

**MULTIVARIATE ANALYSIS OF DRUG SAFETY IN CLINICAL TRIALS WITH
APPLICATION TO ANTIMALARIALS IN PREGNANCY IN MALAWI**



By

NOEL PHUMISA PATSON

A thesis submitted to the Faculty of Health Sciences,
University of Witwatersrand, in fulfilment of the requirements for the
Degree of DOCTOR OF PHILOSOPHY with publications in the

Major Area-Subject:

BIostatISTICS AND EPIDEMIOLOGY

May 2022

Declaration

I Noel Phumisa Patson declare that this thesis is my own original work. It is being submitted for a degree of Doctor of Philosophy in the University of Witwatersrand, Johannesburg. It has not been submitted for any degree or examination at any other University or institution.



Noel Phumisa Patson

10th May 2022

Date

Dedication

I dedicate this thesis to my super wife Chiyanjano aka “Chichi”, my awesome two children, Nikos aka “Zamadenga” and Ruby aka “Effie”, my beloved mother Efita and my great dad Davidson.

Original scientific publications

1. **Patson N**, Mukaka M, Otworld KN, Kazembe L, Mathanga DP, Mwapasa V, Kabaghe AN, Eijkemans MJ, Laufer MK, Chirwa T: Systematic review of statistical methods for safety data in malaria chemoprevention in pregnancy trials. *Malaria journal* 2020, **19**(1):119.

Student`s contribution to the paper

Conceptualized the study idea, conducted a search for articles and collected data, analysed the data and interpreted the results, led the writing of the manuscript

2. **Patson N**, Eijkemans MJC Mukaka M, Mathanga D, Laufer MK, Kazembe L, Chirwa T. Comparison of statistical methods for analysis of recurrent adverse events in the presence of non-proportional hazards and unobserved heterogeneity: a simulation study. *BMC Medical Research Methodology* 2022, **22**(1):24.

Student`s contribution to the paper

Conceptualized the study idea, designed the simulation study, ran the simulations and data analyses, interpreted the results, led the manuscript writing

3. **Patson N**, Mukaka M, D'Alessandro U, Chapotera G, Mwapasa V, Mathanga D, Kazembe L, Laufer MK, Chirwa T. Joint modelling of multivariate longitudinal clinical laboratory safety outcomes, concomitant medication and clinical adverse events: Application to artemisinin-based treatment during pregnancy clinical trial. *BMC Medical Research Methodology* 2021, **21**(1): 208

Student`s contribution to the paper

Conceptualized the study idea, analysed the data and interpreted the results, led the manuscript writing

4. **Patson N**, Mukaka M, Peterson I, Divala T, Kazembe L, Mathanga D, Laufer KL, Chirwa T. Effect of adverse events on non-adherence and study non-completion in malaria chemoprevention during pregnancy trial: a nested case control study. *PLoS ONE* 2022, **17**(1): e0262797.

Student`s contribution to the paper

Conceptualized the study idea, analysed the data and interpreted the results, led the manuscript writing

Conference poster presentations

1. **Patson N**, Mukaka M, Otworld KN, Kazembe L, Mathanga DP, Mwapasa V, Kabaghe AN, Eijkemans MJ, Laufer MK, Chirwa T. Systematic review of statistical methods for analysis of antimalarial safety data in pregnancy clinical trials. Drug Safety Africa Conference, North-West University, Potchefstroom, South Africa from November 20–22, 2018
2. **Patson N**, Mukaka M, Otworld KN, Kazembe L, Mathanga DP, Mwapasa V, Kabaghe AN, Eijkemans MJ, Laufer MK, Chirwa T. Systematic review of statistical methods for analysis of safety data in malaria chemoprevention and treatment in pregnancy trials. Wits School of Public Health Biennial Research

Day, University of the Witwatersrand, Johannesburg, South Africa, August 22, 2019

3. **Patson N**, Mukaka M, Peterson I, Divala T, Kazembe L, Mathanga D, Laufer KL, Chirwa T. Role of adverse events on relationship between malaria chemoprevention during pregnancy and non-adherence: Mediation analysis of a randomized trial. American Society of Tropical Medicine and Hygiene conference, Virtual presentation (due to COVID-19), November 15-19, 2020

Awards

1. PhD Biostatistics fellowship- Forgy International Center of US National Institutes of Health (September 2017- 2021)
2. Travel award - ASTMH 2020 conference travel award (November, 2020)
3. Best poster presentation award- Wits School of Public Health Biennial Research Day, University of the Witwatersrand, Johannesburg, South Africa, August 22, 2019

Abstract

Drug safety assessment is important at every stage of drug development especially in large late-phase clinical trials that have the power to detect relatively rare or common events. Clinical trials evaluating antimalarial trials in pregnancy provide a typical setting where safety data receives non-rigorous analysis compared to efficacy data. In particular, improvements in antimalarial drug safety assessment in pregnancy have focussed on study design issues e.g. definition of safety outcomes and data collection. Safety data is usually collected longitudinally and consist of multiple correlated measured endpoints with potential time-varying observations across the study population, requiring advanced statistical methods for analysis. Analytical challenges with such complex data structure include accounting for heterogeneity, dependency between multiple safety endpoints and differences in follow-up-time, e.g. due to non-adherence or study dropout. This thesis investigates optimal statistical methods for analysis of multivariate safety data in clinical trials with application to trials on malaria chemoprevention and treatment during pregnancy. Firstly, using a systematic review, the thesis ascertained the current practice and methodological issues in statistical analysis and reporting of drug safety for malaria chemoprevention in pregnancy trials. The review showed that although most trials reported that they collected multiple longitudinal safety endpoints, statistical analyses were predominantly descriptive and nonparametric, and ignored multiple occurrence. Furthermore, impact of AEs on treatment non-adherence was rarely quantified.

Motivated by the systematic review findings, statistical methods for analysis of recurrent AEs were investigated. In clinical trials, recurrent AEs are a key multivariate outcomes

that may require complex analytical methods due to correlation that arises from the individual patient unobserved heterogeneity and the dependence between events. Shared frailty with proportional hazards models can be used to address the correlation issue., However, the presence of time-dependent treatment effects can lead to poor unobserved heterogeneity estimates. Flexible univariate parametric survival models are now becoming popular in the presence of time-dependent treatment effects. Using a simulation study, flexible parametric lognormal shared frailty models were assessed whether they can improve frailty variance and hazard ratio estimates in investigating treatment effect on AE recurrence; the models were compared to the inverse Gaussian shared parametric model with proportional hazards assumption. The inverse Gaussian shared frailty model yielded higher coverage probability for the 95% confidence interval, lower mean square error and lower bias of the frailty variance estimates compared to the flexible parametric shared frailty models with non-proportional hazards and restricted cubic splines. The flexible parametric lognormal shared frailty models had unbiased log hazard ratio estimates at the expense of precision loss and high mean square error compared to the conventional inverse Gaussian shared frailty model.

In addition to clinical AEs, longitudinal continuous biochemical laboratory safety data and concomitant medication are important multiple longitudinal outcomes for interpreting drug safety in context. For example, drug safety estimates can be confounded by the exposure to the concomitant medication over the follow-up time. Joint modelling of multivariate longitudinal laboratory safety data, concomitant medication and clinical AEs efficiently harnesses the safety data in order to better understand drug safety profile. Using joint

modelling, the thesis investigated whether clinical AEs vary by treatment and how laboratory outcomes (alanine amino-transferase, total bilirubin) and concomitant medication are associated with clinical AEs over time following artemisinin-based antimalarial therapy. Using the novel joint model, AEs were associated with the concomitant medication (IRR=5.75; 95% CI: 4.70, 7.03; $p<0.001$) but there was no evidence of association of AEs with the total bilirubin (IRR= 0.97; 95% CI: 0.60, 1.56; $p=0.906$) and alanine aminotransferase (IRR=1.12; 95% CI: 0.91, 1.38; $p=0.269$). Generally, the Poisson model yielded lower effect estimates of treatment on AE incidence compared to the joint model.

Finally, this study investigated whether nested case control (NCC) study with incidence density sampling is more efficient than NCC with path set sampling under conditional logistic or weighted cox models in assessing the effect of AEs on treatment non-adherence and participation in preventive antimalarial drug during pregnancy trial. Similar estimates were obtained under incidence density sampling and path set sampling whether weighted cox or conditional logistic models were used. The results indicated no evidence that the clinical AEs impacted treatment non-adherence or study non-completion.

Overall, the thesis work developed and applied novel flexible parametric models to improve contextualized profiling of the antimalarial drug safety when there are multiple or repeatedly-measured and time-to-event safety outcomes. NCC was also adapted to aid assessment of impact of AEs on clinical trials and inform future trial designs.

Acknowledgements

This doctoral work was funded by the Fogarty International Center of the National Institutes of Health [Grant number D43TW010075]; National Institute of Allergy and Infectious Diseases [Grant number K24AI114996] to Prof Miriam Laufer and Prof Don Mathanga.

I would like to thank my supervisors; Prof Tobias Chirwa, Prof MK Laufer, Prof Mavuto Mukaka, Prof Lawrence Kazembe and Prof Marinus JC Eijkemans. What a great dream team! It was a lifetime honour to be supervised by such a great set of five top class professors. Your guidance and support fuelled my passion to continue driving towards the PhD goal with joy.

The PhD journey was decorated with a great opportunity to work with great experts who mentored me and some co-authored my PhD papers. A special vote of thanks should go to Prof Don Mathanga, Prof Victor Mwapasa, Prof Gertrude Chapotera, Prof Umberto D'Alessandro, Dr Alinune Kabaghe, Dr Kennedy Otworld, Dr Ingrid Peterson, and Dr Titus Divala. Thanks for your great ideas that shaped the course of my PhD work. I wish also to thank Prof Ken Maleta and Prof John Phuka who introduced me to the world of clinical trials in 2011. Thanks Prof Maleta for pushing me to improve my original PhD concept. You have truly supported me to grow professionally in so many ways.

I would like to express my gratitude to Wits School of Public Health Staff who greatly supported me in my academic endeavours. A special thank you should go to Prof Jonathan Levin, Prof Eustasius Musenge, Mr Paul Bohloko and Dr Jude Igumbor.

I wish to thank my family and friends for being supportive and enduring my long absence even in times when they needed me most. We did this PhD together! Thank you my siblings Aaron, Jesna, Edna, Resha and Phunziro. Only God knows how I can express my gratitude to my Wife Chiyanjano, son Nikos, daughter Ruby, dad Davidson and mum Efitia. To my friends Limbani and Ruth Chakhoma, Mwiza Singini, Albert Sharra, Okechi Achilonu, Christopher Stanley, John Mbotwa, Ps Saulos Huwa, Francis and Brenda Jiva, Melanie Kaminjolo, and Sam and Julie Ndawala, thanks for being there for me. My brother-in-law, Ngwiro Mchakama, thank you for the unwavering support. A special parcel of historic thank you should go to the mighty Ps Bill Johnson of Bethel Church, Redding, CA, USA and Ps Phil Dooley of Hillsong Church, RSA for the spiritual java that energized me to continue hosting God`s presence. Oh yes! I don`t have the capacity to exaggerate God`s goodness. Jesus honoured me with a great life. Dear Lord, I will let this life sign Your name and thunder for You in everything that I do. Again and again, I will boast that you are the “*captain of my days*”. You preserve my identity wherever I go and in whatever I do. I can now proudly shout, “Go and tell it in Gath! Let it sink among the Philistines. This is my God and I bear his marks”.

Table of Contents

Declaration	ii
Dedication	iii
Abstract	vii
Acknowledgements	x
Table of Contents	xii
List of Tables	xvii
List of Figures	xix
Abbreviations	xx
Chapter 1: Background	1
1.1 Clinical trials and drug safety assessment	1
1.2 Drug safety outcomes data in clinical trials	3
1.3 Antimalarial drug safety in pregnancy	6
1.4 Problem Statement	8
1.5 Justification	9
1.6 Aim and specific objectives	10
1.6.1 Main aim	10
1.6.2 Specific objectives	10
1.7 Thesis outline	11

Chapter 2: Statistical methods for safety data analysis in drug trials: A general overview	13
2.1 Available statistical methods for safety data analysis in drug trials	14
2.2 Summary of statistical issues in drug safety data assessment	17
2.3 Pharmaco-epidemiological approach to assessment of AE burden	22
Chapter 3: Methodology	24
3.1 Motivating Data sources.....	28
3.1.1 Chloroquine for malaria in pregnancy	29
3.1.2 PREGACT trial.....	31
3.2 Ethical considerations	33
3.3 Analytical methods.....	34
3.3.1 Review of statistical methods for safety data analysis	34
3.3.2 Comparison of shared frailty models for recurrent AEs analysis.....	35
3.3.3 Joint modelling of safety outcomes.....	39
3.3.4 Role of AEs on antimalarial drug non-adherence	45
Chapter 4: Systematic review of statistical methods for safety data in malaria chemoprevention in pregnancy trials.....	49
4.0 Abstract.....	50
4.1 Background.....	52
4.2 Methods	55

4.2.1 Search strategy.....	55
4.2.2 Data extraction.....	57
4.2.3 Data synthesis	58
4.3 Results.....	58
4.3.1 Characteristics of the trials.....	61
4.3.2 Characteristics of the reported safety data	61
4.3.3 Statistical analysis for safety data.....	63
4.4 Discussion.....	66
4.5 Conclusion	72
Chapter 5: Comparison of statistical methods for the analysis of recurrent adverse events in the presence of non-proportional hazards and unobserved heterogeneity: A simulation study	
5.0 Abstract.....	74
5.1 Background.....	76
5.2 Methods	80
5.2.1 Notation	81
5.2.2 Statistical methods for recurrent events.....	81
5.2.3 The Shared Frailty Models.....	83
5.2.4 Simulation Study.....	87
5.2.5 Ethical considerations.....	93

5.3 Results	93
5.3.1 Frailty variance estimates and corresponding bias across the shared frailty models	93
5.3.2 Percentage gain in precision for frailty variance estimates of non-proportional hazards shared frailty models with restricted cubic splines relative to ISF-PH model	99
5.3.3 Bias and MSE for the estimated log hazard ratio estimates across the shared frailty models under varying unobserved heterogeneity and sample size	99
5.3.4 Example data application: <i>Chloroquine for malaria in pregnancy</i> trial	102
5.4 Discussion.....	105
Chapter 6: Joint modelling of multivariate longitudinal clinical laboratory safety outcomes, concomitant medication and clinical adverse events: Application to artemisinin-based treatment during pregnancy clinical trial	109
6.0 Abstract.....	110
6.1 Background.....	112
6.2 Methods	115
6.2.1 Data	115
6.2.2 Ethical considerations	116
6.2.3 Notation and Model Specifications.....	117
6.2.4 Statistical data analysis	125
6.3 Results	128

6.3.1 Descriptive results	128
6.3.2 Parameter estimates for the models	133
6.3.3 Sensitivity analyses	137
6.4 Discussion.....	140
6.5 Conclusion	144
Chapter 7: Effect of adverse events on non-adherence and study non-completion in malaria chemoprevention during pregnancy trial: A nested case control study.....	
7.0 Abstract.....	146
7.1 Background.....	148
7.2 Methods	151
7.2.1 Motivating data description	151
7.2.2 Nested case control study designs	153
7.2.3 Statistical analysis	156
7.3 Results.....	159
7.3.1 Comparison of parameter estimates.....	161
7.4 Discussion.....	166
7.5 Conclusion	169
Chapter 8: Discussion	
8.1 Summary of key findings.....	170
8.1.1 Current practice in analysis of safety data	172

8.1.2 Recurrent AEs with time-dependent effects and unobserved heterogeneity	173
8.1.3 Multiple safety outcomes with repeated measurements analysis	173
8.1.4 Impact of AEs in clinical trials	174
8.2 Study strengths	175
8.3 Study limitations	176
8.4 Future work	177
8.5 Conclusions	179
8.6 Recommendations	179
References	181
Appendices	204
Appendix 1: Supplementary tables and figures	204
Appendix 2: Original articles published	210
Appendix 3: Ethical clearance	261

List of Tables

Table 4.1: Overview of randomized clinical trials included in the systematic review	60
Table 4.2: Characteristics and structure of reported safety data (n=18)	62
Table 4.3 Statistical approaches for safety data analysis	64
Table 4.4 Recommendations for analyses in malaria chemoprevention in pregnancy trials based on key safety outcomes from the systematic review	71

Table 5.1: Frailty variance estimates bias, coverage probability and MSE across shared frailty models in the presence of unobserved heterogeneity and non-proportional hazards.	95
Table 5.2: Bias and MSE for log hazard ratio estimates, estimated at 3 months, across shared frailty models in the presence of both unobserved heterogeneity and non-proportional hazards.....	101
Table 5.3: Comparing sharing frailty models for analysis of recurrent AEs among pregnant women on IPTp treatment in Malawi	104
Table 6.1: Baseline characteristics for women enrolled in PREGACT trial in Malawi..	128
Table 6.2: Summary of AE occurrence and concomitant medication use for women enrolled in PREGACT trial.....	129
Table 6.3: Summary of concomitant medication reported by pregnant women in PREGACT trial	131
Table 6.4: Parameter estimates for Clinical AE count incidence using conventional Poisson model.....	134
Table 6.5: Parameter estimates from the joint multivariate longitudinal clinical laboratory, concomitant medication and AE counts model.....	136
Table 6.6: Parameter estimates from the joint multivariate continuous longitudinal clinical laboratory, paracetamol concomitant medication count and AE count model .	140
Table 7.1: Distribution of the study cohort characteristics (n=600).....	160

Table 7.2: Association between AE occurrence and treatment non-adherence; comparison of estimates from risk set sampling and path sampling under multivariable conditional logistic regression model and weighted Cox regression model.....	163
---	-----

Table 7.3: Association between AE occurrence and study non-completion; comparison of estimates from risk set sampling and path sampling under multivariable Conditional logistic regression model and weighted Cox regression model.....	164
--	-----

List of Figures

Figure 4.1: PRISMA flow diagram profiling studies selection process.....	59
--	----

Figure 5.1: Average model standard errors(SE) for the frailty variance estimates across shared frailty models under varying data-generating mechanisms	98
---	----

Figure 6.1: Box plots profiling clinical laboratory safety outcomes for antimalarial drugs over 63-day follow-up among pregnant women in Malawi.....	132
---	-----

Figure 7.1: Time to first AE occurrence among the pregnant women on IPTp up to delivery	159
--	-----

Abbreviations

AEs	Adverse events
RCTs	Randomized controlled trials
WHO	World Health Organization
IPTp	Intermittent Presumptive Treatment in pregnancy
SP	Sulfadoxine-Pyrimethamine
ACTs	Atermisinin Combination Therapies
MEdDRA	Medical Dictionary for Regulatory Activities
MAV	markedly abnormal values
MCAR	missing completely at random
MAR	Missing at random
MNAR	missing not at random
ITT	Intention to treat
PP	Per protocol
LMICs	low and middle income countries
NCC	nested case control
CQ	Chloroquine
MiPL	Malaria in Pregnancy Library
CONSORT	Consolidated Standards of Reporting Trials
ISF-PH	inverse Gaussian shared frailty model with proportional hazards
	Lognormal Shared frailty model with non-proportional hazards and
LSF-NPHS1	restricted cubic splines where the baseline hazard is modelled using

	1 degree of freedom and the time-dependent treatment effects is modelled using 1 degree of freedom
	Lognormal Shared frailty model with non-proportional hazards and restricted cubic splines where the baseline hazard is modelled using 2 degrees of freedom and the time-dependent treatment effects is
LSF-NPHS2	modelled using 1 degree of freedom
	Lognormal Shared frailty model with non-proportional hazards and restricted cubic splines where the baseline hazard is modelled using 3 degrees of freedom and the time-dependent treatment effects is
LSF-NPHS3	modelled using 1 degree of freedom
	Lognormal Shared frailty model with non-proportional hazards and restricted cubic splines where the baseline hazard is modelled using 4 degrees of freedom and the time-dependent treatment effects is
LSF-NPHS4	modelled using 1 degree of freedom
DGM	Data generating mechanism
SE	standard error
SD	standard deviation
MSE	mean square error
MCSE	Monte carlo standard error
PRISMA	Preferred Reporting Items for Systematic Review and Meta-Analyses
ISTp	intermittent screening and treatment in pregnancy (ISTp)
RDT	Rapid diagnostic testing
OR	Odds ratio

IRR	Incidence rate ratio
CI	Confidence interval
HR	Hazard ratio
DSMB	Data Safety and Monitoring Board
TDE	Time-dependent effect
AL	artemether–lumefantrine
ASAQ	amodiaquine–artesunate
DHAPQ	dihydroartemisinin– piperazine
PCR	Polymerase-Chain reaction
IPW	Inverse probability weighting
AIC	Alkaike Information Criterion

Chapter 1: Background

This section provides introduction, problem statement, justification, objectives and the thesis outline. The focus is on antimalarial drug trials in pregnancy.

1.1 Clinical trials and drug safety assessment

Drug safety assessment is of paramount importance at every stage of drug development such that safety data is collected, either actively or passively or both actively and passively, at each phase of drug development. The safety data collected during the premarketing phase of the drug development provides the supporting basis for drug approval and labelling. Volume and structural complexity of data for drug safety assessment continues to increase from pre-clinical to clinical stages of drug development such that statistical analysis of the safety data may depend on the clinical trial phase. Therefore, different trial phases may pose unique issues on how to handle the drug safety data due to variations in properties of the clinical trial phases. The variations are usually influenced by the objectives of the respective phases. The early randomized controlled trial (RCT) phases during drug development are associated with smaller sample sizes where patients are mostly not randomized compared to the later phases of clinical trials. Moreover, in early phase of clinical trials, drug safety conclusions are mainly based on surrogate biomarkers (such as pharmacokinetics) unlike clinical endpoints mostly used in late Phases (Phase III/IV). Therefore, as the sample sizes increase from Phase I to III, more safety data is collected including multiple safety endpoints (e.g. blood chemistry and clinical Adverse Events (AEs)) that are inter-related and measured repeatedly over time [1]. Such multivariate longitudinal nature of the safety data introduces complexity in

the statistical analysis such as how to account for follow-up time, correlation and multiplicity that poses challenge to drug safety assessment [2-4]. The analysis of this complex safety data is rarely pre-specified since most safety endpoints are not known in advance and most trials are not adequately powered to detect safety effect [5]. Therefore, at minimum, the safety analyses should provide answers not only to pre-specified questions only, but also insight into the structure of raw data, new questions and information on safety profile of the drug(s) under investigation [1].

Although characterizing the safety profile of a drug is considered important in clinical trials, statistical analysis of the safety data is traditionally based on descriptive statistics that do not explicitly account for the safety data complexity as indicated above [6]. The traditional inferential methods used in safety data analysis in clinical trials can miss the dependences of the safety outcomes and may not be sufficiently powered to detect an effect if it exists. Sometimes, this can lead to inconclusive drug safety evidence which may affect decisions on future drug use for fear of potential undetected risk. Use of optimal statistical methods that can adequately account for longitudinal structure of the safety data and potential variation in follow-up time for individual patients can help in efficiently characterizing drug safety profile based on the available data. Despite the highlighted importance of drug safety assessment, most of development and application of statistical methods in clinical trials have mainly focus on efficacy data unlike safety data [7]. This work focusses on practical development and application of longitudinal and survival statistical methods for safety data analysis in premarketing clinical trials using antimalarial drugs in pregnancy trials as a case study.

Although Data Monitoring Committees are instituted during RCTs to guard the safety of the patients by continuously reviewing the safety data, their decisions are mainly based on interim data. Furthermore, statistical methods applied during the continuous monitoring of drug safety during drug trial are mainly based on descriptive approaches that sometimes may be limited. Analysis of safety data at the end of a drug trial offers an opportunity to profile drug safety based on as much as available safety data collected during the clinical trial. Therefore, use of statistical methods that can effectively draw evidence of drug safety is considered paramount for improved drug development and safe-guard safety of patients. This work focusses on analysis of drug safety data at the end of a clinical trial.

1.2 Drug safety outcomes data in clinical trials

Clinical trials are a gold standard for determining drug safety and efficacy. Important drug safety endpoints collected during clinical trials include vital signs, clinical laboratory measurements, electrocardiograms, hospitalizations and clinical adverse events. Vital signs are a set of patient's body measurements used to assess general physical health of an individual such as pulse rate [8]. Clinical laboratory tests in RCTs involve testing samples from the human body such as blood with the aim of determining diagnosis, planning and monitoring treatment and disease progression or investigational drug mechanism [9]. The laboratory tests measure actual safety parameter such as hepatotoxicity (e.g. albumin), cardiotoxicity (e.g. QT-interval elongation), neurotoxicity (e.g. electroencephalograms). Based on pre-established reference standards, these

parameters can be used to define laboratory-determined AEs. Sometimes, hospitalization, defined as any sickness episode leading to admission at a hospital over the follow-up time of a patient, is also considered as a safety endpoint. AE occurrence is the main safety endpoint that is evaluated and reported in most clinical trial reports. An AE is defined as any untoward medical occurrence associated with the use of a drug in humans regardless of whether or not it is considered drug related [10]. The AEs can be collected actively or passively as patient reported outcome or based on abnormal laboratory measurement. Usually, the AE information is recorded in text format before coding using Medical Dictionary for Regulatory Activities (MedDRA) [11, 12]. The dictionary is based on systematic five-level hierarchy where system organ class is coded at the highest level (e.g. gastro-intestinal disorders) and lowest level terms at the bottom (e.g. feeling queasy). The AEs can also be graded based on severity: mild = grade 1, moderate = grade 2, severe = grade 3, life threatening = grade 4 and death = grade 5.

Concomitant medication, defined as medications that are taken apart from the investigational drug, are also commonly taken in clinical trials[13, 14]. The concomitant medication has the potential to interact with the investigational drug and eventually affect the safety endpoints. If the concomitant medication interacts with the investigational drug, there is a potential for the drug safety estimates to be confounded by the exposure to the concomitant exposure. Although in some clinical trials use of concomitant medications is included as part of inclusion and exclusion criteria, some patients may start taking concomitant medications during the trial to treat an indication arising in the course of the trial. Therefore, in clinical trials, whether concomitant

medications are part of the inclusion and exclusion criteria or not, information on concomitant medication use is still collected to support interpretation of drug safety estimates [14]. Demographic

characteristics and concurrent disease are also other important data that are useful in supporting computation and interpretation of drug safety estimates since they can potentially confound the estimates obtained.

All the safety outcomes and supporting data described above have complex structure. For example, the AEs can recur and the clinical laboratory measurements are repeatedly measured over the follow-up time. The multiple safety outcomes, are also inter-related such that ignoring the potential correlation of the outcomes can lead to biased estimates. Depending on the length of the follow-up time and frequency of the outcomes, the time to occurrence of a safety outcome is also supposed to be modelled as efficiently as possible. The safety outcomes can also be non-normal and a patient can experience non-linear treatment effects over the follow-up time such that use of the conventional approaches to analyse the data would be inappropriate. These complex issues pose a big challenge in efficient characterising of drug safety profile. Moreover, the scope of these issues can vary depending on the indication under study. Clinical trials for antimalarial drugs in pregnancy presents a typical setting where the highlighted challenges are faced. Hence, premarketing antimalarial drug clinical trials in pregnancy setting is considered as a case study in all the subsequent discussions in this work.

1.3 Antimalarial drug safety in pregnancy

Drug safety in pregnancy has been a public health concern since late 1950s and early 1960s when the drug thalidomide caused more than 10,000 babies to be born with congenital anomalies. The drug was primarily developed as a sedative and was used as a nausea and vomiting inhibitor in pregnancy [15-17]. Since then, during pregnancy, drugs are prescribed and used with caution [18, 19] due to the potential risk to the mother or foetus. However, in malaria-endemic regions it may be inevitable to use an antimalarial drug, therapeutically or prophylactically, during pregnancy. Globally, 125 million pregnant women are at an annual risk of malaria and about 33 million of them are from Sub-Saharan Africa [20-22]. A World Health Organization (WHO) malaria report for 2021 estimated malaria prevalence among pregnant women to be 34% (11 million cases) [20]. The WHO recommends that in high or moderate malaria transmission settings, malaria in pregnancy should be prevented by Intermittent Presumptive Treatment (IPTp) with Sulfadoxine-Pyrimethamine (SP) and treated by Artemisinin Combination Therapies (ACTs). Malaria chemoprevention and treatment in pregnancy reduces risk of anaemia and placental malaria in mothers [23]. Limited evidence of safety for such anti-malarial drugs in pregnancy [23, 24] has affected increased uptake of such treatments. The limited antimalarial drug safety evidence has been attributed to study methodology issues (e.g. uncontrolled for confounders, low power) [25]. Recent literature also indicates that there is inadequate reporting and varied methodological assessments of antimalarial safety in pregnancy [23, 26]. This limits interpretations and conclusions that can be made concerning antimalarial drug safety in pregnancy across clinical trials. Loss in efficacy of SP also calls for alternative malaria prevention drugs during pregnancy [27, 28]. However,

a challenge remains to find an alternative malaria prevention treatment to SP that is adequately safe and tolerable. For example, in recent trials, neither amodiaquine, mefloquine, nor chloroquine–azithromycin are suitable replacements for SP due to poor tolerability [29-35]. However, most analyses from such studies are based on descriptive statistics, largely focussing on reporting the Serious AEs (SAEs). Such reporting approach has a potential to miss information on the progression of the AEs burden e.g. occurrence of mild AE can be a precursor to occurrence SAE. Limited antimalarial drug safety evidence in pregnancy has also been attributed to systematic exclusion of pregnant women from clinical trials due to fear of potential side effects [24]. In the last two decades there has been increased availability of antimalarial drug efficacy evidence due to increased clinical trials including pregnant women. There is a consensus within the clinical trial community that drug safety assessment is equally important as efficacy assessment, as demonstrated by mandatory institution of Data Safety and Monitoring Board. However, there has been limited development and application of the methods for drug safety assessment. Hence, there is still a gap pertaining to antimalarial drug safety evidence since the rigorous drug assessment methods and reporting mainly focus on drug efficacy rather than drug safety, partly, due to limited development of the statistical methods. Improved development and application of drug safety assessment methods and reporting has the potential to make efficient use of the available antimalarial drug safety data for improved decision making.

1.4 Problem Statement

Malaria remains a disease of public health importance, causing high mortality and morbidity among pregnant women in malaria-endemic regions. Identification of safe malaria prevention and treatment therapies is of key interest in malaria chemoprevention and treatment RCTs. The absence of conclusive safety evidence may have been partially due to safety concern of the mother and foetus that may be compromised due to their physiological pattern because of hormonal changes. Additionally, the use of conventional statistical analyses of the available safety data may not have been optimal. The multivariate nature of the safety data introduces complexity such as dependence and multiplicity that poses challenge to drug safety assessment. Fortunately, multivariate statistical methods exist but they are rarely used in literature towards the analysis of safety data. The limited development and use can be due to potential challenges associated with implementation of the methods and lack of knowledge on implications for using the methods that do not explicitly account for multivariate nature of the safety data. Researchers in oncology studies have recognised and documented the importance of rigorous safety analysis [36-38]. However, such information is limited in infectious diseases studies, especially those studying diseases with short latency such as malaria. The disease progression and AEs in antimalarial research can differ from chronic diseases due to biological differences associated with the diseases/AEs. Oncology diseases are longer in duration and slower in progression [39] unlike malaria. Descriptive statistics dominate drug safety assessment in malaria prevention and treatment in pregnancy. Such methods have limited capacity to adequately address issues associated with safety data such as multiplicity, recurrence of the AEs and multivariate longitudinal

structure of safety data. Improved methods need to be identified and investigated to ensure reliable and valid statistical inferences. Furthermore, for long term trials, patient treatment discontinuation due to AEs can be high [40-42]. So far, limited work has been done to address inadequate development, evaluation and application of statistical methods for the analysis of multivariate safety data in malaria chemoprevention and treatment in pregnancy RCTs.

