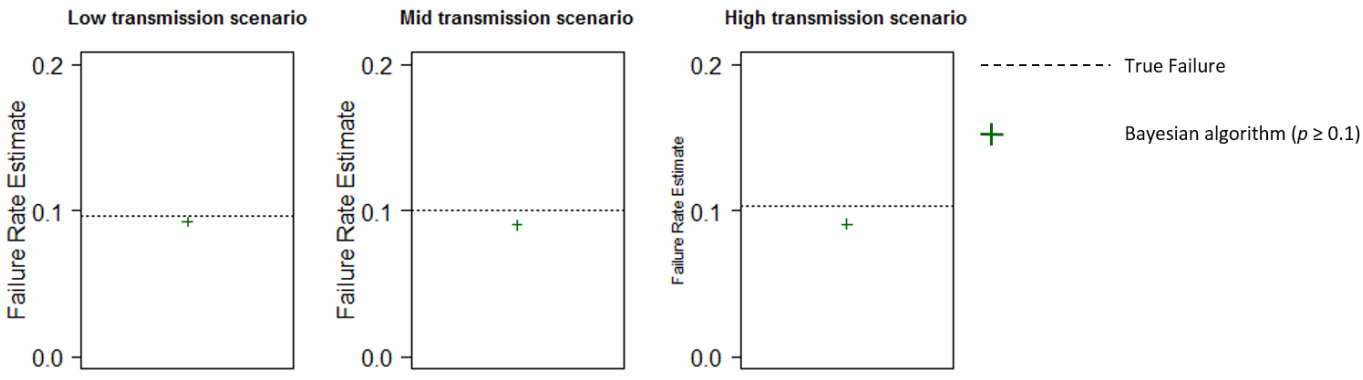



**Figure 7.2** Comparison of failure rate estimates obtained using microsatellite markers calculated with a Bayesian algorithm, with transmission scenarios varied as described in **4.2.1**.

**Table 7.1** Summary of failure rate estimates obtained with each of three types of genetic marker, and how improvements in the accuracy of these estimates may be realized according to the output of modelling work.

| Genetic marker | Current use in vivo and consequences (identified from modelling work) | Proposed improvements (identified from modelling work) | Key sections of thesis |
|---|---|---|---|
| Length-polymorphic markers | WHO/MMV algorithm is extremely conservative at classifying recrudescence and consistently under-estimates true failure rate at all levels of transmission intensity | Use of an allelic family switch algorithm (in areas of low transmission intensity) and a ≥2/3 markers algorithm (in areas of high transmission intensity permits more accurate estimation of failure rates. | **3.3** |
| Microsatellite markers | Match-counting algorithm with stringent thresholds is extremely conservative and consistently under-estimates true failure rate. Lower thresholds produce higher failure rate estimates but are not viable for use as failure rate estimates are extremely sensitivity to changes in transmission intensity. | A novel Bayesian algorithm is capable of estimating failure rate estimates to a high degree of accuracy. The Bayesian algorithm is highly specific and recurrences with posterior probabilities of recrudescence of ≥0.1 should be classified as recrudescence | **4.3** |
| Deep-sequenced Amplicons | Novel methodology not currently under wide-spread use. Modelling work supports the approach in the only existing use of AmpSeq for molecular correction [41] i.e. genotyping three highly diverse loci *ama1-D3*, *cpmp* and *cpp* permits accurate estimation of failure rates. | This methodology accurately estimates failure rates in all transmission intensity scenarios, and steps should be taken for it's wide-spread deployment. It was sufficient to only genotype three highly diverse loci (*ama1-D3*, *cpmp* and *cpp* were modelled here), and classify a recrudescence when the initial and recurrent sample share alleles at 2 or more loci. | **5.3** |

*WHO/MMV: World Health Organization/Malaria for Medicines Venture, AmpSeq: Amplicon Sequencing, cpp: conserved plasmodium protein, cpmp: conserved plasmodium membrane protein, ama1-D3: Apical membrane antigen*

### 7.1.5 Molecular correction for TES: Next steps.

Modelling work such as that undertaken here, while valuable, is not a substitute for real, clinical data, and a clear next step is apparent from this work: A clinical trial (a TES) of sufficient size that genotyped blood samples using all three methods (i.e., length polymorphic markers, microsatellite markers and AmpSeq markers) in all patients would allow, for the first time, the failure rate estimates produced using each methodology (including different algorithms and varied match-counting thresholds) to be

directly compared. Of course, the true failure rate cannot be known in vivo, but the wealth of data available from such a trial, combined with modelling work such as that presented here and in vitro approaches aimed at quantifying detectability of clones in samples (i.e., [40, 41, 106] ) could be combined to create a unified evidence base to inform policy design.

Note that to take a sufficient amount of blood to genotype such a large number of markers (3 length-polymorphic marker loci, 7 microsatellite loci and between 3 and 5 AmpSeq loci) is likely to require more than the single finger prick (a simple and relatively unobtrusive procedure) that is currently taken for TES. Potential options include taking two finger pricks (from separate fingers), but if more blood was required venous blood may have to be taken – a more intrusive procedure with potential implications for obtaining patient consent. I will not attempt, here, to design such a trial, other than to note that an in vivo TES that attempts molecular correction using multiple methodologies will permit direct comparison of failure rate estimates between those methodologies. An obvious, yet important note is that when samples have been genotyped, calculating failure rates using different algorithms (i.e., for length-polymorphic markers) is a purely statistical procedure and does not require any additional blood from patients.

The expectation from this modelling work is that an AmpSeq approach as defined in **chapter 5** (or a more refined version of this approach in the future) should become the gold standard for obtaining failure rate estimates from TES. However, the relative nascency of this methodology and the next-generation technology required for its use is likely to mean its eventual roll-out takes place over the course of years, whilst accurate failure rates estimates are required now to aid the malaria elimination effort. Thus, in the interim period, improvements should be made to existing length-polymorphic marker and microsatellite methodology for TES: Length-polymorphic markers should be analysed with alternate algorithms, with the modelling work undertaken here suggesting use of an allelic family switch algorithm in areas with low ongoing transmission (quantified in the models by FOI) and a ≥2/3 markers algorithm in areas with mid to high ongoing transmission. Microsatellite markers should be analysed as standard with a Bayesian algorithm for accurate failure rate estimates (this has begun to occur [31])– but given the complexity of running this algorithm (**4.2.9**), development of a hub service or software to allow TES sites to submit trial results (microsatellites in initial and recurrent samples and day of recurrence is all that would be needed, preserving anonymity of samples) would be a worthwhile endeavour.

The improvement in failure rate estimates using a Bayesian algorithm to analyse microsatellite markers was notable (**Chapter 4**). Bayesian methodology would be equally applicable to length-polymorphic markers and AmpSeq methods, and improvements in the accuracy of efficacy estimates generated using these markers would be expected from the development and implementation of a Bayesian algorithm. Given the high accuracy of AmpSeq using simple calculations (**Chapter 5**), it is questionable whether developing Bayesian methods for AmpSeq would be a worthwhile endeavour. However, meaningful increases in accuracy could be expected from a Bayesian approach applied to length-polymorphic markers, and if the field is unable (or slow) to implement AmpSeq methodology, development of a Bayesian method for length-polymorphic markers is likely to be useful with relevant implications for policy makers. Note that a Bayesian method could be used to re-analyse historic trial data, and such a method would also be able to account for missing data (i.e., if only one or two of the three length-polymorphic markers were genotyped in a sample, see discussion in **3.2.8**).

Perhaps the major issue in the future of molecular genotyping for malaria TES will be consistency of results across the field. The proposed methods outlined in Table **7.1** will produce relatively comparable results with differences that are negligible relative to the under-estimation of true drug failure rates occurring with current methods. However, inertia in the field is likely to result in all three methods being used in practice with different algorithms (or thresholds for classifying a recrudescence within a match counting algorithm) being employed. Such a situation may lead to confusion where multiple failure rate estimates are available from a single TES. Multiple estimates from a single trial

are useful for modellers and policy makers as they can be cross-referenced against modelled results and permit validation of the models themselves (see **3.2.8** for an example of this). However, having to calculate and report multiple estimates from a single trial is likely to be undesirable for teams conducting these trials. Thus, it will be imperative that regulatory bodies such as the WHO and CDC produce revised guidelines as soon as possible. An exploratory trial genotyping and comparing multiple methodologies will be a crucial asset to allow them to do so.

### 7.1.6 Modelling gametocytes

A gametocyte model (**Figure 7.3** ) was explored for use with the length-polymorphic markers and later not used. This model required three gametocyte-related parameters: The gametocyte number per clone (as a proportion of the asexual parasitaemia of that clone), the lag period of *x* days before gametocyte numbers start to fall (they are not affected by drugs and so, while asexual parasites are still present in high numbers, gametocytes will continue to be produced – then fall when asexual parasites are cleared), and the gametocyte half-life $g_{1/2}$.

It was assumed that gametocytes are present for each clone at a density of 10% of the parasitaemia for that clone. That is to say, if there were 100 asexual parasites, there would be 10 gametocytes. This is likely to be a large over-estimate, it may be as few as 1 gametocyte to every 156 asexual parasites [196]. Any over-estimate of the true gametocyte/asexual parasite ratio will exaggerate the importance of gametocytes on failure rate estimates.

