Optimising statistical methodologies for estimating antimalarial efficacy for uncomplicated *Plasmodium falciparum* malaria

Prabin Dahal

Linacre College

A thesis submitted to the Faculty of Clinical Medicine at the University of Oxford for the degree of Doctor of Philosophy

Hilary Term 2018



Centre for Tropical Medicine and Global Health

Nuffield Department of Clinical Medicine

University of Oxford

Optimising statistical methodologies for estimating antimalarial efficacy for uncomplicated *Plasmodium falciparum* malaria

Prabin Dahal, Linacre College, University of Oxford

Submitted for the degree of Doctor of Philosophy in Clinical Medicine, Hilary Term 2018

Abstract

Plasmodium falciparum malaria remains one of the greatest causes of mortality and morbidity globally. Estimates of treatment efficacy provide critical information for optimising treatment guidelines. The aim of this thesis was to quantify the magnitude of variation in the derived estimates attributable to study design and analytical approaches in uncomplicated *P. falciparum* in patients treated with artemisinin combination therapies (ACTs). This thesis utilised individual patient data collated from 92 clinical trials, including 31,507 patients treated with ACTs extracted from WorldWide Antimalarial Resistance Network (WWARN) repository.

This thesis presents a number of novel findings relevant to clinical researchers and policy makers. The current recommended follow-up period of 28 days with artemether-lumefantrine regimen missed an estimated 58% [95% Confidence Interval (CI): 47% – 74%] of treatment failures in Africa and 32% [95% CI: 15% – 45%] in Asia. For dihydroartemisinin-piperaquine (DP) regimen, a study with 42 day follow-up was estimated to miss 47% [95% CI: 19 – 90%] of the treatment failures in those who were adequately dosed. These results confirmed that studies adhering to the current follow-up recommendations provide an overestimate of the drug efficacy and longer follow-up is warranted for these regimens to adequately capture treatment failures.

The derived estimates were further vulnerable to analytical approaches adopted. Kaplan-Meier (K-M) approach, the currently recommended method, overestimated treatment failure by up to 3.6% compared to a competing risk survival analysis approach. Simulation studies demonstrated that the magnitude of this overestimation increased with declining drug efficacy in high transmission areas, where a substantial fraction of patients develop new infections. Similarly, excluding patients with indeterminant outcomes (i.e. whether parasite recurrence is a treatment failure or new infection cannot be determined) was associated with an upwards bias in the derived estimate of efficacy by up to 1.7% compared to using the sophisticated methods of multiple imputation and inverse probability weighting for dealing with the indeterminant outcomes.

This thesis highlighted that the choice of analytical approaches led to variability around the derived estimates of drug efficacy. The findings of this thesis are discussed with regards to current World Health Organization guidance to change treatment policy.

Statement of contribution

I, Prabin Dahal, confirm that the work presented in this thesis is my own work and take the full responsibility for the content of this thesis. The research presented in this thesis was carried out between Hilary Term 2015 to Hilary Term 2018. I was responsible (with the advice of my supervisors) for the concept, design, analysis, and writing-up of the entire project. Where information has been derived from other sources, I confirm that this has been clearly indicated as appropriate.

The original research performed for this thesis comprises of the characterisation of the hazard and probability density of *P. falciparum* recrudescence post treatment with artemisinin combination therapies, the usage of competing risk survival analysis for handling new infection, and the application of missing data methods for handling indeterminate outcome.

The data used in this thesis (Chapter 3) was standardised by the WWARN informatics and data management team. I was involved with the data standardisation and curation of some of the early studies included in this thesis during my role as a statistician with WWARN from 2011 to 2014. The literature review of all published antimalarial studies was carried out by Prof. Ric N Price (Chapters 3 and 4).

Chapter 2 has been modified from a manuscript which has been published, and manuscripts based on Chapters 5 and 6 are soon to be submitted. I have received editorial and scientific comments on these manuscripts from my co-authors and these have been incorporated in the respective chapters.

Publications

Published

1. Dahal P, Simpson JA, Dorsey G, Guerin 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; 16:430.

   A modified version of this article is presented in **Chapter 2**.

Under peer-review

2. The WWARN Methodology study group. Competing risk events in antimalarial studies of uncomplicated *Plasmodium falciparum* malaria.

   A modified version of this article is presented in **Chapter 5**.

3. Prabin Dahal, Philippe J. Guerin, Ric N. Price, Julie A. Simpson, Kasia Stepniewska. Evaluating antimalarial efficacy in single-armed and comparative drug trials using competing risk survival analysis: A simulation study.

   A modified version of this article is presented in **Chapter 5**.

4. Dahal P, Stepniewska K, Guerin PJ, D'Alessandro U, Price RN, Simpson JA. Dealing with indeterminate outcomes in antimalarial drug efficacy trials: A comparison between complete case analysis, multiple imputation and inverse probability weighting.

   A modified version of this article is presented in **Chapter 6**.

In preparation

5. The WWARN Methodology study group. The distribution of recrudescence post antimalarial treatment regimens: implication for follow-up duration.

   A modified version of this article is presented in **Chapter 4**.

Publications arising during the course of DPhil but not included

The following three papers are my own original work carried out under the supervision of Prof. Ric N Price (AL dosing study), Prof. Philippe Guerin (ASAQ dosing study), and Dr. Kasia Stepniewska (ACT Africa baseline study group). The bulk of the work on these articles was carried out during my role as a statistician with WWARN (until Dec 2014). All the analyses were completed, written-up, submitted, and revised during the first year of my DPhil studies.

1. WWARN AL Dose Impact Study Group. The effect of dose on the antimalarial efficacy of artemether–lumefantrine: a systematic review and pooled analysis of individual patient data. *Lancet Infect Dis*. 2015;15: 692–702.

2. WWARN ASAQ Dose Impact Study Group. The effect of dosing strategies on the therapeutic efficacy of artesunate-amodiaquine for uncomplicated malaria: a meta-analysis of individual patient data. *BMC Med*. 2015;13: 66.

3. WWARN ACT Africa Baseline Study Group. Clinical determinants of early parasitological response to ACTs in African patients with uncomplicated falciparum malaria: a literature review and meta-analysis of individual patient data. *BMC Med 2015, 13:212*.

I have been co-authors on the following papers, all of which were published during the course of my DPhil study.

4. Hoglund RM, Workman L, Edstein MD, Thanh NX, Quang NN, Zongo I, et al. Population Pharmacokinetic Properties of Piperaquine in Falciparum Malaria: An Individual Participant Data Meta-Analysis. *PLoS Med*. 2017;14:1–23.

5. WWARN Gametocyte Study Group: Gametocyte carriage in uncomplicated *Plasmodium falciparum* malaria following treatment with artemisinin combination therapy: a systematic review and meta-analysis of individual patient data. *BMC Med* 2016, 14:79.

6. WorldWide Antimalarial Resistance Network (WWARN) Lumefantrine PK/PD Study Group. Artemether-lumefantrine treatment of uncomplicated *Plasmodium falciparum* malaria: a systematic review and meta-analysis of day 7 lumefantrine concentrations and therapeutic response using individual patient data. *BMC Med*. 2015;13:227.

7. WWARN Parasite Clearance Study Group. Baseline data of parasite clearance in patients with falciparum malaria treated with an artemisinin derivative: an individual patient data meta-analysis. *Malar. J*. 2015;14:359.

I have presented the research included in this thesis at various seminars and conferences which have further contributed to the evolution of these work.

Oral presentation at conferences

1. Dealing with indeterminate outcomes in antimalarial drug efficacy trials: A comparison between complete case analysis, multiple imputation and inverse probability weighting. April 15-20, 2018, 7th Multilateral Initiative in Malaria (MIM) Pan Africa Malaria conference, Dakar, Senegal. ***Accepted***

2. The consequences of censoring new infections when deriving antimalarial efficacy against uncomplicated *P. falciparum* malaria. 2017, ASTMH conference, Maryland, USA.

3. The consequences of censoring new infections when deriving antimalarial efficacy against uncomplicated *P. falciparum* malaria. 2017, ECTMIH conference, Antwerp, Belgium.

4. Quantifying the impact of competing risk events on antimalarial efficacy against uncomplicated *P. falciparum* malaria. 2017, Sixth Annual Survival Analysis for Junior Researchers Conference, Leicester University, UK.

5. The use of fractional polynomials to characterise hazard function. 2016, Fifth Annual Survival Analysis for Junior Researchers Conference, Leeds University, UK.
   (**Prize**: Best oral presentation)

6. On the adequacy of a 28 day follow-up duration with artemether-lumefantrine against uncomplicated *P. falciparum* malaria. 2016, Oxford Tropical Network Conference, University of Oxford, UK.

7. Survival methods in antimalarial clinical trials: current practices and challenges. 2015. Fourth Annual Survival Analysis for Junior Researchers Conference, Keele University, UK.

Poster presentation at conferences

1. The consequences of censoring new infections when deriving antimalarial efficacy against uncomplicated *P. falciparum* malaria. 2017, ASTMH conference, Maryland, USA (for the Young Investigator's award section).

2. On the adequacy of a 28 day follow-up duration with artemether-lumefantrine against uncomplicated *P. falciparum* malaria: Preliminary results. 2016, ASTMH conference, Atlanta, USA.

3. On the adequacy of a 28 day follow-up duration with artemether-lumefantrine against uncomplicated *P. falciparum* malaria. 2016, Oxford Tropical Network Conference, University of Oxford, UK.

4. Investigating the optimal follow-up duration for artemether-lumefantrine against uncomplicated *P. falciparum* malaria. 2016. DPhil Day, John Radcliffe Hospital, Oxford.

5. On the adequacy of a 28 day follow-up period for artemether-lumefantrine against uncomplicated *P. falciparum* malaria. 2016, 8th EDCTP Forum, Lusaka, Zambia. (*Attended by Dr. Georgina Humphreys*).

Seminars

1. Competing risk events in antimalarial studies, Tropical Medicine and Global Health Seminar, University of Oxford, Oxford, 2017.

2. On the distribution of recrudescence post antimalarial treatment, Retreat Day, University of Melbourne, Melbourne, Australia, 2017.

3. Why is it that antimalarial treatments do not always work? Linacre College Seminar, University of Oxford, Oxford, UK, 2016.

4. Optimal duration of follow-up for antimalarial regimens. Tropical Medicine and Global Health seminar, University of Oxford, Oxford, UK, 2015.

Acknowledgements

I have many to thank for this thesis. First and foremost, I would like to thank my supervisory quartet; Prof. Philippe Guerin - for making everything happen; Dr. Kasia Stepniewska - for always being there when needed; Prof. Ric Price – for constantly challenging me; and Prof. Julie Simpson- for instilling discipline, prodding me to keep everything within timelines, and hosting me at Melbourne. All of you have been very generous with your time and I have been privileged to have your constant guidance and support on the research presented in this thesis and beyond.

I would like to thank everybody at WorldWide Antimalarial Resistance Network (WWARN). In particular, Patrice Piola and Christian Nszanabana, my previous line managers who introduced me to different aspects of clinical malaria and antimalarial drug dosing. Thank you Carol, Georgina, Mehul, Rashid, Joel, Anne, Aimee, Aileen, Andrea, Ruth, and Laura for all your direct and indirect help with this thesis. Thank you Sam- my long term colleague at WWARN for all your help and taking care of all the conferences and courses (and for the cakes you bring to the office). Clarissa, Andrew, Georgina, Ignacio, and Hazel- thank you for all the help with data and keeping the repository updated.

Amelie, Shirley, Mandy, and Makoto – thank you for all the lunch and dinner conversations, and sharing all the usual doubts and insecurities a DPhil student goes through.

For the research carried out in Chapter 2, I would like to thank Prof. Grant Dorsey (UCSF), my co-author on the manuscript based on this chapter.

For the research carried out in Chapter 4, I would like to thank Prof. Andreas Martensson (Karolinska) for our several exchanges on this topic. Sophie (Melbourne) - thank you for your help with within-host simulation modelling. Thank you Prof. Ingrid Felger (Swiss TPH) for

sharing your expertise on PCR genotyping. My greatest acknowledgement on this chapter is reserved for Prof. Nick J White (Bangkok). Thank you for your invaluable time, ideas and unwavering enthusiasm on this chapter, to which I have apologetically not been able to keep up. This chapter is in a much better shape for your scientific inputs.

For the research carried out in Chapter 5, I would like to thank Marcel Wolbers (Basel) for the expert guidance on competing risk survival analysis over email. I would like to thank you for thoroughly reviewing the draft manuscript based on this chapter. Hopefully, we will get to meet in person someday.

For the research carried out in Chapter 6, I would like to thank Prof. Ian White (UCL), Margarita Moreno Betancur (Melbourne), and Mavuto Mukaka (Bangkok). Thank you Prof. Umberto D'Alessandro (Antwerp), my co-author on the manuscript based on this chapter, which is ready for submission.

I would like to thank Prof. Oumar Gaye and Prof. Magatte Ndiaye for hosting me at Dakar and taking me to the Pikine clinical study site and for our several conversations over afternoon cups of *attaya*.

I would like to thank Oxford Tropical Network Fund for funding my DPhil (college and tuition fees) and WWARN for paying my stipend. I would like to thank my college for the support throughout, especially with books. Thank you Fiona for purchasing the latest editions of all the statistical books I have ever requested.

Prof. Stephen Senn (Luxembourg) - thank you very much for your helpful replies and our telephone conversation on several statistical issues (from hazard function to missing data). Michael Vaillant (Luxembourg) - thank you for being a friend and going out of your way to solve my statistical queries.

I would like to thank all my collaborators from the 92 clinical studies for generously granting me permission to use the data for the purpose of this thesis.

I am ever so thankful to all my friends and family members who have helped me to make it this far and for always believing in me. Thank you to my parents, who have given everything I have ever asked for, and to my brother and sister-in law, who have been great support throughout. Thank you Rakshya for bearing me and keeping me motivated.

Dedication

To the 31,507 anonymised patients included in this thesis

Contents

Figures ..... **XVIII**

Tables ..... **XXII**

Abbreviations ..... **XXV**

Chapter 1: Introduction **1**

1.1 Aims and objectives of this thesis 1

1.2 History of malaria 2

1.3 Malaria biology and lifecycle 5

1.4 Clinical features of uncomplicated *P. falciparum* malaria 7

1.5 Treatment of uncomplicated *P. falciparum* malaria using artemisinin combination therapies 7

1.6 Antimalarial drug resistance 12

1.7 Monitoring antimalarial efficacy to ACTs *in vivo* 15

1.7.1 Resistance versus treatment failure 15

1.7.2 Surveillance strategy 19

1.8 Outline of chapters 21

Chapter 2: Literature review **23**

2.1 Introduction 23

2.1.1 Aims and objectives 24

2.2 Motivational study 25

2.3 Review of analytical approaches for estimating drug efficacy 26

2.3.1 Proportion cured 26

2.3.2 Survival analysis 27

2.4 The current design and analysis of a therapeutic efficacy study 30

2.5 Challenges in estimating antimalarial efficacy for a single arm 31

2.5.1 Defining optimal duration of follow-up for capturing recrudescence 31

2.5.2 Considering new infection as a competing risk event 35

2.5.3 Dealing with an indeterminate outcome 38

2.5.4 Interval censoring 39

2.5.5 Presenting results from multicentre studies 41

2.6 Challenges for comparative efficacy trials 44

2.6.1 Comparison of antimalarial drugs with different pharmacokinetic profiles 44

2.6.2 Comparing estimates at a fixed point in time in a single centre study 47
2.7 Conclusions 51
Chapter 3: Data acquisition and description 52
3.1 The WWARN repository 52
3.2 Data selection from the WWARN repository 55
3.2.1 Inclusion criteria 55
3.2.2 Desirable criteria 56
3.2.3 Database search 56
3.2.4 Ethical clearance 56
3.3 Study description 58
3.3.1 Design of the studies included 58
3.3.2 PCR genotyping 58
3.3.3 Malaria endemicities in the study sites 59
3.4 Patient description 65
3.5 Parasitological response 67
3.5.1 Early parasitological response 67
3.5.2 Late parasitological response 68
3.6 Concluding remarks 70
Chapter 4: On the optimal follow-up duration for artemisinin combination therapies . 71
4.1 Introduction 71
4.1.1 Aims and objectives 72
4.2 Literature review on time of *P. falciparum* recrudescence 74
4.2.1 Search strategy 74
4.2.2 Results of the literature search 75
4.2.2.1 Artemether-lumefantrine (AL) 75
4.2.2.2 Dihydroartemisinin-piperaquine (DP) 75
4.2.2.3 Artesunate-amodiaquine (ASAQ) 75
4.2.2.4 Artesunate-mefloquine (ASMQ) 75
4.2.3 Conclusions from the literature review 80
4.3 Characterising distribution of *P. falciparum* recrudescence 81
4.3.1 Study and patient data 81
4.3.2 Time of patient visits 81
4.3.3 The time of observed recrudescence 84

4.3.4 Estimation of temporal trend of *P. falciparum* recrudescence 85
4.3.5 Estimation of empirical probability distribution of *P. falciparum* recrudescence 87
4.3.6 Optimal duration of follow-up 88
4.3.7 Software 88
4.3.8 Results 89
4.3.8.1 Characteristics of the studies 89
4.3.8.2 Parasitic recurrences 89
4.3.8.3 The instantaneous hazard function of *P. falciparum* recrudescence 91
4.3.8.4 The empirical distribution of time of *P. falciparum* recrudescence 94
4.3.8.5 Conclusions from the estimated distribution of *P. falciparum* recrudescence 96
4.4 Simulation study 97
4.4.1 Simulation set-up 97
4.4.2 Results 99
4.4.3 Conclusions of the simulation study 101
4.5 Discussion 102
4.5.1 Limitations 105
4.6 Conclusions 107
Chapter 5: Competing risk events in antimalarial studies of uncomplicated *P. falciparum* malaria 108
5.1 Introduction 108
5.1.1 Estimation of failure probabilities 109
5.1.2 Comparing treatment regimens 110
5.1.3 Regression modelling 111
5.1.3.1 Cause-specific hazard 111
5.1.3.2 The sub-distribution hazard 112
5.1.3.3 Regression models in the presence of competing risk events 114
5.1.4 Aims and objectives 116
5.2 Data analysis 117
5.2.1 Study data 117
5.2.2 Statistical analyses 117
5.2.3 Software 119
5.2.4 Results of data analysis 120
5.2.4.1 Study data 120

5.2.4.2 Risk of *P. falciparum* recrudescence 121
5.2.4.3 Risk of *P. falciparum* new infection 122
5.2.4.4 Comparative efficacy studies 126
5.2.4.5 Regression models in the presence of competing risk events 132
5.2.5 Conclusions from data analysis 135
5.3 Simulation studies 136
5.3.1 Methodology 137
5.3.1.1 Simulation set-up for study I 137
5.3.1.2 Simulation set-up for study II 141
5.3.1.3 Sample size considerations 143
5.3.2 Results 147
5.3.2.1 Results of simulation study I 147
5.3.2.2 Results of simulation study II 151
5.3.3 Conclusions from the two simulation studies 158
5.4 Discussion 160
5.4.1 The choice of the method 164
5.4.2 Limitations 164
5.5 Conclusions 166
Chapter 6: Dealing with indeterminate outcomes in antimalarial efficacy trials: a simulation study 167
6.1 Introduction 167
6.1.1 Introduction to missing data 168
6.1.1.1 Multiple Imputation (MI) 169
6.1.1.2 Inverse probability weighting (IPW) 169
6.1.2 Review on the usage of MI and IPW in estimating antimalarial efficacy 170
6.1.3 Aims and objectives 171
6.2 Motivating study: The 4ABC Trial 172
6.3 Simulation study: design and analysis 174
6.3.1 Study design 174
6.3.1.1 Simulation of missingness under mechanism 1 176
6.3.1.2 Simulation of missingness under mechanism 2 176
6.3.2 Analysis of simulated data 179
6.3.2.1 Complete case (CC) analysis 179

6.3.2.2 Multiple Imputation (MI) 179
6.3.2.3 Inverse Probability Weighting (IPW) 181
6.3.3 Performance measures for evaluating different methods 183
6.3.4 Software 185
6.4 Simulation study: Results 186
6.4.1 Mean bias 186
6.4.2 Model-based and empirical standard errors 189
6.4.3 Coverage probability of the true value 192
6.4.4 Root mean squared error (RMSE) 192
6.5 Conclusions from the simulation study 198
6.6 Evaluation of applicability of MI and IPW approaches in small sample antimalarial trials: a resampling study 199
6.6.1 Generating resample data 199
6.6.2 Analysis of created datasets 200
6.6.3 Results 200
6.6.4 Conclusions from resampling study 202
6.7 Discussion 203
6.7.1 Limitations 206
6.8 Conclusions 208
Chapter 7: General discussion and conclusions 209
7.1 Chapter summary 209
7.2 Project strengths 214
7.3 Project limitations 216
7.4 Implications for public health and antimalarial research 218
7.5 Generalisability and future work 221
7.6 Conclusions 222
References 223
Appendices 248
Appendix to Chapter 2 248
Appendix to Chapter 4 250
Appendix to Chapter 5 258
Appendix to Chapter 6 265

Figures

Figure 1.1: Global map of malaria cases in 2016 4

Figure 1.2: Lifecycle of *Plasmodium falciparum* in the human body and the anopheline mosquito. From White et al. Malaria. *Lancet* 2014, 383:723–735. [10]. Reproduced with permission from Elsevier. 6

Figure 1.3: The clinical manifestation of antimalarial drug resistance. From Price RN, Nosten F: Drug resistant falciparum malaria: clinical consequences and strategies for prevention [39]. *Drug Resist Updat* 2001, 4:187–96. , and White NJ: Antimalarial drug resistance. *J Clin Invest* 2004, 113:1084–1092 [13]. Reproduced with permission. 14

Figure 1.4: Therapeutic responses post antimalarial treatment. Adapted from White NJ: The assessment of antimalarial drug efficacy. *Trends Parasitol* 2002, 18:458–464.[58].....17

Figure 1.5: Interpretation of early and late parasitological response from treatment efficacy studies 20

Figure 2.1: The evolution of methodological guidelines for antimalarial studies. From From Dahal et al. Statistical methods to derive efficacy estimates of anti-malarials for uncomplicated Plasmodium *falciparum* malaria: pitfalls and challenges. *Malar. J.* 2017;16:430. [74]. Reproduced under creative commons attribution (CC-BY) license. .29

Figure 2.2: Relationship between detection of treatment failures against half-life of the partner drugs. From Stepniewska K, Taylor WRJ, Mayxay M, Price R, Smithuis F, Guthmann J, et al. In Vivo Assessment of Drug Efficacy against Plasmodium falciparum Malaria : Duration of Follow-Up. *Antimicrob. Agents Chemother*. 2004;48:4271–80. [77]. Reproduced with permission from ASM Journals. 33

Figure 2.3: Recrudescence failure using in Tororo [65]. From From Dahal et al. Statistical methods to derive efficacy estimates of anti-malarials for uncomplicated Plasmodium

falciparum malaria: pitfalls and challenges. *Malar. J.* 2017;16:430. [74]. Reproduced under creative commons attribution (CC-BY) license. 37
Figure 2.4: Cumulative failure for drugs with different terminal elimination half-lives. From From Dahal et al. Statistical methods to derive efficacy estimates of anti-malarials for uncomplicated Plasmodium falciparum malaria: pitfalls and challenges. *Malar. J.* 2017;16:430. [74]. Reproduced under creative commons attribution (CC-BY) license. .46
Figure 2.5: Comparison of estimates at fixed point in time. From Dahal et al. Statistical methods to derive efficacy estimates of anti-malarials for uncomplicated Plasmodium falciparum malaria: pitfalls and challenges. *Malar. J.* 2017;16:430. [74]. Reproduced under creative commons attribution (CC-BY) license. 48
Figure 3.1: The WWARN Process Map 54
Figure 3.2: Study flow diagram 57
Figure 3.3: Study site locations for different drugs 60
Figure 3.4: The distribution of day-3 parasitological response by treatment regimen 68
Figure 3.5: The Kaplan-Meier (K-M) estimate of PCR corrected drug efficacy by follow-up time and drug regimens 69
Figure 4.1: The symptomatic status in patients with recurrent parasitaemia (n=4,699).....82
Figure 4.2: Parasite density at the time of recurrence in symptomatic and asymptomatic patients (n=4,699).....83
Figure 4.3: Distribution of observed time to recrudescence for ACTs stratified by region and duration of follow-up.....90
Figure 4.4: Fractional polynomial estimates of cumulative and instantaneous hazard for AL in Africa and Asia, and for DP regimen in Africa.....93
Figure 4.5: The probability distribution of *P. falciparum* recrudescence for AL in Africa and Asia, and for DP in Africa.....95

Figure 4.6: Absolute overestimation in efficacy due to sub-optimal follow-up duration (compared to day 63 estimates) 100
Figure 4.7: Simulation study comparing AL against DP in children <5 years in Africa 101
Figure 5.1: Calculation of cause-specific and sub-distribution hazard function. Reproduced with permission from Lau et al (2009) [263]. 113
Figure 5.2: The observed proportion of recurrence events in the studies included 120
Figure 5.3: The overestimation of derived failure using 1 minus Kaplan-Meier method compared to the cumulative incidence function method 124
Figure 5.4: Predicted risk of recrudescence from cause-specific Cox model and sub-distribution hazard model 134
Figure 5.5: The instantaneous hazard, cumulative hazard and survival function used for simulating time to recrudescence and new infection in simulation study I 140
Figure 5.6: The baseline hazard function for recrudescence and new infection used for simulation study II 146
Figure 5.7: Overestimation of recrudescence failure using K-M method compared to the CIF in a simulation study I assessing standalone efficacy on day 63 (n=500 subjects) 148
Figure 5.8: Cumulative failure estimates by study follow-up time using extreme examples from simulation study I (n=500 subjects) 149
Figure 5.9: Distribution of simulated hazard ratio (n=500 subjects) 151
Figure 5.10: The ratio of recrudescence and new infections for drug B relative to drug A observed in simulation study II (n=500 subjects per arm) 155
Figure 5.11: Situational competitiveness of newly emergent infections. Adapted from White-2002[58] 159
Figure 6.1: The design of the simulation study 175

Figure 6.2: Average estimate of Kaplan-Meier survival probability on day 28 across 1,000 simulated datasets for different missing data methods, missing data mechanisms and percentages of indeterminate recurrences.....187
Figure 6.3: Relative bias in the K-M estimates on day 28 across 1,000 simulated datasets for different missing data methods, missing data mechanisms and percentages of indeterminate recurrences.....188
Figure 6.4: Empirical standard errors of K-M estimates on day 28 across 1,000 simulated datasets for different missing data methods, missing data mechanisms and percentages of indeterminate recurrences (on complementary log-log scale).....191
Figure 6.5: Coverage probability of K-M estimates on day 28 across 1,000 simulated datasets for different missing data methods, missing data mechanisms and percentages of indeterminate recurrences.....193
Figure 6.6: Root mean squared error of the derived K-M estimates on day 28 across 1,000 simulated datasets for different missing data methods, missing data mechanisms and percentages of indeterminate recurrences.....194

Tables

Table 1.1: Terminal elimination half-life of antimalarial drugs 11

Table 1.2: Determinants of *in vivo* response to antimalarial treatment 18

Table 2.1: Outcome assignment for derivation of antimalarial efficacy 28

Table 2.2: Failure estimates using missing data methods in Machekano et al (2007)[88].....39

Table 2.3: Estimates with and without interval censoring in Uganda data [65].....40

Table 2.4: Kaplan-Meier estimates for artesunate-amodiaquine arm in the motivational study [65].....43

Table 2.5: Comparison of K-M estimates at fixed point in motivational study [65].....49

Table 2.6: Challenges in estimating drug efficacy and some alternatives.....50

Table 3.1: Characteristics of the studies included 61

Table 3.2: Patients treated with antimalarial regimens by transmission settings 64

Table 3.3: Baseline characteristics of the patients included 66

Table 3.4: Early parasitological response in the studies included 67

Table 4.1: Reported timing of recrudescences in artemether-lumefantrine studies with follow-up duration ≥ 42 days 76

Table 4.2: Reported timing of recrudescences in dihydroartemisinin-piperaquine studies with follow-up duration ≥ 63 days.....78

Table 4.3: Reported timing of recrudescences in artesunate-amodiaquine studies with follow-up duration ≥ 42 days 79

Table 4.4: Reported timing of recrudescences in artesunate-mefloquine studies with follow-up duration ≥ 63 days 79

Table 4.5: The area under the curve of the estimated probability distribution.....95

Table 5.1: Absolute overestimation in cumulative recrudescence using the Kaplan-Meier method compared to the Cumulative Incidence Function.....123

Table 5.2: Study sites where cumulative failure estimates were >10% and >5% using the Kaplan-Meier method 125
Table 5.3: Comparing AL against DP using log-rank test and Gray's *k*-sample test 127
Table 5.4: Comparing AL against ASMQ using log-rank test and Gray's *k*-sample test 128
Table 5.5: Comparing AL against ASAQ using log-rank test and Gray's *k*-sample test 129
Table 5.6: Comparing ASAQ against DP using log-rank test and Gray's *k*-sample test 130
Table 5.7: Comparing ASMQ against DP using log-rank test and Gray's *k*-sample test 130
Table 5.8: Comparing ASMQ against ASAQ using log-rank test and Gray's *k*-sample test 131
Table 5.9: Regression models for recrudescence and new infection using data from Burkina Faso (n=810); subset of The 4ABC Trial [139] 133
Table 5.10: Different scenarios for comparing two drug regimens (drug B compared against drug A) in simulation study II 145
Table 5.11: Absolute overestimation in recrudescence failure by K-M method compared to CIF in simulation study I (n=500 subjects) 150
Table 5.12: Probability of rejecting the null hypothesis at two sided 0.05 level (n=500 subjects per arm) 156
Table 5.13: Probability of rejecting the null hypothesis at two sided 0.05 level for different sample sizes (simulation study II) 157
Table 6.1: Antimalarial treatment outcomes for the 4ABC Trial[139] 172
Table 6.2: Baseline predictors of indeterminate outcomes (n=3,431) 173
Table 6.3: Specification of the logistic regression model used to impose missingness 179
Table 6.4: Outline of the imputation model and missingness model 182
Table 6.5: Kaplan-Meier estimates of antimalarial efficacy on day 28 from the full data compared to that obtained from 20,000 bootstrap samples 183
Table 6.6: Model-based standard errors for IPW estimator 190

Table 6.7: Performance measures of different approaches for handling missing outcome for artemether-lumefantrine 195
Table 6.8: Performance measures of different approaches for handling missing outcome for artesunate-amodiaquine 196
Table 6.9: Performance measures of different approaches for handling missing outcome for dihydroartemisinin-piperaquine.....197
Table 6.10: Number of events in resampled data of different sample sizes drawn from observed data for artemether-lumefantrine.....200
Table 6.11: Performance measure against different sample sizes 201

Abbreviations

ACT	Artemisinin Combination Therapies
AL	Artemether Lumefantrine
ASAQ	Artesunate Amodiaquine
AUC	Area Under the Curve
CC	Complete Cases
CI	Confidence Interval
CIF	Cumulative Incidence Function
CP	Coverage Probability
DP	Dihydroartemisinin Piperaquine
FDC	Fixed Dose Combination
glurp	Glutamate Rich Protein
IPD	Individual Participant Data
IPW	Inverse Probability Weighting
IQR	Interquartile Range
K-M	Kaplan Meier
LFU	Lost to Follow Up
MAR	Missing At Random
MCAR	Missing Completely At Random

MI	Multiple Imputation
MIC	Minimum Inhibitory concentration
mice	Multiple Imputation by Chained Equations
MNAR	Missing Not At Random
msp-1	Merozoite Surface Protein-1
msp-2	Merozoite Surface Protein-2
NI	New Infection
NMCP	National Malaria Control Programme
PCR	Polymerase Chain Reaction
Pf	*Plasmodium falciparum*
PIP	Piperaquine
PMF	Parasite Multiplication Factor
PPR	Parasite Positivity Rate
PRR	Parasite Reduction Rate
Pv	*Plasmodium vivax*
RC	Recrudescence
SSA	sub-Saharan Africa
WHO	World Health Organization
WWARN	WorldWide Antimalarial Resistance Network

Chapter 1: Introduction

1.1 Aims and objectives of this thesis

Antimalarial clinical trials provide critical information with which drug developers, policy makers and healthcare providers can optimise treatment guidelines. They have been conducted for over 100 years and the guidelines for the conduct and analyses of antimalarial trials have evolved throughout this period [1]. The overarching aim of this thesis was to identify key study design and analytical challenges that influence the derivation of efficacy estimates against the blood stage parasites of uncomplicated *Plasmodium falciparum* in patients treated with artemisinin combination therapies (ACTs). Alternative statistical approaches to improve the assessment of standalone and comparative efficacy studies were explored.

This thesis had four specific objectives:

I. To review historical and current statistical practices used in estimating antimalarial drug efficacy (Chapter 2),

II. To gauge the optimal duration of follow-up for capturing parasitic recrudescence for currently used ACTs (Chapter 4),

III. To explore the competing risk survival analysis as an alternative approach for handling new infections in the derivation of efficacy estimates (Chapter 5), and

IV. To evaluate sophisticated statistical methods for dealing with indeterminate treatment outcomes when estimating antimalarial efficacy (Chapter 6).

1.2 History of malaria

Malaria is an ancient parasitic disease, that remains one of the greatest causes of mortality and morbidity in the modern world [2]. Descriptions of malaria-like illness (seasonal and intermittent fevers) were found in early writings in Egypt, Greece, India and China [3–5]. Although such fevers were often believed to be punishment of god or vengeance by evil spirits, the Greek physician Hippocrates in the 5th century BC discarded these superstitions and was the first to document clinical features of malaria. He observed the relationship between the disease, seasons and dwelling places of the patients. For over 2,500 years, malarial fever was believed to be caused by miasmas rising from the swamps. The name malaria came from Italy in 1740 from the word *mal'aria* (*mala* and *aria)* meaning "bad air" [6]. The true cause of malaria was however, discovered only in 1880 by Charles Louis Alphonse Laveran who demonstrated the causative agent of malaria was a parasite after observing its pigmented forms in blood smears from patients who had died of malaria [3,7]. In 1897, Ronald Ross discovered the actual mode of transmission of the parasites. He found a developing form of malaria (oocysts) on the stomach wall of an *Anopheles* mosquito that had previously fed on a patient with plasmodia in his blood [4]. The overall development life cycle of the malaria parasites became clearer as a result of further works undertaken by the Italian scientists Amico Bignami, Giuseppe Bastianelli, and in particular by Battisata Grassi in 1898-9 [4,5].

As of today, five species of plasmodium which cause malarial infections in human have been identified: *Plasmodium falciparum*, *Plasmodium vivax*, *Plasmodium ovale*, *Plasmodium malariae* and *Plasmodium knowlesi*. Among these, *Plasmodium falciparum* is the most pathogenic and responsible for the majority of the morbidity and mortality associated with malaria. There was an estimated 216 million cases [95% Confidence Interval (CI): 196 - 263 million] of malaria worldwide in 2016 (Figure 1.1) [2], the majority of which were

uncomplicated, which refers to an illness in which the signs and symptoms of disease do not suggest underlying vital end organ dysfunction [8]. However, if left untreated, the parasites multiply in the peripheral blood, sequestering to the small capillary circulation and causing end organ damage (severe malaria) that can progress rapidly and can lead to death within a few hours. In 2016 alone, there were an estimated 445,000 deaths attributable to malaria, the majority (91%) of which were in young children in sub-Saharan Africa [2]. The scaling up of prevention and control measures and the availability of highly efficacious treatment regimen has led to more than a 50% reduction in malaria mortality globally in the past 10-15 years, thus enthusing optimism regarding the possibility to eliminate malaria. Despite this remarkable progress, there have been worrying signs of late with 4 million more cases reported in 2016 compared to 2015, with 4% increased global mortality [2,9]. Therefore, malaria still remains an overwhelming problem in much of sub-Saharan Africa and continues to exert a significant burden to humankind, most of which is due to the infections with the deadliest *P. falciparum*. The rest of this thesis is concerned with deriving the estimate of antimalarial drug efficacy in clinical studies with that organism, whose complex lifecycle involves two hosts in vertebrates (this thesis is concerned with the human host) and mosquitoes.

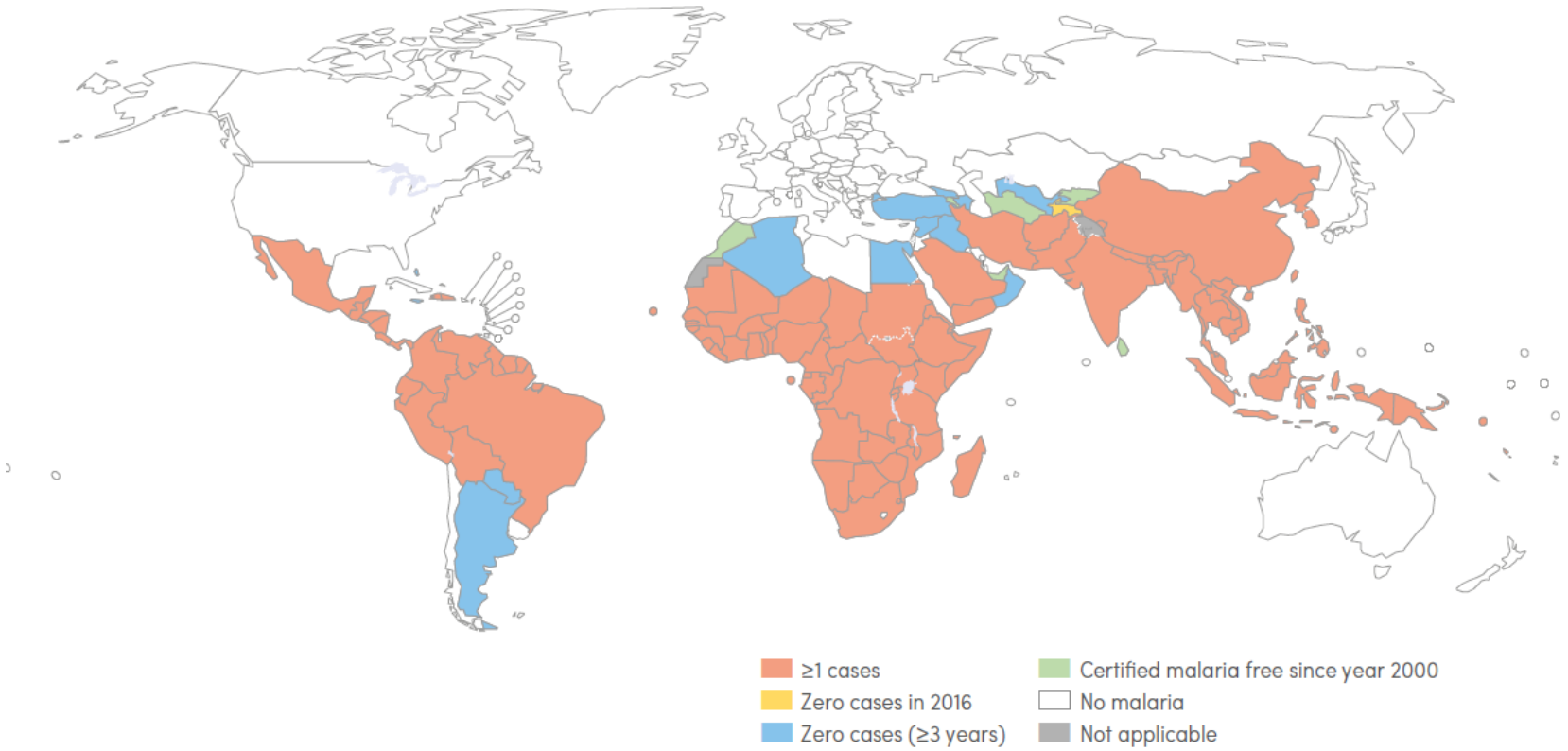


Figure 1.1: Global map of malaria cases in 2016

(Source: World Malaria Report 2017 [2])

1.3 Malaria biology and lifecycle

The malaria life cycle involves two hosts (Figure 1.2). The sexual development of the parasites takes place in the intestine of an anopheline mosquito, whereas the asexual development occurs in a vertebrate host. Infection begins when a female anopheline mosquito inoculates sporozoites into the human host during a blood meal to support the development of eggs. The microscopic sporozoites, which are highly motile, travel around the human body until they reach the liver where they invade the hepatocytes to begin a phase of asexual multiplication. Each sporozoite can produce up to 30,000 daughter merozoites within a period of 5-8 days [10]. The hepatic schizont eventually ruptures releasing $<10^4$ merozoites into the peripheral blood. The released merozoites start the invasion of red blood cells where they undergo a second phase of asexual multiplication.

The intraerythrocytic stages start as a small ring-shaped structure (ring stage) which grows into a larger ring and then into a large solid parasite (trophozoite). The nucleus of the trophozoite undergoes mitosis, each forming 16 small parasites called merozoites. The merozoites break out of the engorged red cells and are released into the blood stream. They further invade a new red blood cell and continue the asexual cycle, resulting in an exponential growth in the parasite biomass. The asexual life cycle lasts up to 24 hours for *P. knowlesi*, 48 hours for *P. falciparum*, *P. vivax* and *P. ovale*, and up to 72 hours for *P. malariae*. As the infection progresses, some of the young merozoites differentiate into male and female gametocytes, the sexual stage of the development. However, their initial presence in the peripheral blood doesn't necessarily result in further transmission even at high densities, as the gametocytes need a further 3 to 4 days to reach functional maturity and become infectious to mosquitoes [11,12]. In general, there is a sigmoid relationship between gametocyte densities and their intrinsic infectivity for mosquitoes [13]. On average, the mosquito ingests a 2 to 3 μL blood meal, and this needs to contain at least one male and one female gametocyte to result in an infection [12]. When the

density of the gametocytes reaches ~5/µL in the peripheral blood, they can infect approximately 3% of female anopheline mosquitos during a blood meal, and the sexual phase of the life cycle progresses [12].

In the mid-gut of a mosquito, a fall in temperature (> $5^{o\ c}$) triggers the maturation of the gametocytes into male and female gametes resulting in fertilisation. The zygote thus formed is motile and penetrates the gut wall to become an oocyst. The oocyst undergoes another phase of multiplication resulting in the formation of sporozoites, which migrate to the salivary gland of the mosquito. The sporozoites are injected into a new vertebrate host when the mosquito takes a blood meal, and the life cycle continues.

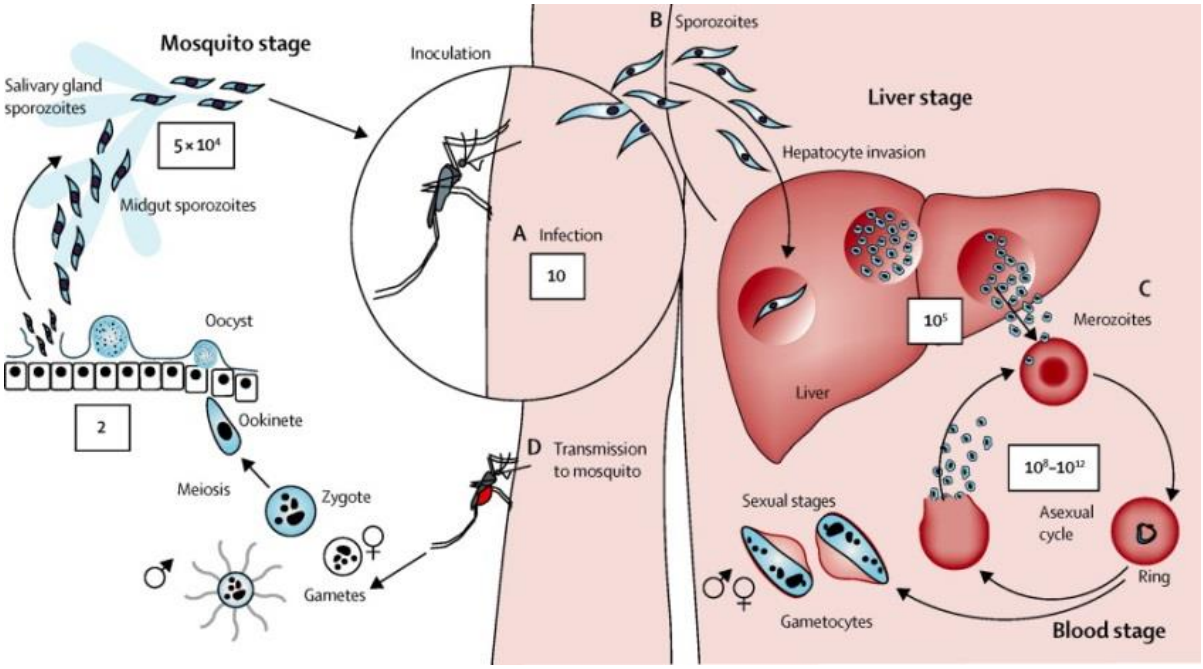


Figure 1.2: Lifecycle of *Plasmodium falciparum* in the human body and the anopheline mosquito. From White et al. Malaria. *Lancet* 2014, 383:723–735. [10]. Reproduced with permission from Elsevier.

1.4 Clinical features of uncomplicated *P. falciparum* malaria

The incubation period, which refers to the time between acquiring the infection and the first patency of clinical sign, is about 7 days for *P. falciparum* malaria [14]. The best known clinical feature of uncomplicated malaria is febrile paroxysms and in a synchronised infection these may occur every 48 or 72 hours, associated with rupture of the infected red blood cells, release of cytokines, and plasmodial waste products during schizogony [7]. For 2 or 3 days before the first paroxysm, patients may experience general feelings of fatigue, muscle ache, headache, dizziness, nausea, vomiting, and abdominal discomfort [10]. Most patients with uncomplicated malaria will develop mild anaemia, splenomegaly, and hepatomegaly. In areas of intense transmission, spontaneous resolution of symptoms is common among adults who have developed protective immunity through repeated exposures, and can tolerate high parasitaemia; these individuals are often asymptomatic. In non-immune children in high transmission areas and in patients of all ages in regions of low transmission, the progression of the disease can be rapid leading to end organ damage and eventually death, if untreated.

1.5 Treatment of uncomplicated *P. falciparum* malaria using artemisinin combination therapies

The current first-line treatments against uncomplicated *P. falciparum* malaria are Artemisinin Combination Therapies (ACTs). These treatments combine partner drugs with long terminal half-lives with short acting artemisinins. Artemisinin was first isolated by the *Artemisinin coordination research group* in China in 1972 from *Artemisia annua L* plant (also known as annual or sweet wormwood), which had been used as a medicinal herb in China for over 2,000 years [15]. Artemisinin and its derivatives have broad stage specificity and rapidly kill the ring stage parasites, allowing only a small fraction of the parasite pool to mature into trophozoites

[16]. The net result is that the drug kills the circulating parasites faster than any other class of antimalarial compounds currently available on the market, reducing the parasite biomass by up to 10^4 per asexual cycle [16]. The artemisinins also have additional gametocytocidal effects thus reducing the transmission and subsequent incidence of malaria [17]. Artemisinin monotherapies are safe and effective; however, due to the short elimination half-life of artemisinins (approximately 1 hour) (Table 1.1), longer treatment courses of 7 days are needed to cure the infection. For pragmatic reasons, a 3-day course of artemisinins are combined with a partner drug that is slowly eliminated, thus shortening the duration of treatment (Table 1.1). The net benefit of such a strategy is that only a small residual pool of parasites (maximum of 10^5) is left to be cleared up by the long-acting, less potent, partner drug [18–21]. The residual parasites, which survive the initial burst of artemisinin, are still exposed to maximum concentrations of the partner drug. Furthermore, for a parasite to escape both drugs, it would require two simultaneous and independent mutational events because of their different modes of action. The overall probability of two independent mutations occurring in a single parasite is the product of the probabilities for each individual mutation and therefore is extremely rare [20]. To demonstrate this, consider an individual with an infection that reaches 10^{12} parasites. If the probability of developing resistance to drug A and drug B are both 1 in 10^{12}, then the overall probability of the parasite simultaneously and independently developing resistance to both drugs is 1 in 10^{24}. There are less than 10^{20} parasites in the entire world in any year, thereby, such resistance will arise spontaneously once in every 10,000 years (if none of the drugs are used in monotherapy) [13]. Hence, the addition of the partner drugs provides protection for the artemisinin (and derivatives) from developing *de novo* resistance [22].

Artemisinins were first deployed on a large scale in a combination therapy as artesunate-mefloquine (ASMQ) in a study carried out on the Thai-Myanmar border in 1992 when mefloquine resistance was rampant [18]. In this study, a 3-day regimen of ASMQ treatment

administered in 652 patients demonstrated a remarkable efficacy of 98% with little side effects. Subsequently, artemisinins (artemether, artesunate, dihydroartemisinin) have been used widely in Asia and Africa in combination with a range of different compounds (lumefantrine, amodiaquine, piperaquine) [23–25]. The World Health Organization (WHO) first endorsed this combination therapy for potentially being used in Africa in 2001 [26]. A large meta-analysis in 2004 highlighted that the addition of 3-days artesunate to existing drugs greatly reduced the treatment failures and gametocyte carriage along with rapidly clearing the peripheral parasites [27]. There was a subsequent push from the research community to use ACTs as a first line treatment in endemic areas [28]. In 2004, the WHO switched the policy and recommended ACTs as treatment of choice in areas where resistance to monotherapies (amodiaquine, chloroquine, and sulphadoxine-pyrimethamine) was widespread [29]. Finally in 2006, ACTs were recommended as the front line treatment against uncomplicated malaria in all endemic countries [30]. There are five artemisinin combination therapies (ACTs) currently recommended for the treatment of uncomplicated malaria by the WHO [8]:

- artemether-lumefantrine (AL)
- artesunate-amodiaquine (ASAQ)
- artesunate-mefloquine (ASMQ)
- dihydroartemisinin-piperaquine (DP)
- artesunate-sulfadoxine-pyrimethamine (ASSP)

A sixth drug combination artesunate-pyronaridine (ASPY) was added to the WHO list of artemisinin combination therapies in 2017 [31]. In the past decade, particular efforts have been made to develop ACTs in a fixed dose combination. This has both drug components co-formulated in the same tablet; thus improving treatment adherence and avoiding exposing artemisinins to monotherapy. All of these currently recommended ACTs are available as a

fixed dose combination, except ASSP. ACTs are widely used today with an estimated 409 million treatment courses administered in 2016 alone [2]. They have demonstrated efficacy consistently higher than 95% [32–34], and the large scale deployment of these highly efficacious ACT regimens together with rolling out of bed-nets has been a key factor contributing to the global reduction of mortality and morbidity associated with malaria [35]. Maintaining these gains is highly dependent on efficient health systems, sustainable global funding and the current status of antimalarial drug resistance. The 2016 WHO report found that globally the number of malaria cases has stopped dropping and mortality has crept up compared to 2015, suggesting that the recent public health gains remain fragile [2,9]. To make things worse, a health calamity is looming large due to the emergence of resistance to artemisinins in South-East Asia, which is threatening to reverse the remarkable progress achieved over the past decade [36].

Table 1.1: Terminal elimination half-life of antimalarial drugs

Drugs	**Elimination half-life (hours)**	**Source: WHO 2015 [8]**
Artemisinin (derivatives)		
Artemether [Dihydroartemisinin]†	0.86 - 5.16 h [1 - 2.3 h]	Table A5.3 [8]
Artesunate [Dihydroartemisinin]†	0.90 h [0.8 - 2.5 h]	Table A5.4 [8]
Dihydroartemisinin	0.85 - 1.40 h	Table A5.8 [8]
Partner drug administered in ACT		
Lumefantrine [Desbutyl-lumefantrine]†	32.7 - 275 h [137 - 141 h]	Table A5.3 [8]
Piperaquine	324 - 672 h	Table A5.8 [8]
Amodiaquine[DEAQ]†	3.3 - 12.4 h [90 - 240 h]	Table A5.1 [8]
Mefloquine	194.4 - 364.8 h	Table A5.10 [8]
Sulfadoxine	98.4 - 261.6 h	Table A5.13 [8]
Pyrimethamine	60 - 450 h	Table A5.13 [8]

†Dihydroartemsinin is the active metabolite of artemether and artesunate, while DEAQ (desethylamodiaquine) and desbutyl-lumefantrine are the active metabolites of amodiaquine and lumefantrine respectively

1.6 Antimalarial drug resistance

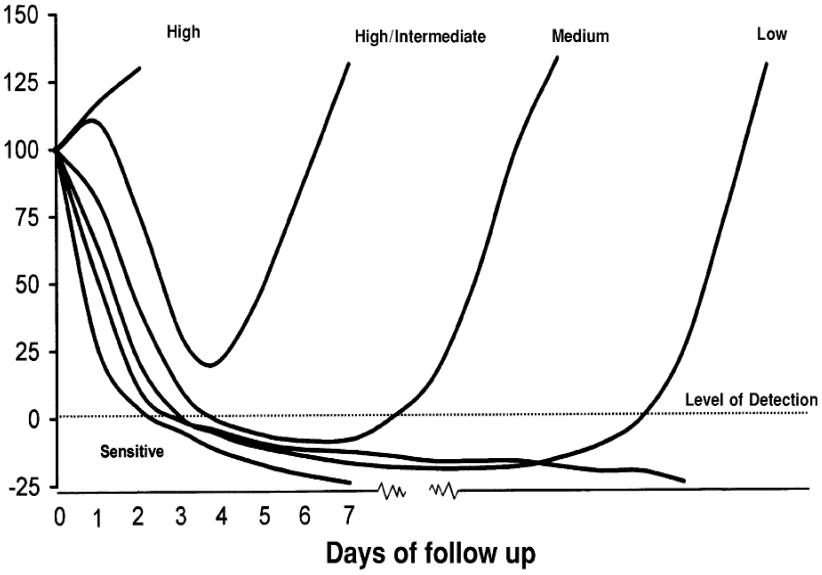
A parasite is defined as being drug resistant if it is able to survive the treatment regimen, multiply and further transmit itself to a new host despite being exposed to adequate drug-levels that should normally kill it [37]. Resistance to antimalarials emerges when a rare mutation or gene duplication occurs, which provides a survival advantage in the presence of drug [38]. This can result from an individual receiving a sub-therapeutic dose or through treatment non-adherence, which increases the selective pressure. This provides an opportunity for resistant mutants to arise *de novo* and confer a survival advantage compared to their sensitive counterparts [39]. In the initial stages, the growth of low grade resistant parasites may be suppressed (or fall below the level of microscopic detection) when the drug levels are above the threshold required to supress parasitic growth. This threshold is known as the minimum inhibitory concentration (MIC). As the drug levels fall below the MIC, the parasite population expands in peripheral blood resulting in *recrudescence*, which manifests clinically as a late treatment failure at low grade resistance. As resistance worsens, the resistant parasites are able to survive and expand even when the drugs concentrations exceed the MIC of a sensitive parasite. This leads to a reduced median time to recrudescence and increased risk of early treatment failure augmented with poor symptomatic resolution (Figure 1.3A). Drug resistance can also be gauged by a right-shift in the dose-response curve. For any given plasma concentration of antimalarial drug, a lesser number of the parasites are killed and the MIC progressively increases eventually resulting in no concentration that can achieve the required maximal effect, rendering the treatment course futile (Figure 1.3B) [13].

Historically, antimalarial drug resistance first emerged in South-East Asia and subsequently spread to Africa through the Indian subcontinent [40]. Chloroquine (CQ) resistance emerged simultaneously and independently in South America and South-East Asia [41]. This arrived at

east Africa in the late 1970s and spread relentlessly across the continent over the next decade. Resistance to sulphadoxine pyrimethamine (SP), the immediate replacement of CQ, rapidly emerged in South-East Asia in less than 5 years of usage [42]. In the same region where CQ and SP resistance had emerged, resistance has now emerged against artemisinin [43]. The concurrent evolution of resistance against partner drugs alongside the reported artemisinin resistance has added insult to the injury. The combination regimen dihydroartemisinin-piperaquine is no longer usable in Cambodia (in Pailin, Pursat and Oddar Meancheay) and in neighbouring Vietnam (in Binh Phuoc province), as the proportion of failures has reached unacceptably high levels [44–46].

The real worry is that artesminin resistance might spread from this region and follow the course of CQ and SP resistance. A key clinical manifestation of artemisinin resistance is slow parasite clearance, and several point mutations in the propeller region of the *Kelch-13* gene have been found to be associated with slow parasite clearance in South-East Asia [36,47]. Multiple foci of origin for those mutations have already been discovered, suggesting that resistance is emerging independently [40]. The pace of resistance has quickened. If the development and registration of new antimalarial drugs does not keep up with the pace at which artemisinin resistance is spreading, the prospect of malaria being untreatable in the future is possible. Preserving the therapeutic efficacy of currently available antimalarial drug regimens is therefore of utmost importance. This demands continuous surveillance, which is currently achieved through Therapeutic Efficacy Studies (TES) carried out by National Malaria Control Programmes (NMCPs). The aim of a TES is to provide information regarding the current status of antimalarial drug resistance, which is assessed based on the derived estimate of drug failure. It is important that resistance be differentiated from treatment failure due to other factors, which is discussed next.

A: Price and Nosten-2001 [39]



B: White-2004 [13]

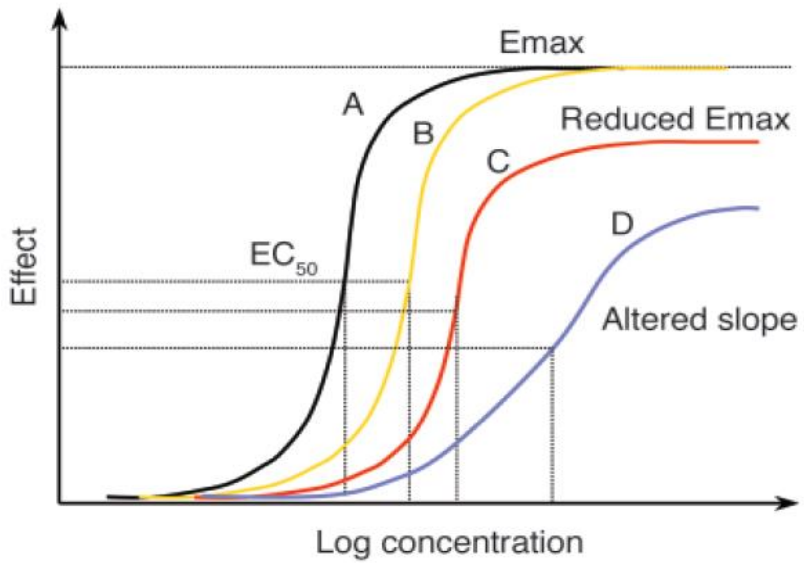


Figure 1.3: The clinical manifestation of antimalarial drug resistance. From Price RN, Nosten F: Drug resistant falciparum malaria: clinical consequences and strategies for prevention [39]. *Drug Resist Updat* 2001, 4:187–96. , and White NJ: Antimalarial drug resistance. *J Clin Invest* 2004, 113:1084–1092 [13]. Reproduced with permission.

Legend: **Panel A**: Initially, at a low grade resistance, late parasitic recrudescences are observed weeks after treatment administration. As resistance worsens, the time to recrudescences gets progressively shorter. At high/intermediate grade resistance, recrudescence appears within a week of treatment. When resistance reaches unacceptably high levels, the treatment has no effect on parasites. **Panel B**: This depicts the pharmacodynamic (drug effect) and drug concentration relationship of the drug regimen. A right shift is observed in the dose-response curve with worsening resistance. The black line represents a sensitive parasite population and yellow, red and blue represents progressively increasing grades of resistance. The y-axes are arbitrary units for both panels.

1.7 Monitoring antimalarial efficacy to ACTs *in vivo*

1.7.1 Resistance versus treatment failure

After the initial reduction of parasite load in the first few days of treatment, a patient can experience recurrent parasitaemia during the follow-up period (Figure 1.4). This can be derived from an inoculation of a new parasite during the follow-up period (**new infection**) or due to reappearance of the same parasite (**recrudescence**). Recrudescence occurs when the peripheral parasitaemia initially falls below the level of detection, but subsequently expands after the concentration of the antimalarial drug falls below the minimum inhibitory concentration (MIC). Recrudescences are true treatment failures, since the parasite has evaded complete cure and the current approach for defining antimalarial efficacy is based on the ability of a drug to prevent the subsequent recrudescence. It is important to distinguish recrudescence from new infection so that drug efficacy attributable to parasite drug resistance can be estimated accurately. Polymerase Chain Reaction (PCR) is used to differentiate these two outcomes by genotyping the pre-treatment and recurrent parasitaemia as they are clinically indistinguishable [48–50]. The current approach for genotyping uses three polymorphic markers: merozoite surface protein (*msp)-1*, *msp-2*, and glutamate rich protein (*glurp)* genes. If at least one allele at each locus is common in pre- and post-recurrence samples, this is defined as a recrudescence, and when the alleles in the recurrent samples are different from those in the baseline samples, it is considered a new infection [51].

However, a recrudescent failure does not necessarily imply drug resistance. Parasite recrudescence can occur for a number of reasons, including sub-optimal dosing, non-adherence to treatment, inadequate drug absorption (due to vomiting), errors in treatment administration, due to the use of sub-standard or falsified medicines, and presenting with a high parasite load (Table 1.2) [1,33,34,38,52–57]. Conversely, a highly drug resistant parasite can sometimes be

cured with a standard antimalarial regimen, if the patient's host immunity plays a significant role in clearance of the parasite. An antimalarial drug might have been administered at the recommended dose, but insufficient for a particularly large patient (resulting in lower weight adjusted doses) with a high parasite load [1]. In such cases, drug concentrations fall below the MIC before the index infection is cleared completely. This increases the likelihood of the parasites being exposed to sub-therapeutic concentrations, which fuels true drug resistance, as recrudescence is essential for the transmission of *de novo* resistance [56].

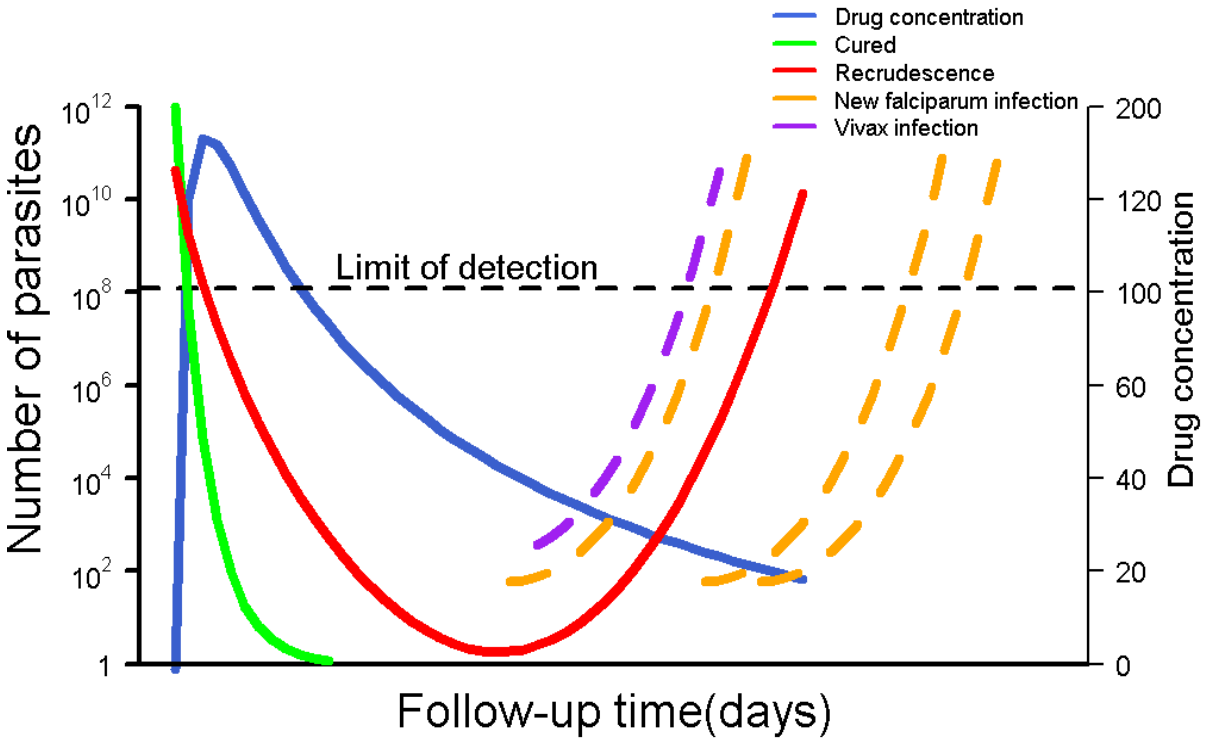


Figure 1.4: Therapeutic responses post antimalarial treatment. Adapted from White NJ: The assessment of antimalarial drug efficacy. *Trends Parasitol* 2002, 18:458–464.[58]

Legend: The blue line represents a hypothetical concentration versus time profile for an antimalarial drug administered orally. The green and red lines represent scenarios for parasite burden versus time profiles following treatment for an infection where all the parasites are completely killed resulting in cure (green) and an infection where parasites are initially killed by high drug levels but with drug levels below the minimum inhibitory concentration (MIC), net parasite growth results in subsequent recrudescence (red). The purple and orange lines represent parasite-time profiles for new infections; either an infection due to a new parasite of the same species (orange) or an infection with a *Plasmodium vivax* parasite (purple) during the follow-up. The left y-axis is for parasite density, and the right y-axis shows drug levels at hypothetical units. The horizontal black line represents the microscopic limit of detection for parasites. The maximum number of parasites a human body can contain is 10^{12}.

Table 1.2: Determinants of *in vivo* response to antimalarial treatment

Host	**Drug**	**Parasite**
Immunity	Pharmacokinetic-pharmacodynamic profile	Parasite biomass
Age	Drug-drug interactions	Sequestration [a]
Pregnancy	Dosing strategy	Mixed infections
Malnutrition status	Duration of treatment	Staging [b]
Co-morbidities	Tolerability	Synchronicity
Treatment seeking behaviour	Drug quality	Polyclonality [c]
Adherence	Drug formulations	Resistance

[a]The process by which infected erythrocytes containing mature parasites adhere to the microvasculature. Their removal from the circulation results in the peripheral parasite count being an underestimate of the true parasite biomass.

[b]The developmental stage of the parasite within the 48 hour intra-erythrocytic life-cycle. The artemisinin compounds have the broadest stage specific action against the parasite (approximately 6 - 44 hrs).

[c]Simultaneous rupture of hepatic schizonts results in a uniform stage distribution of the parasite.

1.7.2 Surveillance strategy

The WHO necessitates that National Malaria Control Programmes (NMCPs) monitor antimalarial efficacy at least once every 24 months at established sentinel sites [59]. The information derived from such therapeutic efficacy studies (TES) forms the basis of policy formulation regarding the current state of antimalarial drug resistance, with *in vitro* and molecular surveillance studies providing complementary information. The current WHO strategy on monitoring antimalarial efficacy (and drug resistance) and optimising treatment policy is based on a series of thresholds. If the estimated failure at the end of a study exceeds 10%, front line treatment should be changed to an alternative ACT regimen (Figure 1.5). This threshold based approach is vulnerable to the choice of analytical approaches adopted to estimate the percentage of failures [60].

The role of host, drug and parasitic factors underpinning drug efficacy have been well studied. However, variations in derived estimates of drug efficacy due to the choice of statistical methods have received little attention [60]. This thesis aimed to explore the magnitude of the variations in derived estimates of drug efficacy due to the choice of analytical approaches.

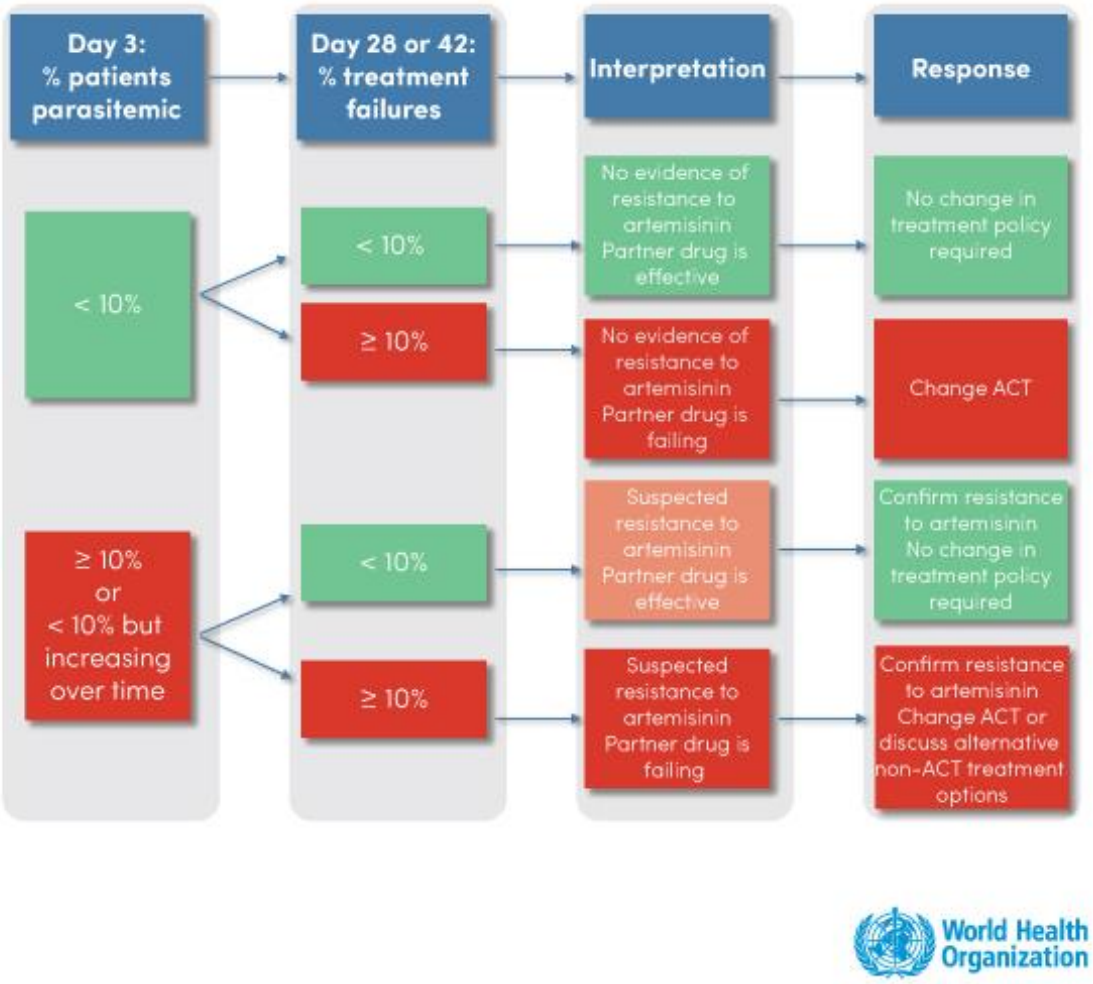


Figure 1.5: Interpretation of early and late parasitological response from treatment efficacy studies

(Source: WHO-2017 [59])

Legend: The percentage of the patients with positive parasitaemia (PPR) on day 3 is a key threshold; when it exceeds 10%, artemisinin resistance is suspected. However, suspected artemisinin resistance may or may not lead to an overall change in treatment policy, which is driven by the derived estimate of drug failure at the end of the study follow-up. When the failure estimate exceeds 10%, the partner drug is deemed to be failing thus prompting a policy change to an alternative treatment regimen.

1.8 Outline of chapters

Chapter 2 provides a review of the historical and current practices in estimating antimalarial efficacy. It provides a critical assessment of the challenges associated with currently adopted statistical methodologies and suggests a number of alternative approaches to overcome such limitations, some of which are explored further in Chapters 4, 5 and 6.

Chapter 3 describes the acquisition, standardisation and ethical clearance of the data used in this thesis. This will describe the characteristics of the studies, and provides the baseline characteristics of the patient population included in the subsequent analyses.

Chapter 4 investigates the optimal duration of follow-up for capturing recrudescences post administration of artemisinin combination therapies. A literature search of the studies with follow-up duration longer than currently recommended is carried out to provide baseline information on the adequacy of the existing recommendation. Data described in Chapter 3 is then used to characterise the hazard and probability distribution of *P. falciparum* recrudescence thus allowing the assessment of the optimal duration of follow-up.

Chapter 5 explores the use of the competing risk survival analysis as an approach for handling new infections. Results derived by considering new infection as a competing risk event is compared against the estimates from the currently recommended Kaplan-Meier analysis, where new infections are considered as censored observations. In addition, results of regression modelling and comparison of drug regimens within the competing risk framework are presented.

In Chapter 6, sophisticated statistical missing data approaches are explored and applied to handling of infections with indeterminate PCR outcomes.

In Chapter 7, the overall research is summarised, and general recommendations are offered to conclude this thesis.

Chapter 2: Literature review

2.1 Introduction

Routine surveillance of *in vivo* antimalarial efficacy in uncomplicated *P. falciparum* involves monitoring of the patient for a fixed period of time after treatment administration and taking clinical and parasitological measurements at regular intervals. The patient is assigned an outcome at the end of the study or on the day when a parasitic recurrence is observed. An overall point estimate of the drug efficacy is then derived based on the outcome assessments from all the patients enrolled at the end of the study. Such routine surveillance studies (called treatment efficacy study, TES) provide critical information to the National Malaria Control Programme (NMCP) to guide and optimise treatment policy.

Standard implementation of a TES follows the protocol recommended by the World Health Organization (WHO). The WHO released its first standardised protocol for assessing *in vivo* efficacy of antimalarial drugs in 1965 as chloroquine resistance began to proliferate (which was identified a few years earlier) [1,61]. The protocol outlined a field test with an objective to establish the sensitivity of the parasites to a standard regimen of chloroquine dose administered under observation. This protocol was revised in 1967 and 1972 recommending that patients be kept in a mosquito free environment to prevent new infections [62]. The focus of these early protocols was on measuring the parasitological sensitivity, and antimalarial drug efficacy was defined based on the outcomes such as recurrent peripheral parasitaemia and time to parasite clearance. This remained largely the same for the next two decades, when in 1996 a major revision was made to the protocol [40]. The focus now shifted from parasitological response to adequate clinical response (ACR), defined as patients without clinical disease and included those with asymptomatic recurrent parasitaemia [63]. The recommendation was

revised again in 2009 when a composite endpoint of both parasitological and clinical assessment was adopted, and this was defined as adequate clinical and parasitological response (ACPR) [64]. The protocol recommended parasitic genotyping be carried out to differentiate a novel parasitic occurrence (new infection) from the recurrence of the original parasite (recrudescence) [64]**.**

In parallel to the evolution of TES protocols (Figure 2.1), the statistical approaches used for measuring antimalarial drug efficacy has also been revised over the years. The 2009 protocol remains the current standard for implementing a TES and some of the limitations of the current statistical approaches outlined in this protocol will be discussed.

2.1.1 Aims and objectives

The aim of this chapter was to carry out a literature review on the evolution of statistical approaches for estimating efficacy for antimalarial regimens and critically assess the limitations of current statistical practices.

The specific aims were:

I. To review the evolution of statistical practices for deriving antimalarial efficacy.

II. To review the challenges associated with derivation of antimalarial efficacy in a standalone TES.

III. To review the challenges associated with comparing two antimalarial regimens.

The rest of the chapter is structured as follows: Section 2.2 describes a motivational study used to highlight some of the challenges currently faced in the derivation of efficacy in an antimalarial study. Section 2.3 reviews the evolution of statistical approaches for deriving efficacy (Objective I); Section 2.4 describes the currently recommended design and analysis of a regular therapeutic efficacy study. Section 2.5 highlights some of the challenges associated with estimating drug efficacy in standalone antimalarial study (Objective II). Challenges specific to comparative studies is explored in Section 2.6 (Objective III) and the chapter is concluded in Section 2.7.

2.2 Motivational study

Data from a large randomised controlled trial which compared three antimalarial regimens in four different sites in Uganda from 2002 to 2004 is used as a motivational study [65]. In this study 2,160 patients aged 6 months or older were randomised to one of the three treatments: chloroquine-sulfadoxine-pyrimethamine (CQSP), amodiaquine-sulfadoxine-pyrimethamine (AQSP) or artesunate-amodiaquine (ASAQ). The primary endpoint in the study was the risk of parasitological failure adjusted by PCR genotyping on day 28. The study was standardised using the WorldWide Antimalarial Resistance Network (WWARN) clinical protocol [66]. Hence, the estimates reported in the original article were slightly different to the estimates derived after standardisation.