1.5 Justification

Many statistical methods have been developed and applied to analyse efficacy data but not much has been done directly for the analysis of safety data [43], despite the fact that a lot of such data (70-80%) is generated in RCTs [2]. Therefore, there is a need to appropriately analyse safety data to ensure balanced drug evaluation. We acknowledge that issues of safety are considered across all phases of clinical trials, more so in Phase I and II studies. This study focusses on Phase III trials because they have largest sample sizes of all pre-market trials and are mostly carried out for a longer period and sometimes across many centres. Determining safety at Phase III trial is beneficial because it prevents majority of the population from being exposed to unsafe drugs. Furthermore, poor-resource settings can benefit a lot when safety data is sufficiently analysed at phase III premarketing stage, since pharmaco-vigilance systems are not well-developed [44]. Hence, the main source of drug safety evidence for such settings is pre-marketing RCTs. Malaria RCTs generate multivariate longitudinal safety data such as repeatedly measured vital signs and AEs. If such data are properly analysed simultaneously they can give sufficient evidence on the safety of the experimental drugs. Such evidence can help to

reduce burden associated with unsafe drugs such as disabilities, morbidity and mortality. This study contributes towards determining appropriate analysis methods of safety data for antimalarial drug trials in pregnancy.

1.6 Aim and specific objectives

1.6.1 Main aim

The main aim of this work was to investigate suitable statistical methods for analysis of multiple safety outcomes in clinical trials in the context of antimalarial drugs for malaria prevention and treatment in pregnancy clinical trials. In order to achieve this, we considered four specific objectives below.

1.6.2 Specific objectives

1. To conduct a systematic review of statistical methods for safety data analysis in malaria chemoprevention during pregnancy trials.
2. To evaluate performance of shared frailty models for recurrent adverse events analysis, in the presence of unobserved heterogeneity and time-dependent effects, using simulations.
3. To determine factors associated with adverse events over time in the presence of multiple safety outcomes.
4. To determine the association between adverse events and non-adherence in malaria chemoprevention during pregnancy trial using nested case control approach.

1.7 Thesis outline

Chapter 2 provides an overview of the drug safety assessment from statistical perspective based on the existing literature that motivated this PhD work. Chapter 3 consists of a summary of methods employed in developing results provided in chapters 4 to 7. Full details of the methods appear in the respective chapters that follow. Chapter 4 of this work presents an overview of the current practice in the safety data analysis in malaria chemoprevention during pregnancy trials based on a systematic review. The chapter highlights the current practice in drug safety data analysis and reporting and key issues that need rigorous methodological attention beyond the current practice. Chapter 5 presents a simulation study comparing shared frailty models for analysis of recurrent AEs especially in long term clinical trials; long term trial was arbitrarily defined as a trial designed to follow-up patients for at least more than two months. The chapter presents work on evaluating properties of shared frailty models focussing on the unobserved heterogeneity in the presence of treatment time-dependent effects. Choosing the model that effectively captures time-dependent treatment effects and unobserved heterogeneity is important in antimalarial drug safety assessment in pregnancy trials. For example, physiological changes that occur during pregnancy can contribute towards the unobserved heterogeneity in drug safety estimates.

Chapter 6 introduces a novel joint model for analysis of multiple safety outcomes collected during drug clinical trials. The chapter demonstrates how to harness laboratory biochemical information, clinical adverse events and concomitant medication in developing the antimalarial drug safety profile.

Chapter 7 focusses on assessing the burden of AEs in clinical trials especially on treatment non-adherence and trial non-completion in the context of IPTp trials. It demonstrates how the nested case-control study design approach can be useful in assessing the AE burden. The chapter provides knowledge to aid in choosing the most appropriate nested case control design and analytical methods, leveraging rich data from clinical trials. Such evidence supports improved practice in interpreting the findings from the safety data in clinical trials especially in the context of safety data analysis since the outcomes are not pre-specified.

Finally, chapter 8 provides a unifying framework for safety data analysis in antimalarial drug trials in pregnancy based on the key findings in the preceding chapters. The chapter highlights key findings of public health relevance, recommendations, the study strengths and weakness. Some directions for potential future work are also suggested.

Chapter 2: Statistical methods for safety data analysis in drug trials: A general overview

Increased calls for improved statistical analysis and reporting of safety data in drug trials have been observed in literature especially in the last two decades [1, 45-47]. Some guidelines and recommendations for safety data analysis have been issued [48-51]. However, there is still limited development and application of improved statistical methods for safety data analysis in drug trials [6].

Since safety data usually consists of multivariate complex data such as interrelated multiple outcomes (mostly) repeatedly measured over time, in choosing statistical methods to analyse such data, one should ensure that they adequately address statistical issues associated with analysis of the respective outcome(s). Therefore, in the subsequent three sections (Sections 2.1-2.3), an overview of the existing proposed and applied statistical methods for safety data analysis in drug clinical trials and key statistical issues associated with the analyses are provided. The sections highlight the scope of importance of the statistical analysis of safety data in drug trials and provides a general broader context that motivated our current work. In particular, this chapter provides an overview of the drug safety assessment from a statistical perspective, based on the existing literature that motivated this PhD work.

2.1 Available statistical methods for safety data analysis in drug trials

Overall, statistical methods for analysing safety data in clinical trials can be considered in two categories based on stage when they are implemented and how they are implemented. These methods are those that attempt to capture safety signals for accumulating safety data for a clinical trial in progress (i.e. dynamic or sequential safety evaluation methods) and those that summarize entire drug safety profile at the end of a clinical trial (i.e. static safety evaluation methods) [52, 53]. The key difference between the dynamic and static safety evaluation methods is based on the time when they are implemented. Therefore, both dynamic and static safety evaluation methods can include longitudinal or non-longitudinal methods, survival or non-survival methods etc. In this work, we focus on the static safety evaluation methods in premarketing drug trials since they provide a detailed evidence on drug safety profile based on data from the completed trial unlike dynamic safety evaluation methods that focus on detecting safety signals based on partial safety data accumulating in an ongoing clinical trial.

There are further ways of categorizing the available statistical methods for drug safety data based on the objective of interest and the analytical approach. A recent review classifies the available statistical methods for drug safety data into four categories; hypothesis testing (i.e. frequentist), estimation (i.e. quantify distribution difference of safety outcomes between the treatment groups), visual summary (i.e. graphical) and decision making probability methods (i.e. Bayesian) [54]. Graphical display simplifies informal comparisons of safety outcomes between the treatment arms especially for large volume of data [55, 56]. In this work, we focus on estimation methods and hypothesis

testing methods within frequentist framework since they are widely accepted methods in analysis of clinical trials, especially in the traditional parallel trial design.

Within the frequentist framework, multivariate methods have been implemented to help in deriving overall safety profile focussing on AEs. For example, Agresti and Klingenberg implemented multivariate approach to assess the global difference for two vectors of proportions based on placebo and treatment effects across different AE types [57]. However, as highlighted in the preceding section, most of the estimation and hypothesis testing methods used in drug safety assessment in clinical trials are predominantly descriptive. On the other hand, Bayesian statistical methods are also useful in safety signal detection since they can borrow information from other events that are assumed medically related, can account for prior information and useful in controlling for multiplicity [58]. For example, a three-level hierarchical logit mixture model by Berry and Berry accounts for multiplicity by explicitly modelling the relationship among the AEs based on their body systems [11]. The model yields posterior probability comparing the treatment group and the control group. Under Poisson likelihood and binomial likelihood Xia et al proposed an extension of the Berry and Berry method by accounting for exposure time for the study participants [59]. Key advantage of such Bayesian hierarchical modelling approach is that it accommodates explicit modelling of the AEs within the MEdDRA coding framework. A main concern for such modelling approach is that it can mask outliers that are important in drug safety assessment. Furthermore, model misspecification may sometimes affect the inference drawn.

Apart from clinical AEs, clinical laboratory measurements present key safety endpoint that is underutilized and present great challenge during analysis. The common practice in the analysis of clinical laboratory data involves use of shift tables, descriptive statistics and markedly abnormal values (MAV) [2]. Shift tables displays number of subjects whose laboratory value(s) are low, normal or high while on-treatment/post-dose compared to the baseline value. Descriptive statistics approach compares mean/median changes from baseline value with on-treatment/post-dose value. The markedly abnormal values approach involves identifying outlying values (i.e. an observation that deviates markedly from the pattern or distribution of the majority of the data) [2]. Such categorization can potentially lead to loss of information. All the three approaches above are inadequate to account for multivariate nature, potential correlation and non-normality of clinical laboratory data values. In order to enhance efficient use of clinical laboratory data, advanced analytical approaches need to be considered that may account for the outlined statistical analysis challenges. Although some advanced approaches to modelling of clinical laboratory data are available, they inefficiently use the available information based on the complex structure of the safety data. For example, the proposed univariate longitudinal analysis of laboratory data using generalized estimating equation and likelihood analyses [60] can further miss inter-outcome correlations. Earlier proposed pseudo-multivariate data dimensionality reduction approach that generates Genie score for overall safety do not also explicitly account for dependence of the outcomes and the score obtained has no explicit statistical or medical meaning. Recently, using extreme value theory, a formal joint modelling of MAV accounting for its multivariate nature of clinical laboratory safety data has been proposed [61]. Such joint modelling is inefficient

since it doesn't use all the available information due to focus on extreme values and it is also mainly appropriate for large sample sizes. Development and application of joint models linking the clinical laboratory safety data and clinical AEs can be efficient in developing comprehensive drug safety profile based on the available data. For example, joint model of multivariate longitudinal clinical laboratory safety data and clinical AEs would be useful in understanding how hazard of AEs is affected by longitudinal process of the continuous clinical safety laboratory-based measurements. The joint modelling can efficiently reveal the pattern of joint changes of multiple outcomes over the follow-up time.

2.2 Summary of statistical issues in drug safety data assessment

Multivariate longitudinal nature of safety data in RCTs poses challenges towards drug safety assessments. For example, the multiple safety parameters such as recurrent AEs, multivariate longitudinal bilirubin and alanine aminotransferase may be repeatedly measured over time and interrelated. Such multiple safety outcomes observed include the unexpected events/outcomes. Therefore, it is sometimes difficult to pre-specify hypothesis for safety events in clinical trials at design stage [3]. This can make drug safety assessment to have the potential for multiple exploratory endpoints. In efficacy analysis, concern is to prevent type I errors while in safety analysis the concern is to prevent type II errors to avoid the possibility of introducing a toxic drug into clinical practice. Due to multiple safety endpoints in clinical trials, false positives may arise from statistical testing without appropriate multiplicity adjustment. Moreover, too strict adjustment for multiplicity may further limit the already finite power available to detect safety signal. Hence, the

statistical analysis for safety data should strive for reasonable method(s) that will not under-adjust or over-adjust for multiplicity.

Currently, safety data analysis and reporting focusses on proportions, cumulative incidence, and sometimes counts of serious AEs [38, 62]. However, in most RCTs, time is an important element as individuals are followed up to measure outcome of interest. In most published trials, reported proportions by intervention arm have been compared descriptively, without explicitly taking into account the time to occurrence [36, 37, 62] and multivariate nature of the outcomes. Sometimes, time to first AE is modelled and this can miss information on the recurrent AEs. Recurrent AEs can provide the progression of the AE burden. Cumulative incidence approach partly accounts for time but it may be misleading if there are multivariate events. The cumulative incidence may yield imprecise results since it accounts for individual follow-up time based on aggregated time. Use of weighted Poisson model to address the problem may be inefficient since it assumes that treatment effect is constant within given bands of follow-up time. Application of the inadequate statistical methods can lead to failure to capture the drugs' associated safety issues at pre-marketing stage [63, 64]. For multivariate survival data such as recurrent AEs, it has been shown that considering the first AEs only can lead to underreporting of safety associated with a drug [65, 66].

The RCT design and outcome of interest determines the most suitable statistical methods for analysing safety data. Patients can drop out in study and this generates missing data that introduces imbalance across the treatment arms. Sometimes the dropout may be due

to an AE which may yield informative censoring. Therefore, assessing dropout pattern can be informative on safety evaluation. Missing data can be in three categories [67]. Firstly, if the probability of dropping out is independent of observed and unobserved measurements, it is defined as missing completely at random (MCAR). Under MCAR loss of statistical power can be expected but does not impact bias [68]. Secondly, if the probability of dropping out is dependent on observed data, but not the unobserved data, it is termed as missing at random (MAR). If missingness depends on observed or unobserved variables, it is defined as missing not at random (MNAR). With the missing data, Intention To Treat (ITT) analysis for either safety or efficacy data may be invalid since it assumes that data from all the participants is available [69]. Opting for complete case analysis can reduce the power of the study.

Another challenge in statistical analysis has been dealing with under-evaluated and rare safety data [70]. Although meta-analyses of the AEs have been proposed as solution [71], the inadequate analysis of the individual RCTs make the meta-analyses unreliable. For the rare outcomes, there are existing penalized methods [71, 72] that can be investigated and extended to improve the predictive and estimation accuracy of safety data. For both rare and common outcomes, research synthesis is another alternative where there is re-analysis of complete original data from previous different studies [73]. Unlike in meta-analysis where the whole study has one statistic, in research synthesis each patient contributes towards the overall estimate.

Heterogeneity in definitions of outcomes also negatively affects standardized and efficient analysis of safety data in clinical trials[74]. Availability of guidelines based on the indication under study ensures systematic definition of safety outcomes. Recently there have been attempts to recommend definitions of safety outcomes in antimalarial drug safety assessment [75]. However, regulatory-body-recommended guidelines on analysis of such standardized safety outcomes remain limited.

Proliferation of multicentre clinical trials where a trial is conducted at more than one centre but following the same protocol, has created additional complexity to analysis of large of safety data that is generated. Multicentre drug trial provides a strong basis for generalization of study findings since the patients can be recruited from wider clinical and population settings. Multicentre trials also help in recruiting the required study participants within a short period of time. Although use of the same protocol across all centres ensures uniform data collection, heterogeneity in both clinical practice and population characteristics can confound interpretation of the study findings. Accounting for the heterogeneity arising due to centre differences poses a great analytical challenge since several considerations need to be made (e.g. adherence to the protocol, sample size, missing data, number of centres). For example, efficient evaluation of AEs may sometimes depend on the skills of study physician within respective study centres. The hierarchical structure of the data generated can also induce correlations since patients are nested within a study centre and for longitudinal outcomes the observations are also nested within a patient. Choice of the most appropriate model to analyse such data may depend on number of centres, number of patients per centre, data missingness and type

of outcomes among others. Presence of missing data and variation in number of patients per centre induces imbalance of observations. Such imbalances are worth consideration during analysis.

Further development in innovative clinical trial designs presents emerging additional challenges towards quantitative drug safety assessment. Several new trial designs have been developed that usually give more room for multiplicity, time to event and correlation adjustment issues e.g. (a) Bayesian adaptive trials that allow modification of a clinical trial after enrolment has already started (b) Umbrella trials that test multiple drugs on a single disease (c) Basket trials that test single drug across multiple diseases. As the case with the traditional parallel trial designs, primary focus when designing the innovative trial is on efficacy outcomes unlike safety outcomes. Consequently, statistical methods developed for analysis of the innovative clinical trial designs are mainly tailored towards the analysis of efficacy outcomes, leaving a gap on methods for quantitative safety assessment.

Some work has been done to improve safety analysis, by suggesting the use of survival methods [36-38] to capture the time factor. However, most of these papers focus on oncology studies and without specific trial phase. Moreover, such previous work did not address multivariate longitudinal safety data issues. Literature has shown existing challenges in the analysis of safety data and the potential role for novel statistical advances to take into account such complexities [3, 73, 76]. This work provides an opportunity to investigate multivariate approach to safety data analysis in the context of

drug safety for malaria treatment or prevention in pregnancy. The multivariate safety data considered includes recurrent AEs and multiple continuous longitudinal outcomes. Paying attention to the statistical issues above in drug safety assessment would provide an assurance to validity and reliability of drug safety evidence developed in clinical trials.

2.3 Pharmaco-epidemiological approach to assessment of AE burden

In preventive drug trials, adverse events can induce treatment non-adherence and discontinuation since the AE burden can outweigh the perceived drug benefit. Understanding the burden of the AEs on treatment non-adherence and discontinuation is useful in developing an informative drug safety profile. However, a challenge exists in quantifying the burden since the AEs that occur before the treatment non-adherence or discontinuation are of interest in such context than those that occur later. In order to obtain comparable sample of participants among those the treatment non-adherent cases and controls, accounting for potential differences in follow-up time, nested case control approach is considered as a useful research tool although the choice of sampling approach can be challenging. Traditionally, nested case control study with risk set sampling is a common approach to cohort sampling that accommodates causal interpretation of the observed effect of an exposure on an outcome [77]. In order to ensure that repeated selection of controls is minimized, other sampling designs have been considered. However, these designs and their respective analytical approaches have received limited attention in the context of drug safety assessment. Traditionally, conditional logistic regression is used in analysis of such nested case control studies. Alternatively, weighted cox model can be considered in order to improve

representativeness of the sampled individuals with respect to the cohort. However, there is limited work comparing the efficiency of these alternatives to the conditional logistic regression under varying nested case control sampling designs. Therefore, there is an opportunity of exploring the utility of nested case control studies approach in assessing impact of AEs on treatment non-adherence or discontinuation.

Chapter 3: Methodology

This chapter describes the data and statistical methods used in addressing the four specific objectives for this thesis. The chapter provides an overview of the study setting, study design, ethical considerations, data generation, data management and statistical analysis approaches. Overall, this chapter highlights how the research work was implemented.

As highlighted earlier, the structure of this thesis was mainly shaped by the first objective that ascertained gaps in the current practice of statistical analysis and reporting safety data in clinical trials (using a systematic review) focussing on antimalarial drugs in pregnancy. There are several methods which can potentially handle these issues but this study narrowed to few specific methods based on the PhD study scope as detailed in the subsequent sections of this chapter.

This work focusses on safety evaluation methods in premarketing drug trials since the low and middle income countries (LMICs) malaria burden is high mainly rely on such data due their poor pharmacovigilance systems. This thesis also focussed on static safety evaluation methods because the motivating data came from trials that were already completed (unlike dynamic safety evaluation methods that focus on detecting safety signals based on partial safety data in an ongoing trial). Furthermore, since the motivating data in this work was also generated from frequentist paradigm, all the statistical methods implemented are based on frequentist approach. Therefore, this work did not consider

Bayesian methods that are more useful especially in context of adaptive trial designs or during interim analyses.

One of the motivating datasets in this thesis came from a multicentre trial (i.e. PREGACT trial). However, the topic of accounting for centre effects was not considered in the current work. The issue of potential centre effects was not considered in this the PhD study, hence is not further explored in the subsequent chapters.

Motivated by the systematic review findings statistical methods for the analysis of recurrent AEs were investigated. Recurrent AEs is a key multivariate outcome encountered in drug safety assessment in clinical trials and its major analytical challenge is on how to account for correlations since the AEs are nested within patients. As highlighted earlier, analysis of AEs is usually descriptive, ignoring censoring, potential time-varying treatment effect and AE dependence. In most cases, mean cumulative function is a popular non-parametric method for analysis of recurrent AEs, although it ignores potential AE-dependence and individual patient heterogeneity. Several other non-parametric, semi-parametric and statistical methods for analysis of recurrent events have been developed and implemented in practice, focussing on efficacy assessment. However, development and application of the recurrent events methods in drug safety assessment setting remains a challenge such that further investigations are needed on their clinical value and applicability [78]. In the framework of benefit-risk assessment, two models (the Andersen Gill model and Prentice, Williams and Peterson model) have been

demonstrated to be useful in providing both direct and indirect effects of treatment on AE recurrence [79]. However, these models provide unbiased estimates if the AEs are uncorrelated and do not efficiently capture both potential time-dependent treatment effects and unobserved heterogeneity. Therefore, most of the commonly used methods are not flexible enough to model the recurrent events in the presence of time-dependent effects since the methods are tailored to be implemented in the context of proportional hazards. Flexible shared frailty models have been developed and seem to be promising in modelling recurrent events even when there are time-dependent treatment effects. However, there is still a gap in understanding their statistical properties compared to the conventional, under a wider range of situations expected in practice. Understanding the properties of the shared frailty models is useful in deciding the most appropriate model to use in practice. However, this can be complicated since the actual distribution of data from clinical trials is usually unknown. The solution to this is using simulation since the true value can be specified in studying the behaviour of the model under varying data generation mechanisms.

Multiple safety outcomes such as concomitant medication and clinical laboratory data are rarely accounted for in determining drug safety. Concomitant medication can potentially interact with the treatment under investigation, eventually affecting the drug safety outcomes including AE occurrence. Clinical laboratory outcomes that are usually longitudinally collected (e.g. bilirubin) also presents useful safety data that are usually highly correlated with the clinical AEs and concomitant medication use. Traditionally, these outcomes are analysed separately, ignoring the correlation of across the outcomes

over the follow-up time. For example, AEs incidence is quantified descriptively using crude incidence rate. Sometimes count models such as Poisson or negative binomial regression are used to quantify AEs incidence in the presence of potential confounders. Similarly, concomitant medication use can be descriptively quantified using crude incidence or can modelled using the Poisson or negative binomial model (if concomitant medication use is recorded as count over the follow-up time). The longitudinal clinical laboratory outcomes can also be separately modelled using linear regression at specified time-points or linear mixed models. Simultaneously modelling the outcomes using joint modelling approach effectively harnesses the rich multiple safety outcomes towards drug safety profile development. Alternative approaches to joint modelling can be incorporating the clinical laboratory safety outcomes as a time-varying covariates. However, the model with time-dependent covariates assumes that the time-dependence is external to the process of AE occurrence. Therefore, the model with time-dependent covariates would not be appropriate since the concomitant medication use and clinical laboratory outcomes are usually correlated with the AE occurrence (i.e. both clinical laboratory outcomes and concomitant medication use are endogenous covariates). Other existing models for simultaneously modelling clinical laboratory data have focussed on extreme value modelling that does not sufficiently account for different types of outcomes and their correlations. Finally, most existing joint models focus on efficacy outcomes. Therefore, this work introduced a novel joint model specifically concentrating on harnessing multiple safety outcomes, including clinical longitudinal laboratory safety outcomes, concomitant medication use and clinical AEs.

Occurrence of AEs can impact clinical trials in many ways. Most clinical trials do not formally assess the impact of the AEs. Probably, limited quantification of the AEs burden can be due to limited statistical tools demonstrating how the impact of AEs can be formally and effectively quantified. This motivated the PhD study to consider nested case control(NCC) approach in analysis of the AEs burden. The NCC approach is helpful especially in handling outcomes that are not pre-specified at design stage e.g. AEs. Therefore, the NCC approach derives a comparable sample of cases and controls.

3.1 Motivating Data sources

The motivating data for this research work came from two clinical trials in pregnancy that were conducted in Malawi. One clinical trial was on evaluating antimalarial drugs for malaria prevention during pregnancy and the other one was on evaluating antimalarial drugs for malaria treatment during pregnancy. It should be noted that AE was the main outcome of interest in both datasets. The PhD study defined AEs as an untoward medical occurrence associated with the use of a drug in humans regardless of whether or not it is considered drug related [10]. This composite endpoint-based definition of AEs was assumed in most of the subsequent analyses, unless otherwise stated. This approach was motivated by the prevailing clinical practice where one of the interest in drug safety assessment focusses on overall incidence of AEs regardless of the type/category. Therefore, this work did not settle for a specific type/category of AEs.

3.1.1 Chloroquine for malaria in pregnancy

The motivating data for addressing objectives 2 and 4 of this PhD study came from an open-label randomised controlled trial conducted in Blantyre, Malawi between February, 2012 and May 2014 (NCT01443130) [80]. The clinical trial recruited 900 pregnant women who were randomly assigned to receive chloroquine-IPTp, Chloroquine chemoprophylaxis or SP-IPTp in 1:1:1 ratio during pregnancy (such that each arm had 300 patients). The inclusion criteria included;

- a) Women in first or second pregnancy before 27th week of gestation
- b) Women attending first antenatal clinic visit and
- c) Women willing to deliver at either the Ndirande Health Centre or Queen Elizabeth Central Hospital in Blantyre, Malawi.
- d) Women willing to remain in the area until 14 weeks after delivery.

The study exclusion criteria included the following;

- a) Women who were HIV positive
- b) Women on any chronic therapy with antimalarial or antifolate drugs,
- c) Women with known high risk pregnancies requiring regular supervision of an obstetrician
- d) Women with known allergies to the study drugs.

Primarily, the clinical trial aimed at evaluating whether chloroquine-IPTp or Chloroquine chemoprophylaxis was better than SP-IPTp in preventing placental malaria. The primary endpoint was histo-pathologically confirmed placental malaria. The procedure of treatment administration was as follows;

- (a) For SP-IPTp two doses of 1500 mg sulfadoxine and 75 mg pyrimethamine were administered at an interval of 4 weeks apart.
- (b) For Chloroquine-IPTp, two doses of 600 mg on day 1, 600 mg on day 2, and 300 mg on day 3 were administered at an interval of 4 weeks apart
- (c) For Chloroquine prophylaxis, participants received 600mg at enrolment followed by 300mg every week until delivery

In this PhD study, we dropped the Chloroquine prophylaxis arm since it had radically different treatment schedule from the IPTp treatment arms. Much as this was a methods development work, handling the extreme radical difference in treatment schedule for the Chloroquine prophylaxis arm compared to the other IPTp treatment arms was not part of this study.

The participants were routinely seen at least once every four weeks until delivery, then at 2, 6, 10 and 14 weeks after delivery. During the follow-up, safety and tolerability assessments were based on unsolicited AEs that were collected from the first day of the treatment administration until the final study visit. The AE grading was based on a standardized protocol of Division of Microbiology and Infectious Diseases (DMID) adult toxicity table of November 2007; graded on a four-point scale as, mild, moderate, severe, or life-threatening. The PhD study focussed on safety and tolerability outcomes. Baseline characteristics that were measured and considered as potential confounders included maternal age in years, body mass index (BMI; calculated as weight in kilograms divided by height in metres squared), gravidity (primigravid or not).

3.1.2 PREGACT trial

The motivating data for addressing objective 3 of this PhD study came from PREGACT trial (ClinicalTrials.gov number, NCT00852423) [81]. PREGACT trial was a multicentre, open-label randomized trial carried out in 4 African countries (Burkina Faso, Ghana, Malawi, and Zambia). For this PhD study, we used data collected in Malawi. In Malawi, the trial was implemented between June 2010 and August 2013 and it enrolled 870 pregnant women in the second or third trimester with falciparum malaria. The participants were randomly allocated in a 1:1:1 ratio to be treated with artemether–lumefantrine (AL), amodiaquine–artesunate (ASAQ) or dihydroartemisinin–diperaquine (DHAPQ). In the trial, polymerase-chain reaction (PCR) adjusted cure rates at day 63 was considered as primary endpoint. The inclusion criteria for the pregnant women was as follows;

- a) Gestational age of at least 16 weeks,
- b) Haemoglobin concentration of at least 7g/dL,
- c) Maternal age of at least 15 years,
- d) Residing within Blantyre district,
- e) Willing to deliver at Queen Elizabeth Central Hospital and adhere to the study requirements
- f) Able to provide consent. For minors the informed consent was additionally obtained from the legal guardian or parent.

The exclusion criteria for the pregnant woman was as follows;

- a) History of any known allergy to the study drugs;
- b) History of any known pregnancy complications
- c) History or presence of illnesses that could influence pregnancy outcomes

- d) Use of cotrimoxazole prophylaxis or ARV treatment
- e) Important illness at enrolment that requires hospitalization, including severe malaria;
- f) Any plans to leave the study area catchment area before delivery.
- g) Prior enrolment in the PREGACT trial or any parallel enrolment in a different trial.
- h) Inability to consume oral medication
- i) Evidence of antimicrobial or antimalarial use one at least week prior to enrolment

In the PREGACT trial, the primary outcome was polymerase-chain reaction (PCR) adjusted cure rates at day 63. As part of safety assessment, safety outcomes were collected including AEs, total bilirubin, alanine aminotransferase, white blood cell and red blood cell counts. The patients were directly observed on days 0-2 (after receiving the dose). The patients were then asked to return to the clinic for follow-up visits on days 3 and 7 and then weekly thereafter until day 63. The safety outcomes were actively collected during the follow-up. The study physician determined the relatedness of the observed AEs with the investigational drug product. For this PhD study, we focus on four outcomes; two classical clinical laboratory safety biomarkers (total bilirubin and alanine aminotransferase), clinical AEs and concomitant medication. The total bilirubin and alanine aminotransferase data was collected at enrolment, day 7, 14, 28 and 63. We dropped creatinine since the actual values were not available. In this PhD study, we focussed on bilirubin and alanine aminotransferase biochemical parameters since they are key biochemical information in antimalarial drug safety assessment during pregnancy

and were also measured in the PREGACT trial. In the current study, we considered the clinical AE count, defined as cumulative number of AEs experienced by the end of follow-up time as a primary outcome of interest. Among the secondary outcomes, we considered bilirubin concentration, alanine aminotransferase and concomitant medication outcome defined as total number of reported concomitant medication count by the end of the follow-up time for each patient. The baseline characteristics considered in this PhD study included trimester, maternal age and gravidity.

3.2 Ethical considerations

Ethical approval for the clinical trial that was evaluating chloroquine-IPTp or Chloroquine chemoprophylaxis as an alternative to SP-IPTp in preventing placental malaria, was obtained from the Institutional Review Board at the University of Maryland, the Research and Ethics Committee at the University of Malawi College Of Medicine, and the Malawi Pharmacy Medicines and Poisons Board [80]. The PREGACT trial received ethical clearance from ethics committee at the Antwerp University Hospital and College of Medicine Research Ethics Committee at the University of Malawi in Malawi [81]. For the current PhD study, ethical approval was also obtained from University of the Witwatersrand's Human Research Ethics Committee. Both data-sets that we used in the PhD study were appropriately de-identified to ensure that the study participants were anonymous.

3.3 Analytical methods

This section outlines the analytical approach for each of the four objectives in the PhD study.

3.3.1 Review of statistical methods for safety data analysis

The first objective of the PhD study involved undertaking a systematic review aimed at establishing the current practice in the statistical analysis of safety data in malaria during malaria chemoprevention trials. From analytical perspective, this objective concentrated on identifying applied statistical methods and their appropriateness in the analysis of safety data in malaria chemoprevention clinical trials during pregnancy, considering the data structure.