It was assumed that the lag period *x* was 4 days after the asexual parasites are killed via ACT for gametocyte numbers to begin to fall. This represents the time taken for gametocytes to circulate and become detectable following their production. This figure could be as high as 7-15 days [197] . A longer time here will increase the window with which meaningful gametocyte numbers can overlap with new infections and make misclassifications due to gametocytes more likely.

The gametocyte half-life $g_{1/2}$ was assumed to be 2.4 days, i.e. half of current gametocytes will perish every 2.4 days. This is in-line with estimates of between 2-4 days for half-life [198] and estimates of gametocyte carriage of between 5-10 days [199]. An increase in half-life will increase the window in which gametocytes can have a meaningful impact on failure rate estimates; a decrease will do the reverse.

Including the above criteria resulted in nearly no change to the failure rate estimates (when calculated using survival analysis) described in **chapter 3** over not modelling gametocytes (the largest change, in simulations of PPQ followed up for 63 days with an FOI of 16 and analysed with length-polymorphic markers as shown in **Figure 3.1** , was less than 0.1% - for lower FOIs, other ACTs and shorter follow-up periods the differences were even smaller (and often zero)). This is because the rate at which gametocytes decay mean they will only have any meaningful signal in the time period between the initial sample and around 13 days. However, for gametocytes to cause a misclassification of a new infection into a recrudescence, a reinfection must also occur and rise to a level at which it becomes detectable in this time period. Due to the prophylactic effects of all three ACTs, this is extremely rare (see **Figure 3.5**, **Figure 3.7** , **Figure 3.10** ). I show some illustrative examples of the likely impact of persisting gametocytes based around the fact that a patient can only diagnosed with malaria once the number of asexual parasites exceeds $10^8$ (the limit of detection via microscopy, fully discussed in **3.2.4**). When minority signals less than around 30% are not detected (i.e., for length-polymorphic markers or microsatellite markers), this means that gametocytes will only cause interference if their number exceeds~$3x10^7$, assuming that gametocytes are present in numbers of 10% of the asexual clone. **Figure 7.3** shows the change in gametocyte number a variety of initial numbers present at time of treatment, and different gametocyte half-lives. These results suggest they are only likely to contribute a signal for the first 10-14 days post-treatment. A longer half-life may extend this period up to 20 days. However, if the starting number of gametocytes is lowered to $10^8$ (reflective of being

1% of asexual parasitaemia of a single clone with a $10^{10}$ initial parasitaemia), they are never at a level at which they can impact results.

It may be that gametocyte signals are more important with the highly sensitive AmpSeq markers (**chapter 5**) where minority signals are detected at a frequency of ~1% of the majority signal. In **chapter 5**, gametocytes were not modelled, as I made the decision to keep the model simplified and consistent with **chapter 3** and **chapter 4**. A future extension of the AmpSeq work will include modelling of gametocytes and a full sensitivity analysis on gametocyte related parameters.

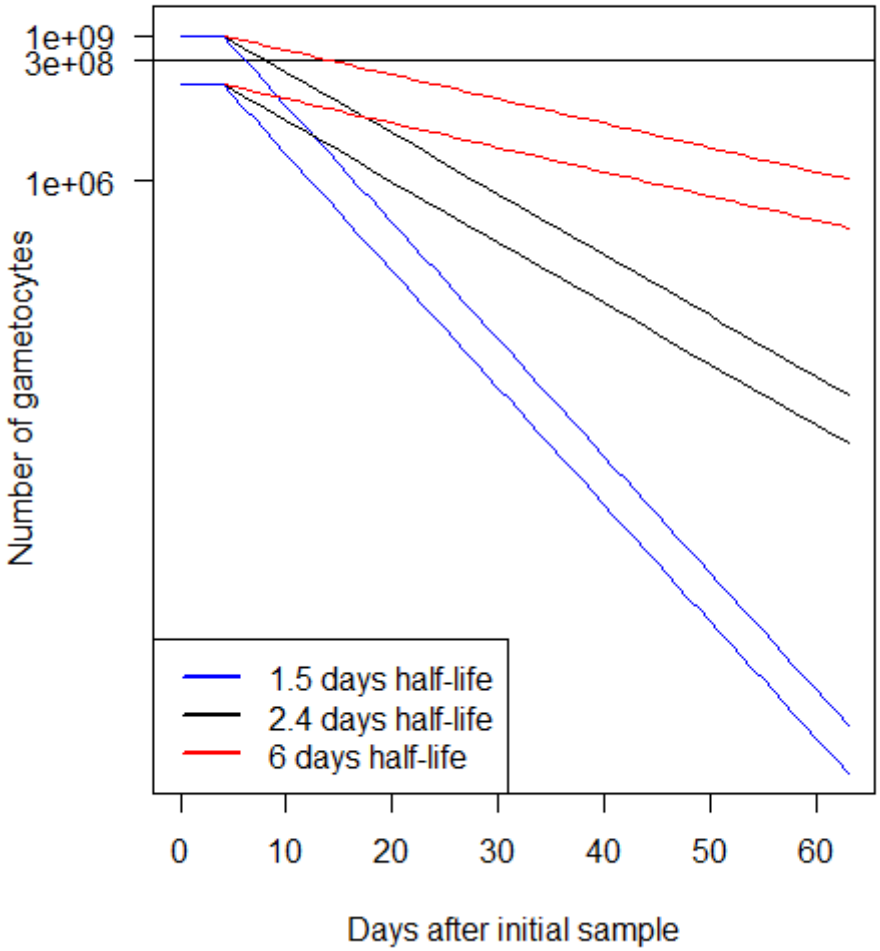



**Figure 7.3** Potential impact of modelling gametocytes with different starting densities (1% or 10% of asexual parasitaemia of $10^{10}$) and varying half-lives. Mathematically, gametocytes must be present in numbers of $3^8$ to have sufficient genetic signal to cause misclassification of reinfection as recrudescence – this only occurs extremely early in follow-up, when reinfection is unlikely to have occurred.

### 7.1.7 Future development of modelling work.

The mPK/PD models presented here for the topic of genetic detection should be easily adaptable in the future. They are easy to calibrate with new MOI distributions, FOI values or allelic frequency distributions. Including alternate loci is simple: For each of the methodologies that exist, additional genetic loci can be simply included providing their frequency distribution is known. Note that the Bayesian analysis method for microsatellite markers may need further adaptation for additional loci and the original authors of this analysis would have to be contacted in the first instance to facilitate this [150]. Additionally, new drugs (i.e., future potential calibrations of AS-PYN or AS-AQ, or other, yet to be discovered drugs) can be analysed provided the parameters required for their mechanistic PK simulation and relevant PD parameters are available. The calibrations of other parameters such as the threshold of the majority allele at which minority alleles were detected, the sampling limit (parasitaemia of a clone required to enter the blood sample) and initial parasite numbers were based on real laboratory processes but are easily variable within this methodology should new practices or information come to light.

One expansion of this work would relate to considerations of patients being lost to follow up (drop-out), follow-up schedule and strategy. Patient drop-out during TES may be relatively high, particularly for the longer periods of follow-up needed for drugs such as DHA-PPQ and AS-MQ with a long post-treatment prophylactic period [200]; a 2009 study comparing 15 TES from African countries and 14 TES from Thailand found that 4.7% and 14.3% of patients were lost to follow-up respectively [201]. Broadly, it takes one of two forms: Firstly, “permanent” drop-out, where a patient is not seen again after a given day of follow-up. Such a patient must be right-censored at that point for statistical analysis. Secondly, “periodic drop-out”, where a patient fails to attend a given day of follow-up but is seen later in the trial. For example, a patient may attend on day 3 and 7, not on day 14, but is seen again on day 21 [43]. Patient drop-out can have consequences on the accuracy of failure rate estimates: Permanent drop-out results in that patient being right-censored from the last day of follow-up on which they were seen – thus, any recrudescence they would later suffer would not be observed. Periodic drop-out carries the consequence that a patient’s recurrence may not be observed promptly as they skip a day of follow-up. This permits more time to pass until they are next seen by a clinician, and reinfections in that patient can grow to detectable levels in this time, and the chance of a recrudescence being misclassified as a reinfection increases. A wide variety of parameterization would have to be explored to include patient drop-out. For example, the chance of both types of drop-out may be correlated with total parasitaemia (if a patient’s parasitaemia drops, symptoms may reduce, and they may feel they do not need to attend future follow-up), and periodic drop-out may be correlated with previous instances of periodic drop-out (this could be mechanistically explained if, for example, a road were to be closed and the patient was unable to travel). Additionally, a patient may have decided not to attend follow-up (permanently) but then return to the clinic when they feel ill, possibly with recurrent malaria.

An mPK/PD approach can be expanded to examine entirely new systems of follow-up. The WHO follow-up schedule has a set routine; patients are followed up on day 3, 7, and every 7 days thereafter for the length of the TES (which differs based on the ACT given). “Sparse sampling” is a potential alternative method of trial design where patients are followed up over a longer period of time, but with more infrequent appointments. Additionally, patients are likely to self-report for an additional, unscheduled appointment if their symptoms worsen (in terms of an mPK/PD model this would be represented by increases in parasitaemia). Such a schedule has theorized operational advantages: it is simpler and more economical for both clinicians (fewer follow-up appointments) and patients (less time spent travelling to and from clinics). However, the question is whether accuracy of failure rate estimates would suffer with this approach; longer time between samples means that, for example, reinfections may emerge and mask the genetic signal of recrudescence. Additionally, the work on length-polymorphic markers in this thesis suggests that most recrudescence have occurred by 42 days for DHA-PPQ and AS-MQ and 28 days for AR-LF (**Figure 3.5**, **Figure 3.8** and **Figure 3.10** ) - longer

periods of follow-up may be un-necessary. However, sparse sampling with shorter periods of follow-up could be examined (i.e., modelling the current total duration of follow-up but with fewer individual days). The methodology contained in this thesis would be a highly appropriate vessel to examine which (if any) methodologies could produce high failure rate accuracy combined with operational advantage of sparse sampling.