2.3 Review of analytical approaches for estimating drug efficacy

Two approaches have been used in deriving efficacy estimates in antimalarial studies:

i. Proportion cured
ii. Survival analysis

2.3.1 Proportion cured

The proportion of patients cured within a specified time period of study follow-up has been the most common method for estimating antimalarial drug efficacy. The cured proportion is often referred to as “cure rate”, and statistically speaking this is misleading since this is not a rate, but a point estimate of cumulative risk at a pre-defined time point. The cured proportion is calculated using a per-protocol (PP) approach or an intention to treat (ITT) principle.

In the PP approach, the proportion cured is derived from all the patients who are followed until treatment failure is observed or reach the planned end of the study follow-up period. Those with protocol deviations plus those who develop a new infection or are lost to follow-up are excluded. Whilst relatively easy to calculate, this approach ignores information provided by patients who experience protocol deviations or are lost to follow-up leading to a loss in analysable sample size and hence precision. Patients failing treatment are more likely to become symptomatic and seek re-treatment and thus be detected passively, whereas patients who are cured are more likely to tire from active detection and be lost to follow-up - these attrition biases result in an underestimation of treatment efficacy [60]. In a more conservative analysis using the ITT principle, the evaluable population includes all patients enrolled in the study. Patients who are lost to follow-up or those who experience protocol deviations are considered as treatment failures (with new infections considered as treatment successes), a strategy leading to an overestimation of failure.

2.3.2 Survival analysis

Survival analysis using the Kaplan-Meier (K-M) method provides an alternative strategy. Technically, this method retains more patients in the statistical analysis and maximises the information available from each patient thus increasing the precision of the derived estimates [60,67]. Patients who do not fail treatment during the study period and those were lost to follow-up are included in the analysis up until the time of last recorded visit at which point they are considered "censored". Similarly, new infections are considered as censored observations on the day of occurrence, while indeterminate outcomes are excluded when deriving the K-M probability of cure ($\hat{S}_{KM}(t)$) (Table 2.1).

Antimalarial efficacy has been reported as proportional cure since the release of the first WHO protocol in 1965. It took almost three decades before survival analysis was first used in the derivation of antimalarial efficacy by Fontanet and Walker (1993) and Price et al (1995) [68,69]. The turn of the millennium found a widespread use of survival analysis and was mentioned in the 2001 and 2003 WHO guidelines [62,70]. Since then several studies have endorsed K-M survival analysis as the preferred approach for deriving antimalarial efficacy. Stepniewska and White (2006) provided a detailed assessment on methodological approaches used in antimalarial studies and strongly advocated using the K-M method [71]. Guthmann et al (2006) pooled datasets from 13 trials (n=2,576 patients) to examine the variability in derived estimates when PP and K-M approach were used [72]. In total, 6% of the samples were lost by day 28 using K-M method when indeterminate outcomes were excluded and new infections were considered as censored, compared to a 25% reduction in sample size using PP method. They reported a lower risk of recrudescence using the K-M method; the risk differences ranged from -2.3% to 2.3% when indeterminate cases were excluded. A similar finding was reported by Ashley et al (2008), where the PP method was associated with a 34% reduction in sample

size as opposed to <10% with K-M survival analysis [73]. The Kaplan-Meier method was finally endorsed as the preferred approach for deriving drug efficacy in the 2009 WHO protocol detailing the design and analysis of therapeutic efficacy studies (Table 2.1)[64]**.**

Table 2.1: Outcome assignment for derivation of antimalarial efficacy

End-point for day X (X= 28 or 42)	Cumulative success or failure probability (Kaplan-Meier analysis)	Proportion (per-protocol analysis)
Adequate clinical and parasitological response at day X	Success	Success
Early treatment failure	Failure	Failure
Late clinical failure before day 7	Failure	Failure
Late clinical failure or late parasitological failure on or after day 7		
• falciparum recrudescence	Failure	Failure
• falciparum reinfection	Censored day of reinfection	Excluded from analysis
• other species with falciparum recrudescence	Failure	Failure
• other species infection	Censored day of infection	Excluded from analysis
• undermined or indeterminate PCR	Excluded from PCR-corrected analysis	Excluded from analysis
Loss to follow-up	Censored last day of follow-up according to timetable	Excluded from analysis
Withdrawal and protocol violation	Censored last day of follow-up according to timetable before withdrawal or protocol violation	Excluded from analysis

(Source: WHO-2009 [64])

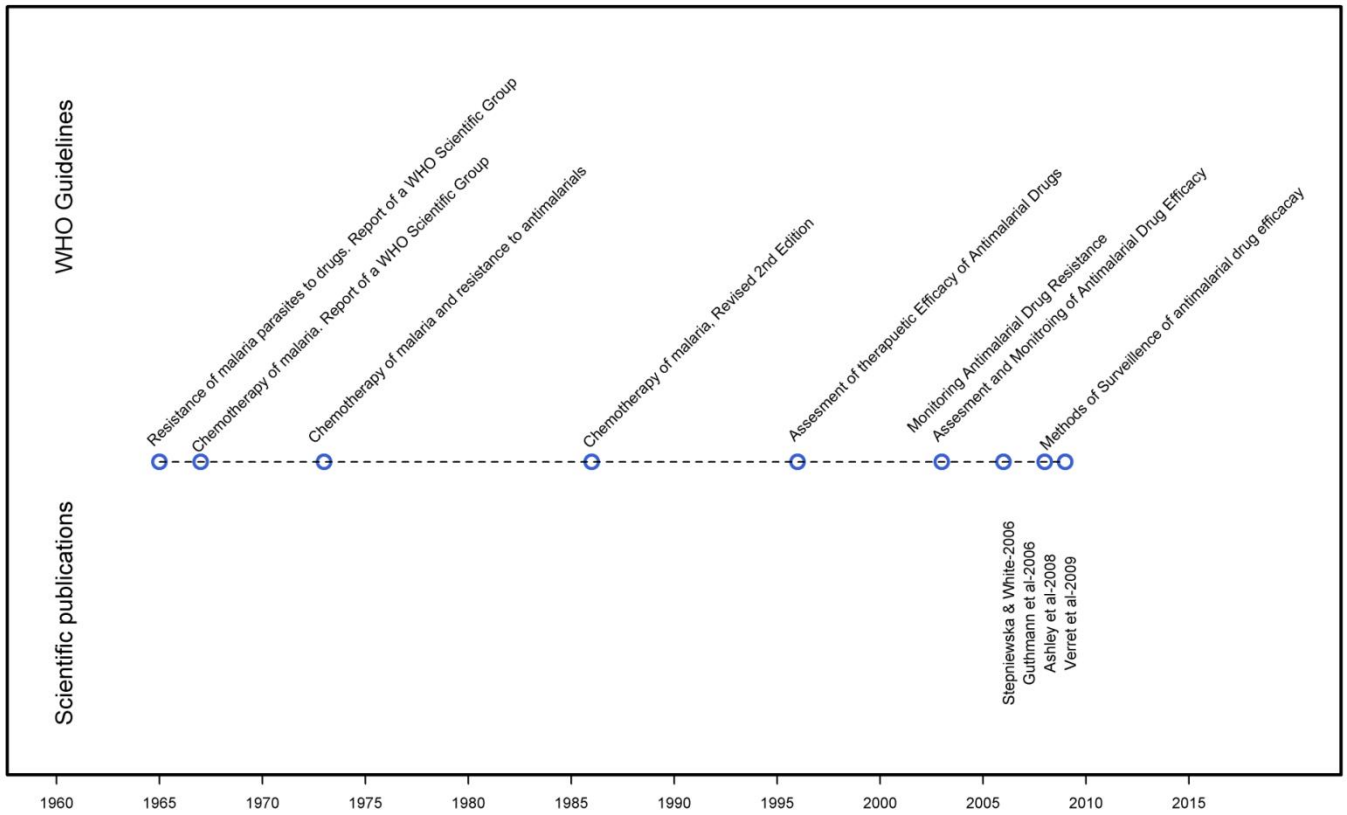


Figure 2.1: The evolution of methodological guidelines for antimalarial studies. From From Dahal et al. Statistical methods to derive efficacy estimates of anti-malarials for uncomplicated Plasmodium *falciparum* malaria: pitfalls and challenges. *Malar. J.* 2017;16:430. [74]. Reproduced under creative commons attribution (CC-BY) license.

2.4 The current design and analysis of a therapeutic efficacy study

The failure estimates derived from a therapeutic efficacy study (TES) provides information on the current status of parasite sensitivity with which national malaria control programmes revise and determine the national treatment policy. TES studies enrol a minimum of 50 patients for assessing the efficacy of the current first line artemisinin combination therapy (ACT) [64]. The ACT is administered as a 3-day regimen, and clinical and parasitological measurements are taken when a patient first comes to the clinic (day 0) and then again on days 1, 2 and 3. Thereafter, patients are actively followed up on day 7 and then weekly until the end of the study follow-up when an outcome is assigned. A patient is classified as treatment success, failure, excluded or censored from the analysis as outlined in Table 2.1.

In subjects with recurrence, treatment success is defined based on the paired analysis of the DNA samples of the parasites collected at baseline and again on the day of recurrence. Three genes are currently used, usually in a sequential approach: merozoite surface protein 1 (*msp*-1), merozoite surface protein 2 (*msp*-2) and glutamate rich protein (*glurp*). Matching DNA fingerprinting in the pre- and post- recurrence sample suggests that a subsequent recurrence is a recrudescent parasite, whereas differences in the paired samples indicates a novel infection. Once treatment outcome is assigned to each of the individuals, an estimate of the drug efficacy is derived using the K-M method. The estimation of per-protocol cure is also encouraged to facilitate comparisons of the results against historical studies.

There are several challenges associated with the design, analysis and interpretation of a standard TES. The remainder of the chapter critically examines the limitations of the current TES protocol from a statistical perspective. I will highlight some of these challenges using the motivational study described in Section 2.2.

2.5 Challenges in estimating antimalarial efficacy for a single arm

2.5.1 Defining optimal duration of follow-up for capturing recrudescence

The drugs partnered with artemisinin (or derivatives) differ in terms of their terminal elimination half-life (Table 1.1, Chapter 1). Hence the study duration is an important consideration in the design and analysis of antimalarial studies. The recommended duration of study follow-up has been regularly updated as the TES protocol has evolved since it was first released in 1965.

In the 1965 protocol, an intensive field test with observation duration of up to 11 days was recommended for assessing the sensitivity of the *P. falciparum* parasites to chloroquine [61]. When a revision was made in 1973, a day 28 follow-up duration was proposed for assessing the efficacy of aminoquinolines [75]. A further revision was carried out in 1996 with a follow-up duration of 14 days recommended in the areas of intense transmission, and 28 days in the areas of low and moderate settings [63]. Owing to the logistical challenges associated with maintaining a study with longer follow-up duration, the 14 days follow-up ended up being widely adopted, even in areas of low and moderate transmission. This was met with a widespread criticism and concerns were raised against the poor sensitivity of a 14 days study protocol in detecting low grade resistance [58,76]. In 2002, a clinical trial carried out in Uganda demonstrated that the risk of failure was 2.1 to 6.8 fold higher on day 28 compared to day 14; the authors proposed that 28 days follow-up should be adopted as the minimum [24]. A subsequent WHO guideline in 2003 reviewed all the accumulated evidence; however did not change the recommendation on follow-up duration in the areas of high transmission [62].

In 2004, Stepniewska et al [77] carried out a large individual patient data pooled analysis (n=13,722 patients), which quantified the degree of underestimation of treatment failures using

a 14 days follow-up. The study included data from 96 treatment arms from randomised controlled trials conducted between 1990 and 2003 with duration of follow-up of 28 days or longer in Asia and Africa. The day 14 assessment had a sensitivity of 0 - 36% for detecting recrudescence, with zero predictive value, whilst the 28-day assessment had a sensitivity of 66% (28 - 100%) and effectively predicted the true risk of failure if transmission intensity or genotyping information were available. The analysis was stratified by the elimination half-lives of the drugs: short (<1 day), medium (1 to <7 days) and long (≥7 days). Non-linear models were fitted to the proportion of failures recorded against time of follow-up (Figure 2.2). The models estimated that a day 14 assessment failed to capture 98%, 90% and 86% of overall recrudescences for drugs with short, medium and long half-lives respectively which would have been observed had the follow-up duration extended up to 63 days. In contrast, a day 28 follow-up was predicted to miss 26%, 36%, and 60% of treatment failures respectively (Figure 2.2).

In addition, Guthmann et al (2006) showed that the estimates of treatment failure were higher on day 28 compared to day 14 following sulphadoxine-pyrimethamine or amodiaquine regimen (but not for chloroquine). For amodiaquine, the risk of failure increased two-fold when follow-up was extended from 14 to 28 days [72]. Resistance to a chloroquine regimen at this time was already very high by day 14, reinforcing that the day 14 test was only be able to detect high-grade resistance [58].

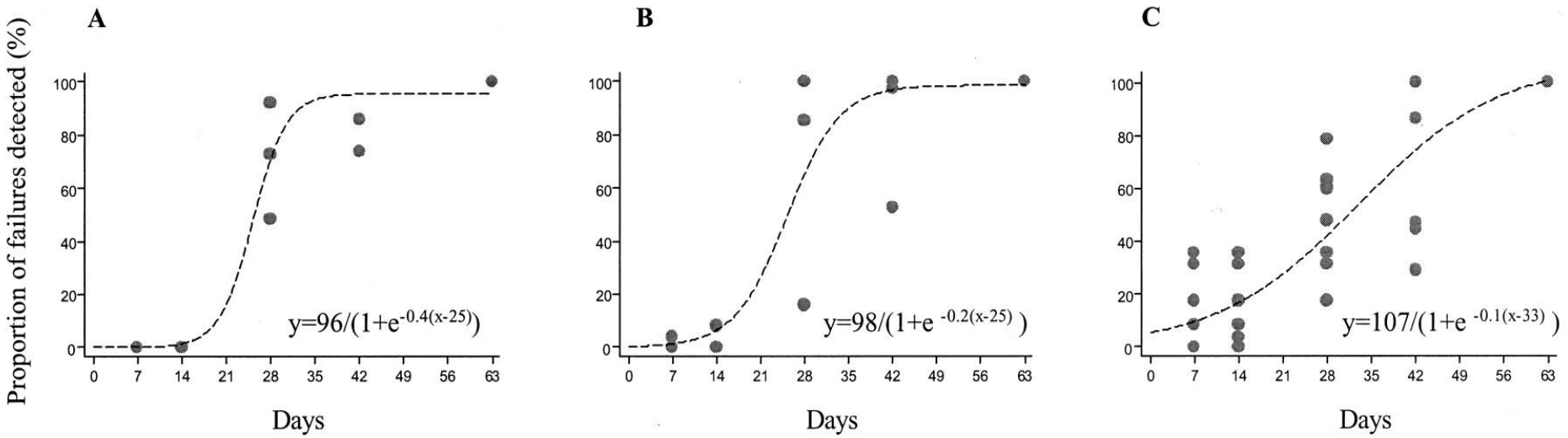


Figure 2.2: Relationship between detection of treatment failures against half-life of the partner drugs. From Stepniewska K, Taylor WRJ, Mayxay M, Price R, Smithuis F, Guthmann J, et al. In Vivo Assessment of Drug Efficacy against Plasmodium falciparum Malaria : Duration of Follow-Up. *Antimicrob. Agents Chemother*. 2004;48:4271–80. [77]. Reproduced with permission from ASM Journals.

Legend: Relationship between the proportion of antimalarial treatment failures detected and the duration of follow-up for (A) short half-life, (B) medium half-life, and (C) long half-life antimalarial drugs. PCR genotype-corrected results are shown.

In 2006, the WHO changed the recommendations to 28 days as the minimum follow-up duration for antimalarial trials, regardless of transmission setting [30]. It was finally the 2009 protocol which provided a definitive recommendation regarding the follow-up duration for drugs with different elimination half-lives [64]. A minimum follow-up duration of 28 days was recommended for ACTs where the partner drugs had elimination half-lives of less than 7 days (amodiaquine, lumefantrine), and 42 days for partner drugs with a longer half-life (piperaquine, mefloquine).

Increasing concerns have been raised over what remains the optimal follow-up duration for ACT regimens since Stepniewska et al (2004) first reported that a day 28 follow-up will miss an estimated 36% of recrudescence for drugs with medium half-lives [77]. The review was largely based on existing data on non-ACT regimens available at the time and generalising the results to ACTs should be done with caution. A clinical trial conducted in Tanzania between 2002 - 2003 demonstrated that a day 28 follow-up missed 28% (5/18) of confirmed recrudescence with artesunate amodiaquine regimen and 58% (7/12) with the artemether lumefantrine regimen, underlying the importance of a longer follow-up duration with these regimens [78]. Similarly, data on artemether lumefantrine from Ethiopia shows that 88% (14/16) of the PCR confirmed recrudescences were detected after day 28 (see Table 2 of [79]). The observed numbers of recrudescences in any single studies are far and few, thus limiting the ability to draw robust and generalisable inference on the optimal duration of follow-up. The large WorldWide Antimalarial Resistance Network (WWARN) database (Chapter 3) provides a unique resource to adequately explore the issue and unravel temporal trends of recrudescence which can be observed during the study follow-up. This will be carried out in Chapter 4.

2.5.2 Considering new infection as a competing risk event

In an antimalarial trial of uncomplicated *P. falciparum* malaria, new infections with this or other species such as *P. vivax* can preclude the occurrence of recrudescence during the ensuing weeks. In statistical literature, events which prevent the primary event of interests from occurring or fundamentally alter their probability of being observed are referred to as competing risk events [80]. In studies conducted in malaria endemic settings, competing risk events of new infections can sometimes occur in more than 30% of enrolled patients [60]. The current approach for dealing with a new infection is to censor them on the day of occurrence in a Kaplan-Meier analysis when deriving antimalarial efficacy $\left(\hat{S}_{KM}(t)\right)$. The complement $\left(1-\hat{S}_{KM}(t)\right)$ is frequently used to report the cumulative estimate of drug failure at the end of the study follow-up.

The K-M method makes a fundamental assumption of non-informative censoring, which means patients who are censored have the same risk of observing the outcome of interest as those who are still being followed-up. Censoring a recurrence due to new infection (a competing risk event) cannot be regarded as non-informative as the patient will no longer be at risk of observing subsequent recrudescence (as alternative treatment is given on the detection of a parasitic recurrence). As such, failure estimates derived from a K-M analysis have been described as providing an estimate of the marginal risk of recrudescence - the risk of recrudescence in a virtual world where new infections do not occur [81]. This is only possible when all enrolled participants are subject to isolation in an environment where it is impossible to get another mosquito bite (thus preventing new infection). In practice, antimalarial trials are almost invariably conducted in endemic settings, where new infections frequently occur. Researchers and policy makers are often interested in real world estimates derived by taking the presence of competing risk events into account.

Survival analysis using Cumulative Incidence Function (CIF) provides an alternative strategy for estimating failure in the presence of competing risk events [82]. The naïve K-M estimator $\left(1-\hat{S}_{KM}(t)\right)$ which is derived by censoring competing risk events has been shown to result in upwards biased estimates compared to the CIF estimate of failures in the real world setting where competing risk events occur [80,83–87]. Despite this, the naïve K-M estimator is commonly used to derive the cumulative probability of failure in antimalarial studies. A comparison of the K-M method and CIF is illustrated using data from 166 patients treated with chloroquine-sulphadoxine-pyrimethamine (CQSP) in the motivational study [65]. The estimate of recrudescence on day 28 was 0.376 [95% CI: 0.264 - 0.470] using the $1-\hat{S}_{KM}(t)$ method compared with 0.265 [95% CI: 0.199 - 0.331] estimated by the CIF method (Figure 2.3A). The corresponding probabilities of new infections on day 28 were 0.761 [95% CI: 0.667 - 0.828] using 1 minus K-M method and 0.654 [95% CI: 0.584 - 0.723] using the CIF. These values represent an absolute overestimation of 0.11 for both recrudescence and new infection. Expressed in relative terms, this represents an overestimation by 41.8% for recrudescence and 24.4% for new infection. Another observation is that the sum of the cumulative probability of recurrence estimated using the 1 minus K-M method for recrudescence and new infection on day 28 was 113.7% (Figure 2.3B). In such situation, the interpretation of the K-M estimates as a probability is no longer valid as the sum exceeds 100%.

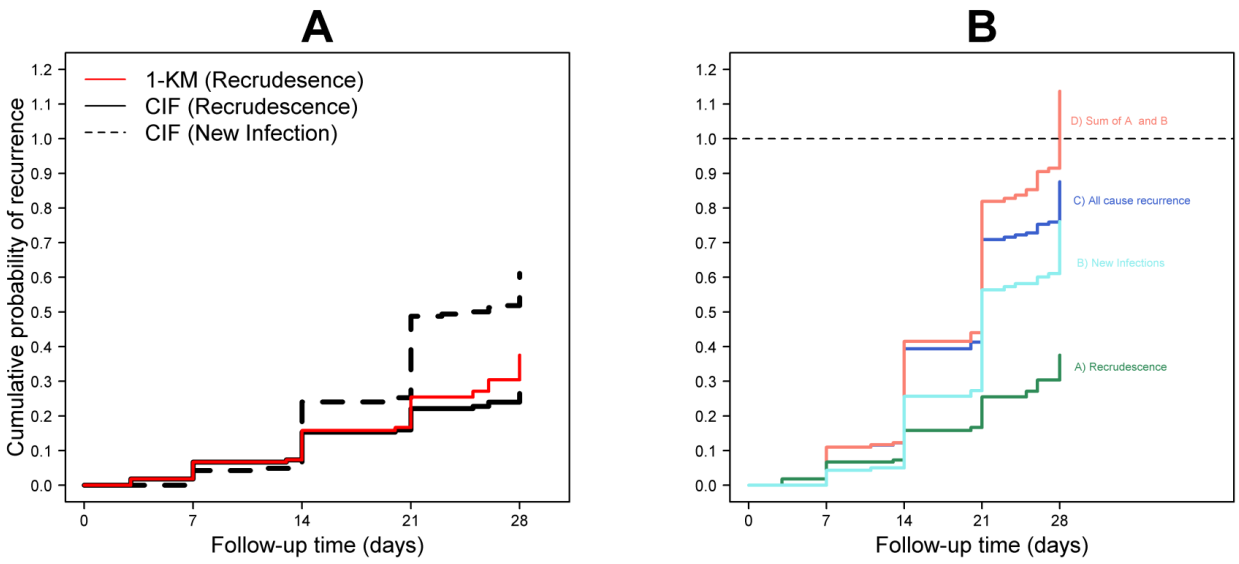


Figure 2.3: Recrudescence failure using in Tororo [65]**.** From Dahal et al. Statistical methods to derive efficacy estimates of anti-malarials for uncomplicated Plasmodium falciparum malaria: pitfalls and challenges. *Malar. J.* 2017;16:430. [74]. Reproduced under creative commons attribution (CC-BY) license.

Legend: **A**) Cumulative probability of failure due to recrudescence derived using Kaplan-Meier approach, which censored new infections (red line), and using Cumulative Incidence Function (solid black line), which accounts for the presence of competing risks (dotted black line). **B**) Estimates of the cumulative probability of recurrences for recrudescence (green), new infections (light blue) and overall recurrences (dark blue line) using K-M method. The sum of the probabilities for recrudescence and new infection is presented as the pink line and exceeds the value of 1 on day 28.

2.5.3 Dealing with an indeterminate outcome

Indeterminate outcomes arise in antimalarial study as a result of failure to amplify the DNA samples, due to the loss of paired samples, the use of untreated filter paper for sample collection which has low success rate, or from a failure to interpret the results [51]. The current approach for dealing with indeterminate outcomes is to exclude them from the analysis in a complete case (CC) approach [64]. This approach leads to a loss in evaluable sample size and precision of the derived estimates. When a large fraction of recurrences are indeterminate, this undermines the validity of the results presented.

Various approaches have been proposed to deal with indeterminate outcomes in the antimalarial literature. Stepniewska and White (2006) proposed using the ratio of the observed proportions of recrudescence to new infection at any given time point to correct the indeterminate outcomes, and classify them as either a recrudescences or a new infection based on this information [71]. However, such approach can only be applied using a pooled database where there are enough recurrences to characterise the temporal profile. Machekano et al (2008) [88] considered indeterminate outcome as missing data and compared the efficacy estimates derived using the well-established methods of multiple imputation (MI), inverse probability weighting (IPW), and doubly robust IPW approaches against the CC estimator using a trial conducted in Uganda. The study reported an overestimation in cured proportion by 3.2% using the CC analysis compared to IPW and MI methods in patients treated with chloroquine-sulphadoxine-pyrimethamine (Table 2.2).

Table 2.2: Failure estimates using missing data methods in Machekano et al (2007)[88]

Method	Chloroquine-SP arm	Amodiaquine-SP arm
Complete Case (CC)	0.499 [0.462-0.535]	0.178 [0.130- 0.238]
Multiple Imputation (MI)	0.531 [0.491-0.569]	0.195 [0.168-0.222]
Inverse Probability Weighting (IPW)	0.531[0.498-0.567]	0.193 [0.143-0.259]
Inverse Probability Weighting (Doubly Robust)	0.522 [0.487-0.556]	0.186 [0.137-0.249]

(Source: Machekano et al (2007)[88])

2.5.4 Interval censoring

In an antimalarial study, patients are scheduled for daily visit in the first 3 days and then weekly from day 7 until the end of the study follow-up. Recurrent parasitaemia thus commonly occurs on pre-scheduled follow-up time-points, particularly in patients who are asymptomatic. However, the true timing of recurrent parasitaemia is often between the scheduled visit days and this gives rise to interval-censored failure times. Ignoring such intervals might lead to an under or overestimation of failures at a given time point, especially when it is assumed that failures occur at the beginning or the end of the interval. The magnitude of the bias in the derived estimates has been shown to accentuate as the length of the interval gets wider [89]. Interval censored methods are now part of standard statistical packages [90–93]. In the motivational study, the Kaplan-Meier estimates derived with and without accounting for interval censoring were similar, and are presented in Table 2.3.

Table 2.3: Estimates with and without interval censoring in Uganda data [65]

Study site	Treatment	Kaplan-Meier estimates [95% CI] (ignoring interval censoring)	Kaplan-Meier estimates [95% CI] (accounting for interval censoring)
Apac	AQSP	0.927 [0.888 - 0.968]	0.921 [0.881 - 0.963]
	ASAQ	0.896 [0.846 - 0.948]	0.889 [0.839 - 0.943]
	CQSP	0.748 [0.677 - 0.827]	0.743 [0.671 - 0.822]
Arua	AQSP	0.860 [0.803 - 0.920]	0.853 [0.796 - 0.915]
	ASAQ	0.921 [0.879 - 0.965]	0.915 [0.872 - 0.96]
	CQSP	0.485 [0.390 - 0.605]	0.480 [0.384 - 0.599]
Jinja	AQSP	0.880 [0.830 - 0.932]	0.874 [0.824 - 0.927]
	ASAQ	0.958 [0.928 - 0.989]	0.953 [0.921 - 0.985]
	CQSP	0.611 [0.534 - 0.698]	0.606 [0.530 - 0.694]
Tororo			
	AQSP	0.820 [0.758 - 0.887]	0.825 [0.765 - 0.890]
	ASAQ	0.876 [0.813 - 0.944]	0.879 [0.818 - 0.945]
	CQSP	0.624 [0.530 - 0.736]	0.631 [0.538 - 0.740]

AQSP = amodiaquine-sulfadoxine-pyrimethamine

ASAQ = artesunate-amodiaquine

CQSP = Chloroquine-sulfadoxine-pyrimethamine

CI = Confidence Interval

2.5.5 Presenting results from multicentre studies

Antimalarial studies often enrol patients from more than one centre in order to achieve adequate sample size. This poses some difficulty when deriving an overall Kaplan-Meier (K-M) estimate of drug efficacy as there might be underlying heterogeneity between the study sites in censoring patterns, attrition rates, patient demographics, and transmission intensity profile. The Cochrane Handbook for Systematic Reviews of Interventions [94] acknowledges the challenges in presenting the K-M estimates from several studies:

"Kaplan Meier plots for all pooled participants across trials in a meta-analysis have previously been presented in medical journals. This practice breaks with the principle of comparing like with like. For this reason, until further discussions have taken place the Statistical Methods Group is unable to recommend inclusion of such plots in Cochrane reviews."

This statement made regarding the presentation of an overall K-M in a meta-analysis of several trials is similar to pooling K-M from several centres. However, there exists no clear consensus among researchers on the best approach to synthesise survival estimates across studies/sites, neither for the aggregate meta-analysis nor for individual patient data meta-analysis.

If cured proportion is used as the measure of drug efficacy, then it is straight forward to pool the estimates of drug efficacy from several centres. The estimates derived at each of the study sites at a fixed time point can be combined using standard meta-analysis methods for pooling proportions using the approach of DerSimonian and Laird (D+L) [95]. However, Kaplan-Meier method remains the preferred method for deriving estimates of antimalarial drug efficacy and the application of the D+L method when pooling K-M curve isn't straight forward.

The simplest approach for deriving Kaplan-Meier estimates of drug efficacy from a multi-centre study is to pool data across the study sites to get an overall estimate under the assumption

of independence (naïve approach), an approach recommended by Srinivasan and Zhou (1993) provided the data from the several sites (studies) are independent [96]. However, such an approach can falsely result in a treatment appearing to be beneficial after pooling data from several sites, in situations where the reverse is in fact true [97]. Comberscure et al (2014) have recently proposed a method to pool K-M estimates from several studies at specific time-points [98]. Under this approach (called the **MetaSurv** approach), the K-M estimate of drug efficacy and the number of patient at risks are derived for each site at a time-point of interest. DerSimonian and Laird's (D+L) method is then applied after carrying out an arc-sine transformation to obtain a pooled estimate. This is available through the MetaSurv subroutine on open source R software [99].

The approach proposed by Comberscure et al (2014) has been illustrated with the dataset on ASAQ arm from the motivational study [65]. The K-M estimate of drug efficacy for each of the study sites were extracted at all pre-scheduled visit-times (days 1, 2, 3, 7, 14, 21 and 28) including the number of subjects at risk. The estimates derived using the naïve approach and MetaSurv approach are presented in Table 2.4. In this example, the pooled estimates on day 28 were similar between the two approaches with the confidence intervals being wider for the estimate derived using the latter approach. However, the MetaSurv approach needs to be used with caution as the bias in estimates could be high when sample size per site is small or when the events are rare [98]. In such a situation, an alternative approach could be to apply the DerSimonian and Laird's procedure after complementary log-log transformation $\{\log(-\log(.))\}$ of the survival function. This approach has the useful statistical property of giving correct coverage probability based on as few as 25 observations [100].

Table 2.4: Kaplan-Meier estimates for artesunate-amodiaquine arm in the motivational study [65]

	Pooled Kaplan-Meier estimate [95% confidence interval]	
Day	Naïve approach [a]	**MetaSurv** approach [98][b]
7	1.000	0.994 [0.987 - 1.000]
14	0.994 [0.989 - 1.000]	0.989 [0.979 - 0.998]
21	0.949 [0.933 - 0.966]	0.943 [0.916 - 0.970]
28	0.918 [0.895 - 0.941]	0.909 [0.873 - 0.948]

[a] Kaplan-Meier estimates were derived assuming the data came from one single study. Patients with new infection, indeterminate outcome and lost to follow-up were censored when deriving the K-M estimates for recrudescence.

[b] Kaplan-Meier estimates and associated number of patients at risk were extracted on pre-scheduled follow-up days 1, 2, 3, 7, 14, 21, and 28. The estimates were pooled using *MetaSurv* package in R. I-Squared statistic for heterogeneity = 0%.

2.6 Challenges for comparative efficacy trials

Investigating suitable alternative treatment regimen requires a randomised clinical trial and these comparative trials raise further difficulties in the analyses and interpretation of data. Some of the challenges are discussed in this section.

2.6.1 Comparison of antimalarial drugs with different pharmacokinetic profiles

Antimalarial drugs with different elimination half-lives may vary considerably in the period of time during which peripheral parasite growth is suppressed. Since post treatment prophylaxis reduces the risk of new infection and delays the timing of recurrent parasitaemia, for drugs with similar efficacy the comparative results may be biased against the drug with the shorter half-life. Conversely, a new drug which has a long elimination half-life may result in the false impression of good efficacy. Consider dihydroartemisinin-piperaquine (DP) and artemether-lumefantrine (AL). Piperaquine is a slowly eliminated regimen with a terminal elimination half-life of 13.5 to 28 days and will suppress parasitic growth for a far longer period compared to lumefantrine, which has an elimination half-life ranging from 1.4 to 11.5 days (Table 1.1) [8]. In a longitudinal study conducted in Uganda, the median time to recrudescence was 21 days (range: 21-50 days) in the AL arm compared to 42 days (range: 33-51 days) in the DP arm [101]. Hence, patients treated with DP will encounter fewer recrudescence/reinfections during the first 28 days of the follow-up and thus a comparison of efficacy at day 28 will be biased against artemether-lumefantrine (see Figure 2.4 for an illustration of this bias).

Ideally, the comparison of overall efficacy of antimalarials should be carried out at a time when drug concentrations cease to suppress parasite growth (when this falls below the minimum inhibitory concentration - MIC) and after allowing the time for the parasites to reach the limit of detection. If the drug concentrations fall below the MIC on day 18, and assuming 10^1 parasites are circulating in the blood, these parasites will reach the threshold for detection (~10^8

parasites) in 7 parasite life cycles (approximately 14 days) assuming an efficient multiplication of 10-fold per cycle. In this scenario recrudescences begin to appear by day 33. Under the same assumption, a drug which provides more prolonged prophylaxis (e.g. with drug concentrations in plasma falling below than MIC on day 31) will result in recrudescent parasites reaching the limit of detection on day 45. Hence, comparing these two drugs on day 42 will result in a biased conclusion. The comparative efficacy should account, therefore, for differences in the pharmacokinetic profile of the drugs. This is further confounded by the transmission intensity, which determines the risk of new infection (competing risk event).

The implication of the follow-up duration on the comparative efficacy (derived estimate of relative risk measure) will be studied in Chapter 4 using artemether-lumefantrine and dihydroartemisinin-piperaquine. Similarly, the influence of competing risk events in comparative efficacy between two regimens for a number of scenarios which can be observed in a field study will be explored through the use of a simulation study in Chapter 5.

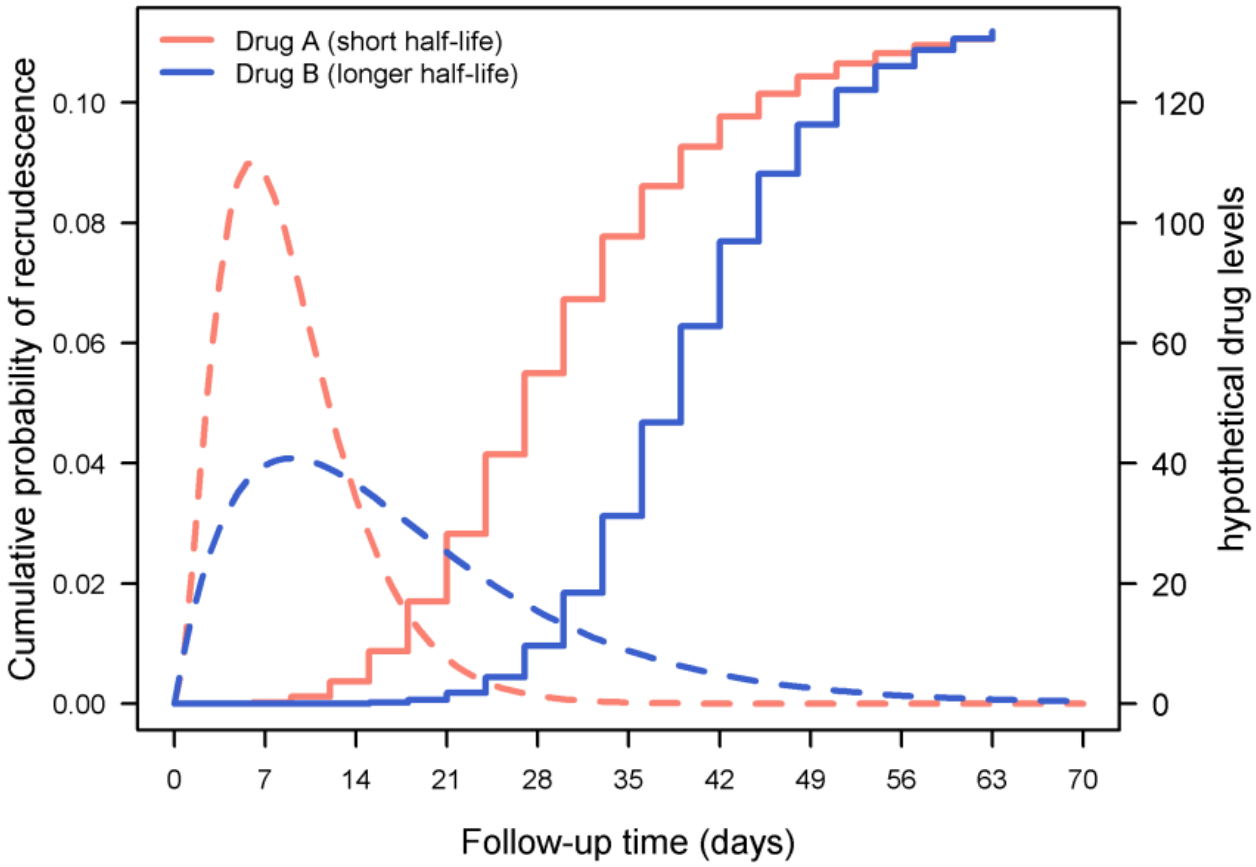


Figure 2.4: Cumulative failure for drugs with different terminal elimination half-lives. From Dahal et al. Statistical methods to derive efficacy estimates of anti-malarials for uncomplicated Plasmodium falciparum malaria: pitfalls and challenges. *Malar. J.* 2017;16:430. [74]. Reproduced under creative commons attribution (CC-BY) license.

Legend: The solid line represents cumulative failure probabilities using the K-M method and the dotted lines are the drug levels with hypothetical units presented on the right y-axis. Drug B (blue) has a longer elimination half-life and the recrudescent failures are patent after day 21 compared to drug A (red) with a short half-life and recrudescences being observed after day 7. However, the cumulative failure estimates for drug B eventually catches up with drug A at around day 63.

2.6.2 Comparing estimates at a fixed point in time in a single centre study

The difference in the derived efficacy estimates between two antimalarial regimens is compared using the log-rank test. The log-rank test uses information throughout the study follow-up period with equal weights given to observed failures at all time points. This test is appropriate under the assumption of proportional hazards (p. 49 of [102]). Intersection of two survival curves may be indicative of non-proportional hazards and the log-rank test will fail to pick up differences. When comparing two antimalarial regimens with elimination half-lives (for example, AL and DP), it might be more relevant to focus on the overall proportion of failures observed at the end of the study follow-up. Consequently, an alternative test is needed which allows comparing the Kaplan-Meier estimates at a specific time-point. Such a test can be constructed from the difference of complementary log-log transformed K-M estimates at a specified time-point and appropriately estimating the standard error for this difference [100].

Let $\hat{S}1(t)$ and $\hat{S}2(t)$ be the two survival estimates for drug A and drug B respectively. Let $\delta = \{\ln\left(-ln\left(\hat{S}1(t)\right)\right) - \ln\left(-ln\left(\hat{S}2(t)\right)\right)\}$ be the difference between these two estimates on a complementary log-log scale. The Chi-squared test statistic (X^2) for comparing the difference in two estimates is given by Klein et al (2007) (referred to as **Klein's test** henceforth) [100]:

$$X^2 = \frac{\{\delta\}^2}{Var(\delta)}$$

where $Var(\delta)$ denotes the variance estimate of δ. The X^2 statistic has an approximate chi-squared distribution with 1 degree of freedom. $Var(\delta)$ can be calculated as:

$$Var(\delta) = \left\{\frac{1}{ln\left(\hat{S}1(t)\right)}\right\}^2 \frac{1}{\hat{S}1(t)^2}.Var\left(\hat{S}1(t)\right) + \left\{\frac{1}{ln\left(\hat{S}2(t)\right)}\right\}^2 \frac{1}{\hat{S}2(t)^2}.Var(\hat{S}2(t))$$

The derivation of this variance is provided in Appendix 2.1. The X^2 statistics is then given by:

$$X^2 = \frac{\{\ln\left(-ln\left(\hat{S}_1(t)\right)\right) - \ln\left(-ln\left(\hat{S}_2(t)\right)\right)\}^2}{\left\{\frac{1}{ln(\hat{S}1(t))}\right\}^2 \frac{1}{\hat{S}1(t)^2}.Var(\hat{S}1(t)) + \left\{\frac{1}{\log(\hat{S}2(t))}\right\}^2 \frac{1}{\hat{S}2(t)^2}.Var(\hat{S}2(t))}$$

The Klein's test approach for comparing estimates at a fixed point in time has been illustrated using example data provided in Appendix 2.2. There was a significant difference between drug A and drug B on day 42 based on the log-rank test ($X^2 = 4.30$, p=0.038) whereas using the Klein's test, no difference in the K-M estimates on day 42 was detected ($X^2 = 0.266$, p=0.606) (Figure 2.5).

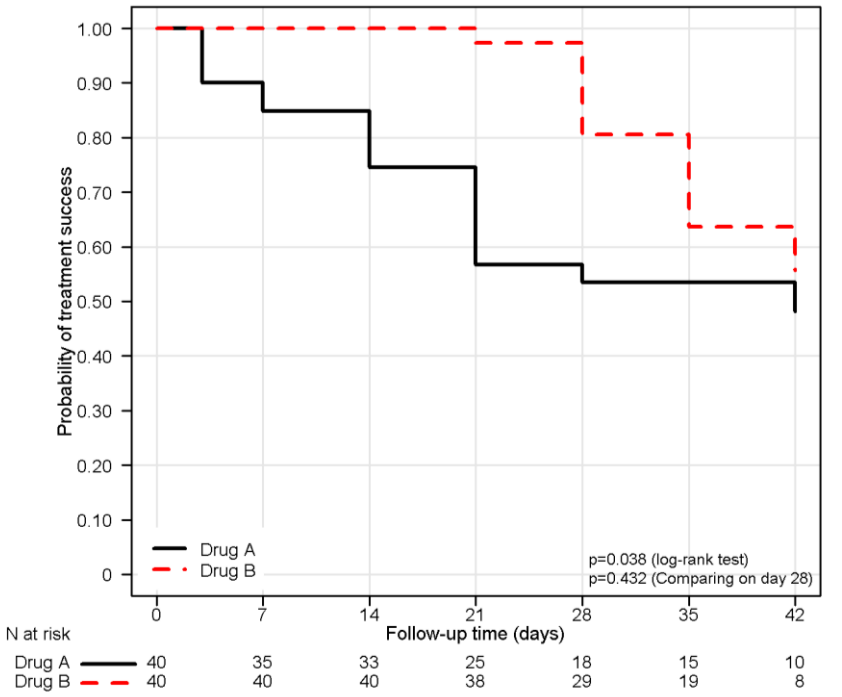


Figure 2.5: Comparison of estimates at fixed point in time. From Dahal et al. Statistical methods to derive efficacy estimates of anti-malarials for uncomplicated Plasmodium falciparum malaria: pitfalls and challenges. *Malar. J.* 2017;16:430. [74]. Reproduced under creative commons attribution (CC-BY) license.

Legend: The data used for this figure is presented in Appendix 2.2.

The Klein's test can be easily extended to the multi-centre studies. Let, $s = 1,2,3, \ldots, m$ be the stratum, and let $\hat{S}_{1s}(t)$ be the survival estimates for drug A in stratum *s* at time *t.* A stratified version of the test has been outlined in Klein et al (2007) [43] and is given by:

$$X^2_{Strat} = \frac{\left(\sum_{s=1}^{m}\{\ln\left(-ln\left(\hat{S}_{1s}(t)\right)\right) - \ln\left(-ln\left(\hat{S}_{2s}(t)\right)\right)\}\right)^2}{\sum_{s=1}^{m}\left\{\frac{1}{ln\left(\hat{S}_{1s}(t)\right)}\right\}^2 \frac{1}{\hat{S}_{1s}{}^2}.Var\left(\hat{S}_{1s}(t)\right) + \left\{\frac{1}{\log\left(\hat{S}_{2s}(t)\right)}\right\}^2 \frac{1}{\hat{S}_{2s}(t)^2}.Var\left(\hat{S}_{2s}(t)\right)}$$

The motivational study data was used to compare the results of the log-rank test against the Klein's test for comparing K-M estimates on day 28. Table 2.5 shows the results of comparing amodiaquine-sulphadoxine-pyrimethamine against artesunate-amodiaquine in four different sites and an overall test taking the sites into account using the two different approaches (stratified Klein's test and stratified log-rank test). In this example, there was no difference in the conclusion based on the two approaches.

Table 2.5: Comparison of K-M estimates at fixed point in motivational study [65]

	χ^2 test for comparing K-M at a fixed time point (day 28) (Klein's test [100]**)**		**Log-rank test for comparing whole curve (day 0 to day 28)**	
Study site	X^2	*p*- value	X^2	*p*- value
Tororo	1.35	0.246	2.00	0.158
Arua	2.72	0.099	2.20	0.142
Jinja	6.18	0.013	6.40	0.011
Apac	0.91	0.340	1.10	0.290
Overall stratified test	4.98	0.026	4.90	0.026

Table 2.6: Challenges in estimating drug efficacy and some alternatives

Challenges	**Current recommended approach**	**Alternative approach**	**Software**
Competing risk event	Censored on the day of event [64]	Cumulative Incidence Function (CIF)	*cmprsk* subroutine R [103]
Interval censoring	Ignored	Interval censored survival estimates	*survival* subroutine R [104]
K-M for multicentre studies	No specific recommendation	Use of meta-analysis approach [98]	*MetaSurv* subroutine R [99]
Comparing K-M estimates	Comparison based on whole survival curve (log-rank test)	Comparison at a fixed point in time based on complementary log-log transformation[100]	R script available as supplementary file from Dahal et al (2017) [74]

2.7 Conclusions

This chapter reviewed the evolution of statistical practices for deriving antimalarial drug efficacy and discussed some of the limitations of the current recommended approaches. Advances in statistical methodologies and their accessibility in standard software have ensured that some of these are easily addressed (Table 2.6). Addressing all the underlying issues identified in this chapter remains beyond the scope of this DPhil study, and only three of the issues will be explored further. First, on the optimal length of follow-up to capture recrudescence failure for currently used ACT regimens (Chapter 4), second, on how to handle new infection within a competing risk framework (Chapter 5), and finally, on handling indeterminate outcomes using missing data theory (Chapter 6).

Chapter 3: Data acquisition and description

The research carried out in this thesis is based on the studies which are part of the WorldWide Antimalarial Resistance Network (WWARN) database. This chapter provides a description of screening and data acquisition of relevant studies from the WWARN database and provides a summary of the studies included.

3.1 The WWARN repository

The WorldWide Antimalarial Resistance Network (WWARN) is a global collaboration of researchers from more than 280 institutions coming together to share expertise and data in a combined effort to answer questions of public health relevance regarding antimalarial drug resistance [105]. The specific questions are answered through a collaborative study group, where principal investigators of the published (or unpublished) studies are invited to share individual participant data (IPD). Researchers agreeing to the terms and conditions of the submission upload anonymised IPD to the WWARN repository through a secure web portal. Clinical trials from different research groups are inevitably heterogeneous in terms of design, execution and methodologies adopted. In order to overcome these methodological heterogeneities, individual studies are then re-analysed using an open and transparent data management and statistical analysis protocol [106]. Such standardisation mitigates the study specific heterogeneity and provides a rational basis for carrying out IPD meta-analysis.

On standardisation, the data are stored in a relational database of several tables containing information on drug regimen, parasitological, clinical, and haematological assessments, genotyping and therapeutic outcomes; all linked by a unique patient identifier. Figure 3.1 shows the WWARN process map which depicts the three different phases of data procurement; data

acquisition from the contributors, data standardisation, and their subsequent re-use in large scale pooled analyses.

The WWARN repository currently holds records on more than 150,000 individual patients treated with antimalarials from 186 studies conducted in Asia, Africa, and South America; the majority of which includes published therapeutic efficacy studies on artemisinin combination therapies [107]. This unparalleled resource has facilitated the comprehensive answering of scientific questions regarding antimalarial drugs which would otherwise have been left unanswered, including providing critical information leading to an optimised drug dosage regimen in a paediatric patient population [34], understanding the effect of drug on early and late parasitological responses [32,33,108], and the effect of antimalarial drugs on blocking onwards transmission of *P. falciparum* malaria [109].

Antimalarial studies are vulnerable to several methodological issues in terms of design and analysis as outlined in Chapter 2. In any given trial, only a few treatment failures are observed which limits the utility of any single study to answer methodological questions of interest, thus necessitating a large database. This thesis is focused on advancing statistical methods in design and analysis of antimalarial trials and facilitating better statistical practices using standardised studies in the WWARN database, the selection of which is described next.

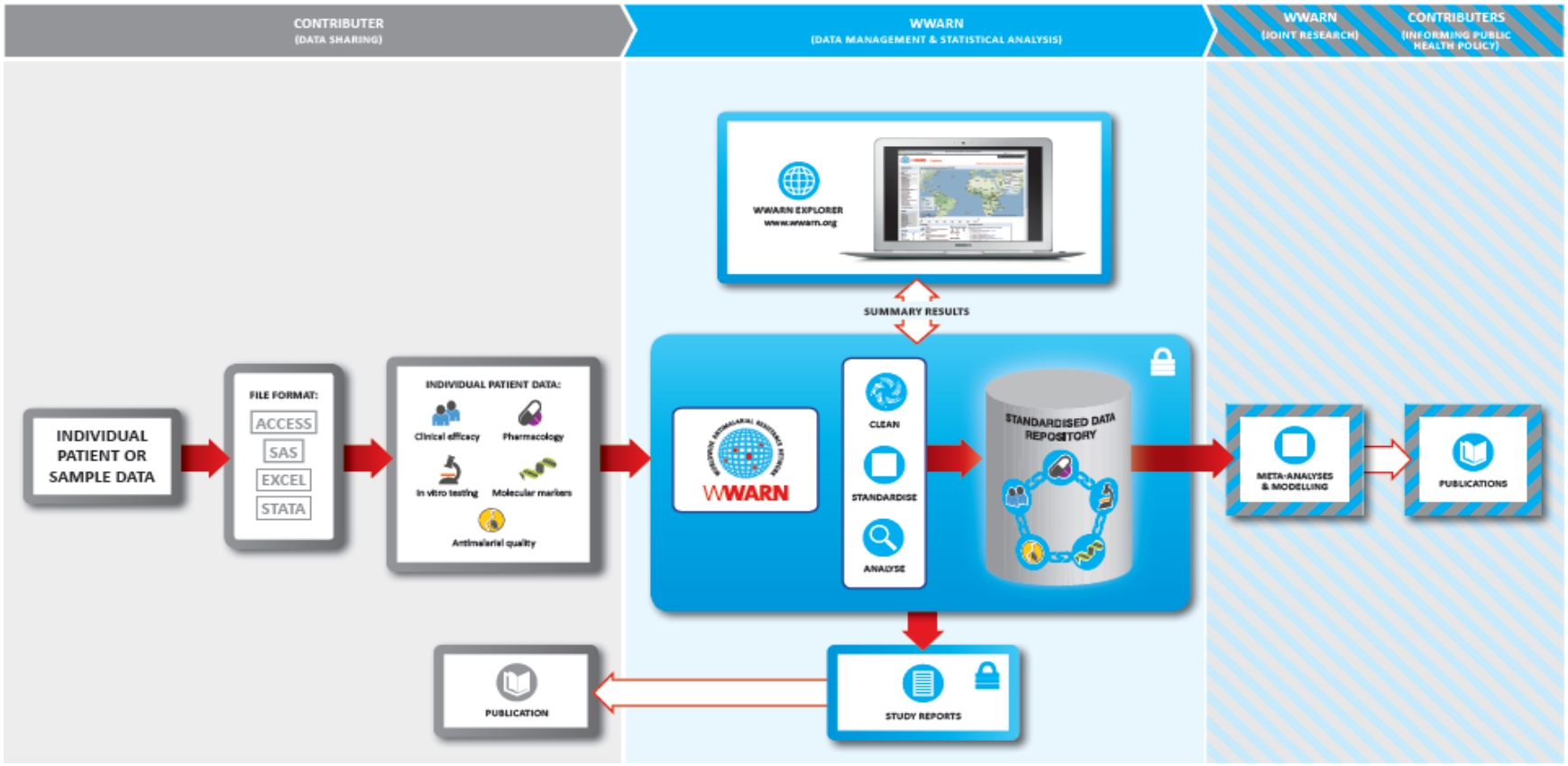


Figure 3.1: The WWARN Process Map

Legend: Individual participant data is shared by researchers to the WWARN secured portal (left panel). The studies are standardised using a common WWARN protocol and the tables of clinical, parasitological and drug measurements are stored in a secured repository (middle panel). On standardisation, the dataset is shared back with the study investigator, and subsequently used in meta-analyses to answer questions of public health importance (right panel).

(Courtesy: Prof. Philippe Guerin)

3.2 Data selection from the WWARN repository

3.2.1 Inclusion criteria

Data in the WWARN repository were eligible for inclusion if they fulfilled the following criteria:

i) Prospective clinical efficacy studies of uncomplicated *P. falciparum* (alone or mixed infections with *P. vivax*).

ii) Clinical trials with one of the following fixed dose combination of artemisinin combination therapies:
   a. artemether-lumefantrine (AL)
   b. dihydroartemisinin-piperaquine (DP)
   c. artesunate-amodiaquine (ASAQ)
   d. artesunate-mefloquine (ASMQ)

iii) Polymerase Chain Reaction (PCR) genotyping performed to distinguish recrudescence from new infection using all or some of the molecular markers merozoite surface protein (*msp)-1, msp-2,* or glutamate rich protein *(glurp*).

iv) A minimum follow-up duration of 28 days.

Studies with only fixed dose combination of the ACTs were considered as these have been shown to provide optimal dosing compared to the loose tablets for the combination ASAQ [32]. Similarly, the loose ASMQ combination varies substantially with regards to the target dose of mefloquine administered; hence the screening was restricted to the fixed dose combination of this regimen. Studies on prophylaxis, severe malaria, pregnant women, and patients with known HIV co-infection, hyperparasitaemia, studies in healthy volunteers and travellers, and pure *P. vivax* malaria formed the exclusion set.

3.2.2 Desirable criteria

- Outcome classification based on each of the individual genotyping markers including the length of pre-and post-recurrence samples for each of the allelic family.
- Multiplicity of infection - the number of different *P. falciparum* strains co-infecting a single host.

3.2.3 Database search

A database search restricted to the WWARN repository was carried out to identify studies fulfilling the inclusion criteria outlined in Section 3.2.1 (dated 06-June-2015). A total of 109 studies with fixed dose formulation of artemisinin combination therapies (ACTs) were identified, of which 15 (13.8%) didn't meet the inclusion criteria (Figure 3.2). An email was sent to the principal investigators (and the corresponding authors) of the 94 eligible studies requesting permission to use these studies for this methodological work. A brief outline of the aims and objectives of the DPhil study was sent along with the email request. In two studies (both were from the same investigator), no response was received and a total of 92 studies where permission was granted formed the DPhil study database (Table 3.1).

3.2.4 Ethical clearance

All studies included in this analysis were obtained in accordance with the laws and ethical approvals applicable to the countries in which the studies were conducted, and were obtained with the knowledge and consent of the individual to which they relate. This individual patient data meta-analysis met the criteria for waiver of ethical review as defined by the Oxford Tropical Research Ethics Committee (OxTREC) as the research consists of secondary analysis of existing, anonymous data [110].

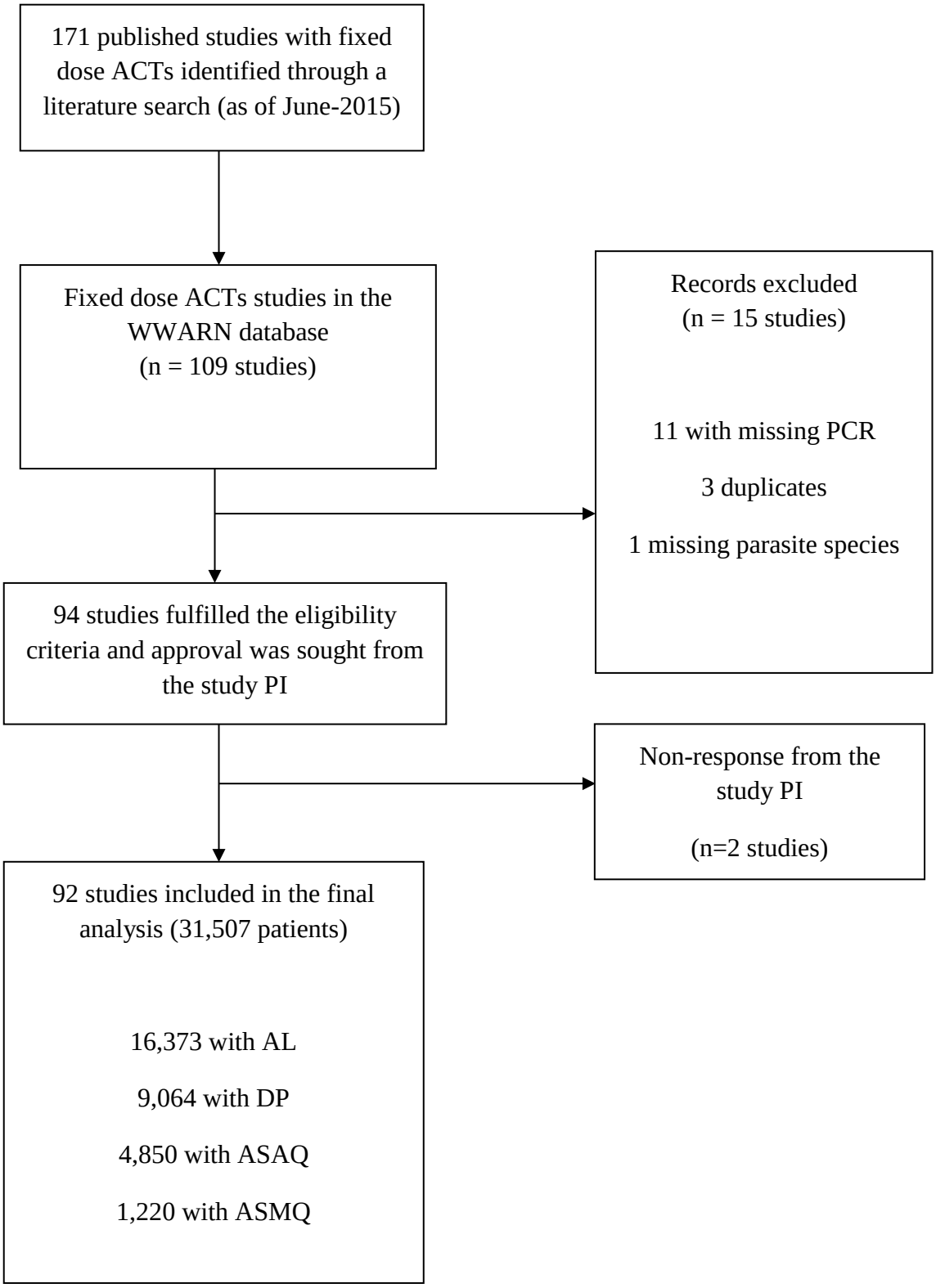


Figure 3.2: Study flow diagram

Legend: PCR = polymerase chain reaction; PI = principal investigator; AL = artemether-lumefantrine; DP = dihydroartemisinin-piperaquine; ASAQ = artesunate-amodiaquine; ASMQ = artesunate-mefloquine; ACT = artemisinin combination therapies.

3.3 Study description

Individual participant data were available from 80 published (n=25,702 patients) and 12 unpublished studies (n=5,033 patients), of which 67 (72.8%) were conducted in Africa, 23 (25.0%) in Asia, and 2 (2.2%) in South America. Figure 3.3 shows the location of sites, where the studies were conducted. The median sample size per treatment arm in the studies included was 160 [Interquartile Range (IQR): 107 - 236, Range: 13 – 1,844].

3.3.1 Design of the studies included

In total, 77 trials (84%) were randomised and the remaining studies were single-armed. The duration of study follow-up was 28 days in 48 studies (52%), 42 days in 32 studies (35%), 56 days in 3 studies (3%), and 63 days in 9 (10%) studies. In 71 studies (77%), treatment administration was fully supervised, 14 studies (15%) were partially supervised, 6 studies (7%) were unsupervised, and one study (1%) had both supervised and unsupervised arms. The characteristics of the studies included in this thesis are presented in Table 3.1.

3.3.2 PCR genotyping

In 43 studies (47%), three molecular markers in *msp-1, msp-2 and glurp* were used for differentiating recrudescence from new infection (Table 3.1). Two markers in *msp-1 and msp-2* were used in 28 (28%) studies, *msp-1, msp-2* along with microsatellites were used in 9 studies (10%), a single marker (either *msp-1* or *msp-2*) was used in 7 studies (8%), one study used *msp-1* and *glurp* (1%), and 3 studies (3%) used microsatellites (5, 6, or 7). In two studies (2%), the genotyping methodology used was not stated, and in one study (1%) no recurrence was observed.

3.3.3 Malaria endemicities in the study sites

The studies included in this thesis were conducted over a wide geographical span with different malaria endemicities from 1996 to 2014. The Malaria Atlas Project (MAP) has developed a metric based on a mathematical model to estimate the transmission intensity of a study site based on its latitude, longitude and the year the study was conducted [111]. The metric is called the *Plasmodium falciparum* parasite rate (PfPR) and is reported as an age standardised rate in 2 to 10 years old children ($Pf\mathrm{PR}_{2-10}$). The $Pf\mathrm{PR}_{2-10}$ estimate (referred to as the **MAP estimate** henceforth in this thesis) of malaria prevalence was obtained from the MAP researchers for the studies included in this thesis. Study sites were categorised as low transmission areas if the MAP estimates were less than or equal to 0.10, moderate if the estimates were >0.10 and ≤ 0.40, and high if the estimates were greater than 0.40. A total of 57 studies (62%) were carried out in the areas of low or moderate settings, 23 (25%) in the areas of high settings, and 12 multi-centre studies (13%) were conducted in areas of different transmission endemicities. The number of patients treated with different drug regimen in areas of different transmission profile is provided in Table 3.2.

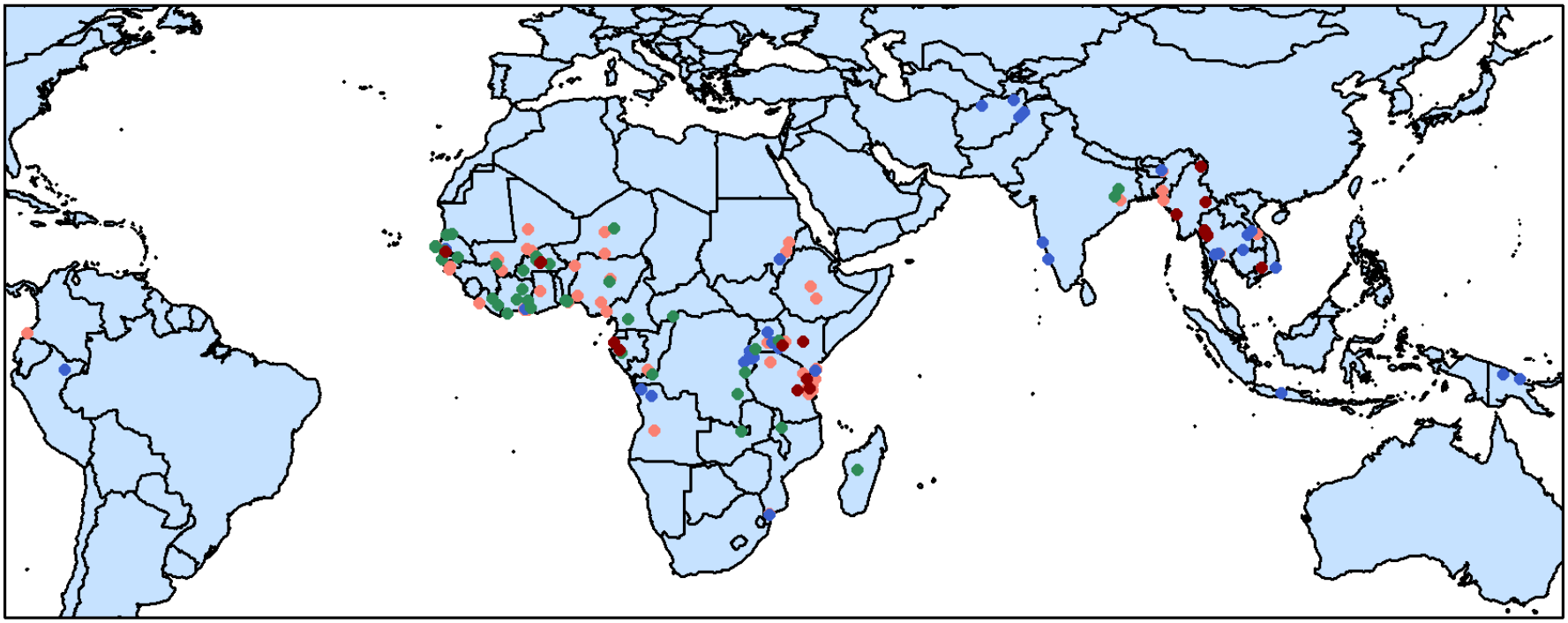


Figure 3.3: Study site locations for different drugs

Legend: •artesunate-mefloquine; •artemether-lumefantrine; •artesunate-amodiaquine; •dihydroartemsinin-piperaquine

Table 3.1: Characteristics of the studies included

WWARN ID	Study	N	Region	Year of enrolment	Age range (months)	Follow-up duration	Genotyping Markers	Treatment Supervision	Randomisation
ACHGC	Juma-2005*	184	Africa	2004—2005	72—708	28	*msp-1, msp-2*	Full	Non
ADXZX	Bouyou-Akotet-2010 [112]	41	Africa	2005—2005	12—72	28	*msp-1, msp-2*	Full	Yes
AJXCU	Menan-2011 [113]	206	Africa	2008—2009	24—744	28	*msp-1, msp-2*	Full	Yes
AKABP	Mutabingwa-2005 [114]	285	Africa	2002—2003	4—59	28	*msp-2*	Unsupervised	Yes
ATMFH	Tran-2012 [115]	55	Asia	2010—2011	120—780	42	*msp-1, msp-2, glurp*	Full	Yes
BQSYZ	Tine-2012 [116]	153	Africa	2010—2011	156—720	28	*msp-1, msp-2*	Full	Yes
CCEPC	Piola-2005 [117]	953	Africa	2002—2004	12—960	28	*msp-1, msp-2, glurp*	Full/Partial	Yes
CDCMJ	Smithuis-2010 [118]	400	Asia	2008—2009	12—792	63	*msp-1, msp-2, glurp*	Unsupervised	Yes
CDTEN	de Wit-2015*	283	Africa	2013—2014	0—252	42	*msp-1, msp-2*	Full	Yes
CRAXS	Thanh-2009 [119]	60	Asia	2006—2007	72—720	42	*msp-1, msp-2*	Full	Yes
CUDNY	Janssens-2007 [120]	182	Asia	2002—2003	24—780	63	*msp-1, msp-2, glurp*	Full	Yes
CXJYT	Falade-2005 [121]	310	Africa	2002—2003	2—119	28	*msp-1, msp-2, glurp*	Full	Non
DAABT	Laminou-2011*	159	Africa	2011—2011	2—168	28	*msp-1, msp-2, glurp*	Partial	Yes
DADPZ	Arinaitwe-2009 [101]	232	Africa	2007—2008	45—221	28	*6 microsatellites*	Partial	Yes
DAXCM	Sawa-2013 [122]	298	Africa	2009—2009	12—120	42	*msp-1, msp-2*	Full	Yes
DBCXT	Clark-2010 [123]	144	Africa	2004—2008	13—144	28	*msp-1, msp-2, 4 microsatellites*	Partial	Yes
DFUZB	Tarning-2012[124]	380	Africa	2007—2007	7—600	42	*msp-1, msp-2, 6 microsatellites*	Full	Non
DJYCQ	Zongo-2007a[125]	261	Africa	2005—2005	6—564	28	*msp-1, msp-2, 4 microsatellites*	Full	Yes
DTDDU	Faye-2007[126]	525	Africa	2002—2003	12—900	28	*msp-1*	Full	Yes
DYFKY	Awab-2010*	59	Asia	2007—2010	24—624	56	*msp-1, msp-2, glurp*	Full	Non
EDBXP	Karema-2006 [127]	252	Africa	2003—2004	12—60	28	*msp-1, msp-2*	Full	Yes
EDPJN	Abdulla-2008 [128]	896	Africa	2006—2007	0—149	42	*msp-1, msp-2, glurp*	Full	Yes
EFTTU	Vugt-1999 [129]	234	Asia	1996—1997	36—744	28	*msp-1, msp-2, glurp*	Full	Yes
EGYMA	Sagara-2012 [130]	680	Africa	2005—2005	6—560	28	*msp-1, msp-2*	Full	Yes

EPDUY	Ashley-2004 [131]	170	Asia	2003—2004	36—684	63	*msp-1, msp-2, glurp*	Full	Yes
FAJXQ	Anvikar-2012 [132]	202	Asia	2007—2007	7—720	28	*msp-1, msp-2, glurp*	Full	Yes
FEGFG	Valecha-2009 [133]	124	Asia	2007—2007	24—660	28	*msp-1, msp-2*	Full	Non
FEKFF	Agarwal-2013 [134]	272	Africa	2010—2011	6—59	42	*msp-1, msp-2, glurp*	Full	Yes
FFNAU	Grande-2007 [135]	252	S. America	2003—2005	60—708	63	*msp-1, msp-2, glurp*	Full	Yes
FSFKM	Kamugisha-2012 [136]	108	Africa	2010—2010	4—60	28	*msp-2*	Partial	Non
GARJK	Plucinski-2015 [137]	399	Africa	2013—2013	6—108	28	*7 microsatellites*	Partial	Non
GHNKU	Lefèvre-2001 [138]	160	Asia	1998—1998	144—852	28	*msp-1, msp-2, glurp*	Full	Yes
GPXJK	4ABC Study -2011 [139]	3,690	Africa	2007—2008	6—59	28	*msp-1, msp-2, glurp*	Full	Yes
GTEFH	Van den Broek-2005 [140]	121	Asia	2003—2003	30—792	42	*msp-1, msp-2, glurp*	Full	Yes
GZQDA	Ngasala-2011a [141]	359	Africa	2007—2008	4—60	56	*msp-1, msp-2, glurp*	Unsupervised	Yes
HCEMT	Carrasquilla-2012 [142]	159	S. America	2007—2008	144—672	42	*Not stated*	Full	Yes
HJNDX	Temu-2010*	102	Africa	2010—2011	7—58	28	*msp-1, msp-2, glurp*	Full	Non
HKNHR	Mayxay-2004 [143]	108	Asia	2002—2002	12—960	42	*msp-1, msp-2, glurp*	Full	Yes
HMPBZ	Adam-2010[144]	71	Africa	2009—2009	12—660	28	*msp-2*	Full	Yes
JFPER	Djallé-2014 [145]	128	Africa	2008—2009	2—59	28	*msp-1, msp-2, glurp*	Full	Yes
JGGNM	Bassat-2009 [146]	1,320	Africa	2005—2006	6—60	42	*msp-1, msp-2, glurp*	Full	Yes
JTXEY	Achan-2009 [147]	86	Africa	2007—2008	6—66	28	*msp-1, msp-2, 4 microsatellites*	Partial	Yes
JXZNZ	Nikiema-2010*	1,073	Africa	2000—2010	0—780	28	*msp-1, msp-2*	Partial	Yes
KJGJT	Sirima-2009 [148]	439	Africa	2004—2006	5—60	28	*msp-1, msp-2, glurp*	Full	Yes
KPSGN	Haque-2007 [149]	67	Asia	2005—2005	216—660	42	*msp-1, glurp*	Full	Non
KRBXE	Smithuis-2006 [150]	319	Asia	2003—2004	12—504	42	*msp-1, msp-2, glurp*	Unsupervised	Yes
KZBZT	Mens-2008 [151]	146	Africa	2007—2007	7—149	28	*msp-1, msp-2, glurp*	Full	Yes
MCZHT	Yeka-2008 [152]	414	Africa	2006—2007	6—120	42	*msp-1, msp-2, 4 microsatellites*	Full	Yes
MEFSC	Premji-2009 [153]	453	Africa	2006—2007	12—156	42	*msp-1, msp-2, glurp*	Full	Yes
MPYTS	Mayxay-2012 [154]	549	Asia	2010—2010	8—876	42	*msp-1, msp-2, glurp*	Full	Yes
MTZZP	Mayxay-2006[155]	107	Asia	2004—2004	16—600	42	*msp-1, msp-2, glurp*	Full	Yes
NBSAE	Sagara-2008 [156]	235	Africa	2004—2004	12—396	28	*msp-1, msp-2,*	Full	Yes

							1 microsatellite		
NJUCK	Menan-2012*	272	Africa	2012—2012	24—756	28	*msp-1, msp-2*	Full	Yes
PCPRC	Gadalla-2011 [157]	86	Africa	2006—2006	18—792	28	*msp-1, msp-2*	Partial	Non
PFJCR	Jullien-2010 [158]	26	Africa	2007—2008	216—720	28	*msp-1, msp-2*	Full	Yes
PKGFP	van den Broek-2006 [159]	106	Africa	2004—2004	12—59	28	*msp-1, msp-2, glurp*	Full	Yes
PUEKP	Juma-2008 [152]	267	Africa	2007—2007	6—62	28	*msp-1, msp-2*	Full	Yes
QRBRC	Espié-2012 [160]	299	Africa	2008—2009	6—59	42	*msp-1, msp-2, glurp*	Full	Yes
QXJGK	Valecha-2010 [161]	765	Asia	2005—2007	8—739	63	*msp-1, msp-2, glurp*	Full	Yes
QZJGM	Borrmann-2011 [162]	461	Africa	2005—2005	6—59	63	*msp-1, msp-2, glurp*	Full	Yes
QZMAG	Sutanto-2013 [163]	180	Asia	2008—2010	48—720	42	*msp-1, msp-2, glurp*	Full	Yes
RDBXS	Grivoyannis-2009*	334	Africa	2009—2009	12—732	28	*msp-1, msp-2*	Partial	Non
REAJS	Ndiaye-2011 [164]	354	Africa	2007—2008	10—780	28	*msp-1, msp-2*	Full	Yes
RMRNH	Sirima-2015*[165]†	927	Africa	2010—2013	6—60	63	*msp-1, msp-2, glurp*	Full	Yes
RNZFN	Faye-2010a [166]	161	Africa	2006—2007	84—792	28	*msp-1, msp-2*	Full	Yes
RSBPS	Toure-2009 [167]	60	Africa	2006—2006	18—600	28	*msp-1, msp-2*	Partial	Yes
SATNJ	Offianan-2011 [168]	242	Africa	2008—2008	6—59	28	*msp-1, msp-2*	Full	Yes
SBCEE	Yeka-2014 [169]	412	Africa	2008—2008	0—65	42	*msp-1, msp-2, glurp*	Full	Yes
SFBSG	Carrara-2009 [170]	197	Asia	2001—2002	36—840	63	*msp-1, msp-2, glurp*	Full	Yes
SUEGP	Zongo-2007b [171]	375	Africa	2006—2006	6—636	42	*msp-1, msp-2,* *4 microsatellites*	Full	Yes
SXGQP	Ursing-2011 [172]	187	Africa	2006—2008	4—175	42	*msp-1, msp-2, glurp*	Full	Yes
SYFQT	Martensson-2007 [173]	50	Africa	2004—2004	6—135	42	*msp-1, msp-2*	Partial	Yes
SZRDK	Ashley-2006 [174]	196	Asia	2004—2005	10—780	63	*msp-1, msp-2, glurp*	Full	Yes
TDFKY	Kamya-2007 [175]	421	Africa	2006—2006	6—108	42	*msp-1, msp-2,* *4 microsatellites*	Full	Yes
TRKFC	Laminou-2008*	192	Africa	2008—2008	5—192	28	*Not stated*	Full	Non
TYKSC	Ndiaye-2009 [176]	940	Africa	2006—2006	11—780	28	*msp-1, msp-2,* *1 microsatellite*	Full	Yes
UANQM	Karunajeewa-2008 [177]	170	Asia	2005—2007	7—62	42	*msp-1, msp-2, glurp*	Partial	Yes
UBTXH	Schramm-2013 [178]	293	Africa	2008—2009	9—69	42	*msp-1, msp-2, glurp*	Full	Yes

UGPAG	Bousema-2006 [179]	75	Africa	2004—2004	6—120	28	*msp-2*	Full	Yes
XEDNN	Yavo-2011 [180]	384	Africa	2006—2007	24—924	28	*msp-1, msp-2*	Full	Yes
XEKED	Faye-2010 [181]	318	Africa	2007—2008	9—96	28	*msp-1, msp-2*	Full	Yes
XTGNB	Hwang-2011 [182]	112	Africa	2009—2009	9—720	42	*msp-2*	Partial	Non
XXFCZ	Ngasala-2011b [183]	244	Africa	2007—2007	5—61	42	*msp-1, msp-2, glurp*	Unsupervised	Non
YGNQS	Sylla-2013 [184]	393	Africa	2010—2010	24—756	42	*msp-1, msp-2*	Full	Yes
YGTAH	Bukirwa-2006 [185]	204	Africa	2004—2005	12—108	28	*msp-1, msp-2*	Full	Yes
YPRHD	Price-2006 [186]	419	Asia	2000—2001	24—792	42	*msp-1, msp-2, glurp*	Full	Yes
YRRSK	Guthmann-2006 [187]	68	Africa	2004—2004	84—660	28	*msp-1, msp-2, glurp*	Full	Yes
YYDSM	Faucher-2009 [188]	185	Africa	2007—2007	6—70	42	*msp-1, msp-2*	Unsupervised	Yes
ZHKTN	Jones-2007*	111	Africa	2007—2007	1—58	28	*No parasitic recurrence*	Full	Yes
ZMNBX	Hien-2006*	1,943	Asia	2002—2004	24—936	56	*msp-1, msp-2, glurp*	Partial	Yes
ZUABH	Sinou-2009 [189]	13	Africa	2007—2007	204—468	28	*5 microsatellites*	Full	Non
ZYBXE	Martensson-2005 [190]	200	Africa	2002—2003	6—60	42	*msp-2*	Full	Yes

*Unpublished studies; †Was unpublished at the time of data acquisition
msp-1 = meroziote surface protein-1; *msp*-2 = meroziote surface protein-2; *glurp* = glutamate rich protein

Table 3.2: Patients treated with antimalarial regimens by transmission settings

Treatment	Low†	Moderate†	High†	Total
Artemether-lumefantrine	4,902	5,304	6,167	16,373
Artesunate-amodiaquine	1,108	1,287	2,455	4,850
Artesunate-mefloquine	796	358	66	1,220
Dihydroartemisinin-piperaquine	4,233	2,935	1,896	9,064
Total	11,039	9,884	10,584	31,507

†Defined based on the MAP estimates as described in Section 3.3.3.