3.3.1.1 Search strategy, data extraction and data synthesis

Five online databases were searched, namely; PubMed, Embase, Scopus, Malaria in Pregnancy Library and Cochrane Central Register of Controlled Trials, for original English articles of phase III randomized controlled trial on drugs for malaria chemoprevention during pregnancy. To ensure a wider coverage of the most recent articles we considered articles published from January 2010 to July 2019. Additionally, we considered articles from 2010 since it is the year when the CONSORT guideline updated and emphasized on the need for improved safety data analysis and reporting in clinical trials. We excluded conference proceedings since they do not contain adequate statistical analysis information and they are usually not rigorously reviewed. The key search terms included *malaria, antimalarial drug, pregnancy, safety, tolerability and adverse events*. The key variables that were extracted and entered in pre-constructed excel sheet included main

author, study design, main efficacy outcome, sample size, safety parameters collected (including laboratory data), nature of safety data collection (i.e., passive or active), statistical methods used for respective safety outcomes and how missing safety or efficacy data were handled. We categorized the statistical methods as descriptive or inferential and univariate or multivariate depending on the purpose and nature of the statistical methods, adapted from previous reviews of similar nature [82-84]. The analysis was done in Stata statistical software version 15.1 (StataCorp, College Station, TX, USA) and results were reported descriptively.

3.3.2 Comparison of shared frailty models for recurrent AEs analysis

The recurrence of the AEs induce correlation within a patient, arising due to unobserved heterogeneity and AE-dependence, such that failure to account for it can yield biased results. Recurrent AEs statistical methods such as shared frailty models can address this issue in quantifying drug safety. However, presence of time-dependent treatment effects can lead to biased estimates since it becomes harder to effectively capture both the time-dependent treatment effects and the unobserved heterogeneity, especially when conventional shared frailty models are used. Therefore, the second objective of the PhD study investigated whether lognormal shared frailty with non-proportional hazards restricted cubic splines (i.e. flexible parametric models) improves frailty variance and hazard ratio estimates compared to the conventional inverse Gaussian shared frailty with proportional hazard model and Weibull baseline hazard, in the presence of time-dependent treatment effects and unobserved individual patient heterogeneity. As indicated above, the shared frailty models are useful in accounting for correlation of

recurrent AEs within patients. The performance of the shared frailty models was formally assessed using simulation. *Survsim* package in Stata [85] was used to generate the data. The following three parametric shared frailty models were evaluated; (a) inverse Gaussian shared frailty model with proportional hazards and Weibull baseline hazard (ISF-PH) (b) Lognormal Shared frailty model with non-proportional hazards and restricted cubic splines where the baseline hazard is modelled using 1 degree of freedom and the time-dependent treatment effects is modelled using 1 degree of freedom (LSF-NPHS1) and (c) Lognormal Shared frailty model with non-proportional hazards and restricted cubic spline (LSF-NPHS3) where the baseline hazard is modelled using 3 degrees of freedom and the time-dependent treatment effects is modelled using 1 degree of freedom. The models were also applied to *Chloroquine for Malaria in pregnancy* trial data. All the analyses were done using Stata statistical software version 15.1 (StataCorp, College Station, TX, USA).

3.3.2.1 Notation

Let T_{ik} be the total time of the k^{th} event for the i^{th} patient followed since enrolment time $t=0$, where $i = 1, 2, \dots, n$. We assume that $T_{i1} < T_{i2} < \dots < T_{ik}$ such that the total number of AEs recorded per patient i over a given time $(t) = \sum_{k=0}^{\infty} I(T_{ik} \leq t)$. Frailty is defined as φ_i and interpreted as an unobserved covariate common for all recurrent AEs within the patient. Covariates and corresponding regression coefficients vectors are defined as X_{ik} and β respectively.

3.3.2.2 Shared frailty models

For i th patient and k th recurrent event, the hazard, λ_{ik} , for the shared frailty model can be specified as;

$$\lambda_{ik}(t_{ij}|\varphi_i, X_{ik}) = \varphi_i \lambda_0(t_{ij}) \exp(\mathbf{X}_{ik}^T(t) \boldsymbol{\beta})$$

If the baseline hazard, λ_0 , is unspecified, the model is called a semi-parametric shared frailty.

3.3.2.2.1 ISF-PH model

Firstly, a shared frailty model with Weibull baseline hazard, inverse Gaussian frailty distribution and proportional hazard assumption (also referred as ISF-PH above) was considered. However, in the presence of both unobserved heterogeneity and time-dependent treatment effect, the frailty estimates obtained can be biased.

3.3.2.2.2 Flexible Parametric shared frailty models

Two flexible parametric shared frailty models (LSF-NPHS1 and LSF-NPHS3) were considered where the baseline hazard is captured using specific number of degrees of freedom and proportionality assumption was relaxed to accommodate time-dependent effects. Generally, the non-proportional shared frailty model can be written as;

$$\log \{ \lambda_{ik}(t_{ij} | \varphi_i, X_{ik}) \} = s\{\log(t) | \gamma, k_0\} + \mathbf{X}_{ik}^T \boldsymbol{\beta} + \log(\varphi_i) + \sum_{f=1}^F s\{\log(t) | \delta_f, k_f\} x_{ikf} \quad (\text{M4})$$

The treatment x_{ik} is interacted with restricted cubic function of log time, $s\{\log(t) | \gamma, k_0\}$, such that γ is parameter vector and k_0 is knot vector. Similarly, $s\{\log(t) | \delta_f, k_f\}$ represents a spline function of log time for the f^{th} time-dependent effect with parameter vector δ_f and knots vector k_f . Since the number of degrees of freedom for restricted cubic splines can

determine the most appropriate model to use we considered to evaluate two flexible parametric shared frailty models as follows;

(i) LSF-NPHS1: This shared frailty model was fitted assuming lognormal frailty distribution and non-proportional hazards with restricted cubic spline where the baseline hazard is modelled using 1 degree of freedom and the time-dependent treatment effect is modelled using 1 degree of freedom

(ii) LSF-NPHS3: This shared frailty model was fitted assuming lognormal frailty distribution and non-proportional hazards restricted cubic spline where the baseline hazard is modelled using 3 degrees of freedom and the time-dependent treatment effect is modelled using 1 degree of freedom

3.3.2.2 Simulation study design

The data generating process involved varying sample size ($n=300$, $n=500$), frailty variance values (0.25, 0.50, 0.75) and time-dependent effect (0.03, -0.03, 0.00). The data arose from an assumed trial of binary treatment arms ($x = 0, 1$) where participant could experience up to 4 recurrent AEs. The data generation specified lognormal frailty distribution, Weibull baseline hazard and average fixed log-treatment effect of 0.5 and 1000 simulations for each scenario. In order to minimize errors and check assumptions, quality checks were implemented before and during the analysis of the simulated data. During the data generation process, convergence of the models was continuously monitored. After the data generation, descriptive summary statistics were also done to check for any abnormal observations. For each scenario, graphical exploration of the

distribution of the estimates was done to ascertain the assumed underlying distribution of the generated data.

3.3.2.3 Statistical data analysis

The generated data was analysed using the ISF-PH, LSF-NPHS1 and LSF-NPHS3 models where hazard ratio and frailty variance estimates with their respective standard errors were captured. The performance of the models for the frailty variance estimates was compared using coverage probability of 95% confidence interval, bias, mean square error and percentage gain in precision relative to ISF-PH model. For the hazard ratios, bias, mean square error and percentage gain in precision relative to ISF-PH model were computed.

The models (ISF-PH, LSF-NPHS1 and LSF-NPHS3) were also applied to *Chloroquine for Malaria in pregnancy* trial data in order to investigate the treatment effect on the AE recurrence. During the application of the models to the real data the treatment was adjusted for BMI, gravidity and maternal age.

3.3.3 Joint modelling of safety outcomes

The third objective of the PhD study concentrated on introducing a novel joint model for multivariate longitudinal clinical laboratory safety data, concomitant medication and clinical AEs where AE outcome is of primary interest. Clinically, the proposed model explored how clinical AEs outcome evolve over follow-up time in relation to the longitudinal laboratory safety outcomes and concomitant medication. The joint model is

useful in simultaneously harnessing the available data towards drug safety profile development. PREGACT trial data was used as motivating data in the joint model development and application.

3.3.3.1 Model formulation and statistical data analysis.

3.3.3.1.1 Notation

Let a patient randomized and received at least a dose be denoted as $i = 1, \dots, n$. and

$\mathbf{V}_i = (v_{i1}^T, v_{i2}^T, \dots, v_{nk}^T)$ be a k-variate continuous clinical laboratory outcome vector.

Each of the two continuous outcome vectors (v_{i1}, v_{i2}) are of $(n_{ik} \times 1)$ dimension for the observed longitudinal measurements of the k-th outcome; $v_{ik} = (v_{i1k}, \dots, v_{ink})^T$

The current scenario considers a bivariate scenario of the clinical laboratory where k=1 is alanine amino transferase and k=2 is total bilirubin, although it is straight forward to extend this to generalized k-variate vector of the continuous clinical laboratory outcome.

The continuous clinical laboratory outcomes are considered to be measured at time t_{ijk} ; the current scenario can flexibly accommodate times where t_{ijk} may differ between

individuals and outcomes ($j = 1, \dots, n_{ik}$). Let $\mathbf{C}_i = (c_{i1}^T, c_{i2}^T)$ represent a bivariate count outcome vector where k=1 is for total number of AEs experienced during their follow-up time T_i and k=2 if for total number concomitant medication used over the follow-up time

T_i ; it also straight forward to extend this to generalized k-variate vector of the count outcomes Vector of p baseline covariates measured in the trial for each patient and considered here is denoted as $\mathbf{X}_i = \{X_{1i}, X_{2i}, \dots, X_{pi}\}$. A vector of random effects covariates for each patient is denoted as $\mathbf{Z}_i = \{Z_{1i}, Z_{2i}, \dots, Z_{qi}\}$.

3.3.3.1.2 Sub-models specification

The joint model was formulated within the extended multivariate generalised linear and non-linear mixed effects models framework [86]. The joint model consisted of two count sub-models (for AEs and concomitant medication) and two longitudinal sub-model for the continuous clinical laboratory safety data (for alanine aminotransferase and total bilirubin).

3.3.3.1.2.1 Longitudinal sub-model

The longitudinal sub-model for both continuous clinical laboratory safety data was based on a flexible linear mixed effects model that could accommodate nonlinear changes of the laboratory safety outcomes through use of restricted cubic splines. The generalized form of the longitudinal sub-model was as shown in model M3.1 below;

$$v_{ik}(t) = \mathbf{X}_{ik}^T(t)\boldsymbol{\beta} + \mathbf{Z}_i(t)\mathbf{b}_i + \boldsymbol{\varepsilon}_{ik}(t) \quad \dots\dots\dots \text{M3.1}$$

Where $i = 1, \dots, n$; $k = 1, 2$ and $\boldsymbol{\varepsilon}_{ik}(t)$ is the independent, identically and normally distributed vector of error terms observed at time t for the k th outcome for patient i . The mean for the vector of the error terms is 0. The mean response model $\mathbf{X}_{ik}^T(t)\boldsymbol{\beta}$, was a flexible linear function of the covariates at a given time for the outcome k . The mean response model could flexibly accommodate time-varying covariates. Without loss of generality, time-varying covariates were not considered in this work. The vector of the shared $\mathbf{b}_i$ random effects is assumed to have a multivariate normal distribution with mean

0 and variance Σ . Therefore, the longitudinal submodel for the alanine aminotransferase is as shown in M3.2 below;

$$v_{i1}(t) = \mathbf{X}_{i1}^T(t)\boldsymbol{\beta} + \mathbf{b}_{0i1} \dots\dots\dots\text{M3.2}$$

and longitudinal sub-models for the total bilirubin is as shown in model M3.3 below;

$$v_{i2}(t) = \mathbf{X}_{i2}^T(t)\boldsymbol{\beta} + \mathbf{b}_{0i2} \dots\dots\dots\text{M3.3}$$

Where $\mathbf{b}_{0i1}$ and $\mathbf{b}_{0i2}$ are the patient specific random effects for alanine aminotransferase and total bilirubin respectively.

3.3.3.1.2.2 Concomitant medication counts sub-model

The concomitant count sub-model was formulated as shown in M3.4 below;

$$\log(\varphi|\mathbf{b}_{03i}) = \mathbf{X}_{ik}^T(t)\boldsymbol{\beta} + \mathbf{b}_{03i}\alpha_3\dots\dots\dots\text{M3.4}$$

Where $\log(\varphi|\mathbf{b}_{03i})$ is the logarithm of the Poisson mean response conditional on the patient-specific random intercept concomitant medication count, $\mathbf{X}_{ik}^T(t)\boldsymbol{\beta}$ is mean response for Poisson and α_3 represents log incidence rate ratio per unit increase in the patient-specific deviation from the mean random intercept of the reported concomitant medication use incidence.

3.3.3.1.2.3 AE count sub-model

The sub-model for primary outcome of interest, AE count, in the joint model was specified as shown in M3.5 below;

$$\log(\varphi|\mathbf{b}_{01i}, \mathbf{b}_{02i}, \mathbf{b}_{03i}) = \mathbf{X}_{ik}^T(t)\boldsymbol{\beta} + \mathbf{b}_{01i}\alpha_1 + \mathbf{b}_{02i}\alpha_2 + \mathbf{b}_{03i}\alpha_3 \dots\dots\dots \text{M3.5}$$

Where $\log(\varphi|\mathbf{b}_{01i}, \mathbf{b}_{02i}, \mathbf{b}_{03i})$ is the logarithm of Poisson mean response conditional on the random intercepts for the two longitudinal clinical laboratory outcomes and the concomitant medication count such that the sub-models are linked to the Poisson linear predictor, $\mathbf{X}_{ik}^T(t)\boldsymbol{\beta}$, via the random effects. The AE count sub-model assumes random effects association structure. The $\alpha_1, \alpha_2, \alpha_3$ represent log incidence rate ratio per unit increase in the patient-specific deviation from the mean random intercept of a respective continuous longitudinal clinical laboratory outcome or the concomitant medication count outcome.

The AE count sub-model was further simplified as shown in M3.6 below;

$$\log(\varphi) = \mathbf{X}_{ik}^T(t)\boldsymbol{\beta} + [\mu_1(t)]\alpha_1 + [\mu_2(t)]\alpha_2 + \mathbf{b}_{03i}\alpha_3 \dots\dots\dots \text{M3.6}$$

Where the linear predictor of the Poisson model was linked the expected value of the respective continuous longitudinal clinical laboratory outcome. This formulation is considered more clinically meaningful such that α_1, α_2 represent log incidence rate ratio per unit increase of the respective continuous longitudinal clinical laboratory outcome, at time t. The estimation for the joint model is achieved by maximizing the marginal likelihood of the joint distribution of the observed data and the random effects.

3.3.3.1.3 Statistical data analysis

Descriptive summary of the baseline characteristics for the pregnant women at enrolment and the outcomes were computed with respective standard deviation for the continuous variables and frequencies with respective percentages for the categorical variables. Box plots for the continuous clinical laboratory safety outcomes were plotted to aid visualization of the data including checking for extreme outliers.

Firstly, a conventional Poisson model was fitted where AE count was an outcome of interest. The model adjusted for treatment, maternal age, gravidity and trimester and was used as a basis for comparison of the treatment effect estimates. Then a joint model with a mixture of expected current value and random effects association structure as described above, was fitted. The model included four outcomes; AE count, concomitant medication count, log-transformed alanine aminotransferase and log-trans-formed to total bilirubin. The joint model assessed the association between the AE incidence and other three outcomes over the follow-up time apart from investigating the effect of treatment on the AE incidence. As part of the sensitivity analyses, the normality assumption of the random effects was relaxed by fitting a joint model with t-distributed random effects. Therefore, in order to ascertain whether the normality assumption of the random effects for was met, the estimates from the joint model with t-distributed random effects were compared with those from the joint model with normally-distributed random effects. AIC was used to assess the relative goodness of fit between the joint models with t-distributed random effects and the joint model with normally-distributed random effects.

3.3.4 Role of AEs on antimalarial drug non-adherence

For the fourth objective, the focus was on empirically investigating whether nested case control (NCC) study with incidence density sampling approach is more efficient than NCC with path set sampling under conditional logistic or weighted cox models in assessing the AE burden on treatment non-adherence and participation in preventive antimalarial drug during pregnancy trial. Since the impact of AEs is rarely quantified in most clinical trials and it is hard to pre-specify at design stage, NCC approach is ethically plausible and statistically sound way of deriving and analysing comparable samples. Using the data from the *Chloroquine for malaria in pregnancy* trial, secondary data analysis using NCC approach was carried out. Two outcomes of interest were considered; (i) Treatment non-adherence defined as missing at least one dose after taking the first dose and (ii) study non-completion defined as dropping out of the study before delivery.

3.3.4.1 Sampling approaches

(i) Under NCC with incidence density sampling, a case was defined as experiencing at least the outcome of interest. Similarly, a control was defined as a pregnant woman yet to experience outcome interest but was still at risk to become a case. For the selection of controls with replacement, each case was matched to a control on follow-up time such that each case was matched to four controls.

(ii) Under NCC with path set sampling the case and controls were defined as under NCC with path incidence density sampling. However, the controls were selected in such a way to minimize repeated selection of the patient using path set concept where a path set is

sampled according to the probabilities given the path set representation in the entire risk set at each event time [87].

Under both the NCC with incidence density sampling and NCC with path set sampling, the exposure was considered as having an AE before selection in the risk-set. The following variables were considered as potential confounders; maternal age in years, body mass index (BMI; calculated as weight in kilograms divided by height in metres squared), gravidity (primigravid or not) and, IPTp treatment arm.

3.3.4.2 Statistical methods and data analysis

Firstly, descriptive analysis was carried out where the baseline characteristics were summarized as mean (standard deviation [SD]) or median (inter-quartile range [IQR]) for continuous variables and proportions for categorical variables.

Let $i = (1, 2 \dots n)$ represent a set of women enrolled in the trial and considered in this analysis. Conventionally, estimates in nested case control studies are obtained by maximizing partial likelihood function that can be written as

$$L_{ncc}(\boldsymbol{\beta}) = \prod_{i \in \Omega} \frac{\exp(\boldsymbol{\beta}' X_i)}{\sum_{j \in \hat{R}_i} \exp(\boldsymbol{\beta}' X_j)} \dots\dots\dots (M3.6)$$

The $L_{ncc}(\boldsymbol{\beta})$ is likelihood for nested case control for parameter $\boldsymbol{\beta}$. $\boldsymbol{\beta}$ represents a vector of the regression coefficients, such that $\boldsymbol{\beta}'$ is a transposed vector corresponding to vector(s) of covariates. The Ω represents a set of all the cases. The sampled risk set $\hat{R}_i$ consists of case i and sampled m controls matched to the case on time since enrolment. The vector of covariates considered in the analysis, X_i and X_j are such that individuals i

$\neq j$. The conditional likelihood function under M3.6 above is the one used when estimating parameters under conditional logistic regression and it coincides with partial likelihood function for cox regression. This is why the results from nested case control study using either conditional logistic regression or cox regression are similar in practice. However, under NCC some information can be lost since not all participants are usually considered in the analysed sample. This is problematic especially when the population being studied is heterogeneous in structure. Inverse probability weighting methods offer an opportunity of accommodating all the information of the participants from the sampled population by weighting the likelihood with the inverse of the inclusion probability of being sampled. For cox regression the partial likelihood used to obtain estimate for the regression coefficient is similar to M3.6 above (see M3.7)

$$L_{cox}(\beta) = \prod_{i \in \Omega} \frac{\exp(\beta' X_i)}{\sum_{j \in R_i} \exp(\beta' X_j)} \dots\dots\dots (\text{M3.7})$$

R_i is a set of all patients who have not failed or been censored but are at risk of failure at that time (i.e. all patients who are still under observation just before time t).

For cox regression model the weighted partial likelihood can be written as;

$$L_{wcox}(\beta) = \prod_{i \in \Omega} \frac{\exp(\beta' X_i) \omega_i}{\sum_{j \in S_i} \exp(\beta' X_j) \omega_j} \dots\dots\dots (\text{M3.8})$$

S_i in model M3.8 is the set of cases and controls in the nested case-control sample who are at risk at time t_i . The weights are defined as $\omega_j = 1/p_j$, where p_j is probability of a woman being in the nested case control sample. For each case, the probability of being included in the nested case control sample is 1. In this work Kaplan Meier method of

computing the weights was considered since it estimates the actual sampling of the controls , based on the population under study.

In modelling the impact of AE occurrence on treatment non-adherence and study non-completion, we fitted both the conditional logistic model (conditioned on risk sets) and weighted cox regression to both the nested case control sample with incidence density sampling and then to the nested case control sample with path set sampling. Such analyses were done separately for the treatment non-adherence and the study non-completion outcomes. To ensure that all the models were comparable, in the multivariable model, we included the AE exposure and IPTp treatment adjusted for maternal age. Using the likelihood ratio test, we also assessed whether inclusion of interaction between the AE occurrence and treatment arm modified the observed effect. Assessment of the relative goodness of fit of the models was done using AIC. We reported all the point estimates supported by the 95% confidence interval at 5% significance level. All the statistical analyses and data management were done in two statistical softwares; Stata version 15 (StataCorp, College Station, TX, USA) and R software version 3.6.0 (R Foundation, Vienna, Austria).

Chapter 4: Systematic review of statistical methods for safety data in malaria chemoprevention in pregnancy trials

Noel Patson^{1,2}, Mavuto Mukaka^{3,4}, Kennedy N. Otworldbe^{1, 5}, Lawrence Kazembe⁶, Don P Mathanga², Victor Mwapasa², Alinune N. Kabaghe², Marinus JC Eijkemans⁷, Miriam K Laufer^{*,8}, Tobias Chirwa¹

¹School of Public Health, University of the Witwatersrand, Johannesburg, South Africa

²University of Malawi, College of Medicine, Blantyre, Malawi

³Mahidol Oxford Tropical Medicine Research unit (MORU), Bangkok, Thailand

⁴Centre for Tropical Medicine, Nuffield Department of Medicine, University of Oxford, Oxford, UK

⁵Perinatal HIV Research Unit, Faculty of Health Sciences, University of the Witwatersrand, Johannesburg, South Africa

⁶Department of Biostatistics, University of Namibia, Windhoek, Namibia

⁷Julius Center for Health Sciences and Primary Care, University Medical Center Utrecht, Utrecht University

⁸Center for Vaccine Development and Global Health, University of Maryland, School of Medicine, 685 W. Baltimore St., HSF-1 Room 480, Baltimore, MD 21201, USA

* Corresponding author

Published in the *Malaria Journal*, (2020) 19:119

4.0 Abstract

Background

Drug safety assessments in clinical trials present unique analytical challenges. Some of these include adjusting for individual follow-up time, repeated measurements of multiple outcomes and missing data among others. Furthermore, pre-specifying appropriate analysis becomes difficult as some safety endpoints are unexpected. Although existing guidelines such as CONSORT encourage thorough reporting of adverse events (AEs) in clinical trials, they provide limited details for safety data analysis. The limited guidelines may influence suboptimal analysis by failing to account for some analysis challenges above. A typical example where such challenges exist are trials of anti-malarial drugs for malaria prevention during pregnancy. Lack of proper standardized evaluation of the safety of antimalarial drugs has limited the ability to draw conclusions about safety. Therefore, a systematic review was conducted to establish the current practice in statistical analysis for preventive antimalarial drug safety in pregnancy.

Methods

The search included five databases (PubMed, Embase, Scopus, Malaria in Pregnancy Library and Cochrane Central Register of Controlled Trials) to identify original English articles reporting Phase III randomized controlled trials (RCTs) on anti-malarial drugs for malaria prevention in pregnancy published from January 2010 to July 2019.

Results

Eighteen trials were included in this review that collected multiple longitudinal safety outcomes including AEs. Statistical analysis and reporting of the safety outcomes in all the trials used descriptive statistics; proportions/counts (n=18, 100%) and mean/median

(n=2, 11.1%). Results presentation included tabular (n=16, 88.9%) and text description (n=2, 11.1%). Univariate inferential methods were reported in most trials (n=16, 88.9%); including Chi-square/Fisher`s exact test (n=12, 66.7%), t-test (n=2, 11.1%) and Mann-Whitney/Wilcoxon test (n=1, 5.6%). Multivariable methods, including Poisson and negative binomial were reported in few trials (n=3, 16.7%). Assessment of a potential link between missing efficacy data and safety outcomes was not reported in any of the trials that reported efficacy missing data (n=7, 38.9%).

Conclusion

The review demonstrated that statistical analysis of safety data in anti-malarial drugs for malarial chemoprevention in pregnancy RCTs is inadequate. The analyses insufficiently account for multiple safety outcomes potential dependence, follow-up time and informative missing data which can compromise anti-malarial drug safety evidence development, based on the available data.

Keywords Pregnancy, Malaria, Safety, Prevention, Clinical trials, Statistical methods

4.1 Background

Drug safety assessment in randomized controlled trials (RCTs) is integral in development of comprehensive drug safety profile. Recently, there is enhanced quality guidance on clinical trial reporting of safety outcomes through adherence to the Consolidated Standards of Reporting Trials (CONSORT) guideline [88, 89]. However, there is scant literature on standardized ways to statistically analyse the safety outcomes in clinical trials. Although there exist some general regulatory guidelines on safety data analysis, such as International Conference on Harmonization which recommend descriptive statistical methods supplemented by confidence intervals [90, 91], the proposed statistical methods rarely account for the complexity of the collected safety data, e.g., recurrent adverse events (AEs). Effective solutions to statistical analysis of safety data in clinical trials may need to be tailored to specific indications (set of diseases with similar characteristics) under study since safety data collected are also influenced by the medical condition under study. Absence of standardized guidelines for safety data analysis in specific settings may limit the ability to draw rich conclusions about the safety of the investigational product, based on collected data. Standardized guidelines can simplify integration of safety information from multiple outcomes across RCTs [5] and would ensure optimal use of data in developing the safety profile of the investigational product.

Statistical analysis of safety data in clinical trials is characterized by a challenge of multiple and related endpoints measured over time. The safety endpoints may include clinical and laboratory defined AEs. Laboratory-based AEs are defined based on standard cut-off points for measures such as vital signs (e.g., body temperature), hepato-toxicity

measures (e.g., bilirubin level), cardio-toxicity measures (e.g., electrocardiograms), and other tests relevant to the medical indication being studied [5]. The safety endpoints may be correlated within patients and over time such that failure to account for this in an analysis may yield biased estimates and false inference. Furthermore, time to occurrence of the safety endpoint may be very informative in profiling the drug safety. Such data present statistical analysis and interpretation challenges due to the complexity in structure [92]. For instance, in the case of multiple, repeatedly measured, safety outcomes, false positives may arise from multiple statistical testing if appropriate longitudinal or time to event methods and/or multiplicity adjustments are not considered.

In clinical trials, AEs may impact compliance and study participation which may further affect treatment efficacy estimates [93, 94]. Occurrence of (even mild) AEs due to a drug would lead to non-adherence, leading to informative censoring. The dropping of the patients from the study generates missing data that may lead to biased results if poorly accounted for. Therefore, safety data analysis accounting for missing data is useful to facilitate identification and characterization of the safety profile of the drug as early as possible. Other analysis challenges include lack of adequate ascertainment and classification of AEs, and limited generalizability of results since some AEs cannot be pre-specified at study design stage.

There are many populations where drug safety assessment is complex. One of the special settings in safety data assessment is the use of drugs to prevent adverse outcomes in pregnancy, currently referred to as intermittent preventive treatment of

malaria in pregnancy (IPTp). For example, the World Health Organization recommends that pregnant women receive routine treatment with anti-malarial drugs to clear any malaria infection that is present and also to prevent infection in the weeks after administration [95]. Recent review indicates that methodological issues in studying anti-malarial drugs in pregnancy have prevented firm conclusions on the safety of new anti-malarial drugs in pregnancy [25]. Previous efforts have attempted to standardize safety assessment methodology for anti-malarial drug trials in pregnancy, including study designs and data collection [26, 96]. However, literature remains limited in describing the standard practice in the statistical analysis of safety data that are collected on anti-malarial drugs during pregnancy trials.

The current review focusses on safety assessment in anti-malarial drugs for chemoprevention in pregnancy trials. Since anti-malarial drug for malaria chemoprevention is given repeatedly to healthy pregnant women, it is critical to improve safety assessment in this vulnerable population. Specifically, appropriate statistical analysis of safety outcomes can improve development of anti-malarial drug safety profile. This can be achieved through sufficient use of the data generated during the RCT which provides a comprehensive drug safety insight. This review, therefore, aims at identifying applied statistical methods and their appropriateness in the analysis of safety data in anti-malarial drugs for malaria prevention during pregnancy clinical trials.

4.2 Methods

The systematic review was conducted according to Preferred Reporting Items for Systematic Review and Meta-Analyses (PRISMA) statement [97] which outlines minimum standards for reporting systematic reviews and meta-analysis (see Table S4.1 in appendix 1). The protocol for this review was registered and published with PROSPERO (CRD42019120916). The study population is pregnant women on any anti-malarial drug for malaria chemoprevention.

4.2.1 Search strategy

4.2.1.1 Inclusion criteria

Primary original articles published in English from Phase III RCTs were considered for inclusion. The articles were from RCTs assessing the efficacy and safety of malaria chemoprevention in pregnancy. This review focussed on Phase III RCTs, since they have the largest sample size among pre-marketing trials and accommodates multidisciplinary support in safety evaluation. Further, the data are systematically collected and have the benefit of being randomized, which aids a fair comparison of treatment groups.

4.2.1.2 Exclusion criteria

Observational studies, case reports, letters to the Editor, narrative reviews, systematic reviews and trials in Phase I or Phase II or Phase IV were excluded from this review. This review did not include clinical trials on malaria prevention in pregnancy using intermittent screening and treatment (ISTp) as an alternative to IPTp. ISTp refers to intermittent rapid

diagnostic testing (RDT) for malaria in pregnant women followed by treatment of RDT-positive cases with an effective artemisinin-based combination therapy, and IPTp is given to pregnant women regardless of their malaria status. Hence, ISTp and IPTp consider different populations which may confound the practice in safety assessment methods (i.e., ISTp considers symptomatic population and IPTp considers both symptomatic and asymptomatic population). Non-English publications were excluded.

4.2.1.3 Selection of studies

Studies published between 1 January 2010 and 31 July, 2019 were searched from five databases (PubMed, Embase, Scopus, Malaria in Pregnancy Library (MiPL) and Cochrane Central Register of Controlled Trials (CENTRAL). The MiPL is an excellent scholarly source of articles on malaria in pregnancy that enabled the review to capture both indexed and non-indexed articles beyond the searched databases. Additional search included reference lists of the identified trials and relevant reviews to identify trials potentially missed in the database search. The year 2010 was selected since it is when CONSORT guideline updated and emphasized on appropriate statistical analysis and reporting of clinical trials [98]. Conference proceedings were not included because they usually contain abstracts that do not give detailed analysis of the presented results and they are not rigorously peer-reviewed. The review focussed on published studies only so no experts or abstract publication authors were contacted for unpublished data. The key search items included: *malaria, anti-malarial drug, pregnancy, efficacy safety or tolerability*. The detailed search strategy is presented in Table S4.2 in appendix 1. The search was customized per database. Based on PRISMA procedure, after removing

duplicates, two reviewers (NP and ANK) independently screened titles and abstracts initially before arriving at a final list of eligible articles. Based on the eligible studies list, full text articles were retrieved. The references were managed using Endnote X7.1 (Thomson Reuters). If there were disagreements, the two reviewers discussed the paper to reach a consensus and reasons for exclusion were provided for ineligible publications/studies.