Finally, there has recently been discussion about the suitability of statistical methods used to calculate failure rate estimates when numbers of recrudescence and reinfection have been determined in the molecular correction process. The WHO guidelines currently recommend the use of survival analysis to calculate failure rate estimates [28]. In the process of calculating failure rate estimates, survival analysis right-censors reinfections (**3.2.7**). Recent publications have suggested a competing risk analysis may be more suitable [108, 131, 202] as a patient having a reinfection emerge may prevent that patient from suffering a recrudescence that otherwise would have occurred. However, as the field moves forward to more accurate methods, inclusion of these novel analyses to further improve the accuracy of failure rate estimates would be prudent, and the mPK/PD work presented here provides a useful platform to quantify the difference in failure rate estimates produced from different statistical techniques (similar to the comparison of the per protocol method and survival analysis presented for length-polymorphic markers in **Chapter 3**).

## 7.2. Severe Malaria.

The modelling work in this thesis for uncomplicated malaria has focused on quantifying the accuracy of failure rate estimates obtained through TES. I also sought to apply a modelling approach to problems faced in clinical trials of severe malaria to demonstrate the role of computer models across the entire spectrum of the disease. It would not have been suitable to investigate long-term clearance of parasites for severe malaria – the key objective is patient survival (not parasite clearance) and so long-term TES are not conducted for the artesunate monotherapy used to treat severe malaria. Instead, I looked at a key issue in severe malaria (described in full in **6.1.1**): Pathology of severe malaria is caused by sequestered parasites, but severe malaria trials generally use metrics relating to circulating parasites to quantify their clinical outcomes.

**Chapter 6** presented a highly adaptable methodology for mPK/PD modelling of treatment of severe malaria that was able to recover key clinical observations (based on circulating parasite numbers), and, with novel metrics, used to investigate the pathology of severe malaria. The model showed that while on a population level a simplified artesunate regimen is non-inferior to the standard WHO regimen, outcomes in a sub-group of patients with infections grouped in late or early initial mean age-bins are notably worse with the simplified regimen. The emergence of artemisinin resistance in early ring stages poses a significant threat to this same group of patients. Neither of these results are particularly obvious from summary statistics of the population and so sub-group analysis is particularly important in devising treatment strategies for severe malaria.

The metrics presented: Area under the pathological load curve ($AUC_{PL}$) and Maximum pathological load (MPL) are intended to be simple for the purpose of designing a novel model to quantify the pathology of sequestered parasites. However, both metrics are highly plausible risk factors for poor outcomes, and it was extremely encouraging that the results comparing two treatment regimens were consistent across a wide range of model assumptions. An obvious expansion to this work is exploring the impact of different, more complicated metrics as fully discussed in **6.2.4**.

A key limitation of the model is that it is monoclonal, and commentators have noted that development of a polyclonal model would be a prudent expansion of this work [203], given that polyclonal sequestered parasites are observed in vivo [203, 204]. Multiple clones are unlikely to be synchronized with each other and so the importance of the initial mean age-bin parameter is likely to change, given

that the chance of a patient's entire infection being grouped in the high-risk early mean age-bins is extremely low. Density-dependent effects should also be modelled, and the presence of multiple desynchronized clones with varying PD parameters could alter the relative performance of drug regimens.

Editorial commentary on the published version this work also noted that alternative measures of validating the model should be considered – the model is validated by its ability to recover, given appropriate PK parameterization, parasite reduction ratios (PRR) from the studies in which PK parameters were drawn (**6.3.1**) [203]. As noted in **6. 1** and by Small and Seydel [203], circulating parasites do not reflect pathology of severe malaria and so alternate metrics with which to validate the model should be considered; Small and Seydel specifically suggested using levels of Plasmodium histidine rich protein-2 (PfHRP-2) and Plasmodium lactate dehydrogenase (pLDH) as they are reflective of total body parasite load. Further development of the model should seek to include a means of mathematically relating the level of total modelled parasitaemia to proxies such as PfHRP-2 and pLDH to allow for validation of the model with a wider range of studies.

In summary, while the model was designed with simplicity in mind given the importance of the central premise (that current methods of using parasite clearance to quantify outcome in severe malaria clinical trials are not fit for purpose), it must now be taken down a path of increasing complexity to improve realism and become more robust. Such a model can then be used as the basis with which to investigate a variety of research questions pertaining tosevere malaria, as has occurred for uncomplicated malaria where mPK/PD models have been used widely and are extensively validated (**Table 1.2**), with the ultimate goal of informing policy and reducing severe malaria mortality.

## 7.3 Concluding remarks: Development of the interface of computer modelling and malaria clinical trials.

The overarching goal of this thesis has been to identify research areas relating to anti-malarial clinical trials where gaps in knowledge or methodology could not be answered using in vivo or in vitro techniques, but which could be addressed using in silico mathematical models. This thesis is by no means an exhaustive pursuit of this ideal; modelling approaches are extremely flexible (**Table 1.2**) and a wealth of future contributions to the scientific evidence base surrounding malaria clearly remain untapped.

However, this thesis has contributed to malaria research using a modelling approach on two topics where the antecedent evidence base has been lacking: Firstly, the quantification of the accuracy of failure rate estimates in uncomplicated malaria TES, with research driven by the ability of an mPK/PD approach to know the true failure rate of a range of drugs, the development of models to calculate which genetic signals (for a range of different markers) are observed, and subsequent analysis of a range of algorithms with which to translate these observed signals into classifications of recurrent infections as recrudescence or reinfection. Consequently, this modelling work will be able to contribute to the ongoing discussion surrounding the best use of molecular correction in TES with evidence grounded using the "gold standard" of the true failure rate obtained using a modelling approach. Secondly, the development of a novel model to quantify the impact of sequestered parasites in severe malaria and the experimental use of that model in the context of investigating different drug regimens and the impact of artemisinin resistance. As described in **6. 1**, sequestered parasites cannot be measured in vivo and so current clinical outcomes of severe malaria trials are methodologically undesirable. In all the work undertaken herein, models are used to quantify either parasites or parameters that cannot be directly observed using currently available diagnostic methods.

These pieces of research have demonstrated the power of using a computer modelling approach to improve malaria clinical trials and created a foundation on which future research can occur. Work on

uncomplicated malaria trials (length-polymorphic markers) has already been published [93], work on microsatellite markers has been submitted for publication, and work on AmpSeq is being prepared for submission. The work on severe malaria presented in thesis has likewise been published [94], with commentators noting that it makes “a significant contribution to the field by incorporating detailed time-specific drug sensitivity as well as incorporating postdeath contribution to pathology into parasite burden models” [203].

I conclude this thesis with pleasure, having contributed meaningful results from computer modelling approaches to two important research areas in the ongoing quest to reduce the global burden of malaria, and excited to continue the development of such approaches in the near future.