3.4 Patient description

The overall median age was 4.1 years [Interquartile Range (IQR): 2.2 - 13.0] with 55.4% (n=17,469) of the patients being < 5 years old. Overall, 16,373 (52%) of the patients included in this thesis were treated with AL, 9,064 (29%) with DP, 4,850 (15%) with ASAQ, and 1,220 (4%) with ASMQ. The majority of the patients (77.2%) were from Africa, 21.5% from Asia, and 1.3% were from South America. Data for AL and ASAQ were mostly from Africa (86.0% and 93.8% respectively), whereas 66.7% of the patients with ASMQ regimen and 53.6% of the patients treated with DP regimen were from Africa. The median parasitaemia at presentation was 18,576/µL [IQR: 5,284 – 49,520] and remained similar between the different drug regimens (Table 3.3).

Table 3.3: Baseline characteristics of the patients included

Baseline characteristics	AL (1996- 2014)	ASAQ (2002-2014)	ASMQ (2002-2013)	DP (2002-2013)	Total (1996-2014)
N	16,373	4,850	1,220	9,064	31,507
Gender					
Female†	7,394 (45.2%)	2,253 (46.5%)	591 (48.4%)	3,464 (38.2%)	13,702 (43.5%)
Age (y) †					
Mean Age ± SD (y)	8.3 ± 10.66	7.0 ± 9.21	11.9 ± 13.37	13.5 ± 14.09	9.8 ± 11.94
<1 y	924 (5.6%)	312 (6.4%)	65 (5.3%)	386 (4.3%)	1,687 (5.4%)
1 to < 5 y	8,629 (52.7%)	2,895 (59.7%)	522 (42.8%)	3,736 (41.2%)	15,782 (50.1%)
5 to <12 y	3,306 (20.2%)	773 (15.9%)	201 (16.5%)	1,264 (13.9%)	5,544 (17.6%)
≥ 12 y	3,491 (21.3%)	862 (17.8%)	426 (34.9%)	3,665 (40.4%)	8,444 (26.8%)
Missing	23 (0.1%)	8 (0.2%)	6 (0.5%)	13 (0.1%)	50 (0.2%)
Continent†					
Africa	14,087 (86.0%)	4,550 (93.8%)	810 (66.4%)	4,862 (53.6%)	24,309 (77.2%)
Asia	2,127 (13.0%)	300 (6.2%)	410 (33.7%)	3,950 (43.6%)	6,787 (21.5%)
South America	159 (1.0%)	0 (0.0%)	0 (0.0%)	252 (2.8%)	411 (1.3%)
Enrolment clinical parameters					
Mean body weight ± SD (kg)	20.9 ± 16.29	20.0 ± 16.02	23.7 ± 17.2	22.4 ± 17.55	21.2 ± 16.63
Mean haemoglobin ± SD (g/dL)	10.2 ± 2.21	9.6 ± 1.98	10.4 ± 2.26	10.2 ± 2.31	10.1 ± 2.22
Median parasitaemia [IQR] (/µL)	19,674 [5,590 – 50,550]	21,730 [7,353 – 53,439]	23,386 [5,953 – 64,103]	14,320 [4,178 – 43,337]	18,576 [5,284 – 49,520]
Elevated temperature (>37.5°)	64.3% [9,705/15,101]	67.9% [3,221/4,747]	74.0% [743/1,004]	60.2% [4,200/6,972]	64.2% [17,869/27,824]

†Column percentages presented for gender and age in parenthesis. AL = artemether-lumefantrine; ASAQ = artesunate-amodiaquine; ASMQ = artesunate-mefloquine; DP = dihydroartemisinin-piperaquine; IQR = Interquartile range; SD = standard deviation; N = number of patients

3.5 Parasitological response

3.5.1 Early parasitological response

Rapid parasitic resolution post treatment administration is the hallmark of artemisinin and their derivatives [16]. The parasite positivity rate (PRR), which is the proportion of the patients with microscopically detectable persistent parasitaemia, is a metric used to measure the current status of parasitic sensitivity to the artemisinins (Chapter 1, Figure 1.5). The overall parasitological resolution in the studies included was rapid; 90.8% [95% confidence interval (CI): 88.4% - 93.2%] of the patients achieved parasitic resolution by day 2 and 98.7% [95% CI: 98.1% - 99.2%] by day 3 (Table 3.4). The distribution of day-3 PPRs for the four ACT regimens is presented in Figure 3.4.

Table 3.4: Early parasitological response in the studies included

	Day2†		Day3†	
Treatment	n/N	PPR [95% CI]	n/N	PPR [95% CI]
AL	1,298/12,872	10.1 [7.2 - 13]	200/13,118	1.5 [0.7 - 2.4]
ASAQ	286/4,385	6.5 [3.6 - 9.4]	34/4,453	0.8 [0.3 - 1.2]
ASMQ	123/985	12.5 [5.3 - 19.7]	17/964	1.8 [0.0 - 3.6]
DP	572/6,544	8.7 [4.7 - 12.8]	79/6,286	1.3 [0.4 - 2.1]
Overall	**2,279/24,786**	**9.2% [6.8 - 11.6]**	**330/24,821**	**1.3% [0.8 - 1.9]**

n = Number of patient with positive parasitaemia; N = Total number of patients with recorded parasitaemia status; PPR = parasite positivity rate; CI = confidence interval; AL = artemether-lumefantrine; ASAQ = artesunate-amodiaquine; DP = dihydroartemisinin-piperaquine; ASMQ = artesunate-mefloquine

†The overall PPRs were calculated as a weighted average of the estimates from each of the individual study sites and associated 95% CIs were calculated by adjusting for within study clustering using the method described in pages 441 - 445 of Fleiss, Levin and Paik (2003) [191]. The R script is openly available through my github [192].

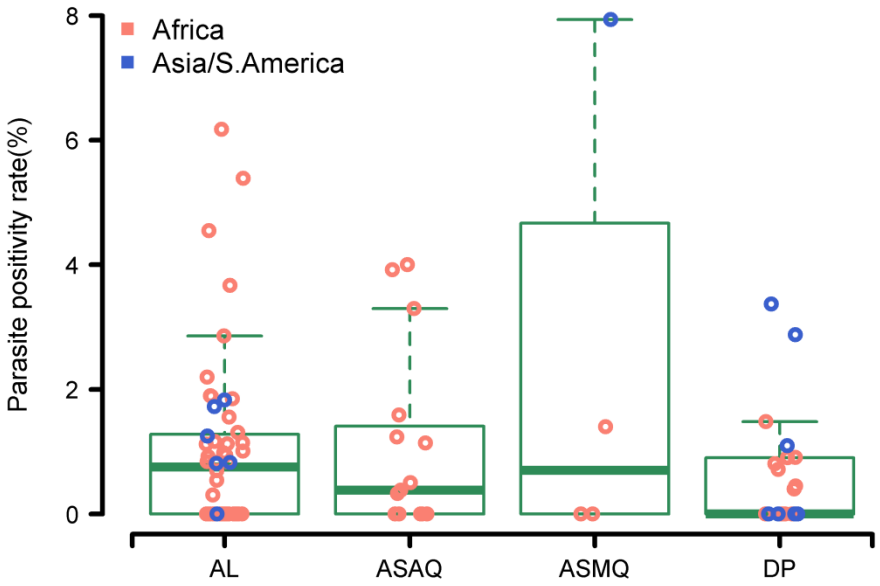


Figure 3.4: The distribution of day-3 parasitological response by treatment regimen

Legend: Parasite positivity rates (PPRs) on day 3 from study with sample size >50 patients. AL = artemether-lumefantrine; ASAQ = artesunate-amodiaquine; DP = dihydroartemisinin-piperaquine; ASMQ = artesunate-mefloquine. In one study conducted in Asia, early parasitological measurements weren't available.

3.5.2 Late parasitological response

After excluding 94 cases of early treatment failures, there were a total of 5,041 (16%) late parasitic recurrences of which 707 (14%, 707/5,041) couldn't be differentiated by the PCR genotyping. Of the remaining 4,334 recurrences, 596 (13.8%) were classified as recrudescence. The Kaplan-Meier estimate of efficacy was consistently greater than 90% in the studies included (Figure 3.5). One study with AL conducted in Tanzania had an efficacy lower than 90% on day 28 [173]. For the ASAQ regimen, one study in Myanmar [118] and one in Benin had efficacy below 90% on day 42 [188].

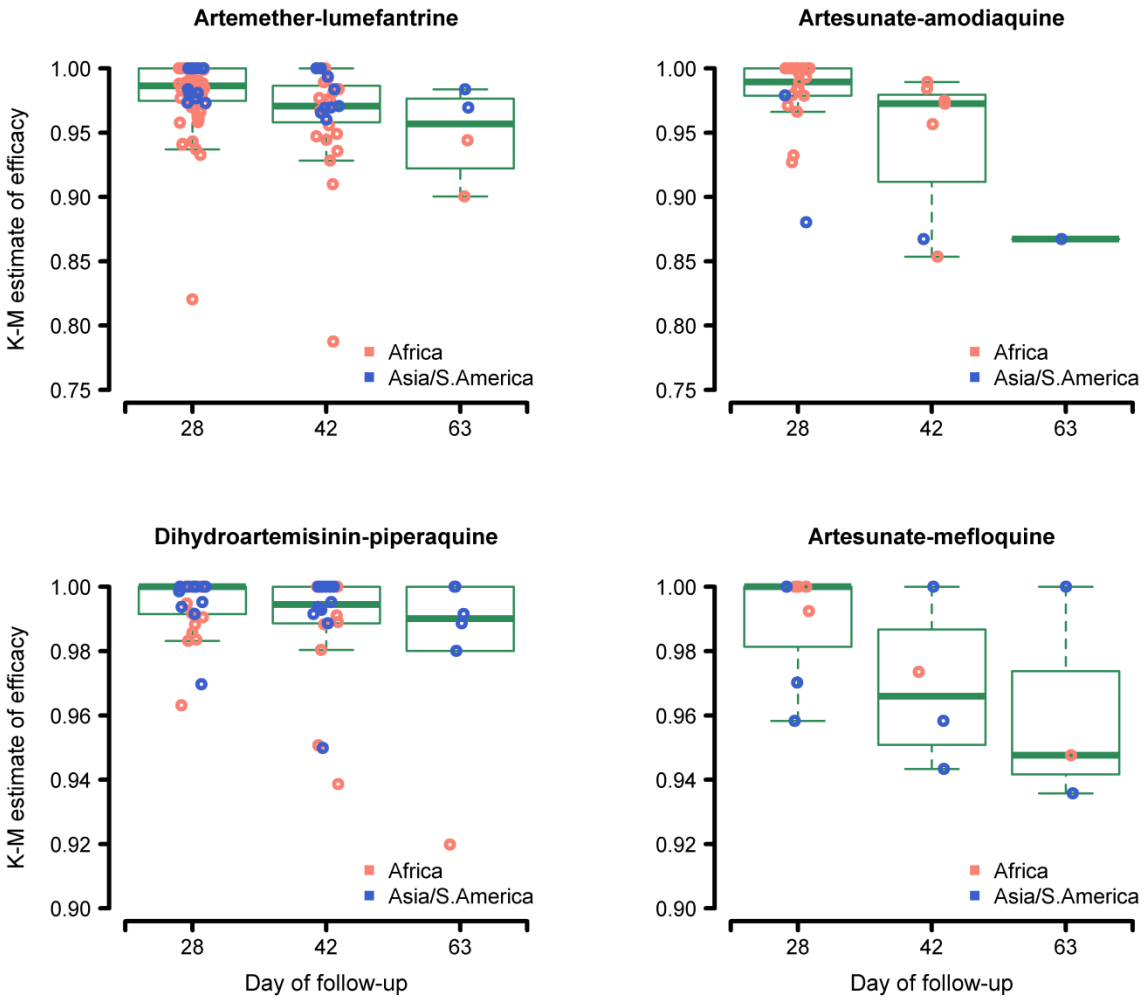


Figure 3.5: The Kaplan-Meier (K-M) estimate of PCR corrected drug efficacy by follow-up time and drug regimens

Legend: Y-axes ranges are different for top and bottom panels.

3.6 Concluding remarks

There is no overall methods chapter *per se* in this thesis as methodologies which are novel to antimalarial literature are explored in the next three chapters which will describe them in full detail. The proceeding chapters will refer back to this data chapter in terms of design and characteristics of the studies included, while providing additional details whenever needed.

Chapter 4: On the optimal follow-up duration for artemisinin combination therapies

4.1 Introduction

Prior to the routine use of PCR genotyping, a short follow-up duration of 14 days was used to assess antimalarial efficacy in areas of high transmission to mitigate confounding from new infections. However, retrospective reviews have highlighted that these studies had almost zero sensitivity in detecting recrudescent infections unless resistance had reached extremely high grades [58]. In the 1990s, the introduction of PCR genotyping paved the way to differentiate recrudescent from novel infections [48]. The 2009 WHO protocol, still in use, recommends that antimalarial drug efficacy should be defined based on PCR genotyping of the pre and post treatment isolates. The follow-up period is determined by the pharmacokinetic profile of the compound partnered with artemisinins or their derivatives. A minimum duration of 28 days is recommended for Artemisinin Combination Therapies (ACTs) with short acting partners drugs such as amodiaquine and lumefantrine, whereas at least 42 days is recommend for ACTs with partner drugs which are more slowly eliminated, such as piperaquine and mefloquine [64].

Clinical studies have shown that recrudescent parasitaemia can be detected in the peripheral blood after 42 days of treatment with artemether-lumefantrine [165,172], and up to 9 weeks after treatment with artesunate-mefloquine and dihydroartemisinin-piperaquine [120,161,165]. A seminal article published by Stepniewska and colleagues in 2004 highlighted that 28 days of follow-up would miss an estimated 36% of recrudescences for regimens rapidly eliminated and 60% for regimens slowly eliminated [77]. These estimates were based mainly on the results from the studies with non-ACT drugs, although the predictions concur with the field observations of ACT regimens. Mårtensson et al (2005) reported that in Tanzania a day 28

follow-up missed 28% (5/18) of recrudescences following artesunate-amodiaquine and 58% (7/12) of those following artemether-lumefantrine [78]. In Ethiopia, 88% (14/16) of the PCR confirmed recrudescences were detected after day 28 with artemether-lumefantrine (See Table 2 of [79]). These studies demonstrate that the current recommendations for the duration of follow-up in antimalarial efficacy studies are sub-optimal for capturing recrudescence and thus overestimate drug efficacy. However, treatment failures are uncommon with the ACTs thus limiting the generalisability of these results. A thorough investigation of the temporal distribution of recrudescent failures with the current ACTs is warranted. This chapter aimed to address this.

4.1.1 Aims and objectives

The overall aim of this chapter was to investigate the adequacy of the currently recommended follow-up period in capturing PCR-confirmed *P. falciparum* recrudescence for different ACTs. The specific objectives were:

I. To determine the proportion of recrudescences captured at standard follow-up time-points (days 28, 42, and 63) in ACT studies with follow-up duration longer than currently recommended,

II. To characterise the temporal trend (hazard function) and the probability distribution of recrudescence following ACT treatment, and from this determine the optimal duration of follow-up, and

III. To quantify the effect of the duration of follow-up on comparative efficacy.

The rest of the chapter is structured as follows: Section 4.2 describes a literature review carried out to identify studies with follow-up duration longer than currently recommended for the widely used ACTs and summarises the proportion of failures captured at each standard follow-up time-point (Objective I). Section 4.3 characterises the distribution of recrudescence using the studies described in Chapter 3 and uses this to gauge the optimal duration of follow-up (Objective II). Section 4.4 presents a simulation study based on the results presented in Section 4.3 to explore the influence of follow-up duration on comparative studies of artemether-lumefantrine and dihydroartemisinin-piperaquine (Objective III). Section 4.5 provides a critical discussion of the results and evaluates methodological strengths and weaknesses. The chapter is concluded in Section 4.6 with recommendations regarding the optimal follow-up duration for currently used ACTs.

4.2 Literature review on time of *P. falciparum* recrudescence

4.2.1 Search strategy

The WorldWide Antimalarial Resistance Network (WWARN) clinical trials publication library is a comprehensive resource detailing information on all published antimalarial trials since 1960 (with data up to Oct 2016) [193]. A search of the openly available WWARN publication library was carried out for eligible studies if they fulfilled the following criteria:

i) Prospective clinical efficacy studies of uncomplicated *P. falciparum* (alone or mixed infections with *P. vivax*),
ii) Study follow-up duration ≥ 42 days with artemether-lumefantrine (6-dose regimen) and artesunate-amodiaquine (fixed dose combination), and ≥ 63 days with artesunate-mefloquine (fixed dose combination) and dihydroartemisinin-piperaquine, and

iii) Polymerase Chain Reaction (PCR) genotyping undertaken to distinguish recrudescent parasites from new infections.

Studies were excluded if they enrolled patients with severe malaria, pregnant women, and patients with known HIV co-infection, healthy volunteers or travellers. Studies of prophylaxis or with passive follow-up after day 28 were also excluded. The number and timing of PCR confirmed parasitic recrudescences were extracted when these were reported in the trial publication at each of the standard time-points: days 7-28, 29-42, 43-63, or beyond.

4.2.2 Results of the literature search

4.2.2.1 Artemether-lumefantrine (AL)

A total of 55 studies with a follow-up duration ≥ 42 days were identified by the literature search (Table 4.1). The number of recrudescences could not be extracted from published data in 13 (24%) studies. In Africa, data were extracted from 27 studies with a total of 240 reported recrudescences of which 139 (58%) were detected by day 28, 90 (38%) between days 29 to 42, and 11 (5%) between days 43 to 63. In Asia and S. America, 15 studies were available with information on the time of recrudescence; there were a total of 38 recrudescent infections of which 22 (58%) were detected by day 28.

4.2.2.2 Dihydroartemisinin-piperaquine (DP)

Of the 10 available studies with an active follow-up duration ≥ 63 days, 8 were undertaken in Asia, one in Africa and one in S. America. A total of 39 recrudescences were reported in 8 publications in which the timing interval of failures could be extracted: 11 (28%) occurred by day 28, 18 (46%) between days 29-42 and 10 (26%) occurred after day 42 (Table 4.2).

4.2.2.3 Artesunate-amodiaquine (ASAQ)

A total of 7 studies with ASAQ had a follow-up duration ≥ 42 days from which data could be extracted. In the 6 African studies, a total of 15 recrudescences were reported of which 10 (67%) occurred by day 28 and 5 (33%) were detected between days 28 and day 42. In a single study in Asia, 10 (91%) of the 11 recrudescences were detected by day 28 (Table 4.3).

4.2.2.4 Artesunate-mefloquine (ASMQ)

Four eligible studies of the fixed dose ASMQ formulation were available; two in Africa and two in Asia. There were 10 recrudescences in Asia of which 9 (90%) were detected by day 42, whereas in Africa, 13 out of 19 recrudescences (68%) were detected by day 42 (Table 4.4).

Table 4.1: Reported timing of recrudescences in artemether-lumefantrine studies with follow-up duration ≥ 42 days

Study	Country	Year studied	Age range	Follow-up (days)	Recrudescence < 28 day	Recrudescence Days 29 to 42	Recrudescence Days 43 to 63
Abudulla-2008 [128]¶	Africa (multicentre)	2006-2007	<12 y	42	12	12	
Adjei-2014 [194]	Ghana	2010-2011	0.5–12 y	42	0	2	
Agarwal-2013 [134]	Kenya	2010-2011	0.5- 5 y	42	3	1	
Ashley-2007 [195]†	Thailand	2002	18-65 y	42	-	-	
Bassat-2009[146] ¶	Africa (multicentre)	2005-2006	0.5- 5 y	42	11	5	
Borrmann-2011[162]¶	Kenya	2005-2008	0.5- 5 y	84	3	12	5
Carrasquilla-2012[142] ¶	Colombia	2007-2008	>12 y	42	0	1	
Chandra-2015[196] †	Africa (multicentre)	2008-2010	5-12 y	42	-	-	
de Wit -2016[197]	DRC	2013-2014	0.5-5 y	42	0	1	
Denoeud-Ndam-2016 (not-SAM arm)[198]	Mali/Niger	2013-2015	<5 y	42	3	1	
Denoeud-Ndam-2016 (SAM arm)[198]	Mali/Niger	2013-2015	<5 y	42	0	0	
Eshetu-2012[79]	Ethiopia	2008-2009	all ages	42	2	14	
Espié-2012[160]	DRC	2008-2009	0.5-5 y	42	1	0	
Gobotosho-2011a[199] †	Nigeria	2009-2010	<12 y	42	-	-	
Gobotosho-2011b[200] †	Nigeria	2005-2010	0.5–12 y	42	-	-	
Haque-2007[149] ¶	Bangladesh	2005	≥18 y	42	0	2	
Hutagalung-2005[201]	Thailand	2001-2002	3-70 y	42	2	1	
Hwang-2011[182]	Ethiopia	2009	>0.5 y	42	0	0	
Kamya-2007 [175]†	Uganda	2006	0.5-10 y	42	-	-	
Karunajeewa-2008[177] ¶	PNG	2005-2007	0.5-5 y	42	0	1	
Laman-2014[202]	PNG	2011-2013	0.5–5 y	42	1	1	
Mayxay-2012[154] ¶	Lao PDR	2010	0.5 -73 y	42	14	2	
Michael-2010[203]	Nigeria†	-	1-11 y	42	-	-	
Mulenga-2006[204]	Zambia	2003-2005	15-50 y	45	13	9	
Nambozi-2011[205]	Zambia	2005-2006	0.5–5 y	42	6	0	
Ngasala-2011 a [141]	Tanzania	2007	<5y	42	11	3	

Ngasala-2011 b[183]	Tanzania	2007-2008	<5y	56	10	6	0
Nji-2015[206] †	Cameroon	2009-2013	0.5-10 y	42	-	-	
Ogouyèmi-Hounto-2016[207]	Benin	2014	0.5–5 y	42	0	0	
Ogutu-2014[208] †	Kenya	2010-2011	<5 y	42	-	-	
Okafor-2010[209] †	Nigeria	2005	0.5–10 y	42	-	-	
Premji-2009[153] ¶	Africa (multicentre)	2006-2007	1 -15y	42	3	3	
Rahman-2008 (DOT)[210]	Bangladesh	2006-2007	≥2 y	42	0	0	
Rahman-2008 (NDOT) [210]	Bangladesh	2006-2007	≥2 y	42	1	0	
Ratcliff-2007[211] ¶	Indonesia	2004-2005	1-60y	42	1	0	
Rujumba-2010	Uganda	2007-2008	<6y	42	2	0	
Saha-2012[212] †	India	2009-2010	all ages	42	-	-	
Saha-2013[213]	India	2010-2011	all ages	42	1	0	
Sawa-2013[122]	Kenya	2009	0.5–10 y	42	2	2	
Schramm-2013[178]	Liberia	2008-2009	0.5-5 y	42	2	4	
Sirima-2016[165] ¶	Africa (multicentre)	2010-2013	0.5-5 y	63	7	4	5
Smithuis-2010[118] ¶	Myanmar	2008-2009	>0.5 y	63	1	0	0
Sow-2016[214] †	Senegal	2010-2011	all ages	42	-	-	
Sowunmi-2007[215] †	Nigeria	2005-2006	≤10y	42	-	-	
Stohrer-2004[216]	Lao PDR	2003	all ages	42	0	3	
Sykes-2009[217]	Tanzania	2008	0.5-5 y	42	9	4	
Thriemer-2010[218] †	Bangladesh	-	8–65y	42	0	2	
Tinto-2014[219]	Burkina Faso	2008-2010	0.5-5 y	42	17	3	
Tiono-2015[220] †	Benin/Burkina Faso	-	Infants	42	0	0	
Toure-2016[221] †	Multicentre (Asia/Africa)	-	12–65y	42	3	2	
Tsefhu-2010[222]	Multicentre (Asia/Africa)	2007-2008	3–60 y	42	2	1	
Ursing-2011[172]	Guinea-Bissau	2006-2008	0.5-15y	70	6	0	1
Ursing-2016[223]	Guinea-Bissau	2012-2015	<15y	42	7	0	
van den Broek-2005 [140]	Bangladesh	2003	>1y	42	1	3	
Yeka-2008[152] ¶	Uganda	2006-2007	0.5-10y	42	5	4	
Yeka-2014[224]	Uganda	2008-2010	0.5-5 y	42	4	0	
Zongo-2007[171]	Burkina Faso	2006	>0.5y	42	-	-	

†Data couldn’t be extracted from the study publication

*The number of failures during the follow-up was estimated from the Kaplan-Meier (K-M) figure presented in the publication. The nine failures are conservative estimates in that each drop in K-M curve was counted as one failure. So, this gives at least 9 failures after day 28

¶From the individual participant outcomes available from the WWARN database

DRC = Democratic Republic of Congo; PNG = Papua New Guinea; Lao PDR = Lao People's Democratic Republic

Table 4.2: Reported timing of recrudescences in dihydroartemisinin-piperaquine studies with follow-up duration ≥ 63 days

Study [Reference]	Country	Year studied	Age range	Follow-up (days)	Recrudescence < 28 day	Recrudescence Days 29 to 42	Recrudescence Days 43 to 63
Ashley-2004 [131]	Thailand	2002-2003	1-65 y	63	3	2	0
Ashley-2005 [225]	Thailand	2003-2004	1 -65 y	63	2	1	1
Jansenns-2007 [120]	Cambodia	2002-2003	1 -65 y	64	0	1	0
Grande-2007 [135]	Peru	2003-2005	> 5 y	63	2	0	0
Valecha-2010 [161] ¶	Thailand/ Lao PDR/India	2005-2007	0.3 - 65 y	63	1	4	7
Smithuis-2010 [118] ¶	Myanmar	2008-2009	>0.5 y	63	0	1	0
Mayxay-2010 [226] †	Lao PDR	2005-2006	< 17 y	63	-	-	-
Borrmann-2011[162]	Kenya	2005-2008	0.5 - 5 y	84	3	9	2
Gargano-2012 [227] †	India	2006-2007	18 - 65 y	63	-	-	-
Amaratunga-2016 [228]†	Cambodia	2012-2013	2 – 65 y	63	-	-	-

†Data couldn't be extracted from the study publication

¶From the individual participant outcomes available from the WWARN database

Lao PDR = Lao People's Democratic Republic

Table 4.3: Reported timing of recrudescences in artesunate-amodiaquine studies with follow-up duration ≥ 42 days

Study [Reference]	Country	Year studied	Age range	Follow-up (days)	Recrudescence < 28 day	Recrudescence Days 29 to 42	Recrudescence Days 43 to 63
de Wit -2016 [197] ¶	DRC	2013-2014	0.5-5 y	42	4	1	
Espié-2012 [229] ¶	DRC	2008-2009	0.5-5 y	42	0	2	
Gbotosho-2011 [200]	Nigeria	2009-2010	≤12 y	42	0	0	
Nji-2015 [206] †	Cameroon	2009-2013	0.5-10 y	42	-	-	
Ratsimbasoa-2012 [230]	Madagascar	2008-2009	<5 y	42	3	0	
Schramm-2013 [178] ¶	Liberia	2008-2009	0.5-5 y	42	1	2	
Smithuis-2010 [118] ¶	Myanmar	2008-2009	>0.5 y	63	10	1	0
Yeka-2014 [224] ¶	Uganda	2008-2010	0.5-5 y	42	2	0	

†Data couldn't be extracted from the study publication

¶From the individual participant outcomes available from the WWARN database

DRC = Democratic Republic of Congo

Table 4.4: Reported timing of recrudescences in artesunate-mefloquine studies with follow-up duration ≥ 63 days

Study [Reference]	Country	Year studied	Age range	Follow-up (days)	Recrudescence < 28 day	Recrudescence Days 29 to 42	Recrudescence Days 43 to 63
Ashley-2006 [174] ¶	Thai-Myanmar border	2004-2005	0.5-65 y	63	5	4	1
Smithuis-2010 [118]	Myanmar	2008-2009	>0.5 y	63	0	0	0
Sirima-2016 [165] ¶	Burkina Faso/Kenya/Tanzania	2010-2013	0.5-5 y	63	3	6	6
Tietche-2010 [231]	Cameroon	2007-2009	0.7–7.9y	63	4	0	0

¶From the individual participant outcomes available from the WWARN database

4.2.3 Conclusions from the literature review

The key results from the review are summarised:

- Published trials suggested that the current recommended follow-up duration of 28 days missed 42.1% [95% CI: 36.0 - 48.4%] of failures after AL treatment in Africa and 42.1% [95% CI: 27.8% - 57.8%] in Asia.

- For ASAQ in Africa, the missed proportion with a 28 day follow-up was 46.7% [95% CI: 22.2% - 72.6%], whereas in a single Asian study 91% (10/11) of recrudescences were detected by day 28.

- Similarly, a day 42 follow-up missed 32% (6/19) of recrudescences following ASMQ in Africa and 32% (8/25) following DP in Asia.

In antimalarial studies, efficacy estimates are reported at standard time-points (days 28, 42 and 63) and hence it was not possible to extract the exact time of recrudescence based on the review data. Individual patient data described in Chapter 3 provided the exact time of recrudescence and this was used to construct the distribution of observed time-to-recrudescence in the next section.

4.3 Characterising distribution of *P. falciparum* recrudescence

4.3.1 Study and patient data

The inclusion criteria are outlined in Chapter 3 (Section 3.2.1). Briefly, fixed dose ACT studies in the WWARN repository were selected if therapeutic outcomes were defined from studies with a minimum follow-up period of 28 days and recrudescence confirmed using PCR genotyping. Early treatment failures (before day 7) were excluded for the ACTs under investigation because they represent either poor drug absorption or progression to severe malaria. Patients with missing or indeterminate PCR outcomes were excluded. In Africa, analysis was also restricted to children less than 5 years who constitute the main burden of the disease and a population with low levels of immunity. In the studies from Asia and S. America, patients of all ages were included.

4.3.2 Time of patient visits

All treatments were administered as 3 day regimens (days 0, 1 and 2). Patients were followed actively on days 3, 7 and weekly thereafter until the end of the follow-up. There were a total of 133,252 visits between days 4 to 63 from a total of 28,061 patients, of which 85.4% (113,768/133,252) took place on the scheduled days, 9.7% in the 3 days prior to the scheduled visit and 4.9% in the 3 days post the scheduled visit. In 128,599 (96.5%) visits where concomitant symptoms were documented, only 7,542 (5.8%) patients were symptomatic. At the time of recurrence, 63.7% (3,054/4,791) of patients were febrile (Figure 4.1); the geometric mean parasitaemia was 4,720/µL [95% CI: 4,355 - 5,115] among febrile patients and 3,081/µL [95% CI: 2,769 - 3,429] in afebrile patients. Patients with recurrent parasitaemia on non-scheduled days were more likely to have febrile illness (72%, 501/703) and have a higher geometric mean parasitaemia of 9,369/µL [95%CI: 8,000 - 10,972] than those attending on a

scheduled visit day who had a geometric mean parasitaemia of 3,124/µL [95% CI: 2,907-3,359].

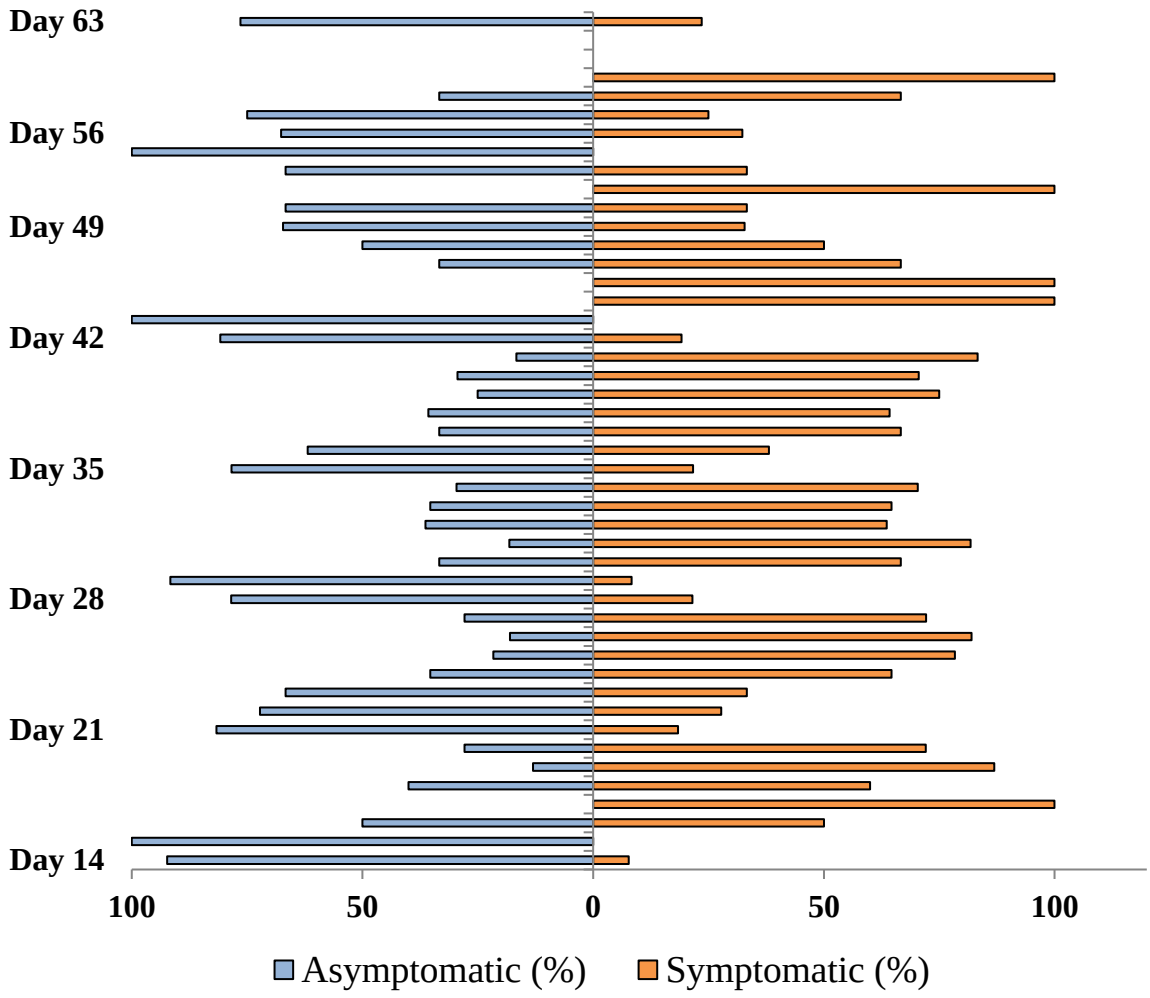


Figure 4.1: The symptomatic status in patients with recurrent parasitaemia (n=4,699)

Legend: The percentage of patients with documented status of febrile illness is given out of the total number of the recurrences recorded on each visit day. The recurrences detected on non-scheduled visit days were 21% more likely to be symptomatic compared to the recurrences detected on the scheduled visit days (adjusted odds ratio = 1.21 [95% CI: 0.97 – 1.50] after controlling for baseline parasitaemia, and age in a logistic regression model where study sites were fitted as random effects). A patient was defined as being symptomatic if the recorded body temperature was >37.5$^{o\,c}$. The data includes those with undifferentiated PCR outcomes.

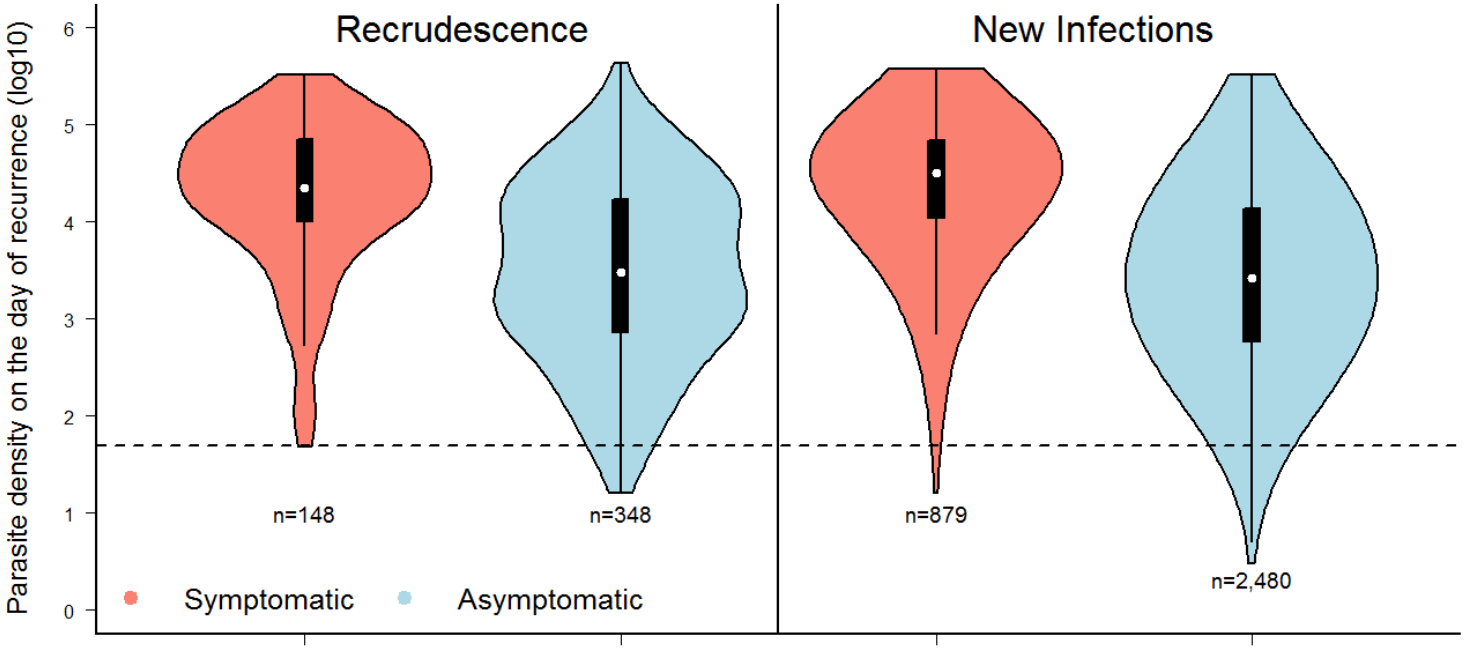


Figure 4.2: Parasite density at the time of recurrence in symptomatic and asymptomatic patients (n=4,699)

Legend: The parasitaemia (log) on the day of recurrence stratified by symptomatic status in patients where febrile status was available. The dotted horizontal line shows the limit of parasitic detection by microscopy (50 parasites/µL). A patient was classified as being symptomatic if the recorded body temperature was >37.5^{o} ^{c}. The data includes those with undifferentiated PCR outcomes.

4.3.3 The time of observed recrudescence

The observed times of recrudescences (t_{obs}) were confounded by active versus passive detection, since symptomatic patients with higher parasitaemia were more likely to seek treatment and be detected passively (Figure 4.1 and Figure 4.2). To control for this confounding factor, a metric was derived to estimate the time at which a recrudescent parasitaemia would exceed the threshold of detection (TOD); this was termed the time to recrudescence (T_R). Thick blood film examination counts parasites per 500 white blood cells, so the threshold for detection is approximately 50 parasites per microlitre. T_R was estimated assuming 10-fold parasite multiplication per 48 hour asexual cycle using the following algorithm:

$$\widehat{T_R} = t_{obs} - ((\log_{10}(\frac{P_R}{50})) \times 2)$$

where t_{obs} is the day when recrudescence was detected in the study, and P_R is the parasite density at the observed time of recrudescence. When the estimated T_R was smaller than the time of last visit when the subject had negative parasite counts on peripheral blood film (t_{neg}), then $\widehat{T_R}$ was replaced as $t_{neg} + 1$. In addition, the following three definitions of observed time to recrudescences were used for carrying out sensitivity analysis when estimating the distribution of recrudescences:

i) $\widehat{T_R} = t_{obs}$; assuming recrudescence occurred exactly on the day when the patient came to the clinic.

ii) $\widehat{T_R} = t_{neg} + 1$; assuming recrudescence occurred the day after the last visit, where the patient recorded negative parasitaemia on blood smear.

iii) $\widehat{T_R} = (\frac{t_{neg}+t_{obs}}{2})$; assuming recrudescence occurred half-way through the last visit with recorded negative smear and the day when it was recorded in the study.

4.3.4 Estimation of temporal trend of *P. falciparum* recrudescence

The temporal trend (instantaneous risk) of recrudescence at any time during the study follow-up was quantified by its hazard function, $h(t)$.

Unadjusted hazard function: The unadjusted hazard function was derived using the Kaplan-Meier method [102]. This assumes the hazard to be constant between successive failure times, and hence the hazard per unit time can be computed by dividing the number of events by the width of the time interval (τ_j). Let d_j be the number of failures occurring within a time interval τ_j and n_j be the number of people at risk at time *j* (called the risk-set), the Kaplan-Meier type hazard is given by Collet (2015) (See equation 2.13 in p.32 of [102]):

$$h_{KM}(t) = \frac{d_j}{n_j \tau_j}$$

The estimate of K-M type hazard is noisy and difficult to discern visible trends. Hence, smoothed estimates of the hazard function using kernel smoothing were computed.

Adjusted hazard function: The adjusted estimate of the hazard function was derived from a Cox proportional hazard model which controlled for the confounding effect of weight adjusted (mg/kg) dose of the partner drug, age and initial parasite load (on a log-scale). Standard software returns the cumulative baseline hazard ($H_0(t)$) instead of the instantaneous baseline hazard, $h_0(t)$. Mathematically, a baseline hazard function is given by the slope of the cumulative baseline hazard, $H_0(t)$, as:

$$h_o(t) = \frac{d}{dt}\big(H_0(t)\big)$$

$H_0(t)$ is a step function and hence not differentiable in strict sense. In order to resolve this, $H_0(t)$ was estimated using fractional polynomial smoothing after log-transformation as outlined by Royston-2011 [232]:

Let,

$$\ln\{H_0(t)\} = \boldsymbol{\alpha} \times f(\ln(time)) = z_0(t)$$

where $\boldsymbol{\alpha}$ is a vector of coefficients for different fractional polynomial transformations. Using the relationship between baseline hazard and cumulative baseline hazard:

$$h_0(t) = \frac{dH_0(t)}{dt}$$

$$\Rightarrow h_0(t) = \frac{dH_0(t)}{dt} \times \frac{d\ ln(H_0(t))}{d\ ln(t)} \times \frac{d\ln(t)}{d\ ln(H_0(t))}$$

$$\Rightarrow h_0(t) = \frac{d\ln(t)}{dt} \times \frac{d\ ln(H_0(t))}{d\ ln(t)} \times \frac{dH_0(t)}{d\ ln(H_0(t))}$$

Differentiating each term of the expression:

$$\frac{d\ln(t)}{dt} = \frac{1}{t}$$

$$\frac{d\ln(H_0(t))}{d\ln(t)} = \frac{dz_0(t)}{d\ln(t)}$$

$$\frac{dH_0(t)}{d\ln(H_0(t))} = \frac{1}{\left(\frac{d\ln(H_0(t))}{dH_0(t)}\right)} = \frac{1}{\frac{1}{H_0(t)}} = H_0(t) = \exp(z_0(t))$$

This gives an estimate of the smoothed baseline hazard:

$$h_0(t) = \frac{1}{t} \times \frac{dz_0(t)}{d\ln(t)} \times \exp(z_0(t))$$

Parametric models and flexible parametric models, which use restricted cubic splines to model the hazard function (Royston-Parmar models) [233], were also considered for comparisons.

4.3.5 Estimation of empirical probability distribution of *P. falciparum* recrudescence

The probability distribution of time to PCR confirmed recrudescence was derived using the following expression:

$$f_0(t) = h_0(t).S_0(t)$$

where $S_0(t)$ is the survival estimate of drug efficacy, and $f_0(t)$ is the probability density function of recrudescence evaluated at baseline hazard and baseline survival function. The following steps outline the construction of the failure time distribution:

i. The baseline hazard function, $h_0(t)$ was estimated as outlined in Section 4.3.4 using the fractional polynomial smoothing.

ii. Baseline survival function was estimated based on the cumulative baseline hazard, $H_0(t)$, estimated from the Cox model, as $S_0(t) = e^{-H_0(t)}$.

iii. The quantities estimated in steps (i) and (ii) were multiplied to obtain a non-normalised probability density of recrudescence. The right hand tail of the density was extrapolated beyond the observed time using a linear regression to estimate the time when the distribution reached 0.

iv. The estimated distribution was normalised so that area under the curve was equal to 1 and satisfied the following properties:

a. $f_0(t) \geq 0$ for every t

b. $\int_0^\infty f_0(t)dt = 1$ i.e. AUC under the curve = 1

c. $P(T \leq t) = \int_0^t f_0(t)dt$

The normalising constant (a) was estimated using a smoothing spline approach as an overall sum of the integrals. The normalised probability density is thus given as $\widetilde{f_0(t)} = \frac{1}{a} f_0(t)$.

4.3.6 Optimal duration of follow-up

The area under the normalised probability density of recrudescence (AUC) constructed in Section 4.3.5 was calculated at a given time-point. The cumulative AUC at a time-point t provides an estimate of the proportion of recrudescences captured by that time. The 95% confidence interval was based on 1,000 bootstrap samples drawn from the data of the same sample size and repeating the algorithm outlined in Section 4.3.5. For the DP regimen, the distribution was constructed for those who received total dose of piperaquine above or below 48 mg/kg, a threshold associated with poorer therapeutic outcomes in a paediatric population [34]. For each treatment regimen, the estimation of distribution was stratified by region (Africa and Asia/S. America), as there is a marked heterogeneity in transmission profiles between the regions.

4.3.7 Software

Kaplan-Meier type hazard and kernel-smoothed hazard function were estimated using the **muhaz** library, and the cumulative baseline hazard was estimated from fitting the Cox proportional hazard model using the **survival** library in R [234]. Fractional polynomial smoothing was carried out using the **fracpoly** command in Stata [235].

4.3.8 Results

4.3.8.1 Characteristics of the studies

Data were available from 60 African studies in 15,509 children aged less than 5 years and from 23 studies in 5,270 patients of all ages from Asia and South America. The duration of follow-up was 28 days in 39 studies (47%), 42 days in 32 studies (38%), 56 days in 3 studies (4%) and 63 days in 9 studies (11%). Further details on the study design and characteristics of the studies included in the analysis are provided in Chapter 3 (Table 3.1).

4.3.8.2 Parasitic recurrences

A total of 5,025 patients had recurrent parasitaemia between days 7 and 63. Three or more loci genotypes (including microsatellites) were attempted in 79% (3,954/5,025) of the cases, 2 loci in 19% (945/5025), one locus in 2% (114/5025), and markers were not documented in 0.2% (12/5025) samples. A total of 3,722 (74%) infections at the time of recurrence were identified as new infections and 596 (12%) as recrudescences; 707 (14%) paired isolates could not be genotyped and were classified as indeterminate. The proportion of recrudescences in studies using single markers were 14.9% (17/114) compared to 13.8% (130/945) in those using two markers, and 11.2% (444/3,954) in those using three or more molecular markers. The proportion of indeterminate outcomes were 27.2% (31/114), 12.3% (89/726), and 14.8% (587/3,954) in the studies which used one, two and three or more markers respectively.

Since the method of genotyping was a major confounding factor, the rest of the analysis was restricted to the studies using at least 3 molecular markers.

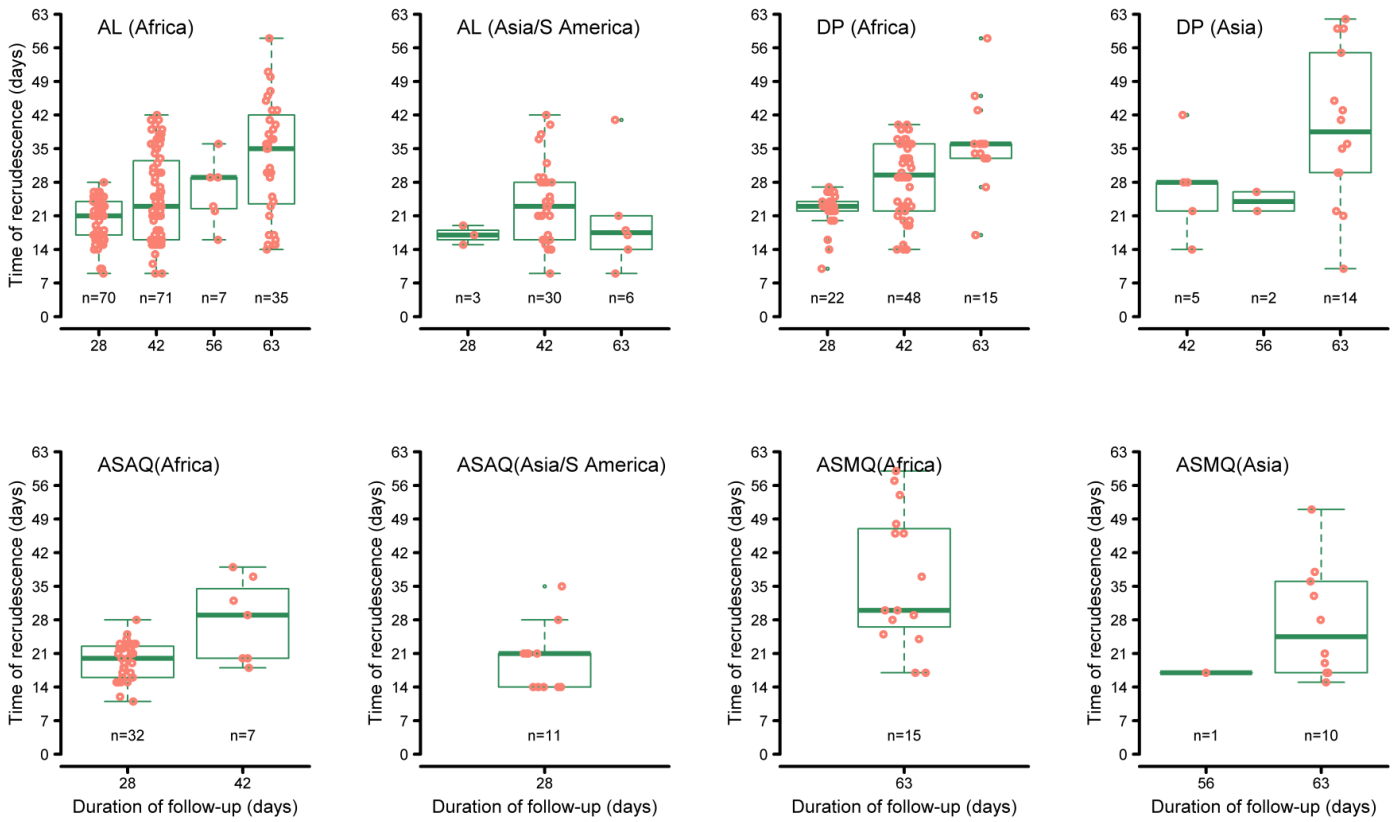


Figure 4.3: Distribution of observed time to recrudescence for ACTs stratified by region and duration of follow-up

Legend: The y-axis depicts the time for the recrudescent infection to reach the parasitic limit of detection (50 parasites/µL). In Africa, the distribution is shown for children <5 years age, whereas in Asia/S.America, data are shown for patients of all ages. Each dot represents an observed recrudescence. AL = artemether-lumefantrine; DP = dihydroartemisinin-piperaquine; ASMQ = artesunate-mefloquine; ASAQ = artesunate-amodiaquine, and n = number of recrudescences. Only data from studies with at least three molecular markers are shown in the graph.

4.3.8.3 The instantaneous hazard function of *P. falciparum* recrudescence

Unadjusted estimates: The unadjusted estimates of hazard function together with the kernel-smoothed estimates are presented in Appendices 4.1 and 4.2. The smoothed estimates of the hazard function exhibited a non-monotonic shape for AL and DP (Appendix 4.2).

Adjusted estimates: The adjusted estimates of hazard function (adjusted for age, mg/kg dose and baseline parasitaemia) derived from a Cox proportional hazard model could be derived only for AL (in Africa and Asia) and DP regimen (in Africa only) as the observed number of failures with ASAQ and ASMQ regimen were too small (Figure 4.3).

Like the unadjusted kernel-smoothed estimates, the baseline hazard function derived from fractional polynomial smoothing approach, exhibited an overall non-monotonic shape for both AL and DP regimen (Right panels; Figure 4.4).

For AL, the instantaneous hazard function was shifted to the right for Africa compared to Asia, and the peak hazard was much lower and earlier in Asia (day 21) compared to Africa (day 27) possibly reflecting the lower overall risk of recrudescence in Asia. For DP, the peak hazard was reached on day 31 in those who were under-dosed (<48 mg/kg of piperaquine) compared to day 39 in those who received ≥ 48 mg/kg of piperaquine dose (Figure 4.4).

The equations for the cumulative baseline hazard functions derived using the fractional polynomial smoothing are given (on log-scale) below.

AL (Africa):

$$\ln(CBH) = -3.680424 + 10.49305 \times \{\ln(t) - 3.427413427\} - 1.258464 \times \{\ln(t)^2 - 11.7471628\}$$

AL (Asia):

$$\ln(CBH) = -3.955812 + 2.812377 \times \{\ln(t)^2 - 12.29012325\} - 1.521957 \times \{\ln(t)^2 \times \ln(\ln(t)) - 15.41670573\}$$

DP (Africa overall):

$$\ln(CBH) = -4.339292 + 0.8416194 \times \{\ln(t)^3 - 37.4570925\} - 0.4869875 \times \{\ln(t)^3 \times \ln(\ln(t)) - 45.23813021\}$$

DP (Africa; (<48 mg/kg of piperaquine):

$$\ln(CBH) = -3.162359 + 0.9712495 \times \{\ln(t)^3 - 47.21631182\} - 0.5727575 \times \{\ln(t)^3 \times \ln(\ln(t)) - 60.66885952\}$$

DP (Africa; (≥48 mg/kg of piperaquine):

$$\ln(CBH) = -4.021096 + 0.5752936 \times \{\ln(t)^3 - 43.50176377\} - 0.3197093 \times \{\ln(t)^3 \times \ln(\ln(t)) - 54.70783962\}$$

CBH = Cumulative baseline hazard

Parametric models, which can accommodate non-monotone hazards (log-normal and log-logistic) were found to be poor fitting based on Cox-Snell's residuals (Appendix 4.3). The hazard functions derived from Royston-Parmar models, which utilises restricted cubic splines, were of similar shape to the ones obtained from fractional polynomial smoothing (Appendix 4.4). Since splines are mathematically complicated equations and difficult to utilise for simulation purposes, they weren't considered any further.

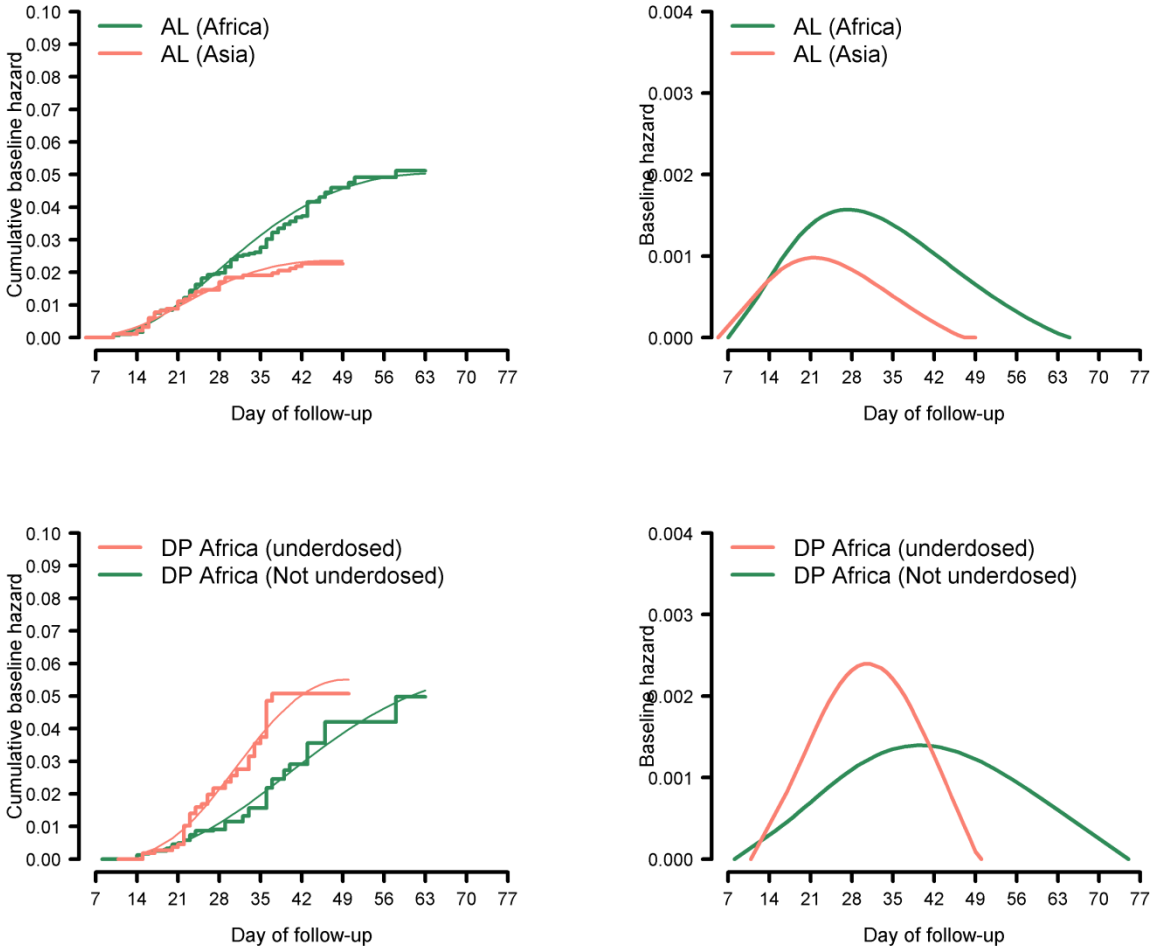


Figure 4.4: Fractional polynomial estimates of cumulative and instantaneous hazard for AL in Africa and Asia, and for DP regimen in Africa

Legend: The cumulative baseline hazard estimated from Cox model (adjusted for age, baseline parasitaemia and mg/kg dosage of partner drug) together with fractional polynomial smoother (left panel). The temporal trend of observing recrudescence during the follow-up period, estimated by the instantaneous baseline hazard function (right panel). Data in Africa was restricted to children <5 years, whereas patients of all ages were included in Asia and S. America. Data from S. America were grouped with Asia. All studies used at least 3 locus genotypes, the usage of microsatellites was considered as a separate locus. Under-dosed defined as total piperaquine dose ≤ 48 mg/kg (those receiving > 48 mg/kg defined as not under-dosed). AL = artemether-lumefantrine; DP = dihydroartemisinin-piperaquine.

4.3.8.4 The empirical distribution of time of *P. falciparum* recrudescence

The distribution of the time to recrudescent *P. falciparum* in patients treated with AL and DP regimen is presented in Figure 4.5. Using data from all the studies (regardless of the follow-up duration of the study), the area under the curve (AUC) for AL in Africa was 0.42 [95% CI: 0.26 – 0.53] on day 28, 0.79 [95% CI: 0.53 – 0.97] on day 42 and 1.00 [95% CI: 0.81 - 1.00] on day 63 (Table 4.5). In Asia, the distribution was shifted to the left compared to that observed in Africa, with an estimated AUC of 0.68 [95% CI: 0.55 – 0.85] on day 28 and 0.98 [95% CI: 0.93 – 1.00] on day 42. The distribution of recrudescence for DP could only be derived in Africa. The distribution of recrudescences shifted to the left in those who were under-dosed (receiving < 48 mg/kg); the AUC by day 42 was 0.91 [95% CI: 0.78 – 1.00] in the under-dosed group compared to 0.53 [95% CI: 0.19 – 0.90] in those who received ≥ 48 mg/kg (Table 4.5).

The duration of follow-up had an impact on the estimated distribution of recrudescence. When the analysis was restricted to day 42 (observations beyond this time-point were censored on day 42), the AL distribution in Africa was shifted to the left compared to one derived from the analysis which included data up to day 63 (solid lines, Figure 4.5). Similarly, the distribution of DP in those who received ≥ 48 mg/kg was also shifted to the left when the analysis was restricted to day 42 (Figure 4.5). The estimated AUCs from the analysis which was restricted to day 42 outcomes are presented in Appendix 4.5.

When different definitions were used for defining the observed time to recrudescence (as described in Section 4.3.3), the estimates of the AUC were altered, however, the overall conclusion remained unchanged (Appendix 4.5).

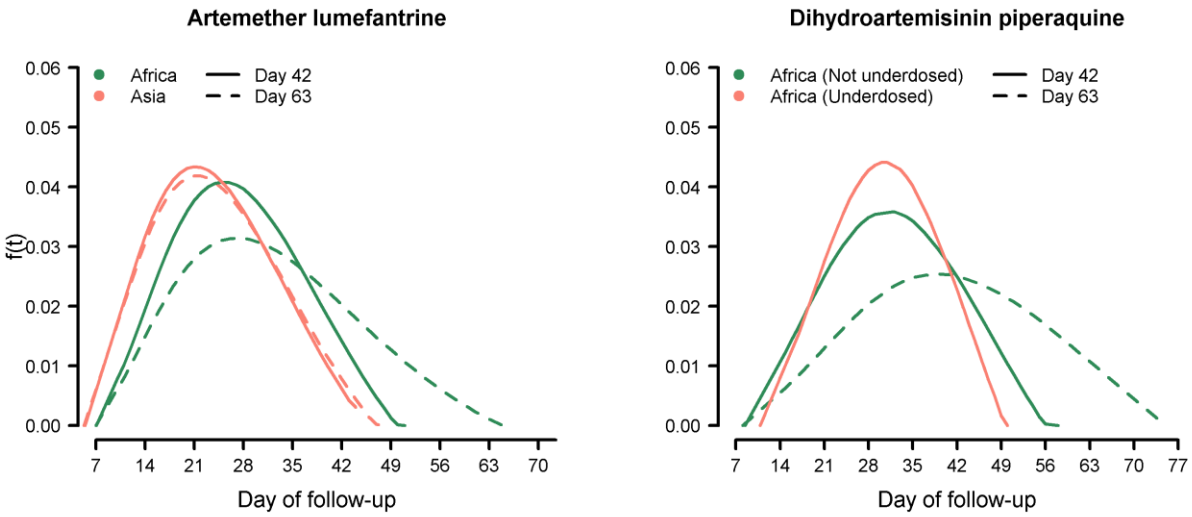


Figure 4.5: The probability distribution of *P. falciparum* recrudescence for AL in Africa and Asia, and for DP in Africa

Legend: Dotted line represents distribution derived from all data, regardless of the follow-up period. Solid line represents the distribution derived by including data up to day 42 (observations beyond day 42 were censored on day 42). For Africa, the data included children <5 years old whereas for Asia, no age restrictions were applied. AL = artemether-lumefantrine; DP = dihydroartemisinin-piperaquine.

Table 4.5: The area under the curve of the estimated probability distribution

Day	AL in Africa [a] [95% CI]	AL in Asia [a] [95% CI]	DP in Africa (PIP mg/kg <48) [95% CI]	DP in Africa (PIP mg/kg ≥48) [95% CI]
28	0.42 [0.26 – 0.53]	0.68 [0.55 - 0.85]	0.39 [0.25 - 0.52]	0.20 [0.00 – 0.36]
35	0.63 [0.40 – 0.78]	0.88 [0.78 - 0.99]	0.69 [0.53 – 0.83]	0.36 [0.03 – 0.59]
42	0.79 [0.53 – 0.97]	0.98 [0.93 - 1.00]	0.91 [0.78 – 1.00]	0.53 [0.10 – 0.81]
49	0.91 [0.64 – 1.00]	1.00 [0.93 - 1.00]	1.00 [0.93 – 1.00]	0.70 [0.22 – 0.99]
56	0.97 [0.74 – 1.00]	1.00 [0.94 - 1.00]	1.00 [0.98 – 1.00]	0.84 [0.36 – 1.00]
63	1.00 [0.81 – 1.00]	1.00	1.00	0.93 [0.47 – 1.00]

[a] Estimated in children <5 years in Africa and in patients of all ages from Asia.

The 95% confidence interval was estimated from 1,000 bootstrap samples of data for AL; for ASAQ this was based on 968 samples, for DP not under-dosed this was from 925 samples, and for DP under-dosed this was based on 987 bootstrap samples drawn from the original dataset. AL = artemether-lumefantrine; DP = dihydroartemisinin-piperaquine; PIP = piperaquine.

4.3.8.5 Conclusions from the estimated distribution of *P. falciparum* recrudescence

The key results derived from the analysis of the data in this section are:

- Constraining the duration of follow-up to 28 days following treatment with AL regimen missed an estimated 58% [95% CI: 47 – 74%] of all recrudescences in children less than 5 years in Africa, and 32% [95% CI: 15 – 45%] in patients of all ages in Asia and South America.

- For the DP regimen in children <5 years in Africa, at day 42, an estimated 47% of recrudescences were missed in those who received ≥48 mg/kg of piperaquine dose, whereas in those who were under dosed, almost all failures were observed by day 42.

- To capture the majority of recrudescent infections patients treated with AL should be followed-up for a minimum period of 42 days and those treated with DP should be followed-up for a minimum period of 63 days.

The translation of missed proportion of failures on to an absolute bias (overestimation) in derived efficacy estimates will be evaluated further in the next section, through simulations of standalone studies using AL and DP regimens. The choice of follow-up duration has implications on the comparative studies; especially the partner components in the ACT regimens which have different terminal elimination half-lives (Chapter 1, Table 1.1). In order to explore how the comparative efficacies are affected in drug regimens with different half-lives, a simulation study was carried out using the baseline hazard function for recrudescence presented in Section 4.3.8.3.

4.4 Simulation study

This simulation study had the following two aims:

- To quantify the absolute overestimation in the efficacy estimates for AL and DP with a day 28 and day 42 follow-up duration, respectively, and

- To investigate the impact of duration of follow-up on the relative risk estimates for drug regimens with different half-lives (using AL and DP as motivating example).

The followings steps define the simulation set-up:

4.4.1 Simulation set-up

- Time to recrudescence (t_{rc}) was generated using the following cumulative baseline hazard (CBH) presented in Section 4.3.8.3:

  $\ln(CBH(t)_{rc}) = -3.680424 + 10.49305 \times \{\ln(t) - 3.427413427\} - 1.258464 \times \{\ln(t)^2 - 11.7471628\}$ for artemether-lumefantrine

  $\ln(CBH\ (t)_{rc}) = -4.021096 + 0.5752936 \times \{\ln(t)^3 - 43.50176377\} - 0.3197093 \times \{\ln(t)^3 \times \ln(\ln(t)) - 54.70783962\}$ for dihydroartemisinin-piperaquine

These functions approximately assigned 4% of the observations to be recrudescences for both AL and DP regimen.

- Simulate time to new infection (t_{ni}) using:

$$\ln(CBH(t)_{ni}) = \alpha_0 + 9501.2150 \times \{\ln(t)^{-2} - 0.0858\} - 31651.33 \times \{\ln(t)^{-2} \times \ln(\ln(t) - 0.1054\} + 29340.83 \times \{lnt^{-2} \times \ln(lnt)^2 - 0.1294\} - 12690.51 \times \{\ln(t)^{-2} \times \ln(\ln(t))^3 - 0.1588\}$$ for artemether-lumefantrine

The parameter α_0 was varied to achieve the desired proportion of new infections.

$\alpha_0 = -4.190984$ for approximately 15% new infections

$\alpha_0 = -3.290984$ for approximately 30% new infections

$$\ln(CBH(t)_{ni}) = \beta_0 + 9501.2150 \times \{\ln(t)^{-2} - 0.0858\} - 31651.33 \times \{\ln(t)^{-2} \times \ln(\ln(t) - 0.1054\} + 29340.83 \times \{lnt^{-2} \times \ln(lnt)^2 - 0.1294\} - 12690.51 \times \{\ln(t)^{-2} \times \ln(\ln(t))^3 - 0.1588\}$$ for dihydroartemisinin-piperaquine

The parameter β_0 was varied to achieve the desired proportion of new infections.

$\beta_0 = -4.800422$ for approximately <10% new infections

$\beta_0 = -4.190422$ for approximately 15% new infections

The respective survival and baseline hazard function used for simulation time to new infection are presented in Appendix 4.6.

- Since early recurrences are unlikely, the minimum time was set to day 14 and administrative censoring was applied on the last scheduled follow-up visit (day 63). For simplicity, no losses to follow-up were assumed.

- For each individual, the observed time (t) was defined as the minimum of the simulated time to recrudescence and new infection. The observed event corresponded to the event with minimum time. The final observed time was rounded to the nearest weekly visit day (7, 14, 21 and so on), reflective of the antimalarial follow-up design.

- Kaplan-Meier estimates for AL and DP were computed on days 28, 42 and 63 and stored. Cox's regression model was fitted and the estimate of hazards ratio (HR) for AL relative to DP on days 28, 35, 42, 56, and 63 were stored.

- This was repeated 1,000 times. The median (interquartile range) of overestimation in efficacy with a day 28 and day 42 follow-up (relative to day 63 estimates) were reported, and the median (interquartile range) of the estimated HRs for AL relative to DP were reported for a simulation study with the following sample sizes: n=200, 500, and 1,000 subjects/arm.

4.4.2 Results

Overestimation of AL and DP efficacy: Compared to the day 63 estimates, a day 28 follow-up period with AL regimen overestimated the drug efficacy by a median of 2.8% [IQR: 2.3 – 3.4%; Range: 0.6 – 5.5%] in the areas of high transmission settings. With a day 42 follow-up, the overestimation was 0.9% [IQR: 0.7-1.3%; Range: 0 - 2.9%] (Figure 4.6). For the DP regimen, a day 42 follow-up duration overestimated the drug efficacy by 2.0% [IQR: 1.7 – 2.6%; Range: 0.0 - 4.6%] in the areas of high settings. The estimates from the areas of low settings were similar (Figure 4.6).

Relative risk estimates: From 1,000 simulation runs, when the observed proportion of new infections were low (8% in the DP arm and 15% in the AL arm) the median hazard ratio (HR) for AL relative to DP was 1.85 [IQR: 1.33 - 3.01] on day 28, and this reduced to 1.30 [IQR: 1.07 - 1.68] on day 42 and fell to parity from day 56 onwards (Figure 4.7A). When the observed proportion of new infection was high (30% in AL arm and 15% in DP arm), similar results were obtained (Figure 4.7B). There was little difference in the results when the simulation was repeated with n=200 and n=1,000 subjects per treatment arm (Appendix 4.6).

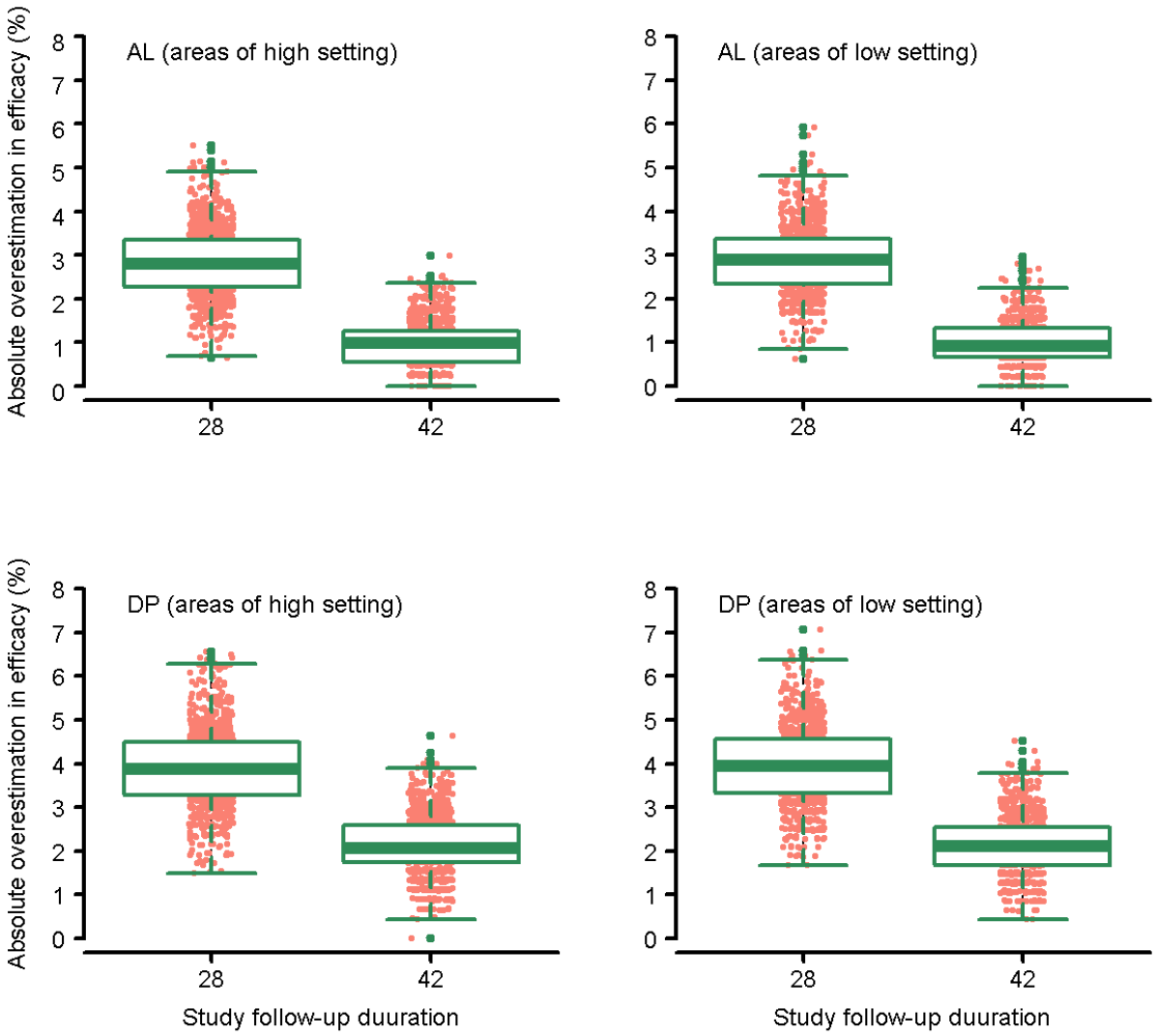


Figure 4.6: Absolute overestimation in efficacy due to sub-optimal follow-up duration (compared to day 63 estimates)

Legend: Overestimation of efficacy relative to day 63 estimates for AL and DP regimen. The survival curves were generated using the distribution of recrudescence presented in Figure 4.5 (using data up to day 63). For DP, the distribution in those who received adequate dose (≥48 mg/kg piperaquine) was used. In the high settings, there was a median of 15% new infection in the DP arm and 30% new infection in AL arm. In the low settings, there was a median of 8% new infection in the DP arm and 15% in the AL arm. Simulation assumed n=500 patients per treatment arm and was repeated for 1,000 runs. AL = artemether-lumefantrine; DP = dihydroartemisinin-piperaquine.