4.2.2 Data extraction

The data extraction file created in Microsoft Excel was used to record all key variables from the selected articles. Some of the collected variables such as mode of safety data collection, participant withdrawal due to AE and handling of continuous measures were based on CONSORT guideline. The following key variables were extracted from the papers: main author, publication date, study design, study location, main efficacy outcome, sample size, list of safety parameters collected (including laboratory data), nature of safety data collection (i.e., passive or active), list of statistical methods used for respective safety outcomes, how the results were presented, retention rate at the end of the follow up and how missing safety or efficacy data were handled. The primary hypothesis type (as superiority, non-inferiority or equivalence) was defined based on what was reported in the actual manuscript or inferred by the lead author (NP), based on how the study framed the primary hypothesis. Superiority hypotheses aim to show whether treatment is better than control, non-inferiority hypotheses intend to show that one treatment is not worse than the other and equivalence hypotheses intend to show that a

given treatment is similar to another for defined characteristics [99]. The statistical methods were classified as descriptive or inferential and univariate or multivariable depending on the purpose and nature of the statistical methods based on previous similar reviews [100, 101], reviewing statistical methods.

4.2.3 Data synthesis

The extracted quantitative data were reported as percentages in tables. The commonly reported safety parameters, suitability of the used statistical methods and other findings were also summarized narratively.

4.3 Results

The search identified 1,103 articles. After removing duplicates, 722 unique articles were identified and considered for possible inclusion in the review. The duplicates (i.e. repeated citations) were the same articles but identified in multiple search databases. Fig. 4.1 presents details of the selection process. During screening, a total of 637 articles were excluded based on relevance of their titles and abstracts. The remaining 85 full text articles were assessed for possible inclusion and 18 articles satisfied the inclusion criteria, and are included in this review as shown in Table 4.1. Reasons for exclusion are shown in Fig. 4.1 below.

Figure 4.1: PRISMA flow diagram profiling studies selection process

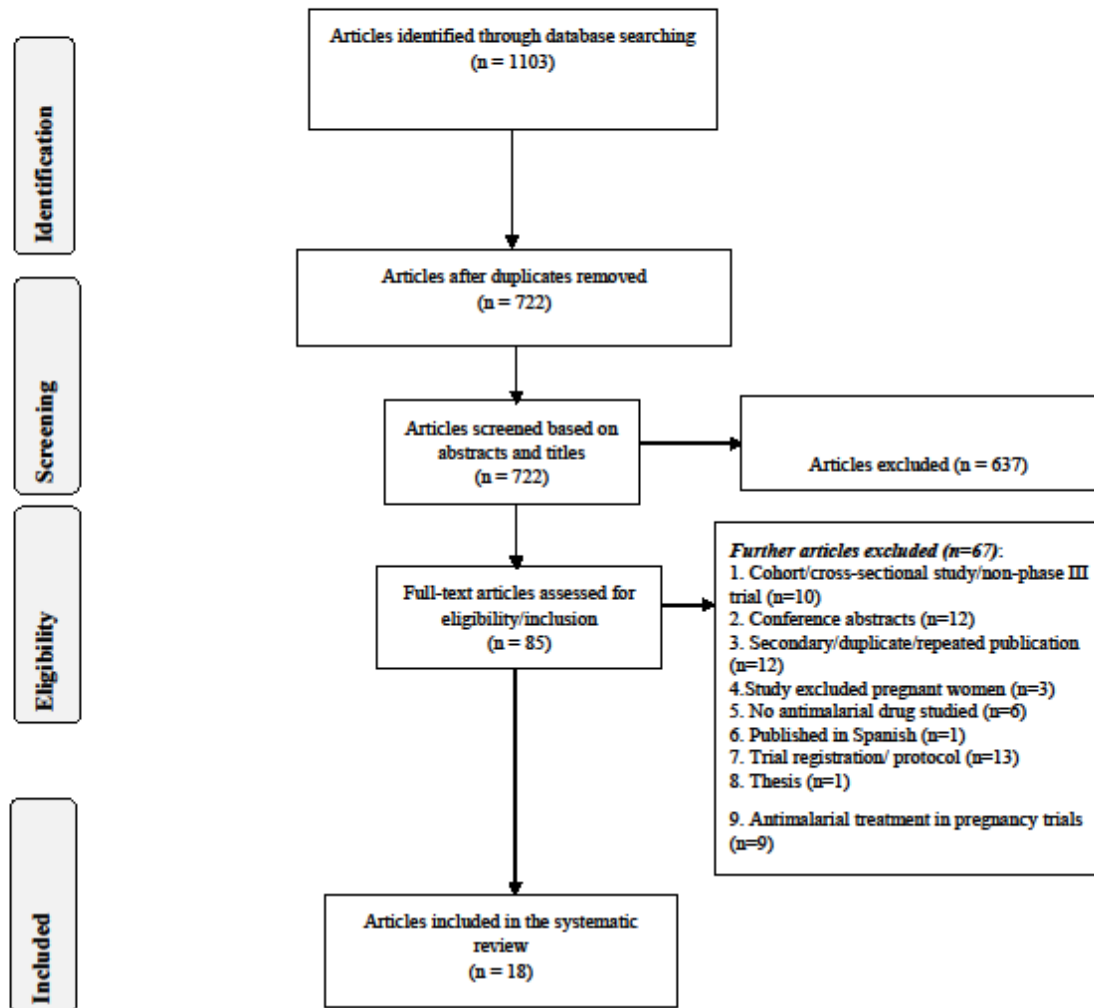


Table 4.1: Overview of randomized clinical trials included in the systematic review

1 st Author, Year, (reference)	Study site	DSMB	Hypothesis	Masking	Comparator drug(s)	Experimental drug(s)
Luntamo. 2010 [102]	Malawi	Yes	Superiority	Single-blinded	two-dose IPTp-SP	AZ-SP
Valea 2010 [103]	Burkina Faso	No	Superiority	Not stated	two-dose IPTp	Monthly SP
Diakite. 2010[104]	Mali	Yes	Superiority	open-label	two-dose IPTp	three IPTp-SP
					IPTp-SP	three IPTp-SP
Ndyomugenyi 2011 [105]	Uganda	Yes	Superiority	double blinded	ITN+placebo	IPTp-SP+ITN
Wini 2013 [106]	Solomon Islands	Yes	Superiority	open-label	IPTp-SP	-CQ prophylaxis
					IPTp-MQ,	CTX(low CD4)
					(IPTp-MQ) and CTX (low CD4)	CTX(high CD4)
Denoeud-Ndam 2014 [107]	Benin	Yes	Noninferiority	open-label		two-dose MQ
González 2014 [108]	Benin, Gabon, Mozambique, and Tanzania	Yes	Superiority	open-label	IPTp-SP	split-dose MQ
González 2014 [109]	Kenya, Mozambique, and Tanzania	Yes	Superiority		CTX and Placebo	CTX and (IPTp-MQ)
Klement 2014 [110]	Togo	Yes	Noninferiority	open-label	IPTp-SP	CTX
Manyando 2014 [111]	Zambia	Yes	Noninferiority	open-label	IPTp-SP	CTX
						IPTp-DP
Desai 2015 [112]	Kenya	No	Superiority	open-label	IPTp-SP	IST-DP
Unger 2015 [113]	Papua New Guinea	Yes	Superiority	single blinded	IPTp-AZCQ	one-dose SP-CQ
						IPTp-DP
Kakuru 2016 [114]	Uganda	No	Superiority	double blinded	IPTp-SP,	IPTp-IPTp-DP
Kimani 2016 [31]	Benin, Kenya, Malawi, Tanzania, and Uganda	Yes	Superiority	open-label	IPTp-SP	IPTp-AZCQ
Natureeba 2017 [115]	Uganda	No	Superiority	double blinded	TMP-SMX	TMP-SMX+ DP
						IPTp-CQ
Divala 2018 [80]	Malawi	Yes	Superiority	open-label	IPTp-SP	CQ prophylaxis
Akinyotu 2018 [116]	Nigeria	No	Superiority	Single-blind	IPTp-SP	IPTp-MQ
Kajubi 2019 [117]	Uganda	No	Superiority	Double-blind	IPTpSP	IPTpDP

AL: Atermeter lumefantrine, CTX: Cotrimoxazole, Q: Quinine, AS: Artesunate, AQ: Amodiaquine, MQ: Mefloquine, CQ: Chloroquine, DP: Dihydroartemisinin-piperaquine, TMP-SMX: Trimethoprim-sulfamethoxazole, PQ: Piperaquine, AZ: Azithromycin

4.3.1 Characteristics of the trials

The trials included in this review were conducted in Oceania (2 trials, 11.1%) and sub-Saharan Africa (16 trials, 88.9%) regions. The 18 RCTs reviewed recruited 26,281 pregnant women with a median sample size of 374 (Interquartile range (IQR): 173, 648) women per treatment arm in a trial. Thirteen trials (72.2%) recruited more than 200 patients per arm. As expected, all trials (18 trials, 100%) computed sample size based on the efficacy outcome(s). The majority of the trials (11 trials, 65.4%) had two treatment arms and the rest had three treatment arms. All 18 trials had an active comparator and IPTp-SP was studied as a standard malaria chemoprevention in the majority of trials (14 trials, 77.8%). Although the review focussed on published trials from 2010 to 2019, the trials were conducted between 2003 and 2017. Based on the primary hypothesis tested, superiority design RCTs were the most common (15 trials, 83.3 %) and the other trials had a non-inferiority hypothesis. Over half of the trials (10 trials, 55.6%) were open-label; one trial did not state the blinding status but the other 7 trials (38.9%) were blinded. Majority of the trials (12 trials, 66.7%) reported that they had a Data Safety and Monitoring Board (DSMB).

4.3.2 Characteristics of the reported safety data

Over half of the trials (10 trials, 55.6%) reported that they collected safety data using a combination of scheduled and non-scheduled visits (Table 4.2), while a third of the trials (6 trials, 33.3%) did not specify the safety data collection approach used. The median retention rate (based on the defined efficacy outcome reported for respective trials) was 89.4% (IQR: 82.5%, 92.4%) and 10 trials (55.6%) had a retention rate below 90%.

Table 4.2: Characteristics and structure of reported safety data (n=18)

Characteristic	Number of articles (%)
Data Monitoring and Safety Board	14 (77.8)
<i>Data collection approach</i>	
Scheduled visits	1 (5.6)
Scheduled and non-scheduled visits	10 (55.6)
Non-scheduled visits	1 (5.6)
Not specified	6 (33.3)
<i>Reported AE severity levels</i>	13 (72.2)
<i>Assessed AE vs adherence</i>	3 (16.7)
<i>Discontinue treatment due to AE</i>	18 (100)
Reported continuous outcome dichotomization	18 (100)
<i>Reported laboratory data</i>	18 (100)
<i>Multiple safety endpoints</i>	18 (100)
<i>Longitudinal safety endpoints</i>	18 (100)
<i>Reported recurrent AEs</i>	0 (0.0)
<i>Retention rate (IQR)^b</i>	89.4% (IQR: 82.5%, 92.4%)

^b IQR = interquartile range.

The retention rate is based on primary outcome reported in trial flow diagram

All the reviewed trials indicated that they had collected multiple longitudinal safety endpoints. As expected, almost all the trials (17 trials, 94.4%) reported obstetric safety outcomes such as foetal loss. Tables S4.3 and S4.4 in appendix 1 provide a detailed list of safety outcomes and respective statistical method reported in each reviewed trial. Despite the reported occurrence of multiple AEs, none of trials seemingly reported recurrence of AEs during pregnancy. In total, 12 trials (66.7%) reported adverse events with different severity levels, e.g., mild, moderate and severe. All trials reported occurrence of AEs by treatment arm. Almost all trials (17 trials, 94.4%) reported laboratory data in their safety assessment of the drug and 16 trials of these (88.9%) dichotomized at least a single continuous safety outcome (e.g., haemoglobin) based on standard cut-off points, to define an AE.

4.3.3 Statistical analysis for safety data

4.3.3.1 Analysis population and missing data

The safety analysis approach (based on treatment allocation and adherence) was specified and reported in 11 trials (61.1%). Per protocol (PP) and intention to treat (ITT) analysis approaches were used in 5 trials (27.8%) and 4 trials (22.2%), respectively. Two trials indicated that they used both PP and ITT to analyse the safety data. Although all the reviewed trials had at least one patient lost to follow-up, only 7 trials (38.9%) reported missing efficacy data and 2 of the 7 trials indicated that the missingness was ignorable after exploring data missingness patterns (Table 4.3). None of the reviewed trials conducted an advanced sensitivity analysis on the relationship between missing data and drug safety. For example, none of the studies assessed the safety outcomes (e.g., AEs) in relation to missing efficacy outcomes. This review found that most trials (16 trials, 88.9%) had at least one participant who experienced an AE leading to discontinuation from the trial although the studies did not formally investigate/quantify the relationship between the AEs and trial completion.

Table 4.3 Statistical approaches for safety data analysis

Statistical method/approach	Number of articles (%)
Descriptive statistics	18 (100)
Proportion/count	18 (100)
Mean/median	1 (5.6)
Incidence rate	6 (33.3)
Inferential methods	
Univariate methods	
Fisher`s exact test/Chi square	12 (66.7)
t-test	4 (22.2)
Linear regression	1 (5.6)
Mann-Whitney/Wilcoxon	1 (5.6)
Multivariable modelling	
Poisson regression	3 (16.7)
Negative binomial	1 (5.6)
Number of inferential methods	
None	3 (16.7)
One method	10 (55.6)
More than one method	5 (27.8)
Analysis approach	
Reporting missing data	
Missing efficacy data reported	7 (38.9)
Handling missing efficacy data	
Imputation	1 (5.6)
Analysis population definition	
ITT	4 (22.2)
PP	5 (27.8)
ITT+PP	2 (11.1)
Not specified	7 (38.9)
Multiplicity adjustment for treatment	0 (0.0)
Results presentation	
Graphs	0 (0.0)
Tables	16 (88.9)
Point estimate and confidence interval	10 (55.6)
P value	16 (88.9)

Note: ITT=Intention to treat; PP=Per protocol;

Entries in this table are mutually exclusive and may not add up to the total 26 articles since some articles used more than one approach

4.3.3.2 Reported statistical methods

All the trials included in the review used descriptive statistics as one of the methods to summarize AEs (Table 4.3). Proportions or counts were the descriptive statistics used in all of the studies to report safety data. Definition of what was safety data depended on respective trials as shown in the additional file (Table S4.2 and Table S4.4 in appendix 1). Incidence rates were reported in 6 trials (33.3%). Most trials (16 trials, 88.9%) reported univariate inferential statistical methods; these included Chi-square or Fisher's exact test (12 trials, 66.7%), t-test (n=4, 22.2%). Only 3 trials (16.7%) reported multivariable statistical methods. The multivariable methods were Poisson regression (n=3, 16.7%), and negative binomial regression (n=1, 5.6%). Usage of at least two inferential statistical methods to compare safety outcomes was reported in 5 trials (27.8%). Although all studies reported multiple safety outcomes, none reported adjustment for multiplicity during analysis.

The review showed that at least a single optimal statistical method was reported in 3 trials (16.7%) that considered multivariable modelling. Even though univariable analysis comparing arms in an RCT were appropriately used, further inferential statistical methods reported in the rest of the trials were suboptimal for the type of data being collected. For further details, additional file under Table S4.4 in appendix 1 provides a detailed list of reported statistical methods with their respective safety outcome(s).

4.3.3.3 Presentation of safety outcomes estimates

In terms of presentation of results, none of the trials presented AEs in a graph. Only 2 trials (11.1%) narratively presented the safety results; the other 16 trials (88.9%)

presented the results in tabular format. A total of 16 trials reported p-values after comparing treatments and there were only 10 trials (55.6%) that reported point estimates with their respective confidence intervals.

4.4 Discussion

This review sought to provide a detailed overview of the actual practice of the statistical analysis of safety data in the unique setting of drug trials for the preventions of malaria in pregnancy as reflected in published literature. The results demonstrate that there is limited reporting of statistical analyses of safety data, at the end of the trial, in these published reports. The findings are useful to advance the development of standardized guidelines in statistical analysis of safety data in anti-malarial drugs in pregnancy trials and related fields. Such guidelines will not replace but rather complement the CONSORT guidelines that are general (i.e., not providing specific statistical methods in analysing harms in RCTs). Based on the authors' knowledge of the available literature, this is the first paper to review statistical methods for safety data in anti-malarial drugs in pregnancy.

Descriptive methods were commonly used to summarize safety data. This review found that each clinical trial used at least one descriptive method to summarize safety data. Univariate statistical methods such as Chi-square or Fishers exact tests were used in two-thirds of the articles reviewed. Such descriptive statistics and univariate statistical inference ignored useful information such as variability in follow-up time, missing data and correlation (for those trials which had their multiple safety outcomes repeatedly measured). Hence there was inefficient data use during analysis that may lead to a loss of useful information for improved and informative conclusions. Although

a third of the reviewed trials attempted to use crude incidence, they failed to adequately account for individual patient follow-up-time and potential confounders.

All trials dichotomized at least a single continuous clinical laboratory safety outcome (i.e., where AE was defined based on standard cut-off points for adult toxicity). Although this aids in providing time-specific drug safety status and easy interpretation, the dichotomized outcome may miss some information on the magnitude of the temporal changes, overtime during the trial. The information loss may lead to reduction in statistical power to detect safety signal if it exists. Valid longitudinal methods (used without restriction on cut-off points) can address the information loss by exploiting potential within-subject correlations for the repeated clinical laboratory measurements [118-120]. Furthermore, the longitudinal methods can provide the basis for developing improved cut-off points tailored to pregnant women in malaria-specific settings. To ensure improved uptake of such methods, future work needs to strive towards making the results from the longitudinal methods feasibly interpretable to the medical practitioners.

Only three studies appropriately used multivariable statistical methods. Adjusting for known prognostic covariates is useful to control for confounding that can be introduced due to imbalance when assessing if treatment is independently associated with safety outcome(s). Of secondary interest, covariate adjustment also preserves type I error [121]. Such adjustment for potential confounders (e.g., age) in safety data analysis are suitable in clinical trials with at least moderate sample size, unlike small sample sizes that lead to unstable estimates. Of specific interest in this review, the Poisson model was more suitable in the context of rare AEs which usually have low event rates [120,

122]. Since Poisson regression assumes a constant rate of occurrence of a rare event, it is not ideal for other multiple transient AEs that were common or recurred and would vary in occurrence overtime [123]. Alternatively, mixed effects models could be considered to characterize the safety events over time since they capture patient-specific effects [124, 125]. Whenever time to AE occurrence information is available, survival analysis models may also be preferred to characterize the time to AE occurrence. For recurrent safety events, that may induce dependence, methods that extend the Cox regression model may be preferred; such models include survival mixed effects models (e.g., frailty models) [91].

Almost half of the reviewed trials did not explicitly define the population on which the safety analysis was based. If per protocol analysis is used to address non-adherence there is potential selection bias since it destroys the balance due to randomization. Although CONSORT recommends ITT, as an alternative for analysis of safety endpoints, non-adherence cannot be explicitly addressed with ITT approach since it ignores dropouts, withdrawal or loss-to-follow up for various reasons including safety concerns; ITT-based inference ignores causal effect of the actual treatment received [126]. Patient withdrawal or dropout due to AEs can induce informative censoring useful in quantifying anti-malarial drug safety. For example, if a patient withdraws due to vomiting after taking an anti-malarial drug, their obstetric efficacy outcomes such as birth weight may appear as missing data. In the context of anti-malarial drug for malaria prevention, even mild AE can lead to drug non-adherence. Since the patient has no disease symptoms, they would judge it less costly for them to discontinue the drug than continue experiencing AEs. Hence, inclusion of information on treatment/trial completion status in relation to anti-malarial safety would enrich development of the

safety profile of anti-malarial drug in pregnancy. Although study completion status, anti-malarial drug safety and missing data may be interlinked, missing data received limited attention such that the few trials that considered efficacy missing data did not explicitly explore the potential link. Studying such complex associations requires statistical methods that can appropriately estimate the pathway from the anti-malarial chemoprevention to study completion. Advantageously, methods based on causal inference framework, such as mediation analysis [127-130] could be adapted/extended to assess the influence of the AEs on non-adherence in RCTs.

Despite about three-quarters of the trials reporting p-values after comparing safety outcomes by treatment arms, only about half of the reviewed trials adhered to International Harmonisation Conference Guideline E9 in reporting of confidence intervals in quantifying the safety effect size [90, 91]. Use of confidence interval aids in interpretation of results by providing a measure of precision. Furthermore, graphical display of safety data to aid in summarizing of safety data was inadequate. Graphs on safety data have a greater ability to convey insight about patterns, trends, or anomalies that may signal potential safety issues compared to presentation of such data in tabular form only [131]. For example, the graphs could help to visualize frequency and changes in AEs over time by treatment arm. The graphs could further help in assessing assumptions for some statistical methods.

Over three-quarters of the reviewed trials were designed as superiority trials based on efficacy outcomes. Although the statistical approach for safety assessment was mainly on superiority hypotheses (for both the superiority and non-superiority trials), clinical and statistical justification of assessing safety based on superiority hypotheses may

be invalid. Superiority hypotheses concentrate on the absence of difference in drug safety effect/risk between or across the treatment arms which may be challenging [99]. For example, when comparing high AE incidences, non-significant difference (when using a superiority hypothesis) would not necessarily translate to a conclusion that a drug is safe and well-tolerated since sometimes all compared treatment arms may have high AE incidence. Perhaps, drug safety evaluation should strive to prove that there is no risk beyond a protocol-defined/hypothesized apriori clinical safety margin (i.e., no excessive safety risk). Based on the findings in this work, researchers are encouraged to consider defining safety margins in safety assessment of anti-malarial drugs. Since safety is mostly a secondary outcome, it is not straightforward on how to define a non-inferiority margin and the appropriate analysis population. Currently, it is still unclear and debatable how to implement this, such that further research is needed [5].

Interestingly, over half of the trials were open-label which may influence physician clinical safety assessment on a patient and patient reporting of AEs based on their expectations since they know the treatment assigned. Appropriate reporting of the AEs would be guided by DSMBs from early stages of the trial. However, availability of DSMBs in over three-quarters of the trials did not translate to improved reporting and analysis. Therefore, DSMB members should advocate for improved analysis approaches for AEs.

Table 4.4 summarizes recommendations to consider on best practices for safety analyses. This provides a general framework for statistical analysis of safety in malaria chemoprevention in pregnancy trials. As highlighted above, the recommendations

assume a context where sample size is moderate or large. For rare events, Bayesian approaches can be considered since they do not depend on asymptotic properties when handling rare events and can incorporate prior/external information [132]. Future research work can further consider adapting/extending recently developed statistical methods for rare disease or small population clinical trials towards analysis of rare safety outcomes in IPTp trials [133-135].

Table 4.4 Recommendations for analyses in malaria chemoprevention in pregnancy trials based on key safety outcomes from the systematic review

Safety outcome type after treatment	Examples	Optimal Statistical method(s)
Time to event	Time to infection/infestation occurrence, time to abortion	Survival methods (e.g. Kaplan Meier plots, log rank test followed by cox regression model)
Longitudinal continuous outcomes	QTc prolongation, haemoglobin, vital signs	Mixed effects linear/nonlinear models
Recurrent events	recurrent hospitalizations, recurrent anaemia, recurrent opportunistic infections	Recurrent event models (e.g. Poisson model, negative binomial, Andersen Gill model)
Safety outcomes with informative missing data,	Treatment switching, Dropout/withdrawal due to AE (i.e. AE-induced non-adherence)	Causal inference methods (e.g. mediation analysis)

This review agrees with other similar publications focusing on drug safety assessment in clinical trials that have noted the need for further improvement in the statistical analysis of the safety data [70, 136]. This review concurs with a recent review that has noted that inappropriate handling of multiple test is prevalent, although their review focussed on four high impact journals, AE in general and a short time of review period [6]. Issues raised in this review include time-dependence of AEs, informative censoring due to discontinuation of treatment because of AEs, safety graphs, and repeated

occurrence of AEs and multivariate longitudinal structure of laboratory data that yields complex correlation. This is an ongoing work whereby further analysis will be explored to address the identified statistical issues above.

The application of the systematic review protocol in gathering the current practice in the context under discussion is more reliable since it exhaustively identified the published anti-malarial drug clinical trials in pregnancy for studied period. However, this review covered only the last decade of publications and may have missed studies published in other languages or that did not appear during literature search. Because the trials reported in the publications spanned for a decade, it was difficult to assess temporal trends. This review represents the most comprehensive review of safety data analysis practice for this important indication.

4.5 Conclusion

Although useful safety data are collected in malaria chemoprevention in pregnancy clinical trials, the analysis remains sub-optimal and this hinders definitive conclusions about drug safety in this setting. Descriptive statistical methods and dichotomization of continuous outcomes are predominant which may lead to loss of useful information. The definition of analysis population and informative presentation of results are not standardized. Overall, the results suggest that choice of a statistical method(s) to use should be jointly driven by the scientific question of interest, epidemiological/clinical plausibility of the method and structure of the raw data. Further work in addressing the highlighted gaps can enhance drug safety decisions and conclusions.

Chapter 5: Comparison of statistical methods for the analysis of recurrent adverse events in the presence of non-proportional hazards and unobserved heterogeneity: A simulation study

Noel Patson^{1,2*}, Mavuto Mukaka^{3,4}, Lawrence Kazembe⁵, Marinus JC Eijkemans⁶, Don Mathanga², Miriam K Laufer⁷, Tobias Chirwa¹

1 School of Public Health, University of the Witwatersrand, Johannesburg, South Africa,

2 School of Public health and Family medicine, College of Medicine, University of Malawi, Blantyre, Malawi,

3 Mahidol Oxford Tropical Medicine Research unit (MORU), Bangkok, Thailand,

4 Centre for Tropical Medicine, Nuffield Department of Medicine, University of Oxford, Oxford, UK

5 Department of Biostatistics, University of Namibia, Windhoek, Namibia,

6 Julius Center for Health Sciences and Primary Care, University Medical Center Utrecht, Utrecht University, Utrecht, The Netherlands

7 Center for Vaccine Development and Global Health, University of Maryland School of Medicine, Baltimore, MD, USA

* Corresponding author

Email: noelpatson@gmail.com

5.0 Abstract

Background

In preventive drug trials such as intermittent preventive treatment for malaria prevention during pregnancy (IPTp), where there is repeated treatment administration, recurrence of adverse events (AEs) is expected. Challenges in modelling the risk of the AEs include accounting for time-to-AE and within-patient correlation, beyond the conventional methods. The correlation comes from two sources, namely; (a) individual patient unobserved heterogeneity (i.e. frailty) and (b) the dependence between AEs characterised by time-dependent treatment effects. Potential within-patient-AE-dependence can be modelled via time-dependent treatment effects, event-specific baseline and event-specific random effect, while heterogeneity can be modelled via subject-specific random effect. Methods that can improve the estimation of both the unobserved heterogeneity and treatment effects can be useful in understanding the evolution of risk of AEs, especially in preventive trials where time-dependent treatment effect is expected.

Methods

Using both simulation study and *Chloroquine for Malaria in Pregnancy* ([NCT01443130](#)) trial data to demonstrate the application of the models, we investigated whether the lognormal shared frailty models with restricted cubic splines and non-proportional hazards (LSF-NPH) assumption can improve estimates for both frailty variance and treatment effect compared to one of the conventional models of inverse Gaussian shared frailty model with proportional hazard (ISF-PH), in the presence of time-dependent treatment effects and unobserved patient heterogeneity. We assessed the bias, precision gain and coverage probability of 95% confidence

interval of the frailty variance estimates for the models under varying known unobserved heterogeneity and sample sizes in the presence of non-proportional hazards for the treatment effects.

Results

The ISF-PH model provided a higher coverage probability of 95% confidence interval, less bias and less precise frailty variance estimates compared to the LSF-NPH models. The LSF-NPH models yielded unbiased hazard ratio estimates at the expense of imprecision and high mean square error compared to the ISF-PH models.

Conclusion

The choice of the shared frailty model for the recurrent AEs analysis should be driven by the study objective. The flexible parametric shared frailty model is appropriate if unbiased hazard ratio estimation is of primary interest in the presence of time-dependent treatment effects. However, ISF-PH model is appropriate when comparing between lognormal Inverse-Gaussian frailty distribution for frailty variance estimation.

Trial registration

ClinicalTrials.gov; [NCT01443130](https://clinicaltrials.gov/ct2/show/study/NCT01443130)

Keywords

Recurrent adverse events, randomised controlled trials, non-proportional hazards, unobserved heterogeneity

5.1 Background

The development of a comprehensive drug safety profile in randomized controlled trials (RCTs) requires statistical methods that can adequately characterize the adverse event (AE) occurrence process (e.g. time of onset, recurrence and duration) [137]. Traditionally, analysis of AEs in IPTp trials, like all clinical trials, is predominantly descriptive and non-parametric such that advanced methods (e.g. parametric and semi-parametric) are rarely used [138-140]. The use of descriptive and non-parametric methods can lead to poor estimates of the effect of treatment on AE recurrence. The descriptive and non-parametric methods may not account for essential aspects of the data structure such as censoring, potential time-varying treatment effect and the AE-dependence. For example, mean cumulative function is the simple and popular non-parametric method used to facilitate the development of trajectories of AEs profiles over the study period, where a patient may experience more than one AEs [46]. However, the method fails to explicitly account for the potential dependence of the AEs or individual patient heterogeneity. The utility of the existing recurrent event methods towards the AEs analysis has been recently considered [79]. However, adopting and interpreting results from such recurrent events methods in drug safety assessment setting remains a challenge. This motivates a need for further investigations on the clinical value and practical applicability of the advanced methods [78]. In the framework of benefit-risk assessment, two models (the Andersen Gill model and Prentice, Williams and Peterson model) have been demonstrated to be useful in providing both direct and indirect effects of treatment on AE recurrence [79]. However, these models yield unbiased estimates if the AEs are uncorrelated and do not efficiently capture (i.e. account for) both potential time-dependent treatment effects and unobserved heterogeneity. This motivates the need to consider using flexible

parametric shared frailty models that can optimally capture both the time-dependent effects and unobserved heterogeneity, to improve drug safety estimates.