# References.


1. World Health Organization. World malaria report 2018. Geneva2018. Available from: http://www.who.int/iris/handle/10665/275867.
2. Gething PW, Casey DC, Weiss DJ, Bisanzio D, Bhatt S, Cameron E, et al. Mapping plasmodium falciparum mortality in africa between 1990 and 2015. New England Journal of Medicine. 2016;375(25):2435-45.
3. World Health Organization. Global technical strategy for malaria 2016–20302016.
4. Ali AS, Thawer NG, Khatib B, Amier HH, Shija J, Msellem M, et al. Artemisinin combination therapy mass drug administration in a setting of low malaria endemicity: Programmatic coverage and adherence during an observational study in zanzibar. Malaria journal. 2017;16(1):332-.
5. von Seidlein L, Greenwood BM. Mass administrations of antimalarial drugs. Trends in Parasitology. 2003;19(10):452-60.
6. Newby G, Hwang J, Koita K, Chen I, Greenwood B, von Seidlein L, et al. Review of mass drug administration for malaria and its operational challenges. Am J Trop Med Hyg. 2015;93(1):125-34. Epub 2015/05/28.
7. Coleman PG, Alphey L. Editorial: Genetic control of vector populations: An imminent prospect. Tropical Medicine & International Health. 2004;9(4):433-7.
8. Benelli G, Beier JC. Current vector control challenges in the fight against malaria. Acta Trop. 2017;174:91-6. Epub 2017/07/08.
9. Karunamoorthi K. Vector control: A cornerstone in the malaria elimination campaign. Clin Microbiol Infect. 2011;17(11):1608-16. Epub 2011/10/15.
10. Ishizuka AS, Lyke KE, DeZure A, Berry AA, Richie TL, Mendoza FH, et al. Protection against malaria at 1 year and immune correlates following pfspz vaccination. Nat Med. 2016;22(6):614-23. Epub 2016/05/10.
11. Olotu A, Fegan G, Wambua J, Nyangweso G, Leach A, Lievens M, et al. Seven-year efficacy of rts,s/as01 malaria vaccine among young african children. New England Journal of Medicine. 2016;374(26):2519-29.
12. Mordmuller B, Surat G, Lagler H, Chakravarty S, Ishizuka AS, Lalremruata A, et al. Sterile protection against human malaria by chemoattenuated pfspz vaccine. Nature. 2017;542(7642):445-9. Epub 2017/02/16.
13. Olliaro P. Mortality associated with severe plasmodium falciparum malaria increases with age. Clinical Infectious Diseases. 2008;47(2):158-60.
14. World Health Organization, Malaria for Medicines Venture. Methods and techniques for clinical trials on antimalarial drug efficacy: Genotyping to identify parasite populations2008 2008. 54 p.
15. Hastings IM, Hodel EM, Kay K. Quantifying the pharmacology of antimalarial drug combination therapy. Scientific Reports. 2016;6:32762.
16. Cunnington AJ, Riley EM, Walther M. Stuck in a rut? Reconsidering the role of parasite sequestration in severe malaria syndromes. Trends in Parasitology. 2013;29(12):585-92.
17. Mackintosh CL, Beeson JG, Marsh K. Clinical features and pathogenesis of severe malaria. Trends in Parasitology. 2004;20(12):597-603.
18. Taylor T, Olola C, Valim C, Agbenyega T, Kremsner P, Krishna S, et al. Standardized data collection for multi-center clinical studies of severe malaria in african children: Establishing the smac network. Trans R Soc Trop Med Hyg. 2006;100(7):615-22. Epub 2006/03/23.
19. World Health Organization. Guidelines for the treatment of malaria.2015.
20. Saunders DL, Vanachayangkul P, Lon C. Dihydroartemisinin–piperaquine failure in cambodia. New England Journal of Medicine. 2014;371(5):484-5.
21. Wongsrichanalai C, Pickard AL, Wernsdorfer WH, Meshnick SR. Epidemiology of drug-resistant malaria. The Lancet Infectious Diseases. 2002;2(4):209-18.

22. Uhlemann A-C, Yuthavong Y, Fidock DA. Mechanisms of antimalarial drug action and resistance. Molecular approaches to malaria: American Society of Microbiology; 2005.
23. Dondorp AM, Yeung S, White L, Nguon C, Day NPJ, Socheat D, et al. Artemisinin resistance: Current status and scenarios for containment. Nature Reviews Microbiology. 2010;8:272.
24. World Health Organization. Eliminating malaria in the greater mekong subregion: United to end a deadly disease2016.
25. World Health Organization. Strategy for malaria elimination in the greater mekong subregion: 2015-2030 2015.
26. Lubell Y, Dondorp A, Guérin PJ, Drake T, Meek S, Ashley E, et al. Artemisinin resistance – modelling the potential human and economic costs. Malaria journal. 2014;13(1):452.
27. World Health Organization. Responding to antimalarial drug resistance 2018. Available from: https://www.who.int/malaria/areas/drug_resistance/overview/en/.
28. World Health Organization. Methods for surveillance of antimalarial drug efficacy2009 2009. 90 p.
29. World Health Organization. Assessment and monitoring of antimalarial drug efficacy for the treatment of uncomplicated falciparum malaria 2003. p. 68.
30. Plucinski MM, Talundzic E, Morton L, Dimbu PR, Macaia AP, Fortes F, et al. Efficacy of artemether-lumefantrine and dihydroartemisinin-piperaquine for treatment of uncomplicated malaria in children in zaire and uige provinces, angola. Antimicrob Agents Chemother. 2015;59(1):437-43. Epub 2014/11/05.
31. Davlantes E, Dimbu PR, Ferreira CM, Florinda Joao M, Pode D, Felix J, et al. Efficacy and safety of artemether-lumefantrine, artesunate-amodiaquine, and dihydroartemisinin-piperaquine for the treatment of uncomplicated plasmodium falciparum malaria in three provinces in angola, 2017. Malaria journal. 2018;17(1):144. Epub 2018/04/05.
32. Fulakeza JRM, Banda RL, Lipenga TR, Terlouw DJ, Nkhoma SC, Hodel EM. Comparison of two genotyping methods for distinguishing recrudescence from reinfection in antimalarial drug efficacy/effectiveness trials. American Journal of Tropical Medicine and Hygiene. 2018;99(1):84-6. Epub 2018/05/23.
33. Fulakeza J, McNitt S, Vareta J, Saidi A, Mvula G, Taylor T, et al. Comparison of *msp* genotyping and a 24 snp molecular assay for differentiating plasmodium falciparum recrudescence from reinfection. Malaria journal. 2019;18(1):84.
34. Assefa SA, Preston MD, Campino S, Ocholla H, Sutherland CJ, Clark TG. Estmoi: Estimating multiplicity of infection using parasite deep sequencing data. Bioinformatics (Oxford, England). 2014;30(9):1292-4. Epub 01/17.
35. Lalremruata A, Jeyaraj S, Engleitner T, Joanny F, Lang A, Bélard S, et al. Species and genotype diversity of plasmodium in malaria patients from gabon analysed by next generation sequencing. Malaria journal. 2017;16(1):398-.
36. Lerch A, Koepfli C, Hofmann NE, Messerli C, Wilcox S, Kattenberg JH, et al. Development of amplicon deep sequencing markers and data analysis pipeline for genotyping multi-clonal malaria infections. BMC Genomics. 2017;18(1):864.
37. Jaki T, Parry A, Winter K, Hastings I. Analysing malaria drug trials on a per-individual or per-clone basis: A comparison of methods. Statistics in Medicine. 2013;32(17):3020-38.
38. Snounou G, Beck HP. The use of pcr genotyping in the assessment of recrudescence or reinfection after antimalarial drug treatment. Parasitol Today. 1998;14(11):462-7. Epub 2006/10/17.
39. Anderson TJ, Su XZ, Bockarie M, Lagog M, Day KP. Twelve microsatellite markers for characterization of plasmodium falciparum from finger-prick blood samples. Parasitology. 1999;119 ( Pt 2):113-25. Epub 1999/08/31.
40. Greenhouse B, Myrick A, Dokomajilar C, Woo JM, Carlson EJ, Rosenthal PJ, et al. Validation of microsatellite markers for use in genotyping polyclonal plasmodium falciparum infections. Am J Trop Med Hyg. 2006;75(5):836-42.

41. Gruenberg M, Lerch A, Beck H-P, Felger I. Amplicon deep sequencing improves plasmodium falciparum genotyping in clinical trials of antimalarial drugs. [Research Article]. In press 2019.
42. Slater M, Kiggundu M, Dokomajilar C, Kamya MR, Bakyaita N, Talisuna A, et al. Distinguishing recrudescences from new infections in antimalarial clinical trials: Major impact of interpretation of genotyping results on estimates of drug efficacy. Am J Trop Med Hyg. 2005;73(2):256-62. Epub 2005/08/17.
43. Stepniewska K, White NJ. Some considerations in the design and interpretation of antimalarial drug trials in uncomplicated falciparum malaria. Malaria journal. 2006;5:127. Epub 2006/12/26.
44. Dondorp A, Nosten F, Stepniewska K, Day N, White N. Artesunate versus quinine for treatment of severe falciparum malaria: A randomised trial. Lancet. 2005;366(9487):717-25. Epub 2005/08/30.
45. Dondorp AM, Fanello CI, Hendriksen IC, Gomes E, Seni A, Chhaganlal KD, et al. Artesunate versus quinine in the treatment of severe falciparum malaria in african children (aquamat): An open-label, randomised trial. Lancet. 2010;376(9753):1647-57. Epub 2010/11/11.
46. Kremsner PG, Adegnika AA, Hounkpatin AB, Zinsou JF, Taylor TE, Chimalizeni Y, et al. Intramuscular artesunate for severe malaria in african children: A multicenter randomized controlled trial. PLOS Medicine. 2016;13(1):e1001938.
47. Kremsner PG, Taylor T, Issifou S, Kombila M, Chimalizeni Y, Kawaza K, et al. A simplified intravenous artesunate regimen for severe malaria–reply. The Journal of Infectious Diseases. 2012;206(4):622-.
48. van Hensbroek MB, Onyiorah E, Jaffar S, Schneider G, Palmer A, Frenkel J, et al. A trial of artemether or quinine in children with cerebral malaria. N Engl J Med. 1996;335(2):69-75. Epub 1996/07/11.
49. Esu E, Effa EE, Opie ON, Uwaoma A, Meremikwu MM. Artemether for severe malaria. The Cochrane Database of Systematic Reviews. 2014;(9):1-83.
50. Dondorp AM, Maude RJ, Hendriksen ICE, Day NP, White NJ. Artesunate dosing in severe falciparum malaria. The Journal of Infectious Diseases. 2012;206(4):618-9.
51. Kay K, Hodel EM, Hastings IM. Altering antimalarial drug regimens may dramatically enhance and restore drug effectiveness. Antimicrob Agents Chemother. 2015;59(10):6419-27.
52. Ferreira PE, Culleton R, Gil JP, Meshnick SR. Artemisinin resistance in plasmodium falciparum: What is it really? Trends Parasitol. 2013;29(7):318-20. Epub 2013/06/19.
53. Krishna S, Kremsner PG. Antidogmatic approaches to artemisinin resistance: Reappraisal as treatment failure with artemisinin combination therapy. Trends Parasitol. 2013;29(7):313-7. Epub 2013/04/30.
54. Dondorp AM, Desakorn V, Pongtavornpinyo W, Sahassananda D, Silamut K, Chotivanich K, et al. Estimation of the total parasite biomass in acute falciparum malaria from plasma *pfhrp2*. PLOS Medicine. 2005;2(8):e204-e. Epub 08/23.
55. Ross R. The prevention of malaria: J. Murray; 1910.
56. Reiner RC, Jr., Perkins TA, Barker CM, Niu T, Chaves LF, Ellis AM, et al. A systematic review of mathematical models of mosquito-borne pathogen transmission: 1970-2010. J R Soc Interface. 2013;10(81):20120921. Epub 2013/02/15.
57. Heffernan JM, Smith RJ, Wahl LM. Perspectives on the basic reproductive ratio. J R Soc Interface. 2005;2(4):281-93. Epub 06/07.
58. Chitnis N, Hyman JM, Cushing JM. Determining important parameters in the spread of malaria through the sensitivity analysis of a mathematical model. Bull Math Biol. 2008;70(5):1272-96. Epub 2008/02/23.
59. Smith DL, Dushoff J, Snow RW, Hay SI. The entomological inoculation rate and plasmodium falciparum infection in african children. Nature. 2005;438(7067):492-5.
60. Abu-Raddad LJ, Patnaik P, Kublin JG. Dual infection with hiv and malaria fuels the spread of both diseases in sub-saharan africa. Science. 2006;314(5805):1603-6. Epub 2006/12/13.