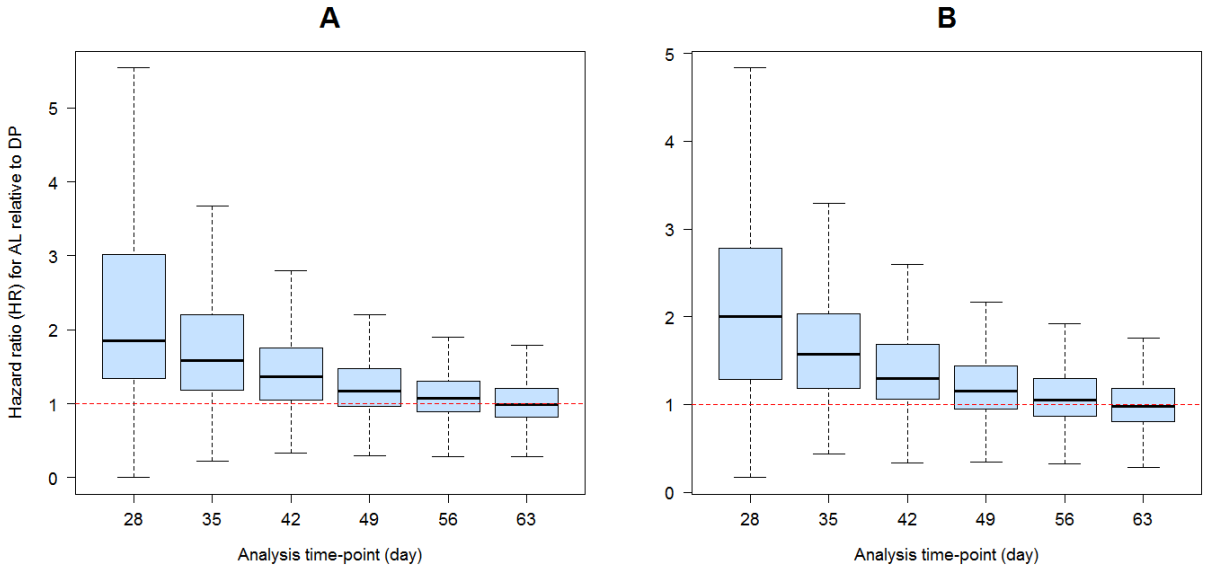


Figure 4.7: Simulation study comparing AL against DP in children <5 years in Africa

Legend: **Panel A**: There were a median (across 1,000 simulation runs) of 8% new infection in DP arm and 15% in AL arm representing areas of low transmission. **Panel B**: There were a median of 15% new infection in DP arm and 30% in AL arm representing areas of high transmission. In both simulation settings, approximately 4% recrudescence was observed by day 63 on AL and DP arm. Data for AL was simulated based on the estimated hazard function of recrudescence in children <5 years Africa, and for DP regimen the data was simulated based on the hazard function for those who received piperaquine dose greater than 48 mg/kg (Section 4.3.8.3, Figure 4.5). The dotted horizontal line shows the line of no effect (hazard ratio = 1). The simulation assumed n=500 patients per treatment arm. AL = artemether-lumefantrine; DP = dihydroartemisinin-piperaquine.

4.4.3 Conclusions of the simulation study

- For AL regimen, a day 28 follow-up overestimated drug efficacy by a median of 2.8% in absolute terms and a day 42 follow-up overestimated efficacy for the DP regimen by a median of 2.0% (relative to day 63 estimates).

- A day 28 comparison of AL against DP is biased against AL as the hazard ratio remained approximately 2-fold higher for the AL arm and this progressively fell to parity only from day 49 onwards.

4.5 Discussion

The distribution of parasite recrudescence following antimalarial treatment is critical for the mathematical and epidemiological modelling of malaria and has important implications for interpretation by both clinicians and public health policy makers. The accuracy of these models relies on realistic assumptions regarding these distributions. The distribution of the natural history of *P. falciparum* is well described by Weibull and Gompertz distributions, using data from malariatherapy studies of patients treated for neurosyphilis under carefully controlled conditions [236–239]. However, the results obtained from untreated infections cannot be generalised to estimate the duration of *P. falciparum* parasitaemia in patients being treated for symptomatic illness. Furthermore, the malariatherapy trials were conducted in the era of monotherapy before the introduction of ACTs and the distribution of *P. falciparum* recrudescence with these regimens is yet to be characterised. In order to explore the distribution of *P. falciparum* recrudescence, data from 92 antimalarial clinical trials conducted in Asia, Africa and Latin America were explored.

First, the temporal trend of *P. falciparum* recrudescence within the study follow-up period was characterised by quantifying the hazard function, which exhibited a positively skewed non-monotonic shape. A study by Fontanet and Walker (1993) found that parasitic recrudescence in patients treated with mefloquine monotherapy was best described with an exponential distribution [67]. This corresponds to a constant underlying hazard function, i.e. the risk of recrudescence remains constant for the duration of follow-up. In contrast, the observations of non-monotonic hazard functions have a strong biological plausibility due to the known pharmacokinetic profile of the drugs. Initially, the blood drug concentrations are above the minimum inhibitory concentration (MIC) and during this period parasitic growth is suppressed [240]. When the drug concentrations fall below the MIC, the sub-microscopic parasite population begins to re-expand, eventually becoming patent and causing symptomatic illness.

During this period, the instantaneous risk of recrudescence remains high. After the sub-microscopic parasite population has emerged, the hazard of recrudescence falls again thus exhibiting the observed non-monotonic shape.

The characterisation of the temporal trend enabled the construction of the probability distribution of the recrudescent infections. This was carried out in children less than 5 years in Africa, a sub-group where the greatest burden of the disease lies, and in populations of all ages in Asia and S. America. As expected, in children less than 5 years in Africa treated with DP, the distribution of recrudescence was shifted to the left in patients who were under-dosed (defined as those receiving < 48 mg/kg of piperaquine dose) compared to those who were adequately dosed (≥ 48 mg/kg). This can be explained by drug concentrations in those receiving a lower dose falling below the MIC earlier, thus allowing the parasite biomass to begin to expand earlier. In the same age group in Africa, the distribution for AL was shifted to the left compared to that of DP with the peak risk being observed on day 27 for AL and day 39 for children receiving an adequate DP dose. This reflects the pharmacokinetic profiles of these drugs; lumefantrine has a much shorter half-life of 3 to 5 days [241,242] and thus concentration of the drug falls below the MIC much earlier than piperaquine (elimination half-life of 20 to 30 days) [124,243–245].

There were significant regional differences in the AL distribution, which was shifted to the left in Asia compared to Africa. This is likely to be a consequence of acquired immunity which is greater in African children who have multiple infections per year and acquire immunity much faster than the patients in the low endemic setting of Asia. [246]. The derived distribution of AL suggested that recrudescent parasites could manifest as late as 9 weeks after AL treatment, long after lumefantrine had fallen to sub-therapeutic concentrations. These findings need to be interpreted with caution. The derived model for AL in Africa was highly vulnerable to 9 recrudescent parasites observed between days 42 to 63 enrolled in two African studies (Kenya

and Burkina Faso) [162,165]. When the analysis was restricted to day 42 outcome, there was a clear left shift in the distribution of recrudescence for AL in Africa (compared to when data up to day 63 was used), whereas the distribution for AL in Asia wasn't affected by the follow-up duration (Figure 4.5).

The primary aim of this chapter was to explore the optimal duration of follow-up required to capture parasite recrudescence following the ACT regimens. The area under the curve (AUC) of the constructed distribution was used to gauge the optimal duration of follow-up for determining antimalarial clinical efficacy (Figure 4.5). A follow-up period of 28 days following AL treatment missed an estimated 58% [95% CI: 47% – 74%] of the recrudescences in Africa and 32% [95% CI: 15% – 45%] in Asia, whereas for DP regimen, a 42 day study missed 9% [95% CI: 0% – 22%] of recrudescences in those who under-dosed and 47% [95% CI: 19 – 90%] in those who were not under-dosed. These estimates are similar to those reported by Martensson et al (2008) for AL regimen [247], but higher than Stepniewska et al (2004) [77] for drugs with moderate and long half-lives. Taken together, these analyses provide clear evidence that a minimum follow-up duration of 28 days for AL and 42 days for DP will result in systematic overestimation of antimalarial drug efficacy and underestimation of treatment failures. In the simulation scenarios explored in this chapter, the overestimation reached as high as 5.5% with a day 28 follow-up for AL and 4.6% with a day 42 follow-up duration with DP in high transmission areas (compared to day 63 estimates) (Figure 4.6).

The duration of follow-up affects not only the efficacy estimates derived in standalone clinical trials, but also the interpretation and conclusions drawn from comparative drug studies. AL and DP are the most widely used antimalarial regimens in the world and are frequently used as a reference arm against which a new antimalarial regimen is compared. Since the partner components of the ACTs have very different terminal elimination half-lives and thus differential suppression of the parasite, the results of the comparative trials are vulnerable to

the duration of follow-up. In a simulation study designed to assess the impact of duration of follow-up on the relative risk estimates between drug regimens AL and DP, the hazard ratio (HR) for AL compared to DP was highest on day 28, and fell to parity only from day 49 onwards. Studies with short duration of follow-up thus favour DP even if both are equally effective. Comparison of any drug regimens with different pharmacokinetic profiles should be interpreted in the light of the potential confounding of the duration of follow-up.

The results presented in this chapter have important public health implications. Recrudescent failures are associated with increased morbidity and are the primary driving factors for the selection and ongoing transmission of *de novo* resistant parasites [56]. There is a period of time ("window of selection") in which resistant parasites are most likely to be selected and propagated [248]. For the AL regimen, this window lies between days 23.8 to 27 for a *de novo* resistance to occur and between 19.8 to 38.7 days for an acquired resistance [248]. The corresponding figures for mefloquine (a drug with a long elimination half-life like piperaquine) are estimated to be between 72.9 to 86.6 days, and 64.8 to 112.3 days respectively for a *de novo* and an acquired resistance [248]. Hence, the current recommended follow-up duration period for AL and DP doesn't fully cover this window where resistant parasites might emerge. It is important that antimalarial researchers and policy makers are aware of this potential pitfall associated with sub-optimal follow-up with these regimens.

4.5.1 Limitations

This analysis has several limitations. First, few of the studies used in the analysis had a duration of follow-up greater than 42 days and the small number of events beyond this period is likely to have had a major effect on the tail of the derived distribution. When the analysis was restricted to studies with 42 days of follow-up, there were some differences in the derived estimates, but the overall conclusion that the current recommended follow-up duration remained sub-optimal for capturing late recrudescences was unchanged. Second,

recrudescence was defined by molecular analysis of the pre and post treatment parasite isolate and it was assumed that the genotyping approach used for outcome classification was consistent despite obvious differences in methodology. To control for this heterogeneity, analysis was limited to studies which used at least three molecular markers. Although this provides better genotyping accuracy, the current methodology remains vulnerable to the population allele frequency of the loci assessed and also the confounding of polyclonal infections. In areas of intense transmission, often the case in Africa, parasite infections are frequently polyclonal, current guidelines recommend a conservative approach to interpreting these infections as recrudescences which inherently overestimates recrudescence [51]. Greenhouse et al (2007) reported that in areas of high transmission as many as 45% of reported recrudescences may have been misclassified and were actually new infections [249]. In the current analysis no information on the multiplicity of infection was available, and hence it was impossible to discern the degree to which the 9 late recrudescences reported following AL may have been confounded and misclassified [162]. This has profound implications for the tail of the temporal distribution particularly following AL in Africa, where new infection is common. Accounting for this error is beyond the scope of this thesis, but emphasises that the tail of the distribution following AL in Africa should be treated with caution.

Finally, the estimation of the hazard and the probability density function is mathematically less stable and vulnerable to the method used for estimation, as highlighted by the different estimates obtained from the Kaplan-Meier type method, the kernel smoothed estimator, the parametric and flexible parametric approach, and the fractional polynomial smoother (Appendix 4.1, 4.2, 4.3, 4.4). Despite these differences, however, all the methods agree on the overall shape of the function.

4.6 Conclusions

In conclusion, this work is the first attempt to characterise the probability density of *P. falciparum* recrudescence following treatment with the ACTs. Despite its limitations, the results from both the literature review and individual patient analysis indicate that the current recommended duration of follow-up for ACT regimens does not capture all recrudescent parasitaemia, and fails to cover the period during which acquired resistance is likely to occur. As such, current study designs have poor sensitivity to detect early signals of resistance, which initially manifests as late recrudescences. The results presented in this chapter suggest that a 42 day duration should be the standard length of follow-up for *in vivo* trials in uncomplicated *P. falciparum* malaria for partner drugs with elimination half-lives comparable to lumefantrine, and 63 days for slowly eliminated compounds such as piperaquine or mefloquine.

Chapter 5: Competing risk events in antimalarial studies of uncomplicated *P. falciparum* malaria

5.1 Introduction

A competing risk is an event such that it's occurrence precludes the occurrence of the primary endpoint of interest or fundamentally alters the probability of observing it [80]. The conceptual foundation of competing risk analysis was laid by Daniel Bernoulli in 1760s when he quantified the effect of smallpox inoculation on the gain in life expectancy by considering each individual as exposed to two risks: death from smallpox, and death from other causes [250,251]. Death from other causes prevented death from smallpox being observed thus constituting a competing risk scenario. Largely ignored by the clinical communities for much of the 20th century, competing risk survival analysis has gathered considerable attention in the clinical and statistical literature in the past decade [80–82,252–255]. However, the application of the competing risk approach in antimalarial research remains novel. In a drug trial of uncomplicated *P. falciparum* malaria, a new infection with *P. falciparum* or *P. vivax* can potentially preclude the occurrence of the primary endpoint of interest (*P. falciparum* recrudescence) during the ensuing weeks, thereby constituting a competing risk event.

The standard approach for dealing with a competing risk event of a new infection in an antimalarial trial is to censor them in survival analysis, routinely performed using the Kaplan-Meier (K-M) method, log-rank test and Cox proportional hazard model. An implicit assumption underlying these analyses is of independent censoring - those who are censored are assumed to have the same risk of developing the event of interest as those who are still being followed-up. However, censoring due to new infection contradicts this assumption because a patient with a new infection (or any recurrence) is given additional antimalarial at this

censoring time. These patients are no longer at subsequent risk of recrudescence as the parasites are exposed to another therapeutic concentration of antimalarials with the second round of treatment. Statistical methods analogous to the K-M method, log-rank test and Cox regression model have been developed to draw inferences in the presence of competing risk events and will be briefly described.

5.1.1 Estimation of failure probabilities

The K-M method is currently recommended by the WHO for deriving antimalarial drug efficacy. The K-M estimate ($\hat{S}_{KM}(t)$) of drug efficacy at time *t* is given by:

$$\hat{S}_{KM}(t) = \prod_{\tau \leq t} \frac{n_\tau - d_\tau}{n_\tau}$$

where d_τ is the number of recrudescences at time τ, n_τ is the total number of patients at risk of experiencing recrudescence just before time τ (the risk-set), and the symbol Π presents the product over all distinct times when recrudescences or losses to follow-up were observed up to time *t*. It is common to report an estimate of cumulative probability of recrudescence at the end of the study. This is usually calculated as the complement of the K-M estimate of drug efficacy [$\hat{F}_{KM}(t) = 1 - \hat{S}_{KM}(t)$]. However, in the presence of a competing risk event, the ($1 - \hat{S}_{KM}(t)$) approach has been suggested to be incorrect [256] and may lead to an overestimation of failure [80] (See also Figure 2.3, Chapter 3).

The **Cumulative Incidence Function (CIF)** estimator proposed by Kalbfleisch and Prentice (p. 255 of [257]) provides the probability of failure by appropriately accounting for competing risk events. The CIF at time *t* for an event *i* is given by:

$$\hat{F}_{CIF\,i}(t) = \sum_{\tau \le t} \frac{d_{i\tau}}{n_\tau} . \widehat{S_{KM}}(\tau - 1)$$

where $d_{i\tau}$ is the number of events of type i that occur at time τ, and n_τ is the number of patients at risk at time τ, and $\hat{S}_{KM}(\tau - 1)$ is the K-M estimator of the survival probability of being free of any event type prior to time τ.

5.1.2 Comparing treatment regimens

In survival analysis, inferences can be drawn on either of the two epidemiological quantities of fundamental interest: rates or risks [258]. The rates are characterised by the hazard function, which is an instantaneous probability of observing the event of interest at a given time-point *t* among those individuals who are still in the risk-set. Risk is a cumulative measure and is defined as the cumulative probability of the event of interest accrued over the time period under consideration. In standard survival analysis (where competing risk events are assumed not to occur), there is a simple one-to-one correspondence between the hazard function (rate) and the cumulative failure function (risk). This means that if the effect of an exposure variable is to increase the instantaneous hazard rate of the event of interest *k*, then its effect on the cumulative risk of the event *k* is also the same. This has implications for comparative studies.

Comparative antimalarial studies use the log-rank test for comparing the equality of the survival curves for recrudescence between two drugs at the end of the study follow-up. The log-rank test is essentially the comparison of the underlying hazard rates between two groups as noted by Bajorunaite and Klein (2007) [259] and Dignam and Kocherginsky (2008) [256]. The one-to-one correspondence between rate and risk means that the log-rank test for

comparing the equality of hazard rates is also equivalent to comparing the equality of survivals. A corollary of this statement is that any inference drawn upon the hazard function holds equivalently true for the survival function and the cumulative incidence. However, in the presence of competing risk events, the one-to-one relationship between the (cause-specific) hazard function and the corresponding cumulative failure function no longer holds true [82], and the effect of the exposure variable on the hazard rate for event *k* may not coincide with its effect on the risk of event *k*.

An alternative approach, which compares the difference between two exposure groups accounting for the presence of competing risk events, is **Gray's *k*-sample test** [260]. This is the usual log-rank test where the cause-specific hazard function is replaced by the hazard of the sub-distribution (explained in Section 5.1.3.2 to follow) [261]. In the absence of competing risk events, the result of a Gray's *k*-sample test will be identical to that derived from a log-rank test [256].

5.1.3 Regression modelling

Regression models for time-to-event data are based on the specification of the underlying hazard function. In the presence of competing risk events, two different hazard functions are defined: the cause-specific hazard, and the sub-distribution hazard.

5.1.3.1 Cause-specific hazard

The cause-specific hazard defines the instantaneous risk of an event occurring per unit time for the specific cause of the event, among subjects without any prior events. This is the probability of failure due to cause *k* (for example, recrudescence) at time *t*, given that no failures of any kind (neither recrudescence nor new infection) have occurred thus far. The cause-specific hazard at a given time point for an event type can be computed by dividing the number of individuals experiencing that event type by the total number of individuals who were in the

risk-set immediately prior to the time *t*. Individuals are removed from the risk-set if they experience either of the two events or those who are lost to follow-up.

5.1.3.2 The sub-distribution hazard

The sub-distribution hazard function, introduced by Fine and Gray [262], is the probability of observing an event *k* in the next time interval, given that no cause *k* failures (neither recrudescence nor new infection) have occurred thus far. The sub-distribution hazard at a given time-point for an event type can be computed by dividing the number of individuals experiencing the event type (e.g. recrudescence) by the total number individuals who were in the risk-set immediately prior to the time *t*, while still maintaining individuals who experienced competing events in the risk-set. This is the main difference between cause-specific hazard and the sub-distribution hazard approach (Figure 5.1). Maintaining subjects with a new infection in the risk-set allows them to act as a place holder that represents the proportion of the risk-set that can never experience the primary outcome [263]. Like with the cause-specific hazard function, those who are lost to follow-up are removed from the risk-set.

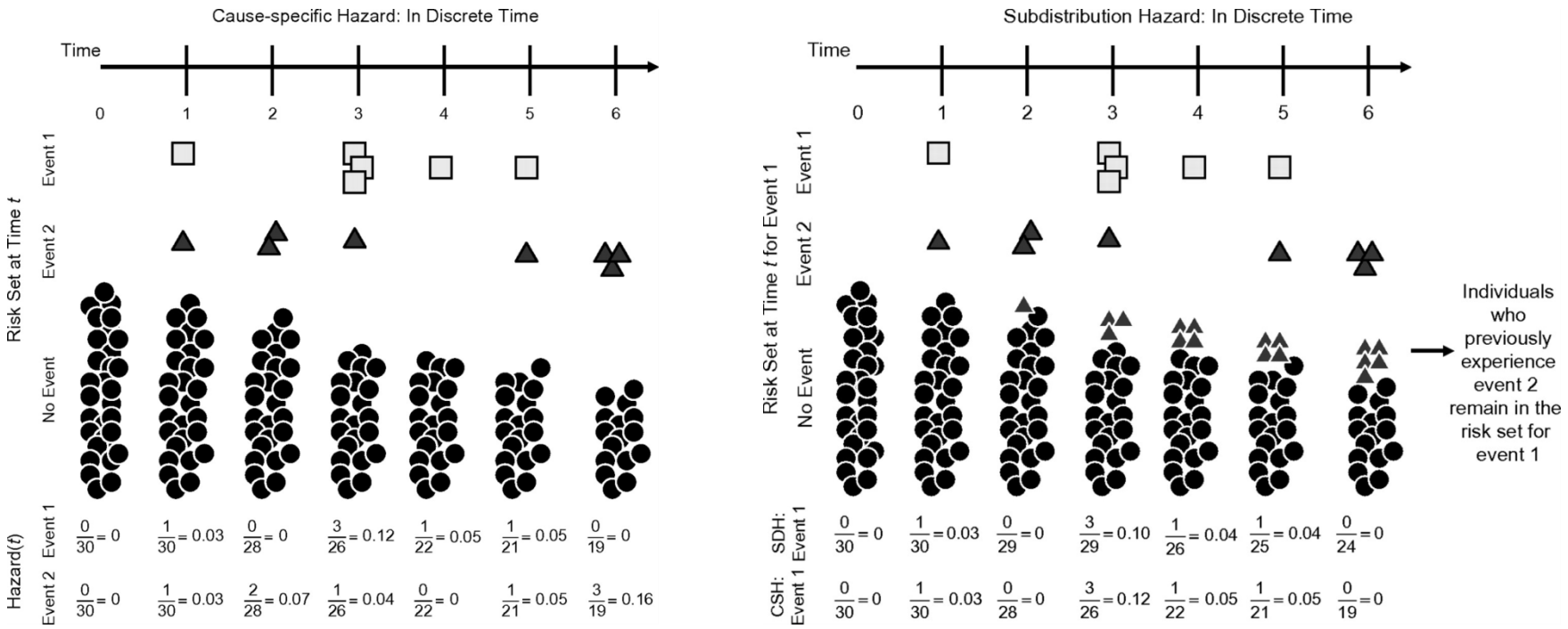


Figure 5.1: Calculation of cause-specific and sub-distribution hazard function. Reproduced with permission from Lau et al (2009) [263]**.**

Legend: Left panel (cause-specific hazard); right panel (sub-distribution hazard). The risk-set has 30 individuals to begin with (solid circles) at time t=0. The primary endpoint of interest is event 1 (squares) and event 2 (triangles) is a competing risk event. For calculating cause-specific hazard for event 1, individuals who experienced event 2 are removed from the risk-set. However, for calculating the sub-distribution hazard function for event 1, individuals who experienced event 2 are still maintained in the risk-set as triangles (right panel). The calculation of the two hazard functions are provided at the bottom of each graph. SDH = sub-distribution hazard; CSH = cause-specific hazard.

5.1.3.3 Regression models in the presence of competing risk events

In the presence of competing risk events, regression modelling can be carried out either on the cause-specific hazard function or the sub-distribution hazard function.

Let there be two events (j), where $j = 1$ indicates recrudescence (the primary event of interest) and $j = 2$ indicates a new infection (a competing risk event). The cause-specific hazard model for an event j can be fitted by using the Cox proportional hazard model and censoring the competing risk event. This is referred to as the cause-specific Cox proportional hazard model. This is expressed as:

$$h_j(t|\boldsymbol{x}) = h_{0j}(t).\exp(\boldsymbol{x}'\beta_j) \qquad j = 1,2$$

where β_j is the vector of the coefficients of the set of explanatory variables $\boldsymbol{x}$, and $h_{0j}(t)$ is the cause-specific baseline hazard function for cause j at time t. The value $\exp(\beta_j)$ is interpreted as the cause-specific hazard ratio (csHR) for the jth event corresponding to a unit change in the value of the corresponding covariate. This model assumes that the hazard remains proportional which means the relative hazard for two groups remains constant throughout the time-period under study. The cause-specific Cox model can be equivalently written in terms of the baseline survival function ($S_0(t)$) for event type j as:

$$S_j(t|\boldsymbol{x}) = S_{0j}(t)^{\exp(\boldsymbol{x}'\beta_j)} \qquad j = 1,2$$

The cause-specific Cox model estimates the relative effect of the covariates on the cause-specific hazard function. This denotes the relative change in the *instantaneous hazard rate* of occurrence of the event of interest in subjects who are currently event-free. In the absence of competing risk events, a one-to-one correspondence is maintained between the rate and risk, thus there is a direct correspondence between the effect of the covariate on the hazard of the outcome and the effect of the covariate on the cumulative incidence (risk) of the outcome.

However, in the presence of competing risk events, the association of a covariate with an increased (or decreased) cause-specific hazard rate doesn't necessarily translate to an increased (or decreased) incidence of the event of interest [82]. This is because one must account for the effect of a covariate on the cause-specific hazard function of each of the competing risk events in order to fully determine the effect of that covariate on the incidence of the event of interest [264]. Hence, cause-specific regression modelling for the event of interest j is limited when the interest is in drawing inference on the cumulative incidence of that event [264].

In the competing risk framework, the influence of covariates on the cumulative incidence of the event of interest can be explored by the sub-distribution hazard model (also called the CIF regression model) proposed by Fine and Gray [262]. The proportional sub-distribution hazard model can be written in a similar way to the Cox proportional hazard model by replacing the cause-specific baseline hazard with the baseline sub-distribution hazard:

$$\lambda_j(t|\boldsymbol{x}) = \lambda_{0j}(t).\exp(\boldsymbol{x}'\varphi_j) \qquad j = 1,2$$

where φ_j is the vector of the coefficients of the set of explanatory variables $\boldsymbol{x}$, and $\lambda_{0j}(t)$ is the baseline sub-distribution hazard function for cause j at time t. The value $\exp(\varphi_j)$ is interpreted as the sub-distribution hazard ratio (sdHR) for a given covariate on the jth event. This model makes an explicit link between the sub-distribution hazard function and allows estimation of the effect of a covariate on the incidence of the event of interest. Following Austin and Fine (2017) [264], Fine and Gray's sub-distribution hazard model can also be written in terms of cumulative incidence function (CIF) as:

$$CIF_j(t|\boldsymbol{x}) = 1 - \left(1 - CIF_{0j}(t)\right)^{\exp(\boldsymbol{x}'\varphi_j)} \qquad j = 1,2$$

where $CIF_{0j}(t)$ is the baseline CIF for an event j.

5.1.4 Aims and objectives

The overarching aim of this chapter was to quantify the influence of competing risk events when drawing inferences from antimalarial trials.

There were three specific objectives:

I. To quantify the magnitude of overestimation in cumulative failure derived by the Kaplan-Meier method compared to the Cumulative Incidence Function in a single-armed antimalarial trial.

II. To quantify the influence of competing risk events on the comparative efficacy between antimalarial drugs, and

III. To compare the predicted risk of recrudescence derived from the cause-specific Cox model and Fine and Gray's sub-distribution hazard model.

The rest of the chapter is structured as follows: Section 5.2 describes the data used and provides results of the statistical analyses performed using the data included in this thesis; Section 5.3 describes two simulation studies designed to explore how competing risk events affect the derived inferences in a range of hypothetical scenarios; and Section 5.4 provides a critical discussion of the results obtained from data analysis and simulation studies. The chapter is concluded with recommendations in Section 5.5.

In this thesis, the terminology "competing risk event" is used to refer to a new *P. falciparum* infection which can potentially prevent *P. falciparum* recrudescence from being observed (unless otherwise stated) and the word "failure" implies parasitic recrudescence as defined by PCR genotyping.

5.2 Data analysis

5.2.1 Study data

The inclusion and exclusion criteria for this analysis have been outlined in Chapter 3 (Section 3.2.1). Briefly, all studies in the WWARN repository which applied molecular genotyping to distinguish recrudescences from new infections using polymerase chain reaction (PCR) with a fixed dose ACT regimen were eligible. Studies on prophylaxis, severe malaria, pregnant women, and patients with hyperparasitaemia, healthy volunteers and travellers were excluded. The details of the studies included are provided in Chapter 3 (Table 3.1).

5.2.2 Statistical analyses

Cumulative failure of recrudescence at the end of the study follow-up was derived using the complement of the Kaplan-Meier estimate of drug efficacy [$\hat{F}_{KM}(t) = 1 - \hat{S}_{KM}(t)$] and the cumulative incidence function (CIF) estimator ($\hat{F}_{CIF}(t)$) as described earlier (Section 5.1.1). New infections and indeterminate outcomes were considered as different categories of competing risk events when estimating $\hat{F}_{CIF}(t)$. They were considered as censored observations when estimating $\widehat{F}_{KM}(t)$. Similarly, recrudescences and indeterminate outcomes were considered as censored observations when deriving the estimates for new infections. The difference in these two estimates was then expressed in absolute and relative terms:

$$Absolute\ difference = \hat{F}_{KM}(t) - \hat{F}_{CIF}(t)$$

$$Relative\ difference = \left(\frac{\hat{F}_{KM}(t) - \hat{F}_{CIF}(t)}{\hat{F}_{CIF}(t)}\right) \times 100\%$$

In comparative efficacy studies, the equality of survival curves for recrudescences between drug regimens was tested using the log-rank test by considering new infections and indeterminate outcomes as censored. The equality of two cumulative incidence functions was tested using Gray's *k*-sample test [87].

The studies included in this analysis had a wide geographical and temporal span. Thus, patients may be correlated within a given study, owing to unobserved shared factors across individuals. In the Cox model, such unobserved heterogeneity is accounted for through the specification of a frailty term shared across the studies, called a shared frailty model (essentially a random effects model). Within the competing risk framework, recent advances have been made to account for such unobserved correlation between subjects within a study by considering it as a clustered data problem [265]. Such approaches are not widely adopted and this is still an active area of statistical research thus limiting the direct comparison of results from cause-specific Cox model (frailties shared across study sites) against the results obtained from Fine and Gray's sub-distribution hazard model. Hence, a subset data from a large multi-centre study (The 4ABC Trial [139]), which enrolled children aged less than 5 years in Africa, was used to illustrate the two regression approaches in the presence of competing risk events. The cause-specific Cox proportional hazard model and Fine and Gray's sub-distribution hazard model were used for modelling time to recrudescence and new infection [262]. The regression parameters of Fine and Gray's model were expressed as sub-distribution hazard (sdHR) ratios, and the output of the cause-specific Cox model as cause-specific hazard ratios (csHR). Regression models were fitted without variable selection as the aim was to use the fitted model for risk prediction (rather than identification of putative factors) using known confounders: age, baseline parasitaemia and treatment regimen. The same set of covariates were used in models for recrudescence and new infection as recommended in Marubini and Valsecchi (p. 347) [266]..The two models were

then used for predicting the risk of recrudescence in a patient with a given combination of covariates.

For a set of *p* covariates (x_i) at time *t*, the predicted risk of recrudescence from Fine and Gray's sub-distribution hazard model ($\hat{F}_{FG}$) was obtained using:

$$\widehat{F_{FG}}(t|x_i) = 1 - \exp\left(- \exp\left(\sum_{k=1}^{p} \beta_k x_{ik} \right) . \int_0^t \overline{h_{1,0}}(s) ds \right)$$

where $\int_0^t \overline{h_{1,0}}(s) ds$ is the cumulative baseline sub-distribution hazard at time *t* and β_k is the regression coefficient for the kth covariate.

The predicted risk from Cox proportional hazard model ($\hat{F}_{Cox}$) was obtained using:

$$\widehat{F_{Cox}}(t|x_i) = 1 - \left(\hat{S}_0(t)^{\exp(\sum_{k=1}^{p} \beta_k x_{ik}))} \right)$$

where $\widehat{S_0}(t)$ is the baseline survival obtained using $\widehat{S_0}(t) = \exp(-\widehat{H_0}(t))$, and $\widehat{H_0}(t)$ is the cumulative baseline hazard at time *t* obtained from the Cox model.

5.2.3 Software

All the analyses were carried out using R software (Version 3.2.4) [267]. The log-rank test was carried out using the **survdiff** function in the **survival** package and Gray's *k*-sample test was performed using the **cuminc** function in the **cmprsk** package.

5.2.4 Results of data analysis

5.2.4.1 Study data

Individual participant data were available from 92 studies (n=31,507 patients) carried out in 169 trial sites investigating 233 treatment arms. A total of 186 arms (79.8%) were from Africa, 45 (19.3%) from Asia and 2 (0.85%) from South America. From these 233 arms, the median observed proportion of recurrence was 1.4% [Interquartile Range (IQR): 0.0 - 3.3%] for recrudescence, 5.4% [IQR: 0.9 - 18.9%] for new infection and 0.9% [IQR: 0.0 - 2.7%] for indeterminate outcome (Figure 5.2). A total of 199 (85.4%) arms had at least one competing risk event with a new infection, and in just over a third (82/233) of the arms the observed proportion of new infections was greater than 10%. The characteristics of the studies and patients included are presented in Chapter 3 (Sections 3.3 and 3.4).

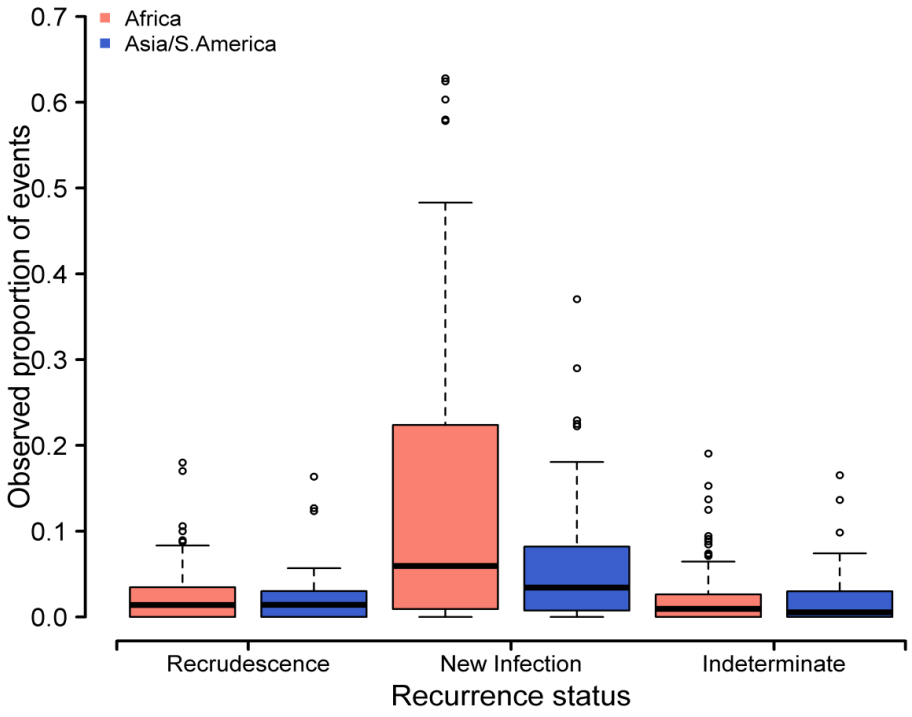


Figure 5.2: The observed proportion of recurrence events in the studies included

5.2.4.2 Risk of *P. falciparum* recrudescence

In 91 arms (39%), either no recrudescences or no competing risk events were observed. In these studies, the failure estimates derived from both methods were identical. In the remaining 142 arms (61%), the 1 minus K-M method was associated with a marginal overestimation of failure compared to the CIF by a median of 0.04% [IQR: 0.00 - 0.27%; Range: 0.00 - 3.60%] (Figure 5.3, Table 5.1). The degree of overestimation got progressively larger with the study follow-up duration; the median overestimation being 0.006% [IQR: 0.00 - 0.07%; Range: 0.00 - 2.54%] on day 28, 0.15% [IQR: 0.00 - 0.57%; Range: 0.00 - 3.23%] on day 42, and 0.56 [IQR: 0.11 - 1.12%; Range: 0.00 - 3.60%] on day 63 (Top panel of Figure 5.3, Table 5.1). The maximum overestimation was 3.6%, which occurred in an artesunate-mefloquine arm in Balonghin site in Burkina-Faso (n=66), an area of high transmission [165]. In this arm, the day 63 failure estimate derived using the K-M method ($\hat{F}_{KM}(t)$) was 20.0% [95% CI: 0.0 - 55.1] and the corresponding CIF estimate ($\hat{F}_{CIF}(t)$) was 16.4% [95% CI: 0.0 - 48.6].

In 9.0% (21/233) of the treatment arms, the overestimation was greater than 1%, in 4.3% (10/233) this was greater than 2% and in 2.6% (6/233) the difference was greater than 3%. All of the 21 study sites where the overestimation exceeded 1% were from Africa except one from Papua New Guinea, where 51% of patients with parasite recurrence experienced a competing event due to *P. vivax* parasitaemia. The magnitude of overestimation was correlated with the observed proportion of the competing events [Pearson's correlation coefficient: 0.43; 95% CI: 0.35 - 0.54] and the observed proportion of recrudescences [Pearson's correlation coefficient: 0.38, 95% CI: 0.30 - 0.46] (Figure 5.3, Table 5.1).

In three (1.3%) study arms, the estimated cumulative risk of recrudescence exceeded 10% based on the K-M method, but the CIF estimates were all less than 10% (Table 5.2). Similarly,

in 9 (3.9%) study arms, the estimated failure were greater than 5% using the K-M method, all of which were less than 5% using the CIF.

5.2.4.3 Risk of *P. falciparum* new infection

The median overestimation in the *P. falciparum* recurrence estimate for new infections derived by the K-M method (which considered recrudescences as censored) compared to the cumulative incidence function was 0.39% [IQR: 0.08 - 1.10%; Range: 0.00 - 10.60%]. The overestimation progressively increased with the follow-up duration and this was 0.10% on day 28, 0.67% on day 42, and 1.40% on day 63 (Appendix 5.1). As in the case of recrudescence, the overestimation increased with increasing proportion of new infection and recrudescence observed in a study (Figure 5.3, bottom panel). The maximum overestimation was 10.6%, observed in a study with artemether-lumefantrine (n=50) carried out in Tanzania. In the study, there were 9 (18%) recrudescences, and 29 (58%) new infections observed. Using the 1 minus K-M method, the risk of new infection on day 42 was 68.6% [95% CI: 54.3 - 83.1] and this was 58.0% [95% CI: 44.2 - 71.7] using the CIF method.

Table 5.1: Absolute overestimation in cumulative recrudescence using the Kaplan-Meier method compared to the Cumulative Incidence Function

	N	Absolute overestimation[a]	Observed proportion of recrudescence[b]	Observed proportion of competing risk events[b]
Day 28				
AL	88	0.02% [IQR : 0.00 - 0.12%; Range : 0.00 - 2.54%]	2.1% [0.1 - 16.0%]	9.8% [0.0 - 51.4%]
ASAQ	21	0.03% [IQR : 0.00 - 0.10%; Range : 0.00 - 0.98%]	2.5% [0.3 - 12.8%]	10.2% [0.0 - 45.4%]
ASMQ	8	0.00% [IQR : 0.00 - 0.03%; Range : 0.00 - 0.08%]	1.9% [0.8 - 7.0%]	11.0% [6.3 - 14.0%]
DP	30	0.00% [IQR : 0.00 - 0.00%; Range : 0.00 - 0.16%]	1.3% [0.4 - 16.4%]	2.7% [0.0 - 36.8%]
Day 42				
AL	48	0.25% [IQR : 0.03 - 0.65%; Range : 0.00 - 3.24%]	3.7% [0.6 - 18.0%]	21.6% [0.0 - 65.4%]
ASAQ	9	0.23% [IQR : 0.02 - 0.69%; Range : 0.00 - 2.24%]	3.3 % [1.0 - 17.0%]	18.4% [2.3 - 65.2%]
ASMQ	8	0.25% [IQR : 0.12 - 0.50%; Range : 0.00 - 1.03%]	3.7% [1.4 - 11.3%]	25.3% [1.0 - 43.3%]
DP	21	0.00% [IQR : 0.00 - 0.19%; Range : 0.00 - 0.73%]	1.9 % [0.8 - 16.4%]	7.1% [0.0 - 34.6%]
Day 63				
AL	9	0.96% [IQR : 0.65 - 1.38%; Range : 0.012 - 3.42%]	3.7% [1.4 - 10.6%]	7.1 % [0.0 - 34.6%]
ASAQ	2	0.18% [IQR : 0.09 - 0.27%; Range : 0.00 - 0.36%]	12.3%	18.5%
ASMQ	7	0.98% [IQR : 0.38 - 2.20%; Range : 0.21 - 3.59%]	4.1% [1.6 - 12.7%]	4.1% [1.6 - 12.7%]
DP	8	0.14% [IQR : 0.01 - 0.38%; Range : 0.00 - 0.89%]	2.2% [1.6 - 12.7%]	7.1 % [3.4 - 30.0%]

[a]Values are median [Interquartile Range (IQR); Range] for $\hat{F}_{KM}(t) - \hat{F}_{CIF}(t)$
[b]Values are median [Range]
N=number of study sites with study sample size ≥25
AL = artemether-lumefantrine; ASAQ = artesunate-amodaiquine; DP = dihydroartemisinin-piperaquine; ASMQ = artesunate-mefloquine.

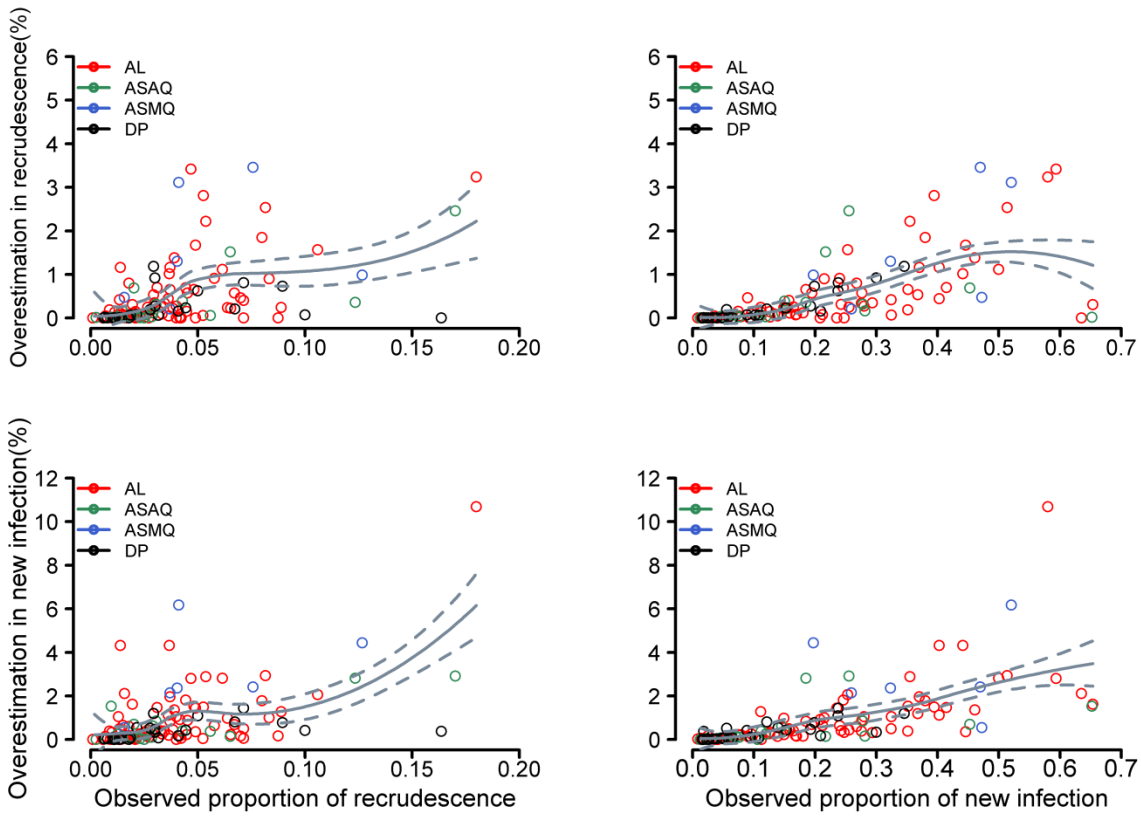


Figure 5.3: The overestimation of derived failure using 1 minus Kaplan-Meier method compared to the cumulative incidence function method

Legend: The overestimation $(\hat{F}_{KM}(t) - \hat{F}_{CIF}(t))$ of cumulative recrudescence (top panel) and new infection (bottom panel) by using the Kaplan-Meier method plotted against observed proportion of recrudescence and proportion of new infections respectively. Estimates presented are at the end of the study follow-up. The grey trend line is a smoothed estimator obtained from local polynomial regression fitting, shown together with 95% confidence interval (outer dotted lines) for the overall data. AL = artemether-lumefantrine, ASAQ = artesunate-amodiaquine; DP = dihydroartemisinin-piperaquine; ASMQ = artesunate-mefloquine. Data is shown from the study arms where at least one recrudescence and at least one competing risk event were observed and from those arms where the number of patients at risk > 25 on the last day of the study follow-up.

Table 5.2: Study sites where cumulative failure estimates were >10% and >5% using the Kaplan-Meier method

Study (site)	N	Number of events (RC/NI/IND)	Day	Drug	$\widehat{F}_{KM}$ estimate of recrudescence [95% confidence interval]	$\widehat{F}_{CIF}$ estimate of recrudescence [(95% confidence interval]
10% threshold†						
Nikiema-2010 (Burkina Faso, Gourcy)*	144	12/28/3	28	AL	10.3 [4.8 - 15.8]	9.4 [4.3 - 14.5]
The 4ABC Trial (Burkina Faso, Nanoro) [139]	294	24/142/9	28	AL	11.1 [6.8 - 15.4]	8.6 [5.3 - 11.8]
Sirima-2015 (Burkina Faso, Balonghin) [165]	66	31/5/0	63	ASMQ	11.5 [7.2 - 15.9]	8.1 [1.2 - 14.9]
5% threshold†						
Yeka-2008 (Uganda, Kanungu) [152]	199	9/49/6	42	AL	5.2 [1.9 - 8.5]	4.6 [1.7 - 7.6]
Sirima-2015 (Tanzania, Korogwe) [165]	27	1/3/2	63	AL	5.0 [0.0 - 14.6]	4.3 [0.0 - 12.9]
Schramm-2013 (Liberia, Nimba) [178]	145	7/45/6	42	AL	5.6 [1.5 - 9.7]	5.0 [1.4 - 8.6]
Agrawal-2013 (Kenya, Siaya) [134]	136	5/48/12	42	AL	5.2 [0.5 - 10.0]	4.2 [0.6 - 7.8]
Karunajeewa-2008 (PNG, Madang) [177]	54	2/20/0	42	AL	5.9 [0.0 - 14.3]	4.7 [0.0 - 11.3]
Sirima-2015 (Kenya, Kisumu) [165]	99	4/23/9	63	ASMQ	6.1 [0.1 - 12.0]	4.8 [0.2 - 9.4]
Sirima-2015 (Burkina Faso, Balonghin) [165]	128	5/58/1	63	AL	5.4 [0.7 - 10.2]	4.0 [0.6 - 7.5]
Bukirwa-2006 (Uganda, Tororo) [185]	204	10/89/2	28	AL	6.6 [2.6 - 10.7]	4.9 [2.0 - 8.0]
Sirima-2015 (Kenya, Ahero) [165]	73	3/28/10	63	ASMQ	7.8 [0.0 - 16.7]	4.7[0.0 - 10.0]

N= study sample size; RC= Recrudescence; NI= New infection; IND= Indeterminate outcomes; *= Unpublished study

$\hat{F}_{KM}(t)$= Cumulative failure estimates derived using 1- Kaplan-Meier method

$\hat{F}_{CIF}(t)$= Cumulative failure estimates derived using Cumulative Incidence Function

AL= artemether-lumefantrine; ASMQ = artesunate-mefloquine

†The 10% threshold is used by the WHO for determining whether the current regimen should be continued to be used as a first line therapy and 5% threshold is used for introducing a new regimen as a first line treatment [59]

5.2.4.4 Comparative efficacy studies

Artemether-lumefantrine was compared against dihydroartemisinin-piperaquine in 14 studies (Table 5.3), against artesunate-mefloquine in 4 studies (Table 5.4), and against artesunate-amodiaquine in 12 studies (Table 5.5). Similarly, artesunate-amodiaquine was compared with dihydroartemisinin-piperaquine in 3 studies (Table 5.6), and against artesunate-mefloquine in one study (Table 5.8). Dihydroartemisinin-piperaquine and artesunate-mefloquine were compared in two studies (Table 5.7).

There were no apparent differences in the results and overall conclusion derived from the two approaches (Tables 5.3, 5.4, 5.5, 5.6, 5.7, and 5.8).

The implication of competing risk events on comparative efficacy is further explored using a simulation study in Section 5.3.

Table 5.3: Comparing AL against DP using log-rank test and Gray's *k*-sample test

Study (site)	Day	$\widehat{F}_{CIF}$ estimate of recrudescence [95% confidence interval]		$\widehat{F}_{KM}$ estimate of recrudescence [95% confidence interval]		Tests for equality of failure (*P*-value)	
		AL	DP	AL	DP	Log-rank test	Gray's *k*-sample test
Arinaitwe-2009 (Uganda, Tororo) [101]	28	0.0%	0.0%	0.0%	0.0%	-	-
Sawa-2013 (Kenya, Mbita) [122]	42	2.7% [0.1 - 5.4]	0.0%	2.8% [0.1 - 5.5]	0.0%	0.050	0.053
Agrawal-2013 (Kenya,Siaya) [134]	42	4.2% [0.6 - 7.8]	3.6% [0.1 – 7.0]	5.2% [0.5 – 10.0]	4.7% [0.2 - 9.3]	0.550	0.707
The 4ABC Trial (Burkina Faso, Nanoro) [139]	28	8.6% [5.3 - 11.8]	3.2% [0.9 - 5.6]	11.1% [6.8 - 15.4]	3.3% [0.9 - 5.7]	0.001	0.012
The 4ABC Trial (Gabon, Fougamou) [139]	28	0.0%	0.0%	0.0%	0.0%	-	-
The 4ABC Trial (Nigeria,Afokang) [139]	28	0.7% [0.0 - 1.9]	0.0%	0.7% [0.0 - 2.0]	0.0%	0.372	0.666
The 4ABC Trial (Rwanda, Mashesha) [139]	28	1.3% [0.0 - 3.9]	0.0%	1.3% [0.0 - 3.9]	0.0%	0.321	0.317
The 4ABC Trial (Rwanda, Rukara) [139]	28	4.2% [0.0 - 8.9]	2.8% [0.0 - 6.7]	4.2% [0.0 - 8.9]	2.8% [0.0- 6.7]	0.643	0.648
The 4ABC Trial (Uganda, Jinja) [139]	28	1.9% [0.1 - 3.8]	1.4% [0.0 - 3]	2.0% [0.1 - 3.9]	1.4% [0.0 – 3.0]	0.716	0.727
The 4ABC Trial (Uganda, Tororo) [139]	28	3.7% [1.2 - 6.2]	4.2% [1.5 - 6.8]	4.1% [1.3 - 7.0]	4.3% [1.6 - 7.1]	0.996	0.819
The 4ABC Trial (Zambia, Ndola) [139]	28	1.4% [0.0 - 4.1]	0.0%	1.4% [0.0 - 4.1]	0.0%	0.331	0.331
Adam 2010 (Sudan,Sinnar)	28	0.0%	0.0%	0.0%	0.0%	-	-
Bassat-2009 (Burkina Faso, Nanoro) [146]	42	8.1% [2.7 - 13.5]	9.1% [5.1 - 13.1]	9.9% [3.2 - 16.6]	9.8% [5.5 - 14.1]	0.675	0.828
Bassat-2009 (Mozambique, Manhica) [146]	42	1.5% [0.0 - 3.6]	6.9% [4.0 - 9.9]	1.6% [0.0 - 3.8]	7.2% [4.1 - 10.2]	0.021	0.018
Bassat-2009 (Uganda,Mbarara) [146]	42	3.4% [0.0 - 7.2]	1.1% [0.0- 2.6]	3.6% [0.0 - 7.7]	1.1% [0.0 - 2.6]	0.189	0.204
Bassat-2009 (Zambia, Ndola) [146]	42	4.1% [0.1 - 8.1]	4.6% [1.7 - 7.6]	4.4% [0.2 - 8.7]	4.9% [1.8 - 7.9]	0.931	0.909
Mens-2008 (Kenya, Mbita) [151]	28	0.0%	0.0%	0.0%	0.0%	-	-
Yeka-2008 (Uganda,Kanungu) [152]	42	4.6% [1.7 - 7.6]	1.9% [0.1 - 3.7]	5.2% [1.9 - 8.5]	1.9% [0.0 - 3.8]	0.082	0.118
Borrmann-2011 (Kenya, Pingilikani) [162]	63	11.4% [7.0 - 15.7]	7.8% [4.0 - 11.7]	12.9% [8.0 - 17.9]	8.7% [4.4 - 12.9]	0.237	0.251
Zongo-2007b (Burkina Faso, Bobo) [171]	42	2.8% [0.4 - 5.3]	1.2% [0.0 - 2.8]	3.1% [0.4 - 5.8]	1.2% [0.0 - 2.8]	0.246	0.274
Kamya-2007 (Uganda, Apac) [175]	42	0.0%	0.0%	0.0%	0.0%	-	-
Karunajeewa-2008 (PNG, Madang) [177]	42	4.7% [0.0 - 11.3]	6.5% [0.2 - 12.7]	5.9% [0.0 - 14.2]	7.1% [0.3 - 13.9]	0.825	0.753
Karunajeewa-2008 (PNG, Maprik) [177]	42	14.3% [0.0 - 33.3]	18.2% [1.7 - 34.7]	14.3% [0.0 - 32.6]	18.2% [2.1 - 34.3]	0.763	0.763
Yavo-2011 (Cameroon,Yaounde) [180]	28	2.1% [0.0 - 6.2]	1.8% [0.0 - 5.3]	2.1% [0.0 - 6.1]	1.8% [0.0 - 5.3]	0.913	0.913

Yavo-2011 (Ivorycoast, Bocabo) [180]	28	1.4% [0.0 - 4.2]	0.0%	1.4% [0.0 - 4.2]	0.0%	0.307	0.307
Yavo-2011 (Senegal,Touba) [180]	28	0.0%	0.0%	0.0%	0.0%	-	-
Sylla-2013 (Senegal, Pikine) [184]	42	2.3% [0.0 - 4.8]	0.9% [0.0 - 2.6]	2.3% [0.0 - 4.9]	0.9% [0.0 - 2.6]	0.372	0.384

$\hat{F}_{KM}(t)$= Cumulative failure estimates derived using 1- Kaplan-Meier approach
$\hat{F}_{CIF}(t)$= Cumulative failure estimates derived using Cumulative Incidence Function
AL= artemether-lumefantrine; DP= dihydroartemisinin-piperaquine

Table 5.4: Comparing AL against ASMQ using log-rank test and Gray's *k*-sample test

		$\widehat{F}_{CIF}$ estimate of recrudescence (95% confidence interval)		$\widehat{F}_{KM}$ estimate of recrudescence (95% confidence interval)		Tests for equality of failure (*P*-value)	
Study (site)	**Day**	**AL**	**ASMQ**	**AL**	**ASMQ**	**Log-rank test**	**Gray's *k*-sample test**
Smithuis-2010 (Myanmar, Rakhine) [118]	63	1.3% [0.0 - 3.9]	0.0%	1.4% [0.0 - 4.0]	0.0%	0.220	0.221
Faye-2007 (Senegal, multisite) [126]	28	0.0%	0.0%	0.0%	0.0%	-	-
Sirima-2015 (Burkina Faso, Balonghin) [165]	63	4.0% [0.6 - 7.5]	1.6% [0.0 - 3.9]	5.4% [0.7 - 10.2]	2.1% [0.0 - 5.0]	0.188	0.255
Sirima-2015 (Burkina Faso, Banfora) [165]	63	5.1% [0.0 - 10.8]	8.1% [1.2 - 14.9]	8.5% [0.0 - 18.1]	11.5% [1.7 - 21.4]	0.699	0.515
Sirima-2015 (Kenya, Kisumu) [165]	63	6.4% [0.9 - 11.9]	4.8% [0.2 - 9.4]	8.6% [1.2 - 16.1]	6.1% [0.1 - 12]	0.507	0.614
Sirima-2015 (Kenya, Ahero) [165]	63	1.8% [0.0 - 5.3]	4.7% [0.0 - 10.0]	2.9% [0.0 - 8.6]	7.8% [0.0 - 16.7]	0.375	0.385
Sirima-2015 (Tanzania, Kilosa) [165]	63	0.0%	5.3% [0.0 - 15.6]	0.0%	5.6% [0.0 - 16.2]	0.361	0.331
Sirima-2015 (Tanzania, Korogwe) [165]	63	4.3% [0.0 - 12.9]	0.0%	5.0% [0.0 - 14.5]	0.0%	0.371	0.343
Sirima-2015 (Tanzania, Bagamayo) [165]	63	2% [0.0 - 5.9]	0.0%	2.8% [0.0 - 8.2]	0.0%	0.405	0.985
Faye-2010 (Senegal, Kaolack) [181]	28	0.0%	0.0%	0.0%	0.0%	-	-

$\hat{F}_{KM}(t)$= Cumulative failure estimates derived using 1- Kaplan-Meier approach
$\hat{F}_{CIF}(t)$= Cumulative failure estimates derived using Cumulative Incidence Function
AL= artemether-lumefantrine; ASMQ= artesunate-mefloquine

Table 5.5: Comparing AL against ASAQ using log-rank test and Gray's *k*-sample test

Study (site)	Day	$\widehat{F}_{CIF}$ estimate of recrudescence [95% confidence interval]		$\widehat{F}_{KM}$ estimate of recrudescence [95% confidence interval]		Tests for equality of failure (*P*-value)	
		AL	ASAQ	AL	ASAQ	Log-rank test	Gray's *k*-sample test
Smithuis-2010 (Myanmar, Rakhine) [118]	63	1.3% [0.0 - 3.9]	11.8% [5.2 - 18.4]	1.4% [0.0- 4.0]	12.3% [5.5 - 19.1]	0.008	0.008
de Wit-2015* (DRC, Baraka)	42	2.4% [0.0 - 5.1]	5% [1.1 - 9.0]	2.6% [0.0 - 5.5]	5.4% [1.2 - 9.7]	0.270	0.281
The 4ABC Trial (Burkina Faso, Nanoro) [139]	28	8.6% [5.3 - 11.8]	3.1% [1.1 - 5.2]	11.1% [6.8 - 15.4]	3.5% [1.2 - 5.7]	0.001	0.006
The 4ABC Trial (Gabon, Fougamou) [139]	28	0.0%	2.8% [0.0 - 6.7]	0.0%	2.9% [0 - 6.8]	0.154	0.159
The 4ABC Trial (Nigeria,Afokang) [139]	28	0.7% [0.0 - 1.9]	0.0%	0.7% [0.0 - 2.0]	0.0%	0.320	0.313
The 4ABC Trial (Zambia, Ndola) [139]	28	1.4% [0.0 - 4.1]	0.0%	1.4% [0.0 - 4.1]	0.0%	0.328	0.328
Nikiema-2010* (Burkina Faso, Bobo)	28	2.1% [0.7 - 3.5]	0.3% [0.0 - 0.8]	2.2% [0.7 - 3.7]	0.3% [0.0 - 0.9]	0.019	0.021
Nikiema-2010* (Burkina Faso, Gourcy)	28	9.4% [4.3 - 14.5]	5.7% [1.9 - 9.6]	10.3% [4.7 - 15.8]	5.8% [1.9 - 9.7]	0.225	0.281
Espié-2012 (DRC, Pweto) [160]	42	0.8% [0.0 - 2.3]	2.2% [0.0 - 4.7]	0.8% [0.0 - 2.4]	2.3% [0.0 - 4.8]	0.366	0.336
Ndiaye-2011 (Senegal, Keur Sosse) [164]	28	0.0%	0.0%	0.0%	0.0%	-	-
Offianan-2011 (Ivorycoast, Ayame) [168]	28	0.0%	0.0%	0.0%	0.0%	-	-
Offianan-2011 (Ivorycoast, Dabakala) [168]	28	0.0%	0.0%	0.0%	0.0%	-	-
Yeka-2014 (Uganda, Tororo) [169]	42	2.0% [0.1 - 3.9]	1.0% [0.0 - 2.4]	2.3% [0.1 - 4.6]	1.0% [0.0 - 2.5]	0.420	0.423
Ndiaye-2009 (Madagascar,Tsiroanomandidy) [176]	28	1.7% [0.0 - 4.9]	0.0%	1.7% [0.0 - 4.9]	0.0%	0.166	0.166
Ndiaye-2009 (Cameroon,Yaoundé) [176]	28	0.0%	0.0%	0.0%	0.0%	-	-
Ndiaye-2009 (Mali, Bancoumana) [176]	28	3.1% [0.0 - 7.5]	1.6% [0.0- 3.7]	3.7% [0.0 - 8.8]	1.7% [0.0 - 4.1]	0.427	0.472
Ndiaye-2009 (Senegal, Mlomp/Keur-Socé) [176]	28	1.3% [0.0 - 3.9]	0.0%	1.5% [0.0 - 4.3]	0.0%	0.150	0.161
Ndiaye-2009 (Senegal, Mlomp/Keur-Socé) [176]	28	2.6% [0.0 - 7.8]	0.0%	2.6% [0.0 - 7.7]	0.0%	0.175	0.175
Schramm-2013 (Liberia, Nimba) [178]	42	5% [1.4 - 8.6]	2.1% [0.0 - 4.4]	5.6% [1.6 - 9.7]	2.7% [0.0 - 5.9]	0.255	0.177
Sylla-2013 (Senegal, Pikine) [184]	42	2.3% [0.0- 4.8]	2.5% [0.0 - 5.3]	2.3% [0.0 - 4.9]	2.5% [0.0 - 5.3]	0.930	0.913
Faucher-2009 (Benin, Allada) [188]	42	10.1% [0.6 - 19.7]	7.0% [0.0 - 14.7]	10.4% [0.7 - 20.1]	8.5% [0.0 - 17.7]	0.624	0.555
Faucher-2009 (Benin, Sekou) [188]	42	7.0% [0.0- 14.7]	18.2% [6.7 - 29.7]	7.2% [0.0 - 15.1]	20.7% [7.8 - 33.6]	0.077	0.113

$\hat{F}_{KM}(t)$= Cumulative failure estimates derived using 1- Kaplan-Meier approach

$\hat{F}_{CIF}(t)$= Cumulative failure estimates derived using Cumulative Incidence Function; AL= artemether-lumefantrine; ASAQ= artesunate-amodiaquine

Table 5.6: Comparing ASAQ against DP using log-rank test and Gray's *k*-sample test

Study (site)	Day	$\widehat{F}_{CIF}$ estimate of recrudescence [95% confidence interval]		$\widehat{F}_{KM}$ estimate of recrudescence [95% confidence interval]		Tests for equality of failure (*P*-value)	
		ASAQ	DP	ASAQ	DP	Log-rank test	Gray's *k*-sample test
Smithuis-2010 (Myanmar, Rakhine) [118]	63	11.8% [5.2 - 18.4]	1.9% [0.0 - 4.6]	12.3% [5.5 - 19.1]	2.0% [0.0 - 4.7]	0.005	0.005
The 4ABC Trial (Burkina Faso, Nanoro) [139]	28	3.1% [1.1 - 5.2]	3.2% [0.9 - 5.6]	3.5% [1.2 - 5.7]	3.3% [0.9 - 5.7]	0.824	0.995
The 4ABC Trial (Gabon, Fougamou) [139]	28	2.8% [0.0 - 6.7]	0.0%	2.9% [0.0 - 6.8]	0.0%	0.180	0.186
The 4ABC Trial (Mozambique, Manhica) [139]	28	3.2% [0.7 - 5.7]	1.6% [0.0 - 3.4]	3.3% [0.7 - 5.9]	1.6% [0.0 - 3.4]	0.284	0.308
The 4ABC Trial (Nigeria,Afokang) [139]	28	0.0%	0.0%	0.0%	0.0%	-	-
The 4ABC Trial (Uganda,Mbarara) [139]	28	1.3% [0.0 - 3.2]	0.7% [0.0 - 2.0]	1.4% [0.0 - 3.3]	0.7% [0.0 - 2.0]	0.554	0.562
The 4ABC Trial (Zambia, Ndola) [139]	28	0.0%	0.0%	0.0%	0.0%	-	-
Sylla-2013 (Senegal, Pikine) [184]	42	2.5% [0.0 - 5.3]	0.9% [0.0 - 2.6]	2.5% [0.0 - 5.3]	0.9% [0.0 - 2.6]	0.334	0.339

$\hat{F}_{KM}(t)$= Cumulative failure estimates derived using 1- Kaplan-Meier approach
$\hat{F}_{CIF}(t)$= Cumulative failure estimates derived using Cumulative Incidence Function
ASAQ= artesunate-amodiaquine; DP= dihydroartemisinin-piperaquine

Table 5.7: Comparing ASMQ against DP using log-rank test and Gray's *k*-sample test

Study (site)	Day	$\widehat{F}_{CIF}$ estimate of recrudescence [95% confidence interval]		$\widehat{F}_{KM}$ estimate of recrudescence [95% confidence interval]		Tests for equality of failure (*P*-value)	
		ASMQ	DP	ASMQ	DP	Log-rank test	Gray's *k*-sample test
Smithuis-2010 (Myanmar, Rakhine) [118]	63	0%	1.9% [0.0- 4.6]	0%	2%[0.0 - 4.7]	0.134	0.137
Hien-2006* (VietNam, Dac O)	56	12.2% [0.0 - 24.5]	2.7% [1.3 - 4.1]	12.5% [0.2 - 24.8]	3% [1.5 - 4.5]	0.001	0.001

$\hat{F}_{KM}(t)$= Cumulative failure estimates derived using 1- Kaplan-Meier approach
$\hat{F}_{CIF}(t)$= Cumulative failure estimates derived using Cumulative Incidence Function
ASMQ= artesunate-mefloquine; DP= dihydroartemisinin-piperaquine

Table 5.8: Comparing ASMQ against ASAQ using log-rank test and Gray's *k*-sample test

Study (site)	Day	$\widehat{F}_{CIF}$ estimate of recrudescence [95% confidence interval]		$\widehat{F}_{KM}$ estimate of recrudescence [95% confidence interval]		Tests for equality of failure (*P*-value)	
		ASMQ	ASAQ	ASMQ	ASAQ	Log-rank test	Gray's *k*-sample test
Smithuis-2010 (Myanmar, Rakhine) [118]	63	0.0%	11.8% [5.2 - 18.4]	0.0%	12.3% [5.5 - 19.1]	$p<0.001$	$p<0.001$

$\widehat{F}_{KM}(t)$= Drug failure estimates derived using 1- Kaplan-Meier approach
$\widehat{F}_{CIF}(t)$= Drug failure estimates derived using Cumulative Incidence Function
ASMQ= artesunate-mefloquine; ASAQ= artesunate-amodiaquine

5.2.4.5 Regression models in the presence of competing risk events

A total of 810 children enrolled in Burkina Faso (a subset of the The 4ABC Trial [139]) treated with artemether-lumefantrine (AL) (n=294), artesunate-amodiaquine (ASAQ) (n=295) and dihydroartemesinin-piperaquine (DP) (n=221) were used to illustrate the regression modelling approach in the presence of competing risk events. The observed proportions of recrudescence were 3.2% in the DP arm, 8.2% in the AL arm, and 3.1% in the ASAQ arm, while the respective proportions of new infections were 10.0%, 48.3%, and 25.1% (Table 5.9). In a multivariable model for recrudescence (which included age, baseline parasitaemia and treatment regimen), age and baseline parasitaemia did not reach conventional statistical significance for recrudescence, neither in the cause-specific hazard model nor in the sub-distribution hazard model (*P*-value > 0.05). Treatment with AL (relative to DP) was associated with increased cause-specific hazard and increased sub-distribution hazard of recrudescence (csHR = 4.02 [95% CI: 1.72 - 9.43]; sdHR=2.85 [95% CI: 1.24 - 6.57]). All the estimates of the sub-distribution hazard ratio (sdHR) were somewhat attenuated and closer to the null value compared to the cause-specific hazard ratio (csHR).

The results presented in Table 5.9 were used for predicting the risk of recrudescence on day 28 for a patient aged 3 years, who on presentation to the clinic had a parasite load of 100,000/µL. For the DP regimen, the predicted risk for a patient with this covariate profile was 4.28% using the cause-specific Cox model, and 3.90% using the sub-distribution hazard model (Figure 5.4). For the AL regimen, the predicted risks were 16.15% and 10.72% using the cause-specific and sub-distribution hazard models respectively. For ASAQ, the estimates of predicted risks were 4.76% using the cause-specific Cox model and 3.94% using the Fine and Gray's sub-distribution hazard model (Figure 5.4).

Table 5.9: Regression models for recrudescence and new infection using data from Burkina Faso (n=810); subset of The 4ABC Trial [139]

		Cause-specific Cox proportional hazard model		**Fine and Gray's sub-distribution hazard model**	
Model for recrudescence	Observed proportion of events	cause-specific HR [95% confidence interval]	*P*-value	sub-distribution HR [95% confidence interval]	*P*-value
Age (/yearly increase)	4.9% (40/810)	0.99 [0.77 - 1.29]	0.965	1.00 [0.80 - 1.25]	0.990
Baseline parasitaemia (/µL) (10-fold increase)	4.9% (40/810)	1.67 [0.88 - 3.16]	0.114	1.44 [0.84 - 2.46]	0.190
Treatment (ref= DP)	3.2% (7/221)				
AL	8.2% (24/294)	4.02 [1.72 - 9.43]	0.001	2.85 [1.24 - 6.57]	0.014
ASAQ	3.1% (9/295)	1.12 [0.42 - 2.99]	0.829	1.01 [0.38 - 2.69]	0.980
Model for new infection					
Age (/yearly increase)	29.4% (238/810)	1.08[0.98 - 1.21]	0.121	1.08[0.99 - 1.19]	0.093
Baseline parasitaemia (/µL) (10-fold increase)	29.4% (238/810)	1.52[1.18 - 1.97]	0.001	1.40[1.10 - 1.78]	0.006
Treatment (ref= DP)	10.0% (22/221)				
AL	48.3% (142/294)	8.05[5.12 - 12.67]	$p<0.001$	6.51[4.27 - 9.94]	$p<0.001$
ASAQ	25.1% (74/295)	3.03[1.88 - 4.88]	$p<0.001$	2.83[1.80 - 4.44]	$p<0.001$

AL= artemether-lumefantrine; DP = dihydroartemisinin-piperaquine; ASAQ = artesunate-amodiaquine

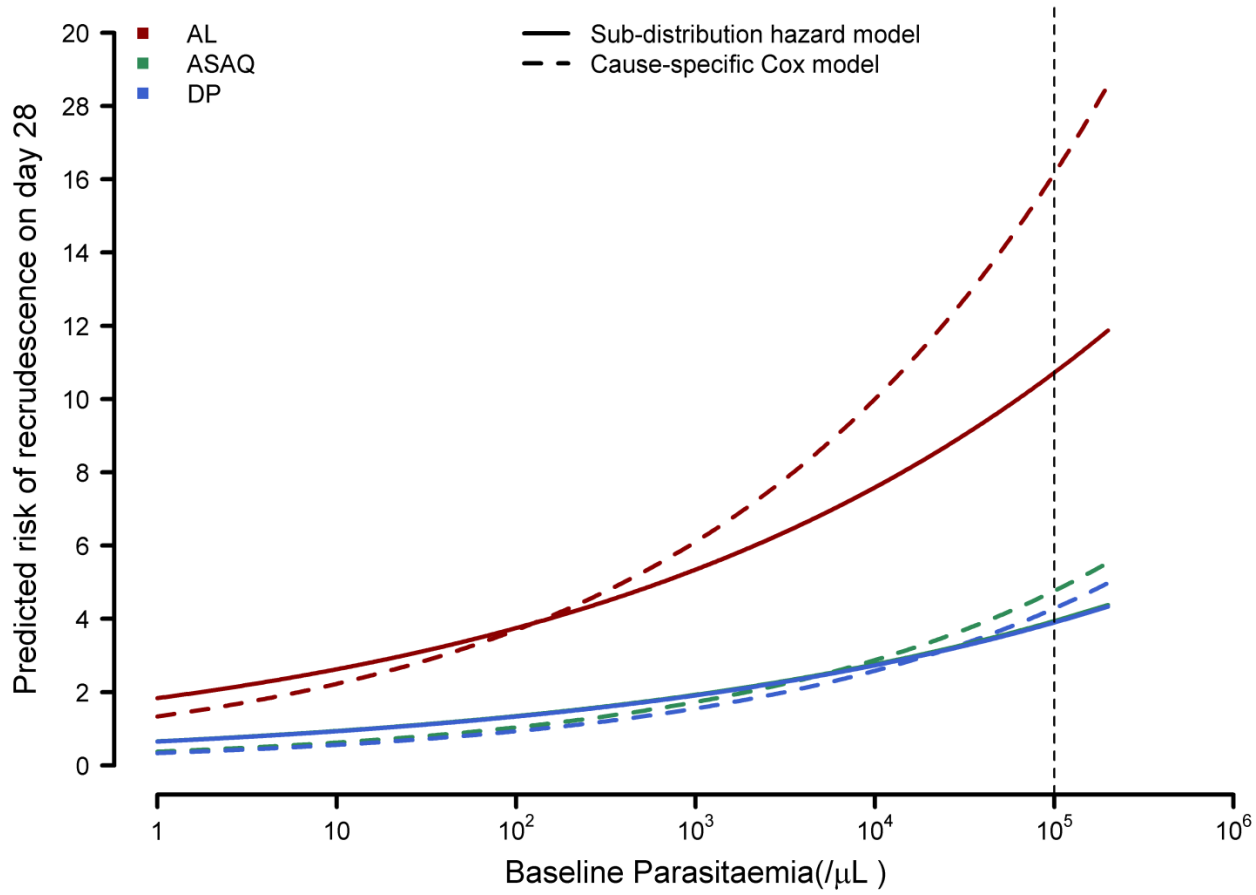


Figure 5.4: Predicted risk of recrudescence from cause-specific Cox model and sub-distribution hazard model

Legend: The graph was generated using the regression coefficients presented in Table 5.9 and the estimate of baseline hazard obtained from the respective sub-distribution and cause-specific hazard model (for a 3 year old child). The cumulative baseline sub-distribution hazard on day 28 from Fine and Gray's model was 0.0064768082; the cumulative baseline hazard on day 28 from the cause-specific Cox model was 0.003401961. The vertical dotted line represents the parasitaemia of 100,000/μL for the child described in the main text. On day 28, the predicted risk of recrudescence for this patient was 16.15%, 4.76% and 4.28% using the cause-specific Cox model. The corresponding figures were 10.72%, 4.76% and 4.28% with the Fine and Gray's sub-distribution hazard model. AL = artemether-lumefantrine; ASAQ = artesunate-amodiaquine; DP = dihydroartemisinin-piperaquine.