In IPTp clinical trials, the potential of recurrent AEs is high because the preventive antimalarial drugs are repeatedly administered. Recurrent AEs may induce within-patient correlation that originates from two sources: the individual patient unobserved heterogeneity and the dependence between events [141]. The individual patient heterogeneity cannot be explained by the observed/known factors [142, 143], while the between-AE-dependence can be explained based on the observed AEs. The unobserved heterogeneity arises from unmeasured information that partly explains the treatment effect. For example, in IPTp trials, unobserved heterogeneity can arise from the unmeasured inherent characteristics (e.g. physiological changes associated with foetus development). The underlying individual patient-specific unobserved heterogeneity can influence the susceptibility of patients to AE recurrence. In IPTp trials, as the case with some drug trials, the treatment effect depends on follow-up time. Such treatment effect that is a function of follow-up time is called time-varying or time-dependent treatment effect. In the presence of the time-dependent treatment effect, the relative effect of treatment changes over the follow-up time such that the proportional assumption is violated (on log hazard scale). Ignoring the time-dependent effects can lead to biased estimates. Therefore, accounting for the potential time-dependent treatment effects in the analysis of recurrent events is necessary. However, in practice, identifying a model that can account for both unobserved heterogeneity and time-dependent effects can present a challenge. Since the recurrence of AEs can be a reflection of both time-dependent treatment effects and underlying individual

patient unobserved heterogeneity. Therefore, it is vital to consider models that account for both issues.

Currently, shared frailty is one of the prevalent models for the analysis of recurrent events. This model can account for both the observed and unobserved heterogeneity. However, the common current applications of the shared frailty model assume that the treatment effect estimate (i.e. hazard ratio) is constant over the follow-up time [144]. In preventive drug trial settings, where there is repeated treatment administration and a long patient follow-up time, the proportional hazard assumption can be invalid when treatment hazard rates, between/across arms, for recurrent events can vary over the follow-up time. For example, in the context of IPTp drug safety assessment, more recurrent AEs are expected close to the day(s) of taking the drugs. The potential presence of time-dependent treatment effects can pose a challenge in applying the shared frailty model to analyse recurrent events since it becomes complicated to separate time-dependent effects from unobserved heterogeneity [145].

Recent developments in flexible parametric survival models can effectively capture both time-dependent effects and unobserved heterogeneity [86]. Researchers have established, mainly in univariate survival data context, that flexible parametric survival models with restricted cubic splines can accurately model time-dependent effects [146, 147]. However, there is limited literature investigating the efficiency and utility of such methods in the context of multivariate survival data and drug safety assessment. Within multivariate survival data setting, Gasparini *et al.* recently compared the standard and spline-based shared frailty models under varying heterogeneity and

baseline hazard specification assuming proportional hazards [148]. The shared parametric models with splines emerged to be more robust to model misspecification than the standard approaches. However, their work is limited to the context of the proportional hazards assumption. As highlighted above, a challenge arises in applying the shared frailty model when there are time-dependent effects. Using the standard semi-parametric shared frailty models, Balan and Putter demonstrated that the bias of frailty variance estimate, in the presence of time-dependent effects, can be mitigated as the cluster size (i.e. number of recurrent events within an individual) increases [145]. Based on the available literature, it is not clear whether the use of flexible parametric shared frailty with splines models can improve the frailty estimates compared to the standard shared frailty models in the presence of both time-dependent treatment and unobserved heterogeneity.

Our work aims at investigating whether the flexible parametric shared frailty model with non-proportional hazards and restricted cubic splines can improve both the frailty variance and hazard ratio estimates compared to the standard parametric shared frailty models, in the presence of both unobserved heterogeneity and time-dependent treatment effects. The optimal shared frailty model can efficiently capture both unobserved heterogeneity and time-dependent effects. We evaluated the following models; (a) inverse Gaussian shared frailty (ISF-PH) model with proportional hazards and Weibull baseline hazard (b) Lognormal shared frailty model with non-proportional hazard and restricted cubic splines where we model the baseline hazard function with 1 degree of freedom and model the time-dependent treatment effect with 1 degree of freedom (LSF-NPHS1) and (c) Lognormal shared frailty model with non-proportional

hazard and restricted cubic splines where we model the baseline hazard function with 3 degrees of freedom and model the time-dependent treatment effect with 1 degree of freedom (LSF-NPHS3). We compare the models using a simulation study assuming an application area of recurrent AEs in IPTp trials. Understanding the statistical properties would enhance the informed choice of statistical methods for the analysis of recurrent AEs. The simulation offers an opportunity to effectively understand the properties of the statistical models since the true data distribution is known, unlike in a practical clinical trial setting. We present the remaining sections of the paper as follows: In the methods section, we provide a detailed review of the models under study, outline the simulation study design and describe the analysis approach. In the results section, we present the findings of our simulation study and apply the models to the analysis of recurrent AEs observed in *Chloroquine for Malaria in pregnancy* trial ([NCT01443130](#)). In the discussion section, we discuss the key findings of our work and provide a conclusion.

5.2 Methods

This section outlines notation and provides an overview of the three models that were compared in our simulation study. Specifically, we briefly introduce and discuss the general inferential methodological framework of the models, focussing on model specification, parameter estimation and respective assumptions. The presentation of this section is customized to the context of drug safety assessment in clinical trials. For readers interested in detailed concepts, on recurrent events analysis, they can consult excellent introductory books on recurrent events analysis and frailty models methodology [143, 149, 150], written for a wider audience.

5.2.1 Notation

We assume that the observation begins at enrolment, denoted as time $t=0$. Let T_{ik} be the total time of the k^{th} event for the i^{th} patient where $i = 1, 2, \dots, n$, $k = 0, 1, 2, \dots, p$ and we assume $T_{i1} < T_{i2} < \dots < T_{ik}$. The AEs are assumed to be recorded over a period of time interval $[0, t]$ such that the total number of AEs recorded per patient i over that time t is $\sum_{k=0}^p \mathbb{I}(T_{ik} \leq t)$, where $\mathbb{I}$ is of sequence of occurrence of AEs. The history of the recorded number of events over given time t , denoted as $N(t), 0 \leq t$, is defined as a counting process; this forms the basis for easy-to-implement recurrent events analysis. The frailty is defined as φ_i and interpreted as an unobserved covariate common for all recurrent AEs within the patient. Covariate and regression coefficients vectors are defined as X_{ik} and β respectively. The covariates are considered to be fixed. Under the simulation study design described below, one covariate is considered (i.e. treatment). For the non-proportional hazard models, the treatment effect is allowed to be dependent on time (i.e. $\beta(t)$). These are further discussed in context in the section below.

5.2.2 Statistical methods for recurrent events

Recurrent events modelling is popularly based on three main approaches; event counts, event intensity function and joint distribution of gap times between successive events. Event counts-based models are suitable when there are frequent events and event occurrence does not impact the occurrence of the subsequent events [149, 151]. In the context of AEs, some mild AEs may not alter the event process (e.g. mild headache). When events are infrequent and the prediction of the next event is of interest, gap times methods are appropriate [152, 153]. To ensure that the event process history is appropriately incorporated, the intensity function is used; the

function defines the instantaneous probability of an event occurring at time t given the event process history [149].

$$\lambda(t|H(t)) = \lim_{\Delta t \rightarrow 0} \frac{\Pr(\Delta N(t)=1|H(t))}{\Delta t} \quad (\text{M5.1})$$

Where $\Delta N(t)$ is the number of events in the time interval $[t, t+\Delta t)$, $H(t)$ is the history of the event process at time t . Since event counts (i.e. Poisson process) modelling approach assumes that the events are independent such that the event history process at t , $H(t)$, does not affect the instantaneous probability of an event at time t , the intensity function is reduced to the rate function $\rho(t)$ below;

$$\lambda(t|H(t)) = \lim_{\Delta t \rightarrow 0} \frac{\Pr(\Delta N(t)=1)}{\Delta t} = \rho(t) \quad (\text{M5.2})$$

This implies; $\Delta t \rho(t) = E(\Delta N(t))$. Hence if we let the expected cumulative rate at time t to be $\mu(t)$, we can derive the expected cumulative rate by integrating the rate function over the interval $[0, t)$

$$\mu(t) = \int_0^t \rho(s) ds \quad (\text{M5.3})$$

The gap/waiting times method is another alternative approach that focusses on modelling time between consecutive events. Methods based on gap time are suitable mainly in settings where the occurrence of the events is considered relatively infrequent [149]. The work in this paper is confined to multiplicative intensity-based models because it permits deriving and interpretation of failure intensity relative to the treatment under investigation. Shared frailty models are considered as important intensity-based models for the analysis of recurrent events that can account for both measured and unmeasured heterogeneity.

5.2.3 The Shared Frailty Models

In practice, the common approach to account for unobserved heterogeneity in recurrent events analysis is to use a robust variance (i.e. robust sandwich covariance matrix is used in estimating the log hazard ratio). Adjusting for within-subject unobserved heterogeneity using robust variance has been shown to be inadequate in some settings [154]. One of the solutions to this is explicitly modelling the correlation using patient-specific random effect. Such a model is defined as a frailty model. Frailty models are hazard models having a multiplicative frailty factor; these are conditional models. Overall, the model has frailty, linear predictor and the baseline hazard function as the multiplicative factors. The random effect introduced in frailty models describes the excess risk that cannot be explained by the measured/observed variables [143]. The unmeasured observations may introduce heterogeneity between the analysis units e.g. for recurrent AEs, observations within a patient may be more similar compared to different patients. Hence frailty model attempts to account for the within-patient correlation. Ignoring the potential heterogeneity in modelling recurrent events can lead to underestimation of treatment effect [155, 156]

For recurrent events, when the frailty is modelled as a common random effect for all events within a patient, it is defined as a shared frailty model (see model M5.4). The dependence accounted for in the shared frailty model is only due to the unobserved individual heterogeneity, φ_i . This measure of the extent of heterogeneity among the patients can follow several distributions [150]. Gamma and inverse Gaussian distributions are the common frailty distribution specifications, possibly due to their mathematical tractability and software availability. The only difference between the gamma and frailty distribution is that gamma frailty yields time-independent

heterogeneity while the inverse Gaussian frailty yields heterogeneity that decays over time. Inverse Gaussian frailty is useful when the modelled outcome involves early occurrence of events during the follow-up. Motivated by this fact, in this simulation study, we considered shared frailty models with inverse Gaussian and lognormal frailty distributions. Lognormal frailty was opted for due to its flexibility to accommodate account for the non-proportional hazards, based on the recent theoretical and software developments. Across all the patients, in all the shared frailty models evaluated, the respective random effects are assumed to be independent and identically distributed.

$$\lambda_{ik}(t_{ij}|\varphi_i, X_{ik}) = \varphi_i \lambda_0(t_{ij}) \exp(\mathbf{X}_{ik}^T(t) \boldsymbol{\beta}) \quad (\text{M5.4})$$

If the baseline hazard, $\lambda_0(t_{ij})$, is unspecified, the model is called a semi-parametric frailty model. Specifying the baseline hazard function ensures that an appropriate shape of the hazard function is captured. However, miss-specification of the baseline hazard distribution may induce bias on regression coefficients [148]. Therefore, caution needs to be exercised on the choice of the baseline hazard distribution by paying attention to the most clinically plausible hazard function. Traditionally, the hazards are assumed proportional over the whole follow-up time. Within the flexible parametric survival methods framework, the proportional hazards assumption can be relaxed to accommodate time-dependent treatment effects (i.e. non-proportional hazards [86]). This can be implemented by interacting treatment with splines of log time as part of the linear predictor. Fitting such flexible non-proportional shared frailty model on log hazard scale is helpful since it can effectively accommodate interpretation and presentation of results. The shared frailty model with non-proportional hazards and restricted cubic splines can be written as;

$$\log \{ \lambda_{ik}(t_{ij}|\varphi_i, X_{ik}) \} = s\{\log(t) | \gamma, k_0\} + \mathbf{X}_{ik}^T \boldsymbol{\beta} + \log(\varphi_i) + \sum_{f=1}^F s\{\log(t) | \delta_f, k_f\} x_{ikf} \quad (\text{M5.5})$$

The treatment x_{ik} is interacted with restricted cubic function of log time, $s\{\log(t) | \gamma, k_0\}$, such that γ is parameter vector and k_0 is knot vector. Similarly, $s\{\log(t) | \delta_f, k_f\}$ represents a spline function of log time for the f^{th} time-dependent effect with parameter vector δ_f and knots vector k_f . Since the number of restricted cubic splines can determine the most appropriate model to use, we considered to evaluate two flexible parametric shared frailty models to be described below. Both models accounted for non-proportional hazard effect of the treatment by interacting the treatment with log time and including them in the linear predictor; therefore 1 degree of freedom was chosen to model the time-dependent treatment effect. The models were defined as follows;

- (i) **LSF-NPHS1:** We considered a shared frailty model assuming lognormal frailty distribution and non-proportional hazards with restricted cubic splines where the baseline hazard function is modelled using 1 degree of freedom and the time-dependent treatment effect is modelled using 1 degree of freedom.
- (ii) **LSF-NPHS3:** We also considered a shared frailty model, fitted assuming lognormal frailty distribution and non-proportional hazards with restricted cubic splines where the baseline hazard function is modelled using 3 degrees of freedom and the time-dependent treatment effect is modelled using 1 degree of freedom.

The models (LSF-NPHS1 and LSF-NPHS3) only differed in number of degrees of freedom used to model the baseline hazard. The more complex the shape of the baseline hazard, the more degrees of freedom are needed to effectively capture the shape of the baseline function. In practice, degrees of freedom ranging from 1 to 5 are

considered as practically sufficient. Therefore, exploring models with 1 and 3 degrees of freedom respectively was considered enough considering the intermediate complexity of time-dependent effect of antimalarial treatment that would be expected in practice (based on personal experience).

Generally, under shared frailty models, depending on the model specification, parameter estimation can be done using expectation-maximization algorithm, maximum likelihood penalized partial likelihood and Markov chain Monte-Carlo methods [150]. Since in this paper we focus on parametric shared frailty models from a Frequentist perspective, the parameter estimation is based on maximum likelihood estimation. It should be noted that not all patients may experience the AE(s), such that others can be censored. Censorship in this paper is considered independent of the AE occurrence. Specifically, the model parameters can be estimated by maximizing marginal log-likelihood as shown below; Let Z be the vector containing the observed patient-specific information such that

$$Z = ((y_{ij}, \xi_{ij}, x_{ij}) \mid j = 0, 1, \dots, k) \dots \dots \dots \text{M5.6}$$

where y_{ij} , ξ_{ij} and x_{ij} are time to AE occurrence, AE occurrence indicator (1 for AE occurrence and 0 otherwise) and fixed covariate for patient i and j th AE occurrence, respectively. Let U be a vector containing the latent information (the frailty terms) for n patients.

$$U = (u_1, u_2, \dots, u_n),$$

therefore, the marginal log-likelihood of the observed data Z can be written as:

$$l_{marg}(\xi, \beta, Z|\tau)$$

$$= \sum_{i=1}^n \left\{ \left[\sum_{j=0}^k \xi_{ij} (\log(h_0 y_{ij})) + (X_{ij}^T \beta) \right] \right. \\ \left. + \log[(-1)^{d_i} (\mathcal{L}^{(d_i)} \left(\sum_{j=0}^k H_0(y_{ij}) \exp(X_{ij}^T \beta) \right)) \right] \\ - \log \left[\mathcal{L} \left(\sum_{j=0}^k H_0(\tau_{ij}) \exp(X_{ij}^T \beta) \right) \right] \right\}$$

.....M5.7

where d_i represents the total number of events for patient and $\mathcal{L}^{(v)}(.)$ represents the v^{th} derivative of the Laplace transform for the derivative. Availability of the Laplace transform simplifies the maximization of the log-likelihood in order to compute the desired model parameters. Under the time-dependent treatment effect, the fixed treatment effect β can be replaced by the time-dependent treatment effect $\beta(t)$.

5.2.4 Simulation Study

This simulation study was conducted and reported based on “Aims Data generating process, Method of analysis, Estimands and Performance” approach since it provides a scientifically coherent structured framework for designing, interpreting and reporting simulation studies.

5.2.4.1 Aim

There is a dearth of literature highlighting the performance of recurrent statistical methods when both AE-dependence and heterogeneity is accounted for in analysis of recurrent AEs in clinical trials. The aim of the simulation study was to evaluate whether the use of Lognormal shared frailty model with restricted splines and non-proportional

hazards improves both the frailty variance and hazard ratio estimates in the presence of time-dependent effects and unobserved heterogeneity under varying unobserved heterogeneity levels and sample sizes. The performance of this flexible model was compared against the standard proportional hazards-based inverse Gaussian shared frailty model with proportional hazard and Weibull baseline hazard distribution assumptions.

5.2.4.2 Data Generating Mechanisms

The data generating mechanism (DGM) in this simulation study focussed on the specification and estimation of frailty variance and hazard ratio in the presence of time-dependent treatment effects and unobserved heterogeneity (Table 1). For each scenario, we considered 1000 simulations, lognormal frailty distribution, Weibull baseline hazard, fixed effect treatment effect of log hazard ratio 0.5 and up to a maximum of 4 AEs experienced by a patient. The choice of possible maximum of 4 AEs per patient was aimed at mimicking a setting for moderately high occurrence of recurrent AEs, in IPTp trials. The choice of lognormal frailty distribution was driven by the mathematical tractability of the distribution, availability of software for implementation and interpretability of the results in the context of AEs. The DGMs varied sample size ($n=300$, $n=500$), frailty distribution variance (0.25, 0.50, 0.75) and time-dependent effect (0.03, -0.03). The choice of the sample size was intended to reflect sample sizes that are commonly employed in most IPTp trials. Choice of frailty variance (i.e. a measure of the strength of the unobserved heterogeneity within an individual) intended to represent a setting 0.25, 0.5 and 0.75 represented weak variance, moderate variance and heavy variance respectively. The time-dependent treatment effect allowed the fixed effect treatment effect of log hazard ratio 0.5 to

decay over time (when set at -0.03) or increase over time (when set at 0.03) or remain constant when set at 0.0). All the DGMs were parametric in order to accommodate the investigation of several scenarios unlike resampling that confines investigation to data with properties only for the application data under study[157]. Hence, we simulated recurrent AE data similar to common scenarios that are practically and theoretically expected in IPTp trials. In all the simulation scenarios, the data generation mimicked a randomized controlled trial with two treatment arms. In all scenarios under study, we assumed that there were no substantial competing risks observed during the trial such that censoring was considered independent. The individual follow-up times were simulated based on uniform distribution and maximum follow-up time of 12 months (i.e. administrative censoring was set at one year), using *Survsim* package in stata [85]. The choice of 12 months of follow-up time was motivated by the fact that most IPTp trials follow-up the mother for 12 months from enrolment to beyond delivery.

The hazard functions were simulated using inversion method, under Weibull baseline hazard function, accommodating both time-varying treatment effect and unobserved heterogeneity with lognormal frailty distribution. The inversion method enabled sampling from a given distribution using uniform distribution and quantile function where uniform distribution was assumed to have minimum 0 and maximum of 1. An overview of the investigated scenarios appears in Table 5.1 below.

During the data generation process, convergence of the models was also continuously monitored. After the data generation, descriptive summary statistics were also produced to check for any abnormal observations. For each scenario, graphical

exploration of the distribution of the estimates was done to ascertain the assumed underlying distribution of the generated data.

5.2.4.3 Simulation study targets

Of primary interest here was the studied shared frailty models' ability to effectively capture the unobserved heterogeneity. Therefore, the primary estimand of interest was log frailty variance φ_i for each of the investigated models. Since our second interest was to assess the ability of the models in capturing time-dependent treatment effect, the log hazard ratio β computed at 3 months of follow-up time was considered as a secondary estimand. We opted for log hazard ratio at the specific follow-up time in order to ensure that the estimates from both the proportional and non-proportional hazard models are comparable.

5.2.4.4 Methods for data analysis

Each simulated data set was analysed using the three methods under investigation: ISF-PH, LSF-NPHS1 and LSF-NPHS3. Based on our simulation scenarios, the ISF-PH model miss-specified the proportionality of the hazards although it correctly specified the baseline hazard function (i.e. the baseline hazard function for the ISF-PH was modelled using Weibull distribution). The LSF-NPHS1 and LSF-NPHS3 models correctly specified the proportionality of the hazards although the baseline hazard function is approximated by the restricted cubic splines. When there are non-proportional hazards, reporting of a single average HR is considered not practically useful. Therefore, in order to ensure that the performance measures for the HR estimates are comparable across the three models, we computed hazard ratio at 3 months of follow-up time since enrolment (i.e. LSF-NPHS1 and LSF-NPHS3 are non-

proportional hazard models and ISF-PH is a proportional hazard model). The choice of the time at which the HR was calculated was motivated by our practical experience where most patients take approximately half of their scheduled IPTp doses by third month of follow-up time.

The three shared frailty models (ISF-PH, LSF-NPHS1 and LSF-NPHS3) are also used in the analysis of recurrent adverse events that were observed in the *Chloroquine for Malaria Prevention in Pregnancy* trial ([NCT01443130](#)) data. The motivating data description, data analysis and results appear in the application section of this paper, after simulation results presentation.

5.2.4.5 Performance measures

In this paper, our model performance measures of interest were defined as described in Morris *et al* [157]. Bias and MSE were main performance measures for our estimands. For frailty variance estimand, we assessed the model performance based on precision, bias, MSE and coverage probability of 95% confidence interval of the frailty variance estimates. For the hazard ratio estimand, we assessed the model performance based on precision, bias and MSE. Point estimates of performance are reported including Monte Carlo standard errors in order to highlight the uncertainty of our point estimates.

We used bias in order to quantify whether the estimator of interest effectively targeted and estimated the true value of interest. Bias is described as the difference between the expected value of an estimator and the true parameter value; $E(\hat{\theta}) - \theta$. Here, θ is the true parameter value, $\hat{\theta}$ as an estimator of the parameter value and $E(\hat{\theta})$ the

expected value of the estimator. Secondly, in order to integrate bias and variance of an estimator into one performance measure, we computed the mean square error (MSE). MSE is described as sum of the squared bias and variance of the estimator; $E(\hat{\theta} - \theta)^2$. In defining relative precision gain, ISF-PH model was considered as a reference model. Therefore, precision gain is described as the inverse squared ratio of the empirical SE of the respective flexible parametric model (i.e. LSF-NPHS1 or LSF-NPHS3) to the empirical SE of the ISF-PH model. For example, relative % increase in precision for LSF-NPHS1 versus ISF-PH would be

$$100\left(\left(\frac{\widehat{EmpSE}_{LSF-NPHS1}}{\widehat{EmpSE}_{ISF-PH}}\right)^2 - 1\right).$$

Additionally, coverage was also computed, defined as proportion of confidence intervals for an estimator that contain the true parameter value (i.e. probability that the confidence interval contains the true parameter value); $\Pr(\hat{\theta}_{low} \leq \theta \leq \hat{\theta}_{upper})$. Here, $\hat{\theta}_{low}$ and $\hat{\theta}_{upper}$ represent the upper and lower confidence interval limits

5.2.4.6 Software and computer hardware specification

The simulation study was conducted using user-written *survsim* and *merlin* packages in Stata version 15.1 [158-160]. The analyses of the simulated and application data were also done using *streg* built-in function in Stata for the ISF-PH model.

The simulation study was performed using the Stata version 15.1 that was installed in a purposely-built Threadripper 3990X GT 1030 workstation, equipped with a Threadripper 3990X (288 MB Cache, 64x Cores, 4.3GHz) turbo processor, MSI TRX40 PRO 10G AMD Ryzen Threadripper Motherboard, Nvidia GeForce RTX 2070

SUPER OC Ed. 8GB GDDR6 GPU, Corsair Vengeance PRO RGB 3200 MHz 64 GB High-Performance Gaming, Hikvision E2000 1 TB M.2 SSD that reads up to 3.5 GB/s and a 4 TB High-Performance Hard Disk Drive (HDD).

5.2.5 Ethical considerations

The *Chloroquine for malaria in pregnancy* trial was conducted in accordance with the Declaration of Helsinki and Good Clinical Practice guidelines. The trial obtained ethical approval from the Institutional Review Board at the University of Maryland, the College of Medicine Research and Ethics Committee at the University of Malawi and the Malawi Pharmacy Medicines and Poisons Board. Ethical approval was also obtained from the University of the Witwatersrand Human Research Ethics Committee, prior to conducting the current study.

5.3 Results

5.3.1 Frailty variance estimates and corresponding bias across the shared frailty models

Based on 1000 simulations for each scenario, we observed a 100% convergence rate of all the models that were fitted. Across all the scenarios, the magnitude of bias for the frailty variance estimates increased as the known unobserved heterogeneity variance increased from 0.25 to 0.75 whether the sample size was 300 or 500 (Table 5.1). The observed bias across all the scenarios was negative suggesting underestimation of the unobserved heterogeneity. The magnitude of bias for the frailty variance estimates were slightly higher when the sample size was 300 compared to the sample size of 500. Overall, the coverage probability of the frailty variance

estimates consistently decreased as the known unobserved heterogeneity variance increased from 0.25 to 0.75.

Table 5.1: Frailty variance estimates bias, coverage probability and MSE across shared frailty models in the presence of unobserved heterogeneity and non-proportional hazards.

Data Generating Mechanisms				ISF-PH Model	LSF-NPHS1 Model	LSF-NPHS3 Model	% gain in precision relative to ISF-PH model	
Scenario ^a	Sample size	Frailty variance	TDE	Bias (MCSE), Coverage (MCSE), MSE (MCSE)	Bias (MCSE), Coverage (MCSE), MSE (MCSE)	Bias (MCSE), Coverage (MCSE), MSE (MCSE)	LSF-NPHS1 % gain in precision (MCSE)	LSF-NPHS3 % gain in precision (MCSE)
1	500	0.25	0.03	-0.0389 (0.0064), 96.40 (0.5891), 0.0423 (0.0021)	-0.0556 (0.0060), 95.90 (0.6270), 0.0386 (0.0020)	-0.0263 (0.0061), 96.20 (0.6046), 0.0374 (0.0019)	14.6985 (0.7269)	11.1644 (1.6768)
2	500	0.50	0.03	-0.0431 (0.0051), 95.00 (0.6892), 0.0278 (0.0012)	-0.1329 (0.0045), 87.10 (1.0600), 0.0375 (0.0015)	-0.1098 (0.0045), 90.20 (0.9402), 0.0323 (0.0014)	30.9883 (0.9472)	27.8901 (1.8680)
3	500	0.75	0.03	-0.0689 (0.0048), 92.80 (0.8174), 0.0280 (0.0013)	-0.2151 (0.0041), 61.80 (1.5365), 0.0628 (0.0020)	-0.1939 (0.0041), 69.10 (1.4612), 0.0544 (0.0019)	40.7088 (1.0645)	38.4618 (1.9530)
4	300	0.25	0.03	-0.0363 (0.0079), 97.20 (0.5217), 0.0637 (0.0029)	-0.0542 (0.0073), 97.50 (0.4937), 0.0569 (0.0027)	-0.0248 (0.0075), 96.70 (0.5649), 0.0561 (0.0027)	12.5407 (1.8701)	12.5407 (1.8701)
5	300	0.50	0.03	-0.0557 (0.0069), 94.90 (0.6957), 0.0504 (0.0025)	-0.1448 (0.0061), 90.60 (0.9228), 0.0575 (0.0027)	-0.1221 (0.0061), 92.50 (0.8329), 0.0518 (0.0025)	29.3473 (0.9645)	28.3377 (1.8182)
6	300	0.75	0.03	-0.0674 (0.0062), 94.50 (0.7209), 0.0431 (0.0020)	-0.2141 (0.0052), 76.60 (1.3388), 0.0729 (0.0026)	-0.1950 (0.0053), 80.90 (1.2431), 0.0660 (0.0025)	42.0668 (1.1767)	37.7659 (2.0100)
7	500	0.25	-0.03	-0.0212 (0.0063), 96.10 (0.6122), 0.0405 (0.0019)	-0.0694 (0.0060), 95.20 (0.6760), 0.0413 (0.0021)	-0.0348 (0.0061), 95.90 (0.6270), 0.0384 (0.0020)	9.7913 (0.6757)	7.8448 (1.5502)
8	500	0.50	-0.03	-0.0317 (0.0051), 95.30 (0.6693), 0.0268 (0.0012)	-0.1425 (0.0045), 85.60 (1.1102), 0.0405 (0.0016)	-0.1163 (0.0045), 89.00 (0.9894), 0.0341 (0.0014)	27.3902 (0.9049)	25.1537 (1.7685)
9	500	0.75	-0.03	-0.0605 (0.0048), 93.30 (0.7906), 0.0269 (0.0013)	-0.2236 (0.0041), 59.20 (1.5541), 0.0668 (0.0021)	-0.1998 (0.0041), 67.70 (1.4788), 0.0569 (0.0019)	38.1980 (1.0547)	37.1299 (1.8588)
10	300	0.25	-0.03	-0.0190 (0.0079), 97.00 (0.5394), 0.0621 (0.0027)	-0.0680 (0.0075), 97.70 (0.4740), 0.0602 (0.0028)	-0.0335 (0.0075), 97.00 (0.5394), 0.0577 (0.002)	11.1739 (0.7791)	9.1266 (1.7207)
11	300	0.50	-0.03	-0.0454 (0.0069), 94.80 (0.7021), 0.0493 (0.0024)	-0.1549 (0.0061), 89.50 (0.9694), 0.0614 (0.0029)	-0.1286 (0.0061), 92.50 (0.8329), 0.0537 (0.0026)	26.3505 (0.9108)	26.8512 (1.7308)
12	300	0.75	-0.03	-0.0588 (0.0062), 94.80 (0.7021), 0.0418 (0.0019)	-0.2219 (0.0053), 75.70 (1.3563), 0.0768 (0.0027)	-0.2005 (0.0053), 79.90 (1.2673), 0.0684 (0.0026)	39.0878 (1.1318)	36.0944 (1.9042)
13	500	0.25	0	-0.0296 (0.0064), 96.50 (0.5812), 0.0412 (0.0020)	-0.0619 (0.0060), 95.70 (0.6415), 0.0398 (0.0020)	-0.0300 (0.0061), 95.90 (0.6270), 0.0378 (0.0019)	12.1870 (0.6812)	9.4918 (1.6201)
14	500	0.50	0	-0.0373 (0.0051), 95.00 (0.6892), 0.0272 (0.0012)	-0.1373 (0.0045), 86.80 (1.0704), 0.0389 (0.0016)	-0.1127 (0.0045), 89.70 (0.9612), 0.0331 (0.0014)	29.0623 (0.8985)	26.5314 (1.8048)
15	500	0.75	0	-0.0644 (0.0048), 93.00 (0.8068), 0.0274 (0.0013)	-0.2188 (0.0041), 61.30 (1.5402), 0.0646 (0.0020)	-0.1964 (0.0041), 68.80 (1.4651), 0.0554 (0.0019)	39.4216 (1.0400)	37.9288 (1.8940)
16	300	0.25	0	-0.0273 (0.0079), 97.10 (0.5307), 0.0627 (0.0028)	-0.0605 (0.0074), 97.70 (0.4740), 0.0583 (0.0028)	-0.0283 (0.0075), 96.70 (0.5649), 0.0567 (0.0027)	13.2530 (0.7841)	10.8368 (1.7814)
17	300	0.50	0	-0.0500 (0.0069), 94.80 (0.7021), 0.0495 (0.0024)	-0.1491 (0.0061), 90.20 (0.9402), 0.0591 (0.0028)	-0.1249 (0.0061), 92.60 (0.8278), 0.0526 (0.0026)	27.7522 (0.9164)	27.0827 (1.7637)
18	300	0.75	0	-0.0627 (0.0062), 94.50 (0.7209), 0.0423 (0.0020)	-0.2174 (0.0052), 76.20 (1.3467), 0.0745 (0.0027)	-0.1972 (0.0053), 80.60 (1.250), 0.0668 (0.0026)	40.6654 (1.1293)	37.0383 (1.9578)

TDE=time dependent effect, a; each scenario is based on 1000 simulations

As expected, overall, among the non-proportional hazard models with restricted cubic splines, the LSF-NPHS3 had lower bias and higher coverage for the 95% CI of the frailty variance estimates compared to the LSF-NPHS1. This result suggests that for the non-proportional hazard models with restricted cubic splines, increasing the number of degrees of freedom modelling the baseline hazard function from 1 to 3 improves the frailty variance estimates. For the scenarios where the known unobserved heterogeneity was 0.50 or 0.75, we observed a higher magnitude of bias of frailty variance estimates for the non-proportional hazard models with restricted cubic splines models compared to the ISF-PH model. For the scenarios where the known unobserved heterogeneity variance was 0.25, we observed the lowest magnitude of bias for frailty variance estimates in LSF-NPHS3 model followed by ISF-PH model then LSF-NPHS1 model such that LSF-NPHS1 model yielded the highest magnitude of bias for frailty variance estimates. For example, for scenario 1, frailty variance estimate bias (MCSE) was -0.0389 (0.0064), -0.0556 (0.0060), -0.0263 (0.0061) for ISF-PH, LSF-NPHS1 and LSF-NPHS3 models respectively. Pertaining to coverage, although the frailty variance estimates coverage decreased with increase unobserved heterogeneity variance, in most of the scenarios, decreasing the sample size from 500 to 300 yielded results with higher coverage: the decreased frailty variance estimates with increased unobserved heterogeneity was more drastic among the non-proportional hazards with restricted cubic splines models (LSF-NPHS1 or LSF-NPHS3) compared to the ISF-PH model. This is because the non-proportional hazards model cannot separate the time-dependent treatment effect from unobserved heterogeneity although increasing the number of degrees of freedom modelling the baseline hazard would improve the situation. The ISF-PH model

consistently yielded higher coverage probability for frailty variance estimates compared to the non-proportional hazards with restricted cubic splines models (LSF-NPHS1 or LSF-NPHS3). For example, for scenario 1, coverage(MCSE) was 96.40 (0.5891), 95.90 (0.6270) and 96.20 (0.6046) for ISF-PH, LSF-NPHS1 and LSF-NPHS3 models respectively. These results suggest the robustness of the ISF-PH even when the model frailty distribution is misspecified compared to the non-proportional hazard models in estimating the frailty. This is expected since the non-proportional hazards are more tailored to improving the estimation of the time-dependent treatment effect rather than the frailty variance.