61. Mtisi E, Rwezaura H, Michel Tchuenche J. A mathematical analysis of malaria and tuberculosis co-dynamics2009. 827-64 p.
62. SMITH T, KILLEEN GF, MAIRE N, ROSS A, MOLINEAUX L, TEDIOSI F, et al. Mathematical modeling of the impact of malaria vaccines on the clinical epidemiology and natural history of plasmodium falciparum malaria: Overview. Am J Trop Med Hyg. 2006;75(2_suppl):1-10.
63. Penny MA, Verity R, Bever CA, Sauboin C, Galactionova K, Flasche S, et al. Public health impact and cost-effectiveness of the rts,s/as01 malaria vaccine: A systematic comparison of predictions from four mathematical models. Lancet. 2016;387(10016):367-75. Epub 2015/11/10.
64. Goodman CA, Coleman PG, Mills AJ. Cost-effectiveness of malaria control in sub-saharan africa. Lancet. 1999;354(9176):378-85. Epub 1999/08/07.
65. O'LEARY PA. Treatment of neurosyphilis by malaria: Report on the three years' observation of the first one hundred patients treated. JAMA. 1927;89(2):95-100.
66. Collins WE, Jeffery GM. A retrospective examination of the patterns of recrudescence in patients infected with plasmodium falciparum. Am J Trop Med Hyg. 1999;61(1 Suppl):44-8. Epub 1999/08/04.
67. Collins WE, Jeffery GM. A retrospective examination of secondary sporozoite- and trophozoite-induced infections with plasmodium falciparum: Development of parasitologic and clinical immunity following secondary infection. Am J Trop Med Hyg. 1999;61(1 Suppl):20-35. Epub 1999/08/04.
68. Collins WE, Jeffery GM. A retrospective examination of sporozoite- and trophozoite-induced infections with plasmodium falciparum: Development of parasitologic and clinical immunity during primary infection. Am J Trop Med Hyg. 1999;61(1 Suppl):4-19. Epub 1999/08/04.
69. Gambino M. Fevered decisions: Race, ethics, and clinical vulnerability in the malarial treatment of neurosyphilis, 1922–1953. Hastings Center Report. 2015;45(4):39-50.
70. Childs LM, Buckee CO. Dissecting the determinants of malaria chronicity: Why within-host models struggle to reproduce infection dynamics. J R Soc Interface. 2015;12(104):20141379. Epub 2015/02/13.
71. Molineaux L, Diebner HH, Eichner M, Collins WE, Jeffery GM, Dietz K. Plasmodium falciparum parasitaemia described by a new mathematical model. Parasitology. 2002;122(4):379-91. Epub 06/18.
72. Johnston GL, Smith DL, Fidock DA. Malaria's missing number: Calculating the human component of r0 by a within-host mechanistic model of plasmodium falciparum infection and transmission. PLOS Computational Biology. 2013;9(4):e1003025.
73. Maire N, Smith T, Ross A, Owusu-Agyei S, Dietz K, Molineaux L. A model for natural immunity to asexual blood stages of plasmodium falciparum malaria in endemic areas. Am J Trop Med Hyg. 2006;75(2 Suppl):19-31. Epub 2006/08/26.
74. Felger I, Maire M, Bretscher MT, Falk N, Tiaden A, Sama W, et al. The dynamics of natural plasmodium falciparum infections. PLOS ONE. 2012;7(9):e45542.
75. Molineaux L, Dietz K. Review of intra-host models of malaria. Parassitologia. 1999;41:221-31.
76. Dietz K, Raddatz G, Molineaux L. Mathematical model of the first wave of plasmodium falciparum asexual parasitemia in non-immune and vaccinated individuals. American Journal of Tropical Medicine and Hygiene. 2006;75(2 Suppl):46-55. Epub 2006/08/26.
77. Hoshen MB, Heinrich R, Stein WD, Ginsburg H. Mathematical modelling of the within-host dynamics of plasmodium falciparum. Parasitology. 2000;121 ( Pt 3):227-35. Epub 2000/11/21.
78. Langhorne J, Ndungu FM, Sponaas A-M, Marsh K. Immunity to malaria: More questions than answers. Nature Immunology. 2008;9(7):725-32.
79. Recker M, Buckee CO, Serazin A, Kyes S, Pinches R, Christodoulou Z, et al. Antigenic variation in plasmodium falciparum malaria involves a highly structured switching pattern. PLOS Pathogens. 2011;7(3):e1001306.
80. Roberts DJ, Craig, Berendt AR, Pinches R, Nash G, Marsh K, et al. Rapid switching to multiple antigenic and adhesive phenotypes in malaria. Nature. 1992;357(6380):689-92.

81. Cunnington AJ, Walther M, Riley EM. Piecing together the puzzle of severe malaria. Science Translational Medicine. 2013;5(211):211ps18-ps18.
82. Patel K, Simpson JA, Batty KT, Zaloumis S, Kirkpatrick CM. Modelling the time course of antimalarial parasite killing: A tour of animal and human models, translation and challenges. Br J Clin Pharmacol. 2015;79(1):97-107. Epub 2013/11/21.
83. Slater HC, Okell LC, Ghani AC. Mathematical modelling to guide drug development for malaria elimination. Trends in Parasitology. 2017;33(3):175-84. Epub 2016/10/30.
84. Czock D, Keller F. Mechanism-based pharmacokinetic-pharmacodynamic modeling of antimicrobial drug effects. J Pharmacokinet Pharmacodyn. 2007;34(6):727-51. Epub 2007/10/02.
85. Hodel EM, Kay K, Hayes DJ, Terlouw DJ, Hastings IM. Optimizing the programmatic deployment of the anti-malarials artemether-lumefantrine and dihydroartemisinin-piperaquine using pharmacological modelling. Malaria journal. 2014;13:138. Epub 2014/04/09.
86. Simpson JA, Watkins ER, Price RN, Aarons L, Kyle DE, White NJ. Mefloquine pharmacokinetic-pharmacodynamic models: Implications for dosing and resistance. Antimicrob Agents Chemother. 2000;44(12):3414-24.
87. Kay K, Hodel EM, Hastings IM. Improving the role and contribution of pharmacokinetic analyses in antimalarial drug clinical trials. Antimicrob Agents Chemother. 2014;58(10):5643-9. Epub 2014/07/02.
88. Dini S, Zaloumis S, Cao P, Price RN, Fowkes FJI, van der Pluijm RW, et al. Investigating the efficacy of triple artemisinin-based combination therapies for treating plasmodium falciparum malaria patients using mathematical modeling. Antimicrob Agents Chemother. 2018;62(11). Epub 2018/08/29.
89. Challenger JD, Bruxvoort K, Ghani AC, Okell LC. Assessing the impact of imperfect adherence to artemether-lumefantrine on malaria treatment outcomes using within-host modelling. Nat Commun. 2017;8(1):1373. Epub 2017/11/11.
90. Kay K, Hastings IM. Measuring windows of selection for anti-malarial drug treatments. Malaria journal. 2015;14:292.
91. Stepniewska K, White NJ. Pharmacokinetic determinants of the window of selection for antimalarial drug resistance. Antimicrob Agents Chemother. 2008;52(5):1589-96. Epub 2008/02/27.
92. Klonis N, Xie SC, McCaw JM, Crespo-Ortiz MP, Zaloumis SG, Simpson JA, et al. Altered temporal response of malaria parasites determines differential sensitivity to artemisinin. Proceedings of the National Academy of Sciences. 2013;110(13):5157-62.
93. Jones S, Kay K, Hodel EM, Chy S, Mbituyumuremyi A, Uwimana A, et al. Improving methods for analysing anti-malarial drug efficacy trials: Molecular correction based on length-polymorphic markers *msp-1*, *msp-2* and *glurp*. Antimicrob Agents Chemother. 2019:AAC.00590-19.
94. Jones S, Hodel EM, Sharma R, Kay K, Hastings IM. Optimal treatments for severe malaria and the threat posed by artemisinin resistance. The Journal of Infectious Diseases. 2018;219(8):1243-53.
95. Johnston GL, Gething PW, Hay SI, Smith DL, Fidock DA. Modeling within-host effects of drugs on plasmodium falciparum transmission and prospects for malaria elimination. PLOS Computational Biology. 2014;10(1):e1003434.
96. Gerardin J, Eckhoff P, Wenger EA. Mass campaigns with antimalarial drugs: A modelling comparison of artemether-lumefantrine and dha-piperaquine with and without primaquine as tools for malaria control and elimination. BMC Infectious Disease. 2015;15:144. Epub 2015/04/19.
97. Nguyen TD, Olliaro P, Dondorp AM, Baird JK, Lam HM, Farrar J, et al. Optimum population-level use of artemisinin combination therapies: A modelling study. The Lancet Global Health. 2015;3(12):e758-e66.
98. Hwang J, Alemayehu BH, Reithinger R, Tekleyohannes SG, Takele T, Birhanu SG, et al. In vivo efficacy of artemether-lumefantrine and chloroquine against plasmodium vivax: A randomized open label trial in central ethiopia. PLOS ONE. 2013;8(5):e63433.