5.2.5 Conclusions from data analysis

The key findings of the analyses carried out in this section are:

- The cumulative probability of failure derived using the K-M method was overestimated by a median of 0.04% [IQR: 0.00 - 0.27%, Range: 0.00 - 3.60%] (compared to the CIF). This overestimation progressively increased with increasing proportions of recrudescence and new infection observed in a study.

- In 27 comparative studies, the comparison of equality of the survival curves using the log-rank test and the comparison of the cumulative incidence using the Gray's *k*-sample test provided similar conclusion.

- The predicted risks of recrudescence derived from the cause-specific Cox model were larger (by up to 50% in the example studied) than those obtained from the Fine and Gray's sub-distribution hazard model.

The results presented in this section demonstrate that antimalarial studies are susceptible to competing risk bias as just over a third of the studies had observed proportions of new infections greater than 10%. The efficacy of artemisinin combination therapies remained consistently high, with the median observed proportion of recrudescence in the studies included being 1.4% [IQR: 0.0 - 3.3%]. This limited the ability to draw inferences on scenarios of falling antimalarial drug efficacies. In order to explore the variations in derived efficacy in hypothetical scenarios of declining ACT efficacy and to further understand how the comparative efficacy would be affected, two simulation studies were carried out.

5.3 Simulation studies

Simulation study I

The first study aimed to quantify the magnitude of overestimation in failure by the K-M method compared to the CIF in a non-comparative efficacy trial. This started with a base case simulation of 5% recrudescence to represent current observations with ACTs and to replicate the data used in this thesis. To model situations likely to be observed when antimalarial drug resistance worsens, scenarios of 10% and 15% recrudescences were also studied. For each of these scenarios, different proportions of new infections were simulated to represent areas of different transmission settings.

Simulation study II

The second study was set-up to explore the impact of competing risk events in a comparative trial. Nine different scenarios, which could be potentially observed when comparing two drug regimens, were simulated. The results of log-rank test for comparing the equality of the survival curves for recrudescence and Gray's *k*-sample test for comparing the equality of the cumulative incidence of recrudescence, for each of these nine settings were summarised.

5.3.1 Methodology

5.3.1.1 Simulation set-up for study I

The following combinations of recurrences were generated: recrudescent proportion (5%, 10%, and 15%) and new infection proportion (<10%, 10-20%, 20-40% and >40%). To replicate the antimalarial data structure, the times to parasitic recurrences were simulated from the empirical baseline hazard function derived from the cause-specific Cox model for recrudescence and new infection under the assumption that time to recrudescence and new infections were independent (thus simplifying the simulation set-up). A slight variant of the empirical hazard function of recrudescence for dihydroartemisinin-piperaquine (DP) presented in Section 4.3.8.3 was used as a case study in order to achieve the desired proportion of recrudescence and new infection but the result is equally generalizable for any antimalarials.

The following steps describe the simulation protocol:

i. Simulate time to recrudescence (t_1) based on the following cumulative baseline hazard (CBH) on log-scale using **survsim** stata package:

$$\ln(CBH(t)_{rc}) = \beta_0 - 63.6284 \times \{\ln(t)^{-1} - 0.2849\} - 0.3800 \times \{\ln(t)^2 - 12.3188\}$$

The parameter β_0 was varied to achieve the desired proportion of recrudescence.

$\beta_0 = -3.7092$ for approximately 5% recrudescence

$\beta_0 = -3.0160$ for approximately 10% recrudescence

$\beta_0 = -2.6105$ for approximately 15% recrudescence

The hazard and survival function used for this simulation are depicted in Figure 5.5.

ii. Simulate time to new infections (t_2) using **survsim** stata package based on the following CBH function:

$$\ln(CBH(t)_{ni}) = \alpha_0 + 9501.2150 \times \{\ln(t)^{-2} - 0.0858\} - 31651.33 \times \{\ln(t)^{-2} \times \ln(\ln(t) - 0.1054\} + 29340.83 \times \{lnt^{-2} \times \ln(lnt)^2 - 0.1294\} - 12690.51 \times \{\ln(t)^{-2} \times \ln(\ln(t))^3 - 0.1588\}$$

The parameter α_0 was varied in order to achieve the desired proportion of new infections.

$\alpha_0 = -5.6004$ for approximately <10% new infections

$\alpha_0 = -3.9909$ for approximately 10 to <20% new infections

$\alpha_0 = -3.2978$ for approximately 20 to <40% new infections

$\alpha_0 = -2.8924$ for approximately >40% new infections

The hazard and survival function used for these scenarios are depicted in Figure 5.5.

iii. Since early recurrences are unlikely, the minimum time was set to day 14 and administrative censoring was applied on the last scheduled follow-up visit (day 63). For simplicity, no losses to follow-up were assumed.

iv. For each individual (n=500 subjects), the observed time (t) was defined as the minimum of the simulated time to recrudescence (t_1) and new infection (t_2).

$$t = \min(t_1, t_2)$$

v. The final observed time was rounded to the nearest weekly visit day (7, 14, 21 and so on), reflective of the antimalarial follow-up design.

vi. The observed event corresponded to the event with minimum time, *t.*

vii. For each simulated dataset, the cumulative probability of failure was estimated using the 1 minus Kaplan-Meier method, and the Cumulative Incidence Function.

viii. The absolute and relative differences in the two estimators derived in step (vii) were calculated.

ix. For each scenario, steps (i)-(viii) were repeated 1,000 times using an **acceptance sampling** procedure where only datasets fulfilling the study criteria were kept (e.g. 5% recrudescence, <10% new infection). Studies where 4-6%, 9-11% and 14-16% of recrudescences were observed were defined to have 5, 10 and 15% recrudescence, respectively. In order to achieve the desired proportion of recrudescences (approximately 5%, 10% and 15%), this required a large number of simulation runs, and the first 1,000 datasets fulfilling the criteria were kept for analysis.

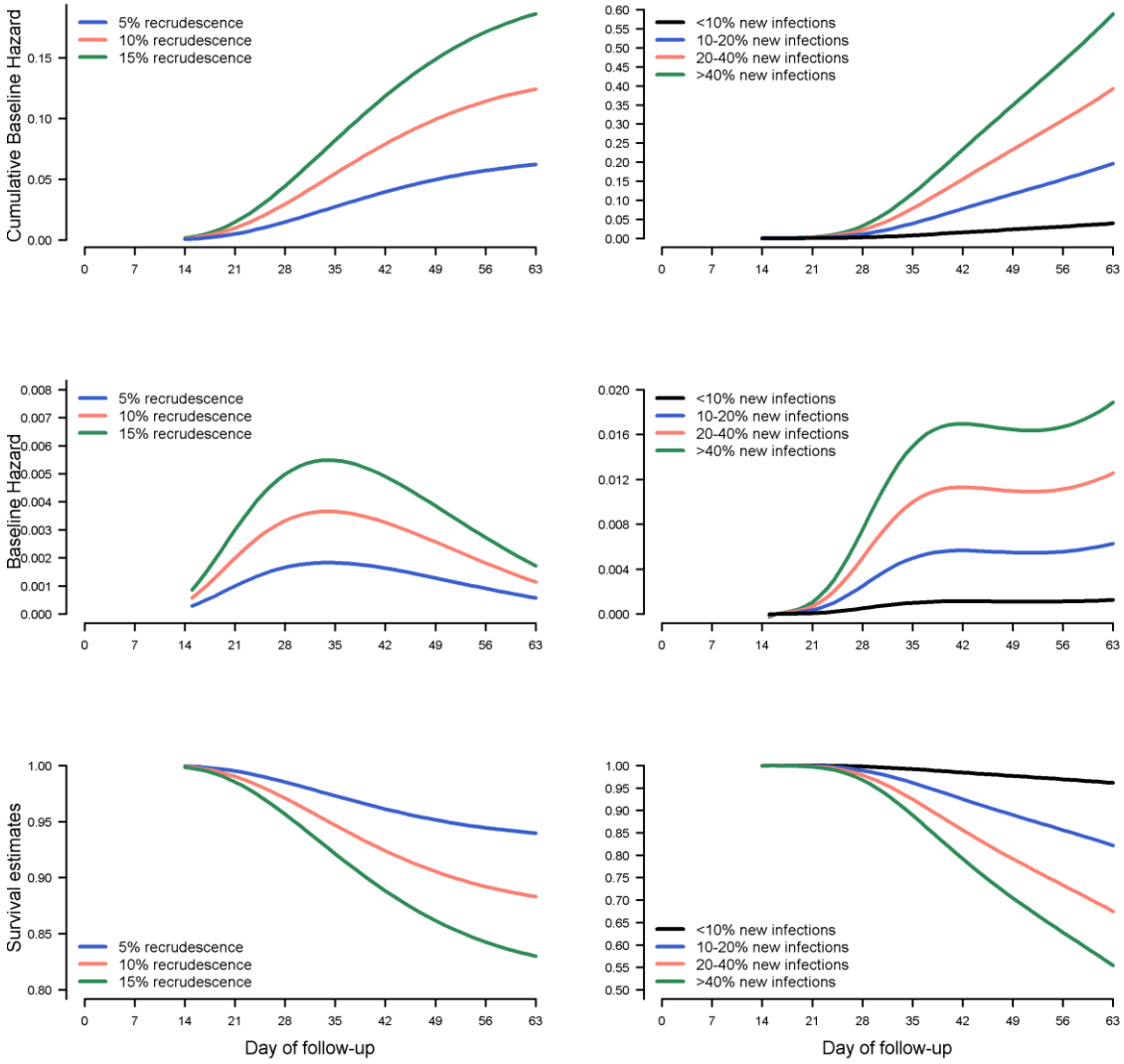


Figure 5.5: The instantaneous hazard, cumulative hazard and survival function used for simulating time to recrudescence and new infection in simulation study I

Legend: Cumulative baseline hazard for recrudescence and new infection (top panel), respective baseline hazard function (middle panel) and survival function (bottom panel) used for generating time to recrudescence and new infection for simulation study-I. The equations for the log cumulative baseline hazard corresponding to these graphs are given in Section 5.3.1.1. The middle panel is the numerical derivative of the equation used for the top panel. Note that y-axes are on different scales for each plot.

5.3.1.2 Simulation set-up for study II

Let drug A (current first line treatment) and drug B (new drug) be two antimalarials under comparison. There are 9 different possible scenarios which can be observed in terms of the effect the new drug can exert on recrudescence and new infection compared to drug A (Table 5.10). The following steps describe the simulation protocol:

i. Simulate time to recrudescence (t_1) (n=500 subjects/arm) for drug *i* using:

$$\ln\left(CBH(t)_{rc_i}\right) = \beta_{0i} - 63.6284 \times \{\ln(t)^{-1} - 0.2849\} - 0.3800 \times \{\ln(t)^2 - 12.3188\}$$

where $i = A, B$ for drug A and drug B respectively. Since drug A is the reference treatment, its intercept (β_{0i}) parameter was held constant for all the simulation scenarios. For drug B, this parameter was varied to simulate increased, decreased or null effect on recrudescences compared to drug A (Table 5.10). This corresponded to a hazard ratio of 2.72 for increased scenario and 0.37 for the decreased scenario. The hazard functions used for each of the simulations performed are presented in Figure 5.6.

ii. Simulate time to new infections (t_2) (n=500 subjects/arm) for drug *i* using:

$$\ln\left(CBH(t)_{ni_i}\right) = \alpha_{0i} + 9501.2150 \times \{\ln(t)^{-2} - 0.0858\} - 31651.33 \times \{\ln(t)^{-2} \times \ln(\ln(t)) - 0.1054\} + 29340.83 \times \{lnt^{-2} \times \ln(lnt)^2 - 0.1294\} - 12690.51 \times \{\ln(t)^{-2} \times \ln(\ln(t))^3 - 0.1588\}$$

where $i = A, B$ for drug A and drug B respectively. For drug B, the intercept parameter (α_{0i}) was varied to simulate increased, decreased or null effect on new infections compared to drug A, while for drug A it was held constant. This corresponded to a hazard ratio of 2.72 for increased scenario and 0.37 for the decreased scenario (Table 5.10, Figure 5.6).

iii. Generate observed times and event indicators as outlined in steps (iii-vi) in simulation study I (Section 5.3.1.1).

iv. For each simulated dataset, drug A and drug B were compared at the planned end of the study follow-up (day 63) using log-rank test (where new infections were censored) for testing the equality of the survival curves. To test the equality of the cumulative incidence of recrudescences between the two groups Gray's *k*-sample test was carried out. *P*-values and the associated chi-squared test statistic were extracted.

v. This was repeated 1,000 times and the proportion of times log-rank and Gray's *k*-sample test were significant at conventional 5% significance level were calculated (**rejection probability**). This is equal to the rejection of null hypothesis that there is no difference between the two treatment regimens in terms of the risk of recrudescence.

5.3.1.3 Sample size considerations

Simulation study I: A standard sample size calculation isn't relevant for the methodological comparison presented in this simulation study and hence a sample size of n=500 subjects per trial was chosen arbitrarily. In order to examine the effect of using different sample sizes, the simulation was repeated with n= 100, 200, and 1,000 patients per arm per trial. The sample size of 100 and 200 is typical of antimalarial studies.

Simulation study II: The simulation study was set-up to compare two drugs. Based on the hazard ratio (of 2.72 and 0.37) used to simulate the data, the number of patients needed per arm to detect a difference for a given log-hazard ratio (θ_R) was calculated in a two-step procedure.

Step I: In the first step, the number of events required per arm to detect the desired log-relative risk (θ_R) was calculated. The number of recrudescences (d) required per group assuming a one-one treatment allocation ratio to detect this difference is given by (Equation 15.1 in p. 473 of Collet (2015) [268]):

$$d = \frac{4(Z_{\alpha/2} + Z_{\beta})^2}{{\theta_R}^2}$$

where $Z_{\alpha/2}$ and Z_{β} are the upper $\alpha/2$- and β-points of the standard normal distribution, and θ_R is the log-hazard ratio.

Step II: In the second step, the total number of patients (N) required per arm is computed by dividing the number of events computed in Step I by the overall probability of observing recrudescence during the study follow-up. For accrual duration *a*, and follow-up period *f*, the overall probability of observing recrudescence can be computed using (Equation 15.11 in p. 479 of Collet (2015) [268]) as:

$$P(d) = \frac{1}{6}\{\bar{S}(f) + 4\bar{S}(0.5a + f) + \bar{S}(a + f)\}$$

where $\bar{S}(f)$ is the average of the estimated value of survival functions of two drugs. The accrual duration for antimalarial studies is often much longer than the follow-up period with studies often running for two to three malaria seasons. The follow-up (*f*) period, in contrast lasts only up to 9 weeks. For simplicity, *a* was assumed to be 0. The sample size required per group to detect a statistically significant difference for a given log-hazards ratio was then obtained as:

$$N = \frac{\left(4\left(Z_{\frac{\alpha}{2}} + Z_{\beta}\right)^2 \Big/ {\theta_R}^2\right)}{1 - \bar{S}(f)}$$

In simulation study II, the assumed efficacy for drug A (reference), $S_A(t)$ on day 63 was 0.940. The assumed efficacies for drug B, $S_B(t)$, for different scenarios is given in Appendix 5.2 together with the sample size required to detect the given hazards ratio with 80 and 90% power.

A sample size of 500 patients per arm was found to be adequate across all the 9 simulation scenarios studied assuming 80% power. Nonetheless, the simulation was repeated with n= 100, 200, and 1,000 patients per trial to explore the extent of variations in results which could be observed due to the differences in study sample sizes.

Table 5.10: Different scenarios for comparing two drug regimens (drug B compared against drug A) in simulation study II

		True effect size[a]	**Recrudescence parameter (β_{0i}) [b]**		**New infection parameter (α_{0i}) [b]**	
Scenario	Description		Drug A (reference)	Drug B (new drug)	Drug A (reference)	Drug B (new drug)
1	**Drug B has same effect on RC as Drug A, and**	$HR_{rc} = 1.00, HR_{ni} = 1.00$				
1A	Drug B has same effect on NI	$HR_{rc} = 1.00, HR_{ni} = 2.72$	-3.7092	-3.7092	-2.8924	-2.8924
1B	Drug B Increases NI	$HR_{rc} = 1.00, HR_{ni} = 0.37$	-3.7092	-3.7092	-2.8924	-1.8924
1C	Drug B Decreases NI		-3.7092	-3.7092	-2.8924	-3.8924
2	**Drug B has same effect on NI as Drug A, and**	$HR_{rc} = 2.72, HR_{ni} = 1.00$				
2A	Drug B increases RC	$HR_{rc} = 0.37, HR_{ni} = 1.00$	-3.7092	-2.7092	-2.8924	-2.8924
2B	Drug B decreases RC		-3.7092	-4.7092	-2.8924	-2.8924
3	**Drug B has different effect on both RC and NI relative to Drug A, and**					
3A	Drug B increases RC and increases NI	$HR_{rc} = 2.72, HR_{ni} = 2.72$	-3.7092	-2.7092	-2.8924	-1.8924
3B	Drug B increases RC and decreases NI	$HR_{rc} = 2.72, HR_{ni} = 0.37$	-3.7092	-2.7092	-2.8924	-3.8924
3C	Drug B decreases RC and increases NI	$HR_{rc} = 0.37, HR_{ni} = 2.72$	-3.7092	-4.7092	-2.8924	-1.8924
3D	Drug B decreases RC and decreases NI	$HR_{rc} = 0.37, HR_{ni} = 0.37$	-3.7092	-4.7092	-2.8924	-3.8924

RC = Recrudescence; NI = New infection; HR = Hazard Ratio

[a] Hazard ratio for recrudescence and new infections (as the ratio of the respective cause-specific hazard function) used for simulating data

[b] Coefficients used for simulating time to recrudescence and new infection for the equations providing in Section 5.3.1.2

HR_{rc} = Hazard ratio for recrudescence for drug B relative to drug A

HR_{ni} = Hazard ratio for new infection for drug B relative to drug A

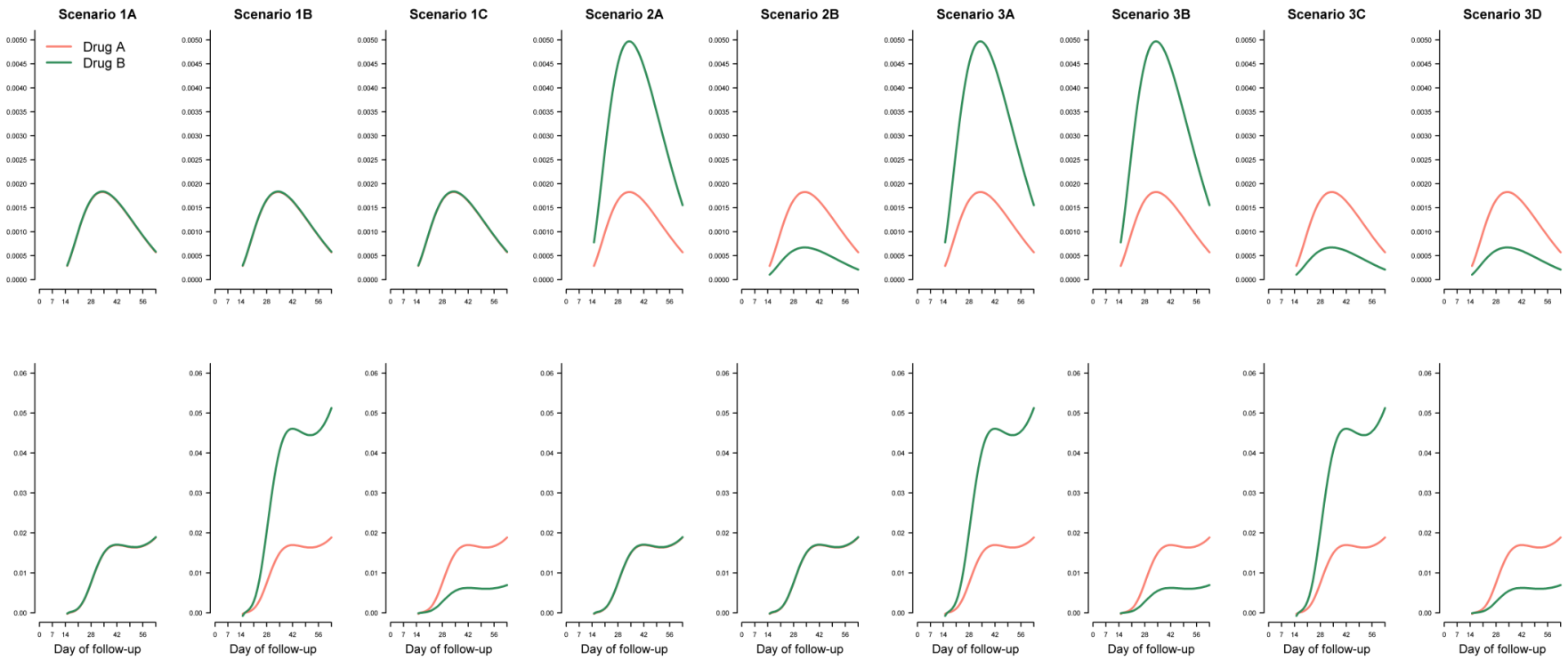


Figure 5.6: The baseline hazard function for recrudescence and new infection used for simulation study II

Legend: Top panel (recrudescence); bottom (new infection). Drug A (orange) is the reference arm and its hazard function for recrudescence and new infections is kept constant across all the simulation scenarios studied. Drug B (green) is a new regimen which is being compared against drug A. Scenario 1 (1A, 1B and 1C) is the null scenario where there is no difference in hazard function of recrudescence between these two drugs. In scenario 2, the two regimens have same hazard function for new infection, but drug B has either increased or decreased hazard of recrudescence with respect to drug A. In scenario 3, the two drugs differ in terms of both recrudescence and new infection. The baseline hazard function was derived as the first derivative of the cumulative baseline hazards presented on log-scale in Section 5.3.1.2. The description of the different scenarios can be found in Table 5.10.

5.3.2 Results

5.3.2.1 Results of simulation study I

The findings of this simulation study are presented in Figure 5.7. In the base case simulation, which assumed 5% recrudescence, the magnitude of overestimation was similar to the results from the observed data included in this analysis. The overestimation of failure by the K-M method increased with increasing proportions of new infections i.e. with increasing malaria endemicities. In the low transmission areas (<10% observed new infection), the maximum difference in the derived cumulative risk of recrudescence was 0.2% when 5% recrudescence was observed, however as the drug failure rose to 15%, the difference in estimates increased to 0.5%. In the high transmission areas (>40% new infections), the maximum absolute overestimation was 0.75% for the base case simulation and this rose to 3.1% and 4.3% when the drug failure rose to 90% and 85% respectively (Table 5.11).

The overestimation with the K-M method increased with the study follow-up duration (Table 5.11, Figure 5.7, and Figure 5.8). The results when expressed on the relative scale exhibited the same trend and conclusion as observed on the absolute scale (Appendix 5.3).The results remained unaffected when the simulation was repeated with sample sizes of n=100, 200 and 1,000 patients per trial (Appendix 5.4).

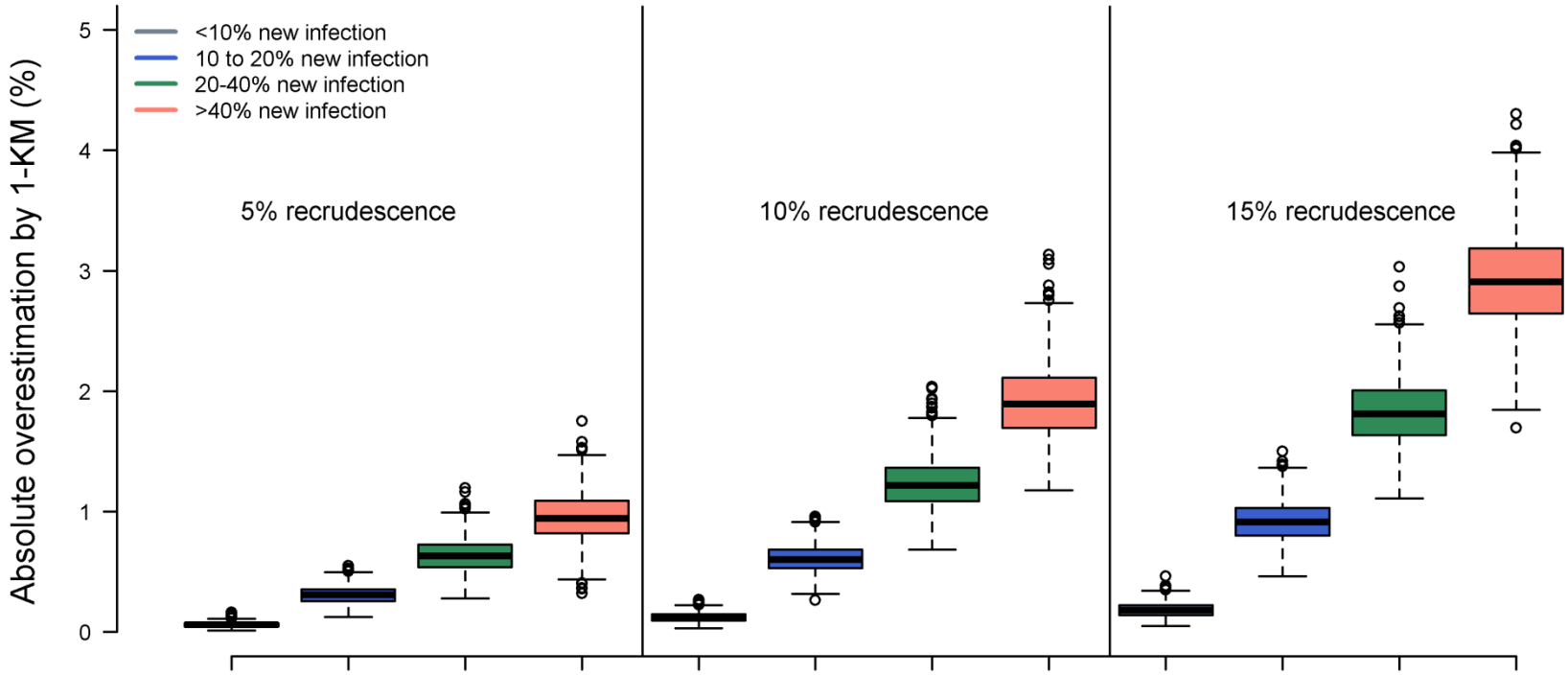


Figure 5.7: Overestimation of recrudescence failure using K-M method compared to the CIF in a simulation study I assessing standalone efficacy on day 63 (n=500 subjects)

Legend: The overestimation $(\hat{F}_{KM}(t) - \hat{F}_{CIF}(t))$ of cumulative recrudescence by the K-M method. Each panel represents different underlying status of drug efficacy on average (~5%, 10% and 15% recrudescence observed) in a study with a sample size of 500 subjects/trial. The results are presented from 1,000 independent simulation runs. The variation in absolute overestimation within each boxplot is due to varying proportion of new infections observed within the simulation scenario. Within each panel, the colours indicate different simulated scenarios of proportions of new infections: <10% new infections (grey), 10-20% new infections (blue), 20-40% new infections (green) and >40% new infections (orange), representing areas of progressively increasing malaria transmission from low to very high.

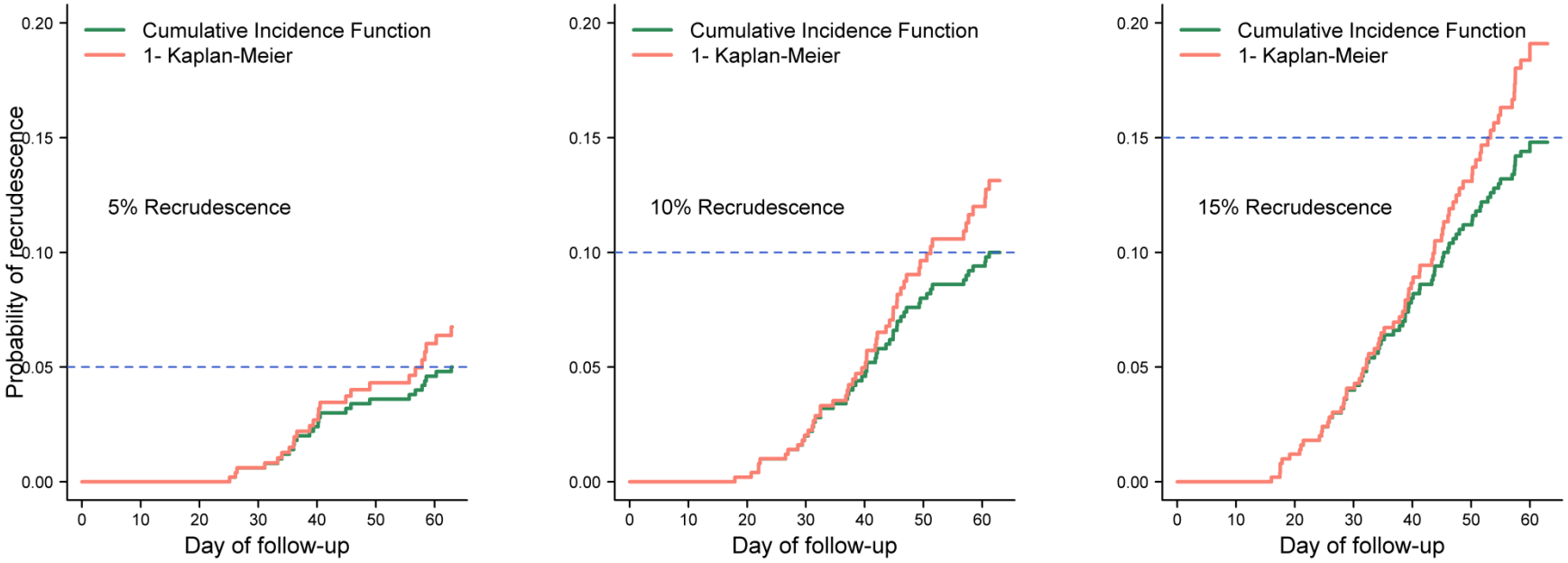


Figure 5.8: Cumulative failure estimates by study follow-up time using extreme examples from simulation study I (n=500 subjects)

Legend: The figure shows the derived cumulative estimates of recrudescence in three cases from simulation study I where the maximum difference was observed between 1-(K-M) and CIF for 5%, 10% and 15% respectively in the areas of very high transmission (>40% new infections). The absolute difference between the two estimators was 1.8%, 3.1% and 4.3% on day 63 respectively for 5%, 10% and 15% recrudescences. These three cases are the extreme cases presented in Figure 5.7 for the scenarios where >40% new infections were observed.

Table 5.11: Absolute overestimation in recrudescence failure by K-M method compared to CIF in simulation study I (n=500 subjects)

		Median absolute overestimation[IQR; Range]		
5% recrudescence	**Observed proportion of new infections[a]**	**Day 28**	**Day 42**	**Day 63**
<10% NI	3.8% [1.0 - 6.6]	0.00% [0.00 - 0.00 ; Range:0.00 - 0.01]	0.02% [0.01 - 0.02 ; Range:0.00 - 0.06]	0.06% [0.05 - 0.07 ; Range:0.01 - 0.16]
10-20% NI	17.0% [12.8 - 19.8]	0.00% [0.00 - 0.01 ; Range:0.00 - 0.02]	0.08% [0.07 - 0.10 ; Range:0.01 - 0.22]	0.31% [0.26 - 0.36 ; Range:0.13 - 0.55]
20-40% NI	31.2% [25.0 - 37.8]	0.01% [0.00 - 0.01 ; Range:0.00 - 0.04]	0.18% [0.14 - 0.22 ; Range:0.04 - 0.42]	0.63% [0.54 - 0.73 ; Range:0.28 - 1.20]
40+% NI	43.0% [40.0 - 50.0]	0.01% [0.01 - 0.02 ; Range:0.00 - 0.06]	0.28% [0.23 - 0.34 ; Range:0.09 - 0.60]	0.94% [0.82 - 1.09 ; Range:0.32 - 1.75]
10% recrudescence				
<10% NI	3.6% [1.2 - 6.2]	0.00% [0.00 - 0.00 ; Range:0.00 - 0.02]	0.03% [0.02 - 0.04 ; Range:0.00 - 0.11]	0.12% [0.10 - 0.15 ; Range:0.03 - 0.27]
10-20% NI	16.4% [10.8 - 19.8]	0.01% [0.00 - 0.01 ; Range:0.00 - 0.05]	0.17% [0.14 - 0.21 ; Range:0.05 - 0.36]	0.60% [0.53 - 0.68 ; Range:0.26 - 0.96]
20-40% NI	30.0% [24.4 - 36.2]	0.02% [0.01 - 0.02 ; Range:0.00 - 0.07]	0.36% [0.31 - 0.42 ; Range:0.13 - 0.89]	1.22% [1.09 - 1.37 ; Range:0.69 - 2.04]
40+% NI	42.0% [40.0 - 48.0]	0.03% [0.02 - 0.04 ; Range:0.00 - 0.08]	0.56% [0.48 - 0.65 ; Range:0.28 - 1.07]	1.90% [1.69 - 2.11 ; Range:1.18 - 3.13]
15% recrudescence				
<10% NI	3.4% [1.0 - 6.2]	0.00% [0.00 - 0.00 ; Range:0.00 - 0.02]	0.05% [0.03 - 0.07 ; Range:0.00 - 0.16]	0.18% [0.14 - 0.22 ; Range:0.05 - 0.46]
10-20% NI	16.0% [10.0 - 19.8]	0.01% [0.01 - 0.02 ; Range:0.00 - 0.06]	0.26% [0.22 - 0.31 ; Range:0.10 - 0.54]	0.92% [0.80 - 1.03 ; Range:0.46 - 1.50]
20-40% NI	28.8% [23.0 - 36.6]	0.02% [0.02 - 0.03 ; Range:0.00 - 0.08]	0.54% [0.46 - 0.62 ; Range:0.25 - 1.02]	1.81% [1.64 - 2.01 ; Range:1.11 - 3.03]
40+% NI	41.0% [40.0 - 45.8]	0.04% [0.03 - 0.06 ; Range:0.00 - 0.14]	0.88% [0.77 - 1.00 ; Range:0.44 - 1.60]	2.91% [2.64 - 3.18 ; Range:1.69 - 4.30]

NI= New infections. The simulation was repeated with n=100, 200, and 1,000 subjects per trial and the results are presented in Appendix 5.4.

[a] Values presented are median [Range]

5.3.2.2 Results of simulation study II

For each simulated dataset, the hazard ratio of recrudescence and new infection (for drug B relative to drug A) was estimated using a Cox model with treatment group as a covariate. The distributions of estimated hazard ratios from 1,000 simulation runs were plotted to check the simulated data (Figure 5.9).

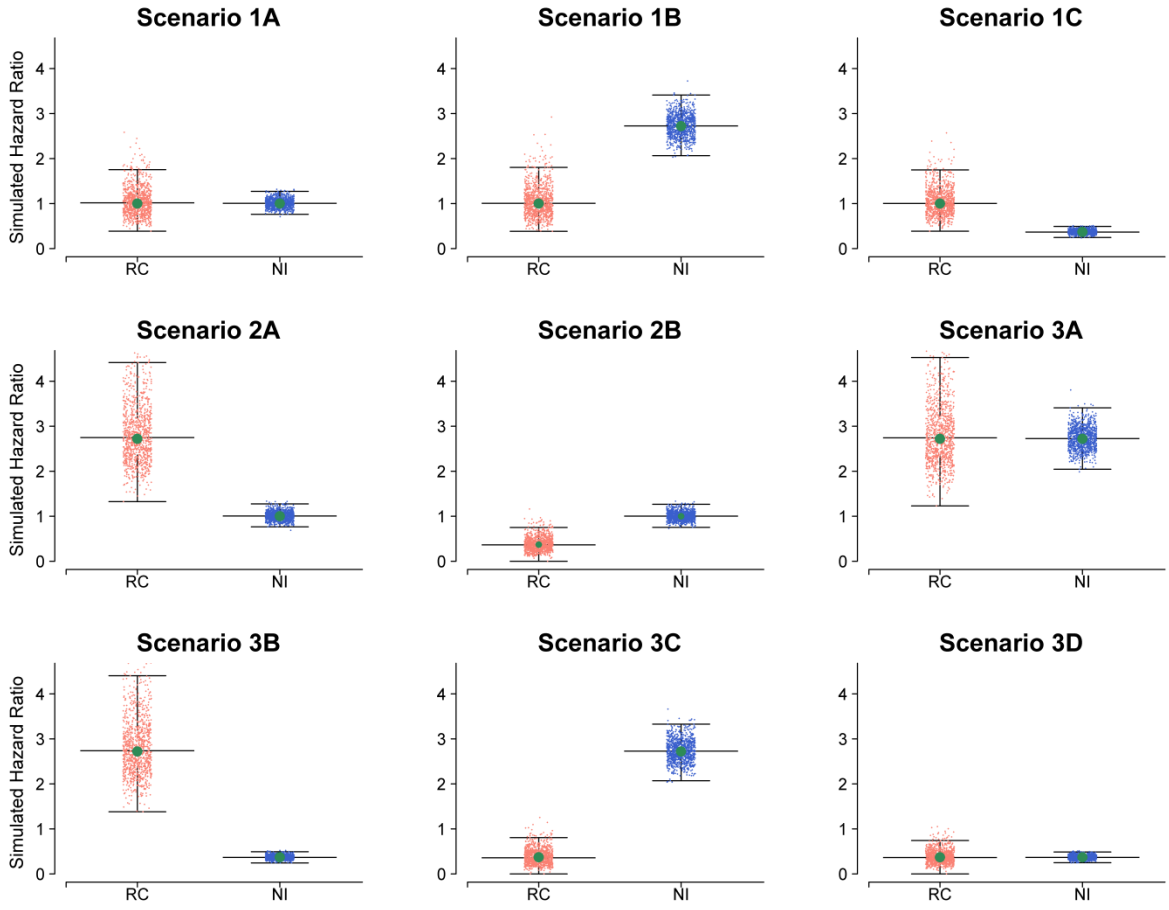


Figure 5.9: Distribution of simulated hazard ratio (n=500 subjects)

Legend: The scatterplot of estimated hazard ratio for recrudescence and new infection for drug B relative to drug A from 1,000 simulation runs. The median and interquartile range is shown. The centre green dot depicts the true hazard ratio which was used to simulate the respective datasets. RC = recrudescence, NI = New infection. The description of each of the individual scenario is provided in Table 5.10.

Table 5.12 presents the results for the different scenarios considered with a sample size of 500 subjects per arm, which had at least 80% power to detect the desired hazard ratio for recrudescence between the two drugs across all the scenarios studied.

Scenario 1: In the null situation (Scenario 1A), where it was postulated there was no difference in recrudescence and new infection between the two drug regimens, both test achieved their correct size (α) i.e. rejection rate was close to nominal 5%, as expected. Despite there being no difference between the two drugs for both events (as the respective hazard functions for recrudescence and new infections were identical for both drugs), stochastic variations will lead to a rejection of the null hypothesis approximately 5% of the time when the converse is true (See Table 5.12).

In the partially null scenario of 1C i.e. drug B had the same effect on recrudescence as drug A but was associated with decreased relative risk of new infection, both tests achieved their correct α. In Scenario 1B, where drug B was associated with increased recrudescence by a hazard ratio of 2.72, the log-rank test correctly achieved its nominal size (5% rejection), but the Gray's *k*-sample test led to a slightly higher rejection rate (11.9%). This could be explained by the fact that increased new infections prevented recrudescence from being observed (as the drug had no effect on recrudescence) as suggested by a lower observed proportion of recrudescence with drug B (median observed recrudescence = 1.9%) compared to drug A (median observed recrudescence = 2.5% recrudescence). This led to the situation where the null hypothesis was rejected more than expected.

Scenario 2: When there was no difference between the drug A and drug B in terms of their post-treatment prophylaxis but drug B was associated with increased recrudescence with a hazard ratio of 2.72 (Scenario 2A), both test had similar rejection probability close to 100%. The median proportion of recrudescence observed in this scenario was 6.5% in drug B

compared to 2.5% for drug A. In scenario 2B, where the drug B decreased recrudescence relative to drug A (hazard ratio = 0.37), both tests led to rejection of the null hypothesis 80% of the time.

Scenario 3: The most relevant and biologically plausible scenario in an antimalarial trial occurs when a new treatment exerts unidirectional effect on recrudescence and new infection (compared to the reference drug), corresponding to scenarios 3A and 3D. In scenario 3A, where drug B was associated with approximately 2-fold increase in both recrudescence and new infection, the log-rank test appeared to be the more powerful of the two approaches with rejection probability of 99% compared to 90% with Gray's *k*-sample test. In situation 3D, where drug B was associated with a median reduction in recrudescence and new infection by approximately 60%, the log-rank test again proved to be superior by rejecting the null hypothesis of no difference (between drug A and drug B) 82.8% of the time compared to 71.3% by the Gray's *k*-sample test (Figure 5.10, Panel D).

The most interesting difference was observed when drug B exerted a differential effect on recrudescence and new infection i.e. reduced recrudescence but increased new infection compared to drug A (Scenario 3C). In this situation, the Gray's *k*-sample test appeared to be more powerful of the two tests (Figure 5.10, Panel C). In Scenario 3B, where drug B was associated with increased recrudescence but reduced new infection, the results of the two tests were again very similar.

Test of proportional hazard: The log-rank test is said to be the most appropriate test under the assumption of a proportional hazard (p. 49 of [102]). In the simulation scenarios studied, the assumption of PH was violated in 5.4% (490/9,000) of the simulated datasets for the comparison of recrudescence, and 4.5% (407/9,000) for new infection. The violation of this

assumption didn't seem to affect the results of the tests as the proportion of times this assumption was violated were similar across different scenarios (Appendix 5.5).

Effect of sample size and simulation run: Increasing the number of simulation runs to 10,000 (from 1,000) didn't change the result (Table 5.12, results from 10,000 simulation runs shown in parenthesis). Hence, the simulation study was repeated with sample size of n=100, 200 and 1,000 subjects only using 1,000 runs.

In studies with n=100, and 200 (which were known to be under-powered from the sample size calculation), both tests achieved their nominal 5% level under scenario 1 (Table 5.13). In scenarios 2 and 3 (for n=100 and 200), where the hazards ratio for recrudescence between the two drugs was 2.72 and 0.37, the rejection probability was much lower than 80%. As expected, when the sample size was increased to 1,000 patients per arm, both tests correctly rejected the null hypothesis (close to 100%) for Scenarios 2 and 3, where there was difference between the two treatments in terms of recrudescence (hazard ratio of 2.72 or 0.37). Both tests achieved their nominal level in the null scenario with the exception of Gray's *k*-sample test for scenario 1B, which rejected the null hypothesis 21.7% despite there being no difference between two drugs. In this scenario, the influence of sample size was apparent as the rejection probability using Gray's *k*-sample test progressively increased with an increase in study sample size (Table 5.13).

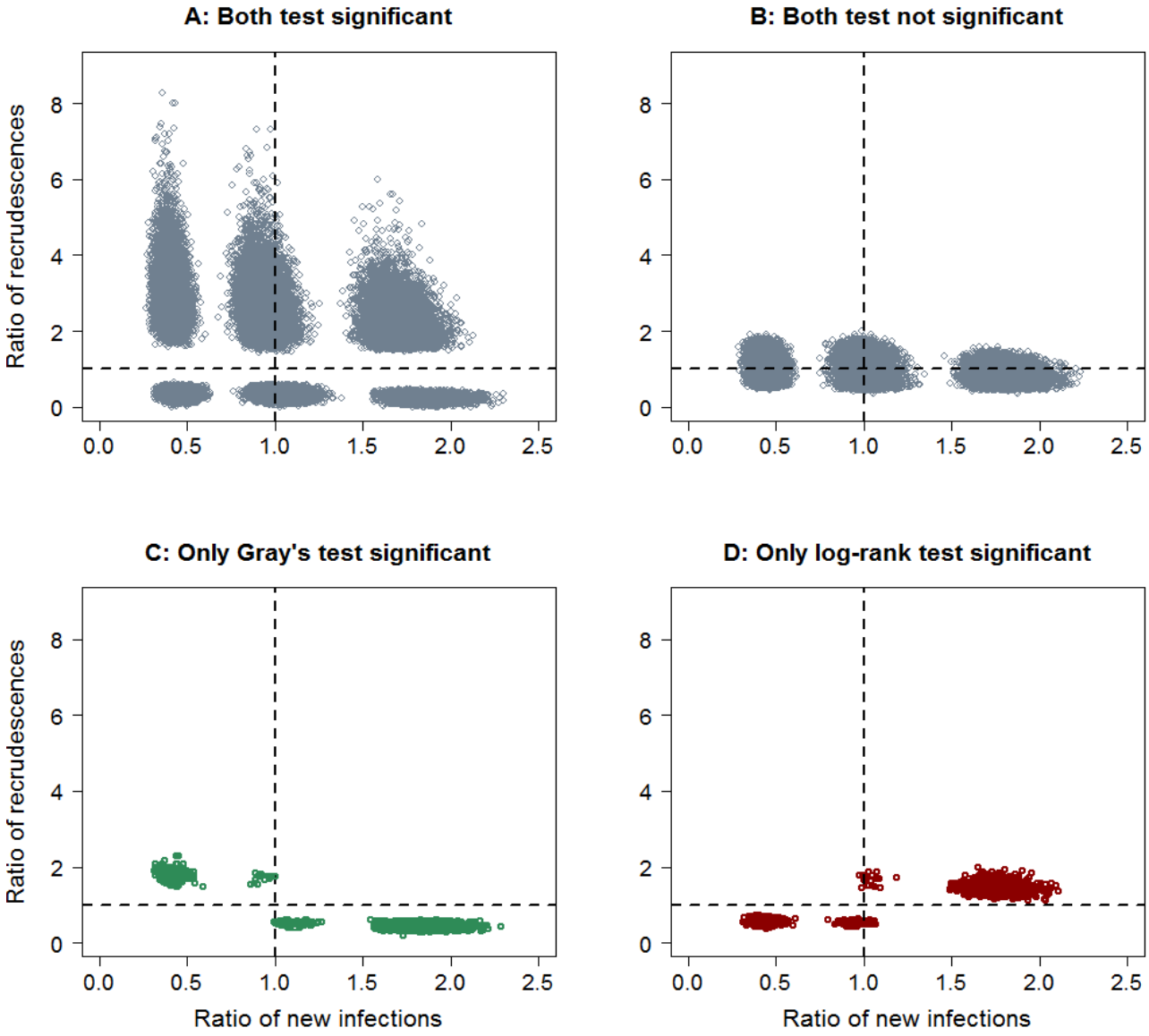


Figure 5.10: The ratio of recrudescence and new infections for drug B relative to drug A observed in simulation study II (n=500 subjects per arm)

Table 5.12: Probability of rejecting the null hypothesis at two sided 0.05 level (n=500 subjects per arm)

Scenario	True effect size from which data was simulated [a]	Median observed proportions of RC and NI in drug A [b]	Median observed proportions of RC and NI in drug B [b]	Rejection probability from 1,000 simulation runs (10,000 simulation runs)	
1. Drug B has same effect on RC as Drug A				**Log-rank test**	**Gray's test**
A. Drug B has same effect on NI	$HR_{rc} = 1.00, HR_{ni} = 1.00$	2.5% RC; 21.4% NI	2.5% RC; 21.4% NI	0.047 (0.045)	0.047 (0.045)
B. Drug B Increases NI	$HR_{rc} = 1.00, HR_{ni} = 2.72$	2.5% RC; 21.4% NI	1.9% RC; 38.6% NI	0.052 (0.048)	0.119 (0.125)
C. Drug B Decreases NI	$HR_{rc} = 1.00, HR_{ni} = 0.37$	2.5% RC; 21.4% NI	2.8% RC; 9.4% NI	0.045 (0.047)	0.062 (0.062)
2. Drug B has same effect on NI as Drug A					
A. Drug B increases RC	$HR_{rc} = 2.72, HR_{ni} = 1.00$	2.5% RC; 21.4% NI	6.5% RC; 20.0% NI	0.991 (0.996)	0.995 (0.996)
B. Drug B decreases RC	$HR_{rc} = 0.37, HR_{ni} = 1.00$	2.5% RC; 21.4% NI	0.9% RC; 22.0% NI	0.801 (0.797)	0.804 (0.797)
3. Drug B has different effect on both RC and NI relative to Drug A					
A. Drug B increases RC and increases NI	$HR_{rc} = 2.72, HR_{ni} = 2.72$	2.5% RC; 21.4% NI	5.1% RC; 36.3% NI	0.991 (0.990)	0.897 (0.896)
B. Drug B increases RC and decreases NI	$HR_{rc} = 2.72, HR_{ni} = 0.37$	2.5% RC; 21.4% NI	7.2% RC; 8.7% NI	0.996 (0.723)	0.999 (1.000)
C. Drug B decreases RC and increases NI	$HR_{rc} = 0.37, HR_{ni} = 2.72$	2.5% RC; 21.4% NI	0.7% RC; 39.5% NI	0.714 (0.723)	0.903 (0.910)
D. Drug B decreases RC and decreases NI	$HR_{rc} = 0.37, HR_{ni} = 0.37$	2.5% RC; 21.4% NI	1.0% RC; 9.6% NI	0.828 (0.820)	0.713 (0.718)

[a] Hazard ratio for recrudescence and new infections derived as the ratio of the respective cause-specific hazard function (Figure 5.6)

HR_{rc}= Hazard ratio for recrudescence for drug B relative to drug A

HR_{ni}= Hazard ratio for new infection for drug B relative to drug A

[b] median observed proportion from 1,000 simulation runs

RC=Recrudescence; NI= New infection

Table 5.13: Probability of rejecting the null hypothesis at two sided 0.05 level for different sample sizes (simulation study II)

	n=100 subjects per arm		n=200 subjects per arm		n=500 subjects per arm		n=1,000 subjects per arm	
Scenario	LR	G	LR	G	LR	G	LR	G
1A	0.043	0.042	0.055	0.045	0.047	0.047	0.042	0.040
1B	0.043	0.055	0.052	0.082	0.052	0.119	0.051	0.217
1C	0.041	0.052	0.047	0.052	0.045	0.062	0.044	0.080
2A	0.554	0.548	0.846	0.838	0.997	0.995	1.000	1.000
2B	0.198	0.187	0.391	0.395	0.801	0.804	0.982	0.983
3A	0.501	0.312	0.787	0.543	0.991	0.897	1.000	0.997
3B	0.570	0.653	0.854	0.911	0.996	1.000	1.000	1.000
3C	0.151	0.251	0.328	0.501	0.714	0.903	0.964	0.996
3D	0.231	0.168	0.422	0.353	0.828	0.713	0.988	0.959

LR= Log-rank test; G= Gray's *k*-sample test

5.3.3 Conclusions from the two simulation studies

The key findings of the two simulation studies are:

- In the areas of high transmission (>40% new infections), the K-M method overestimated failure at the end of study follow-up by up to 1.8% when drug efficacy was assumed to be 95%. As the efficacy fell to 90%, overestimation rose to 3.1%.

- In comparative studies, when a new drug influenced both recrudescence and new infection in opposite directions i.e. increased recrudescence but decreased new infection (or vice versa) compared to the reference drug, Gray's *k*-sample test was more powerful compared to the log-rank test for detecting the difference between the two drugs.

- When a new drug influenced recrudescence and new infection in the same direction (for example, increased recrudescence and increased new infection), the log-rank test was more powerful compared to Gray's *k*-sample test for detecting difference between the treatment regimens (Scenarios 3A and 3D).

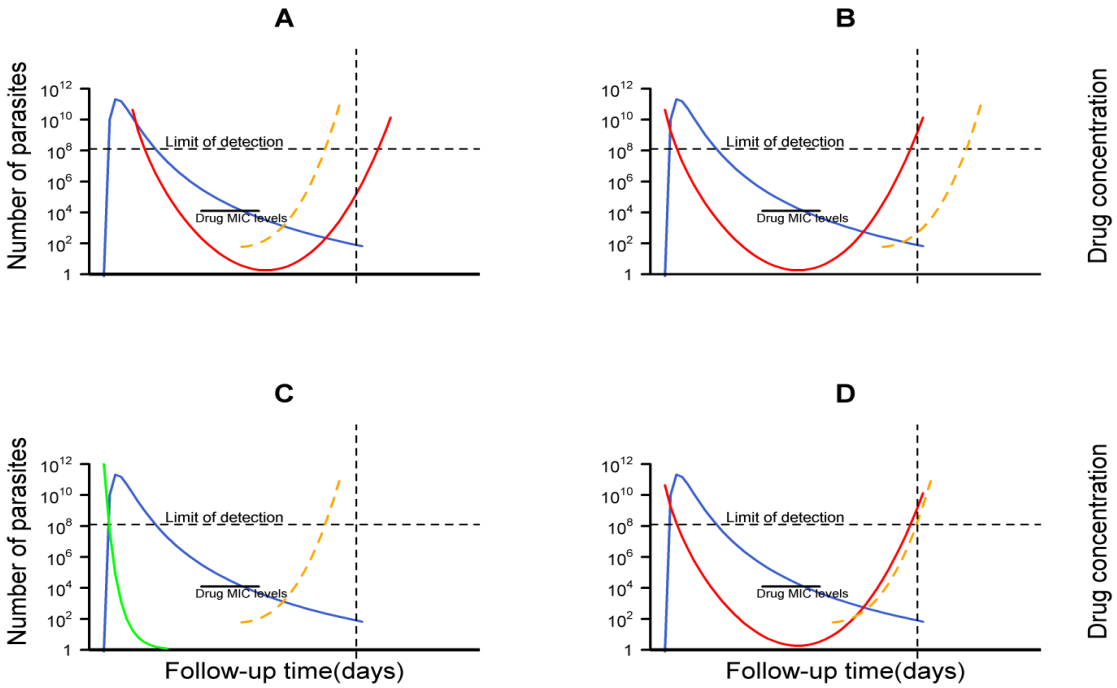


Figure 5.11: Situational competitiveness of newly emergent infections. Adapted from White-2002[58].

Legend: The blue line represents a hypothetical drug concentration of partner component, the green and red lines represent scenarios for parasite burden versus time profiles following treatment for an infection where all the parasites are completely killed resulting in cure (green) and an infection where parasites are initially killed by high drug levels but with drug levels below the minimum inhibitory concentration (MIC), net parasite growth results in subsequent recrudescence (red). The orange line represents parasite-time profiles for a new infection. The left y-axis is for parasite density, and the right y-axis shows drug levels at hypothetical units. The vertical dotted line is the administrative end of the study follow-up. The horizontal dotted line represents the microscopic limit of detection for parasites. **A)** Parasite population from a new inoculation out-competes the parasite population which caused the disease thus precluding recrudescence. In this situation, new infection is a biologically competing risk event. **B)** In this situation a new infection can be thought of as a biologically competing risk event which doesn't prevent recrudescence being observed. **C)** The parasite population which caused the disease is completely eliminated. Here, a new infection is not a competing event. **D)** In this situation, the parasite population which caused the disease and which is derived from a novel inoculation appear at the same time.

5.4 Discussion

Recent reviews have pointed out that the vast majority of studies published in high impact medical journals are susceptible to competing risk biases [84,87,269]. Malaria is no exception as just over a third (83/233) of the treatment arms from the data analysed in this chapter had an observed proportion of new infection greater than 10%, a threshold considered to make studies vulnerable to competing risk bias [255]. In the presence of competing risk events, the use of the Kaplan-Meier method for deriving failure probabilities has been shown to exhibit an upwards bias [80]. The extent of the influence competing risk events exhibit when deriving failure estimates in standalone and comparative studies of uncomplicated *P. falciparum* malaria is currently unknown.

This analysis of individual patient data from 92 clinical efficacy studies conducted in Asia, Africa, and South America explored the degree to which the failure estimates derived using the complement of Kaplan-Meier estimate was affected by competing risk events. The K-M method was associated with a marginal overestimation of the failure compared to the CIF. In only 9.0% (21/233) of treatment arms the overestimation was greater than 1%, in 4.3% (10/233) the difference was greater than 2% and in 2.6% (6/233) of the arms the difference was greater than 3%. All, but one of the 21 study sites where the difference exceeded 1% were from Africa (areas of high malarious transmission), the exception being a study from Papua New Guinea, where a very high proportion of patients experienced recurrent parasitaemia due to *P. vivax*. The degree to which K-M overestimated treatment failure was correlated with the observed proportion of new infection and recrudescence, and the follow-up duration; these findings are consistent with the existing literature [80].

The current WHO guideline recommends that when the estimates of recrudescence at the end of the follow-up exceed 10%, a series of detailed clinical, pharmacological and *in vitro*

investigations be undertaken to examine the possibility of parasite drug resistance [64]. In three study arms, the estimated failure was greater than 10% (the WHO threshold for withdrawing antimalarials) when the K-M method was used, but remained below 10% when using a competing risk survival analysis method. This highlights that ignoring competing risk events can result in potentially misleading conclusions being drawn from an efficacy trial. Furthermore, inappropriate withdrawal of a treatment regimen which is perceived to be failing has important financial ramifications given the paucity of alternative regimens currently available, and hence it is important that researchers are aware of the presence of competing risk events when drawing conclusion from the studies.

ACTs have been widely introduced over the last 15 years following documentation of high treatment failures following conventional treatment with chloroquine, sulphadoxine-pyrimethamine and mefloquine monotherapy [27]. In general, their efficacy against uncomplicated *falciparum* malaria remains high, with treatment failure reported in less than 5% of patients in the vast majority of ACT studies included in this chapter (Chapter 3, Figure 3.5). Hence, the data used in this re-analysis didn't allow the investigation of what would happen when the antimalarial efficacy declines. In order to quantify the magnitude of the potential biases in situations of falling antimalarial efficacy (now observed for dihydroartemisinin-piperaquine in Cambodia and Vietnam [44–46]), simulation studies were carried out which demonstrated that the overestimation in failure rose with an increasing proportion of competing risk events observed and declining drug efficacy. When the antimalarial efficacy declined to 90%, the absolute deviation in derived estimates reached 0.3% in the areas of low transmission and 3.1% in the areas of high transmission, suggesting that estimation of failure in areas of high transmission requires careful attention.

The effect of competing events has implications not only for standalone studies, but also on comparative drug trials, particularly when respective partner components of the ACTs are eliminated at different rates. For example, lumefantrine, the partner drug in AL, has an elimination half-life of 4 days and hence almost all antimalarial activity is sub-therapeutic within 16 days [8]. Conversely, the elimination half-life of piperaquine (partner drug in DP) is four weeks and it exerts prolonged post treatment prophylaxis, reducing the risk of recurrent infections for up to 42 days [8]. Hence, the observed proportion of competing risk events is expected to be significantly lower following DP compared to AL, especially in the areas of high transmission. When a large fraction of patients develop new infections, fewer patients are available from which recrudescences can be observed. Hence, it is important that the proportion of competing risk events be taken into consideration when comparing two regimens with different pharmacological properties. To explore whether these pharmacological differences affected the comparative analyses of these two drugs, the equality of the survival curves were compared in 27 randomised studies using the log-rank test and Gray's *k*-sample test. There were no apparent differences in the derived conclusions using these two approaches in the data considered in this analysis, as there were very few observed recrudescences.

In order to further investigate the comparison between two antimalarial regimens, simulation studies were carried out for comparing the results of the log-rank and Gray's *k*-sample test for nine different hypothetical scenarios (simulation study II). The results from the two tests were similar in most of the scenarios explored, but there were contrasting and interesting differences in two particular scenarios. When the effect of a new drug on recrudescence and new infection were both in the same direction as the reference drug, it was found that the log-rank test was more powerful compared to Gray's *k*-sample test for detecting differences between treatment regimen (Table 5.12), whereas in the scenario when the new drug influenced recrudescence and new infection in opposite directions (relative to the reference drug), Gray's *k*-sample test

was found to be more powerful. This finding is consistent with the results reported by two previous simulation studies in the statistical literature by Williamson et al (2007), and Varadhan et al (2010) [270, 271].. However, it must be stressed that the latter scenario is less likely to be observed within the context of comparing antimalarial regimens in a real-life situation.

Finally, the results from two different approaches to regression modelling in the presence of competing risks were considered: regression model on cause-specific hazard and on the sub-distribution hazard. Although the relative risk measures (hazard ratio and sub-distribution hazard ratio) obtained from these two regression models were similar (Table 5.9), they aren't directly comparable as they have a different interpretation [81,270]. In the illustrated example, the csHR of 1.67 for baseline parasitaemia implies that every 10-fold rise in parasite load was associated with a 1.67-fold higher risk of recrudescence, among patients who had not experienced any recurrence yet by the end of the follow-up. The sdHR of 1.44 (higher than 1) means that the cumulative incidence of recrudescence is 1.44-fold higher with every 10-fold increase in parasite load. Since sdHR is not a hazard ratio, the interpretation of the numeric value of 1.44 is not straight forward [270]. This is because subjects who have experienced new infections are still maintained in the risk-set when computing a sub-distribution hazard, even though they are no longer at risk of experiencing recrudescence (as illustrated in Figure 5.1). Their utility is, however, in terms of drawing inferences of the effect of covariates on the risk of incidence of the event of interest (in terms of risk prediction). In the example studied in Section 5.2.4.5, the predicted risk from the conventional cause-specific Cox model was higher than the one estimated from the Fine and Gray's sub-distribution hazard model, a result consistent with literature [87].

5.4.1 The choice of the method

What should be the correct analysis (log-rank *vs* Gray's, K-M *vs* CIF, Cox *vs* Fine and Gray) in the presence of competing risk events has gathered considerable attention in the statistical and medical literature [256,259,271–273]. It has been advocated that if the research interest is in understanding the biological mechanism of how a treatment affects recrudescence (hazard rate), the log-rank test is appropriate [80,274,275]. But when the interest is in comparing an overall risk, for example, to determine if subjects receiving a particular drug are more likely to experience recrudescence before the end of the study, the comparison of CIF through Gray's *k*-sample test is considered appropriate [80,274,275].

In terms of regression modelling, many authors advocate that if the aim of a trial is to explore the biological effect of a covariate on treatment outcome, then the cause-specific hazard model remains the correct approach [80,81,263,270]. If the aim is to estimate probability (absolute risk) and provide evidence to inform medical decision making, the use of proportional sub-distribution hazard model has been advocated [81,82,263]. Presenting results of both approaches would provide a complete biological understanding of the treatment on the different endpoints and provide further insights into the complex interplay between the events and competing risk events [80,276]. It may be a clichéd remark, but it is important that researchers are aware that the choice of the analytical method in the presence of competing risk events should be guided by the research question of interest.

5.4.2 Limitations

There were several statistical and biological limitations in this analysis. First, indeterminate outcomes were considered as censored in the K-M analysis and as an extra category of competing events when estimating the CIF. Strictly speaking, indeterminate outcomes are not competing events and treating them as an extra category for CIF analysis might have introduced

bias. A better approach might be to use the recent advances in handling missing causes of failure which utilises the sophisticated procedure of multiple imputation [277–279]. The suitability and utility of missing data approaches for handling such indeterminate outcomes will be considered in Chapter 6. Second, the simulation studies aimed at studying the impact of competing risk events assumed independence between time to recrudescence and new infections - an assumption unlikely to be verified. Simulation studies accounting for a correlation between recrudescence and new infections were beyond the scope of this project and will be considered as a future work. Similarly, the simulation studies assumed no losses to follow-up for simplicity. A loss to follow-up of approximately 20% is anticipated in antimalarial studies and this can be incorporated in the simulation studies, again as a future work. For simulation study II, the sample size calculation was based on the cause-specific hazard function for recrudescence without taking competing risk events into account. However, similar results were obtained when the simulation study II was repeated with 1,000 subjects per arm (which most likely would be powered to detect the desired hazard ratio even after taking competing risks into account).

Finally, the most important limitation is that it is difficult to disentangle whether a new infection constitutes a truly competing endpoint in a biological sense. The use of the competing risk approach is straightforward in the evaluation of fatal outcomes such as cancer mortality; where death due to cardiovascular disease will clearly represent a competing risk event. However, this is less obvious in antimalarial studies in uncomplicated *P. falciparum* malaria, where the plausibility of a new infection as a competing risk event for recrudescence most likely depends upon a range of underlying circumstances. In situations where the parasite causing the initial infection is *subdued* at a low density, and the newly emerging infection has a higher density than the *subdued* original infection, then on the assumption of an identical parasite multiplication rate for both infections, this constitutes a situation where the occurrence

of the new infection precludes recrudescence (Figure 5.11A). However, if the parasite density of the newly acquired infection is lower compared to the *subdued* infection, the occurrence of recrudescence will not be affected (Figure 5.11B). When the initial infection is completely eradicated after exposure to antimalarials, the host is no longer at risk of subsequent recrudescence and the competing risk situation doesn't occur (Figure 5.11D). In reality, it is impossible to disentangle the underlying *in vivo* parasitological circumstances from our dataset. Thus, I am of the view that new infections be considered as a "situational competing risk event" primarily dependent on the density, fitness, and the multiplication rates (efficiency) of a newly emergent infection and of the existing recrudescent parasites coupled with host immunity.

However, when the interest is in understanding new infections, it has to be noted that the recurrence due to recrudescence will always preclude them (should there be one). Hence, it is important that any inferences drawn on the post-treatment prophylaxis effect of an antimalarial drug should appropriately account for the presence of competing risk events.

5.5 Conclusions

In conclusion, competing risk events are the rule rather than exception and should be appropriately accounted for when analysing antimalarial trials. The use of the Kaplan-Meier method for deriving estimates of drug failure was associated with a marginal overestimation in the studies included in this chapter. Simulation studies demonstrated that such overestimation increased significantly with declining antimalarial efficacy and the ramifications of ignoring competing risk events were far greater in high transmission settings. In the areas of high malarious transmission where a large proportion of recurrences are attributable to novel infections, the use of CIF should be considered as an alternative approach.

Chapter 6: Dealing with indeterminate outcomes in antimalarial efficacy trials: a simulation study

6.1 Introduction

The preceding two chapters explored methodologies for dealing with recrudescences and new infections when deriving the estimates of PCR adjusted drug efficacy. This chapter is concerned with statistical approaches for dealing with indeterminate cases of parasitic recurrences which arise when the paired analysis of pre and post treatment parasite genotypes cannot reliably determine the cause of parasitic recurrence.

The current WHO recommendation for dealing with indeterminate outcomes when deriving antimalarial efficacy is to exclude them in a complete case (CC) analysis [64]. There are two major concerns with this approach. First, this approach leads to a loss in sample size and hence leads to a potentially inefficient analysis. Second, the validity of this approach relies on an implicit assumption that indeterminate outcomes are a representative random sample of the recurrences; an assumption which cannot be verified. The CC analysis is usually supplemented with two extreme sensitivity analyses representing worst (or best) scenarios, where all indeterminate recurrences are imputed to be a recrudescence (or new infection). Such *ad hoc* imputation approach considers imputed datum as the true value and this yields inferences that are over-precise (standard errors are too small) rendering the associated hypothesis tests as invalid [280–282]. An alternative approach for dealing with such indeterminate outcomes is to consider these events as missing, which allows exploring the suitability of some of the statistically principled approaches developed for handling missing data.

6.1.1 Introduction to missing data

Missing data are values which were intended to be recorded in a study. Missing data are inevitable in a clinical study and may arise due to one of the three mechanisms postulated by Rubin [284, 285]: missing completely at random (MCAR), missing at random (MAR) and missing not at random (MNAR). The missingness mechanism is said to be MCAR if the missingness is independent of any observed and unobserved data. When the missingness doesn't depend on any unobserved data conditional on observed data, it is said to be a MAR mechanism. When the missingness is dependent on the unobserved observation itself, then it is said to be MNAR.

In an antimalarial study, different missingness mechanisms can lead to a recurrent parasitaemia being indeterminate. An indeterminate outcome could arise due to failure in amplification of the parasitic DNA, due to the loss of PCR samples or due to a low parasite count [51,283]. This can potentially occur in a truly random process (i.e. independently of any observed or unobserved data) among patients with recurrent infections. Alternatively, the missingness may be conditional on some observed data (but not on any unobserved data) such as transmission intensity, the number of molecular markers used, and /or the density of the parasites on the day of recurrence [283]. In addition, the probability of a recurrence being indeterminate may equally depend on some unobserved variables (for example, background allelic diversity of the parasitic strains, which are rarely reported) even after conditioning on the observed data. Insights regarding the mechanisms which could have led to missing values are often unknown and cannot be addressed based on observed data [284]. Hence, all statistical analyses in the presence of missing data rely on the validity of some untestable assumption regarding the mechanism which led to the generation of missingness. Generally, MAR missingness is assumed in the standard implementation of sophisticated approaches for handling missing data, such as multiple imputation (MI).

6.1.1.1 Multiple Imputation (MI)

The basic principle of MI, first proposed by Rubin in 1977, is to impute the missing values based on the distribution of the observed data and repeat this *M* times in order to account for the uncertainty in the missing values [285,286]. This is essentially a two-step procedure. In the first step, incomplete data are estimated multiple times from a suitable imputation model in which values are drawn from the posterior predictive distribution (**Imputation Step)** [287]. In the second stage (**Analysis Step**), the substantive model (target analysis) of interest is carried out on each of the individual completed datasets (observed plus imputed data). The final estimates and standard errors are then derived by combining estimates across each of the multiply imputed datasets using Rubin's combination rules, which incorporate uncertainties within and between imputations [286]. The within-imputation variability reflects uncertainty about the results from one imputed dataset and the between-imputation variability accounts for uncertainty due to missing information; thus rectifying the disadvantages associated with single imputation. When correctly implemented, MI leads to asymptotically unbiased estimates and standard errors [288]. Further details regarding the Rubin's rules are outlined in Section 6.3.2.

6.1.1.2 Inverse probability weighting (IPW)

The results derived from MI are frequently compared to those obtained from the inverse probability weighting (IPW) approach [289]. In this approach, complete cases are weighted by the inverse of their probability of being a complete case, i.e. up-weighting the data from participants who have a low probability of being observed thus creating a pseudo-population [289]. The final analysis is then carried out using only the complete observations (indeterminate outcomes are not included), but with weighting to rebalance the set of complete cases so that it is representative of the whole sample. The idea of bias correction by weighting the complete cases by the inverse of the probability of being observed was put forward by Horvitz and Thomson formally in 1952 [290]. This was further popularised by Rosenbaum and

Rubin in 1983 when they used this approach to correct bias in observational studies and has now found application in the causal inference literature (called propensity score modelling) [291]. Like MI, the IPW approach is also a two-step estimator. In the first step, a missingness model is constructed to estimate the probability of an observation being a complete case. This describes the process by which the observations are filtered or selected to produce the observed sample of data. In the second step, the inverse of the probability that an observation is a complete case is used as a weight in the target analysis.

6.1.2 Review on the usage of MI and IPW in estimating antimalarial efficacy

MI and IPW approaches for handling missing data have been increasingly used in the medical and statistical literature in the past decade [289,292]. Yet, only a handful of studies have explored the suitability of these methods when dealing with indeterminate outcomes in the estimation of antimalarial efficacy. There has been only three studies as of date (to my knowledge), which have explicitly used these approaches when deriving antimalarial efficacy in uncomplicated *P. falciparum* malaria [88,293,294].

The first study, by Machekano et al (2008) [88] carried out a comprehensive simulation study and compared the performance of the MI and IPW approaches against the CC analysis for estimating cured proportions in a randomised study in Uganda. Their study demonstrated that the CC approach overestimated the cured proportion and was inefficient compared to the MI and IPW methods. The second study by Mukaka et al (2016) [293] used a simulation study to explore the use of MI when handling an incomplete binary outcome (using an antimalarial trial from Malawi as the motivation) to estimate the risk difference between treatment regimens. More recently, a large multi-centre trial carried out in pregnant women (the PREGACT study-2017) [294] used MI as a sensitivity analysis to assess the robustness of the derived estimate of cured proportions. However, no study to date has compared the utility of the MI and IPW

approaches in handling indeterminate outcomes when deriving Kaplan-Meier (K-M) estimates of drug efficacy for artemisinin combination therapies (ACTs).

6.1.3 Aims and objectives

The aim of this chapter was to compare the estimates of antimalarial drug efficacy derived from the MI and IPW approaches against those from the CC analysis.

There were two specific objectives:

I. To compare the performance measures (bias, efficiency, coverage and accuracy) of the K-M estimates derived from Multiple Imputation (MI) and Inverse Probability Weighting (IPW) against those from the complete case (CC) analysis using The 4ABC Trial as the motivational study [139], and

II. To assess the applicability of MI and IPW in small sample antimalarial studies (n=100, n=200 subjects per trial).

The rest of the chapter is structured as follows: Section 6.2 provides a brief description of the motivational study used (The 4ABC Trial); Sections 6.3, 6.4, and 6.5 describe the design, analysis and results of the simulation study designed to achieve objective I. Section 6.6 presents a resampling case-study designed to assess the applicability of MI and IPW in regular antimalarial studies with small sample sizes (n=100 and 200) (objective II). Section 6.7 provides a discussion of the results obtained from the simulation and resampling studies. The chapter is concluded with recommendations in Section 6.8.

6.2 Motivating study: The 4ABC Trial

The 4ABC Trial was a large multi-centric study (4,116 patients enrolled) conducted at 12 sites in seven sub-Saharan African countries between 2007 and 2009 [295]. Four antimalarial regimens were assessed: artemether-lumefantrine (AL), artesunate-amodiaquine (ASAQ), dihydroartemisinin-piperaquine (DP) and chlorproguanil-dapsone-artesunate (CDA). Patients were followed actively for 28 days. CDA was discontinued partway through the study due to reports of severe haemolysis from another phase III study [296]. For this reason, only data from the AL, ASAQ and DP arms were considered in the current analysis. Furthermore, for the purpose of this work, only patients with defined outcomes (cured, or had a recurrent parasitaemia) were kept for the analysis (early treatment failures and losses to follow-up were excluded). There were a total of 598 (17.4%, 598/3,431) recurrences of which 81 were recrudescences, 455 new infections and 62 indeterminate outcomes after performing PCR genotyping (Table 6.1). While the percentage of indeterminate outcomes out of the total sample size is low (1.8%, 62/3,431), this comprises 10.4% (62/598) of the total recurrences.

Table 6.1: Antimalarial treatment outcomes for the 4ABC Trial[139]

Treatment	Cured	Missing outcome (indeterminate PCR)	Recrudescence	New Infection	Total
AL	847 (73.0%)	29 (2.5%)	41 (3.5%)	243 (21.0%)	1,160
ASAQ	744 (81.8%)	20 (2.2%)	18 (2.0%)	127 (14.0%)	909
DP	1,242 (91.2%)	13 (1.0%)	22 (1.6%)	85 (6.2%)	1,362

AL= artemether-lumefantrine, ASAQ = artesunate-amodiaquine, DP = dihydroartemisinin-piperaquine

In order to inform the design of the simulation study reported in Section 6.3, baseline variables which were associated with indeterminate outcomes were identified (for simulating covariate dependent missingness) and additional predictors on the day of recurrence were also identified (auxiliary variables). Among the study population (n=3,431), two baseline variables were associated with a decreased odds of indeterminate outcomes in a univariable analysis carried out using logistic regression with study site fitted as random effects: areas of low transmission settings [Odds Ratio (OR): 0.22; 95% Confidence Interval (CI): 0.08 - 0.67, compared to high settings], and treatment with DP regimen (compared to AL) [OR: 0.36; 95% CI: 0.19 - 0.71]. These relationships stayed in the multivariable analysis (Table 6.2).