Overall, the MSE for the ISF-PH model frailty variance estimates were lower compared to the LSF-NPHS1 and LSF-NPHS3 models, except when the known unobserved heterogeneity variance was 0.25. Across all the 18 scenarios studied, we observed that the MSE estimates for the ISF-PH model decreased as the unobserved heterogeneity variance increased. However, the MSE estimates for the LSF-NPHS1 and LSF-NPHS3 models increased as the unobserved heterogeneity variance increased: Notably, the MSE for the LSF-NPHS3 model yielded lower MSE estimates compared to the LSF-NPHS1 model. As expected, increasing the sample size from 300 to 500 reduced the MSE estimates for the corresponding scenarios.

Our investigation showed that LSF-NPHS1 and LSF-NPHS3 models were more efficient than the ISF-PH model as demonstrated by the lower mean of model-based standard errors (Figure 5.1). The difference in model-based standard errors between that LSF-

NPHS1 and LSF-NPHS3 models was negligible. As expected, increasing sample size from 300 to 500 improved the efficiency of the shared frailty models under study such that the model-based standard errors for the frailty variance estimates decreased.

Figure 5.1: Average model standard errors(SE) for the frailty variance estimates across shared frailty models under varying data-generating mechanisms



5.3.2 Percentage gain in precision for frailty variance estimates of non-proportional hazards shared frailty models with restricted cubic splines relative to ISF-PH model

Overall, as the unobserved heterogeneity increased, percentage gain in precision for frailty variance estimates of non-proportional hazards shared frailty models with restricted cubic splines relative ISF-PH model also increased (Table 5.1). The gain in precision was higher when time-dependent treatment effects were increasing (i.e. $TDE=0.03$) compared to the decaying time-dependent treatment effects ($TDE=-0.03$). Across the models, the gain in precision was higher in the LSF-NPHS1 model compared to LSF-NPHS3 model. The observed gain in precision for LSF-NPHS1 and LSF-NPHS3 models compared to ISF-PH model under proportional hazard scenarios (i.e. $TDE=0$) was similar to the scenarios when time-dependent treatment effects were increasing (i.e. $TDE=0.03$).

5.3.3 Bias and MSE for the estimated log hazard ratio estimates across the shared frailty models under varying unobserved heterogeneity and sample size

Overall, across all the three models under study as shown in Table 5.2, the MSE for the HR estimates increased as the unobserved heterogeneity variance increased. The ISF-PH model consistently yielded lowest MSE estimates for the HRs across the three models regardless of the nature of the time-dependent treatment effect. Increasing the sample size from 300 to 500 also led to reduction in the estimated MSE for the HRs. We observed a loss in precision of hazard ratio estimates for the ISF-PH model compared to the LSF-NPHS1 and LSF-NPHS3 model. The precision loss ranged from 9% to 20%. The highest precision loss was observed when the unobserved heterogeneity was low (i.e. frailty

variance equal to 0.25). The precision loss estimates were similar for both LSF-NPHS1 and LSF-NPHS3 models. There was low precision loss in the absence of time-dependent treatment effects ($TDE=0$). In the presence of time-dependent treatment effects (i.e. $TDE=0.03$ and $TDE=-0.03$), we observed higher magnitude of bias in the ISF-PH model compared to the LSF-NPHS1 and LSF-NPHS3 models. In the absence of the time-dependent treatment effects, the bias for the HR estimates was lower in the ISF-PH model compared to the LSF-NPHS1 and LSF-NPHS3 models. However, in the absence of the time-dependent treatment effects, the bias across all the three models was relatively lower (i.e. under 0.01) compared to scenarios when time-dependent effects are present.

Table 5.2: Bias and MSE for log hazard ratio estimates, estimated at 3 months, across shared frailty models in the presence of both unobserved heterogeneity and non-proportional hazards

Data Generating Mechanisms				ISF-PH Model	LSF-NPHS1 Model	LSF-NPHS3 Model	% gain in precision relative to ISF-PH model	
Scenario ^a	Sample size	Frailty variance	TDE	Bias (MCSE), MSE (MCSE)	Bias (MCSE), MSE (MCSE)	Bias (MCSE), MSE (MCSE)	LSF-NPHS1 % gain in precision (MCSE)	LSF-NPHS3 % gain in precision (MCSE)
1	500	0.25	0.03	-0.0284 (0.0021), 0.0050 (0.0002)	-0.0033 (0.0023), 0.0051 (0.0002)	0.0021 (0.0023), 0.0052 (0.0002)	-17.4231 (2.1208)	-18.7145 (2.0875)
2	500	0.50	0.03	-0.0285 (0.0024), 0.0068 (0.0003)	-0.0023 (0.0026), 0.0067 (0.0003)	0.0031 (0.0026), 0.0068 (0.0003)	-11.1578 (1.9985)	-12.3537 (1.9639)
3	500	0.75	0.03	-0.0233 (0.0027), 0.0076 (0.0004)	0.0035 (0.0029), 0.0081 (0.0004)	0.0093 (0.0029), 0.0083 (0.0004)	-12.7599 (1.9452)	-14.1747 (1.8756)
4	300	0.25	0.03	-0.0259 (0.0028), 0.0085 (0.0004)	-0.0015 (0.0030), 0.0093 (0.0004)	0.0041 (0.0031), 0.0096 (0.0004)	-15.4391 (2.0856)	-18.0137 (2.0061)
5	300	0.50	0.03	-0.0215 (0.0032), 0.0105 (0.0005)	0.0047 (0.0033), 0.0111 (0.0005)	0.0103 (0.0034), 0.0114 (0.0005)	-9.4606 (2.1449)	-11.4682 (2.0577)
6	300	0.75	0.03	-0.0252 (0.0035), 0.0126 (0.0006)	0.0019 (0.0037), 0.0134 (0.0006)	0.0069 (0.0037), 0.0137 (0.0006)	-10.3721 (1.9615)	-12.1647 (1.8949)
7	500	0.25	-0.03	0.0287 (0.0021), 0.0051 (0.0002)	-0.0075 (0.0023), 0.0053 (0.0002)	0.0018 (0.0023), 0.0053 (0.0002)	-18.3401 (2.1438)	-19.5151 (2.1016)
8	500	0.50	-0.03	0.0289 (0.0024), 0.0068 (0.0003)	-0.0072 (0.0026), 0.0069 (0.0003)	0.0021 (0.0026), 0.0069 (0.0003)	-12.1807 (2.0392)	-13.6435 (1.9931)
9	500	0.75	-0.03	0.0341 (0.0027), 0.0083 (0.0004)	-0.0011 (0.0029), 0.0083 (0.0004)	0.0086 (0.0029), 0.0085 (0.0004)	-14.1771 (1.9844)	-15.7796 (1.9005)
10	300	0.25	-0.03	0.0311 (0.0028), 0.0088 (0.0004)	-0.0059 (0.0031), 0.0095 (0.0004)	0.0035 (0.0031), 0.0098 (0.0004)	-16.7676 (2.1182)	-19.2818 (2.0272)
11	300	0.50	-0.03	0.0361 (0.0032), 0.0114 (0.0005)	0.0004 (0.0033), 0.0112 (0.0005)	0.0096 (0.0034), 0.0116 (0.0005)	-10.0998 (2.1977)	-12.4175 (2.0889)
12	300	0.75	-0.03	0.0316 (0.0035), 0.0130 (0.0006)	-0.0032 (0.0037), 0.0136 (0.0006)	0.0052 (0.0037), 0.0139 (0.0007)	-11.6808 (2.0013)	-13.7017 (1.9235)
13	500	0.25	0	-0.0009 (0.0021), 0.0042 (0.0002)	-0.0051 (0.0023), 0.0052 (0.0002)	0.0020 (0.0023), 0.0053 (0.0002)	-17.8005 (2.1303)	-19.0529 (2.0932)
14	500	0.50	0	-0.0009 (0.0024), 0.0059 (0.0003)	-0.0044 (0.0026), 0.0067 (0.0003)	0.0027 (0.0026), 0.0068 (0.0003)	-11.4262 (2.0238)	-12.6627 (1.9860)
15	500	0.75	0	0.0044 (0.0027), 0.0071 (0.0003)	0.0016 (0.0029), 0.0082 (0.0004)	0.0091 (0.0029), 0.0084 (0.0004)	-13.4164 (1.9621)	-14.9156 (1.8853)
16	300	0.25	0	0.0017 (0.0028), 0.0079 (0.0004)	-0.0032 (0.0031), 0.0094 (0.0004)	0.0041 (0.0031), 0.0097 (0.0004)	-15.8830 (2.0985)	-18.3944 (2.0120)
17	300	0.50	0	0.0061 (0.0032), 0.0101 (0.0005)	0.0027 (0.0033), 0.0112 (0.0005)	0.0097 (0.0034), 0.0115 (0.0005)	-9.7878 (2.1647)	-11.9642 (2.0672)
18	300	0.75	0	0.0021 (0.0035), 0.0120 (0.0006)	-0.0004 (0.0037), 0.0134 (0.0006)	0.0062 (0.0037), 0.0138 (0.0006)	-11.0009 (1.9807)	-12.9017 (1.9108)

TDE=time dependent effect, a; each scenario is based on 1000 simulations

5.3.4 Example data application: *Chloroquine for malaria in pregnancy* trial

We illustrated the application of the three investigated shared frailty models above (ISF-PH, LSF-NPHS1 and LSF-NPHS3) to data on recurrent AEs collected from a clinical trial evaluating antimalarial drugs to prevent malaria in pregnancy. The trial is called *Chloroquine for malaria in pregnancy*.

The current modelling focussed on investigating the effect of treatment on recurrent AEs. We included a population of 600 pregnant women in their first or second trimesters who were randomized to receive IPTp of (a) sulfadoxine-pyrimethamine (SP, two doses, four weeks apart), (b) chloroquine (CQ, 600 mg on day 1, 600 mg on day 2, and 300 mg on day 3 two days apart). Although the original trial had three treatment arms, in the current analysis, we dropped the arm of CQ prophylaxis (to focus on IPTp arms) since its treatment schedule was radically different from the IPTp arms. Full details of the trial are published elsewhere [161].

Of the 600 pregnant women analysed in the current analysis, 474 (79%) experienced at least one AE. Cumulatively, more AEs were experienced in the CQ arm (895 AEs) than those in the SP arm (650 AEs). We fitted ISF-PH model followed by the LSF-NPHS1 and LSF-NPHS3 models where IPTp treatment was an exposure of interest. As part of sensitivity analysis to further explore how the data best fitted our models, we further considered two additional models lognormal shared frailty models, modelling the baseline hazard function with 2 (LSF-NPHS2) and 4 (LSF-NPHS4) degrees respectively. The SP

treatment was considered as a reference treatment in the interpretation of the results. ISF-PH assumed the baseline hazard function followed Weibull distribution, unobserved heterogeneity followed inverse Gaussian distribution and the hazard rates between the CQ and SP treatments were proportional over the follow-up time. The non-proportional hazard with restricted cubic splines models (LSF-NPHS1 and LSF-NPHS3) assumed that unobserved heterogeneity followed lognormal distribution and time-dependent effects could be modelled using one degree of freedom. NPHS1 and LSF-NPHS3 approximated the baseline hazard function using 1 and 3 degrees of freedom respectively. We fitted adjusted and unadjusted model for all the three models under investigation. The adjusted model included the following covariates: treatment, maternal age, maternal body mass index and gravidity (i.e. whether the pregnant woman had previous pregnancy or not). We adjusted for the same covariates for the three models to ensure that the estimates were comparable. Since the results from the adjusted and unadjusted models were similar, our results reported in Table 5.3 below are based on the unadjusted models since we considered them more parsimonious. The reported hazard ratios across the three models are based on marginal hazard ratio, predicted hazard ratio at 14 days and 60 days of follow-up. Additionally, we plotted the hazard ratio estimates over the follow-up time.

Table 5.3: Comparing sharing frailty models for analysis of recurrent AEs among pregnant women on IPTp treatment in Malawi

Estimate	ISF-PH	LSF-NPHS1	LSF-NPHS2 ^a	LSF-NPHS3	LSF-NPHS4 ^b
Log HR at 14 days (SE)	0.2281 (0.0499)	0.4758 (0.3572)	0.4343 (0.3724)	0.4648 (0.4033)	0.4404 (0.3710)
log HR at 60 days (SE)	0.2281 (0.0499)	0.0236 (0.1240)	0.2011 (0.2562)	0.2251 (0.2831)	0.1837 (0.2441)
Marginal log HR (SE)	0.2281 (0.0499)	0.2328 (0.0517)	0.3090 (0.0769)	0.3360 (0.0850)	0.3024 (0.0746)
Frailty variance	0.0324	0.0476	0.4591	0.6440	0.3977
AIC	6562.57	19922.40	19658.30	19651.20	19570.70

a: Lognormal shared frailty model with non-proportional hazard and restricted cubic splines where the baseline hazard is modelled with 2 degrees of freedom and the time-dependent treatment effects is modelled with 1 degree of freedom

b: Lognormal shared frailty model with non-proportional hazard and restricted cubic splines where the baseline hazard is modelled with 4 degrees of freedom and the time-dependent treatment effects is modelled with 1 degree of freedom

In this clinical example for recurrent AEs data analysis, the estimated marginal log hazard ratios(SE) were 0.2281 (0.0499), 0.2328 (0.0517), 0.3090 (0.0769), 0.3360 (0.0850) and 0.3024(0.0746) for ISF-PH, LSF-NPHS1, LSF-NPHS2, LSF-NPHS3 and LSF-NPHS4 models respectively; the results indicate an increased risk of AEs among the women who took the CQ treatment arm compared to those in the SP treatment arm (Table 5.3). At 14 days, the Log HR across the non-proportional hazard models were similar ranging between 0.43 to 0.47 and were higher compared to the ISF-PH model. At 60 days, LSF-NPHS2 and LSF-NPHS4 yielded similar log HR estimates; approximately 0.20 and considerably lower compared to the ISF-PH model estimates. The frailty variance estimates were generally higher among the non-proportional hazard models compared to the ISF-PH model. The unobserved heterogeneity in this clinical example was low within patient as reflected by ISF-PH model low frailty variance estimate in Table 5.3: based on the results from the simulation study above ISF-PH estimates the frailty variance better than the lognormal non-proportional hazard. Therefore, in this trials potential variation of treatment effect due to some unmeasured information would not be considered as a

critical issue in interpretation of the IPTp safety. Among the non-proportional hazards models with restricted cubic splines, the AIC estimates show that the model fit was good when the baseline hazard function was approximated with minimum of 2 degrees.

5.4 Discussion

The analysis of recurrent AEs in clinical trials rarely account for potential unobserved heterogeneity and time-dependent effects [138]. Using a simulation study, we investigated the performance of three statistical methods for recurrent events that can be adapted towards the analysis of recurrent AEs in clinical trials, in the presence of both time-dependent effects and unobserved heterogeneity. Our paper mainly focused on evaluating the ability of the flexible parametric shared frailty models with non-proportional hazards and restricted cubic splines to capture both the unobserved heterogeneity and time-dependent treatment effects. We established that the inverse Gaussian shared frailty with proportional hazard model consistently yielded higher coverage, lower bias and lower MSE of the frailty variance estimates at the expense of imprecision compared to the flexible shared frailty models with non-proportional hazards and restricted cubic splines. Interestingly, the frailty variance estimates coverage and bias, for the shared frailty models with non-proportional hazards and restricted cubic splines, improved upon increasing the number of degrees of freedom used to model the baseline hazard function: The simulation was confined to investigating 1 and 3 degrees of freedom although the clinical example further explored models with 2 and 4 degrees of freedom. As expected, the flexible parametric shared frailty model with non-proportional hazards and restricted cubic splines yielded better hazard ratio estimates, reflecting the time-dependent

treatment effects, compared to the inverse Gaussian shared frailty with proportional hazard model. The flexible parametric shared frailty models with non-proportional hazards and restricted cubic splines yielded unbiased estimates of HRs at the expense of imprecision and inaccuracy (i.e. high MSE) compared to the ISF-PH model. Therefore, our findings show that the more precise estimators were biased and imprecise estimators were unbiased. In drug safety assessment using methods that yield an unbiased hazard ratio would be of primary interest such that the flexible parametric shared frailty model should be considered if there is the presence of both unobserved heterogeneity and time-dependent effects. However, such a decision should be made at the cost of imprecision and loss of accuracy. Hence, we emphasize that the flexible shared frailty with non-proportional hazards and restricted cubic splines should be applied only when the proportional hazards assumption is violated. If interest is in estimating frailty variance, the inverse Gaussian model with proportional hazards can be considered sufficient even in the presence of both time-dependent treatment effect and unobserved heterogeneity

The inverse Gaussian shared frailty with proportional hazard probably effectively captured the unobserved heterogeneity because we did not mis-specify the baseline hazard (i.e. Weibull distribution). Secondly, the longer follow-up period of 12 months made the population more homogeneous with time since the unobserved heterogeneity for inverse Gaussian frailty decays over the follow-up time[162]. Thirdly, the inverse Gaussian frailty is practically close to the lognormal frailty where the simulated data arose. The structure of the inverse Gaussian frailty would be considered more appropriate in the context of recurrent AEs in antimalarial drug trials where there are more frequent AEs early in the

clinical trial [150]. Unfortunately, the conditional hazard ratio derived from the inverse Gaussian shared frailty with proportional hazard is constant over the follow-up time which does not reflect the non-proportionality of the hazards in the presence of time-dependent treatment effect.

The interpretation of our results is confined to scenarios where the censoring is assumed to be non-informative. In the presence of informative censoring, it will be interesting to understand how the models perform. Failure to account for time to medication presents a limitation in all the models considered in this work. Another key limitation of our study is relying on the assumption of constant (i.e. fixed) unobserved heterogeneity that did not vary with time. Future work can consider evaluating the performance of the shared frailty models accounting for both actual time to medication and time to AE occurrence under complicated scenarios where there is both time-varying treatment effects and time-varying unobserved heterogeneity

Generally, in drug clinical trials, understanding drug safety profile requires efficient modelling of the recurrent AEs including accounting for both unobserved heterogeneity and any non-linear effects of the investigated drug. The methods studied in this paper provide key knowledge on choosing the most appropriate model for analysing recurrent AE data presenting the unobserved heterogeneity and time-dependent effects issues. The flexible parametric shared frailty models accounting for both time-dependent effects and unobserved heterogeneity are directly useful in IPTp trials since physiological

changes that occur during pregnancy are mostly unobserved and can contribute to the heterogeneity of the treatment effect.

Based on our findings, we provide some practical recommendations on the studied shared frailty models to analyse the recurrent AEs when the hazards are non-proportional and there is potential unobserved heterogeneity. Overall, the choice of the most appropriate model should be driven by the objective of the analysis. For instance, if interest is mainly in estimating hazard ratios, a shared frailty model with non-proportional hazards and restricted cubic splines is appropriate. When interest is in obtaining the unobserved heterogeneity estimates, inverse Gaussian with proportional hazard can serve the purpose. In the absence of time-dependent effects, using a shared frailty model with non-proportional hazards and restricted cubic splines should be avoided since it leads to biased estimates due to model overfitting.

**Chapter 6: Joint modelling of multivariate longitudinal clinical laboratory
safety outcomes, concomitant medication and clinical adverse events:**

Application to artemisinin-based treatment during pregnancy clinical trial

Noel Patson^{1,2*}, Mavuto Mukaka^{3,4}, Umberto D'Alessandro⁵, Gertrude Chapotera² Victor
Mwapasa², Don Mathanga², Lawrence Kazembe⁶, Miriam K Laufer⁷, Tobias Chirwa¹

1 School of Public Health, University of the Witwatersrand, Johannesburg, South Africa,

2 School of Public health and Family medicine, College of Medicine, University of Malawi,
Blantyre, Malawi,

3 Mahidol Oxford Tropical Medicine Research unit (MORU), Bangkok, Thailand,

4 Centre for Tropical Medicine, Nuffield Department of Medicine, University of Oxford,
Oxford, UK

5 Medical Research Council Unit, The Gambia at the London School of Hygiene and
Tropical Medicine, Fajara, Gambia

6 University of Namibia, Department of Biostatistics, Windhoek, Namibia,

7 Center for Vaccine Development and Global Health, University of Maryland, School of
Medicine, Baltimore, MD, USA

* Corresponding author

Email: noelpatson@gmail.com

Published in *BMC Medical Research Methodology* 2021, **21**(1): 208

6.0 Abstract

Background

In drug trials, clinical adverse events (AEs), concomitant medication and laboratory safety outcomes are repeatedly collected to support drug safety evidence. Despite the potential correlation of these outcomes, they are typically analysed separately, potentially leading to misinformation and inefficient estimates due to partial assessment of safety data. Using joint modelling, we investigated whether clinical AEs vary by treatment and how laboratory outcomes (alanine amino-transferase, total bilirubin) and concomitant medication are associated with clinical AEs over time following artemisinin-based antimalarial therapy.

Methods

We used data from a trial of artemisinin-based treatments for malaria during pregnancy that randomized 870 women to receive artemether–lumefantrine (AL), amodiaquine–artesunate (ASAQ) and dihydroartemisinin–piperaquine (DHAPQ). We fitted a joint model containing four sub-models from four outcomes: longitudinal sub-model for alanine aminotransferase, longitudinal sub-model for total bilirubin, Poisson sub-model for concomitant medication and Poisson sub-model for clinical AEs. Since the clinical AEs was our primary outcome, the longitudinal sub-models and concomitant medication sub-model were linked to the clinical AEs sub-model via current value and random effects association structures respectively. We fitted a conventional Poisson model for clinical AEs to assess if the effect of treatment on clinical AEs (i.e. incidence rate ratio (IRR)) estimates differed between the conventional Poisson and the joint models, where AL was

reference treatment. Global test assessed differences in treatment effect across the three arms.

Results

Out of the 870 women, 564 (65%) experienced at least one AE. Using joint model, AEs were associated with the concomitant medication (log IRR=1.7487; 95% CI: 1.5471, 1.9503; $p<0.001$) but not the total bilirubin (log IRR=-0.0288; 95% CI: -0.5045, 0.4469; $p=0.906$) and alanine aminotransferase (log IRR=0.1153; 95% CI: -0.0889, 0.3194; $p=0.269$). Global test showed no treatment effect differences on AE incidence across the three arms in the Poisson ($p=0.149$) and joint ($p=0.376$) models. The Poisson model underestimated the effects of treatment on AE incidence such that log IRR for ASAQ was 0.2118 (95% CI: 0.0082, 0.4154) for joint model compared to 0.1838 (95% CI: 0.0574, 0.3102) for Poisson model.

Conclusion

Although the AEs did not vary across the treatments, the joint model yielded efficient estimates compared to the Poisson model. The joint model showed a positive relationship between the AEs and concomitant medication but not with laboratory outcomes.

Trial registration

ClinicalTrials.gov: NCT00852423.

Keywords: Adverse events, randomised controlled trials, joint model, drug safety, concomitant medication

6.1 Background

There has been limited formal analysis of safety data being conducted in drug clinical trials compared to efficacy data. During drug clinical trials, scheduled and non-scheduled visits are conducted where multiple safety outcomes are collected including longitudinal clinical laboratory outcomes and clinical AEs [1, 5]. Data on concomitant medications defined as any medication/supplement taken by the trial participant other than the investigational product, are also repeatedly collected in order to support drug safety evidence. The repeatedly measured clinical laboratory outcomes are considered as objective and proxy indicators for drug-induced human body organ dysfunction and tend to be strongly associated with incidence of clinical AEs. For example, classic liver function biomarkers such as total bilirubin and alanine aminotransferase are usually used in assessment of drug safety [163]. However, each of the clinical laboratory safety outcomes is usually analysed separately based on summary statistics such as means or proportion of patients with elevated values for the safety markers [45, 164]. The clinical AE data is also usually analysed separately using the descriptive and non-parametric methods (e.g. cumulative incidence function and crude incidence rate) [46, 138]. Although ignored in most safety analyses, concomitant medication taken during the trial period can also influence the magnitude of drug safety estimates since the concomitant medication can interact with the treatment under investigation. Concomitant medication is usually also analysed separately (e.g. using mean cumulative function) [165]. The standard approaches that analyse each of the clinical laboratory safety data, clinical AEs and concomitant medication separately are inefficient and may yield biased estimates of AE incidence rate (i.e. clinical AEs occurrence over follow-up time) since they ignore the

correlation structure of the multiple outcomes over the follow-up time, leading to information loss. Furthermore, the separate analyses present challenges in interpreting overall drug safety profile. Integrated approach in analysis of the clinical laboratory safety data, concomitant medication and clinical AEs through the use of joint models, can offer an opportunity to efficiently harness the available information towards enriching the drug safety estimates. The joint models can efficiently quantify how the incidence rate of AEs is associated with the multivariate longitudinal process of the multiple continuous laboratory-based measurements, accounting for baseline covariates, concomitant medications and unobserved heterogeneity that is not captured in the conventional analytical approaches of safety data, all of which can potentially confound the estimates.

Recently, there have been increased proposals for improved analysis of multivariate clinical laboratory data, concomitant medications and clinical AEs. Schildrout et al. and Rosenkranz. discuss the utility of longitudinal modelling of clinical laboratory data using estimating equations and maximum likelihood. However, such separate longitudinal models ignore correlation between the multiple safety outcomes. Southworth and Heffernan proposed joint modelling approach that captures the joint behaviour of clinical laboratory safety data elevated values indicative of potential liver damage [61]. However, this was limited as the model did not use all the available information due to focus on extreme values and, further, such modelling is mainly suited for large sample sizes. Merz et al. also recommended accounting for both multivariate longitudinal clinical laboratory data and clinical AEs, although they do not explicitly discuss how this can be practically achieved. Barker [165] demonstrates how mean cumulative function can be used as an

exploratory tool in analysis of number of concomitant medications taken over the follow-up time in order to support drug safety evidence interpretation. Building on these recent developments, this paper explores whether jointly modelling multivariate longitudinal clinical laboratory, concomitant medication and clinical AEs adjusted for baseline characteristics can be efficient in estimating AE incidence rate and useful in understanding joint evolution of the potentially correlated multiple safety outcomes.

In clinical trials with frequent AE occurrence (e.g. antimalarial treatment in pregnancy trials), event count models such as Poisson or negative binomial regression are considered important in quantifying the AE rate [149, 166-168]. However, the AEs are also highly correlated with the cumulative concomitant medications and clinical laboratory safety outcomes. Joint modelling can offer an opportunity to explicitly model the correlations and account for the follow-up time, ensuring efficient and unbiased estimates. Despite the advanced development and evident benefits of joint modelling of multivariate longitudinal data[169-171], joint modelling has received limited attention in drug safety assessment. Recent statistical software developments have provided an opportunity for extending joint models to more different types of outcomes , beyond the traditionally implemented joint models with single continuous longitudinal data and single survival outcome. In this paper, we present a joint model for longitudinal continuous clinical laboratory safety data (alanine amino transferase and total bilirubin), individual clinical AEs and concomitant medication. We focus on demonstrating the utility of the proposed model in assessing the association between the clinical AEs in relation to these outcomes controlling for baseline covariates. Furthermore, we empirically assessed whether the

joint model improved the estimates in log incidence rate ratio (IRR) on effect of treatment on clinical AEs compared to the conventional Poisson model. The proposed joint model is useful in providing evidence-based decisions in determining the most important clinical safety measurements to focus on for drug safety assessment especially in resource-constrained settings.

6.2 Methods

6.2.1 Data

Our proposed joint model was applied to data from PREGACT trial as the motivating data. PREGACT trial (ClinicalTrials.gov number, NCT00852423) was a multicentre, open-label randomized trial carried out in 4 African countries (Burkina Faso, Ghana, Malawi, and Zambia). The current work uses the data from Malawi. In Malawi, the trial was implemented between June 2010 and August 2013 and enrolled 870 pregnant women during their second or third trimester with falciparum malaria. The women were randomly allocated in a 1:1:1 ratio to be treated with artemether–lumefantrine (AL), amodiaquine–artesunate (ASAQ) or dihydroartemisinin–dipepaquine (DHAPQ). The primary outcome was polymerase-chain reaction (PCR) adjusted cure rates at day 63.