99. Stepniewska K, Taylor W, Sirima SB, Ouedraogo EB, Ouedraogo A, Gansané A, et al. Population pharmacokinetics of artesunate and amodiaquine in african children. Malaria journal. 2009;8(1):200.
100. Hong KB, Hastings I, Kay K, Hodel EM. Evaluating artesunate-amodiaquine deployment, efficacy and safety: An in silico pharmacological model. bioRxiv. 2019:567008.
101. Htay MNN. Pharamcological modelling of intermittent preventative treatment in pregnancy with sulpadoxine-pyrimethamine [MSc thesis]. Liverpool: Liverpool School of Tropical Medicine; 2014.
102. Sagara I, Beavogui AH, Zongo I, Soulama I, Borghini-Fuhrer I, Fofana B, et al. Pyronaridine-artesunate or dihydroartemisinin-piperaquine versus current first-line therapies for repeated treatment of uncomplicated malaria: A randomised, multicentre, open-label, longitudinal, controlled, phase 3b/4 trial. Lancet. 2018;391(10128):1378-90.
103. Croft SL, Duparc S, Arbe-Barnes SJ, Craft JC, Shin C-S, Fleckenstein L, et al. Review of pyronaridine anti-malarial properties and product characteristics. Malaria journal. 2012;11(1):270.
104. Kay K, Hastings IM. Improving pharmacokinetic-pharmacodynamic modeling to investigate anti-infective chemotherapy with application to the current generation of antimalarial drugs. PLOS Computational Biology. 2013;9(7):e1003151.
105. Winter K, Hastings IM. Development, evaluation, and application of an in silico model for antimalarial drug treatment and failure. Antimicrob Agents Chemother. 2011;55(7):3380-92. Epub 2011/05/04.
106. Messerli C, Hofmann NE, Beck HP, Felger I. Critical evaluation of molecular monitoring in malaria drug efficacy trials and pitfalls of length-polymorphic markers. Antimicrob Agents Chemother. 2017;61(1). Epub 2016/11/09.
107. Ken-Dror G, Hastings IM. Markov chain monte carlo and expectation maximization approaches for estimation of haplotype frequencies for multiply infected human blood samples. Malaria journal. 2016;15(1):430. Epub 2016/08/26.
108. Dahal P, Guerin PJ, Price RN, Simpson JA, Stepniewska K. Evaluating antimalarial efficacy in single-armed and comparative drug trials using competing risk survival analysis: A simulation study. BMC medical research methodology. 2019;19(1):107-.
109. White MT, Karl S, Koepfli C, Longley RJ, Hofmann NE, Wampfler R, et al. Plasmodium vivax and plasmodium falciparum infection dynamics: Re-infections, recrudescences and relapses. Malaria journal. 2018;17(1):170-.
110. Post TM, Freijer JI, Ploeger BA, Danhof M. Extensions to the visual predictive check to facilitate model performance evaluation. J Pharmacokinet Pharmacodyn. 2008;35(2):185-202. Epub 01/16.
111. Kay K, Hodel EM, Hastings IM. Altering antimalarial drug regimens may dramatically enhance and restore drug effectiveness. Antimicrobial Agents and Chemotherapy. 2015.
112. Staehli Hodel EM, Guidi M, Zanolari B, Mercier T, Duong S, Kabanywanyi AM, et al. Population pharmacokinetics of mefloquine, piperaquine and artemether-lumefantrine in cambodian and tanzanian malaria patients. Malaria journal. 2013;12:235-.
113. Tarning J, Rijken MJ, McGready R, Phyo AP, Hanpithakpong W, Day NP, et al. Population pharmacokinetics of dihydroartemisinin and piperaquine in pregnant and nonpregnant women with uncomplicated malaria. Antimicrob Agents Chemother. 2012;56(4):1997-2007. Epub 2012/01/19.
114. World Health Organization. Malaria parasite counting, malaria microscopy standard operating procedure – mm-sop-09. 2016.
115. Mueller I, Schoepflin S, Smith TA, Benton KL, Bretscher MT, Lin E, et al. Force of infection is key to understanding the epidemiology of plasmodium falciparum malaria in papua new guinean children. Proc Natl Acad Sci U S A. 2012;109(25):10030-5. Epub 2012/06/06.
116. Smith T, Maire N, Dietz K, Killeen G, Vounatsou P, Molineaux L. Relationships between the entomological inoculation rate and the force of infection for plasmodium falciparum malaria. Am J Trop Med Hyg. 2006;75(Suppl 2):11 - 8.

117. Garnham PCC. Malaria parasites and other haemosporidia. 5, Alfred Street, Oxford: Blackwell Scientific Publications Ltd.; 1966. xviii + 1114 pp. p.
118. Adu-Gyasi D, Adams M, Amoako S, Mahama E, Nsoh M, Amenga-Etego S, et al. Estimating malaria parasite density: Assumed white blood cell count of 10,000/µl of blood is appropriate measure in central ghana. Malaria journal. 2012;11:238-.
119. Bousema T, Okell L, Shekalaghe S, Griffin JT, Omar S, Sawa P, et al. Revisiting the circulation time of plasmodium falciparum gametocytes: Molecular detection methods to estimate the duration of gametocyte carriage and the effect of gametocytocidal drugs. Malaria journal. 2010;9(1):136.
120. Greenhouse B, Dokomajilar C, Hubbard A, Rosenthal PJ, Dorsey G. Impact of transmission intensity on the accuracy of genotyping to distinguish recrudescence from new infection in antimalarial clinical trials. Antimicrob Agents Chemother. 2007;51(9):3096-103. Epub 2007/06/27.
121. Juliano JJ, Gadalla N, Sutherland CJ, Meshnick SR. The perils of pcr: Can we accurately 'correct' antimalarial trials? Trends in Parasitology. 2010;26(3):119-24. Epub 2010/01/20.
122. Plucinski MM, Morton L, Bushman M, Dimbu PR, Udhayakumar V. Robust algorithm for systematic classification of malaria late treatment failures as recrudescence or reinfection using microsatellite genotyping. Antimicrob Agents Chemother. 2015;59(10):6096-100. Epub 2015/07/22.
123. CATTAMANCHI A, KYABAYINZE D, HUBBARD A, ROSENTHAL PJ, DORSEY G. Distinguishing recrudescence from reinfection in a longitudinal antimalarial drug efficacy study: Comparison of results based on genotyping of *msp-1*, *msp-2*, and *glurp*. Am J Trop Med Hyg. 2003;68(2):133-9.
124. Liljander A, Wiklund L, Falk N, Kweku M, Mårtensson A, Felger I, et al. Optimization and validation of multi-coloured capillary electrophoresis for genotyping of plasmodium falciparum merozoite surface proteins (msp1 and 2). Malaria journal. 2009;8:78-.
125. Hastings IM, Nsanzabana C, Smith TA. A comparison of methods to detect and quantify the markers of antimalarial drug resistance. Am J Trop Med Hyg. 2010;83(3):489-95.
126. Schoepflin S, Valsangiacomo F, Lin E, Kiniboro B, Mueller I, Felger I. Comparison of plasmodium falciparum allelic frequency distribution in different endemic settings by high-resolution genotyping. Malaria journal. 2009;8(1):250.
127. Duru V, Witkowski B, Menard D. Plasmodium falciparum resistance to artemisinin derivatives and piperaquine: A major challenge for malaria elimination in cambodia. American Journal of Tropical Medicine and Hygiene. 2016;95(6):1228-38. Epub 2016/12/09.
128. Slater HC, Ross A, Felger I, Hofmann NE, Robinson L, Cook J, et al. The temporal dynamics and infectiousness of subpatent plasmodium falciparum infections in relation to parasite density. Nat Commun. 2019;10(1):1433-.
129. Hastings IM, Kay K, Hodel EM. How robust are malaria parasite clearance rates as indicators of drug effectiveness and resistance? Antimicrob Agents Chemother. 2015;59:6428-36.
130. Leang R, Canavati SE, Khim N, Vestergaard LS, Borghini Fuhrer I, Kim S, et al. Efficacy and safety of pyronaridine-artesunate for treatment of uncomplicated plasmodium falciparum malaria in western cambodia. Antimicrob Agents Chemother. 2016;60(7):3884-90.
131. Dahal P, Simpson JA, Dorsey G, Guérin PJ, Price RN, Stepniewska K. Statistical methods to derive efficacy estimates of anti-malarials for uncomplicated plasmodium falciparum malaria: Pitfalls and challenges. Malaria journal. 2017;16(1):430.
132. Tinto H, Diallo S, Zongo I, Guiraud I, Valea I, Kazienga A, et al. Effectiveness of artesunate-amodiaquine vs. Artemether-lumefantrine for the treatment of uncomplicated falciparum malaria in nanoro, burkina faso: A non-inferiority randomised trial. Tropical Medicine & International Health. 2014;19(4):469-75.
133. Ndounga M, Mayengue PI, Casimiro PN, Loumouamou D, Basco LK, Ntoumi F, et al. Artesunate-amodiaquine efficacy in congolese children with acute uncomplicated falciparum malaria in brazzaville. Malaria journal. 2013;12:53.
134. Shayo A, Mandara CI, Shahada F, Buza J, Lemnge MM, Ishengoma DS. Therapeutic efficacy and safety of artemether-lumefantrine for the treatment of uncomplicated falciparum malaria in north-eastern tanzania. Malaria journal. 2014;13:376.