Table 6.2: Baseline predictors of indeterminate outcomes (n=3,431)

	Observed proportion[a]	**Univariable analysis**		**Multivariable analysis**	
Baseline variables		Odds Ratio (OR) [95% CI]	*P*-value	Adjusted OR [95% CI]	*P*-value
Age (y)	1.8% (62/3,431)	0.97 [0.78 - 1.21]	0.796	0.99 [0.79 - 1.25]	0.929
Log(baseline parasitaemia) (for every 10-fold increase)	1.8% (62/3,431)	1.01 [0.81 - 1.24]	0.963	1.01 [0.81 - 1.26]	0.941
Haemoglobin (g/dL)	1.8% (62/3,431)	0.97 [0.81 - 1.16]	0.748	1.00 [0.83 - 1.21]	0.988
Baseline fever (>37.5) No (ref)	1.8%(27/1,470)	1	-	1	-
Yes	1.8%(35/1,961)	1.07 [0.64 - 1.78]	0.789	1.12 [0.65 - 1.91]	0.690
Transmission (ref= High) [b]	2.5% (42/1687)	1	-	1	-
Moderate	1.5% (15/1,049)	0.61 [0.3 - 1.21]	0.158	0.69 [0.31 - 1.56]	0.372
Low	0.6% (4/695)	0.22 [0.08 - 0.67]	0.007	0.23 [0.07 - 0.71]	0.011
Treatment (ref= AL)	2.5% (29/1160)	1	-	1	-
ASAQ	2.2%(20/909)	0.90 [0.47 - 1.72]	0.755	0.88 [0.47 - 1.65]	0.679
DP	1.0%(13/1,362)	0.36 [0.19 - 0.71]	0.003	0.37 [0.19 - 0.72]	0.004
Gender (ref=Female)	1.8%(29/1619)				
Male	1.8%(33/1812)	1.01 [0.61 - 1.66]	0.981	1.01 [0.61 - 1.68]	0.960

[a] Observed proportion of indeterminate outcomes; [b] Defined based on the MAP estimates
AL = artemether-lumefantrine; DP = dihydroartemisinin-piperaquine; ASAQ = artesunate-amodiaquine; ref = reference category

6.3 Simulation study: design and analysis

For the purpose of the simulation study, the 62 indeterminate outcomes were excluded and the remaining data (n=3,369 of which 536 were recurrences) were considered as a complete dataset (referred to as the **full data** here onwards). Missingness was artificially induced to the full dataset under different mechanisms to study the performance of different methods for handling them. The terms missing completely at random (MCAR) and missing at random (MAR) are widely used in the statistical literature to describe the missingness mechanisms (Section 6.1.1). Since missingness of outcomes in antimalarial studies is conditional on a recurrence being observed, I restrain from using the standard terminology used in the literature and refer to missingness with the term "*mechanism*" instead.

6.3.1 Study design

The missingness was simulated by repeatedly setting a proportion of the recurrent cases as missing under two different mechanisms (mechanisms 1 and 2, description to follow). The following proportion of the recurrent cases was set to be missing: 10, 30 and 45%, and for each of these missing scenarios 1,000 datasets were simulated. The value of 10% was chosen to mimic the percentage of indeterminate recurrences observed in the motivational The 4ABC Trial, while 30% and 45% were chosen to represent moderate and extreme scenarios, occasionally observed in standard antimalarial studies.

Missingness was induced on the resampled (bootstrap resampling) observed recurrent cases (Figure 6.1). The bootstrap resampling was carried out with replacement using the 536 recurrent cases (284 with AL, 145 with ASAQ and 107 with DP) and this was carried out by treatment regimen so that there were exactly 284 recurrences with AL, 145 with ASAQ and 107 with the DP regimen observed in each resampled data.

The decision to resample the recurrent cases only, prior to assigning missingness, was based on the performance measures of the estimator derived when the missingness was simply introduced on the same (not resampled) recurrent samples to create the 1,000 datasets (Appendix 6.1). This original approach resulted in smaller empirical standard errors (relative to model-based standard errors) suggesting that there was little between simulation variation (Appendix 6.1).

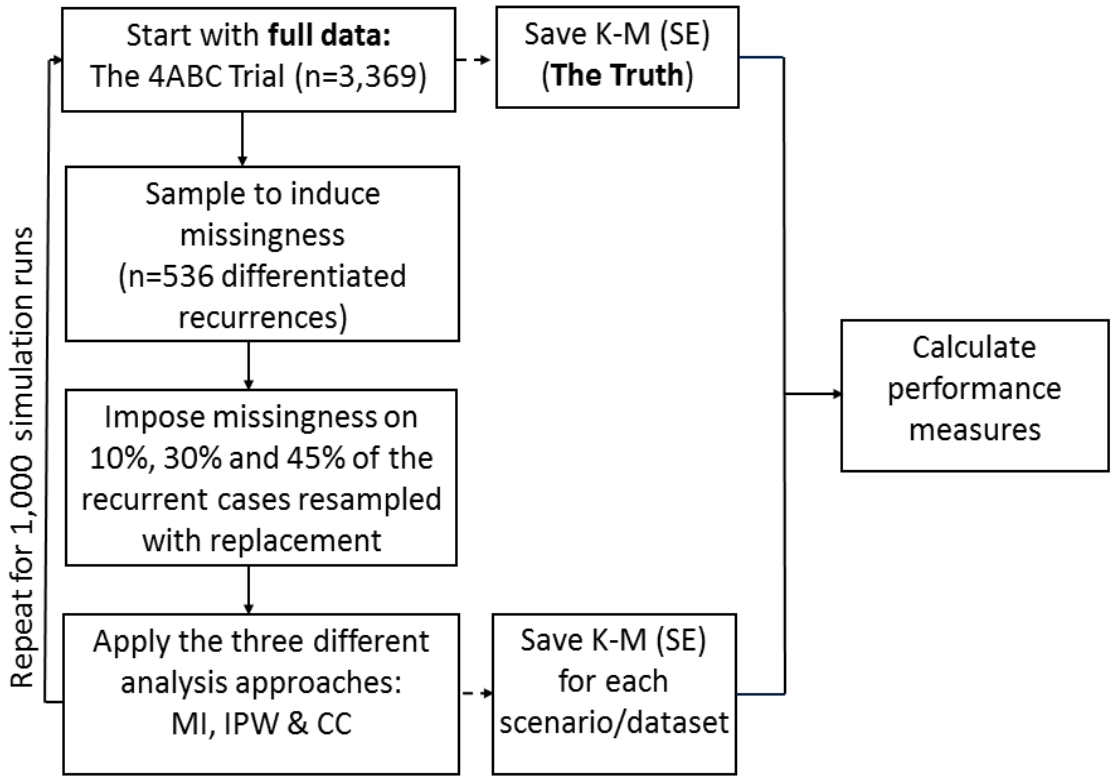


Figure 6.1: The design of the simulation study

Legend: MI = Multiple Imputation; IPW = Inverse Probability Weighting; K-M = Kaplan-Meier estimate; SE = Standard Error. The truth was defined as the estimates obtained from the full data before missingness was induced.

6.3.1.1 Simulation of missingness under mechanism 1

For mechanism 1, it was posited that missingness was a truly random process among the 536 patients who observed parasite recurrence, thus, 10, 30, and 45% of the patients with parasite recurrence were randomly reclassified as having indeterminate outcomes [293].

Missingness was induced as follow:

i. For each subject i with recurrent parasitaemia, generate a random number (u_i) from the uniform distribution [0,1]

$$u_i \sim U\,[0,1] \qquad\qquad i = 1,2,3,\ldots,536$$

ii. The desired proportion (p) of the smallest u_i were set as having their outcome as missing; $p = 0.10, 0.30, 0.45$

6.3.1.2 Simulation of missingness under mechanism 2

For mechanism 2, the probability of an indeterminate outcome was assumed to depend on transmission intensity and treatment regimen (covariate dependent missingness). This assumption was based on regression modelling of the original dataset from The 4ABC Trial (62 indeterminate recurrences, 81 recrudescences, 455 new infections and 2,833 cured) to determine the predictors associated with indeterminate outcomes (Table 6.2). Under this generating model, patients treated with ASAQ and DP were progressively less likely to be assigned indeterminate outcomes, reflecting the increasingly longer terminal elimination half-lives with these drug regimens which prevents some of the recurrences to be fully observed by day 28; and similarly, patients in the moderate and low transmission settings were progressively less likely to have indeterminate outcomes compared to those in high settings as the PCR method is less likely to fail as clonal competition is lower due to reduced parasitic diversity in low transmission areas.

Missingness was induced as follow:

i. For each subject i with recurrent parasitaemia, the probability of their outcome being missing (π_i) was estimated using the generating model based on the treatment regimen received and the transmission level of the site the subject came from. With missingness indicator R_i for an individual i, the following logistic regression model was used as a generating model:

$$logit\big(p(R_i = 1)\big) = \beta_0 + \beta_{1k} * transmission_{ik} + \beta_{2j} * treatment_{ij}$$

where *k*=1 and 2 for low and moderate transmission (with high settings as a reference), respectively, and *j*=1 and 2 for antimalarial treatments ASAQ and DP, respectively (with AL as a reference arm).

ii. Generate a Bernoulli outcome (y_i) for subject i based on their missingness probability (π_i)

$$y_i \sim Bernoulli(\pi_i) \qquad i = 1,2,3,\dots,536$$

$$y_i = 1\ if\ missing, 0\ otherwise$$

Two different scenarios were further studied: **weak and strong**. For the weak scenario (**mechanism 2a**), the strength of the association between transmission, treatment and missingness was halved to that of the strong scenario (**mechanism 2b**) to represent a more extreme case. The coefficients (β_{1k}, β_{2j}) used for assigning missingness for the weak scenario and strong scenario are detailed in Table 6.3.

The intercept of the logistic regression (β_0) was chosen by iteration to achieve the desired proportions of missing outcomes as outlined below:

Mechanism 2a

$logit\big(p(R_i = 1)\big) = -2.10 + \psi$ for approximately 10% missingness

$logit\big(p(R_i = 1)\big) = -0.75 + \psi$ for approximately 30% missingness

$logit\big(p(R_i = 1)\big) = -0.11 + \psi$ for approximately 45% missingness

where $\psi = \beta_{1k} * transmission_{ik} + \beta_{2j} * treatment_{ij}$

Mechanism 2b

$logit\big(p(R_i = 1)\big) = -2.03 + \psi$ for approximately 10% missingness

$logit\big(p(R_i = 1)\big) = -0.68 + \psi$ for approximately 30% missingness

$logit\big(p(R_i = 1)\big) = -0.02 + \psi$ for approximately 45% missingness

where $\psi = \beta_{1k} * transmission_{ik} + \beta_{2j} * treatment_{ij}$

Table 6.3: Specification of the logistic regression model used to impose missingness

	Log (Odds Ratio)		Odds Ratio	
	Mechanism 2a	Mechanism 2b	Mechanism 2a	Mechanism 2b
Treatment	β_{2j}	β_{2j}	$\exp(\beta_{2j})$	$\exp(\beta_{2j})$
AL (reference)	0.00	0.00	1.00	1.00
ASAQ	-0.05	-0.10	0.95	0.90
DP	-0.20	-0.40	0.82	0.67
Transmission	β_{1k}	β_{1k}	$\exp(\beta_{1k})$	$\exp(\beta_{1k})$
High (reference)	0.00	0.00	1.00	1.00
Moderate	-0.15	-0.30	0.86	0.74
Low	-0.25	-0.50	0.78	0.61

AL= artemether-lumefantrine; ASAQ = artesunate-amodiaquine; DP=dihydroartemisinin-piperaquine

6.3.2 Analysis of simulated data

Three different approaches were used for handling missing data: complete case (CC) analysis, multiple imputation (MI) and inverse probability weighting (IPW). Each of these three methods was applied to the same 1,000 independent samples generated. For each simulation dataset, the target estimand of Kaplan-Meier estimate and associated standard errors were extracted and stored. The three different approaches are briefly described here.

6.3.2.1 Complete case (CC) analysis

In the complete case approach, outcomes which were assigned as missing were excluded from the K-M survival analysis.

6.3.2.2 Multiple Imputation (MI)

Missing outcomes were imputed using a logistic regression restricted to patients with recurrent parasitaemia (recrudescence *vs* new infection) and including all the variables in the target analysis as covariates in the imputation model [297]. In addition, predictors of missingness (time of recurrence and the parasite density on the day of recurrence) were also added to the imputation model as auxiliary variables to ensure that the assumption of missing at random

(assumed for MI) was more plausible and to reduce the between-imputation variability [298,299]. Since the study was carried out in multiple centres, study site was also included in the imputation model (Table 6.4). For each of the analyses, 50 multiply imputed datasets were created and the derived estimates of K-M and their associated standard errors were pooled using Rubin's rules to obtain an overall MI estimate [287]. The number of imputations ($m = 50$) was selected following the recommendation that m should be at least equal to the percentage of missing cases when the fraction of missing information is less than 50% [288,300].

With $\boldsymbol{M}$ sets of imputations, let $\widehat{S_m}$ be the estimate of the Kaplan-Meier (K-M) survival probability for the m^{th}imputation, *m*=1,2,3,..,M. The final survival estimate was then obtained by averaging over all M imputations as $\bar{S}(t) = \sum_{m=1}^{M} \frac{\widehat{S_m}}{M}$. The variance of $\bar{S}(t)$ was estimated by accounting for within-imputation (W) variability and between-imputation (B) variability as:

$$Var\big(\bar{S}(t)\big) = W + \left(1 + \frac{1}{M}\right) B$$

where $W = \sum_{m=1}^{M} \frac{w_m}{M}$; $w_j = var(\widehat{S_m})$; $B = \frac{1}{(M-1)} \sum_{m=1}^{M} (\widehat{S_m} - \overline{S(t)})^2$

Rubin's combination rules require that the estimated parameter are asymptotically Normally distributed, therefore the K-M derived estimates were complementary log-log transformed (cloglog) $\{\log\big(-\log\big(S(t)\big)\big)\}$ [288,301,302]. An estimate of the variance of the cloglog transformed K-M was obtained using Taylor's series approximation given by the expression in Appendix 2.1. Since the missingness was restricted to recurrences only, imputation wasn't carried out separately by treatment arms; rather an overall imputation was carried out.

6.3.2.3 Inverse Probability Weighting (IPW)

A missingness model was constructed to estimate the probability of a patient having an observed treatment outcome (cure/recrudescence/new infection) using a logistic regression (Table 6.4). Patients who were cured (i.e. did not observe any recurrence) received a weight of one, while those who had a recurrence status received weights which were the inverse of their estimated missingness probability. For the estimate of the standard errors of the IPW approach to be valid, uncertainties regarding the estimation of the weight should be fully accounted for. Therefore, bootstrapping with 200 resamples was undertaken to obtain the standard error as described by Austin et al (2016) [303].

Efficacy studies for antimalarials are unique in that indeterminate outcomes can arise only if a patient experiences parasitic recurrence - and thus recurrence is a perfect predictor. Two different strategies for handling this perfect predictor were considered:

- Including recurrence status as a predictor in the missing model (**IPW**)
  (referred to as the **standard IPW** approach henceforth in this thesis**)**
- Excluding recurrence status as a predictor in the missing model (**IPW-E**)

Table 6.4: Outline of the imputation model and missingness model

Model	Response	Predictors
Imputation Model	$Y = \begin{cases} 1 & if\ recrudescence \\ 0 & if\ new\ infection \end{cases}$	• age (years) • mg/kg dose of partner drug • transmission intensity [a] • treatment regimen • time of recurrence • parasitaemia (log) • study sites [b] • parasite density (log) on the day of recurrence
Missingness model	$Y_{obs} = \begin{cases} 1 & if\ outcome\ is\ observed \\ 0 & if\ outcome\ is\ indeterminate \end{cases}$	• age (years) • mg/kg dose of partner drug • transmission intensity [a] • treatment regimen • time of recurrence • recurrence status (yes/no) [c]

[a] Transmission settings were categorised as low if Malaria Atlas Project estimate were less than or equal to 0.10, moderate if >0.10 and ≤ 0.40 and high if greater than 0.4. [b] Study site was not added in the missingness model as it led to convergence issues. [c] Excluded in the IPW-E approach.

6.3.3 Performance measures for evaluating different methods

The performance measures for different missing data methods were evaluated by comparing the derived estimates against the "true" values. A previous resampling study had defined true values as the average estimates obtained from 20,000 bootstrap replications [304]. The Kaplan-Meier estimates and their standard errors derived from the "full data" before inducing missingness (full data estimates), together with the estimates obtained from 20,000 bootstrap samples are presented in Table 6.5. The estimates obtained from both approaches were similar and the full data estimates derived from 3,369 available cases are referred to hereafter as the "true" values against which the results of the simulation study will be compared in this chapter.

Table 6.5: Kaplan-Meier estimates of antimalarial efficacy on day 28 from the full data compared to that obtained from 20,000 bootstrap samples

	Full data estimate (N= 3,369)			20,000 bootstrap samples		
Treatment	K-M	SE	SE †	K-M	SE	SE †
AL (n=1,131)	0.960	0.0061	0.1567	0.960	0.0061	0.1579
ASAQ (n=889)	0.979	0.0049	0.2358	0.979	0.0048	0.2410
DP (n=1,349)	0.983	0.0035	0.2132	0.983	0.0035	0.2213

†Standard error after complementary log-log (cloglog) transformation of Kaplan-Meier estimate. The cloglog transformation was applied as the MI estimates were computed on complentary log-log scale for the application of Rubin's combination rules to be valid; K-M = Kaplan-Meier estimates; SE = standard error; AL = artemether-lumefantrine; ASAQ = artesunate-amodiaquine, DP = dihydroartemisinin-piperaquine.

The performance measures of the different methodologies were explored in terms of bias, precision, coverage and overall accuracy. Bias targets the parameter of interest (θ) on average and was defined as the difference between the full data estimate (true value) and the average of the estimates obtained from the 1,000 datasets with simulated missingness, and was expressed in terms of a percentage as relative bias. This (relative risk) provides a more straightforward comparison of bias across treatment regimens as the true value of θ is different.

$$Bias = \frac{1}{n_{sim}} \sum_{i=1}^{n_{sim}} \widehat{\theta_i} - \theta \qquad n_{sim} = 1, 2, 3, \dots, 1000$$

Precision was measured based on the standard errors from the simulation studies. The empirical standard error (EmpSE) was calculated as the square root of the variance of the estimated K-M across 1,000 datasets. The empirical standard error depends only on the estimated $\hat{\theta}_i$ and doesn't require knowledge of the true value of θ. The model-based SE (ModSE) or the average estimated standard error, which targets the empirical SE, was calculated as the square root of the average variance across 1,000 simulated datasets.

$$Empirical\ SE = \sqrt{\frac{1}{(n_{sim} - 1)} \sum_{i=1}^{n_{sim}} (\widehat{\theta_i} - \overline{\theta})^2}$$

$$Model\ based\ SE = \sqrt{\frac{1}{n_{sim}} \sum_{i=1}^{n_{sim}} \widehat{Var(\widehat{\theta_i})}}$$

$$n_{sim} = 1, 2, 3, \dots, 1000$$

The coverage probability (CP) was estimated as the proportion of the 1,000 datasets where the estimated 95% confidence interval (CI) included the full data ("true") estimate of the K-M. For an advertised 95% CI, theoretical coverage is expected to fall between 93.6% to 96.4% with

1,000 simulated datasets, and a drop in coverage below 90% is regarded as problematic [298]. Finally, the overall accuracy of the estimator was measured using root mean squared error (RMSE) which combines bias and variance.

$$RMSE = \sqrt{\frac{1}{n_{sim}} \sum_{i=1}^{n_{sim}} \left(\widehat{\theta_i} - \theta\right)^2}$$

6.3.4 Software

Multiple imputation was carried out using the **mice** library and survival analyses were performed using the **survival** library in R.

6.4 Simulation study: Results

The result of the different performance measures for the three methods used for handling missing data are presented in Figures 6.2, 6.3, 6.4, 6.5 and 6.6, and in Tables 6.6, 6.7, 6.8, and 6.9.

6.4.1 Mean bias

In all of the scenarios studied, the CC analysis resulted in an upwards biased estimate of antimalarial efficacy (Figure 6.2). The bias was incremental with increasing percentage of indeterminate recurrences, irrespective of the treatment regimen and missingness mechanism studied. For example, in the AL arm, the mean bias was 0.34%, 1.04% and 1.59% when respectively 10, 30 and 45% of the recurrences were set to missing under mechanism 1. The IPW approach, which included recurrence status as a predictor (IPW) in the missingness model, produced estimates closest to the true value across all scenarios. IPWs calculated from a missingness model that excluded recurrence status (IPW-E) produced larger absolute biases compared to the standard IPW and MI approaches. In general, MI estimates were slightly negatively biased whereas the IPW estimates exhibited positive bias. The MI and IPW approaches led to smaller biases under mechanisms 2a and 2b compared to mechanism 1, although the magnitude of the bias was negligible. Similarly, there was no clear trend in the direction of bias for mechanisms 2a and 2b. When expressed in relative terms, the conclusion remained similar (Figure 6.3).

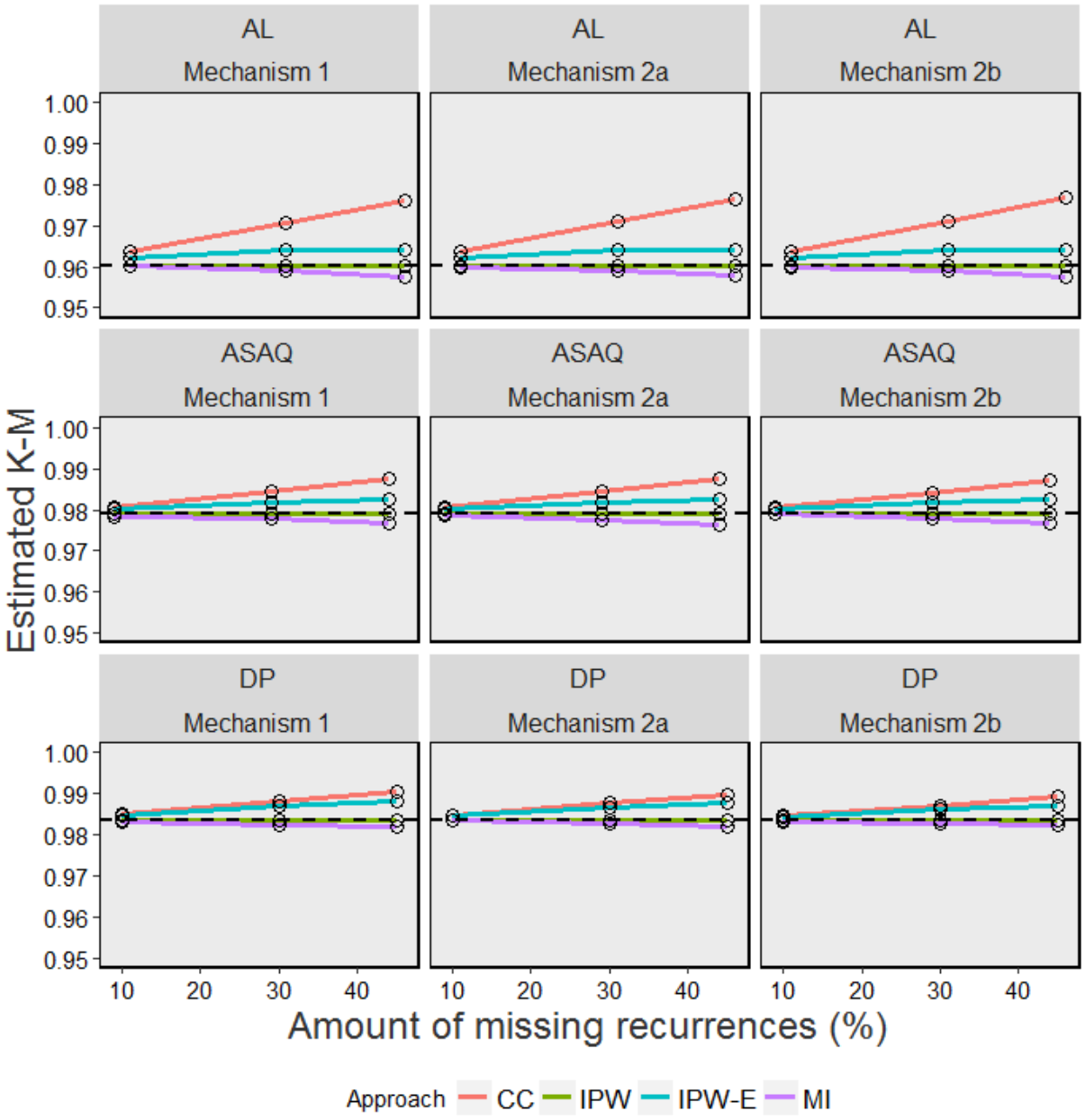


Figure 6.2: Average estimate of Kaplan-Meier survival probability on day 28 across 1,000 simulated datasets for different missing data methods, missing data mechanisms and percentages of indeterminate recurrences

Legend: CC = complete case; IPW = inverse probability weighting; IPW-E = inverse probability weighting with recurrence status excluded; MI = multiple imputation; AL= artemether-lumefantrine; ASAQ = artesunate-amodiaquine; DP= dihydroartemisinin-piperaquine. The dotted line represents the estimates derived from the full data estimate before missingness was induced.

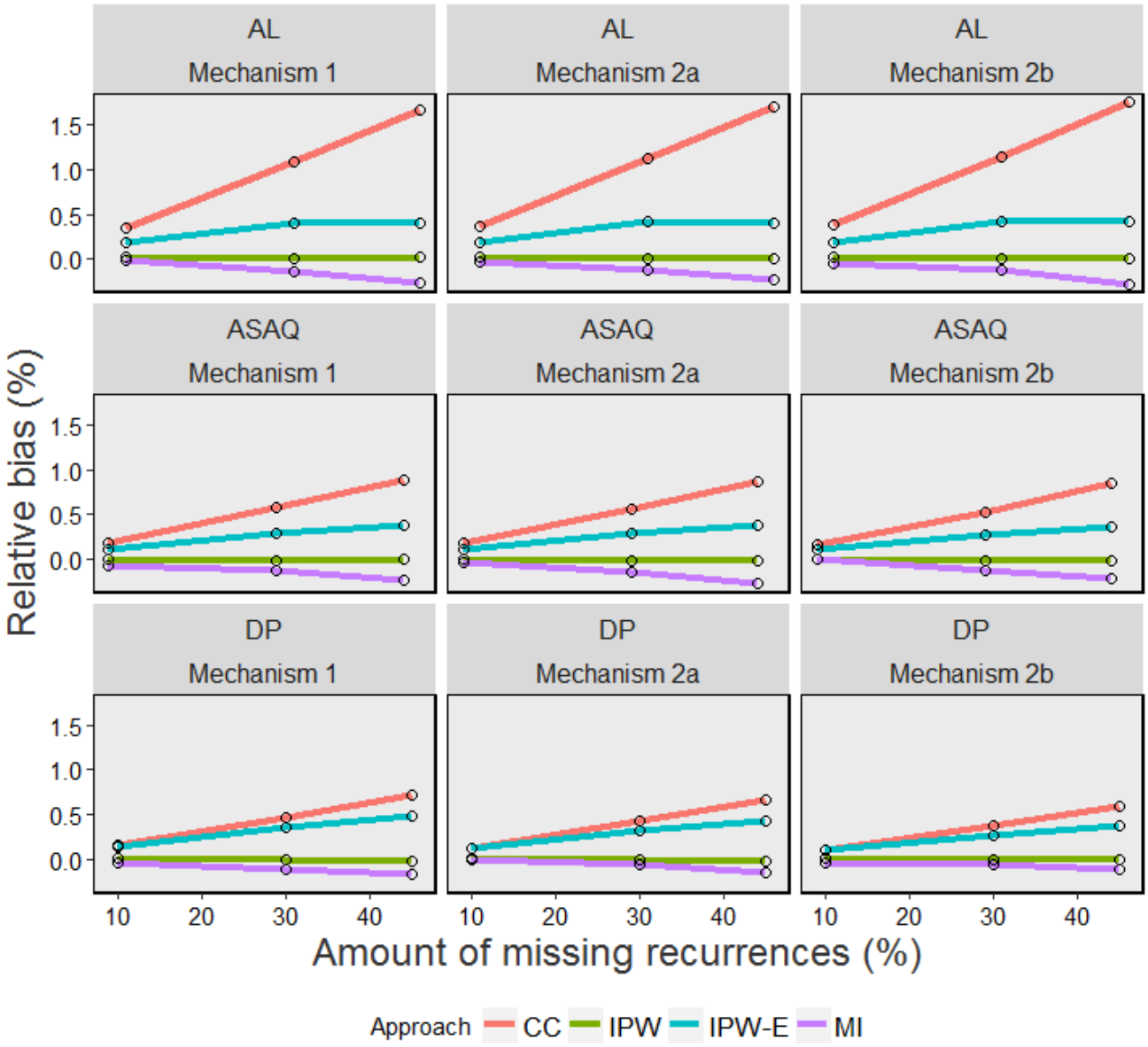


Figure 6.3: Relative bias in the K-M estimates on day 28 across 1,000 simulated datasets for different missing data methods, missing data mechanisms and percentages of indeterminate recurrences

Legend: CC = complete case; IPW = inverse probability weighting; IPW-E = inverse probability weighting with recurrence status excluded; MI = multiple imputation; AL= artemether-lumefantrine; ASAQ = artesunate-amodiaquine; DP= dihydroartemisinin-piperaquine.

6.4.2 Model-based and empirical standard errors

All approaches used for handling missing data became increasingly inefficient as the proportion of missing outcomes increased. When 10% of the recurrences were missing, estimates of model-based SE were similar across the methods, with differences observed only at the third decimal place. When missingness was ≥ 30%, there was a clear trend in efficiency. The IPW-E had the largest SE followed by the CC analysis. MI and standard IPW had smaller standard errors compared to other approaches with MI performing the best across all scenarios studied.

It was found that the IPW implementation which didn't account for uncertainty associated with estimating weights (the naïve estimator) resulted in SEs which were (falsely) smaller than the estimates obtained from MI. The comparison of SEs for IPW using the naïve approach and bootstrap method are presented in Table 6.6.

The average gain in precision (model-based SE) with IPW compared to the CC analysis over all missingness mechanisms were 1.9%, 4.7%, and 6.9% for 10, 30 and 45% of missingness respectively. With MI, these were 2.9%, 9.0%, 16.9% for 10, 30 and 45% missingness respectively. Similar results were observed with empirical SEs (Figure 6.4, Tables 6.7, 6.8, and 6.9).

Table 6.6: Model-based standard errors for IPW estimator

	10% missing		30% missing		45% missing	
	Naïve SEs	Bootstrap SEs	Naïve SEs	Bootstrap SEs	Naïve SEs	Bootstrap SEs
AL						
Mechanism 1	0.1584	0.1656	0.1586	0.1861	0.1593	0.2097
Mechanism 2a	0.1585	0.1661	0.1586	0.1879	0.1592	0.2123
Mechanism 2b	0.1585	0.1664	0.1587	0.1889	0.1593	0.2151
ASAQ						
Mechanism 1	0.2418	0.2532	0.2426	0.2858	0.2450	0.3230
Mechanism 2a	0.2417	0.2527	0.2425	0.2846	0.2446	0.3198
Mechanism 2b	0.2415	0.2520	0.2424	0.2826	0.2443	0.3173
DP						
Mechanism 1	0.2178	0.2274	0.2186	0.2542	0.2190	0.2841
Mechanism 2a	0.2177	0.2260	0.2182	0.2490	0.2187	0.2754
Mechanism 2b	0.2177	0.2246	0.2181	0.2444	0.2187	0.2682

AL= artemether-lumefantrine; ASAQ = artesunate-amodiaquine; DP = dihydroartemisinin-piperaquine
Naïve SEs = SEs for the IPW estimator without incorporating uncertainty in the estimated weights
Bootstrap SEs = SEs derived as the standard deviation of the Kaplan-Meier estimate across 200 bootstrap samples;
SE = standard error after complementary log-log (cloglog) transformation of K-M estimate
The description of the mechanisms is presented in Section 6.3.1.1 and 6.3.1.2

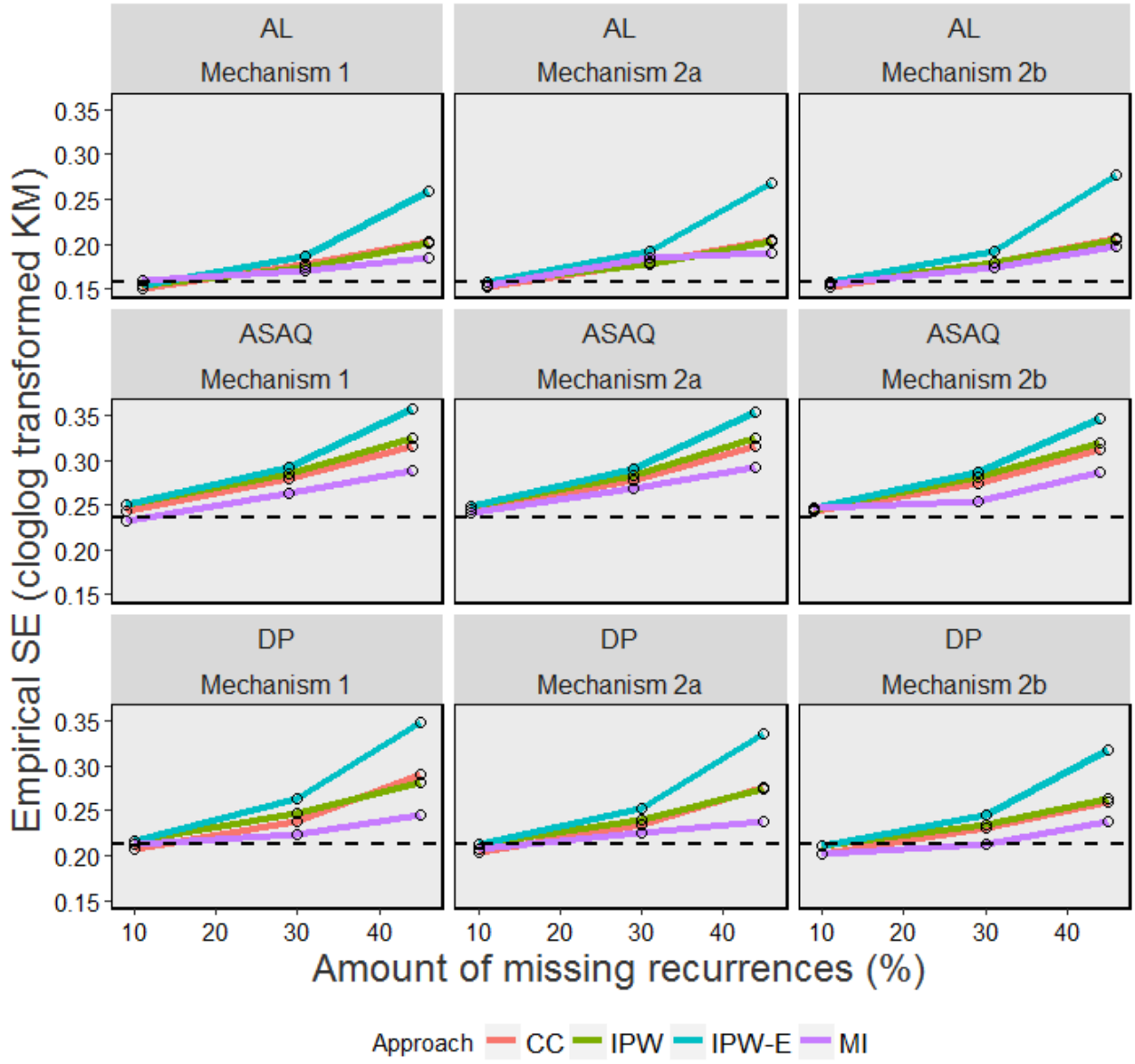


Figure 6.4: Empirical standard errors of K-M estimates on day 28 across 1,000 simulated datasets for different missing data methods, missing data mechanisms and percentages of indeterminate recurrences (on complementary log-log scale)

Legend: CC = complete case; IPW = inverse probability weighting; IPW-E = inverse probability weighting with recurrence status excluded; MI = multiple imputation; AL= artemether-lumefantrine; ASAQ = artesunate-amodiaquine; DP= dihydroartemisinin-piperaquine. The dotted line represents the estimates derived from the full data estimate before missingness was induced.

6.4.3 Coverage probability of the true value

The complete case approach suffered from poor coverage in all the scenarios under consideration and this deteriorated rapidly with increasing proportions of missingness (Figure 6.5). The IPW and MI implementations resulted in coverage probabilities close to the nominal 95% level irrespective of the missingness scenarios studied. The IPW-E approach was also associated with coverage that was lower than the nominal level across the entire simulation scenarios studied.

6.4.4 Root mean squared error (RMSE)

The complete case approach had the least overall accuracy of all the missing data methods followed by IPW-E across all missingness mechanisms (Figure 6.6). For all the missing data methods, the overall accuracy decreased with increasing proportions of missingness. The MI and IPW approach both provided similar estimates of accuracy. With the MI and IPW approaches, the accuracy was higher under missingness mechanism 2 compared to mechanism 1 (Tables 6,7, 6.8, and 6.9).

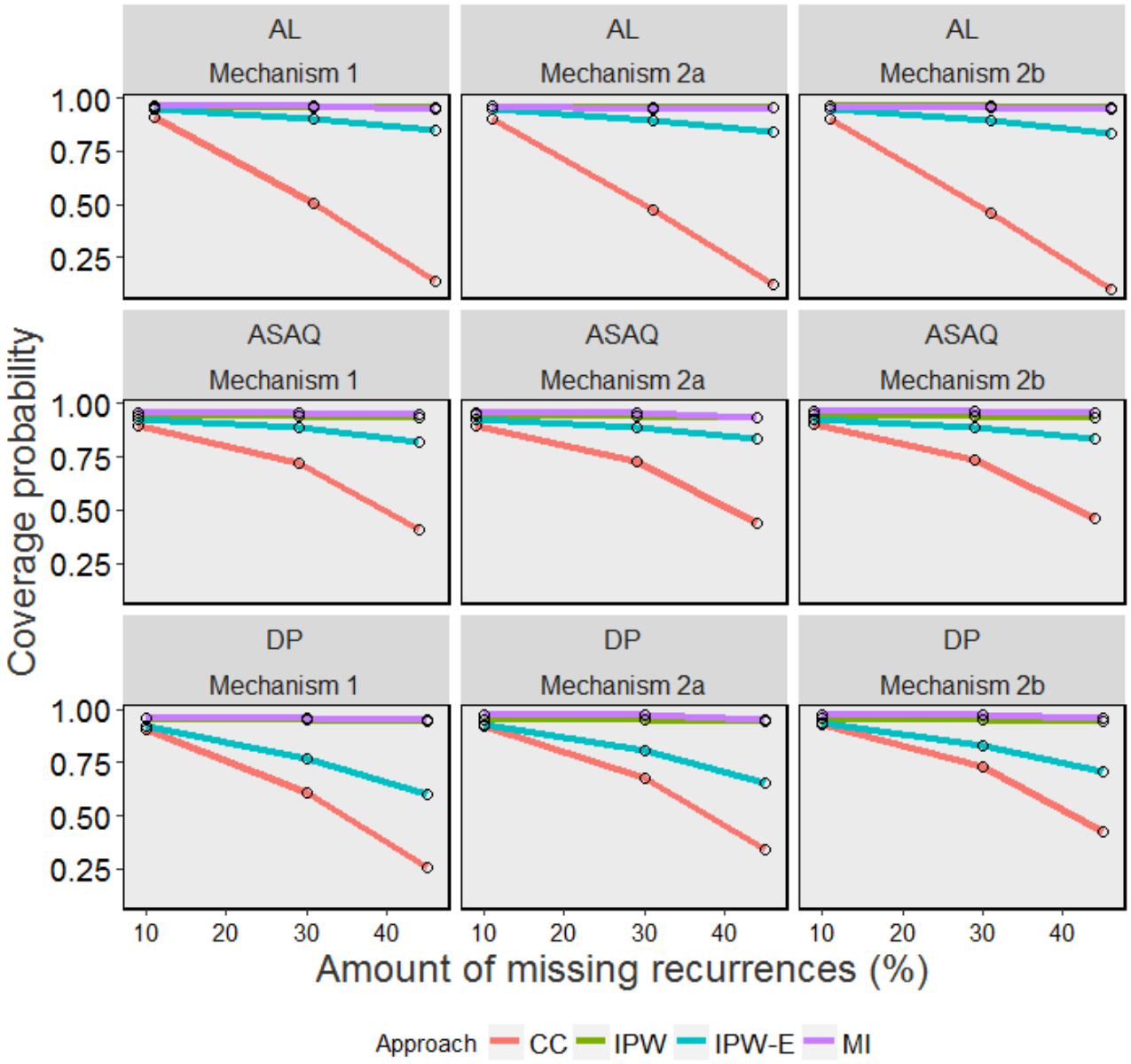


Figure 6.5: Coverage probability of K-M estimates on day 28 across 1,000 simulated datasets for different missing data methods, missing data mechanisms and percentages of indeterminate recurrences

Legend: CC = complete case; IPW = inverse probability weighting; IPW-E = inverse probability weighting with recurrence status excluded; MI = multiple imputation; AL= artemether-lumefantrine; ASAQ = artesunate-amodiaquine; DP= dihydroartemisinin-piperaquine.

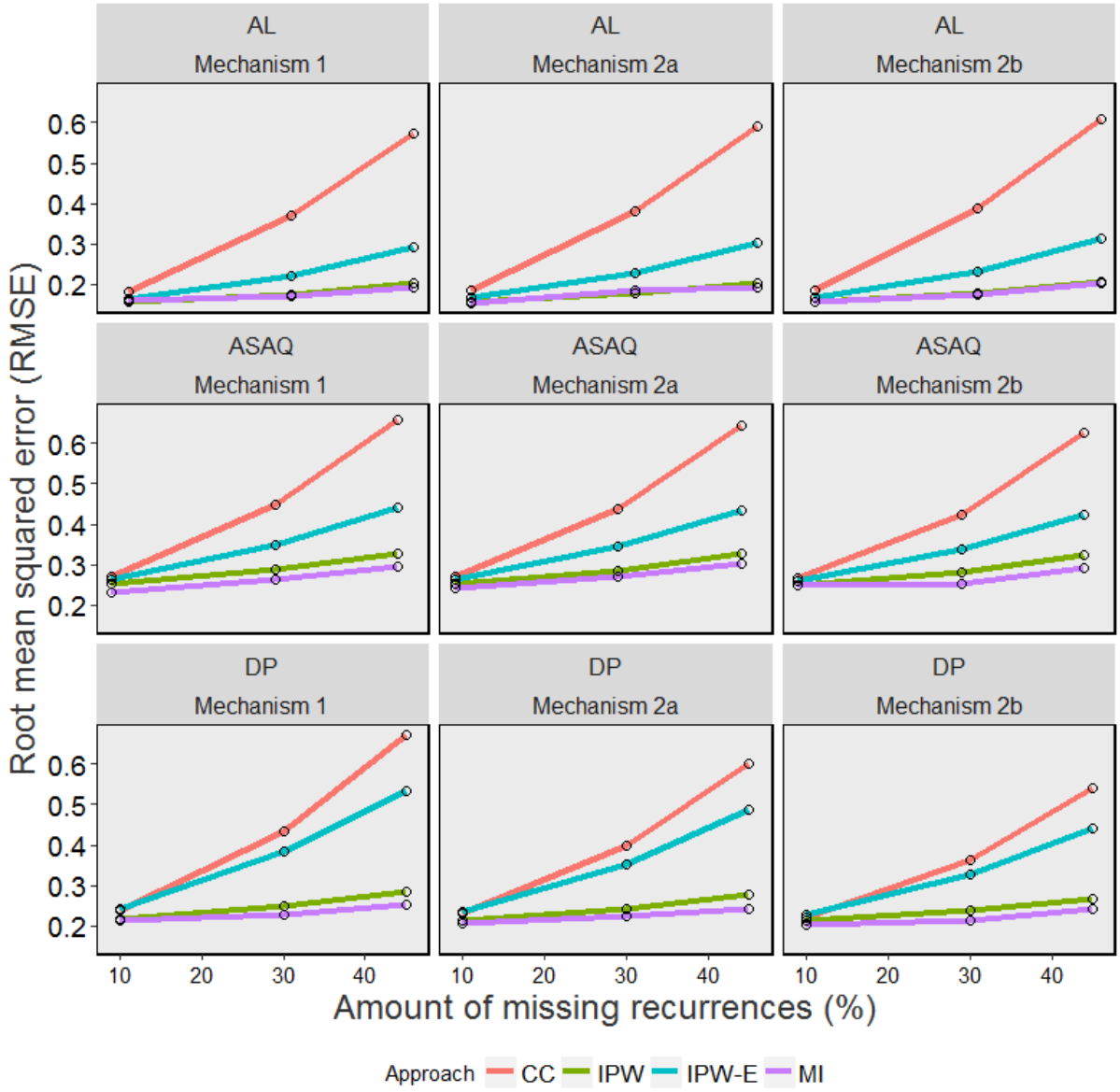


Figure 6.6: Root mean squared error of the derived K-M estimates on day 28 across 1,000 simulated datasets for different missing data methods, missing data mechanisms and percentages of indeterminate recurrences

Legend: CC = complete case; IPW = inverse probability weighting; IPW-E = inverse probability weighting with recurrence status excluded; MI = multiple imputation; AL= artemether-lumefantrine; ASAQ = artesunate-amodiaquine; DP= dihydroartemisinin-piperaquine.

Table 6.7: Performance measures of different approaches for handling missing outcome for artemether-lumefantrine

	10% Missing						**30% missing**						**45% missing**					
	Bias	rBias	ModSE†	EmpSE	CP	RMSE	Bias	rBias	ModSE†	EmpSE	CP	RMSE	Bias	rBias	ModSE†	EmpSE	CP	RMSE
M1																		
CC	0.0034	0.35	0.1672	0.1505	0.909	0.1815	0.0104	1.08	0.1905	0.1774	0.504	0.3691	0.0159	1.65	0.2152	0.2032	0.134	0.5732
IPW	0.0002	0.02	0.1656	0.1543	0.958	0.1551	0.0001	0.01	0.1861	0.1742	0.957	0.1749	0.0002	0.02	0.2097	0.2015	0.954	0.2028
MI	-0.0002	-0.02	0.1654	0.1600	0.962	0.1601	-0.0012	-0.13	0.1823	0.1707	0.961	0.1715	-0.0026	-0.27	0.2001	0.1854	0.944	0.1914
IPW-E	0.0018	0.19	0.1660	0.1544	0.946	0.1652	0.0039	0.40	0.1968	0.1867	0.903	0.2224	0.0039	0.41	0.2536	0.2582	0.851	0.2918
M2a																		
CC	0.0035	0.37	0.1676	0.1511	0.905	0.1840	0.0108	1.12	0.1920	0.1788	0.471	0.3809	0.0163	1.69	0.2176	0.2053	0.123	0.5907
IPW	0.0002	0.02	0.1661	0.1566	0.96	0.1574	0.0001	0.01	0.1879	0.1769	0.957	0.1776	0.0001	0.01	0.2123	0.2028	0.957	0.2039
MI	-0.0003	-0.04	0.1651	0.1535	0.966	0.1535	-0.0012	-0.12	0.1848	0.1846	0.947	0.1850	-0.0022	-0.23	0.2041	0.1901	0.955	0.1938
IPW-E	0.0018	0.19	0.1664	0.1571	0.947	0.1681	0.0040	0.42	0.2001	0.1910	0.893	0.2281	0.0039	0.41	0.2594	0.2680	0.837	0.3019
M2b																		
CC	0.0036	0.38	0.1679	0.1518	0.902	0.1861	0.0110	1.14	0.1928	0.1800	0.458	0.3874	0.0167	1.74	0.2200	0.2060	0.101	0.6085
IPW	0.0002	0.02	0.1664	0.1563	0.961	0.1570	0.0001	0.01	0.1889	0.1787	0.961	0.1794	0.0001	0.01	0.2151	0.2051	0.957	0.2063
MI	-0.0005	-0.05	0.1654	0.1561	0.964	0.1560	-0.0011	-0.11	0.1857	0.1741	0.954	0.1745	-0.0026	-0.27	0.2060	0.1978	0.945	0.2031
IPW-E	0.0019	0.19	0.1668	0.1568	0.946	0.1680	0.0040	0.42	0.2016	0.1927	0.894	0.2304	0.0040	0.41	0.2659	0.2772	0.831	0.3119

†The model based standard errors for IPW approaches were estimated using 200 bootstrap samples

M1 = Missingness Mechanism 1; M2a = Missingness Mechanism 2a; M2b = Missingness Mechanism 2b; rBias = Relative bias (%); ModSE = Model based standard error; EmpSE = Empirical standard error; CP = Coverage Probability; RMSE = Root Mean Squared Error; CC = Complete Case; IPW = Inverse probability weighting model; MI = Multiply Imputed; IPW E = IPW model which excluded recurrence as a predictor in missingness model

Table 6.8: Performance measures of different approaches for handling missing outcome for artesunate-amodiaquine

	10% Missing						**30% missing**						**45% missing**					
	Bias	rBias	ModSE†	EmpSE	CP	RMSE	Bias	rBias	ModSE†	EmpSE	CP	RMSE	Bias	rBias	ModSE†	EmpSE	CP	RMSE
M1																		
CC	0.0018	0.26	0.2560	0.2438	0.896	0.2716	0.0056	0.57	0.2922	0.2787	0.720	0.4469	0.0086	0.88	0.3329	0.3166	0.407	0.6594
IPW	0.0001	0.25	0.2532	0.2499	0.941	0.2519	-0.0001	-0.01	0.2858	0.2850	0.940	0.2870	0.0000	0.00	0.3230	0.3249	0.934	0.3284
MI	-0.0006	0.25	0.2505	0.2326	0.958	0.2325	-0.0012	-0.13	0.2823	0.2639	0.955	0.2649	-0.0021	-0.22	0.3128	0.2878	0.947	0.2937
IPW-E	0.0012	0.25	0.2546	0.2505	0.926	0.2647	0.0029	0.30	0.2942	0.2921	0.885	0.3470	0.0038	0.39	0.3496	0.3579	0.818	0.4405
M2a																		
CC	0.0018	0.26	0.2556	0.2442	0.897	0.2706	0.0055	0.56	0.2906	0.2774	0.725	0.4383	0.0085	0.86	0.3301	0.3154	0.437	0.6455
IPW	0.0001	0.25	0.2527	0.2490	0.946	0.2509	-0.0001	-0.01	0.2846	0.2827	0.942	0.2846	-0.0001	-0.01	0.3198	0.3241	0.931	0.3270
MI	-0.0003	0.25	0.2522	0.2418	0.959	0.2422	-0.0014	-0.15	0.2802	0.2683	0.954	0.2702	-0.0025	-0.25	0.3090	0.2929	0.934	0.3017
IPW-E	0.0011	0.25	0.2540	0.2492	0.928	0.2632	0.0029	0.29	0.2924	0.2899	0.886	0.3437	0.0037	0.37	0.3447	0.3536	0.833	0.4323
M2b																		
CC	0.0017	0.25	0.2548	0.2437	0.902	0.2679	0.0052	0.53	0.2877	0.2733	0.736	0.4224	0.0082	0.84	0.3267	0.3125	0.459	0.6277
IPW	0.0001	0.25	0.2520	0.2457	0.946	0.2475	-0.0001	-0.01	0.2826	0.2804	0.941	0.2822	-0.0001	-0.01	0.3173	0.3199	0.936	0.3226
MI	0.0000	0.25	0.2538	0.2467	0.965	0.2480	-0.0011	-0.12	0.2790	0.2533	0.965	0.2542	-0.0021	-0.21	0.3095	0.2870	0.959	0.2923
IPW-E	0.0011	0.25	0.2532	0.2459	0.930	0.2591	0.0028	0.28	0.2892	0.2868	0.891	0.3385	0.0036	0.37	0.3397	0.3457	0.837	0.4219

†The model based standard errors for IPW approaches were estimated using 200 bootstrap samples

M1 = Missingness Mechanism 1; M2a = Missingness Mechanism 2a; M2b = Missingness Mechanism 2b; rBias = Relative bias (%); ModSE = Model based standard error; EmpSE = Empirical standard error; CP = Coverage Probability; RMSE = Root Mean Squared Error; CC = Complete Case; IPW = Inverse probability weighting model; MI = Multiply Imputed; IPW E = IPW model which excluded recurrence as a predictor in missingness model

Table 6.9: Performance measures of different approaches for handling missing outcome for dihydroartemisinin-piperaquine

	10% Missing						30% missing						45% missing					
	Bias	rBias	ModSE†	EmpSE	CP	RMSE	Bias	rBias	ModSE†	EmpSE	CP	RMSE	Bias	rBias	ModSE†	EmpSE	CP	RMSE
M1																		
CC	0.0016	0.16	0.2296	0.2071	0.904	0.2391	0.0047	0.48	0.2618	0.2385	0.605	0.4334	0.0076	0.73	0.2997	0.2903	0.256	0.6726
IPW	0.0001	0.01	0.2274	0.2158	0.957	0.2179	0.0001	0.01	0.2542	0.2470	0.95	0.2493	0.0044	-0.01	0.2841	0.2812	0.946	0.2831
MI	-0.0003	-0.03	0.2247	0.2131	0.961	0.2131	-0.0010	-0.10	0.2447	0.2234	0.957	0.2257	0.0045	-0.15	0.2667	0.2457	0.948	0.2528
IPW-E	0.0014	0.14	0.2298	0.2156	0.923	0.2415	0.0036	0.37	0.2703	0.2635	0.766	0.3851	0.0062	0.49	0.3342	0.3490	0.601	0.5332
M2a																		
CC	0.0013	0.14	0.2276	0.2042	0.917	0.2292	0.0042	0.43	0.2562	0.2340	0.673	0.3979	0.0070	0.66	0.2882	0.2755	0.337	0.6028
IPW	0.0002	0.02	0.2260	0.2120	0.955	0.2141	0.0001	0.01	0.2490	0.2397	0.951	0.2416	0.0043	0.00	0.2754	0.2737	0.943	0.2754
MI	0.0000	0.00	0.2252	0.2064	0.972	0.2073	-0.0005	-0.05	0.2441	0.2254	0.971	0.2253	0.0042	-0.13	0.2609	0.2381	0.953	0.2432
IPW-E	0.0012	0.12	0.2281	0.2121	0.927	0.2335	0.0032	0.32	0.2626	0.2528	0.804	0.3521	0.0058	0.43	0.3208	0.3353	0.653	0.4856
M2b																		
CC	0.0011	0.11	0.2260	0.2025	0.928	0.2217	0.0037	0.38	0.2506	0.2306	0.733	0.3632	0.0064	0.60	0.2782	0.2598	0.424	0.5397
IPW	0.0002	0.02	0.2246	0.2108	0.959	0.2128	0.0001	0.01	0.2444	0.2354	0.951	0.2374	0.0041	0.00	0.2682	0.2643	0.946	0.2662
MI	-0.0003	-0.03	0.2218	0.2010	0.974	0.2009	-0.0006	-0.06	0.2382	0.2128	0.972	0.2130	0.0041	-0.10	0.2569	0.2380	0.960	0.2401
IPW-E	0.0010	0.11	0.2263	0.2103	0.933	0.2275	0.0028	0.28	0.2555	0.2462	0.830	0.3254	0.0054	0.38	0.3082	0.3178	0.709	0.4418

†The model based standard errors for IPW approaches were estimated using 200 bootstrap samples

M1 = Missingness Mechanism 1; M2a = Missingness Mechanism 2a; M2b = Missingness Mechanism 2b; rBias = Relative bias (%); ModSE = Model based standard error; EmpSE = Empirical standard error; CP = Coverage Probability; RMSE = Root Mean Squared Error; CC = Complete Case; IPW = Inverse probability weighting model; MI = Multiply Imputed; IPW E = IPW model which excluded recurrence as a predictor in missingness model

6.5 Conclusions from the simulation study

The key findings of this simulation study are:

- The widely used complete case (CC) approach for handling indeterminate outcomes overestimated antimalarial efficacy by up to 1.7%. The CC estimator was also associated with increased inefficiency and sub-optimal coverage probabilities across all the scenarios studied.

- In the IPW approach:

    - Excluding recurrence status as predictor in the missingness model was associated with increased biases and poor coverage probability.

    - The calculation of the standard error of the IPW estimator should appropriately account for uncertainties in estimating the weights through the use of bootstrap resampling. Ignoring the uncertainties in estimated weights led to the IPW approach falsely appearing more efficient compared to the MI method.

- MI and correctly implemented IPW approach (recurrence status included as a predictor for deriving the weights and bootstrap resamples used for estimating standard errors) remained approximately unbiased, improved the efficiency of the derived estimate and provided coverage probabilities close to the advertised 95% level.

6.6 Evaluation of applicability of MI and IPW approaches in small sample antimalarial trials: a resampling study

The 4ABC Trial was a large multi-centric study with more than 1,000 patients enrolled per treatment arm. However, most of the antimalarial trials have sample sizes of about 200 subjects (Chapter 3, Table 3.1). To evaluate the applicability of the MI and IPW approaches in antimalarial studies with small sample sizes, a bootstrap resampling study was carried out using data from the artemether-lumefantrine arm (n=1,131) from the motivational The 4ABC Trial as a '**synthetic population**'. Since the magnitude of the bias was similar for all the missingness mechanisms studied (Tables 6.7, 6.8, and 6.9), only missingness mechanism 1 (a truly random process after conditioning on recurrence) was further considered. Furthermore, only the worst case scenario of 45% missingness was considered.

6.6.1 Generating resample data

The following sample sizes were studied: n=100, 200 to represent the typical size of antimalarial studies. An additional scenario of n=500 was also studied for comparison. Datasets of the required size were created by randomly sampling (with replacement) from the data of 1,131 patients, and missingness was induced in each of the sampled data; this was repeated 1,000 times. The samples were drawn to replicate the observed outcome distribution of recrudescence (4%), new infection (22%), and cured (75%) in 1,131 subjects treated with artemether-lumefantrine (Table 6.10).

Table 6.10: Number of events in resampled data of different sample sizes drawn from observed data for artemether-lumefantrine

Outcome	Full data	n=100	n=200	n=500
Cured	847	75	150	374
Recrudescence	41	4	7	18
New Infection	243	21	43	107

6.6.2 Analysis of created datasets

Each of the datasets was analysed using the CC, IPW and MI approaches as described in Section 6.3.2. The "true value" was defined as the full data estimates derived from the synthetic population of 1,131 patients (Table 6.5). The performance measures were evaluated from a total of 1,000 replications in terms of bias, coverage and precision.

6.6.3 Results

The mean bias for the CC approach remained unaffected by the study sample size, being 1.5% for n=100 and 1.8% for n=200 (Table 6.11). The MI approach was biased for small sample sizes, the bias being -1.9% for n=100, -1.0% for n=200, and -0.22% for n=500. The IPW approach produced very marginally unbiased estimates across all of the sample sizes studied.

As expected, there were gains in precision with increases in sample size, with both the empirical and model-based standard errors being smallest with MI and largest for the CC method. As the RMSEs for the CC approach were larger for n=100, and n=200, this resulted in over-coverage. In contrast, the MI and IPW approaches, both provided coverage greater than 90% across all sample sizes. Finally, the MI approach was found to be have the best overall accuracy (measured by RMSE) followed by the IPW approach.

Table 6.11: Performance measure against different sample sizes

	Complete Case	Multiple Imputation	Inverse probability weighting
Mean bias			
n=100	0.0151	-0.0189	0.0005
n=200	0.0185	-0.0101	0.0048
n=500	0.0187	-0.0022	0.0050
Empirical SE			
n=100	0.5142	0.3921	0.6091
n=200	0.5229	0.3891	0.5689
n=500	0.3632	0.2766	0.3670
Model-based SE (Full data estimate = 0.1567)¶			
n=100	0.7897	0.6432	0.8964 (0.6303)†
n=200	0.6073	0.5139	0.6326 (0.4651)†
n=500	0.3612	0.3248	0.3591 (0.2694)†
Root mean square error			
n=100	0.8014	0.5100	0.6356
n=200	0.9168	0.4215	0.6280
n=500	0.7935	0.2772	0.4163
Coverage probability			
n=100	1.000	0.897	0.988
n=200	0.989	0.906	0.986
n=500	0.508	0.969	0.957

¶Full data estimate = the estimate from the full data of the synthetic population of 1,131 patients (Table 6.5)
†SE for the IPW estimate was derived from 200 bootstrap resamples. The naïve estimate of standard errors which didn't account for uncertainties in estimating weights is shown in parenthesis.
SE = standard error

6.6.4 Conclusions from resampling study

The key findings of this resampling study are:

- Complete case (CC) approach remained biased for all sample sizes studied and overestimated antimalarial drug efficacy by at least 1.5%.

- MI was biased for small sample sizes (n=100, and 200) and underestimated the efficacy by 1.9% and 1.0% for n=100 and n=200 respectively.

- IPW approach remained approximately unbiased, but was inefficient compared to the MI estimates.

6.7 Discussion

Missing data in clinical trials can arise due to the study design or happenstance, and can pose analytical challenges including undermining the validity and interpretation of the results. In antimalarial studies, indeterminate recurrences resulting from genotyping failure are frequently encountered, especially in areas of high transmission intensity, where multiple infections are common. Principled approaches for handling missing data have proliferated the medical and statistical literature in recent years [289,292,305], yet the most common approach used by malaria researchers and which is recommended by the WHO for handling indeterminate cases is to simply exclude these from the statistical analysis. In this chapter, the performance of multiple imputation (MI) and inverse probability weighing (IPW) methods were evaluated for handling indeterminate outcomes in the context of estimating antimalarial efficacy using The 4ABC Trial as a motivational study [139]. The 4ABC Trial is one of the largest antimalarial studies ever conducted and was well suited to studying the utility of the MI and IPW approaches for handling indeterminate recurrences. The use of a real dataset to represent the complete (full) data avoided arbitrary choices usually made in simulating covariates and survival data, and provided a rich dataset from multiple endemic settings, with auxiliary covariates for implementation of IPW and MI approaches, thus making the generalisability of results for antimalarial trials more plausible.

Two different missingness mechanisms were investigated and differences in the estimates obtained were compared for scenarios in which 10, 30 and 45% of the recurrences were reclassified as having a missing outcome. In all of these scenarios, the current WHO recommendation of excluding indeterminate cases resulted in an upwards biased estimate of drug efficacy by up to a maximum of 1.7%, the magnitude of which was correlated with the proportion of recurrent outcomes classified as indeterminate. Similar findings were observed by Machekano et al (2008) [88] who reported an absolute overestimation of 3.2% using the

CC approach compared to the IPW and MI methods for the antimalarial regimen of chloroquine-sulphadoxine-pyrimethamine. The overestimation was 1.7% for the regimen amodiaquine-sulphadoxine-pyrimethamine when the observed proportions of missing recurrences were 33% and 17% respectively in the chloroquine- sulphadoxine-pyrimethamine and amodiaquine- sulphadoxine-pyrimethamine arms.

While the CC analysis was associated with overestimation of drug efficacy, the MI and IPW approaches were generally unbiased under all missingness scenarios and provided an increased precision. The IPW approach provided consistently the least biased estimate of all the approaches for all proportions of missingness; however, it came at a price of marginally inflated standard errors compared to the MI method. This observation corroborates well with the findings reported by Machekano and colleagues [88]. However, the analysis carried out in this chapter had two important differences. First, the Kaplan-Meier method, which is currently the preferred approach for estimating drug efficacy , was used (as opposed to the cured proportion reported in Machekano and colleagues [88]). Second, when constructing the missingness model for the IPW implementation, recurrence status was included as a predictor in this analysis (not included by Machekano and colleagues in their missingness model). In antimalarial studies, the observation of a missing outcome is only possible once a patient experiences recurrent parasitaemia - thus leading to a scenario where recurrence status perfectly predicts missing outcome. The results of the simulation study demonstrated that exclusion of recurrence status from the missingness model was associated with an increased bias, inflated standard error and poor coverage probability (Tables 6.7, 6.8, 6.9). This suggests that recurrence status should always be included in the missingness model to obtain valid inferences for the IPW method.

As for the IPW approach, the validity of inferences derived from MI relies on the correct implementation of the imputation model, hence this approach should include the correct

functional form and specify any interactions (not fully considered in this chapter). Failure to do so could lead to invalid inferences being drawn, especially when the fraction of missing information is large [288,300,306]. In practice, all imputation models are likely to be mis-specified to some extent. Arguably specifying the missingness model correctly is an easier task compared to specifying a correct imputation model [289,307], thus making the IPW approach a feasible alternative for handling indeterminate outcomes when estimating efficacy in antimalarial studies. However, it is important to account for the uncertainty associated with the estimation of weights when using the IPW method, as the naïve estimate of the standard error ignores this uncertainty, leading to the IPW approach paradoxically appearing to be far more efficient than the MI approach (Table 6.6) and potentially producing small *P*-values [303,308,309].

In addition to being biased and inefficient, the CC estimates also suffered from poor coverage compared to the MI and IPW methods. When the missingness was greater than 30%, the coverage for the CC approach deteriorated rapidly and fell below 90% whereas for both the MI and IPW approaches, the coverage remained near the nominal 95% level. The fall in coverage below the 90% level with the CC approach is explained by the fact that the CC estimator became progressively biased with increasing missing fraction. Similarly, the coverage probability for the IPW estimator which excluded recurrence status as a predictor deteriorated rapidly with increasing missingness.

The current WHO guidelines require that a new regimen should demonstrate at least 95% efficacy in order to be included in the antimalarial treatment policy, and further investigations are warranted when treatment failure exceeds 10% when examining for drug resistance [64]. The results of this study, taken together with the findings of Machekano et al [88] highlights that the CC approach provides an optimistic view of the treatment efficacy and this can have

potentially deleterious consequences when the estimate is at the cusp of these WHO thresholds (especially in a study where a large proportion of outcomes are indeterminate). From a public health perspective, the false sense of confidence generated by these studies regarding the current status of antimalarial regimens can have important ramifications for the evolution of antimalarial drug resistance. The prolonged usage of a less optimal regimen provides a constant drug selection pressure to the parasites; a scenario highly conducive for the emergence of *de novo* drug resistance. Given the paucity of alternative regimens currently available and the spread of artemisinin resistance across South-East Asia [43], it is important that researchers and policy makers alike are aware of the pitfalls associated with the CC estimate of efficacy when drawing conclusions from routine therapeutic surveillance studies (TES).

6.7.1 Limitations

This simulation study has a number of limitations. First, as with all simulation studies, this study also does not represent every missing-data problem which can be encountered in practice. A single method cannot be universally recommended but rather the choice of the method should be guided by the research question and the context of the study. Second, IPW and MI are not the only available approaches for handling missing data. Likelihood based approaches, which use expectation-maximisation (EM) algorithms are alternative approaches but at present are not implemented in standard software [310]. Finally, the applicability of the sophisticated missing data approaches in a relatively small sized therapeutic efficacy study needs attention. The resampling study presented in Section 6.6 suggested that MI and IPW should be used with caution in small sample trials. MI was biased but had greater precision while IPW was unbiased but inefficient (Table 6.11). MI requires a logistic imputation model to be fitted to only the recurrences, thereby likely to encounter difficulties when the observed numbers of recurrences are small. For the IPW approach, although the missingness logistic model is relatively straight forward to fit to all the data, this might result in unstable weights when the number of

recurrences is small. Hence, MI and IPW approaches cannot be recommended in small sample trials based on the results presented in Section 6.6.

In the presence of missing data, no statistical methods, simple or sophisticated, can supersede the result, which would have been derived had the data been fully observed. Thus, best possible efforts should be made to minimise missingness through careful design, study management, and adherence to standardised protocols [311–314]. Diligence is essential in sample collection in the field, and the use of better genotyping method (e.g. capillary based) including appropriate quality control measures through a regular proficiency testing program should be deployed [315]. Missing data should be anticipated in advance and researchers should strive to collect data on variables which might be related to variables expected to exhibit missingness (such as background allelic frequency of the markers used for parasite genotyping). When using MI and IPW, researchers should clearly report the details of modelling approaches including the construction of imputation and missingness models and any associated sensitivity analyses [299,316].

The definition of recrudescence and new infection depends on how different sized bands are binned and classified as being the same or different alleles. For example, Cattamanchi et al (2003) [317] considered the alleles to be the same if the molecular weights were within 10 base-pair length for *msp-2* genes whereas Rouse et al (2008) reported that an identical *msp-2* allele could be different by up to 18 base pairs [318]. The definition adopted for defining recrudescence *vs* new infection is critical and researchers should always endeavour to publish the fragment length of the alleles in the pre- and post-treatment samples as done by Plucinski et al (2017) (See supplemental table 1 of [137][319]).

6.8 Conclusions

The widely used approach of excluding indeterminate outcomes invariably results in overestimation of antimalarial efficacy. In the example studied, the incorporation of missing data through correctly implemented IPW (including recurrence status as the predictor and using bootstrapping to estimate the standard error) and MI methods greatly reduced bias, enhanced efficiency and provided superior coverage probability of the derived efficacy estimates compared to the CC method. IPW and MI approaches are easily implementable in standard statistical software and should be considered for handling indeterminate outcomes in the derivation of antimalarial efficacy for large trials. However, in regular antimalarial studies (n < 200 subjects) with small number of recurrences, they should be cautiously applied with the help from an expert statistician.

Chapter 7: General discussion and conclusions

The aim of this thesis was to optimise statistical methodologies for handling different causes of parasitic recurrences (recrudescence, new infection and indeterminate) in the quantification of antimalarial drug efficacy for artemisinin combination therapies (ACTs) in uncomplicated *P. falciparum* malaria. To do so, the unique WorldWide Antimalarial Resistance Network (WWARN) database of individual patient data collated from 92 clinical trials was utilised, and a comprehensive investigation of the methodological issues currently faced by antimalarial researchers was undertaken. This project had four specific objectives: first, to review the evolution of statistical approaches used in the derivation of antimalarial efficacy and critically assess the challenges associated with current practices; second, to investigate the optimal duration of patient follow-up for ACTs to capture recrudescent infections; third, to explore the competing risk approach for dealing with new infections; and finally, to evaluate the suitability of missing data techniques for handling indeterminate parasitic recurrences.

This chapter summarises the salient findings of the thesis and their public health implications, discusses the methodological strengths and limitations, and outlines future work.

7.1 Chapter summary

Chapter 1 started with an introduction to malaria and provided a rationale for the work conducted in this thesis, in the wake of artemisinin resistance. In Chapter 2, I undertook a review of the statistical approaches used for the quantification of antimalarial drug efficacy since the release of the first standardised WHO protocol in 1965 for conducting antimalarial surveillance studies. This review presented some of the limitations associated with current practices for measuring antimalarial efficacy; some of which were explored further in this thesis. Chapter 3 described the screening of studies from the WWARN database that were used

in this thesis. The remaining three chapters (Chapters 4 to 6) addressed the objectives laid out in this thesis and were concerned with optimising analytical approaches for handling recrudescence, new infection and undifferentiated recurrence, respectively.

Chapter 4 dealt with parasitic recrudescence with the aim of assessing whether the currently recommended follow-up duration for different ACT regimens are likely to adequately capture all recrudescences. In this chapter, the empirical distribution of the time to PCR confirmed recrudescence post ACT treatment was constructed. The derived probability distribution provided clear evidence that the current recommended follow-up durations for the two widely used ACTs, artemether-lumefantrine (AL) and dihydroartemisinin-piperaquine (DP), result in systematic underestimation of drug failure. In African children aged less than 5 years old, a clinical trial of DP with the current recommended follow-up duration (42 days) was predicted to miss 47% [95% CI: 19 – 90%] of all recrudescences in those who received ≥ 48 mg/kg of piperaquine dose. In the same population, with a current recommended follow-up period of 28 days with an AL regimen, a clinical trial was predicted to miss 58% [95% CI: 47 – 74%] of all failures and in Asia (in all ages) the prediction was that 32% [95% CI: 15 – 45%] of failures would be missed. These results provide convincing evidence that a longer follow-up duration should be adopted to capture late recrudescences following treatment with AL and DP.

The challenges of adopting a longer follow-up need to be weighed up against the increased cost, complexity and logistical difficulties that often result in a larger proportion of patients being lost to follow-up. One approach to circumvent these issues would be to adopt a pragmatic study design with staggered follow-up, for instance actively following patients for a minimum period (for example, 28 days with AL and 42 days with DP) and then passively for a much longer duration. This strategy has been implemented successfully in The 4ABC trial where patients were followed passively from days 29 to 63 [139], and in another study undertaken in

Kenya, where passive follow-up was used to detect failures between days 63 to 84 [162]. Whilst such an approach will capture symptomatic recurrent infections assuming the patient comes back to the same clinic, this will not capture asymptomatic illness where a patient is unaware of their peripheral parasitaemia. Another pragmatic approach could be to decrease the number of follow-up visits by collecting clinical and parasitological samples every 10-12 days instead of 7 days, and coupling a longer gap between follow-up visits with the use of mobile phone technology [320]. The use of mobile phone messaging has shown promising results in a recent randomised trial conducted in Kenya, which showed that children who received a text reminder were 1.58 times more likely to return for their clinical visits (although it didn't affect treatment adherence) [321]. A thorough economic analysis is warranted to assess the clinical benefits of capturing such late recrudescences and the practicality of using a staggered or a passive follow-up design.

While capturing all possible recrudescent parasites is important for obtaining an accurate estimate of drug efficacy, a substantial proportion of patients are expected to develop new infections during the ensuing weeks of follow-up. The occurrence of a new infection can potentially preclude subsequent recrudescent infection, thus representing a competing risk event. In Chapter 5, a competing risk survival analysis approach was used to handle these new infections. The estimation of the cumulative incidence function (CIF) allows partitioning of the risks into different recurrence categories thus appropriately accounting for competing risk events, as opposed to the conventional Kaplan-Meier (K-M) analysis which considers new infections as censored. It was found that failure estimates derived using a 1 minus Kaplan-Meier approach were a marginal (absolute) overestimation of cumulative risk of recrudescence compared to the CIF, and this overestimation was correlated with the proportion of new infections and recrudescences observed in a study.

The ACTs retained extremely good efficacy in the studies included in this thesis and hence despite the very large sample size, there were few recrudescent infections recorded. In order to explore scenarios of falling drug efficacy, simulation studies which utilised the failure-time distributions constructed in Chapter 4 confirmed that the impact of competing risk events was far greater in the areas of high transmission compared to the low transmission settings. The magnitude of the overestimation of drug failure increased as the drug efficacy started to wane - this being 1.75%, 3.13% and 4.30% when 5%, 10% and 15% recrudescences were observed, respectively in the areas of high transmission (Table 5.11). This has important clinical consequences. The current WHO strategy for monitoring and evaluation of antimalarial drug efficacy uses a threshold-based approach. It must be emphasised that the associated 95% confidence interval around the derived point estimates should be given equal importance as the point estimates themselves. For new drugs to be eligible for introduction as a first line treatment, derived failure estimates should be less than 5%, and for current first line treatment failure, estimates should not exceed 10% [59]. The conclusions of a clinical report are far less conclusive when the point estimates are on either side of these thresholds. For example, in three out of 233 study sites (of the data included in this thesis), the estimated failure was greater than 10% when the K-M method was used, but remained below 10% when using the competing risk survival analysis approach. This highlights that ignoring competing risk events can result in potentially misleading conclusions being drawn from a clinical study, particularly in high transmission settings where a large fraction of patients may develop new infections during the follow-up period, thus confounding the derived efficacy estimates.