The safety outcomes collected at baseline and during follow-up were AEs, total bilirubin, alanine aminotransferase, white blood cell and red blood cell counts. The patients were directly observed on days 0-2 (after receiving the dose). The patients were then asked to return to the clinic for follow-up visits on days 3 and 7 and then weekly thereafter until day 63. For this analysis, we focus on four outcomes; two clinical laboratory safety biomarkers

(total bilirubin and alanine aminotransferase), clinical AEs and concomitant medication. The total bilirubin and alanine aminotransferase data was collected at enrolment, day 7, 14, 28 and 63. We focussed on bilirubin and alanine aminotransferase biochemical parameters since they are key biochemical parameters in antimalarial drug safety assessment during pregnancy and were also measured in the PREGACT trial.

The clinical AE count, defined as cumulate number of AEs experienced by the end of follow-up time is the primary outcome of interest in the current analysis. In all the subsequent discussions, the clinical AEs that were defined as *definitely not related* (by the study physician) to the antimalarial drug treatment are not considered in developing the joint model to avoid spurious results. The concomitant medication outcome is also defined as cumulative number of reported concomitant medication use by the end of the follow up time for each patient. In the current study, we considered the reported concomitant medication use regardless of its intended use.

6.2.2 Ethical considerations

The PREGACT trial received ethical clearance from ethics committee at the Antwerp University Hospital in Belgium and College of Medicine Research Ethics Committee from the University of Malawi [81]. Prior to enrolment, informed consent was also sought from the mother. Ethical approval was also obtained from University of the Witwatersrand Human Research Ethics Committee, prior to access and utilization of the data for the current analysis.

6.2.3 Notation and Model Specifications

Let each patient who was randomized and received at least a dose be denoted as $i = 1, \dots, n$. Let $V_i = (v_{i1}^T, v_{i2}^T)$ be a bivariate continuous clinical laboratory outcome vector. Specifically, in our context we consider a bivariate scenario; $k=1$ is alanine amino transferase and $k=2$ is total bilirubin. Each of the two continuous outcome vectors (v_{i1}, v_{i2}) are of $(n_{ik} \times 1)$ dimension for the observed longitudinal measurements of the k -th outcome; $v_{ik} = (v_{i1k}, \dots, v_{in_k})^T$. We accommodate the situations where observation times, t_{ijk} may differ between individuals and outcomes ($j = 1, \dots, n_{ik}$) (i.e. formulation of the joint models to be discussed below is on continuous scale rather than a discrete scale in order to even accommodate values obtained from non-scheduled visits). For each patient, we let $C_i = (c_{i1}^T, c_{i2}^T)$ represent a bivariate count outcome vector where $k=1$ is for total number of AEs experienced during their follow-up time T_i and $k=2$ for total number concomitant medication used over the follow-up time T_i . Without loss of generality, the concomitant medication outcome is also referred to as outcome variable V_3 when defining the sub-models of the joint model. A set of covariates that were collected at baseline for each individual are defined as $X_i = \{X_{1i}, X_{2i}, \dots, X_{pi}\}$.

6.2.3.1 Conventional Poisson model for clinical AEs

In drug trials, where multiple safety outcomes are repeatedly collected to support safety evidence, clinical AEs are usually of primary interest. The occurrence of the clinical AEs is typically correlated with other safety-related outcomes such as alanine

aminotransferase, total bilirubin and concomitant medication. Modelling these outcomes separately is inefficient since it insufficiently accounts for the correlations. Traditionally, modelling of clinical AEs count is done using the conventional Poisson model, where the Poisson mean response (of clinical AEs in this context), φ , is assumed to be equal to the variance. Since the clinical AEs count is non-negative the logarithm of φ can naturally be linked with the baseline variables such that the conventional Poisson model for the clinical AE count can be presented as;

$$\log(\varphi) = \mathbf{X}_{ik}^T \boldsymbol{\beta} \quad \text{M6.1}$$

Where φ represents Poisson mean response and $\boldsymbol{\beta}$ is a vector of coefficients (i.e. log IRR) for the corresponding vector of baseline covariates, $\mathbf{X}_{ik}^T$, that included maternal age, gravidity, trimester at enrolment and treatment arm. The model M6.1 has log link of the mean response of clinical AEs count and assumes a Poisson distribution of the clinical AEs. The coefficients of the covariates are interpreted on logarithm scale as expected change in the log of mean clinical AEs count per unit change in the covariate. Exponentiating the coefficients (i.e. log IRR) yields IRR. However, the estimated effect of treatment on clinical AEs from model M6.1 may be insufficient since it does not account for other factors that can confound the estimates e.g. concomitant medication and clinical laboratory outcomes. Alternatively, one can consider incorporating this additional information (of concomitant medication and clinical laboratory outcomes) as part of covariates in model M6.1. This can yield model M6.2 below.

$$\log(\varphi) = \mathbf{X}_{ik}^T(\mathbf{t}) \boldsymbol{\beta} \quad \text{M6.2}$$

Since the concomitant medication and clinical laboratory outcomes (alanine aminotransferase, total bilirubin) are added in the model as time-varying covariates (with time t_0 , representing baseline covariates in M6.1), the vector of covariates adjusted in model M6.2 is $\mathbf{X}_{ik}^T(t)$ denoting the vector of covariates value at time t . However, both model M6.1 and M6.2 do not account for correlation overtime across the longitudinal alanine aminotransferase, longitudinal total bilirubin, concomitant medication and clinical AEs outcomes. In this context of multiple outcomes that are correlated overtime, joint modelling offers an opportunity to obtain improved and efficient estimates since it can efficiently account for the potential correlations and baseline covariates. The joint model also enables formal quantification of the relationship between correlated outcomes by estimating the strength of the association between the outcomes.

Most previously proposed joint models consist of two sub-models; longitudinal sub-model and time-to-event sub-model (where the time to event outcome is of primary interest). Here, we propose a joint model where the primary outcome of interest is count outcome (i.e. clinical AEs count) such that time to event outcome sub-model is replaced with count outcome sub-model. Since we assumed Poisson distribution of the clinical AEs count, the count outcome sub-model was specified as Poisson-distributed. Therefore, the joint model presented in this paper consist of two longitudinal continuous laboratory safety biomarkers sub-models (one for total bilirubin and one for alanine aminotransferase) and two Poisson sub-models (one for the individual clinical AEs count and one for the individual concomitant medication count). The connecting of these outcomes sub-models

is very flexible such that we could use random effects or expected value of outcomes as detailed in the model specification below.

6.2.3.2 Joint model formulation

As highlighted above, the joint model is formulated in such a way that clinical AE count is a primary outcome. In this section we describe the structures of the sub-models that made the joint model considered in this work. The estimates in all the models are considered on a logarithm scale.

6.2.3.2.1 Longitudinal sub-model

We modelled two continuous longitudinal clinical laboratory safety outcomes (i.e. total bilirubin and alanine aminotransferase) using a longitudinal sub-model based on a flexible linear mixed effects model that could accommodate nonlinear changes of the laboratory safety outcomes. The flexibility was achieved through the use of restricted cubic splines. Each continuous longitudinal clinical laboratory outcome v_{i1}, v_{i2} was assumed to be normally distributed with mean μ_k and variance δ_k^2 ($k=1, 2$) on a logarithm scale. Each continuous longitudinal clinical laboratory outcome is modelled using the sub-model below

$$v_{ik}(t) = \mathbf{X}_{ik}^T(t)\boldsymbol{\beta} + \mathbf{Z}_i(t)\mathbf{b}_i + \varepsilon_{ik}(t), i = 1, \dots, n; k = 1, 2 \quad (\text{M6.3})$$

where $\varepsilon_{ik}(t)$ is the error term observed at time t for the k th outcome from patient i . The error terms for the model are assumed to be independent, identical and normally distributed with mean 0. The mean response model $\mathbf{X}_{ik}^T(t)\boldsymbol{\beta}$, specified as a linear function

of the covariates at a given time for the outcome k is flexible such that it can accommodate time-varying covariates. The $\mathbf{Z}_i(t)$ is an indicator vector of random effects for k th outcome for patient i at time t such that it takes the value of 1 when there is a random effect and 0 otherwise. The vector of the patient-specific shared random effects $\mathbf{b}_i$ is assumed to have a multivariate normal distribution with mean 0 and variance-covariance matrix Σ , such that $(\mathbf{b}_i) \sim MVN(\mathbf{0}, \Sigma)$.

(i) Alanine aminotransferase sub-model

Based on M6.1, considering patient-specific random intercept as our shared parameter of interest, linking the outcomes, the longitudinal sub-model for alanine aminotransferase can be written simply as M4 below;

$$v_{i1}(t) = \mathbf{X}_{i1}^T(t)\boldsymbol{\beta} + \mathbf{b}_{0i1} \quad (\text{M6.4})$$

where b_{0i1} is the patient specific random intercept corresponding to the alanine aminotransferase outcome.

(ii) Total bilirubin sub-model

Similarly, the longitudinal sub-model for total bilirubin, M6.5, can be formulated as;

$$v_{i2}(t) = \mathbf{X}_{i2}^T(t)\boldsymbol{\beta} + \mathbf{b}_{0i2} \quad (\text{M6.5})$$

where b_{0i2} is the patient specific random intercept corresponding to total bilirubin outcome.

6.2.3.2.2 Count sub-models

We considered two event count sub-models; one for concomitant medication count and the other for the AE count. The two count sub-models were derived from a modified form of the generic Poisson model. The count sub-model was as follows;

(i) Concomitant medication count sub-model

Assuming that the concomitant medication count had a Poisson distribution, its sub-model was defined as

$$\log(\varphi|\mathbf{b}_{03i}) = \mathbf{X}_{ik}^T(t)\boldsymbol{\beta} + \mathbf{b}_{03i}\alpha_3 \quad (\text{M6.6})$$

such that $\log(\varphi|\mathbf{b}_{03i})$ represents the logarithm of the Poisson mean response, φ , conditional on the patient-specific random intercept, $\mathbf{b}_{03i}$ for concomitant medication count (i.e. the concomitant medication count sub-model is linked with the AE count sub-model in the joint model via the random intercept). The $\mathbf{X}_{ik}^T(t)\boldsymbol{\beta}$ is the linear predictor for the Poisson model and has a logarithm link function and contains a vector of fixed effects $\boldsymbol{\beta}$ corresponding to the baseline covariates. The α_3 represents log incidence rate ratio per unit increase in the patient-specific deviation from the mean random intercept of the reported concomitant medication use.

(ii) Clinical AE count sub-model

Considering that AE count was the primary outcome of interest and assuming a Poisson distribution for the AE count, the AE count sub-model was formulated as;

$$\log(\varphi|\mathbf{b}_{01i}, \mathbf{b}_{02i}, \mathbf{b}_{03i}) = \mathbf{X}_{ik}^T(t)\boldsymbol{\beta} + \mathbf{b}_{01i}\alpha_1 + \mathbf{b}_{02i}\alpha_2 + \mathbf{b}_{03i}\alpha_3 \quad (\text{M6.7})$$

The $\log(\varphi|\mathbf{b}_{01i}, \mathbf{b}_{02i}, \mathbf{b}_{03i})$ represents the logarithm of Poisson mean response, φ , conditional on the shared random intercepts for the two longitudinal clinical laboratory outcomes and the concomitant medication count. Each random intercept for the longitudinal sub-models and concomitant medication sub-model ($b_{01i}, b_{02i}, b_{03i}$) is linked to the linear predictor $\mathbf{X}_{ik}^T(t)\boldsymbol{\beta}$ for the Poisson model. This yields random effects association structure where $\alpha_1, \alpha_2, \alpha_3$ estimate the strength of the association between the respective continuous longitudinal clinical laboratory outcome, concomitant medication count and the AE count; the $\alpha_1, \alpha_2, \alpha_3$ represent log incidence rate per unit increase in the patient-specific deviation from the mean random intercept of a respective continuous longitudinal clinical laboratory outcome or the concomitant medication count outcome. For example, α_1 represents log incidence rate ratio per unit increase in the patient-specific deviation from the mean random intercept of a respective log alanine aminotransferase.

Alternative clinically meaningful formulation of the AE count sub-model could be linked to the expected value of the respective continuous longitudinal clinical laboratory outcome with the linear predictor of the Poisson model. This yields the current value association structure where the $\alpha_1, \alpha_2, \alpha_3$ represent log incidence rate ratio per unit increase of the respective continuous longitudinal clinical laboratory outcome, at time t. Current value association structure is very important and clinically plausible when linking continuous outcomes. In our cases, given the two continuous longitudinal clinical laboratory outcomes (i.e. log alanine aminotransferase and log total bilirubin) v_1, v_2 , under the current value association structure, we can modify the AE count sub-model as;

$$\log(\varphi) = \mathbf{X}_{ik}^T(t)\boldsymbol{\beta} + E[\mathbf{v}_1(t)|\mu_1(t)]\boldsymbol{\alpha}_1 + E[\mathbf{v}_2(t)|\mu_2(t)]\boldsymbol{\alpha}_2 + \mathbf{b}_{03i}\boldsymbol{\alpha}_3 \quad (\text{M6.8})$$

The $E[\mathbf{v}_1(t)|\mu_1(t)]$ and $E[\mathbf{v}_2(t)|\mu_2(t)]\boldsymbol{\alpha}_2$ are the expected current values of log alanine aminotransferase and total bilirubin respectively. It can be noted from M6.6 that the concomitant medication count, $\mathbf{v}_3$, sub-model is still linked to the AE count sub-model via a random intercept $\mathbf{b}_{03i}$. We maintained the random intercept ($\mathbf{b}_{03i}\boldsymbol{\alpha}_3$) parameterization since this efficiently deals with any potential over-dispersion problem encountered in Poisson models. Therefore, model M6.8 contains a mixture of both the expected current value and random effects association structures. Since both continuous outcomes (the log alanine aminotransferase and log total bilirubin) are assumed to be normally distributed (i.e. with identity link), the model can further be simplified as;

$$\log(\varphi) = \mathbf{X}_{ik}^T(t)\boldsymbol{\beta} + [\mu_1(t)]\boldsymbol{\alpha}_1 + [\mu_2(t)]\boldsymbol{\alpha}_2 + \mathbf{b}_{03i}\boldsymbol{\alpha}_3 \quad (\text{M6.9})$$

In this paper, we focus on reporting results from the M6.9 association structure since it is more clinically meaningful since it has association parameters that are directly interpretable in clinical practice.

6.2.3.3 Joint model likelihood and estimation

The estimates for the joint model are obtained by maximizing the marginal likelihood of the joint distribution of the observed data and the random effects[86]. The likelihood function for the observed data can be specified as;

$$L_i(\theta) = \prod_{i=1}^n f(\mathbf{V}_i|\mathbf{b}_i, \boldsymbol{\theta}) f(\mathbf{T}_i, \mathbf{C}_i|\mathbf{b}_i, \boldsymbol{\theta}) f(\mathbf{b}_i|\boldsymbol{\theta}) d\mathbf{b}_i \quad (\text{M6.10})$$

where θ represents the set of parameters of interest to be estimated including both the fixed and random effects (e.g. association parameters $\alpha_1, \alpha_2, \alpha_3$, and coefficients for the baseline covariates, β). Component $f(\mathbf{b}_i|\theta)$ is a normal density function for the random effects conditional on θ and $f(\mathbf{V}_i|\theta)$ is a normal density function for the log total bilirubin and log alanine aminotransferase conditional on θ and $\mathbf{b}_i$. The $f(T_i, \mathbf{C}_i|\mathbf{b}_i, \theta)$ represents Poisson density function conditional on θ and $\mathbf{b}_i$; patient-specific total follow up days is T_i and $\mathbf{C}_i$ represents total events recorded (for AE count sub-model and concomitant count sub-model). Since obtaining the log likelihood requires integrating out the random effects, it becomes challenging as number of random effects increases such that the traditionally used adaptive Gauss-Hermite quadrature would be limited (to handle the large number of random effects in our complex joint model). To mitigate this problem, we employed Monte Carlo integration technique since the number of draws it makes from the random effects do not need to change with increase in number of random effects (hence reducing the computation burden). Estimation of our models was done using merlin in Stata [158]. Only 7 patients had missing data points and the data missing mechanism was considered completely at random.

6.2.4 Statistical data analysis

Firstly, we computed the summary of the baseline characteristics for the women at enrolment. We presented means with respective standard deviation for the continuous variables (e.g. maternal age, gestational age) and frequencies with respective percentages for the categorical variables (e.g. trimester at enrolment, bed-net use). These summaries were also computed for the clinical AE occurrence and concomitant

medication counts. The concomitant medication was also summarized based on frequency of the specific patients and the number of patients who took the medication. In order to aid visualization of the trend of the data, we plotted box plots for the two continuous longitudinal clinical laboratory data (total bilirubin and alanine aminotransferase) outcomes by study visit day.

We considered two different models to empirically compare how they efficiently estimated effect of treatment on clinical AEs over the follow-up time. The models compared were conventional Poisson model M6.1 and joint model of M6.9 formulation above. Both the models adjusted for the same set of baseline covariates (maternal age, gravidity and trimester at enrolment). The choice of the potential confounders was driven by clinical relevance. Maternal age is a key confounder that determines antimalarial drug effect; as the age increases risk of adverse outcomes during pregnancy also increases. Therefore, in assessment of drug safety age is considered, in order to inform safe dosing quantity and schedules that appropriately accounts for the maternal age category. Additionally, trimester in which a pregnant woman has received an antimalarial treatment requires a special attention since drug safety can be very critical in the first two trimesters mainly due to the fetal development [26]. Therefore, both trimester and maternal age need to be considered in modelling antimalarial drug safety outcomes as key confounders since they can affect the clinical laboratory parameters as well as the clinical AEs incidence [172, 173]

The traditional Poisson model was the first model to be fitted and we used logarithm of the total follow-up time for each patient as an offset. Then we fitted a joint model with a mixture of expected current value of outcome and random effects association structure as shown in M6.9. We considered the joint model to assess the joint evolution (or association) between the clinical AEs and other three outcomes (alanine aminotransferase, total bilirubin and concomitant medication) where $\alpha_1, \alpha_2, \alpha_3$ as defined in M6.9 are the parameters quantifying the between-outcome association. Secondly, we were also interested in investigating the efficiency of the joint model in improving AE incidence rate estimate of the treatment effect obtained from the conventional Poisson model M6.1. In order to assess whether the treatment effect on AE incidence differed across the three arms we used global test. As part of sensitivity analysis we also investigated how the association structure of the joint model affects the magnitude of the treatment effect estimates. Comparing the association structures was achieved by fitting a joint model with random effects association structure only as specified in model M6.7 above and joint model with mixture of current value and random effects association structure of joint model M6.9. For both joint models, in order to flexibly model nonlinear changes of the continuous longitudinal clinical laboratory outcomes we used restricted cubic splines with 3 degrees of freedom considered sufficient to enhance flexibility of the model. We compared the fit of the two joint models (M6.9 versus M6.7) using the Akaike Information Criterion (AIC), with small values of AIC suggesting better model.

Additionally, we did an exploratory analysis to identify how the frequently reported concomitant medications are associated with the AEs. Such exploratory analysis was

helpful in understanding how specific frequently used concomitant medication influence the overall impact of concomitant medication on AEs. Concomitant medication was defined as frequent if it constituted at least 10% of the reported medications. A joint model of a similar structure to M6.9 but where the concomitant medication count was confined to the specific frequently reported concomitant medication (i.e. paracetamol) was used during the exploratory analysis.

6.3 Results

6.3.1 Descriptive results

The trial recruited 870 women who were randomly assigned to received AL, ASAQ, DHAPQ; each arm recruited 290 women. Baseline characteristics were similar across and between the trial arms (Table 6.1). Maternal age was skewed to the right. Overall, median maternal age was 20 years (IQR: 18.0-24.0). Overall mean haemoglobin concentration at enrolment (g/dl) was 10.1 (SD: 1.3) and distribution was also consistently similar across all the treatment arms. The overall reported bed net use in a night before enrolment was 23.2% and it was similar across and between the study arms (Table 6.1)

Table 6.1: Baseline characteristics for women enrolled in PREGACT trial in Malawi

Characteristic	Overall	AL	ASAQ	DHAPQ
Maternal age(years), median(IQR)	20.0 (18.0-24.0)	20.0 (18.0-24.0)	20.0 (18.0-25.0)	20.0 (19.5-24.0)
Gestation age (weeks), median(IQR)	21.0 (18.0-25.0)	21.0 (18.0-25.0)	21.0(18.0-25.0)	21.0 (18.0-24.0)
Haemoglobin(g/dl), mean(SD)	10.1 (1.3)	10.2 (1.4)	10.0 (1.3)	10.1 (1.3)
Primigravida, n (%)	444 (51.1)	144 (49.8)	144 (49.7)	156 (53.8)
Trimester				
Second, n (%)	730 (83.9)	242 (83.4)	244 (84.1)	244 (84.1)
Third, n (%)	140 (16.1)	48 (16.6)	46 (15.9)	46 (15.9)
Bed net use, n (%)	202 (23.2)	65 (22.4)	67 (23.1)	70 (24.1)

In total 1512 AEs were observed over the whole follow-up time and 108 of these were definitely not related to the treatment. Proportion of women who experienced at least one AE in the AL, ASAQ and DHAPQ treatment arms were 64.5% (187 of 290 women), 70.7% (205 of 290 women) and 59.3% (172 of 290 women), respectively. Almost all patients took at least a concomitant medication over the follow-up such that for each arm 99% (287 out of 290) took at least a concomitant medication (Table 6.2). Across all the treatment arms, the reported median concomitant medications taken was 3; for AL 3(IQR: 2-5), for ASAQ 3(IQR: 1-4) and for DHAPQ 3(IQR: 2-5).

Table 6.2: Summary of AE occurrence and concomitant medication use for women enrolled in PREGACT trial

Characteristic	AL	ASAQ	DHAPQ
Total AEs	475	569	468
AE drug-relatedness totals			
Definitely not	36	34	38
Possibly/probably	26	96	26
Unlikely	413	443	404
At least one AE ^a , n (%)	187 (64.5)	205 (70.7)	172 (59.3)
Total concomitant medications	1229	1033	1042
Median concomitant medications (IQR)	3 (2, 5)	3 (1, 4)	3 (2, 5)
At least one concomitant medication, n(%)	287(99.0)	287(99.0)	287(99.0)

a; The AE n(%) is based on total number of enrolled in each arm (i.e. 290 patients).

We observed that iron supplement (26.3%), paracetamol (26.2%) and albendazole (10.7%) were frequently reported concomitant medications (Table 6.3). Almost all the patients analysed, 99.2%, took the iron supplements and approximately two-thirds, 60.3%, of the patients took paracetamol apart from the antimalarial drug under investigation. During the follow-up period, a total of 353 patients reported having used albendazole once.

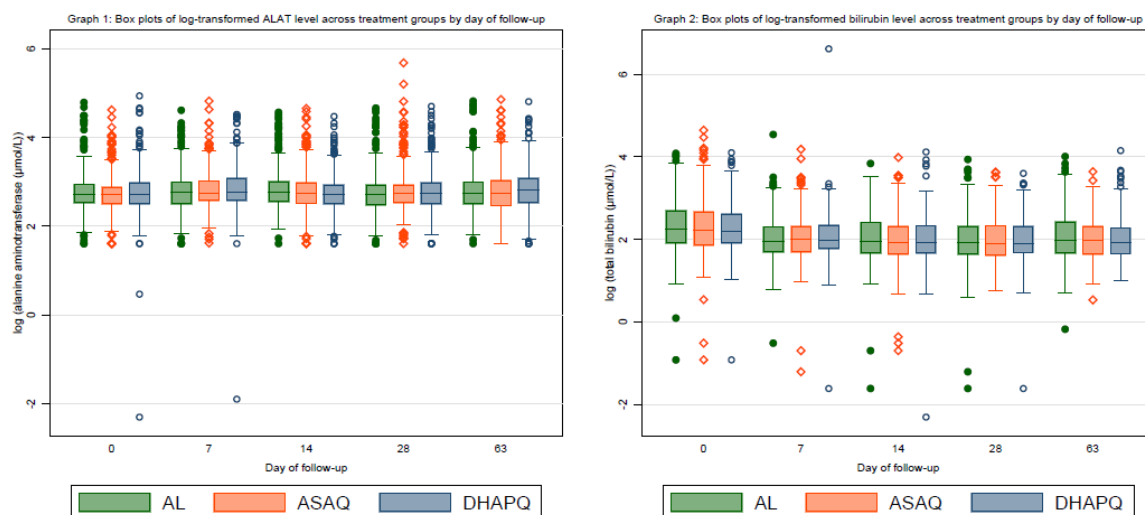
Table 6.3: Summary of concomitant medication reported by pregnant women in PREGACT trial

Concomitant medication	Number of times reported ^b , n(%)	Number of patients ^a , n(%)
Iron supplement	869 (26.3)	856 (99.2)
Paracetamol	864 (26.2)	520 (60.3)
Others	433 (13.1)	302 (35.0)
Albendazole	353 (10.7)	353 (40.9)
Amoxicillin	220 (6.7)	192 (22.3)
Benzathine	108 (3.3)	68 (7.9)
Pethidine	49 (1.5)	49 (5.7)
Chloramphenicol	48 (1.5)	43 (5.0)
Oral Rehydration Salts (ORS)	47 (1.4)	44 (5.1)
Dextrose	47 (1.4)	45 (5.2)
Lignocaine	39 (1.2)	39 (4.5)
Metronidazole	38 (1.2)	36 (4.3)
Quinine	35 (1.1)	26 (3.0)
Gentamycin	35 (1.1)	35 (4.1)
Sulfadoxine-Pyrimethamine	31 (0.9)	27 (3.1)
Aspirin	16 (0.5)	14 (1.6)
Pitocin	16 (0.5)	16 (1.9)
Erythromycin	13 (0.4)	12 (1.4)
Piriton	12 (0.4)	12 (1.4)
Methyldopa	12 (0.4)	9 (1.0)
Cotrimoxazole	8 (0.2)	8 (0.9)
Promethazine	7 (0.2)	7 (0.8)
Multivitamin Tabs	4 (0.1)	3 (0.4)

a: This is out of 863 patients who used concomitant medication; some patients took multiple concomitant medications; b: This is out of 3304 times of reported concomitant medication use

Overall median of log-transformed alanine aminotransferase was 2.7 (IQR: 2.5-3.0) such that the medians for AL was 2.8 (IQR: 2.5-3.0), for ASAQ was 2.7 (IQR: 2.5-3.0) and for DHAPQ was 2.8 (IQR: 2.5-3.0). Overall median for log-transformed total bilirubin was 2.0 (IQR: 1.7-2.4). Fig. 6.1 below shows the detailed distribution of each of clinical laboratory safety outcomes over the follow-up days by antimalarial drug treatment arm. As shown in graph 1 within Fig 6.1, the alanine amino-transferase did not vary much throughout the follow-up days. The general trend indicated that across the follow-up days, the alanine amino-transferase concentration was slightly higher among the women who received DHAPQ compared to those women who received either AL or ASAQ. After day 0 the total bilirubin concentration gradually decreased across all the treatment arms and, on average, women treated with AL had the highest median bilirubin concentration across the follow-up days (see Graph 2 within Fig 6.1).

Figure 6.1: Box plots profiling clinical laboratory safety outcomes for antimalarial drugs over 63-day follow-up among pregnant women in Malawi



6.3.2 Parameter estimates for the models

In this section, we report the estimates from the conventional Poisson model (where AEs count is an outcome) (Table 6.4) and the proposed joint model (Table 6.5). Our interest was to empirically establish whether the joint model yielded different results compared to the conventional Poisson model and how the jointly modelled outcomes evolved over the follow-up time in relation to clinical AEs; the results from the AE count sub-model in the joint model (Table 6.5) are of main interest in our reporting.

6.3.2.1 Parameter estimates from the conventional Poisson model for the clinical AE count

Using the conventional Poisson model, M6.1, the global test showed that the clinical AE incidence rate did not differ significantly across the treatment groups ($p=0.149$). However, the estimates suggested that the AE incidence rate was higher in the ASAQ treatment group compared to the AL treatment group (log IRR=0.1838; 95% CI: 0.0574, 0.3102; $p=0.004$) (Table 6.4). Exponentiating the point estimate of the log IRR yields IRR of 1.2018, implying that the patients treated with ASAQ were expected to have 20.18% higher rate of clinical AEs occurrence compared to those treated with AL. There was no evidence that the AE incidence rate differed between DHAPQ and AL treatment groups (log IRR=-0.0255; 95% CI: -0.1587, 0.1076; $p=0.707$). A year increase in maternal age was associated with 0.0208 increase in log IRR of the AEs (95% CI: 0.0068, 0.0347; $p=0.004$). The AE incidence rate was significantly higher among the those who enrolled in the second trimester compared to those who enrolled in the third trimester (log IRR=0.3768; 95% CI: 0.1943, 0.5192; $p<0.001$)

Table 6.4: Parameter estimates for Clinical AE count incidence using conventional Poisson model

Characteristic	Coefficient	SE	95% CI	P-value
Treatment				
ASAQ	0.1838	0.0645	0.0574, 0.3102	0.004
DHAPQ	-0.0255	0.0679	-0.1587, 0.1076	0.707
Maternal age	0.0208	0.0071	0.0068, 0.0347	0.004
Primigravid	0.1208	0.0731	-0.0224, 0.2641	0.098
Second trimester	0.3768	0.0829	0.1943, 0.5192	<0.001
Constant	-0.3901	0.2021	-0.7862, 0.0060	0.054

6.3.2.2 Parameter estimates from the joint model

We report the fixed and random parameter estimates for the final fitted joint model based on mixed association structure, M6.9, (i.e. random effects and current value association structure) in Table 6.5. Using the global test, AE count sub-model indicated that there was no evidence that the AE incidence varied across all the treatment groups ($p=0.376$); this is despite the fact that the estimates indicated that AE incidence rate was higher in ASAQ treatment group compared to AL (log IRR=0.2118, 95% CI: 0.0082, 0.4154; $P=0.041$) suggesting that patients treated with ASAQ were expected to had 0.2118 higher log IRR of clinical AEs occurrence compared to those treated with AL. There was also no evidence that AE incidence rate significantly differed in DHAPQ treatment compared to AL (log IRR: -0.0508; 95% CI: -0.2559, 0.1542). Our results show that the magnitude of the treatment effect was higher in the joint model, M6.9, compared to the estimates in the conventional Poisson model, M6.1, above. Among the association parameters describing joint evolution between the outcomes, only concomitant medication was associated with clinical AEs. The association parameter, α_3 , linking

concomitant medication and clinical AEs suggested that the clinical AEs outcome was positively and strongly associated with the concomitant medications (log IRR=1.7487; 95% CI: 1.5471, 1.9503; $p<0.001$). A unit increase for a patient in deviation from the mean random intercept of concomitant medication use was associated with 1.7487 increase in log IRR of clinical AEs. We found no evidence that the clinical AEs were associated with the alanine aminotransferase nor total bilirubin over the follow-up time as demonstrated by their respective association parameters that were non-significant. For the association between alanine aminotransferase and clinical AEs over follow-up time the estimated log IRR was -0.0288 (95% CI: -0.5045, 0.4469; $p=0.906$) and for the association between total bilirubin and clinical AEs the log IRR was 0.1153 (95%CI: -0.0889, 0.3194; $p=0.269$). As also observed in the conventional Poisson model above, the clinical AE sub-model indicated that those women who enrolled in the second trimester had a higher AE incidence rate compared to those who enrolled in the third trimester (log IRR=0.3986; 95% CI: 0.1592, 0.6381; $p=0.001$); the women who enrolled in the second trimester had 0.3986 higher log IRR of clinical AEs compared to those who enrolled in the third trimester.