135. Ogouyemi-Hounto A, Azandossessi C, Lawani S, Damien G, de Tove YS, Remoue F, et al. Therapeutic efficacy of artemether-lumefantrine for the treatment of uncomplicated falciparum malaria in northwest benin. Malaria journal. 2016;15:37.
136. Salvador C, Rafael B, Matsinhe F, Candrinho B, Muthemba R, De Carvalho E, et al. Efficacy and safety of artemether-lumefantrine for the treatment of uncomplicated falciparum malaria at sentinel sites in mozambique, 2015. Acta Trop. 2017;171:146-50.
137. Bharti PK, Shukla MM, Ringwald P, Krishna S, Singh PP, Yadav A, et al. Therapeutic efficacy of artemether-lumefantrine for the treatment of uncomplicated plasmodium falciparum malaria from three highly malarious states in india. Malaria journal. 2016;15(1):498.
138. de Wit M, Funk AL, Moussally K, Nkuba DA, Siddiqui R, Bil K, et al. In vivo efficacy of artesunate-amodiaquine and artemether-lumefantrine for the treatment of uncomplicated falciparum malaria: An open-randomized, non-inferiority clinical trial in south kivu, democratic republic of congo. Malaria journal. 2016;15:455.
139. Kay K, Hastings IM. Measuring windows of selection for anti-malarial drug treatments. Malaria journal. 2015;14(1):292.
140. Hodel EM, Kay K, Hastings IM. Incorporating stage specificity into pharmacological modelling of antimalarial drug treatment. Antimicrob Agents Chemother. 2016;60:2747-56.
141. Nyachieo A, C VANO, Laurent T, Dujardin JC, D'Alessandro U. Plasmodium falciparum genotyping by microsatellites as a method to distinguish between recrudescent and new infections. Am J Trop Med Hyg. 2005;73(1):210-3. Epub 2005/07/15.
142. Greenhouse B, Dokomajilar C, Hubbard A, Rosenthal PJ, Dorsey G. Impact of transmission intensity on the accuracy of genotyping to distinguish recrudescence from new infection in antimalarial clinical trials. Antimicrobial Agents and Chemotherapy. 2007;51(9):3096-103.
143. Malvy D, Torrentino-Madamet M, L'Ollivier C, Receveur M-C, Jeddi F, Delhaes L, et al. Plasmodium falciparum recrudescence two years after treatment of an uncomplicated infection without return to an area where malaria is endemic. Antimicrob Agents Chemother. 2018;62(2):e01892-17.
144. Russo G, L'Episcopia M, Menegon M, Souza SS, Dongho BGD, Vullo V, et al. Dihydroartemisinin-piperaquine treatment failure in uncomplicated plasmodium falciparum malaria case imported from ethiopia. Infection. 2018;46(6):867-70. Epub 2018/07/08.
145. Plucinski MM, Huber CS, Akinyi S, Dalton W, Eschete M, Grady K, et al. Novel mutation in cytochrome b of plasmodium falciparum in one of two atovaquone-proguanil treatment failures in travelers returning from same site in nigeria. Open forum infectious diseases. 2014;1(2):ofu059-ofu.
146. Mwangi JM, Omar SA, Ranford-Cartwright LC. Comparison of microsatellite and antigen-coding loci for differentiating recrudescing plasmodium falciparum infections from reinfections in kenya. International Journal for Parasitology. 2006;36(3):329-36.
147. Hosseinzadeh-Colagar A, Haghighatnia MJ, Amiri Z, Mohadjerani M, Tafrihi M. Microsatellite (ssr) amplification by pcr usually led to polymorphic bands: Evidence which shows replication slippage occurs in extend or nascent DNA strands. Molecular biology research communications. 2016;5(3):167-74.
148. Pasqualotto AC, Denning DW, Anderson MJ. A cautionary tale: Lack of consistency in allele sizes between two laboratories for a published multilocus microsatellite typing system. Journal of Clinical Microbiology. 2007;45(2):522-8.
149. Davison A, Chiba S. Laboratory temperature variation is a previously unrecognized source of genotyping error during capillary electrophoresis. Molecular Ecology Notes. 2003;3(2):321-3.
150. Plucinski MM, Morton L, Bushman M, Dimbu PR, Udhayakumar V. Robust algorithm for systematic classification of malaria late treatment failures as recrudescence or reinfection using microsatellite genotyping. Antimicrobial Agents and Chemotherapy. 2015;59(10):6096-100.
151. Plucinski MM, Dimbu PR, Macaia AP, Ferreira CM, Samutondo C, Quivinja J, et al. Efficacy of artemether–lumefantrine, artesunate–amodiaquine, and dihydroartemisinin–piperaquine for

5096 treatment of uncomplicated plasmodium falciparum malaria in angola, 2015. Malaria journal.
5097 2017;16(1):62.
5098 152. Siahaan L. Laboratory diagnostics of malaria. IOP Conference Series: Earth and
5099 Environmental Science. 2018;125:012090.
5100 153. Zweig MH, Campbell G. Receiver-operating characteristic (roc) plots: A fundamental
5101 evaluation tool in clinical medicine. Clinical Chemistry. 1993;39(4):561-77.
5102 154. Miller RH, Hathaway NJ, Kharabora O, Mwandagalirwa K, Tshefu A, Meshnick SR, et al. A
5103 deep sequencing approach to estimate plasmodium falciparum complexity of infection (coi) and
5104 explore apical membrane antigen 1 diversity. Malaria journal. 2017;16(1):490.
5105 155. Talundzic E, Ravishankar S, Kelley J, Patel D, Plucinski M, Schmedes S, et al. Next-generation
5106 sequencing and bioinformatics protocol for malaria drug resistance marker surveillance. Antimicrob
5107 Agents Chemother. 2018;62(4):e02474-17.
5108 156. Rao PN, Uplekar S, Kayal S, Mallick PK, Bandyopadhyay N, Kale S, et al. A method for
5109 amplicon deep sequencing of drug resistance genes in plasmodium falciparumclinical isolates from
5110 india. Journal of Clinical Microbiology. 2016;54(6):1500-11.
5111 157. Neafsey DE, Juraska M, Bedford T, Benkeser D, Valim C, Griggs A, et al. Genetic diversity and
5112 protective efficacy of the rts,s/as01 malaria vaccine. New England Journal of Medicine.
5113 2015;373(21):2025-37.
5114 158. Early AM, Daniels RF, Farrell TM, Grimsby J, Volkman SK, Wirth DF, et al. Detection of low-
5115 density plasmodium falciparum infections using amplicon deep sequencing. Malaria journal.
5116 2019;18(1):219.
5117 159. Juliano JJ, Porter K, Mwapasa V, Sem R, Rogers WO, Ariey F, et al. Exposing malaria in-host
5118 diversity and estimating population diversity by capture-recapture using massively parallel
5119 pyrosequencing. Proc Natl Acad Sci U S A. 2010;107(46):20138-43. Epub 10/19.
5120 160. Lerch A, Koepfli C, Hofmann NE, Kattenberg JH, Rosanas-Urgell A, Betuela I, et al.
5121 Longitudinal tracking and quantification of individual plasmodium falciparum clones in complex
5122 infections. Scientific Reports. 2019;9(1):3333.
5123 161. Gupta V, Dorsey G, Hubbard AE, Rosenthal PJ, Greenhouse B. Gel versus capillary
5124 electrophoresis genotyping for categorizing treatment outcomes in two anti-malarial trials in
5125 uganda. Malaria journal. 2010;9(1):19.
5126 162. Loman NJ, Misra RV, Dallman TJ, Constantinidou C, Gharbia SE, Wain J, et al. Performance
5127 comparison of benchtop high-throughput sequencing platforms. Nat Biotechnol. 2012;30(5):434-9.
5128 Epub 2012/04/24.
5129 163. Dondorp AM, Kager PA, Vreeken J, White NJ. Abnormal blood flow and red blood cell
5130 deformability in severe malaria. Parasitology Today. 2000;16(6):228-32. Epub 2000/05/29.
5131 164. Clark IA, Alleva LM. Is human malarial coma caused, or merely deepened, by sequestration?
5132 Trends in Parasitology. 2009;25(7):314-8. Epub 2009/06/23.
5133 165. Medana IM, Turner GDH. Plasmodium falciparum and the blood-brain barrier—contacts and
5134 consequences. The Journal of Infectious Diseases. 2007;195(7):921-3.
5135 166. Smith JD, Rowe JA, Higgins MK, Lavstsen T. Malaria's deadly grip: Cytoadhesion of
5136 plasmodium falciparum-infected erythrocytes. Cell Microbiol. 2013;15(12):1976-83. Epub
5137 2013/08/21.
5138 167. Hughes KR, Biagini GA, Craig AG. Continued cytoadherence of plasmodium falciparum
5139 infected red blood cells after antimalarial treatment. Mol Biochem Parasitol. 2010;169(2):71-8. Epub
5140 2009/10/06.
5141 168. Udomsangpetch R, Pipitaporn B, Krishna S, Angus B, Pukrittayakamee S, Bates I, et al.
5142 Antimalarial drugs reduce cytoadherence and rosetting plasmodium falciparum. The Journal of
5143 Infectious Diseases. 1996;173(3):691-8. Epub 1996/03/01.
5144 169. Team RC. R: A language and environment for statistical computing. R foundation for
5145 statistical