However, the use of the competing risk approach in antimalarial studies is not without limitations. While the recurrences can be considered as competing risk events in a strictly statistical sense, the plausibility of them being biological competing risk events should be carefully considered as discussed in Chapter 5. In order for a new infection to preclude

recrudescence, it should exhibit far more efficient parasitic multiplication compared to an already existing infection. A scenario when this can occur is when the host develops immunity against the existing parasites of a specific strain after the first malarial attack and the new infection might be due to a different parasite strain against which the host has no prior exposure. In reality, it is impossible to unravel these underlying *in vivo* scenarios from the dataset included in this analysis and the plausibility of a new infection outcompeting existing recrudescent parasites needs to be considered with healthy scepticism. However, it should be noted that when the endpoint of interest is a new infection, the occurrence of recrudescence will always constitute a competing risk event.

Inferences drawn from treatment efficacy studies can be further confounded by the ambiguity of genotyping used to classify the origin of a recurrence, and this can lead to an indeterminate outcome. The current WHO recommendation regarding dealing with an indeterminate outcome is to exclude these individuals from the Kaplan-Meier survival analysis, commonly termed as a complete case (CC) analysis. Excluding individuals with an indeterminate outcome leads to a loss in statistical power and precision, especially when a large fraction of recurrences are indeterminate. In Chapter 6, such inconclusive outcomes were treated as missing observations and the well-established techniques of multiple imputation (MI) and inverse probability weighting (IPW) were used to handle them; the large multi-centre The 4ABC Trial including 4,116 patients was used to highlight this scenario [139]. In a hypothetical scenario where 10, 30 and 45% of the recurrences observed in this trial were artificially induced as missing under different mechanisms, it was found that the CC approach resulted in an upwards biased estimate of drug efficacy by up to 1.7%. The magnitude of this overestimation was correlated with the proportion of recurrent outcomes classified as indeterminate. In contrast, the MI and IPW approaches were generally unbiased under all missingness scenarios and provided an increased precision. However, the feasibility of applying the missing data methods considered in this

thesis when estimating antimalarial drug efficacy from standard therapeutic efficacy studies (TES) warrants further attention. TES are often small-sized (enrolling at least 50 patients), and with highly effective treatment regimens, very few recurrences may be observed in any given trial. In these scenarios, the fitting of the imputation model (for MI) and the missingness model (for IPW) to a small number of recurrences may be unstable, as suggested by the results of a resampling study carried out in Chapter 6. The real benefit of these missing-data approaches can be harnessed in large studies (like the 4ABC Trial) and in large individual participant data meta-analysis (IPD-MA), where a substantial fraction of patients experience recurrences and indeterminate outcomes.

7.2 Project strengths

This project has several methodological strengths. First, the results are based on the large standardised database of antimalarial studies captured in the WWARN repository, representing more than two-thirds of patients enrolled in published studies with ACTs since the year 2000. This made it possible to carry out comprehensive research on optimising statistical approaches for dealing with parasitic recurrences. The data included in this study represent a wide range of geographical locations across areas of very different malaria endemicities. The sheer volume of the data allowed the delineation of trends for each of the individual drug regimens within each transmission setting, thus avoiding unnecessary speculation regarding the generalisability of the results. For example, the available data provided the largest number of recrudescences to characterise the failure-time distribution for ACTs in children less than 5 years in Africa and in population of all ages in Asia.

Second, this thesis provides a number of findings, novel to ACTs. This is the first study to characterise the temporal trend (hazard function) and construct the probability distribution of recrudescence following ACT regimens, thus making this work unique and valuable for

rational planning of malaria control activities. The use of the competing risk approach to handle new infections, and the use of MI and IPW in handling indeterminate outcomes when deriving K-M estimate of drug efficacy, both remain novel within the context of antimalarial research for ACTs.

Third, the results presented in this thesis are consistent with those reported in the existing medical and statistical literature. Stepniewska et al (2004) [77] and Martensson et al (2008) [247] previously explored the adequacy of the recommended follow-up duration for antimalarials using the per protocol cured proportion and highlighted the inadequacy of a 28 day follow-up with an AL regimen thus substantiating the findings presented in Chapter 4. The unique contribution of this thesis is that a survival analysis approach, the current recommended approach by the WHO for deriving antimalarial failure estimates, was applied using non-linear and fractional polynomial modelling to quantify failures which were allowed to occur beyond the follow-up period. Similarly, the results presented in Chapter 5 showed that the 1 minus K-M method overestimated failure compared to the CIF, which is well-documented in the non-malaria literature [80,322]. The results presented in Chapter 6 are consistent with Machekano et al (2008) [88] who reported similar findings regarding the utility of missing data approaches for handling indeterminate PCR outcomes, but for a non-ACT regimen.

Finally, the proposed alternative approaches for dealing with new infections and indeterminate PCR approaches are readily available in open source software (Table 2.6) and can easily be implemented by researchers in low resource settings. To facilitate this, I have made a detailed tutorial available together with analytical scripts in an open-source publication [74].

7.3 Project limitations

Limitations specific to the individual analyses have been discussed already in the respective chapters. There were some limitations applicable throughout this thesis which are discussed next.

First, it was assumed that parasites were sensitive to the ACTs in the studies included in this thesis. The majority of the studies included in this analysis were conducted in Africa where artemisinin resistance hasn't been reported yet, and the included studies from Asia were from the time before the emergence of artesminin resistance. The proportion of patients remaining parasitic on day 3 after treatment initiation is used as a proxy for establishing the sensitivity of the parasites to artemisinins. The observed parasite clearance in the studies included was fast with day 3 PPR being <10% in all of the studies included (Figure 3.4). Hence, the assumption of parasites being sensitive to the ACT regimen in the included studies can generally be considered true.

Second, there was heterogeneity in the studies included in terms of study design (supervision, randomisation, blinding), PCR genotyping procedures and transmission settings (Table 3.1). Best possible efforts were made to handle heterogeneity in the data considered in each individual chapter. In Chapter 4, the analysis was stratified by region (Asia, Africa) and was further restricted to a paediatric population less than five years in Africa. The confounding factors of age, mg/kg drug dosage and baseline parasite load were adjusted as fixed effects in the Cox model and unknown heterogeneity was accounted for by considering the study sites as random effects (shared frailty) in the Cox regression analysis (in Chapter 4). In Chapter 5, the simulation studies were stratified by transmission settings to control for the inevitable differential fractions of new infections which would be observed, and in Chapter 6, missing

outcomes were simulated conditional on the transmission intensities. However, despite the best possible efforts made, the presence of residual confounding cannot be ruled out.

Third, it was assumed that the genotyping outcome reflects the true treatment outcome. However, the genotyping procedure is prone to misclassification error, particularly in areas of intense transmission [50,249,323,324]. Greenhouse et al (2007) estimated that up to 45% of the recrudescences were actually misclassified new infections in two studies conducted in Africa [249]. They found that increasing the number of molecular markers improved the discriminatory ability of the PCR genotyping in the areas of moderate transmission. In the areas of high transmission, even using six markers was found to be inadequate. In order to account for this, the construction of the distribution of recrudescence in Chapter 4 was restricted to studies which used at least three molecular markers. However, a limitation of this approach is that this doesn't guarantee that the studies would have used all the three molecular markers for classifying the outcomes as the current WHO guidelines recommends using a sequential approach, starting with the highest discriminatory marker (*msp-2* or *glurp*) followed by *msp-1* [51]. This means that the genotyping procedure is stopped at the first detection of a new infection. All three markers are used only if each of the markers individually classifies the recurrence as a recrudescence. In a recent article, Prof. Felger and colleagues argue that such a sequential strategy will lead to an underestimation of drug failure [324]. In 44 field samples studied, when all three markers were used for genotyping, an additional 10 recrudescences were detected compared to the sequential approach currently recommended (see Table 2 of [324]). This highlights the vulnerability of derived estimates of efficacy to the definition and genotyping methodologies adopted. The data included in this thesis lacked the fragment length for each of the individual markers used for outcome classification, thus limiting an independent evaluation of the approach proposed by Prof. Felger and colleagues. A thorough consideration of this problem should incorporate the population allelic diversity and recurrences should be

classified by taking the probability of being re-infected with the same alleles into account, as was done in an early study [49]. However, information on the allelic diversity in the population is often unmeasured and the length of the markers often not presented. Addressing the question of potential misclassification error in PCR defined outcomes remained beyond the scope of the work conducted as a part of this thesis.

7.4 Implications for public health and antimalarial research

Considering the strengths and limitations discussed in sections 7.2 and 7.3, this thesis has a number of implications for clinicians, antimalarial researchers, and policy makers.

A longer follow-up duration is warranted to adequately capture recrudescence failure with artemether-lumefantrine (AL) and dihydroartemisinin-piperaquine (DP). Failure to adequately capture recrudescences has important consequences for an individual patient as this is associated with increased morbidity and mortality. The consequences are far greater from a public health perspective as recrudescence is an essential component for the selection and further transmission of *de novo* resistant parasites. Drug resistance evolves gradually and usually manifests as a late recrudescence in the early stages. Such late recrudescences are more likely to produce gametocytes thus contributing to the onwards transmission of potentially or partially drug resistant parasites. Early detection and appropriate treatment of late recrudescences is critical from both medical and prevention perspectives. A longer follow-up duration enables more of these late recrudescences to be captured, and provide effective case-management for clinicians treating patients.

Population level modelling studies designed to explore the utility of different ACT regimens for mass drug administration and targeted malaria elimination often use the data on the duration of *P. falciparum* infection derived from malariatherapy data in non-immune patients [325].

However, the duration of infection in non-immune patients may not be generalisable to the patient population living in malaria endemic countries. The characterisation of the empirical distribution of recrudescence for each ACT provides important information for mathematical and epidemiological modelling studies for rational planning and deployment of ACTs. In addition, the distribution of recrudescences provides important information to malaria researchers for designing comparative studies, especially when the terminal elimination half-lives of drug regimens under comparison are different. For example, when all other sources of variability are held constant, a day 28 comparison of AL against DP is severely biased against AL. Simulation studies in Chapter 4 comparing these two regimens in a comparative trial demonstrated that the relative risk of recrudescence failures fell almost to parity only from day 49 onwards. The comparative efficacy of two ACTs is further confounded by the differential fraction of new infections observed when the partner drugs have different terminal half-lives. The use of the competing risk approach to account for new infections provides a useful approach for effectively dealing with this issue.

Inadequate follow-up duration and the conventional approach of discarding the inconclusive genotypes both led to underestimation of treatment failure, whereas ignoring competing events led to an overestimation of failure. From a public health perspective, the underestimation of failure has potentially important clinical consequences as compared to overestimation. Overestimation of failure is more conservative in terms of monitoring drug resistance. However, such overestimation might prevent promising antimalarial drugs from being introduced as alternative treatment measures. Antimalarial researchers and national malaria control programmes looking for alternative regimens for replacing the existing (failing) first line therapy should be aware of these pitfalls associated with the current approach for deriving antimalarial efficacy.

Three drugs antimalarial regimens (Triple ACTs), where an additional partner drug is added to an ACT, are currently being considered in South East Asia. The results presented in this thesis can have important implications for the Triple ACTs and generalisation of the results from ACTs to Triple ACTs likely depends on the drug interaction between two partner drugs. If the two partner drugs are assumed to act synergistically, the net effect will be that the parasites are killed faster and that the overall proportion of recurrences will be lower with these three drug regimens. In this synergistic scenario, the optimal duration of follow-up would be determined by the partner drug with the longest terminal half-life, which would supress the parasites for a longer period of time. This would mean that a Triple ACT regimen of AL+ piperaquine should be followed up for 63 days. Detailed pharmacokinetics-pharmacodynamics modelling studies are required to explore this further [326]. Similarly, like for the ACTs, the usage of competing risk methodologies for handling new infections and the use of missing data approaches for handling indeterminate outcomes would provide a more accurate estimate of the efficacy for the triple therapy.

The results presented in Chapter 4 of this thesis can have further implications in terms of registration of antimalarial studies with regulatory bodies. For example, the registration of AL with the US FDA was based on data presented on the efficacy of this regimen at day 28. The results of this thesis clearly suggest that a day 42 estimate of the efficacy should be considered for AL and day 63 estimates for DP. Similarly, the results presented in Chapter 5 can have further design implications in terms of sample size calculations. Current sample size calculation when designing antimalarial studies ignore the presence of competing risk events. When the ACTs under comparison have partner drugs with different terminal elimination half-lives, the two arms will be different in terms of the proportion of new infections which would be observed and hence these should be appropriately accounted for when estimating the sample sizes required [327]. A recent FDA guidance has recognised that competing risk events poses challenges in clinical trials with multiple endpoints [328]. Although there is no malaria specific recommendation for regulatory purposes, results from the competing risk analysis can be presented as a part of sensitivity analyses.

7.5 Generalisability and future work

The findings of this thesis are generalisable to other patient populations with *P. falciparum*, *P. vivax*, and indeed other infectious diseases. For example, the characterisation of the distribution of failure times presented in Chapter 4 is currently being adopted for gauging the optimal duration of follow-up for pregnant women with uncomplicated *P. falciparum* malaria [329]. Such an approach can also be used to answer the questions regarding optimal duration of follow-up for other parasitic diseases such as schistosomiasis where the time of outcome assessment is unclear [330]. Similarly, the competing risk approach considered in Chapter 5 has application when the endpoint of interest is *P vivax* recurrence in patients treated for index *P. falciparum* infection [331]. The use of competing risk methodologies is particularly suitable in this scenario as *P. falciparum* infection has been suggested to acutely supress the *P. vivax* infection [332]. Finally, the implications of this research are particularly relevant for large pooled analyses [32–34], which would benefit more from the recommended methods proposed, since the impact on individual smaller studies with few recrudescences was modest.

This research opened up several avenues for further methodological developments in evaluation of antimalarial drug efficacy. As stated earlier, a thorough consideration of PCR adjusted efficacy should incorporate the population allele diversity, which is often unmeasured or not presented. This would require building a database of allelic frequencies and collecting raw data on the fragment length of the molecular markers used. Some progress has been made on this front, as a PCR database from antimalarial studies carried out at the Mahidol Oxford Tropical Medicine Research Unit (MORU) has been acquired. I envisage taking this further to build a comprehensive database of molecular markers to thoroughly explore the issues

regarding PCR genotyping. Establishing such a database would be critical for further refining the distribution of recrudescences presented in Chapter 4.

The simulation studies presented in Chapter 5 assumed independence between time to recrudescence and new infection thus simplifying the simulation set-up. These simulation studies can be expanded by introducing dependencies between the competing risk events as outlined by Beyersmann et al (2009) [333], and will be considered as a future work. In Chapter 6, I explored the suitability of advanced missing data methods in handling indeterminate outcomes in derivation of Kaplan-Meier estimates of drug efficacy [292]. I plan to extend this work and explore their applicability when estimating the CIF, and in estimating regression parameters for the cause-specific Cox model and sub-distribution hazard models.

7.6 Conclusions

The importance of the correct use of statistical methods when evaluating drug efficacy should not be overlooked. This thesis highlighted that the choice of analytical approaches led to variability around the derived estimates of drug efficacy. Antimalarial researchers and policy makers alike should be aware of the pitfalls associated with the commonly employed statistical methods. For a large part of the 20th century, per protocol cured proportions were used as the primary measure of antimalarial efficacy before Kaplan-Meier analysis was introduced as the preferred method in early 2000s. Such policy change was pioneered by large scaled pooled analyses of data from several studies. The WWARN standardised repository has ensured maximum utility of the primary data collected for a different purpose to answer the methodological questions proposed in this thesis.

References

1. Covell G, Coatney GR, Field JW SJ. Chemotherapy of malaria. Monogr Ser World Heal. Organ. 1955.

2. World Health Organization. World Malaria Report 2017. 2017.

3. Bruce-Chwatt LJ. Chemotherapy of Malaria. Second edition. Geneva: World Heal. Organ. Monogr. Ser. No. 27. 1986. p. 261.

4. Sinden RE, Gilles HM. Historical Outline. Essent. Malariol. 2002. p. 1–7.

5. Cox FE. History of the discovery of the malaria parasites and their vectors. Parasit. Vectors. 2010;3:5.

6. Bruce-Chwatt LJ. Alphonse Laveran's discovery 100 years ago and today's global fight against malaria. J. R. Soc. Med. 1981;74:531–6.

7. Hawking F. The clock of the malaria parasite. Sci. Am. 1970;222:123–31.

8. World Health Organization. Guidelines for the treatment of malaria: Third edition. Geneva, Switzerland; 2015.

9. World Health Organization. World Malaria Report 2016. World Heal. Organ. 2016.

10. White NJ, Pukrittayakamee S, Hien TT, Faiz MA, Mokuolu OA, Dondorp AM. Malaria. Lancet. 2014;383:723–35.

11. Lensen A, Bril A, van de Vegte M, van Gemert GJ, Eling W, Sauerwein R. Plasmodium falciparum: infectivity of cultured, synchronized gametocytes to mosquitoes. Exp. Parasitol. 1999;103:101–3.

12. 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;24:377–410.

13. White NJ. Antimalarial drug resistance. J. Clin. Invest. 2004;113:1084–92.

14. Warrell DA. Clinical features of malaria. Essent. Malariol. 2002. p. 191–205.

15. Qinghaosu Antimalaria Coordinating Research Group. Antimalaria studies on Qinghaosu. Chin. Med. J. (Engl). 1979;92:811–6.

16. White NJ. Clinical pharmacokinetics and pharmacodynamics of artemisinin and derivatives. Trans R Soc Trop Med Hyg. 1994;1141–3.

17. Price R., Nosten F, Luxemburger C, ter Kuile F., Paiphun L, Chongsuphajaisiddhi T, et al. Effects of artemisinin derivatives on malaria transmissibility. Lancet. 1996;347:1654–8.

18. Nosten F, Luxemburger C, ter Kuile F, Woodrow C, Eh JP, Chongsuphajaisiddhi T, et al. Treatment of multidrug-resistant Plasmodium falciparum malaria with 3- day artesunate-mefloquine combination. J. Infect. Dis. 1994;170:971–7.

19. White NJ, Olliaro PL. Strategies for the prevention of antimalarial drug resistance: rationale for combination chemotherapy for malaria. Parasitol. Today. 1996;12:399–401.

20. White NJ. Preventing antimalarial drug resistance through combinations. Drug Resist. Updat. 1998;1:3–9.

21. White N. Antimalarial drug resistance and combination chemotherapy. Philos Trans R Soc L. B Biol Sci. 1999;354:739–49.

22. White NJ. Drug resistance in malaria. Br. Med. Bull. 1998;54:703–15.

23. Adjuik M, Agnamey P, Babiker A, Borrmann S, Brasseur P, Cisse M, et al. Amodiaquine-artesunate versus amodiaquine for uncomplicated Plasmodium falciparum malaria in African children: A randomised, multicentre trial. Lancet. 2002;359:1365–72.

24. Dorsey G, Njama D, Kamya MR, Cattamanchi A, Kyabayinze D, Staedke SG, et al. Sulfadoxine/pyrimethamine alone or with amodiaquine or artesunate for treatment of uncomplicated malaria: a longitudinal randomised trial. Lancet. 2002;360:2031–8.

25. von Seidlein L, Milligan P, Pinder M, Bojang K, Anyalebechi C, Gosling R, et al. Efficacy of artesunate plus pyrimethamine-sulphadoxine for uncomplicated malaria in Gambian children: a double-blind, randomised, controlled trial. Lancet. 2000;355:352–7.

26. World Health Organization. Antimalarial drug combination therapy: Report of a WHO Technical Consultation. 2001.

27. International Artemisinin Study Group. Artesunate combinations for treatment of malaria: Meta-analysis. Lancet. 2004;363:9–17.

28. Attaran A, Barnes KI, Curtis C, D'Alessandro U, Fanello CI, Galinski MR, et al. WHO, the Global Fund, and medical malpractice in malaria treatment. Lancet. 2004;363:237–40.

29. Maude RJ, Woodrow CJ, White LJ. Artemisinin antimalarials: preserving the magic bullet. Drug Dev. Res. 2010;71:12–9.

30. World Health Organization. Guidelines for the treatment of malaria. Who/Htm/Mal/2006.1108 Isbn 978 92 4 154694 2. 2006;266 p.

31. World Health Organization. WHO Model List of Essential Medicines. Essent. Med. Heal. Prod. [Internet]. 2017;1–39. Available from: http://www.who.int/entity/medicines/publications/essentialmedicines/20th_EML2017_FINAL_amendedAug2017.pdf?ua=1

32. The WorldWide Antimalarial Resistance Network (WWARN) AS-AQ Study Group. The effect of dosing strategies on the therapeutic efficacy of artesunate-amodiaquine for uncomplicated malaria: a meta-analysis of individual patient data. BMC Med. 2015;13:66.

33. Worldwide Antimalarial Resistance Network (WWARN) AL Dose Impact Study Group. The effect of dose on the antimalarial efficacy of artemether–lumefantrine: a systematic review and pooled analysis of individual patient data. Lancet Infect. Dis. 2015;15:692–702.

34. The WorldWide Antimalarial Resistance Network (WWARN) DP Study Group. The Effect of Dosing Regimens on the Antimalarial Efficacy of Dihydroartemisinin-Piperaquine: A Pooled Analysis of Individual Patient Data. PLoS Med. Public Library of Science;

2013;10:1–17.

35. Bhatt S, Weiss DJ, Cameron E, Bisanzio D, Mappin B, Dalrymple U, et al. The effect of malaria control on *Plasmodium falciparum* in Africa between 2000 and 2015. Nature. 2015;1–9.

36. Ashley EA, Dhorda M, Fairhurst RM, Amaratunga C, Lim P, Suon S, et al. Spread of Artemisinin Resistance in Plasmodium falciparum Malaria. N. Engl. J. Med. 2014;371:411–23.

37. Sibley CH, Price RN. Monitoring antimalarial drug resistance: Applying lessons learned from the past in a fast-moving present. Int. J. Parasitol. Drugs Drug Resist. Australian Society for Parasitology; 2012;2:126–33.

38. Barnes KI, Watkins WM, White NJ. Antimalarial dosing regimens and drug resistance. Trends Parasitol. 2008. p. 127–34.

39. Price RN, Nosten F. Drug resistant falciparum malaria: clinical consequences and strategies for prevention. Drug Resist. Updat. 2001;4:187–96.

40. Nsanzabana C, Djalle D, Guerin PJ, Menard D, González IJ. Tools for surveillance of anti-malarial drug resistance: an assessment of the current landscape. Malar. J. 2018;17:1–16.

41. Payne D. Spread of chloroquine resistance in Plasmodium falciparum. Parasitol. Today. 1987;3:241–6.

42. Hurwitz ES, Johnson D, Campbell CC. Resistance of Plasmodium Falciparum Malaria To Sulfadoxine-Pyrimethamine (Fansidar) in a Refugee Camp in Thailand. Lancet. 1981;317:1068–70.

43. Dondorp AM, Nosten F, Yi P, Das D, Phyo AP, Tarning J, et al. Artemisinin resistance in Plasmodium falciparum malaria. N. Engl. J. Med. 2009;361:455–67.

44. Leang R, Barrette A, Bouth DM, Menard D, Abdur R, Duong S, et al. Efficacy of dihydroartemisinin-piperaquine for treatment of uncomplicated plasmodium falciparum and plasmodium vivax in Cambodia, 2008 to 2010. Antimicrob. Agents Chemother. 2013;57:818–26.

45. Saunders DL, Vanachayangkul P, Lon C. Dihydroartemisinin–Piperaquine Failure in Cambodia. N. Engl. J. Med. 2014;371:484–5.

46. Phuc BQ, Rasmussen C, Duong TT, Dong LT, Loi MA, Tarning J, et al. Treatment Failure of Dihydroartemisinin/Piperaquine for Plasmodium falciparum Malaria, Vietnam. Emerg. Infect. Dis. 2017;23:715–7.

47. Ariey F, Witkowski B, Amaratunga C, Beghain J, Langlois A-C, Khim N, et al. A molecular marker of artemisinin-resistant Plasmodium falciparum malaria. Nature. 2013;505:50–5.

48. Snounou G, Beck HP. The use of PCR genotyping in the assessment of recrudescence or reinfection after antimalarial drug treatment. Parasitol. Today. 1998;14:462–7.

49. Brockman A, Paul R, Anderson T, Hackford I, Phaiphun L, Looareesuwan S, et al. Application of genetic markers to the identification of recrudescent Plasmodium falciparum

infections on the northwestern border of Thailand. Am J Trop Med Hyg. 1999;60:14–21.

50. Juliano JJ, Gadalla N, Sutherland CJ, Meshnick SR. The perils of PCR: can we accurately "correct" antimalarial trials? Trends Parasitol. 2010;26:119–24.

51. World Health Organization. Methods and techniques for clinical trials on antimalarial drug efficacy : genotyping to identify parasite populations. Geneva, Switzerland; 2007.

52. Lell B, Lehman LG, Schmidt-Ott JR, Sturchler D, Handschin J, Kremsner PG. Malaria chemotherapy trial at a minimal effective dose of mefloquine/sulfadoxine/pyrimethamine compared with equivalent doses of sulfadoxine/pyrimethamine or mefloquine alone. Am. J. Trop. Med. Hyg. 1998;58:619–24.

53. Bloland PB. Drug resistance in malaria. World Heal. Organ. 2001.

54. Terlouw DJ, Nahlen BL, Courval JM, Kariuki SK, Rosenberg OS, Oloo AJ, et al. Sulfadoxine-pyrimethamine in treatment of malaria in Western Kenya: increasing resistance and underdosing. Antimicrob. Agents Chemother. 2003;47:2929–32.

55. Terlouw DJ, Courval JM, Kolczak MS, Rosenberg OS, Oloo AJ, Kager P a, et al. Treatment history and treatment dose are important determinants of sulfadoxine-pyrimethamine efficacy in children with uncomplicated malaria in Western Kenya. J. Infect. Dis. 2003;187:467–76.

56. White NJ, Pongtavornpinyo W, Maude RJ, Saralamba S, Aguas R, Stepniewska K, et al. Hyperparasitaemia and low dosing are an important source of anti-malarial drug resistance. Malar. J. 2009;8:253.

57. Elliott I, Mayxay M, Yeuichaixong S, Lee SJ, Newton PN. The practice and clinical implications of tablet splitting in international health. Trop. Med. Int. Heal. 2014;

58. White NJ. The assessment of antimalarial drug efficacy. Trends Parasitol. 2002;18:458–64.

59. World Health Organization. Responding to antimalarial drug resistance [Internet]. World Heal. Organ. 2017 [cited 2017 Dec 5]. Available from: http://www.who.int/malaria/areas/drug_resistance/overview/en/

60. Verret WJ, Dorsey G, Nosten F, Price RN. The effect of varying analytical methods on estimates of anti-malarial clinical efficacy. Malar. J. 2009;8:77.

61. World Health Organization. Resistance of Malaria Parasite to Drugs. Geneva, Switzerland; 1965.

62. World Health Organization. Assessment and Monitoring of Antimalarial Drug Efficacy for the Treatment of Uncomplicated Falciparum Malaria. Geneva, Switzerland; 2003.

63. World Health Organization. Assesment of Therapeutic Efficacy of Antimalarial Drugs For Uncomplicated Falciparum Malaria in Areas with Intense Transmission. Geneva, Switzerland; 1996.

64. World Health Organization. Methods for Surveillance of Antimalarial Drug Efficacy. Geneva. Switzerland; 2009.

65. Yeka A, Banek K, Bakyaita N, Staedke SG, Kamya MR, Talisuna A, et al. Artemisinin versus nonartemisinin combination therapy for uncomplicated malaria: Randomized clinical trials from four sites in Uganda. PLoS Med. 2005;2:0654–62.

66. WorldWide Antimalarial Resistance Network. Data Management and Statistical Analysis Plan Version 1.2 [Internet]. www.wwarn.org. 2012 [cited 2017 Feb 10]. Available from: http://www.wwarn.org/sites/default/files/ClinicalDMSAP.pdf

67. 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;68:133–9.

68. Fontanet AL, Walker AM. Predictors of treatment failure in multiple drug-resistant falciparum malaria: results from a 42-day follow-up of 224 patients in eastern Thailand. Am-J-Trop-Med-Hyg. 1993;49:465–72.

69. Price RN, Nosten F, Luxemburger EK, Kham A, Brockman A, Chongsuphajaisiddhi T, et al. Artesunate versus artemether in combination with mefloquine for the treatment of multidrug-resistant falciparum malaria. Trans. R. Soc. Trop. Med. Hyg. 1995;89:523.

70. World Health Organization. Monitoring Antimalarial Drug Resistance. Geneva, Switzerland; 2001.

71. Stepniewska K, White NJ. Some considerations in the design and interpretation of antimalarial drug trials in uncomplicated falciparum malaria. Malar. J. 2006. p. 127.

72. Guthmann J-P, 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;50:3734–9.

73. Ashley EA, Pinoges L, Turyakira E, Dorsey G, Checchi F, Bukirwa H, et al. Different methodological approaches to the assessment of in vivo efficacy of three artemisinin-based combination antimalarial treatments for the treatment of uncomplicated falciparum malaria in African children. Malar. J. 2008;7:154.

74. 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. Malar. J. 2017;16:430.

75. World Health Organization. Chemotherapy of Malaria and Resistance to Antimalarials. World Heal. Organ. Techical Rep. Ser. 1973. p. 1–124.

76. Kremsner PG, Krishna S. Antimalarial combinations. Lacent. 2004;364:285–94.

77. Stepniewska K, Taylor WRJ, Mayxay M, Price R, Smithuis F, Guthmann J, et al. In Vivo Assessment of Drug Efficacy against Plasmodium falciparum Malaria : Duration of Follow-Up. Antimicrob. Agents Chemother. 2004;48:4271–80.

78. Mårtensson A, Strömberg J, Sisowath C, Msellem MI, Gil JP, Montgomery SM, et al. Efficacy of artesunate plus amodiaquine versus that of artemether-lumefantrine for the treatment of uncomplicated childhood Plasmodium falciparum malaria in Zanzibar, Tanzania.

Clin. Infect. Dis. 2005;41:1079–86.

79. Eshetu T, Abdo N, Bedru KH, Fekadu S, Wieser A, Pritsch M, et al. Open-label trial with artemether-lumefantrine against uncomplicated Plasmodium falciparum malaria three years after its broad introduction in Jimma Zone, Ethiopia. Malar. J. 2012;11:1–11.

80. Gooley TA, Leisenring W, Crowley J, Storer BE. Estimation of failure probabilities in the presence of competing risks: New representations of old estimators. Stat. Med. 1999;18:695–706.

81. Wolbers M, Koller MT, Stel VS, Schaer B, Jager KJ, Leffondre K, et al. Competing risks analyses: objectives and approaches. Eur. Heart J. 2014;35:2936–2941.

82. Andersen PK, Geskus RB, De witte T, Putter H. Competing risks in epidemiology: Possibilities and pitfalls. Int. J. Epidemiol. 2012;41:861–70.

83. Andersen PK, Abildstrom SZ, Rosthøj S. Competing risks as a multi-state model. Stat. Methods Med. Res. 2002;11:203–15.

84. Koller MT, Raatz H, Steyerberg EW, Wolbers M. Competing risks and the clinical community: Irrelevance or ignorance? Stat. Med. 2012;31:1089–97.

85. Lacny S, Wilson T, Clement F, Roberts DJ, Faris PD, Ghali WA, et al. Kaplan-Meier Survival Analysis Overestimates the Risk of Revision Arthroplasty: A Meta-analysis. Clin. Orthop. Relat. Res. 2015;473:3431–3442.

86. Lacny S, Wilson T, Clement F, Roberts DJ, Faris PD, Ghali WA, et al. Kaplan-Meier Survival Analysis Overestimates the Risk of Revision Arthroplasty: A Meta-analysis. Clin. Orthop. Relat. Res. 2015;473:3431–42.

87. Austin PC, Fine JP. Accounting for competing risks in randomized controlled trials : a review and recommendations for improvement. Stat. Med. 2017;36:1203–9.

88. Machekano R, Dorsey G, Hubbard A. Efficacy studies of malaria treatments in Africa: efficient estimation with missing indicators of failure. Stat. Methods Med. Res. 2007;17:191–206.

89. Giolo SR. Turnbull's nonparametric estimator for interval-censored data. Dep. Stat. Fed. Univ. Paraná. 2004. p. 1–10.

90. Turnbull BW. The Empirical Distribution Function with Arbitrarily Grouped , Censored and Truncated Data point in time is to be incorporated. J. R. Stat. Soc. 1975;38:290–5.

91. Lindsey JC, Ryan LM. Tutorial in biostatistics. Methods for interval-censored data. Stat. Med. 1998;17:219–38.

92. Gomez G, Calle ML, Oller R, Langohr K. Tutorial on methods for interval-censored data and their implementation in R. Stat. Modelling. 2009;9:259–97.

93. Singh RS, Totawattage DP. The Statistical Analysis of Interval-Censored Failure Time Data with Applications. Open J. Stat. 2013;2013:155–66.

94. Higgins JP, Green S. Cochrane Handbook for Systematic Reviews of Interventions Version 5.1.0. Cochrane Collab. 2008.

95. DerSimonian R, Laird N. Meta-analysis in clinical trials. Control. Clin. Trials. 1986;7:177–88.

96. Srinivasan C, Zhou M. A note on pooling Kaplan-Meier estimates. Biometrics. 1993;49:861–4.

97. Rubin DB. An Alternative to Pooling Kaplan-Meier Curves in Time-to-Event Meta-Analysis. Int. J. Biostat. 2011;7:1–26.

98. Combescure C, Foucher Y, Jackson D. Meta-analysis of single-arm survival studies: A distribution-free approach for estimating summary survival curves with random effects. Stat. Med. 2014;33:2521–37.

99. Foucher Y, Combescure C, Jackson D. R Package MetaSurv. 2014.

100. Klein JP, Logan B, Harhoff M, Andersen PK. Analyzing survival curves at a fixed point in time. Stat. Med. 2007;26:4505–19.

101. Arinaitwe E, Sandison TG, Wanzira H, Kakuru A, Homsy J, Kalamya J, et al. Artemether-Lumefantrine versus Dihydroartemisinin- Piperaquine for Falciparum Malaria: A Longitudinal, Randomized Trial in Young Ugandan Children. Clin. Infect. Dis. 2009;49:1629–37.

102. Collett D. Modelling Survival Data in Medical Research, Third Edition. Boca Raton: CRC Press Book; 2015.

103. Scrucca L, Santucci a, Aversa F. Competing risk analysis using R: an easy guide for clinicians. Bone Marrow Transplant. 2007;40:381–7.

104. Fay MP. Weighted Logrank Tests and NPMLE for interval censored data. R package "interval." CRAN Repos. 2015.

105. WorldWide Antimalarial Resistance Network. The WWARN Collaborator map [Internet]. www.wwarn.org. 2018 [cited 2018 Feb 10]. Available from: http://www.wwarn.org/working-together/collaborator-network

106. WorldWide Antimalarial Resistance Network. WWARN Clinical Data Management and Statistical Analysis Plan [Internet]. www.wwarn.org. 2011 [cited 2016 Jan 27]. Available from: http://www.wwarn.org/sites/default/files/ClinicalDMSAP.pdf

107. Pisani E, Botchway S. Sharing individual patient and parasite-level data through the WorldWide Antimalarial Resistance Network platform: A qualitative case study. Wellcome Open Res. 2017;2:63.

108. WWARN ACT Africa Baseline Study Group. Clinical determinants of early parasitological response to ACTs in African patients with uncomplicated falciparum malaria: a literature review and meta-analysis of individual patient data. BMC Med. 2015;13:212.

109. WWARN Gametocyte Study Group. Gametocyte carriage in uncomplicated Plasmodium falciparum malaria following treatment with artemisinin combination therapy: a systematic review and meta-analysis of individual patient data. BMC Med. 2016;14:79.

110. OxTREC. Oxford Tropical Research Ethics Committee (OxTREC) [Internet]. 2016 [cited 2016 Jan 7]. Available from: http://www.tropicalmedicine.ox.ac.uk/oxtrec

111. Gething PW, Patil AP, Smith DL, Guerra CA, Elyazar IRF, Johnston GL, et al. A new world malaria map: Plasmodium falciparum endemicity in 2010. Malar. J. 2011;10:378.

112. Bouyou-Akotet MK, Ramharter M, Ngoungou EB, Mamfoumbi MM, Mihindou MP, Missinou MA, et al. Efficacy and safety of a new pediatric artesunatemefloquine drug formulation for the treatment of uncomplicated falciparum malaria in Gabon. Wien. Klin. Wochenschr. 2010;122:173–8.

113. Menan H, Faye O, Same-Ekobo A, Oga ASS, Faye B, Kiki Barro CP, et al. Comparative study of the efficacy and tolerability of dihydroartemisinin - piperaquine - trimethoprim versus artemether - lumefantrine in the treatment of uncomplicated Plasmodium falciparum malaria in Cameroon, Ivory Coast and Senegal. Malar. J. 2011;10:185.

114. Mutabingwa TK, Anthony D, Heller A, Hallett R, Ahmed J, Drakeley C, et al. Amodiaquine alone, amodiaquine+sulfadoxine-pyrimethamine, amodiaquine+artesunate, and artemether-lumefantrine for outpatient treatment of malaria in Tanzanian children: A four-arm randomised effectiveness trial. Lancet. 2005;365:1474–80.

115. Hien TT, Thuy-Nhien NT, Phu NH, Boni MF, Thanh NV, Nha-Ca NT, et al. In vivo susceptibility of plasmodium falciparum to artesunate in Binh Phuoc Province, Vietnam. Malar. J. Malaria Journal; 2012;11:355.

116. Tine R, Faye B, Sylla K, Ndiaye JL, Ndiaye M, Sow D, et al. Efficacy and tolerability of a new formulation of artesunate-mefloquine for the treatment of uncomplicated malaria in adult in Senegal: Open randomized trial. Malar. J. 2012;11:415.

117. Piola P, Fogg C, Bajunirwe F, Biraro S, Grandesso F, Ruzagira E, et al. Supervised versus unsupervised intake of six-dose artemether-lumefantrine for treatment of acute, uncomplicated Plasmodium falciparum malaria in Mbarara, Uganda: A randomised trial. Lancet. 2005;365:1467–73.

118. Smithuis F, Kyaw MK, Phe O, Win T, Aung PP, Oo APP, et al. Effectiveness of five artemisinin combination regimens with or without primaquine in uncomplicated falciparum malaria: An open-label randomised trial. Lancet Infect. Dis. 2010;10:673–81.

119. Thanh NX, Trung TN, Phong NC, Thien NX, Dai B, Shanks GD, et al. Open label randomized comparison of dihydroartemisinin-piperaquine and artesunate-amodiaquine for the treatment of uncomplicated Plasmodium falciparum malaria in central Vietnam. Trop Med Int Heal. 2009;14:504–11.

120. Janssens B, van Herp M, Goubert L, Chan S, Uong S, Nong S, et al. A randomized open study to assess the efficacy and tolerability of dihydroartemisinin-piperaquine for the treatment of uncomplicated falciparum malaria in Cambodia. Trop Med Int Heal. 2007;12:251–9.

121. Falade C, Makanga M, Premji Z, Ortmann CE, Stockmeyer M, Ibarra de Palacios P. Efficacy and safety of artemether-lumefantrine (Coartem®) tablets (six-dose regimen) in African infants and children with acute, uncomplicated falciparum malaria. Trans. R. Soc. Trop. Med. Hyg. 2005;99:459–67.

122. Sawa P, Shekalaghe SA, Drakeley CJ, Sutherland CJ, Mweresa CK, Baidjoe AY, et al. Malaria transmission after artemether-lumefantrine and dihydroartemisinin- piperaquine: A randomized trial. J. Infect. Dis. 2013;207:1637–45.

123. Clark TD, Njama-Meya D, Nzarubara B, Maiteki-Sebuguzi C, Greenhouse B, Staedke SG, et al. Incidence of malaria and efficacy of combination antimalarial therapies over 4 years in an urban cohort of Ugandan children. PLoS One. 2010;5:1–8.

124. Tarning J, Zongo I, Somé F a, Rouamba N, Parikh S, Rosenthal PJ, et al. Population Pharmacokinetics and Pharmacodynamics of Piperaquine in Children With Uncomplicated Falciparum Malaria. Clin. Pharmacol. Ther. Nature Publishing Group; 2012;91:497–505.

125. Zongo I, Dorsey G, Rouamba N, Tinto H, Dokomajilar C, Guiguemde RT, et al. Artemether-lumefantrine versus amodiaquine plus sulfadoxine-pyrimethamine for uncomplicated falciparum malaria in Burkina Faso: a randomised non-inferiority trial. Lancet. 2007;369:491–8.

126. Faye B, Ndiaye J-L, Ndiaye D, Dieng Y, Faye O, Gaye O. Efficacy and tolerability of four antimalarial combinations in the treatment of uncomplicated Plasmodium falciparum malaria in Senegal. Malar. J. 2007;6:80.

127. Karema C, Fanello CI, Van Overmeir C, Van geertruyden JP, van Doren W, Ngamije D, et al. Safety and efficacy of dihydroartemisinin/piperaquine (Artekin®) for the treatment of uncomplicated Plasmodium falciparum malaria in Rwandan children. Trans. R. Soc. Trop. Med. Hyg. 2006;100:1105–11.

128. Abdulla S, Sagara I, Borrmann S, D'Alessandro U, Gonzalez R, Hamel M, et al. Efficacy and safety of artemether-lumefantrine dispersible tablets compared with crushed commercial tablets in African infants and children with uncomplicated malaria: a randomised, single-blind, multicentre trial. Lancet. 2008;372:1819–27.

129. Vugt M V, Wilairatana P, Gemperli B, Gathmann I, Phaipun L, Brockman A, et al. Efficacy of six doses of artemether-lumefantrine (benflumetol) in multidrug-resistant Plasmodium falciparum malaria. Am J Trop Med Hyg. 1999;60:936–42.

130. Sagara I, Fofana B, Gaudart J, Sidibe B, Togo A, Toure S, et al. Repeated artemisinin-based combination therapies in a malaria hyperendemic area of Mali: efficacy, safety, and public health impact. Am. J. Trop. Med. Hyg. 2012;87:50–6.

131. Ashley EA, Krudsood S, Phaiphun L, Srivilairit S, McGready R, Leowattana W, et al. Randomized, controlled dose-optimization studies of dihydroartemisinin-piperaquine for the treatment of uncomplicated multidrug-resistant falciparum malaria in Thailand. J. Infect. Dis. 2004;190:1773–82.

132. Anvikar AR, Sharma B, Shahi BH, Tyagi PK, Bose TK, Sharma SK, et al. Artesunate-amodiaquine fixed dose combination for the treatment of Plasmodium falciparum malaria in India. Malar. J. 2012. p. 97.

133. Valecha N, Srivastava P, Mohanty SS, Mittra P, Sharma SK, Tyagi PK, et al. Therapeutic efficacy of artemether-lumefantrine in uncomplicated falciparum malaria in India. Malar. J. 2009;8:107.

134. Agarwal A, McMorrow M, Onyango P, Otieno K, Odero C, Williamson J, et al. A randomized trial of artemether-lumefantrine and dihydroartemisinin-piperaquine in the treatment of uncomplicated malaria among children in western Kenya. Malar. J. Malaria Journal; 2013;12:254.

135. Grande T, Bernasconi A, Erhart A, Gamboa D, Casapia M, Delgado C, et al. A randomised controlled trial to assess the efficacy of dihydroartemisinin-piperaquine for the treatment of uncomplicated falciparum malaria in Peru. PLoS One. 2007;2:e1101.

136. Kamugisha E, Jing S, Minde M, Kataraihya J, Kongola G, Kironde F, et al. Efficacy of artemether-lumefantrine in treatment of malaria among under-fives and prevalence of drug resistance markers in Igombe-Mwanza, north-western Tanzania. Malar. J. BioMed Central Ltd; 2012;11:58.

137. 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;59:437–43.

138. Lefèvre G, Looareesuwan S, Treeprasertsuk S, Krudsood S, Silachamroon U, Gathmann I, et al. A clinical and pharmacokinetic trial of six doses of artemether-lumefantrine for multidrug-resistant Plasmodium falciparum malaria in Thailand. Am. J. Trop. Med. Hyg. 2001;64:247–56.

139. The Four Artemisinin-Based Combinations (4ABC) Study group. A head-to-head comparison of four artemisinin-based combinations for treating uncomplicated malaria in african children: A randomized trial. PLoS Med. 2011;8.

140. van den Broek I V, Maung UA, Peters A, Liem L, Kamal M, Rahman M, et al. Efficacy of chloroquine + sulfadoxine-pyrimethamine, mefloquine + artesunate and artemether + lumefantrine combination therapies to treat Plasmodium falciparum malaria in the Chittagong Hill Tracts, Bangladesh. Trans. R. Soc. Trop. Med. Hyg. 2005;99:727–35.

141. Ngasala BE, Malmberg M, Carlsson AM, Ferreira PE, Petzold MG, Blessborn D, et al. Effectiveness of artemether-lumefantrine provided by community health workers in under-five children with uncomplicated malaria in rural Tanzania: An open label prospective study. Malar. J. 2011;10:1–10.

142. Carrasquilla G, Barón C, Monsell EM, Cousin M, Walter V, Lefèvre G, et al. Randomized, prospective, three-arm study to confirm the auditory safety and efficacy of artemether-lumefantrine in Colombian patients with uncomplicated Plasmodium falciparum malaria. Am J Trop Med Hyg. 2012;86:75–83.

143. Mayxay M, Khanthavong M, Lindegardh N, Keola S, Barends M, Pongvongsa T, et al. Randomized Comparison of Chloroquine plus Sulfadoxine-Pyrimethamine versus Artesunate plus Mefloquine versus Artemether-Lumefantrine in the Treatment of Uncomplicated Falciparum Malaria in the Lao People's Democratic Republic. Clin. Infect. Dis. 2004;39:1139–47.

144. Adam I, Salah MT, Eltahir HG, Elhassan a H, Elmardi K a, Malik EM. Dihydroartemisinin-piperaquine versus artemether-lumefantrine, in the treatment of uncomplicated Plasmodium falciparum malaria in central Sudan. Ann. Trop. Med. Parasitol. 2010;104:319–26.

145. Djallé D, Njuimo SP, Manirakiza A, Laganier R, Le Faou A, Rogier C. Efficacy and safety of artemether + lumefantrine, artesunate + sulphamethoxypyrazine-pyrimethamine and artesunate + amodiaquine and sulphadoxine-pyrimethamine + amodiaquine in the treatment of uncomplicated falciparum malaria in Bangui, Central African Re. Malar. J. 2014;13:1–9.

146. Bassat Q, Mulenga M, Tinto H, Piola P, Borrmann S, Menéndez C, et al. Dihydroartemisinin-piperaquine and artemether-lumefantrine for treating uncomplicated malaria in African children: A randomised, non-inferiority trial. PLoS One. 2009;4.

147. Achan J, Tibenderana JK, Kyabayinze D, Wabwire Mangen F, Kamya MR, Dorsey G, et al. Effectiveness of quinine versus artemether-lumefantrine for treating uncomplicated falciparum malaria in Ugandan children: randomised trial. Br. Med. J. 2009;339:b2763.

148. Sirima SB, Tiono AB, Gansané A, Diarra A, Ouédraogo A, Konaté AT, et al. The efficacy and safety of a new fixed-dose combination of amodiaquine and artesunate in young African children with acute uncomplicated Plasmodium falciparum. Malar. J. 2009;8:48.

149. Haque R, Thriemer K, Wang Z, Sato K, Wagatsuma Y, Salam MA, et al. Therapeutic efficacy of artemether-lumefantrine for the treatment of uncomplicated Plasmodium falciparum malaria in Bangladesh. Am J Trop Med Hyg. 2007;76:39–41.

150. Smithuis F, Kyaw MK, Phe O, Aye KZ, Htet L, Barends M, et al. Efficacy and effectiveness of dihydroartemisinin-piperaquine versus artesunate-mefloquine in falciparum malaria : an open-label randomised comparison. Lacent. 2006;367:2075–85.

151. Mens PF, Sawa P, Van Amsterdam SM, Versteeg I, Omar SA, Schallig HDFH, et al. A randomized trial to monitor the efficacy and effectiveness by QT-NASBA of artemether-lumefantrine versus dihydroartemisinin-piperaquine for treatment and transmission control of uncomplicated Plasmodium falciparum malaria in western Kenya. Malar. J. 2008;7:237.

152. Yeka A, Dorsey G, Kamya MR, Talisuna A, Lugemwa M, Rwakimari JB, et al. Artemether-lumefantrine versus dihydroartemisinin-piperaquine for treating uncomplicated malaria: a randomized trial to guide policy in Uganda. PLoS One. 2008;3:e2390.

153. Premji Z, Umeh RE, Owusu-Agyei S, Esamai F, Ezedinachi EU, Oguche S, et al. Chlorproguanil-dapsone-artesunate versus artemether-lumefantrine: a randomized, double-blind phase III trial in African children and adolescents with uncomplicated Plasmodium falciparum malaria. PLoS One. 2009;4:e6682.

154. Mayxay M, Khanthavong M, Chanthongthip O, Imwong M, Pongvongsa T, Hongvanthong B, et al. Efficacy of artemether-lumefantrine, the nationally-recommended artemisinin combination for the treatment of uncomplicated falciparum malaria, in southern Laos. Malar. J. 2012;11:1–10.

155. Mayxay M, Thongpraseuth V, Khanthavong M, Lindegårdh N, Barends M, Keola S, et al. An open, randomized comparison of artesunate plus mefloquine vs. dihydroartemisinin-piperaquine for the treatment of uncomplicated Plasmodium falciparum malaria in the Lao People's Democratic Republic (Laos). Trop. Med. Int. Heal. 2006;11:1157–65.

156. Sagara I, Diallo A, Kone M, Coulibaly M, Diawara SI, Guindo O, et al. A randomized trial of artesunate-mefloquine versus artemether-lumefantrine for treatment of uncomplicated Plasmodium falciparum malaria in Mali. Am J Trop Med Hyg. 2008;79:655–61.

157. Gadalla NB, Adam I, Elzaki SE, Bashir S, Mukhtar I, Oguike M, et al. Increased pfmdr1 copy number and sequence polymorphisms in Plasmodium falciparum isolates from Sudanese malaria patients treated with artemether-lumefantrine. Antimicrob. Agents Chemother. 2011;55:5408–11.

158. Jullien V, Ogutu B, Juma E, Carn G, Obonyo C, Kiechel JR. Population pharmacokinetics and pharmacodynamic considerations of amodiaquine and desethylamodiaquine in Kenyan adults with uncomplicated malaria receiving artesunate-amodiaquine combination therapy. Antimicrob. Agents Chemother. 2010;54:2611–7.

159. van den Broek I, Kitz C, Al Attas S, Libama F, Balasegaram M, Guthmann JP. Efficacy of three artemisinin combination therapies for the treatment of uncomplicated Plasmodium falciparum malaria in the Republic of Congo. Malar. J. 2006;5:113.

160. Espié E, Lima A, Atua B, Dhorda M, Flévaud L, Sompwe EM, et al. Efficacy of fixed-dose combination artesunate-amodiaquine versus artemether-lumefantrine for uncomplicated childhood Plasmodium falciparum malaria in Democratic Republic of Congo: A randomized non-inferiority trial. Malar. J. 2012;11:174.

161. Valecha N, Phyo AP, Mayxay M, Newton PN, Krudsood S, Keomany S, et al. An Open-Label, Randomised Study of Dihydroartemisinin-Piperaquine Versus Artesunate-Mefloquine for Falciparum Malaria in Asia. PLoS One. 2010;5:e11880.

162. Borrmann S, Sasi P, Mwai L, Bashraheil M, Abdallah A, Muriithi S, et al. Declining responsiveness of plasmodium falciparum infections to Artemisinin-Based combination treatments on the Kenyan coast. PLoS One. 2011;6.

163. Sutanto I, Suprijanto S, Kosasih A, Dahlan MS, Syafruddin D, Kusriastuti R, et al. The effect of primaquine on gametocyte development and clearance in the treatment of uncomplicated falciparum malaria with dihydroartemisinin- piperaquine in South Sumatra, Western Indonesia: An open-label, randomized, controlled trial. Clin. Infect. Dis. 2013;56:685–93.

164. Ndiaye JLA, Faye B, Gueye A, Tine R, Ndiaye D, Tchania C, et al. Repeated treatment of recurrent uncomplicated Plasmodium falciparum malaria in Senegal with fixed-dose artesunate plus amodiaquine versus fixed-dose artemether plus lumefantrine: A randomized, open-label trial. Malar. J. 2011;10:237.

165. Sirima SB, Ogutu B, Lusingu JPA, Mtoro A, Mrango Z, Ouedraogo A, et al. Comparison of artesunate-mefloquine and artemether-lumefantrine fixed-dose combinations for treatment of uncomplicated Plasmodium falciparum malaria in children younger than 5 years in sub-Saharan Africa: A randomised, multicentre, phase 4 trial. Lancet Infect. Dis. 2016;16:1123–33.

166. Faye B, Offianan AT, Ndiaye JL, Tine RC, Touré W, Djoman K, et al. Efficacy and tolerability of artesunate-amodiaquine (Camoquin plus) versus artemether-lumefantrine (Coartem) against uncomplicated Plasmodium falciparum malaria: multisite trial in Senegal and Ivory Coast. Trop. Med. Int. Health. 2010;15:608–13.

167. Toure OA, Penali LK, Yapi JD, Ako BA, Toure W, Djerea K, et al. A comparative, randomized clinical trial of artemisinin/naphtoquine twice daily one day versus artemether/lumefantrine six doses regimen in children and adults with uncomplicated falciparum malaria in Cote d'Ivoire. Malar. J. 2009;8:148.

168. Offianan AT, Assi SB, Coulibaly A, N'guessan LT, Ako AA, Kadjo FK, San MK PL. Assessment of the efficacy of first-line antimalarial drugs after 5 years of deployment by the National Malaria Control Programme in Côte d'Ivoire. Open Access J. Clin. Trials. 2011;3:Pages 67—76.

169. Yeka A, Lameyre V, Afizi K, Fredrick M, Lukwago R, Kamya MR, et al. Efficacy and safety of fixed-dose artesunate-amodiaquine vs. artemether-lumefantrine for repeated treatment of uncomplicated malaria in Ugandan children. PLoS One. Public Library of Science; 2014;9:e113311.

170. Carrara VI, Zwang J, Ashley EA, Price RN, Stepniewska K, Barends M, et al. Changes in the Treatment Responses to Artesunate-Mefloquine on the Northwestern Border of Thailand during 13 Years of Continuous Deployment. PLoS One. 2009;4:e4551.

171. Zongo I, Dorsey G, Rouamba N, Dokomajilar C, Sere Y, Rosenthal PJ, et al. Randomized Comparison of Amodiaquine plus Sulfadoxine-Pyrimethamine, Artemether-Lumefantrine, and Dihydroartemisinin-Piperaquine for the Treatment of Uncomplicated Plasmodium falciparum Malaria in Burkina Faso. Clin. Infect. Dis. 2007;45:1453–61.

172. Ursing J, Kofoed PE, Rodrigues A, Blessborn D, Thoft-Nielsen R, Björkman A, et al. Similar efficacy and tolerability of double-dose chloroquine and artemether-lumefantrine for treatment of Plasmodium falciparum infection in Guinea-Bissau: A randomized trial. J. Infect. Dis. 2011;203:109–16.

173. Martensson A, Ngasala B, Ursing J, Isabel Veiga M, Wiklund L, Membi C, et al. Influence of consecutive-day blood sampling on polymerase chain reaction-adjusted parasitological cure rates in an antimalarial-drug trial conducted in Tanzania. J. Infect. Dis. 2007;195:597–601.

174. Ashley EA, Lwin KM, McGready R, Simon WH, Phaiphun L, Proux S, et al. An open label randomized comparison of mefloquine-artesunate as separate tablets vs. a new co-formulated combination for the treatment of uncomplicated multidrug-resistant falciparum malaria in Thailand. Trop. Med. Int. Heal. 2006;11:1653–60.

175. Kamya MR, Yeka A, Bukirwa H, Lugemwa M, Rwakimari JB, Staedke SG, et al. Artemether-lumefantrine versus dihydroartemisinin-piperaquine for treatment of malaria: A randomized trial. PLoS Clin. Trials. 2007;2.

176. Ndiaye J, Randrianarivelojosia M, Sagara I, Brasseur P, Ndiaye I, Faye B, et al. Randomized, multicentre assessment of the efficacy and safety of ASAQ – a fixed-dose artesunate-amodiaquine combination therapy in the treatment of uncomplicated Plasmodium falciparum malaria. Malar. J. 2009;8:125.

177. Karunajeewa HA, Mueller I, Senn M, Lin E, Law I, Gomorrai PS, et al. A Trial of Combination Antimalarial Therapies in Children from Papua New Guinea. N. Engl. J. Med. 2008;359:2545–57.

178. Schramm B, Valeh P, Baudin E, Mazinda CS, Smith R, Pinoges L, et al. Tolerability and safety of artesunate-amodiaquine and artemether- lumefantrine fixed dose combinations for the treatment of uncomplicated Plasmodium falciparum malaria: Two open-label, randomized trials in Nimba County, Liberia. Malar. J. 2013;12:250.

179. Bousema JT, Schneider P, Gouagna LC, Drakeley CJ, Tostmann A, Houben R, et al. Moderate effect of artemisinin-based combination therapy on transmission of Plasmodium falciparum. J. Infect. Dis. 2006;193:1151–9.

180. Yavo W, Faye B, Kuete T, Djohan V, Oga SA, Kassi RR, et al. Multicentric assessment of the efficacy and tolerability of dihydroartemisinin-piperaquine compared to artemether-

lumefantrine in the treatment of uncomplicated Plasmodium falciparum malaria in sub-Saharan Africa. Malar. J. BioMed Central Ltd; 2011;10:1–8.

181. Faye B, Ndiaye JL, Tine R, Sylla K, Gueye A, Lô AC, et al. A randomized trial of artesunate mefloquine versus artemether lumefantrine for the treatment of uncomplicated Plasmodium falciparum malaria in Senegalese children. Am. J. Trop. Med. Hyg. 2010;82:140–4.

182. Hwang J, Alemayehu BH, Hoos D, Melaku Z, Tekleyohannes SG, Teshi T, et al. In vivo efficacy of artemether-lumefantrine against uncomplicated Plasmodium falciparum malaria in Central Ethiopia. Malar. J. 2011;10:209.

183. Ngasala BE, Malmberg M, Carlsson AM, Ferreira PE, Petzold MG, Blessborn D, et al. Efficacy and effectiveness of artemether-lumefantrine after initial and repeated treatment in children < 5 years of age with acute uncomplicated plasmodium falciparum malaria in rural Tanzania: A randomized trial. Clin. Infect. Dis. 2011;52:873–82.

184. Khadime Sylla, Annie Abiola, Roger Clément Kouly Tine, Babacar Faye, Doudou Sow, Jean Louis Ndiaye, Magatte Ndiaye, Aminata Colé LO, Kuaku Folly LAN and OG. Monitoring the efficacy and safety of three artemisinin combinations therapies (ACT) in Senegal: Results from two years surveillance. Am. J. Trop. Med. Hyg. 2013;89:251.

185. Bukirwa H, Yeka A, Kamya MR, Talisuna A, Banek K, Bakyaita N, et al. Artemisinin Combination Therapies for Treatment of Uncomplicated Malaria in Uganda. PLoS Clin. Trials. 2006;1:e7.

186. Price RN, Uhlemann A-C, van Vugt M, Brockman A, Hutagalung R, Nair S, et al. Molecular and pharmacological determinants of the therapeutic response to artemether-lumefantrine in multidrug-resistant Plasmodium falciparum malaria. Clin. Infect. Dis. 2006;42:1570–7.

187. Guthmann JP, Cohuet S, Rigutto C, Fortes F, Saraiva N, Kiguli J, et al. High efficacy of two artemisinin-based combinations (artesunate + amodiaquine and artemether + lumefantrine) in Caala, Central Angola. Am. J. Trop. Med. Hyg. 2006;75:143–5.

188. Faucher J-F, Aubouy A, Adeothy A, Cottrell G, Doritchamou J, Gourmel B, et al. Comparison of sulfadoxine-pyrimethamine, unsupervised artemether-lumefantrine, and unsupervised artesunate-amodiaquine fixed-dose formulation for uncomplicated plasmodium falciparum malaria in Benin: a randomized effectiveness noninferiority trial. J. Infect. Dis. 2009;200:57–65.

189. Sinou V, Malaika LTM, Taudon N, Lwango R, Alegre SS, Bertaux L, et al. Pharmacokinetics and pharmacodynamics of a new ACT formulation: Artesunate/Amodiaquine (TRIMALACT) following oral administration in African malaria patients. Eur. J. Drug Metab. Pharmacokinet. 2009;34:133–42.

190. Mårtensson A, Strömberg J, Sisowath C, Msellem MI, Gil JP, Montgomery SM, et al. Efficacy of artesunate plus amodiaquine versus that of artemether-lumefantrine for the treatment of uncomplicated childhood Plasmodium falciparum malaria in Zanzibar, Tanzania. Clin. Infect. Dis. 2005;41:1079–86.

191. Fleiss J Paik M LB, Hoboken. Statistical methods for rates and proportions. 3rd ed. New Jersey: Wiley; 2003.

192. Dahal P. R script for confidence interval for porpotion accounting for cluster effect [Internet]. [cited 2018 Mar 23]. Available from: https://github.com/PrabinDahal/FleissInterval

193. WorldWide Antimalarial Resistance Network. WWARN Clinical Trials Publication Library [Internet]. www.wwarn.org. 2018 [cited 2018 Feb 16]. Available from: http://www.wwarn.org/tools-resources/literature-reviews/wwarn-clinical-trials-publication-library

194. Adjei GO, Goka BQ, Enweronu-Laryea CC, Rodrigues OP, Renner L, Sulley AM, et al. A randomized trial of artesunate-amodiaquine versus artemether-lumefantrine in Ghanaian paediatric sickle cell and non-sickle cell disease patients with acute uncomplicated malaria. Malar. J. 2014;13:1–14.

195. Ashley EA, Stepniewska K, Lindegårdh N, McGready R, Annerberg A, Hutagalung R, et al. Pharmacokinetic study of artemether-lumefantrine given once daily for the treatment of uncomplicated multidrug-resistant falciparum malaria. Trop. Med. Int. Health. 2007;12:201–8.

196. Chandra R, Ansah P, Sagara I, Sie A, Tiono AB, Djimde AA, et al. Comparison of azithromycin plus chloroquine versus artemether-lumefantrine for the treatment of uncomplicated Plasmodium falciparum malaria in children in Africa: A randomized, open-label study. Malar. J. 2015;14:1–10.

197. 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. Malar. J. BioMed Central; 2016;15:1–10.

198. Denoeud-Ndam L, Dicko A, Baudin E, Guindo O, Grandesso F, Diawara H, et al. Efficacy of artemether-lumefantrine in relation to drug exposure in children with and without severe acute malnutrition: An open comparative intervention study in Mali and Niger. BMC Med. BMC Medicine; 2016;14:1–14.

199. Gbotosho GO, Sowunmi A, Happi CT, Okuboyejo TM. Therapeutic efficacies of artemisinin-based combination therapies in nigerian children with uncomplicated falciparum malaria during five years of adoption as first-line treatments. Am. J. Trop. Med. Hyg. 2011;84:936–43.

200. Gbotosho GO, Sowunmi A, Okuboyejo TM, Happi CT, Folarin OA, Michael OS, et al. Therapeutic efficacy and effects of artemether-lumefantrine and artesunate-amodiaquine coformulated or copackaged on malaria-associated anemia in children with uncomplicated Plasmodium falciparum malaria in Southwest Nigeria. Am. J. Trop. Med. Hyg. 2011;84:813–9.

201. Hutagalung R, Paiphun L, Ashley EA, McGready R, Brockman A, Thwai KL, et al. A randomized trial of artemether-lumefantrine versus mefloquine-artesunate for the treatment of uncomplicated multi-drug resistant Plasmodium falciparum on the western border of Thailand. Malar. J. 2005;4:1–6.

202. Laman M, Moore BR, Benjamin JM, Yadi G, Bona C, Warrel J, et al. Artemisinin-naphthoquine versus artemether-lumefantrine for uncomplicated malaria in Papua New Guinean children: an open-label randomized trial. PLoS Med. 2014;11:e1001773.

203. Michael OS, Gbotosho GO, Folarin OA, Okuboyejo T, Sowunmi A, Oduola AM, et al. Early variations in plasmodium falciparum dynamics in Nigerian children after treatment with two artemisinin-based combinations: Implications on delayed parasite clearance. Malar. J. 2010;9:1–8.

204. Mulenga M, Van Geertruyden JP, Mwananyanda L, Chalwe V, Moerman F, Chilengi R, et al. Safety and efficacy of lumefantrine-artemether (Coartem) for the treatment of uncomplicated Plasmodium falciparum malaria in Zambian adults. Malar. J. 2006;5:1–7.

205. Nambozi M, Van Geertruyden JP, Hachizovu S, Chaponda M, Mukwamataba D, Mulenga M, et al. Safety and efficacy of dihydroartemisinin-piperaquine versus artemether-lumefantrine in the treatment of uncomplicated Plasmodium falciparum malaria in Zambian children. Malar. J. 2011;10:1–9.

206. Nji AM, Ali IM, Moyeh MN, Ngongang EO, Ekollo AM, Chedjou JP, et al. Randomized non-inferiority and safety trial of dihydroartemisin-piperaquine and artesunateamodiaquine versus artemether-lumefantrine in the treatment of uncomplicated Plasmodium falciparum malaria in Cameroonian children. Malar. J. 2015;14:1–10.

207. Ogouyèmi-Hounto A, Azandossessi C, Lawani S, Damien G, De Tove YSS, Remoue F, et al. Therapeutic efficacy of artemether-lumefantrine for the treatment of uncomplicated falciparum malaria in northwest Benin. Malar. J. BioMed Central; 2016;15:1–8.

208. Ogutu BR, Onyango KO, Koskei N, Omondi EK, Ongecha JM, Otieno GA, et al. Efficacy and safety of artemether-lumefantrine and dihydroartemisinin-piperaquine in the treatment of uncomplicated Plasmodium falciparum malaria in Kenyan children aged less than five years: results of an open-label, randomized, single-centre study. Malar. J. 2014;13:1–10.

209. Okafor HU, Shu EN, Oguonu T. Therapeutic efficacy and effect on gametocyte carriage of an artemisinin and a non-based combination treatment in children with uncomplicated P. falciparum malaria, living in an area with high-level chloroquine resistance. J. Trop. Pediatr. 2010;56:398–406.

210. Rahman MM, Dondorp AM, Day NPJ, Lindegardh N, Imwong M, Faiz MA, et al. Adherence and efficacy of supervised versus non-supervised treatment with artemether/lumefantrine for the treatment of uncomplicated Plasmodium falciparum malaria in Bangladesh: a randomised controlled trial. Trans. R. Soc. Trop. Med. Hyg. 2008;102:861–7.

211. Ratcliff A, Siswantoro H, Kenangalem E, Maristela R, Wuwung R, Laihad F, et al. Two fixed-dose artemisinin combinations for drug-resistant falciparum and vivax malaria in Papua, Indonesia: an open-label randomised comparison. Lancet. 2007;369:757–65.

212. Saha P, Guha SK, Das S, Mullick S, Ganguly S, Biswas A, et al. Comparative efficacies of artemisinin combination therapies in Plasmodium falciparum malaria and polymorphism of pfATPase6, pfcrt, pfdhfr, and pfdhps genes in tea gardens of Jalpaiguri District, India. Antimicrob. Agents Chemother. 2012;56:2511–7.

213. Saha P, Naskar A, Ganguly S, Das S, Guha SK, Biswas A, et al. Therapeutic efficacy of artemisinin combination therapies and prevalence of S769N mutation in PfATPase6 gene of Plasmodium falciparum in Kolkata, India. Asian Pac. J. Trop. Med. Hainan Medical College; 2013;6:443–8.

214. Sow D, Ndiaye JL, Sylla K, Ba MS, Tine RCK, Faye B, Pene M, Ndiaye M, Seck A, Lo A C, Abiola A, Dieng Y GO. Evaluation of the efficacy and safety of three 2-drug combinations for the treatment of uncomplicated Plasmodium falciparum malaria in Senegal: artesunate-amodiaquine, dihydroartemisinin-piperaquine, and artemether- lumefantrine. Med. Sante Trop. 2016;26:45–50.

215. Sowunmi A, Gbotosho GO, Happi CT, Adedeji AA, Fehintola FA, Folarin OA, et al. Therapeutic efficacy and effects of artemether-lumefantrine and amodiaquine-sulfalene-pyrimethamine on gametocyte carriage in children with uncomplicated Plasmodium falciparum malaria in southwestern Nigeria. Am. J. Trop. Med. Hyg. 2007;77:235–41.

216. Stohrer M, Dittrich S, Thongpaseuth V, Vanisaveth V, Phetsouvanh R. Therapeutic efficacy of artemether-lumefantrine and artesunate-mefloquine for treatment of uncomplicated Plasmodium falciparum malaria in Luang Namtha Province, Lao People's Democratic Republic. Trop. Med. Int. Heal. 2004;9:1175–83.

217. Sykes A, Hendriksen I, Mtove G, Mandea V, Mrema H, Rutta B, et al. Azithromycin plus artesunate versus artemether-lumefantrine for treatment of uncomplicated malaria in Tanzanian children: a randomized, controlled trial. Clin. Infect. Dis. 2009;49:1195–201.

218. Thriemer K, Starzengruber P, Khan WA, Haque R, Marma ASP, Ley B, et al. Azithromycin combination therapy for the treatment of uncomplicated falciparum malaria in Bangladesh: an open-label randomized, controlled clinical trial. J. Infect. Dis. 2010;202:392–8.

219. 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. Trop. Med. Int. Heal. 2014;19:469–75.

220. Tiono AB, Tinto H, Alao MJ, Meremikwu M, Tshefu A, Ogutu B, et al. Increased systemic exposures of artemether and dihydroartemisinin in infants under 5 kg with uncomplicated Plasmodium falciparum malaria treated with artemether-lumefantrine (Coartem®). Malar. J. 2015;14:1–10.

221. Toure OA, Valecha N, Tshefu AK, Thompson R, Krudsood S, Gaye O, et al. A Phase 3, Double-Blind, Randomized Study of Arterolane Maleate-Piperaquine Phosphate vs Artemether-Lumefantrine for Falciparum Malaria in Adolescent and Adult Patients in Asia and Africa. Clin. Infect. Dis. 2016;62:964–71.

222. Tshefu AK, Gaye O, Kayentao K, Thompson R, Bhatt KM, Sesay SSS, et al. Efficacy and safety of a fixed-dose oral combination of pyronaridine-artesunate compared with artemetherlumefantrine in children and adults with uncomplicated plasmodium falciparum malaria: A randomised non-inferiority trial. Lancet. Elsevier Ltd; 2010;375:1457–67.

223. Ursing J, Rombo L, Rodrigues A, Kofoed P-E. Artemether-Lumefantrine versus Dihydroartemisinin-Piperaquine for Treatment of Uncomplicated Plasmodium falciparum Malaria in Children Aged Less than 15 Years in Guinea-Bissau – An Open-Label Non-Inferiority Randomised Clinical Trial. PLoS One. 2016;11:e0161495.

224. Yeka A, Lameyre V, Afizi K, Fredrick M, Lukwago R, Kamya MR, et al. Efficacy and safety of fixed-dose artesunate-amodiaquine vs. artemether-lumefantrine for repeated treatment of uncomplicated malaria in Ugandan children. PLoS One. 2014;9:e113311.

225. Ashley EA, Mcgready R, Hutagalung R, Phaiphun L, Slight T, Proux S, et al. A Randomized , Controlled Study of a Simple , Once- Daily Regimen of Dihydroartemisinin-Piperaquine for the Treatment of Uncomplicated , Multidrug- Resistant Falciparum Malaria. Clin. Infect. Dis. 2005;41:425–32.

226. Mayxay M, Keomany S, Khanthavong M, Souvannasing P, Stepniewska K, Khomthilath T, et al. A phase III, randomized, non-inferiority trial to assess the efficacy and safety of dihydroartemisinin-piperaquine in comparison with artesunate-mefloquine in patients with uncomplicated Plasmodium falciparum malaria in southern Laos. Am. J. Trop. Med. Hyg. 2010;83:1221–9.

227. Gargano N, Ubben D, Tommasini S, Bacchieri A, Corsi M, Bhattacharyya PC, et al. Therapeutic efficacy and safety of dihydroartemisinin-piperaquine versus artesunate-mefloquine in uncomplicated Plasmodium falciparum malaria in India. Malar. J. 2012;11:1–12.

228. Amaratunga C, Lim P, Suon S, Sreng S, Mao S, Sopha C, et al. Dihydroartemisinin-piperaquine resistance in Plasmodium falciparum malaria in Cambodia: A multisite prospective cohort study. Lancet Infect. Dis. Elsevier Ltd; 2016;16:357–65.

229. Espie E, Lima A, Matindii BA, Flevaud LL, Sompwe EM, Urrutia PPP, et al. Efficacy of fixed-dose combination artesunate-amodiaquine versus arthemether-lumefantrine for uncomplicated Plasmodium falciparum malaria in children under five: A Randomized non inferiority trial in democratic Republic of Congo. Am. J. Trop. Med. Hyg. 2011;Conference:357.

230. Ratsimbasoa A, Ravony H, Vonimpaisomihanta JA, Raherinjafy R, Jahevitra M, Rapelanoro R, et al. Compliance, safety, and effectiveness of fixed-dose artesunate-amodiaquine for presumptive treatment of non-severe malaria in the context of home management of malaria in Madagascar. Am. J. Trop. Med. Hyg. 2012;86:203–10.

231. Tietche F, Chelo D, Mina Ntoto NK, Djoukoue FM, Hatz C, Frey S, et al. Tolerability and efficacy of a pediatric granule formulation of artesunate-mefloquine in young children from Cameroon with uncomplicated falciparum malaria. Am. J. Trop. Med. Hyg. 2010;82:1034–40.

232. Royston P. Estimating a smooth baseline hazard function for the Cox model [Internet]. London; 2011. Available from: https://www.ucl.ac.uk/statistics/research/pdfs/rr314.pdf [Access date: 27/01/2016]

233. Royston P, Parmar MKB. Flexible parametric proportional-hazards and proportional-odds models for censored survival data, with application to prognostic modelling and estimation of treatment effects. Stat. Med. 2002;21:2175–97.

234. R Development Core Team R. R: A Language and Environment for Statistical Computing. R Found. Stat. Comput. 2017.

235. StataCorp. Stata Statistical Software: Release 13. 2013. 2013.

236. Collins WE, Geoffrey M. A retrospective examination of the patterns of recrudescence in patients infected with plasmodium falciparum. Am. J. Trop. Med. Hyg. 1999;61:44–8.

237. Simpson JA, Aarons L, Collins WE, Jeffery GM, White NJ. Population dynamics of

untreated Plasmodium falciparum malaria within the adult human host during the expansion phase of the infection. Parasitology. 2002;124:247–63.

238. Sama W, Dietz K, Smith T. Distribution of survival times of deliberate Plasmodium falciparum infections in tertiary syphilis patients. Trans. R. Soc. Trop. Med. Hyg. 2006;100:811–6.

239. Bretscher MT, Maire N, Chitnis N, Felger I, Owusu-Agyei S, Smith T. The distribution of Plasmodium falciparum infection durations. Epidemics. Elsevier B.V.; 2011;3:109–18.

240. White NJ. Assessment of the pharmacodynamic properties of antimalarial drugs in vivo. Antimicrob. Agents Chemother. 1997;41:1413–22.

241. Ezzet F, Vugt M Van, Nosten F, Looareesuwan S. Pharmacokinetics and Pharmacodynamics of Lumefantrine (Benflumetol) in Acute Falciparum Malaria Pharmacokinetics and Pharmacodynamics of Lumefantrine (Benflumetol) in Acute Falciparum Malaria. Antimicrob. Agents Chemother. 2000;44:697–704.

242. Salman S, Page-Sharp M, Griffin S, Kose K, Siba PM, Ilett KF, et al. Population Pharmacokinetics of Artemether, Lumefantrine, and Their Respective Metabolites in Papua New Guinean Children with Uncomplicated Malaria. Antimicrob. Agents Chemother. 2011;55:5306–13.

243. Tarning J, Lindegårdh N, Annerberg A, Singtoroj T, Day NPJ, Ashton M, et al. Pitfalls in estimating piperaquine elimination. Antimicrob. Agents Chemother. 2005;49:5127–8.

244. Tarning J, Ashley E a., Lindegardh N, Stepniewska K, Phaiphun L, Day NPJ, et al. Population pharmacokinetics of piperaquine after two different treatment regimens with dihydroartemisinm-piperaquine in patients with Plasmodium falciparum malaria in Thailand. Antimicrob. Agents Chemother. 2008;52:1052–61.

245. Tarning J, Lindegardh N, Lwin KM, Annerberg A, Kiricharoen L, Ashley E, et al. Population pharmacokinetic assessment of the effect of food on piperaquine bioavailability in patients with uncomplicated malaria. Antimicrob. Agents Chemother. 2014;58:2052–8.

246. Covell G. Relationship between malarial parasitaemia and symptoms of the disease: a review of the literature. Bull. World Health Organ. 1960;22:605–19.

247. Mårtensson A. PCR adjusted cure rates in clinical trials of antimalarial drugs in Africa-influence of extended follow-up and consecutive day blood sampling [Internet]. karolinska institutet; 2008. Available from: https://openarchive.ki.se/xmlui/bitstream/handle/10616/37930/thesis.pdf;sequence=1

248. Stepniewska K, White NJ. Pharmacokinetic determinants of the window of selection for antimalarial drug resistance. Antimicrob. Agents Chemother. 2008;52:1589–96.

249. 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;51:3096–103.

250. Bernoulli D. Essai d'une nouvelle analyse de la mortalite causee par la petite verole. Mem Math Phy Acad Roy Sci Paris. 1766;1.

251. Blower S, Bernoulli D. An attempt at a new analysis of the mortality caused by

smallpox and of the advantages of inoculation to prevent it. Rev. Med. Virol. 2004;14:275–88.

252. Cornfield J. The estimation of the probability of developing a disease in the presence of competing risks. Am. J. Public Health. 1957;47:601–7.

253. Prentice RL, Kalbfleisch JD, Peterson A V, Flournoy N, Farewell VT, Breslow NE. The analysis of failure times in the presence of competing risks. Biometrics. 1978;34:541–54.

254. Putter H, Fiocco M, Geskus RB. Tutorial in biostatistics: Competing risks and multi-state models. Stat. Med. 2007;26:4817–34.

255. Austin PC, Lee DS, Fine JP. Introduction to the Analysis of Survival Data in the Presence of Competing Risks. Circulation. 2016;133:601–9.

256. Dignam JJ, Kocherginsky MN. Choice and interpretation of statistical tests used when competing risks are present. J. Clin. Oncol. 2008;26:4027–34.

257. Kalbfleisch JD, Prentice RL. Competing Risks and Multistate Models. Stat. Anal. Fail. Time Data. 2nd ed. New York, USA: John Wiley and Sons Inc; 2002. p. 247–77.

258. Wolkewitz M, Cooper BS, Bonten MJM, Barnett AG, Schumacher M. Interpreting and comparing risks in the presence of competing events. BMJ. 2014;349:g5060.

259. Bajorunaite R, Klein JP. Comparison of failure probabilities in the presence of competing risks. J. Stat. Comput. Simul. 2008;78:951–66.

260. Gray RJ. A Class of K-Sample Tests for Comparing the Cumulative Incidence of a Competing Risk. Ann. Stat. 1988;16:1141–54.

261. Klein JP. Competing risks. Wiley Interdiscip. Rev. Comput. Stat. 2010. p. 333–9.

262. Fine JP, Gray RJ. A Proportional Hazards Model for the Subdistribution of a Competing Risk. J. Am. Stat. Assoc. 1999;94:196–509.

263. Lau B, Cole SR, Gange SJ. Competing risk regression models for epidemiologic data. Am. J. Epidemiol. 2009;170:244–56.

264. Austin PC, Fine JP. Practical recommendations for reporting Fine-Gray model analyses for competing risk data. Stat. Med. 2017;36:4391–400.

265. Zhou B, Fine J, Latouche A, Labopin M. Competing risks regression for clustered data. Biostatistics. 2012;13:371–83.

266. Ettore M, Valsecchi MG. Competing Risks. Anal. Surviv. Data from Clin. Trials Obs. Stud. Cichester, England: John Wiley & Sons, Inc.; 2004. p. 357.

267. R Foundation for Statistical Computing. R: A language and environment for statistical computing [Internet]. Vienna, Austria; 2017. Available from: https://www.r-project.org/

268. Collett D. Sample size requirements for a survival study. Model. Surviv. data Med. Res. Third. Boca Raton: Chapman & Hall/CRC Biostatistics Series; 2015. p. 471–84.

269. Van Walraven C, McAlister FA. Competing risk bias was common in Kaplan-Meier risk estimates published in prominent medical journals. J. Clin. Epidemiol. Elsevier;

2016;69:170–3.

270. Wolbers M, Koller MT, Witteman JCM, Steyerberg EW. Prognostic Models With Competing Risks. Epidemiology. 2009. p. 555–61.

271. Freidlin B, Korn EL. Testing treatment effects in the presence of competing risks. Stat. Med. 2005;24:1703–12.

272. Williamson PR, Kolamunnage-Dona R, Tudur Smith C. The influence of competing-risks setting on the choice of hypothesis test for treatment effect. Biostatistics. 2007;8:689–94.

273. Rotolo F, Michiels S. Testing the treatment effect on competing causes of death in oncology clinical trials. BMC Med. Res. Methodol. 2014;14:1–11.

274. Pintilie M. Analysing and interpreting competing risk data. Stat. Med. 2007;26:1360–7.

275. Tai B-C, Wee J, Machin D. Analysis and design of randomised clinical trials involving competing risks endpoints. Trials. BioMed Central Ltd; 2011;12:127.

276. Latouche A, Allignol A, Beyersmann J, Labopin M, Fine JP. A competing risks analysis should report results on all cause-specific hazards and cumulative incidence functions. J. Clin. Epidemiol. Elsevier Inc; 2013;66:648–53.

277. Nicolaie MA, Houwelingen H Van, Putter H. Vertical modeling: Analysis of competing risks data with missing causes of failure. Stat. Methods Med. Res. 2011;

278. Moreno-Betancur M, Latouche A. Regression modeling of the cumulative incidence function with missing causes of failure using pseudo-values. Stat. Med. 2013;32:3206–23.

279. Moreno-Betancur M, Rey G, Latouche A. Direct likelihood inference and sensitivity analysis for competing risks regression with missing causes of failure. Biometrics. 2015;71:498–507.

280. Heitjan DF, Little RJA. Multiple Imputation for the Fatal Accident Reporting System. J. R. Stat. Soc. Stat. 1991;40:13–29.

281. Donders ART, van der Heijden GJMG, Stijnen T, Moons KGM. Review: A gentle introduction to imputation of missing values. J. Clin. Epidemiol. 2006;59:1087–91.

282. Greenland S, Finkle WD. A Critical Look at Methods for Handling Missing Covariates in Epidemiologic Regression Analyses. Am. J. Epidemiol. 1995;142:1255–64.

283. Mugittu K, Adjuik M, Snounou G, Ntoumi F, Taylor W, Mshinda H, et al. Molecular genotyping to distinguish between recrudescents and new infections in treatment trials of Plasmodium falciparum malaria conducted in Sub-Saharan Africa: Adjustment of parasitological outcomes and assessment of genotyping effectiveness. Trop. Med. Int. Heal. 2006;11:1350–9.

284. Rubin DB, Schenker N. Multiple imputation in health-care databases : An overview and some applications. Stat. Med. 1991;10:585–98.

285. Rubin DB. Inference and missing data. Biometrika. 1976;63:581–92.

286. Rubin DB. Basic Ideas of Multiple Imputation for Nonresponse. Surv. Methodol. 1986;12:37–47.

287. Lee KJ, Simpson JA. Introduction to multiple imputation for dealing with missing data. Respirology. 2014;19:162–7.

288. White IR, Royston P, Wood AM. Multiple imputation using chained equations: Issues and guidance for practice. Stat. Med. 2011;30:377–99.

289. Seaman SR, White IR. Review of inverse probability weighting for dealing with missing data. Stat. Methods Med. Res. 2013;22:278–95.

290. Horvitz DG, Thompson D. A generalization of sampling without replacement from a finite universe. J. Am. Stat. Assoc. 1952;44:663–85.

291. Rosenbaum PR, Rubin DB. The Central Role of the Propensity Score in Observational Studies for Causal Effects. Biometrika. 1983;70:41–55.

292. Hayati Rezvan P, Lee KJ, Simpson JA. The rise of multiple imputation: a review of the reporting and implementation of the method in medical research. BMC Med. Res. Methodol. 2015;15:30.

293. Mukaka M, White SA, Terlouw DJ, Mwapasa V, Kalilani-Phiri L, Faragher EB. Is using multiple imputation better than complete case analysis for estimating a prevalence (risk) difference in randomized controlled trials when binary outcome observations are missing? Trials. 2016;17:341.

294. The PREGACT Study Group. Four artemisinin-based treatments in African pregnant women with malaria. N. Engl. J. Med. 2016;374:913–27.

295. The Four Artemisinin-Based Combinations Study group. A head-to-head comparison of four artemisinin-based combinations for treating uncomplicated malaria in african children: A randomized trial. PLoS Med. 2011;8.

296. Fanello CI, Karema C, Avellino P, Bancone G, Uwimana A, Lee SJ, et al. High risk of severe anaemia after chlorproguanil-dapsone+artesunate antimalarial treatment in patients with G6PD (A-) deficiency. PLoS One. 2008;3:e4031.

297. Shaukat AM, Gilliams EA, Kenefic LJ, Laurens MB, Dzinjalamala FK, Nyirenda OM, et al. Clinical manifestations of new versus recrudescent malaria infections following anti-malarial drug treatment. Malar. J. 2012;11:207.

298. Collins LM, Schafer JL, Kam C-M. A comparison of inclusive and restrictive strategies in modern missing data procedures. Psychol. Methods. 2001;6:330–51.

299. Sterne JAC, White IR, Carlin JB, Spratt M, Royston P, Kenward MG, et al. Multiple imputation for missing data in epidemiological and clinical research: potential and pitfalls. BMJ. 2009;338:b2393–b2393.

300. Bodner TE. What Improves with Increased Missing Data Imputations? Struct. Equ. Model. A Multidiscip. J. 2008;15:651–75.

301. Molenberghs G, Kenward MG. 9. Multiple Imputation. Missing Data Clin. Stud. First. Chichester, England: John Wiley & Sons, Inc.; 2007. p. 108.

302. Buuren S van. Flexible imputation of missing data. Chapman Hall/CRC Interdiscip. Stat. Ser. 2012.

303. Austin PC. Variance estimation when using inverse probability of treatment weighting (IPTW) with survival analysis. Stat. Med. 2016;35:5642–55.

304. Marshall A, Altman DG, Holder RL. Comparison of imputation methods for handling missing covariate data when fitting a Cox proportional hazards model: a resampling study. BMC Med. Res. Methodol. BioMed Central Ltd; 2010;10:112.

305. Harel O, Zhou X-H. Multiple imputation: Review of theory, implementation and software. Stat. Med. 2007;26:3057–77.

306. Allison PD. Multiple Imputation for Missing Data: A Cautionary Tale. Sociol. Methods Res. 2000;28:301–9.

307. Carpenter JR, Kenward MG, Vansteelandt S. A comparison of multiple imputation and inverse probability weighting for analyses with missing data. J. R. Stat. Soc. Ser. A. 2006;169:571–84.

308. Höfler M, Pfister H, Lieb R, Wittchen HU. The use of weights to account for non-response and drop-out. Soc. Psychiatry Psychiatr. Epidemiol. 2005;40:291–9.

309. Williamson EJ, Forbes A, White IR. Variance reduction in randomised trials by inverse probability weighting using the propensity score. Stat. Med. 2014;33:721–37.

310. Graham JW. Missing data analysis: making it work in the real world. Annu. Rev. Psychol. 2009;60:549–76.

311. Wittes J. Missing inaction: preventing missing outcome data in randomized clinical trials. J. Biopharm. Stat. 2009;19:957–68.

312. White IR, Horton NJ, Carpenter J, Pocock SJ. Strategy for intention to treat analysis in randomised trials with missing outcome data. Br. Med. J. 2011;342:910–2.

313. Little RJ, Agostino RD, Cohen ML, Dickersin K, Emerson SS, Farrar JT, et al. The Prevention and Treatment of Missing Data in Clinical Trials. N. Engl. J. Med. 2012;367:1355–60.

314. Little RJ, Cohen ML, Dickersin K, Emerson SS, Farrar JT, Neaton JD, et al. The design and conduct of clinical trials to limit missing data. Stat. Med. 2012;31:3433–43.

315. Lourens C, Lindegardh N, Barnes KI, Guerin PJ, Sibley CH, White NJ, et al. Benefits of a pharmacology antimalarial reference standard and proficiency testing program provided by the Worldwide Antimalarial Resistance Network (WWARN). Antimicrob. Agents Chemother. 2014;58:3889–94.

316. Mustillo S. The effects of auxiliary variables on coefficient bias and efficiency in multiple imputation. Sociol. Methods Res. 2012;41:335–61.

317. 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;68:133–9.

318. Rouse P, Mkulama MAP, Thuma PE, Mharakurwa S. Distinction of Plasmodium falciparum recrudescence and re-infection by MSP2 genotyping: A caution about unstandardized classification criteria. Malar. J. 2008;7:185.

319. Plucinski MM, Dimbu PR, Macaia AP, Ferreira CM, Samutondo C, Quivinja J, et al. Efficacy of artemether-lumefantrine, artesunate-amodiaquine, and dihydroartemisinin-piperaquine for treatment of uncomplicated Plasmodium falciparum malaria in Angola, 2015. Malar. J. BioMed Central; 2017;16:1–10.

320. Zurovac D, Talisuna AO, Snow RW. Mobile phone text messaging: Tool for malaria control in Africa. PLoS Med. 2012;9:1–6.

321. Talisuna AO, Oburu A, Githinji S, Malinga J, Amboko B, Bejon P, et al. Efficacy of text-message reminders on paediatric malaria treatment adherence and their post-treatment return to health facilities in Kenya: A randomized controlled trial ISRCTN39512726 ISRCTN. Malar. J. BioMed Central; 2017;16:1–11.

322. Verduijn M, Grootendorst DC, Dekker FW, Jager KJ, Le Cessie S. The analysis of competing events like cause-specific mortality - beware of the Kaplan-Meier method. Nephrol. Dial. Transplant. 2011;26:56–61.

323. Gatton ML, Cheng Q. Can estimates of antimalarial efficacy from field studies be improved? Trends Parasitol. 2008;24:68–73.

324. Messerli C, Hofmann NE, Beck H-P, Felger I. Critical evaluation of molecular monitoring in malaria drug efficacy trials: pitfalls of length polymorphic markers. Antimicrob. Agents Chemother. 2017;61:e01500-16.

325. Malaria Policy Advisory Committee Meeting. Consensus modelling evidence to support the design of mass drug administration programmes. World Heal. Organ. 2015;1–20.

326. Dini S, Zaloumis S, Cao P, Price RN, Fowkes FJI, Mccaw JM, et al. Investigating the efficacy of triple artemisinin-based combination therapies (TACTs) for treating Plasmodium falciparum malaria patients using mathematical modelling [PrePrint]. bioRxiv. 2018;

327. Latouche A, Porcher R. Sample size calculations in the presence of competing risks. Stat. Med. 2007;26:5370–80.

328. Food and Drug Administration. Multiple endpoints in clinical trials: Guidance for Industry [Internet]. FDA Guid. 2017. p. 1–57. Available from: https://www.fda.gov/downloads/drugs/guidancecomplianceregulatoryinformation/guidances/ucm536750.pdf

329. WorldWide Antimalarial Resistance Network. Malaria in Pregnancy Treatment Efficacy Study Group [Internet]. www.wwarn.org. 2018 [cited 2018 Feb 24]. Available from: http://www.wwarn.org/working-together/study-groups/malaria-pregnancy-treatment-efficacy-study-group

330. Julé AM, Vaillant M, Lang TA, Guérin PJ, Olliaro PL. The Schistosomiasis Clinical Trials Landscape: A Systematic Review of Antischistosomal Treatment Efficacy Studies and a Case for Sharing Individual Participant-Level Data (IPD). PLoS Negl. Trop. Dis. 2016;10:1–17.

331. Douglas NM, Nosten F, Ashley E a., Phaiphun L, Van Vugt M, Singhasivanon P, et al. Plasmodium vivax recurrence following falciparum and mixed species malaria: Risk factors and effect of antimalarial kinetics. Clin. Infect. Dis. 2011;52:612–20.

332. Looareesuwan S, White NJ, Chittamas S, Bunnag D, Harinasuta T. High rate of Plasmodium vivax relapse following treatment of falciparum malaria in Thailand. Lancet. 1987;2:1052–5.

333. Beyersmann J, Latouche A, Buchholz A, Schumacher M. Simulating competing risks data in survival analysis. Stat. Med. 2009;28:956–71.

Appendices

Appendix to Chapter 2

Appendix 2.1: Expression for computing variance of δ

Let $\hat{S}1(t)$ and $\hat{S}2(t)$ be the two survival estimates for drug A and drug B respectively. Let $\delta = \{\ln\big(-ln\big(\hat{S}1(t)\big)\big) - \ln\big(-ln\big(\hat{S}2(t)\big)\big)$ be the difference between these two estimates on complementary log-log scale. The interest is in estimating the variance of δ.

The Taylor's series approximations to the variance of a function g(X) of a random variable is given by the following expression (outlined in equation 2.8, page 26 of Collet (2015)[102]):

$$Var\{g(X)\} \approx \left\{\frac{d\,g(X)}{d\,X}\right\}^2 Var(X) \quad (1)$$

Using (1), we obtain the variance of $\log(-\log(S(t)))$ is equal to

$$Var\{ln(-ln(S(t)))\} \approx \left\{\frac{1}{ln(S(t))}\right\}^2 Var(ln(S(t))) \quad (2)$$

Using the following properties of variance for two independent random variables

$$Var\,(X+Y) = Var(X) + Var(Y)$$
$$Var(aX+b) = a^2.Var(X)$$

We get that,

$$Var(\delta) = Var\big[ln\big(-ln\big(\hat{S}1(t)\big)\big) - \log\big(-ln\big(\hat{S}2(t)\big)\big)\big]$$
$$= Var\big[ln\big(-ln\big(\hat{S}1(t)\big)\big)\big] + Var\,\big[ln\big(-ln\big(\hat{S}2(t)\big)\big)\big]$$

From equation (2), we obtain:

$$Var(\delta) = \left\{\frac{1}{ln(\hat{S}1(t))}\right\}^2 \frac{1}{\hat{S}1(t)^2}.Var\big(\hat{S}1(t)\big) + \left\{\frac{1}{ln(\hat{S}2(t))}\right\}^2 \frac{1}{\hat{S}2(t)^2}.Var(\hat{S}2(t)) \quad (3)$$

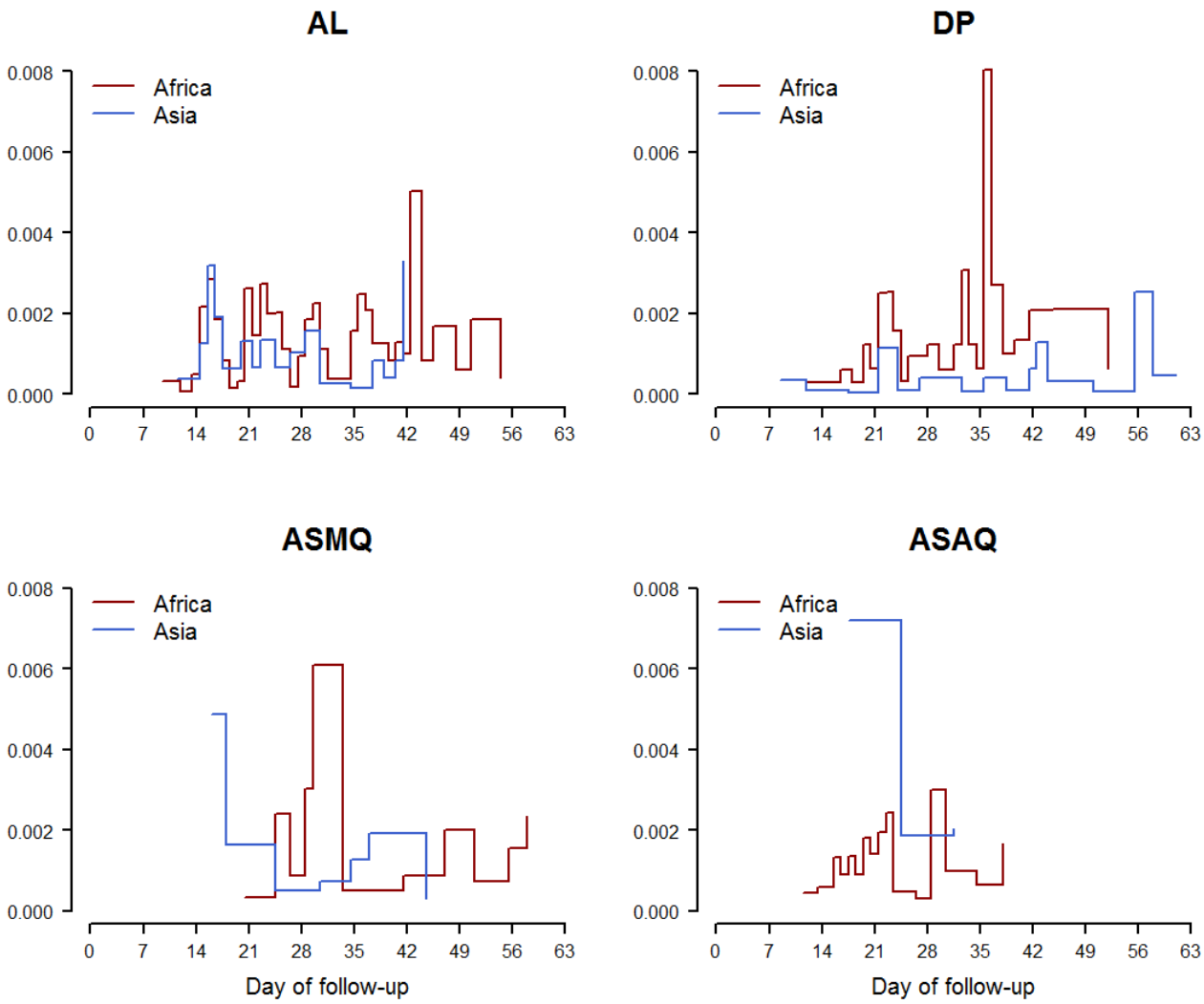
Appendix 2.2: Example data

Drug A		Drug B	
time	censor	time	censor
3	1	14	0
3	1	14	0
3	1	21	0
3	1	21	0
3	0	21	0
7	1	21	0
7	1	21	0
14	0	21	0
14	0	21	0
14	0	21	0
14	1	21	1
14	1	28	1
14	0	28	0
14	1	28	1
14	1	28	0
21	1	28	0
21	1	28	0
21	1	28	1
21	1	28	1
21	1	28	1
21	1	28	0
21	0	35	0
28	1	35	0
28	0	35	0
28	0	35	0
35	0	35	0
35	0	35	0
35	0	35	0
35	0	35	1
35	0	35	1
42	0	35	1
42	0	35	1
42	0	42	0
42	0	42	0
42	0	42	0
42	0	42	0
42	0	42	0
42	0	42	0
42	1	42	0
42	0	42	1

0 = censored observation; 1= recrudescence failures

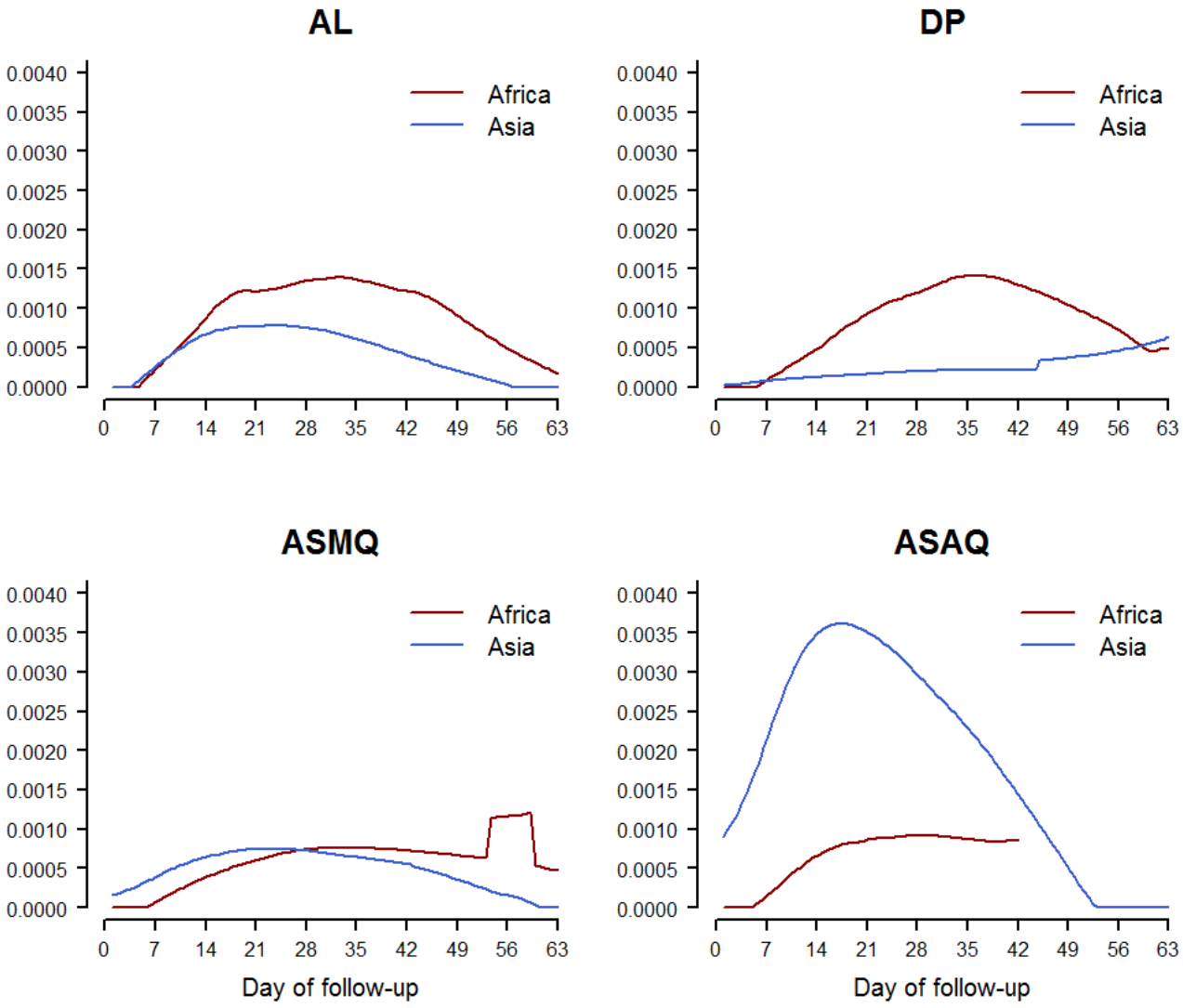
Appendix to Chapter 4

4.1 Kaplan-Meier type estimate of hazard function of *P. falciparum* recrudescence



Legend: The hazard function was estimated using **kphaz.fit** function in **muhaz** package in R software. AL = artemether-lumefantrine; DP = dihydroartemisinin-piperaquine; ASMQ = artesunate-mefloquine; ASAQ = artesunate-amodiaquine, and n = number of recrudescences. Only data from studies with at least three molecular markers are shown in the graph.

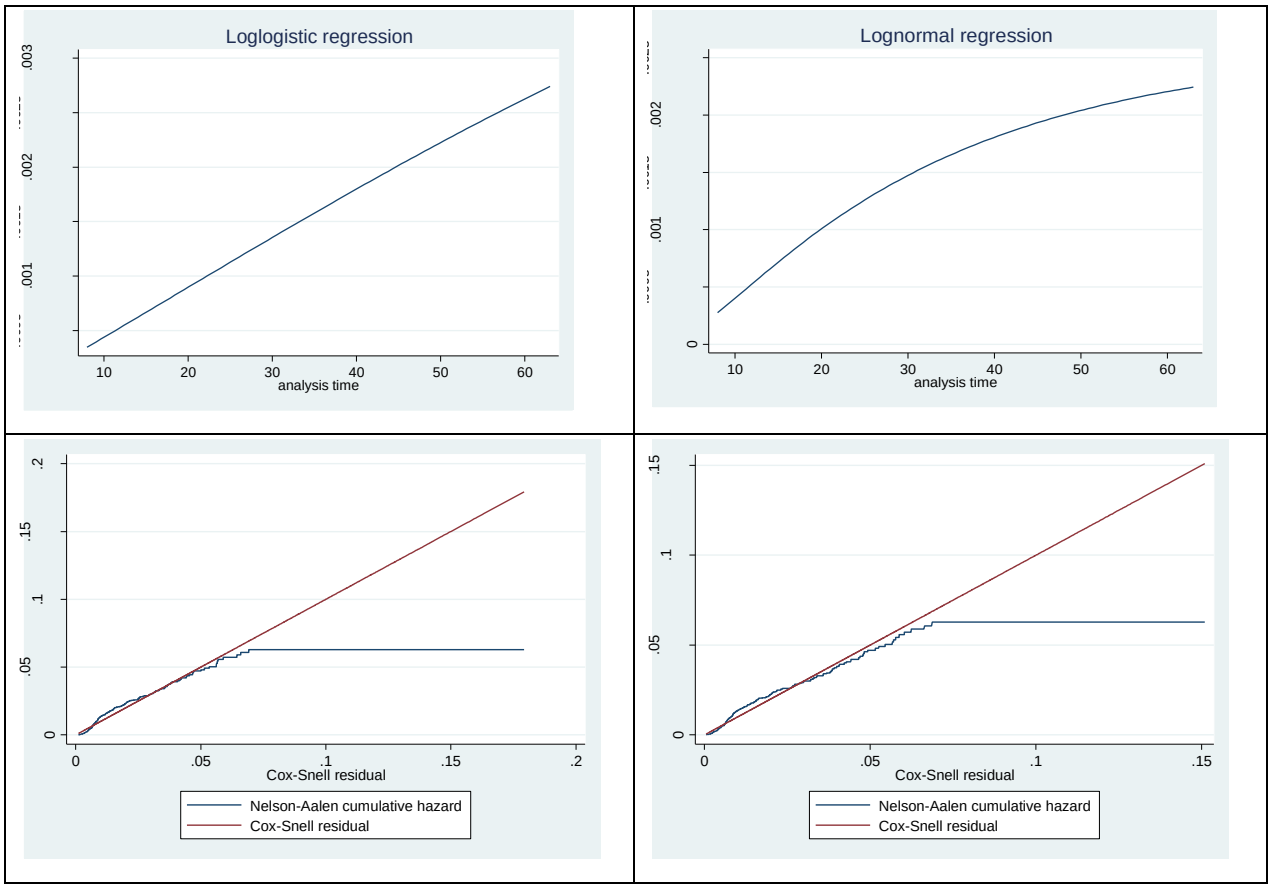
4.2 Kernel-smoothed estimate of hazard function of *P. falciparum* recrudescence



Legend: The hazard function was estimated using muhaz function in **muhaz** package in R software using epacnechnikov kernel. AL = artemether-lumefantrine; DP = dihydroartemisinin-piperaquine; ASMQ = artesunate-mefloquine; ASAQ = artesunate-amodiaquine, and n = number of recrudescences. Only data from studies with at least three molecular markers are shown in the graph.

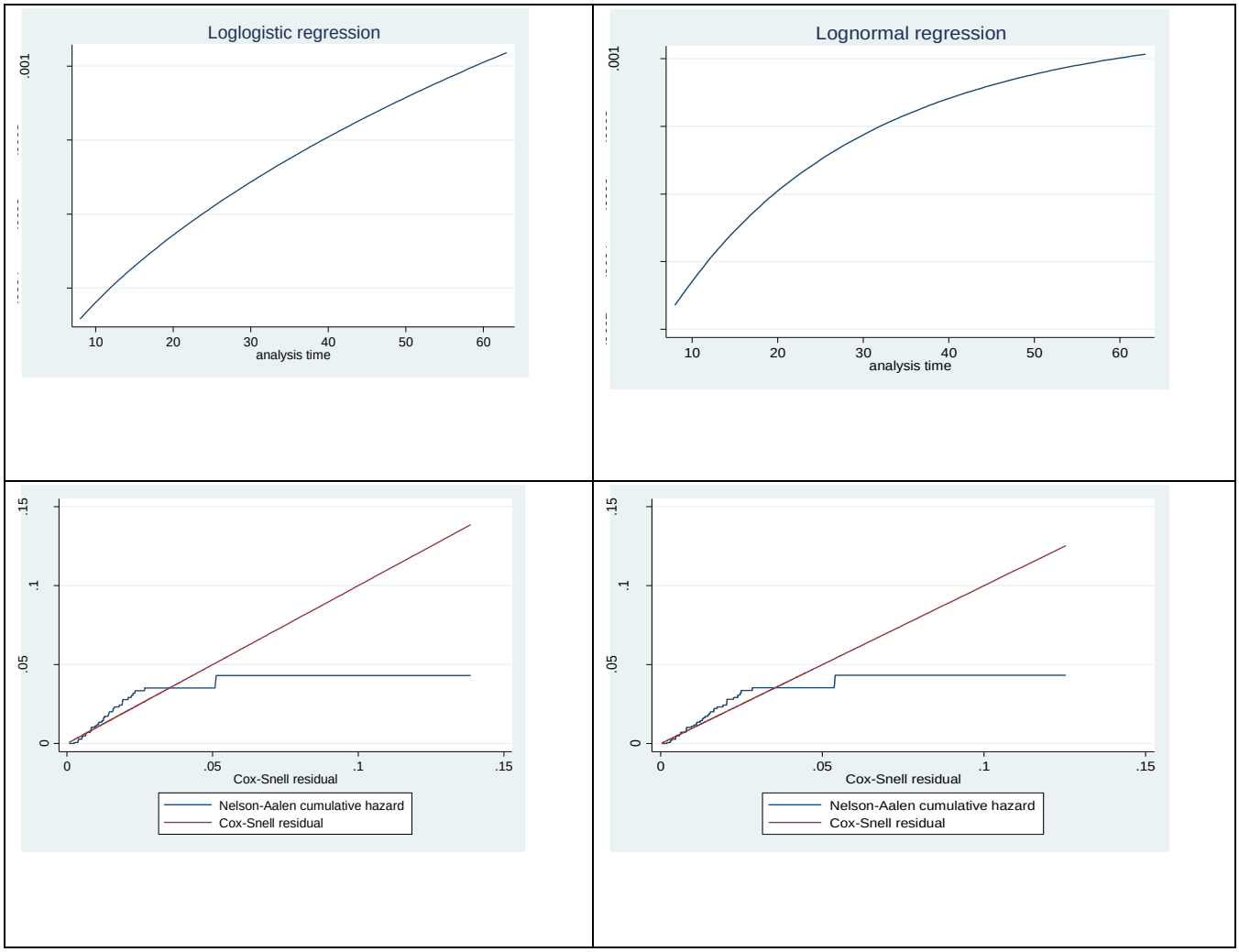
4.3. Parametric survival analysis approach for estimating hazard function of *P. falciparum* recrudescence

4.3.1 Estimates of hazard function of *P. falciparum* recrudescence in children <5 years treated with artemether-lumefantrine (AL) in Africa



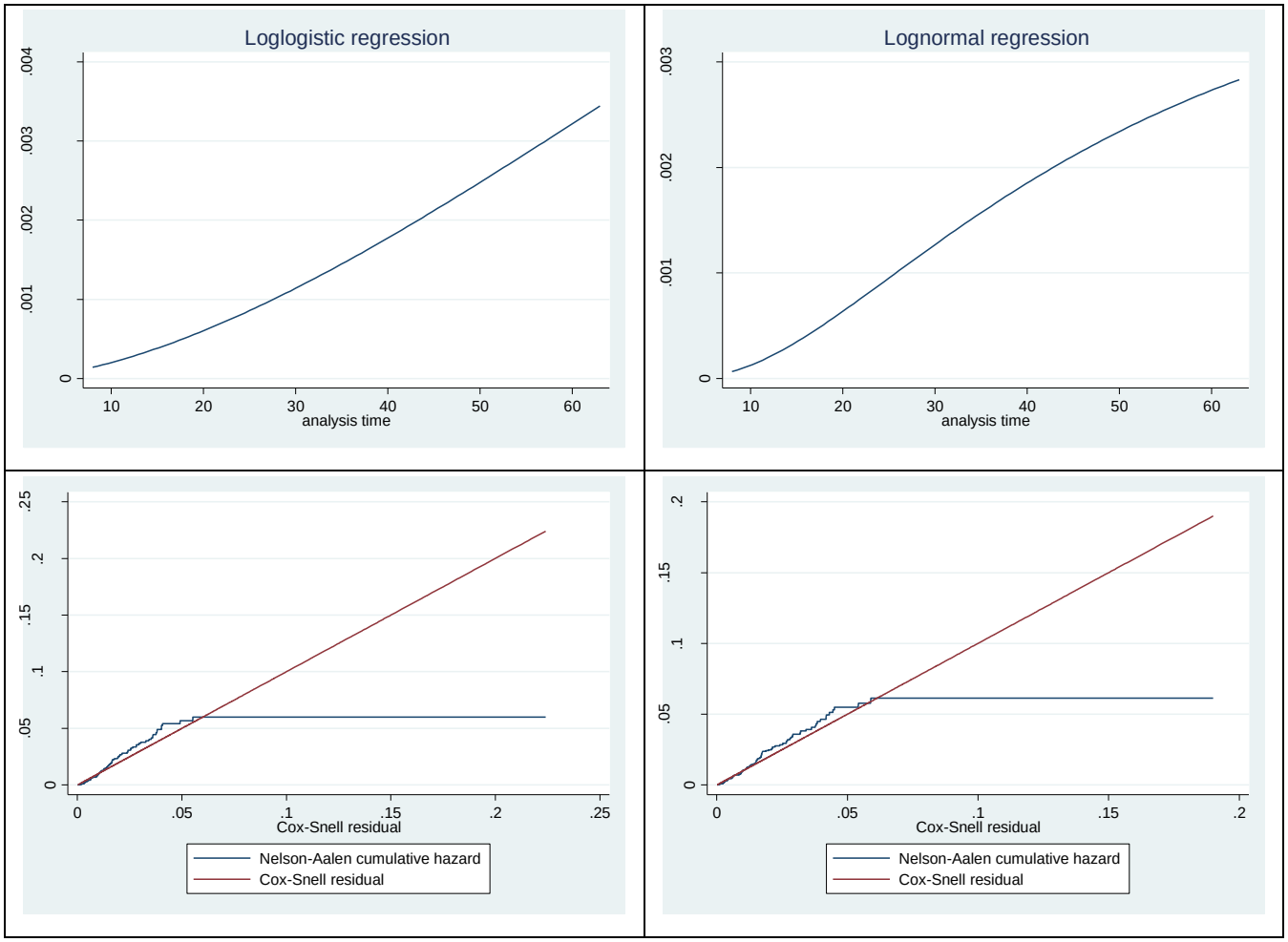
Legend: **Top panel**: The hazard function estimated using log-logistic regression and log-normal regression using **stata streg** command. These models adjusted for age, baseline parasitaemia and mg/kg drug dose of partner component. **Bottom panel**: Assessment of goodness of fit of the parametric models using Cox-Snell's residuals. The residuals should fall along the 45 degrees line for a good-fitting model.

4.3.2 Estimates of hazard function of *P. falciparum* recrudescence in patient of all ages treated with artemether-lumefantrine (AL) in Asia



Legend: **Top panel**: The hazard function estimated using log-logistic regression and log-normal regression using **stata streg** command. These models adjusted for age, baseline parasitaemia and mg/kg drug dose of partner component. **Bottom panel**: Assessment of goodness of fit of the parametric models using Cox-Snell's residuals. The residuals should fall along the 45 degrees line for a good-fitting model.

4.3.3 Estimates of hazard function of *P. falciparum* recrudescence in children <5 years treated with dihydroartemisinin-piperaquine (DP) in Africa



Legend: **Top panel**: The hazard function estimated using log-logistic regression and log-normal regression using **stata streg** command. These models adjusted for age, baseline parasitaemia and mg/kg drug dose of partner component. **Bottom panel**: Assessment of goodness of fit of the parametric models using Cox-Snell's residuals. The residuals should fall along the 45 degrees line for a good-fitting model.

4.4 Flexible parametric (Royston-Parmar model) approach for estimating the hazard of *P. falciparum* recrudescence in patients treated with artemether-lumefantrine and dihydroartemisinin-piperaquine

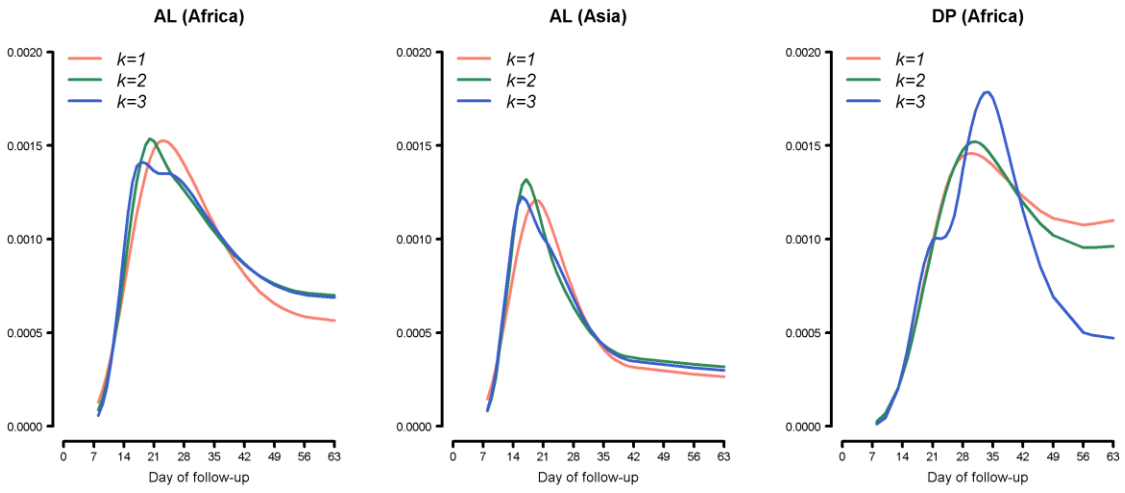
In Africa, the analysis was carried out in children < 5 years for artemether-lumefantrine and dihydroartemisinin-piperaquine. For AL in Asia, patient of all ages were included. The AIC estimates for Royston-Parmar models with different number of knots are shown below:

Table 4A.1: Akaika's Information Criteria (AIC) for Royston-Parmar model with different number of knots

Knots	AL Africa†	AL Asia	DP Africa
k=0	2759.1	647.2	1319.8
k=1	2707.9	630.6	1305.9
k=2	2707.7	631.4	1307.5
k=3	2707.4	633.6	1306.6

†For AL Africa, *k*=1 chosen as the model of choice for parsimonious reasons. The Royston-Parmar model for AL Africa was adjusted for age, baseline parasitaemia and mg/kg drug dosage with study sites fitted as shared frailty term. For AL Asia, the model adjusted for age and baseline parasitaemia. For DP Africa, the model adjusted for age, parasitaemia and mg/kg dose.

The hazard functions corresponding to knots = 1,2, and 3 presented in Table 4A.1 are plotted below:



AL = artemether-lumefantrine; DP = dihydroartemisinin-piperaquine

4.5 Sensitivity analyses: The estimated area under the curve (AUC) of constructed distribution for *P. falciparum* recrudescence

In Africa, the analysis was carried out in children < 5 years for artemether-lumefantrine and dihydroartemisinin-piperaquine. For AL in Asia, patient of all ages were included.

Table 4A.2: The AUC of the probability distribution from day 42 studies

	Day 42 studies			
Day	AL Africa	AL Asia	DP (UD)	DP (nUD)
28	0.55	0.70	0.37	0.35
35	0.80	0.90	0.67	0.59
42	0.95	0.99	0.90	0.80
49	1.00	1.00	0.99	0.94
56	1.00	1.00	1.00	1.00
63	1.00	1.00	1.00	1.00

AL = artemether-lumefantrine; DP = dihydroartemisinin-piperaquine; UD = Under-dosed (piperaquine dose ≤48 mg/kg); nUD= Not Under-dosed (piperaquine dose > 48 mg/kg)

Table 4A.3: The AUC of the probability distribution using different definition of observed time of P. falciparum recrudescence

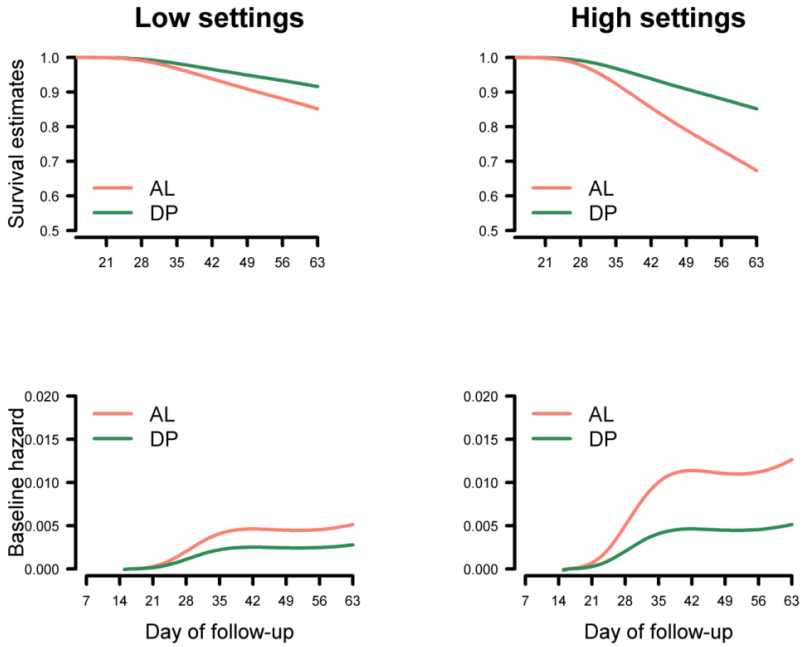
	Assumption 1[a]			Assumption 2[b]		
Day	AL Africa	AL Asia	DP (UD)	AL Africa	AL Asia	DP (UD)
28	0.52	0.72	0.44	0.40	0.72	0.41
35	0.74	0.91	0.71	0.63	0.92	0.72
42	0.89	0.99	0.91	0.81	1.00	0.93
49	0.97	1.00	0.99	0.93	1.00	1.00
56	1.00	1.00	1.00	0.99	1.00	1.00
63	1.00	1.00	1.00	1.00	1.00	1.00

AL = artemether-lumefantrine; DP = dihydroartemisinin-piperaquine; UD = Under-dosed (piperaquine dose <= 48 mg/kg); nUD= Not Under-dosed (piperaquine dose > 48 mg/kg)

[a]Assuming recrudescence occurred the day after the last visit where patient recorded negative parasitaemia on blood smear; [b] Assuming recrudescence occurred half-way through the last visit with recorded negative smear and the day when it recorded in the study.

Appendix 4.6: Simulated hazard ratio for artemether-lumefantrine relative to dihydroartemisinin-piperaquine

The description of simulation study is provided in Section 4.4.1. The survival and hazard function used for the simulation study is given below. The hazard ratio derived from Cox regression model is presented in the table below.



AL = artemether-lumefantrine; DP = dihydroartemisinin-piperaquine

Table 4A.4: Simulated hazard ratio for AL relative to DP at different follow-up time

	Low Settings			**High Settings**		
Day	n=200 subjects	n=500 subjects	n=1,000 subjects	n=200 subjects	n=500 subjects	n=1,000 subjects
28	2.01 [1.00 - 4.04]	2.00 [1.29 - 2.78]	1.89 [1.46 - 2.51]	2.01 [1.01 - 4.04]	1.85 [1.34 - 3.02]	1.90 [1.45 - 2.53]
35	1.62 [1.01 - 2.69]	1.57 [1.19 - 2.03]	1.57 [1.27 - 1.91]	1.67 [1.01 - 2.74]	1.59 [1.18 - 2.21]	1.58 [1.28 - 1.94]
42	1.35 [0.89 - 2.04]	1.30 [1.07 - 1.68]	1.33 [1.12 - 1.59]	1.39 [0.91 - 2.06]	1.36 [1.05 - 1.76]	1.35 [1.11 - 1.63]
49	1.16 [0.83 - 1.67]	1.15 [0.95 - 1.43]	1.16 [1.01 - 1.38]	1.22 [0.85 - 1.72]	1.17 [0.96 - 1.48]	1.19 [1.00 - 1.39]
56	1.03 [0.76 - 1.44]	1.05 [0.87 - 1.29]	1.05 [0.92 - 1.23]	1.08 [0.77 - 1.5]	1.07 [0.89 - 1.30]	1.08 [0.93 - 1.24]
63	0.94 [0.71 - 1.32]	0.98 [0.81 - 1.19]	0.97 [0.84 - 1.12]	1.01 [0.73 - 1.39]	0.99 [0.82 - 1.21]	1.00 [0.87 - 1.15]

AL = artemether-lumefantrine; DP = dihydroartemisinin-piperaquine

Low settings = approximately 8% new infections with AL and 15% new infections in DP arm

High settings = approximately 15% new infections with AL and 30% new infections in DP arm

Appendix to Chapter 5

Appendix 5.1: Overestimation of the risk of P. falciparum new infections at the end of the study follow-up by Kaplan-Meier approach

	n	Absolute overestimation [a]	Observed proportion of recrudescence [b]	Observed proportion of new infection [b]
Day 28				
AL	38	0.13% [IQR :0.02 - 0.36 ; Range: 0.00 - 2.93]	2.4% [0.1 - 8.8]	9.1% [0.6 - 48.3]
ASAQ	10	0.18% [IQR :0.07 - 0.34 ; Range: 0.00 - 0.84]	2.5% [0.3 - 5.6]	10.2% [1.3 - 27.4]
DP	9	0.03% [IQR :0.00 - 0.08 ; Range: 0.00 - 0.42]	1.4% [0.6 – 10.0]	2.7% [1.8 - 19.5]
Day 42				
AL	34	0.78% [IQR :0.28 - 1.52 ; Range: 0.05 - 10.68]	3.8% [0.7 – 18.0]	20.4% [1.4 - 62.4]
ASAQ	7	0.39% [IQR :0.14 - 1.11 ; Range: 0.04 - 2.91]	2.3% [1.0 – 17.0]	21.7% [2.3 - 62.8]
DP	11	0.38% [IQR :0.08 - 0.79 ; Range: 0.01 - 1.2]	2.9% [1.1 - 16.4]	9.2% [1.1 - 31.6]
Day 63				
AL	9	2.06% [IQR :0.83 - 2.80 ; Range: 0.59 - 4.32]	3.7% [1.4 - 10.6]	24% [7.6 - 57.8]
ASAQ	1	2.82% [IQR :2.82 - 2.82 ; Range: 2.82 - 2.82]	12.3% [12.3 - 12.3]	11.1% [11.1 - 11.1]
ASMQ	6	2.39% [IQR :2.19 - 3.93 ; Range: 0.55 - 6.17]	4.1% [1.6 - 12.7]	30.8% [14.1 - 47]
DP	10	0.53% [IQR :0.37 - 1.13 ; Range: 0.02 - 4.31]	1.6% [0.7 - 7.1]	7.4% [1.6 - 29]

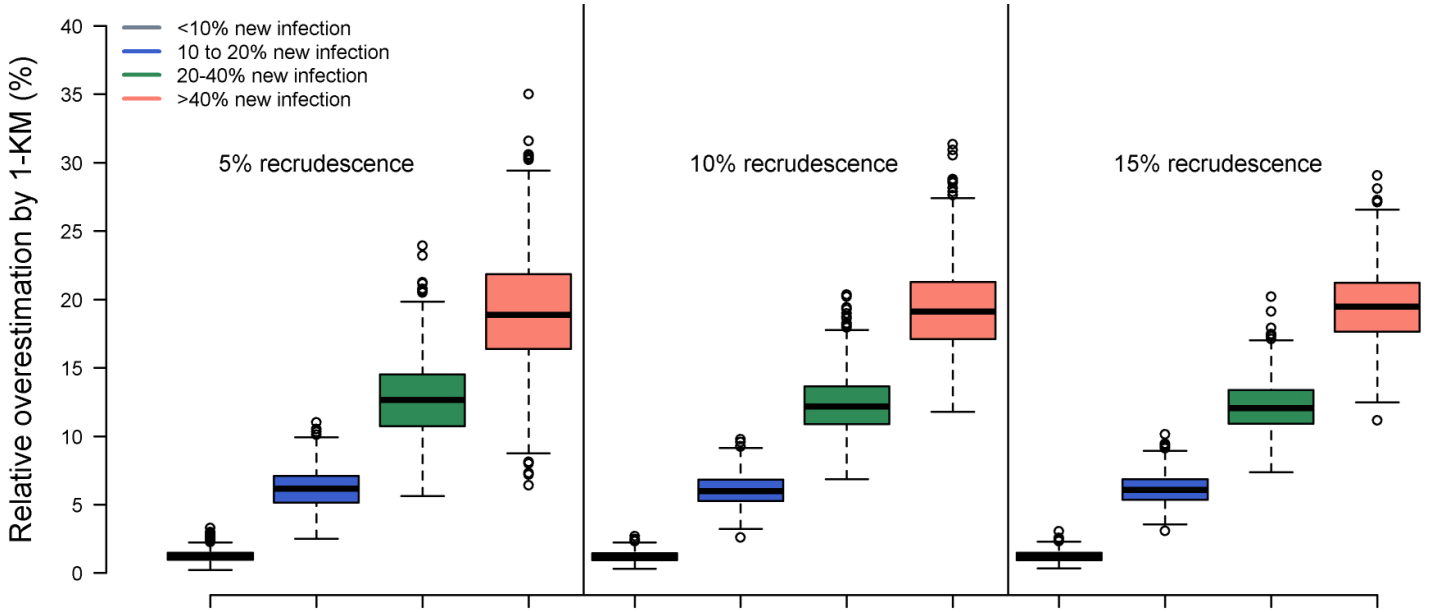
[a]Values are median [Interquartile Range (IQR); Range]
[b]Values are median [Range]
n=number of study sites with study sample size ≥25; IQR=Interquartile range
AL= artemether-lumefantrine, ASAQ= artesunate-amodaiquine; DP= dihydroartemisinin-piperaquine; ASMQ= artesunate-mefloquine.

Appendix 5.2: Sample size required per group to detect a difference in given hazards ratio for simulation study II

					80% power		90% power	
Scenario	$S_A(t)$	$S_B(t)$	$1-\bar{S}(t)$	HR to detect	Events/arm required	SS/arm required	Events/arm required	SS/arm required
1A	0.94	0.94	0.06	1	-	-	-	-
1B	0.94	0.94	0.06	1	-	-	-	-
1C	0.94	0.94	0.06	1	-	-	-	-
2A	0.94	0.84	0.11	2.72	16	146	21	195
2B	0.94	0.98	0.04	0.37	16	384	21	514
3A	0.94	0.84	0.11	2.72	16	146	21	195
3B	0.94	0.84	0.11	2.72	16	146	21	195
3C	0.94	0.84	0.11	0.37	16	147	21	197
3D	0.94	0.98	0.04	0.37	16	384	21	514

$S_A(t)$= Survival estimate at the end of study follow-up for drug A
$S_B(t)$= Survival estimate at the end of study follow-up for drug B
$\bar{S}(t)$ = average survival between two groups

Appendix 5.3: Relative overestimation of recrudescence failure using K-M approach compared to the CIF in a simulation study assessing standalone efficacy on day 63 (n=500 subjects)



The overestimation $(\hat{F}_{KM}(t) - \hat{F}_{CIF}(t))$ of cumulative recrudescence using K-M approach. Each panel represents different underlying status of drug efficacy on average (~5%, 10% and 15% recrudescence observed) in a study with a sample size of 500. The results are presented from 1,000 independent simulation runs. The variation in absolute overestimation within each boxplot is due to varying proportion of new infections. Within each panel, the colours indicate different proportion of new infections: <10% new infections (grey), 10-20% new infections (blue), 20-40% new infections (green) and >40% new infections (orange), representing areas of progressively increasing malaria transmission.

Appendix 5.4: Results for simulation study I with different sample sizes

Table A5.1: Absolute overestimation in recrudescence failure by K-M method compared to CIF in simulation study I (n=100 subjects)

		Median absolute overestimation[IQR; Range]		
5% recrudescence	**Observed proportion of new infections**	**Day28**	**Day 42**	**Day 63**
<10% NI	3% [0 - 9]	0.00% [0.00 - 0.00 ; Range:0.00-0.04]	0.01% [0.00 - 0.03 ; Range:0.00-0.15]	0.05% [0.02 - 0.09 ; Range:0.00-0.34]
10-20% NI	16% [10 - 19]	0.00% [0.00 - 0.00 ; Range:0.00-0.08]	0.08% [0.04 - 0.12 ; Range:0.00-0.39]	0.30% [0.21 - 0.41 ; Range:0.00-0.79]
20-40% NI	31% [20 - 39]	0.00% [0.00 - 0.01 ; Range:0.00-0.12]	0.18% [0.11 - 0.27 ; Range:0.00-0.89]	0.66% [0.45 - 0.9 ; Range:0.01-2.05]
40+% NI	44% [40 - 59]	0.01% [0.00 - 0.02 ; Range:0.00-0.29]	0.31% [0.17 - 0.45 ; Range:0.00-1.09]	1.08% [0.71 - 1.49 ; Range:0.00-3.3]
10% recrudescence				
<10% NI	4% [0 - 9]	0.00% [0.00 - 0.00; Range:0.00-0.05]	0.02% [0.00 - 0.05 ; Range:0.00-0.29]	0.11% [0.06 - 0.18 ; Range:0.00-0.53]
10-20% NI	16% [10 - 19]	0.00% [0.00 - 0.01; Range:0.00-0.12]	0.15% [0.09 - 0.22 ; Range:0.00-0.80]	0.55% [0.41 - 0.72 ; Range:0.00-1.52]
20-40% NI	30% [20 - 39]	0.01% [0.00 - 0.02 ; Range:0.00-0.15]	0.32% [0.20 - 0.47 ; Range:0.00-1.51]	1.19% [0.90 - 1.53 ; Range:0.12-3.33]
40+% NI	43% [40 - 56]	0.02% [0.00 - 0.04 ; Range:0.00-0.23]	0.58% [0.40 - 0.78 ; Range:0.00-2.58]	1.97% [1.48 - 2.45 ; Range:0.35-5.58]
15% recrudescence				
<10% NI	3% [0 - 9]	0.00% [0.00 - 0.00 ; Range:0.00-0.09]	0.03% [0.00 - 0.08 ; Range:0.00-0.40]	0.16% [0.07 - 0.25 ; Range:0.00-0.66]
10-20% NI	15% [10 - 19]	0.00% [0.00 - 0.02 ; Range:0.00-0.15]	0.22% [0.14 - 0.33 ; Range:0.00-1.04]	0.82% [0.64 - 1.06 ; Range:0.11-1.93]
20-40% NI	29% [20 - 39]	0.02% [0.00 - 0.04 ; Range:0.00-0.25]	0.52% [0.36 - 0.70 ; Range:0.00-1.65]	1.73% [1.37 - 2.21 ; Range:0.38-4.56]
40+% NI	43% [40 - 57]	0.03% [0.01 - 0.07 ; Range:0.00-0.32]	0.84% [0.61 - 1.15 ; Range:0.04-2.66]	2.90% [2.27 - 3.62 ; Range:0.38-6.19]

NI= New infections

Table A5.2: Absolute overestimation in recrudescence failure by K-M method compared to CIF in simulation study I (n=200 subjects)

		Median absolute overestimation[IQR; Range]		
5% recrudescence	**Observed proportion of new infections**	**Day28**	**Day 42**	**Day 63**
<10% NI	4% [0 - 8]	0.00% [0.00 - 0.00; Range:0.00-0.02]	0.01% [0.01 - 0.03 ;Range:0.00-0.09]	0.06% [0.04 - 0.09 ; Range:0.00-0.23]
10-20% NI	16.5% [10 - 19.5]	0.00% [0.00 - 0.01 ; Range:0.00-0.05]	0.08% [0.05 - 0.10 ; Range:0.00-0.29]	0.29% [0.22 - 0.36 ; Range:0.00-0.74]
20-40% NI	31.5% [22 - 39]	0.01% [0.00 - 0.01 ; Range:0.00-0.09]	0.18% [0.12 - 0.24 ; Range:0.01-0.83]	0.62% [0.47 - 0.78 ; Range:0.09-1.37]
40+% NI	43.5% [40 - 54.5]	0.01% [0.00 - 0.02 ; Range:0.00-0.09]	0.28% [0.20 - 0.38 ; Range:0.00-0.8]	0.98% [0.73 - 1.21 ; Range:0.16-2.79]
10% recrudescence				
<10% NI	3.5% [0.5 - 8.5]	0.00% [0.00 - 0.00 ; Range:0.00-0.04]	0.03% [0.01 - 0.05 ; Range:0.00-0.23]	0.11% [0.07 - 0.16 ; Range:0.00-0.40]
10-20% NI	16% [10.5 - 19.5]	0.01% [0.00 - 0.01 ; Range:0.00-0.08]	0.16% [0.12 - 0.22 ; Range:0.00-0.47]	0.58% [0.47 - 0.69 ; Range:0.09-1.15]
20-40% NI	30% [21 - 39.5]	0.01% [0.01 - 0.03 ; Range:0.00-0.09]	0.36% [0.26 - 0.45 ; Range:0.04-0.92]	1.24% [1.01 - 1.49 ; Range:0.34-2.8]
40+% NI	42.5% [40 - 54]	0.02% [0.01 - 0.04 ; Range:0.00-0.19]	0.55% [0.44 - 0.70 ; Range:0.12-1.34]	1.91% [1.57 - 2.22 ; Range:0.63-3.79]
15% recrudescence				
<10% NI	3.5% [0 - 8]	0.00% [0.00 - 0.00 ; Range:0.00-0.05]	0.04% [0.02 - 0.07 ; Range:0.00-0.25]	0.17% [0.12 - 0.24 ; Range:0.00-0.63]
10-20% NI	16% [10 - 19.5]	0.01% [0.00 - 0.02 ; Range:0.00-0.09]	0.25% [0.18 - 0.33 ; Range:0.03-0.66]	0.89% [0.73 - 1.06 ; Range:0.31-1.85]
20-40% NI	28.5% [20 - 39]	0.02% [0.01 - 0.04 ; Range:0.00-0.19]	0.53% [0.41 - 0.65 ; Range:0.11-1.34]	1.80% [1.51 - 2.11 ; Range:0.70-3.35]
40+% NI	42% [40 - 49.5]	0.04% [0.02 - 0.06 ; Range:0.00-0.19]	0.87% [0.69 - 1.06 ; Range:0.19-2.34]	2.91% [2.47 - 3.36 ; Range:1.21-5.59]

NI= New infections

Table A5.3: Absolute overestimation in recrudescence failure by K-M method compared to CIF in simulation study I (n=1,000 subjects)

		Median absolute overestimation[IQR; Range]		
5% recrudescence	**Observed proportion of new infections**	**Day28**	**Day 42**	**Day 63**
<10% NI	3.7% [2 - 5.7]	0.00% [0.00 - 0.00 ; Range:0.00-0.01]	0.02% [0.01 - 0.02 ; Range:0.00-0.06]	0.06% [0.05 - 0.08 ; Range:0.03-0.13]
10-20% NI	17.2% [13.5 - 19.9]	0.00% [0.00 - 0.01 ; Range:0.00-0.02]	0.09% [0.07 - 0.10 ; Range:0.02-0.15]	0.32% [0.28 - 0.36 ; Range:0.16-0.55]
20-40% NI	31.2% [26.4 - 36.7]	0.01% [0.01 - 0.01 ; Range:0.00-0.03]	0.18% [0.15 - 0.21 ; Range:0.07-0.33]	0.63% [0.57 - 0.70 ; Range:0.34-0.95]
40+% NI	42.7% [38.1 - 47.2]	0.01% [0.01 - 0.02 ; Range:0.00-0.04]	0.27% [0.24 - 0.32 ; Range:0.11-0.54]	0.93% [0.83 - 1.04 ; Range:0.57-1.47]
10% recrudescence				
<10% NI	3.6% [1.8 - 5.6]	0.00% [0.00 - 0.00 ; Range:0.00-0.01]	0.04% [0.03 - 0.04 ; Range:0.01-0.08]	0.13% [0.11 - 0.15 ; Range:0.05-0.25]
10-20% NI	16.4% [13.1 - 19.9]	0.01% [0.01 - 0.01 ; Range:0.00-0.03]	0.18% [0.16 - 0.20 ; Range:0.08-0.29]	0.63% [0.57 - 0.69 ; Range:0.34-1.01]
20-40% NI	30% [25.1 - 34.5]	0.02% [0.01 - 0.02 ; Range:0.00-0.04]	0.37% [0.32 - 0.41 ; Range:0.22-0.59]	1.25% [1.15 - 1.36 ; Range:0.76-1.87]
40+% NI	41.6% [40 - 46.3]	0.03% [0.02 - 0.03 ; Range:0.01-0.06]	0.57% [0.51 - 0.63 ; Range:0.35-0.90]	1.87% [1.71 - 2.02 ; Range:1.22-2.57]
15% recrudescence				
<10% NI	3.5% [1.8 - 6.3]	0.00% [0.00 - 0.00 ; Range:0.00-0.02]	0.05% [0.04 - 0.06 ; Range:0.01-0.13]	0.19% [0.16 - 0.22 ; Range:0.09-0.43]
10-20% NI	15.8% [12.9 - 19.4]	0.01% [0.01 - 0.02 ; Range:0.00-0.04]	0.27% [0.24 - 0.30 ; Range:0.14-0.48]	0.92% [0.85 – 1.00 ; Range:0.61-1.34]
20-40% NI	28.6% [24.5 - 33.5]	0.03% [0.02 - 0.03 ; Range:0.01-0.08]	0.54% [0.48 - 0.60 ; Range:0.30-0.81]	1.82% [1.68 - 1.95 ; Range:1.14-2.42]
40+% NI	40.8% [40 - 44.7]	0.04% [0.03 - 0.05 ; Range:0.01-0.09]	0.85% [0.77 - 0.94 ; Range:0.53-1.30]	2.82% [2.63 - 3.02 ; Range:2.01-3.85]

NI= New infections

Appendix 5.5: The assumption of proportional hazard in simulated datasets in Simulation study II (n=500/arm)

Recrudescence	Assumption of proportional hazard [a] Satisfied	Violated	Proportion of times PH assumption violated
Both tests significant	4930	280	5.4%
Only Gray's test significant	313	17	5.2%
Only log-rank test significant	252	9	3.4%
Both test not significant	3015	184	5.8%
Overall	8510	490	5.4%
New Infection			
Both tests significant	4961	249	4.8%
Only Gray's test Significant	315	15	4.5%
Only log-rank test significant	250	11	4.2%
Both test not significant	3067	132	4.1%
Overall	8593	407	4.5%

[a] The assumption of PH tested using `cox.zph` function in R survival package from a cox model with drug regimen as the only factor. The assumption of PH was defined to be violated if *p*-value <0.05 for the test using cox.zph

Appendix to Chapter 6

Appendix 6.1: Performance measures of missing data methods when recurrences were kept as fixed

Table A6.1: Performance measures of different approaches for handling missing outcome for artemether-lumefantrine

	10% Missing						30% missing						45% missing					
	Bias	rBias	ModSE¶	EmpSE	CP	RMSE	Bias	rBias	ModSE	EmpSE	CP	RMSE	Bias	rBias	ModSE	EmpSE	CP	RMSE
M1																		
CC	0.0033	0.35	0.1655	0.0518	0.999	0.1041	0.0103	1.07	0.1880	0.0992	0.537	0.3237	0.0159	1.66	0.2130	0.1354	0.050	0.5439
IPW†	0.0007	0.07	0.1646	0.0524	1.000	0.0962	0.0015	0.15	0.1896	0.1102	0.992	0.2059	0.0016	0.17	0.2287	0.1629	0.967	0.2972
MI	-0.0004	-0.04	0.1636	0.0484	1.000	0.0697	-0.0016	-0.16	0.1801	0.0905	1.000	0.1377	-0.0024	-0.26	0.1970	0.1176	0.995	0.1832
M2a																		
CC	0.0034	0.36	0.1658	0.0541	0.999	0.1082	0.0106	1.11	0.1896	0.1056	0.490	0.3385	0.0163	1.70	0.2154	0.1432	0.044	0.5637
IPW†	0.0007	0.07	0.1650	0.0538	1.000	0.0988	0.0015	0.15	0.1916	0.1144	0.993	0.2125	0.0017	0.18	0.2331	0.1698	0.965	0.3111
MI	-0.0003	-0.04	0.1642	0.0511	1.000	0.0730	-0.0016	-0.16	0.1814	0.0934	1.000	0.1413	-0.0029	-0.30	0.1982	0.1243	0.995	0.1978
M2b																		
CC	0.0035	0.37	0.1661	0.0550	1.000	0.1109	0.0108	1.13	0.1904	0.1083	0.474	0.3454	0.0167	1.74	0.2177	0.1429	0.028	0.5808
IPW†	0.0007	0.07	0.1653	0.0552	1.000	0.1009	0.0015	0.16	0.1929	0.1161	0.991	0.2161	0.0017	0.17	0.2368	0.1766	0.959	0.3219
MI	-0.0004	-0.04	0.1643	0.0517	1.000	0.0742	-0.0014	-0.15	0.1832	0.0964	1.000	0.1433	-0.0029	-0.31	0.2005	0.1281	0.992	0.2030

†The model based standard errors for IPW approaches were estimated using 200 bootstrap samples

¶Full data estimate of model-based standard error: 0.1567 (Table 6.5 of main-text).

M1 = Missingness Mechanism 1; M2a = Missingness Mechanism 2a; M2b = Missingness Mechanism 2b; rBias = Relative bias (%); ModSE = Model-based standard error; EmpSE = Empirical standard error; CP = Coverage Probability; RMSE = Root Mean Squared Error; CC = Complete Case; IPW = Inverse probability weighting model; MI = Multiply Imputed

Table A6.2: Performance measures of different approaches for handling missing outcome for artesunate-amodiaquine

	10% Missing						30% missing						45% missing					
	Bias	rBias	ModSE¶	EmpSE	CP	RMSE	Bias	rBias	ModSE	EmpSE	CP	RMSE	Bias	rBias	ModSE	EmpSE	CP	RMSE
M1																		
CC	0.0019	0.20	0.2499	0.0801	1.000	0.1274	0.0058	0.59	0.2860	0.1608	0.842	0.3708	0.0087	0.89	0.3257	0.2215	0.399	0.6036
IPW†	0.0005	0.05	0.2485	0.0787	1.000	0.1434	0.0011	0.11	0.2834	0.1669	0.988	0.3104	0.0014	0.15	0.3286	0.2398	0.959	0.4486
MI	-0.0007	-0.07	0.2465	0.0754	1.000	0.1149	-0.0021	-0.22	0.2711	0.1322	1.000	0.2252	-0.0037	-0.38	0.2933	0.1758	0.984	0.3265
M2a																		
CC	0.0018	0.19	0.2494	0.0817	0.999	0.1253	0.0056	0.57	0.2841	0.1600	0.860	0.3591	0.0086	0.87	0.3232	0.2207	0.440	0.5900
IPW†	0.0004	0.04	0.2477	0.0767	1.000	0.1386	0.0010	0.11	0.2822	0.1654	0.986	0.3070	0.0014	0.14	0.3257	0.2301	0.967	0.4310
MI	-0.0006	-0.06	0.2467	0.0778	1.000	0.1163	-0.0021	-0.21	0.2707	0.1396	1.000	0.2326	-0.0037	-0.38	0.2912	0.1782	0.986	0.3293
M2b																		
CC	0.0017	0.17	0.2484	0.0777	1.000	0.1174	0.0053	0.54	0.2814	0.1551	0.893	0.3414	0.0083	0.85	0.3197	0.2132	0.468	0.5696
IPW†	0.0004	0.04	0.2472	0.0747	1.000	0.1348	0.0010	0.10	0.2797	0.1617	0.989	0.2989	0.0013	0.14	0.3226	0.2290	0.968	0.4269
MI	-0.0006	-0.06	0.2462	0.0737	1.000	0.1101	-0.0020	-0.21	0.2690	0.1334	0.999	0.2239	-0.0036	-0.37	0.2907	0.1752	0.986	0.3212

†The model based standard errors for IPW approaches were estimated using 200 bootstrap samples

¶Full data estimate of model-based standard error: 0.2358 (Table 6.5 of main-text).

M1 = Missingness Mechanism 1; M2a = Missingness Mechanism 2a; M2b = Missingness Mechanism 2b; rBias = Relative bias (%); ModSE = Model-based standard error; EmpSE = Empirical standard error; CP = Coverage Probability; RMSE = Root Mean Squared Error; CC = Complete Case; IPW = Inverse probability weighting model; MI = Multiply Imputed

Table A6.3: Performance measures of different approaches for handling missing outcome for dihydroartemisinin-piperaquine

	10% Missing						30% missing						45% missing					
	Bias	rBias	ModSE¶	EmpSE	CP	RMSE	Bias	rBias	ModSE	EmpSE	CP	RMSE	Bias	rBias	ModSE	EmpSE	CP	RMSE
M1																		
CC	0.0016	0.16	0.2253	0.0687	1.000	0.1224	0.0047	0.48	0.2571	0.1381	0.669	0.3728	0.0072	0.73	0.2944	0.2073	0.182	0.6297
IPW†	0.0005	0.05	0.2241	0.0764	1.000	0.1440	0.0013	0.13	0.2571	0.1827	0.959	0.3595	0.0018	0.18	0.3061	0.2703	0.878	0.5345
MI	0.0000	0.00	0.2222	0.0651	1.000	0.0921	-0.0003	-0.03	0.2428	0.1248	1.000	0.1769	-0.0006	-0.06	0.2628	0.1693	0.995	0.2416
M2a																		
CC	0.0014	0.14	0.2237	0.0662	1.000	0.1107	0.0042	0.43	0.2515	0.1334	0.793	0.3322	0.0066	0.67	0.2838	0.1984	0.267	0.5619
IPW†	0.0004	0.04	0.2223	0.0701	1.000	0.1295	0.0011	0.11	0.2510	0.1661	0.975	0.3212	0.0015	0.15	0.2945	0.2453	0.901	0.4737
MI	0.0000	0.00	0.2209	0.0634	1.000	0.0896	-0.0002	-0.02	0.2402	0.1154	1.000	0.1632	-0.0005	-0.05	0.2582	0.1557	0.996	0.2215
M2b																		
CC	0.0011	0.12	0.2220	0.0600	1.000	0.0953	0.0037	0.37	0.2455	0.1200	0.880	0.2860	0.0060	0.61	0.2743	0.1803	0.392	0.4969
IPW†	0.0003	0.03	0.2210	0.0631	1.000	0.1157	0.0010	0.10	0.2454	0.1481	0.985	0.2851	0.0013	0.13	0.2836	0.2173	0.926	0.4151
MI	0.0000	0.00	0.2202	0.0584	1.000	0.0826	-0.0001	-0.01	0.2366	0.1058	1.000	0.1496	-0.0003	-0.03	0.2542	0.1430	0.999	0.2023

†The model based standard errors for IPW approaches were estimated using 200 bootstrap samples

¶Full data estimate of model-based standard error: 0.2132 (Table 6.5 of main-text).

M1 = Missingness Mechanism 1; M2a = Missingness Mechanism 2a; M2b = Missingness Mechanism 2b; rBias = Relative bias (%); ModSE = Model-based standard error; EmpSE = Empirical standard error; CP = Coverage Probability; RMSE = Root Mean Squared Error; CC = Complete Case; IPW = Inverse probability weighting model; MI = Multiply Imputed