5146 computing, vienna, austria. 2014.

170. Kremsner PG, Adegnika AA, Hounkpatin AB, Zinsou JF, Taylor TE, Chimalizeni Y, et al. Intramuscular artesunate for severe malaria in african children: A multicenter randomized controlled trial. PLoS Med. 2016;13(1):e1001938.
171. Kendjo E, Agbenyega T, Bojang K, Newton CRJC, Bouyou-Akotet M, Pedross F, et al. Mortality patterns and site heterogeneity of severe malaria in african children. PLOS ONE. 2013;8(3):e58686.
172. Maitland K. Management of severe paediatric malaria in resource-limited settings. BMC Medicine. 2015;13(1):42.
173. Saralamba S, Pan-Ngum W, Maude RJ, Lee SJ, Tarning J, Lindegardh N, et al. Intrahost modeling of artemisinin resistance in plasmodium falciparum. Proc Natl Acad Sci U S A. 2011;108(1):397-402.
174. White NJ. The parasite clearance curve. Malaria journal. 2011;10:278. Epub 2011/09/24.
175. White NJ. Malaria parasite clearance. Malaria journal. 2017;16(1):88. Epub 2017/02/25.
176. Gordi T, Xie R, Jusko WJ. Semi-mechanistic pharmacokinetic/pharmacodynamic modelling of the antimalarial effect of artemisinin. Br J Clin Pharmacol. 2005;60(6):594-604. Epub 2005/11/25.
177. Hietala SF, Martensson A, Ngasala B, Dahlstrom S, Lindegardh N, Annerberg A, et al. Population pharmacokinetics and pharmacodynamics of artemether and lumefantrine during combination treatment in children with uncomplicated falciparum malaria in tanzania. Antimicrob Agents Chemother. 2010;54(11):4780-8. Epub 2010/08/18.
178. Milner DA, Vareta J, Valim C, Montgomery J, Daniels RF, Volkman SK, et al. Human cerebral malaria and plasmodium falciparum genotypes in malawi. Malaria journal. 2012;11(1):35.
179. Oyedeji SI, Awobode HO, Kun J. Limited genetic diversity and low multiplicity of plasmodium falciparum infections in children with severe malaria in lafia, north-central nigeria. Journal of Experimental & Clinical Medicine. 2013;5(4):143-7.
180. Hoshen MB, Stein WD, Ginsburg H. Modelling the chloroquine chemotherapy of falciparum malaria: The value of spacing a split dose. Parasitology. 1998;116 ( Pt 5):407-16. Epub 1998/06/06.
181. World Health Organization. Management of severe malaria2012.
182. Hendriksen IC, Mtove G, Kent A, Gesase S, Reyburn H, Lemnge MM, et al. Population pharmacokinetics of intramuscular artesunate in african children with severe malaria: Implications for a practical dosing regimen. Clin Pharmacol Ther. 2013;93(5):443-50. Epub 2013/03/21.
183. World Health Organization. World malaria report 20162016 13 December 2016. 186 p.
184. Zaloumis SG, Tarning J, Krishna S, Price RN, White NJ, Davis TME, et al. Population pharmacokinetics of intravenous artesunate: A pooled analysis of individual data from patients with severe malaria. CPT: Pharmacometrics & Systems Pharmacology. 2014;3(11):e145.
185. Hastings IM, Hodel EM, Kay K. Quantifying the pharmacology of antimalarial drug combination therapy. Sci Rep. 2016;6:32762.
186. Fairhurst RM, Dondorp AM. Artemisinin-resistant plasmodium falciparum malaria. Microbiol Spectr. 2016;4(3). Epub 2016/06/24.
187. Färnert A, Lebbad M, Faraja L, Rooth I. Extensive dynamics of plasmodium falciparum densities, stages and genotyping profiles. Malaria journal. 2008;7(1):241.
188. Paget-McNicol S, Gatton M, Hastings I, Saul A. The plasmodium falciparum var gene switching rate, switching mechanism and patterns of parasite recrudescence described by mathematical modelling. Parasitology. 2002;124(Pt 3):225-35. Epub 2002/04/02.
189. Chua CLL, Brown G, Hamilton JA, Rogerson S, Boeuf P. Monocytes and macrophages in malaria: Protection or pathology? Trends in Parasitology. 2013;29(1):26-34.
190. Kremsner PG, Taylor T, Issifou S, Kombila M, Chimalizeni Y, Kawaza K, et al. A simplified intravenous artesunate regimen for severe malaria. The Journal of Infectious Diseases. 2011.
191. White NJ. Assessment of the pharmacodynamic properties of antimalarial drugs in vivo. Antimicrob Agents Chemother. 1997;41(7):1413-22.
192. Hinkle DE WW, Jurs SG. Applied statistics for the behavioral sciences. 5th ed: Boston: Houghton Mifflin; 2003.

193. Zaloumis S, Humberstone A, Charman SA, Price RN, Moehrle J, Gamo-Benito J, et al. Assessing the utility of an anti-malarial pharmacokinetic-pharmacodynamic model for aiding drug clinical development. Malaria journal. 2012;11(1):303.
194. Moxon CA, Chisala NV, Wassmer SC, Taylor TE, Seydel KB, Molyneux ME, et al. Persistent endothelial activation and inflammation after plasmodium falciparum infection in malawian children. The Journal of Infectious Diseases. 2014;209(4):610-5.
195. Woodrow CJ, Taylor WRJ. Reappraisal of the efficacy of a simplified artesunate regimen in falciparum malaria. The Journal of Infectious Diseases. 2012;206(4):619-21.
196. Eksi S, Czesny B, van Gemert GJ, Sauerwein RW, Eling W, Williamson KC. Malaria transmission-blocking antigen, *pfs230*, mediates human red blood cell binding to exflagellating male parasites and oocyst production. Mol Microbiol. 2006;61(4):991-8. Epub 2006/08/02.
197. Bousema T, Drakeley C. Epidemiology and infectivity of plasmodium falciparum and plasmodium vivax gametocytes in relation to malaria control and elimination. Clin Microbiol Rev. 2011;24(2):377-410. Epub 2011/04/13.
198. Smalley ME, Sinden RE. Plasmodium falciparum gametocytes: Their longevity and infectivity. Parasitology. 1977;74(1):1-8. Epub 1977/02/01.
199. Eziefula AC, Bousema T, Yeung S, Kamya M, Owaraganise A, Gabagaya G, et al. Single dose primaquine for clearance of plasmodium falciparum gametocytes in children with uncomplicated malaria in uganda: A randomised, controlled, double-blind, dose-ranging trial. The Lancet Infectious Diseases. 2014;14(2):130-9. Epub 2013/11/19.
200. Guthmann JP, Pinoges L, Checchi F, Cousens S, Balkan S, van Herp M, et al. Methodological issues in the assessment of antimalarial drug treatment: Analysis of 13 studies in eight african countries from 2001 to 2004. Antimicrob Agents Chemother. 2006;50(11):3734-9. Epub 2006/09/07.
201. Verret WJ, Dorsey G, Nosten F, Price RN. The effect of varying analytical methods on estimates of anti-malarial clinical efficacy. Malaria journal. 2009;8(1):77.
202. Dahal P, Simpson JA, Abdulla S, Achan J, Adam I, Agarwal A, et al. Competing risk events in antimalarial drug trials in uncomplicated plasmodium falciparum malaria: A worldwide antimalarial resistance network individual participant data meta-analysis. Malaria journal. 2019;18(1):225.
203. Small D, Seydel K. Malaria modeling to evaluate treatment for severe disease. The Journal of Infectious Diseases. 2018;219(8):1176-7.
204. Montgomery J, Milner DA, Jr., Tse MT, Njobvu A, Kayira K, Dzamalala CP, et al. Genetic analysis of circulating and sequestered populations of plasmodium falciparum in fatal pediatric malaria. The Journal of Infectious Diseases. 2006;194(1):115-22. Epub 2006/06/03.