Overall, the standard errors are at least twice in the joint model compared to the biomarkers` submodels. The difference would be attributed to lack of evidence of difference in the biomarker values across the treatments or the other covariates. In order to arrive at firm theoretical conclusion, potential future work (i.e. simulation study) can probably explore how different levels of the biomarkers correlations and evolution would affect the standard error values and eventually the hypothesis being tested.

Table 6.5: Parameter estimates from the joint multivariate longitudinal clinical laboratory, concomitant medication and AE counts model

Variable	Coefficient(SE)	95% Confidence interval	P-value
<i>Log alanine amino transferase longitudinal sub-model</i>			
Maternal age	0.0006 (0.0023)	-0.00394, 0.0052	0.793
Treatment			
ASAQ	-0.0036 (0.0275)	-0.0575, 0.0504	0.897
DHAPQ	0.0031 (0.0275)	-0.0508, 0.0570	0.909
Second trimester	-0.0212 (0.0311)	-0.0820, 0.0397	0.496
Constant	2.8221 (0.0609)	2.7027, 2.9415	<0.001
<i>Log total bilirubin longitudinal sub-model</i>			
Maternal age	-0.0026 (0.0035)	-0.0095, 0.0043	0.464
Treatment			
ASAQ	-0.0229 (0.0424)	-0.1059, 0.0601	0.589
DHAPQ	-0.0413 (0.0422)	-0.1240, 0.0414	0.328
Second trimester	-0.1786 (0.0493)	-0.2752, -0.0820	<0.001
Constant	2.3103 (0.0963)	2.1216, 2.4990	<0.001
<i>Concomitant medication count sub-model</i>			
Maternal age	0.0086 (0.0052)	-0.0017, 0.0189	0.101
Treatment			
ASAQ	-0.1841 (0.0619)	-0.3054, -0.0629	0.003
DHAPQ	-0.1723 (0.0610)	-0.2918, -0.0528	0.005
Second trimester	0.2861 (0.0718)	0.1454, 0.4268	<0.001
Constant	0.9147 (0.1395)	0.6414, 1.188	<0.001
<i>Adverse events count sub-model</i>			
Maternal age	0.0189 (0.0088)	0.0017, 0.0362	0.031
Treatment			
ASAQ	0.2118 (0.1039)	0.0082, 0.4154	0.041
DHAPQ	-0.0508 (0.1046)	-0.2559, 0.1542	0.627
Second trimester	0.3986 (0.1222)	0.1592, 0.6381	0.001
$\alpha_{\log(ALT)}$	-0.0288 (0.2427)	-0.5045, 0.4469	0.906
$\alpha_{\log(bilirubin)}$	0.1153 (0.1042)	-0.0889, 0.3194	0.269
$\alpha_{concmcd}$	1.7487 (0.1029)	1.5471, 1.9503	<0.001
Constant	-0.8220 (0.7739)	-2.3388, 0.6949	0.288
<i>Patient-specific random effects standard deviations</i>			
Standard deviations			
$\delta_{\log(ALT)}$	0.2547 (0.0119)	0.2324, 0.2792	
$\delta_{\log(bilirubin)}$	0.5025 (0.0154)	0.4732, 0.5336	
$\delta_{\log(concmcd)}$	0.5378 (0.0248)	0.4914, 0.5887	

ALT=alanine aminotransferase, concmcd=concomitant medication

In the sub-model for log alanine aminotransferase, we observed no evidence of any statistically significant difference across all the variables included in that sub-model. For the log total bilirubin sub-model, we observed 0.1786 lower mean log total bilirubin among the women who enrolled in the second trimester compared to those who enrolled in the third trimester (mean difference log bilirubin: -0.1786; 95% CI: -0.2752, -0.0820, $p < 0.001$).

The concomitant sub-model showed that concomitant medication was associated with the treatment group. Women enrolled in ASAQ treatment group reported lower incidence rate of concomitant medication use compared to those in AL treatment group (log IRR=-0.1841; 95% CI: -0.3054, -0.0629; $p=0.003$). Women in DHAPQ treatment group also reported lower incidence rate of concomitant medication use compared to those in AL treatment group. Those women who enrolled in the second trimester reported a higher incidence rate of concomitant medication use than those who enrolled in the trial while in the third trimester (log IRR=0.2861; 95% CI: 0.1454, 0.4268; $p<0.001$).

The standard deviations for patient-specific random intercepts, across the outcomes, suggested that concomitant medication had the highest unobserved heterogeneity with standard deviation of 0.5378 (95% CI: 0.4914, 0.5887). The lowest unobserved heterogeneity was observed in log alanine aminotransferase outcome with random intercept standard deviation of 0.2547 (0.2324, 0.2792).

6.3.3 Sensitivity analyses

As part of sensitivity analysis, we attempted to relax the normality assumption of the random effects. We fitted a robust joint model for the mixture association structure, M9, where the random effects were assumed to have a t-distribution. The model yielded similar results as those in Table 6.5 and Akaike Information Criterion (AIC) of 18230.28 which was comparable to an AIC of 18278.53 for the model in Table 6.5. Secondly, fitting

a joint model with random effects association structure, M6.5, we obtained similar estimates as those in Table 6.5 and with negligibly lower AIC of 18277.87.

Following a reviewer's suggestion, we further investigated the effect of the concomitant medications with high frequencies on AEs. Although iron supplement was the most commonly reported concomitant medication, we did not consider it in assessing the impact of specific concomitant medication on AEs since it was taken by almost all the patients, 99.2%. Instead, we considered paracetamol since it was the second-highest reported concomitant medication where 60.3% of the patients reported it at least once. Assessment of the effect of treatment on clinical AEs using a joint model of a similar structure to M6.9 but where the concomitant medication count was confined to paracetamol yielded similar results as those obtained in Table 6.5 (see Table 6.6). The paracetamol concomitant medication was positively and strongly associated with clinical AEs (log IRR=1.2969; 95% CI: 1.1328, 1.4609; $p<0.001$). However, this observed effect of paracetamol concomitant medication was slightly lower compared to the overall concomitant medication effect as reported in Table 6.5 (log IRR=1.7487; 95% CI: 1.5471, 1.9503; $p<0.001$). No interaction was observed between the paracetamol concomitant medication and the treatment ($p=0.078$) such that we focussed on reporting results with the main effect only model as shown in Table 6.6. The model with interaction terms had a negligibly lower AIC of 16556.19 than the model without the interaction terms with an AIC of 16557.29.

The interest and structure of the proposed joint model could not permit us to assess the impact of albendazole on AEs. The patients who took albendazole took it once; therefore, this concomitant medication arose as a binary outcome that does not meet the joint model specification (i.e. where the concomitant medication is incorporated in the model as a count outcome). Since the objective of the exploratory analysis was to establish the impact of specific frequent concomitant medication on AEs, we did not undertake any additional analysis on the effect of the “other” concomitant medication on AEs.

Table 6.6: Parameter estimates from the joint multivariate continuous longitudinal clinical laboratory, paracetamol concomitant medication count and AE count model

Variable	Coefficient(SE)	95% Confidence interval	P-value
<i>Log alanine amino transferase longitudinal sub-model</i>			
Maternal age	0.0006 (0.0023)	-0.0040, 0.0051	0.810
Treatment			
ASAP	-0.0016 (0.0276)	-0.0557, 0.0525	0.953
DHAPQ	0.0031 (0.0275)	-0.0508, 0.0570	0.909
Second trimester	-0.0223 (0.0311)	-0.0833, 0.0387	0.474
Constant	2.8236 (0.0608)	2.7043, 2.9428	<0.001
<i>Log total bilirubin longitudinal sub-model</i>			
Maternal age	-0.0027 (0.0034)	-0.0094, 0.0039	0.418
Treatment			
ASAP	-0.0346 (0.0419)	-0.1168, 0.0476	0.409
DHAPQ	-0.0483 (0.0416)	-0.1299, 0.0334	0.246
Second trimester	-0.1747 (0.0471)	-0.2669, -0.0824	<0.001
Constant	2.3118 (0.0911)	2.1394, 2.4966	<0.001
<i>Concomitant medication count sub-model</i>			
Maternal age	0.0144 (0.0088)	-0.0030, 0.0318	0.103
Treatment			
ASAP	-0.3469 (0.1068)	-0.5563, -0.1376	0.001
DHAPQ	-0.2850 (0.1046)	-0.4901, -0.0798	0.006
Second trimester	0.3850 (0.1273)	0.1355, 0.6344	0.002
Constant	-0.6712 (0.2385)	-1.1387, -0.2038	0.005
<i>Adverse events count sub-model</i>			
Maternal age	0.0161 (0.0092)	-0.0019, 0.0340	0.079
Treatment			
ASAP	0.1687 (0.1099)	-0.0466, 0.3840	0.125
DHAPQ	-0.0876 (0.1103)	-0.3038, 0.1286	0.427
Second trimester	0.3986 (0.1270)	0.1497, 0.6474	0.002
$\alpha_{\log(ALT)}$	0.0016 (0.2392)	-0.4673, 0.4704	0.995
$\alpha_{\log(bilirubin)}$	0.0218 (0.1207)	-0.2148, 0.2585	0.856
$\alpha_{concmcd}$	1.2969 (0.0837)	1.1328, 1.4609	<0.001
Constant	-0.6805 (0.7793)	-2.2079, 0.8469	0.383
<i>Patient-specific random effects standard deviations</i>			
Standard deviations			
$\delta_{\log(ALT)}$	0.2555 (0.0119)	0.2331, 0.2800	
$\delta_{\log(bilirubin)}$	0.5028 (0.0153)	0.4736, 0.5338	
$\delta_{\log(concmcd)}$	0.7931 (0.0480)	0.7043, 0.8931	

ALT=alanine aminotransferase, concmcd=paracetamol concomitant medication

6.4 Discussion

In clinical trials, multiple safety outcomes collected over the follow-up time require advanced analyses in order to develop a valid drug safety profile that makes full use of the information from the multiple safety outcomes. This paper introduced a framework for joint modelling of multivariate continuous longitudinal clinical laboratory safety outcomes,

concomitant medication counts and adverse event count, exploiting the advantages of the existing extended multivariate generalised linear and non-linear mixed effects models framework [86]. Our work focussed on investigating whether clinical AEs varied by treatment arm and how the laboratory outcomes and concomitant medication were associated with clinical AEs over follow-up time in artemisinin-based treatment of malaria during pregnancy clinical trial. We found that the concomitant medication was strongly associated with the clinical AEs incidence, suggesting that the patients who experienced more AEs were more likely to report the use of more concomitant medications. If not well-accounted for, such increased use of concomitant medication can distort the treatment effect on the clinical AE occurrence. From methodological perspective, we observed that the conventional Poisson model yielded lower treatment effect on the AE incidence rate ratio; the effect of treatment on AE incidence rate estimates for the joint model were higher than those from conventional Poisson model. The difference in the models can be attributed to the fact that the joint model uses more of the available information through the explicit modelling of the observed correlation, unobserved heterogeneity and any nonlinear effects [174-176]. The joint model efficiently uses the available information, including the unobserved heterogeneity, leading to improved estimates compared to the conventional Poisson model involving separate analysis.

Controlling for baseline covariate to address any potential confounding in AE incidence rate ratio estimates, we found that maternal age and enrolling in the second trimester were factors associated with the AE occurrence. A year increase in maternal age was associated with 2% increased rate of AE incidence. This could be as a result that as the

women become older they are more likely to report AEs out of experience than the younger women. As expected women who enrolled in the second trimesters had 49% increased rate of AE incidence compared to those who enrolled in the third trimester. This could probably be due to the fact that physiological changes that take place during pregnancy are more impactful in the second than in the third trimester. Hence, the susceptibility to AE occurrence after taking either the antimalarial drug or the concomitant medication should be more expected during the second semester.

In assessment of the impact of concomitant medication on AEs, identifying specific concomitant medication is usually of interest. In this paper, we observed that paracetamol greatly contributed to the overall association between concomitant medication and AEs. Although the difference in the effect of overall concomitant medication on AEs and the effect of paracetamol concomitant medication on AEs was not big, the difference could not be considered clinically negligible. Therefore, undertaking both the analyses (i.e. effect of overall concomitant medication and effect of specific most frequent concomitant medication on AEs) should be considered useful in understanding antimalarial drug safety profile. Alternatively, an interesting future joint model development can consider quantifying the contribution of each of the specific frequent concomitant medication on AEs occurrence. This can be helpful in identifying and handling any potential noisy concomitant medications that contribute less to the AEs occurrence.

Within the joint modelling framework, our proposed model provides additional knowledge by extending the number of outcomes that can be analysed simultaneously, including multiple longitudinal outcomes and multiple count outcomes, focussing on drug safety assessment. We acknowledge that other researchers have previously proposed similar kind of modelling but applied in the efficacy context[177-181]. Buu et al. considered a joint model of count and binary outcome[179]. Li et al. considered jointly modelling proportion, count and continuous longitudinal outcomes[180]. Yang and Kang introduced a joint model for mixed Poisson outcome and continuous outcome [181]. Our model is considered as a special case of these previously proposed models focussing on drug safety assessment. Key unique features of our proposed joint model include incorporating mixed association structure of random effects and expected value for linking the outcomes, use of restricted cubic splines to accommodate the capturing of any potential time-dependent effects, accommodating multiple safety outcomes and novel application to drug safety assessment. A case study based paper like ours can further facilitate the adoption of improved statistical analysis of safety data in clinical trials [33].

In this paper, data missingness mechanism was ignorable. Our future work will consider investigating the impact of missing data on the model performance under varying follow-up times and model association structures. Secondly, in this paper we worked with a large sample size. An interesting further methodological work would be to assess how the sample size impacts the performance of the model that we applied. Unfortunately, the key limitation of the joint model introduced and applied in this paper is that it is computationally demanding such that it takes longer time to converge. This directly affects the scope of

the simulations that can be done to investigate the highlighted potential methodological issues associated with such complicated models. An extension can also be made to the number of outcomes included the joint model depending on the availability of the data. For example, aspartate amino transferase, highly correlated with alanine amino transference, can be added in our joint model.

6.5 Conclusion

Jointly modelling of multivariate longitudinal clinical laboratory safety data, concomitant medication and clinical AEs efficiently harnesses the safety data in order to better understand drug safety profile. This paper has demonstrated the utility of joint modelling to yield improved AE incidence rate estimates compared to the conventional Poisson model. The proposed joint model helps to quantify the association between the laboratory safety outcomes, concomitant medication, and clinical AEs over follow-up time. Key public health relevance of the joint modelling of the safety outcomes includes facilitating the choice of the most important clinical laboratory safety and other safety outcomes for profiling drug safety, especially in resource-constrained settings. In this exploratory analysis, we have also established that the AEs that were observed in the PREGACT trial in Malawi were more linked to concomitant medications that the patients took. Our findings provide an assurance that artemisinin-based treatments can be safely used in second and third trimester during pregnancy.

Chapter 7: Effect of adverse events on non-adherence and study non-completion in malaria chemoprevention during pregnancy trial: A nested case control study

Noel Patson^{1,2*}, Mavuto Mukaka^{3,4}, Ingrid Peterson⁵, Titus Divala^{6,7}, Lawrence Kazembe⁸, Don Mathanga², Miriam K Laufer⁵, Tobias Chirwa¹

1 School of Public Health, University of the Witwatersrand, Johannesburg, South Africa,

2 School of Public health and Family medicine, College of Medicine, University of Malawi, Blantyre, Malawi,

3 Mahidol Oxford Tropical Medicine Research unit (MORU), Bangkok, Thailand,

4 Centre for Tropical Medicine, Nuffield Department of Medicine, University of Oxford, Oxford, UK

5 Center for Vaccine Development and Global Health, University of Maryland, School of Medicine, Baltimore, MD, USA

6 TB Centre, London School of Hygiene and Tropical Medicine, London, UK

7 Helse Nord Tuberculosis initiative, College of Medicine, University of Malawi, Blantyre, Malawi

8 Department of Biostatistics, University of Namibia, Windhoek, Namibia.

* Corresponding author

Email: noelpatson@gmail.com

7.0 Abstract

Background

In drug trials, adverse events (AEs) burden can induce treatment non-adherence or discontinuation. The non-adherence and discontinuation induce selection bias, affecting drug safety interpretation. Nested case control(NCC) study can efficiently quantify the impact of the AEs, although choice of sampling approach is challenging. We investigated whether NCC study with incidence density sampling is more efficient than NCC with path sampling under conditional logistic or weighted Cox models in assessing the effect of AEs on treatment non-adherence and participation in preventive antimalarial drug during pregnancy trial.

Methods

Using data from a trial of medication to prevent malaria in pregnancy that randomized 600 women to receive chloroquine or sulfadoxine-pyrimethamine during pregnancy, we conducted a NCC study assessing the role of prospectively collected AEs, as exposure of interest, on treatment non-adherence and study non-completion. We compared estimates from NCC study with incidence density against those from NCC with path sampling under conditional logistic and weighted Cox models.

Results

Out of 599 women with the outcomes of interest, 474 (79%) experienced at least one AE before delivery. For conditional logistic model, the odds ratio for the effect of AE occurrence on treatment non-adherence was 0.70 (95% CI: 0.42, 1.17; $p=0.175$) under incidence density sampling and 0.68 (95% CI: 0.41, 1.13; $p=0.137$) for path sampling. For

study non-completion, the hazard ratio was 1.02 (95% CI: 0.56, 1.83; $p=0.955$) under incidence density sampling and 0.85 (95% CI: 0.45, 1.60; $p=0.619$) under path sampling. We obtained similar hazard ratios and odds ratios with standard errors under incidence density sampling and path sampling whether weighted Cox or conditional logistic models were used.

Conclusion

NCC with incidence density sampling and NCC with path sampling are practically similar in efficiency whether conditional logistic or weighted Cox analytical methods are used although path sampling uses more unique controls to achieve the similar estimates.

Trial registration

ClinicalTrials.gov: NCT01443130

7.1 Background

Treatment non-adherence and treatment discontinuation, in which patients either do not take their medication as prescribed or do not take it at all, is a challenge in analysis and interpretation of randomized controlled trial results since it can lead to underestimation of treatment effects [182]. Adverse events (AEs) occurrence is one of the factors that influence treatment non-adherence and discontinuation. Therefore, treatment non-adherence and discontinuation can be considered as proxy outcomes for a measure of the impact of AEs on patients in clinical trials [6]. However, most trials descriptively report treatment non-adherence and discontinuation information in relation to AEs under safety and tolerability section and participant flow chart of the main study report [138, 183, 184]. Such kind of reporting and analysis is less informative since it does not provide complete estimates on how the AEs affect non-adherence over the follow-up time, after accounting for potential confounders e.g. patient age. Further, AEs that impact treatment non-adherence and study discontinuation outcomes precede these outcomes. Ignoring the fact that the AEs that impact treatment non-adherence and study non-completion outcomes occur before the observation of the treatment non-adherence or discontinuation can lead to underestimation of the effect of AEs [185, 186]. Establishing association between the AEs and treatment non-adherence or discontinuation is also both statistically and ethically challenging since the AE exposure cannot be randomized and it is difficult to pre-specify at design stage. Use of a nested case-control approach to understand the impact of AEs can be useful in this context since it derives a sample of cases and controls that are comparable, accounting for potential differences in follow-up time. However, the choice of an efficient nested case-control design and analytical method can present

challenges. In order to address this challenge, empirical comparison of competing nested case-control designs is useful especially towards the choice of the most efficient nested case-control sampling and analytical approach.

Traditionally incidence density sampling (i.e. risk-set sampling) is used in most nested case-control studies. Under a nested case-control study with incidence density sampling design, each index case is matched to a specified number of controls sampled at the time of index case occurrence. The nested case-control study with incidence density sampling design is considered as an unbiased and statistically efficient approach to cohort sampling that accommodates causal interpretation of the observed effect of an exposure on an outcome [77]. The matching on time ensures that the cases are comparable to control by minimizing confounding that would arise due to the differences in follow-up time between the cases and controls. Hence, the obtained estimates are unbiased and reflect the original trial cohort [187]. The selection of controls is done with replacement such that a subject could be selected as a control multiple times and may later become a case. However, it has also been argued that greater efficiency can also be achieved if the set of controls for different cases are as disjoint as possible. For instance, Langholz and Thomas proposed a modified alternative sampling approach that aims at ensuring that minimum number of controls are used more than once [87]. Under this second alternative nested case-control sampling approach, firstly a path set is constructed that includes all subjects who entered the study before t_i in time interval $(t_{i-1}, t_i]$ and exited after time t_j in time interval $(t_j, t_{j+1}]$. Under this second sampling approach the specified number of controls are randomly sampled from the path set without replacement until all the

controls in that path set are selected and the path set can be replenished. Once the desired number of controls is achieved the path sampling is stopped. As defined elsewhere, in this paper we refer to this variation of nested case-control with incidence density sampling design as nested case-control study with *path sampling* design [188]. Based on the provided descriptions of the nested case-control with incidence density sampling and nested case-control with path sampling, path sampling samples more distinct patients than the traditional incidence density sampling. In practice, it is important to choose a nested case-control sampling design that can optimally estimate the effect of AEs on treatment non-adherence and study non-completion. However, there is limited work highlighting whether inclusion of more distinct controls via path sampling offers an added advantage over incidence density sampling.

Conventionally, conditional logistic regression or Cox regression stratified on case-control sets are used in analysis of most nested case-control studies [189]. Recently inverse probability weighting (IPW) Cox models have been shown to have higher power and efficiency than conditional logistic regression analysing nested case-control studies [190-193]. Under IPW Cox models, individual log-likelihood is weighted by the inverse probability of being included in the nested case-control study. The weighting can even offer more advantages of adapting the modelling, generally, beyond proportional hazards regression models [194]. Furthermore, the weighted Cox model also breaks the match to accommodate the use of cases and controls at all-time points when they are at risk.

Most of the studies discussed above have demonstrated the utility of the nested case-control with incidence density sampling under conditional logistic regression model and IPW models such as weighted Cox regression model. However, there is limited work demonstrating the utility and comparing the efficiency of the two alternative sampling designs (i.e. incidence density sampling and path sampling) under IPW Cox models and the traditional conditional logistic regression. In this paper, we investigated whether nested case-control with path sampling is more efficient than nested case-control study with incidence density sampling under conditional logistic regression model and weighted Cox regression model. We empirically compared nested case-control with incidence density sampling design against nested case-control with path sampling design under weighted Cox model and conditional logistic regression model, in investigating the effect of AEs on treatment non-adherence or study non-completion in the context of intermittent preventive treatment of malaria in pregnancy (IPTp) trial as case study. We hypothesized that path sampling design and weighted Cox regression will yield more efficient estimates with lower standard errors and AE exposure will increase the risk of non-adherence or discontinuation.

7.2 Methods

7.2.1 Motivating data description

Using nested case-control study approach, we assessed the effect of AEs on treatment non-adherence and study non-completion. The data used in the current study was collected in clinical trial of preventive treatment for malaria in pregnancy Malawi (NCT01443130). This open-label randomised trial aimed at evaluating whether

chloroquine for intermittent preventive treatment to prevent malaria in pregnancy(IPTp-CQ) or chloroquine chemoprophylaxis was better than IPTp using sulfadoxine-pyrimethamine (IPTp-SP). The trial included women in first or second trimester before 27 weeks of gestational age, attending first antenatal visit and willing to remain in the study area until 14 weeks after delivery. The trial excluded women who were HIV positive, on any known chronic therapy with antimalarial or anti-folate drugs and had known high risk pregnancies.

Between February 2012 and May 2014 a total of 600 pregnant women in their first or second trimester were recruited and randomized to receive (a) sulfadoxine-pyrimethamine(SP) IPTp (two doses, four weeks apart), (b) chloroquine IPTp (600 mg on day 1, 600 mg on day 2, and 300 mg on day 3 two days apart), (see Fig S7.1 in appendix 1). In the current analysis, we excluded a third arm consisting of 300 women randomized to receive weekly chloroquine prophylaxis since that arm had a radically different treatment schedule compared to the IPTp arms. The study treatments were administered under directly observed therapy. Detailed procedures, participant demographics and treatment effects for the trial are described in the primary trial manuscript [80].The clinical trial protocol was registered with [ClinicalTrials.gov](https://clinicaltrials.gov/ct2/show/study/NCT01443130) (NCT01443130)

Safety and tolerability assessments were based on unsolicited adverse events collected from the first day of the treatment administration until the final study visit. AEs were graded based on a standardized protocol , (i) graded on a four-point scale as, mild, moderate, severe, or life-threatening and (ii) further classified by System Organ Class category. The

clinical trial was reviewed and approved by Institutional Review Board at the University of Maryland, the College of Medicine Research and Ethics Committee at the University of Malawi and the Malawi Pharmacy Medicines and Poisons Board. From every participant, a written informed consent was obtained just before undergoing any trial procedure. During the current analysis we only considered follow-up time from enrolment up to delivery because medication was administered until delivery.

7.2.2 Nested case control study designs

We considered two types of nested case-control study designs; (a) Nested case-control study with risk-set sampling (i.e. incidence density sampling) (b) Nested case-control study with path sampling. The samples for the two designs were derived from the cohort $\varsigma = \{i, \dots, n\}$ of women who had taken at least their first dose of IPTp administered in the trial. Both nested case-control designs were based on two study outcomes: (i) treatment non-adherence and (ii) study non-completion. However, if the only outcome of interest (e.g. treatment non-adherence or study non-completion) occurred only at enrolment time ($t=0$), the woman was considered ineligible for the nested case-control study since matching was done on follow-up time ($t>0$). For both designs we identified same set of baseline characteristics as potential confounders. The potential confounders included maternal age in years, body mass index (BMI; calculated as weight in kilograms divided by height in metres squared), gravidity (primigravid or not) and, IPTp treatment arm.

7.2.2.1 Nested case control with incidence density sampling:

Definition of a case was based on experiencing an outcome of interest as defined in the preceding section. Firstly, we defined cases of *treatment non-adherence* as those missing at least one dose of treatment after the first dose. Although there was a possibility of observing more than one non-adherence events per woman, we focused on time to the first treatment non-adherence after enrolment. Secondly, we defined *study non-completion* as discontinuing participation in the study before delivery.

The controls were defined as women who had not yet experienced the outcome interest (i.e. *treatment non-adherence* or *study non-completion*) but were at risk to become cases. For each treatment non-adherence and study non-completion cases, we randomly selected four controls using risk set sampling (i.e. incidence density sampling). The selection of the controls (and corresponding analyses) was done separately for each endpoint. Each case was matched on follow-up time since enrolment by selecting controls from the set of women in the clinical trial cohort who had not experienced the outcome up until the time that the case arose; controls were selected without regard to outcomes that might occur after the time of matching. Hence a control could be sampled multiple times and could later become a case. We considered the selection of four controls per case as sufficient since there is negligible gain in statistical efficiency when more than four controls are selected per case [195-197].

The exposure in this analysis was having an AE before selection in risk-set sampling. We dichotomized the AEs as (having an AE versus not having one before the index case time). Due to the exploratory nature of our study in demonstrating the efficiency of

incidence density sampling over path sampling under different analytical methods, we considered only any AE exposure regardless of the relatedness to the treatment or severity.

7.2.2.2 Nested case control with path set sampling

The definition of cases and exposure are as in the incidence density sampling above. Every aspect for the path sampling approach is similar to the standard nested case-control with incidence density sampling except that this approach (i.e. path sampling) attempted to minimize the number of times that a control could be selected more than once. This was achieved using the path set concept where we sampled a path set according to the probabilities given the path set representation in the entire risk set at each event time [87]. As highlighted in the introduction section above, Langholz and Thomas indicate that the path set Q_{ij} consist of all subjects who entered the study before t_i in time interval $(t_{i-1}, t_i]$ and exited after time t_j in time interval $(t_j, t_{j+1}]$. In our context, the path set consisted of women who enrolled before t_i and exited the trial after t_j , in the respective time intervals. After the path set was defined and sampled, the controls were randomly sampled from the selected path set until the whole path set was emptied when the set can be replenished (i.e. once a woman was selected as a control from the path set, she could not be selected again until all women in that path set were selected). We implemented this sampling design using the recently illustrated implementation of the design in R statistical software [188]. This sampling approach yields as unique number of patients as possible included in the nested case-control study. Therefore, path sampling attempts to use more information from the controls by sampling

more unique controls from the cohort. As the case with nested case-control with incidence density sampling, the selection of the controls (and corresponding analyses) was done separately for each endpoint.

7.2.3 Statistical analysis

We conducted all the data management and statistical analyses in Stata version 15 (StataCorp, College Station, TX, USA) and R software version 3.6.0 (R Foundation, Vienna, Austria).

Descriptive analysis: Firstly, we summarized the baseline characteristics of the study participants by case-control status. For continuous variables, we reported mean (standard deviation [SD]) or median (inter-quartile range [IQR]) and proportion for categorical variables. We also summarized the distribution of AE occurrence within respective treatment arms based on AE grade and frequency. We plotted Kaplan Meier graph to characterise time to first AE occurrence between the IPTp arms.

7.2.3.1 Modelling the effect of AEs on non-adherence and study non-completion

Let likelihood for the nested case-control for parameter β be $L_{ncc}(\beta)$. Traditionally, estimates in nested case-control studies are derived by maximizing partial likelihood function that can be written as

$$L_{ncc}(\beta) = \prod_{i \in \Omega} \frac{\exp(\beta' X_i)}{\sum_{j \in \mathbb{R}_i} \exp(\beta' X_j)} \quad (\text{M7.1})$$

Where $L_{ncc}(\beta)$ is likelihood for nested case-control for parameter β . The β represents a vector of the regression coefficients, such that β' is a transposed vector corresponding to vector(s) of covariates. The X_i and X_j are the vectors of covariates for the individuals

i and j . The sampled risk set $\hat{R}_i$ consists of case i and sampled m controls matched to the case on given time t since enrolment. The Ω represents the set of all the cases. The conditional likelihood under logistic regression coincides with Cox regression partial likelihood [198]. Therefore, the analysis of nested case-control study using Cox regression model or conditional logistic stratified on case-control set yield similar results. In this study we firstly considered, conditional logistic models whose results are always the same as the traditional stratified Cox regression model (i.e. we stratified the model by case-control set indicator $\acute{s}$ (see the equation **M7.2** below). The baseline hazard for the stratum at follow-up time t was considered as $h_{0\acute{s}}(t)$.

$$h_i(t) = h_{0\acute{s}}(t) \exp(\beta' X_i) \quad (\text{M7.2})$$

As already defined, the partial likelihood that is maximized to obtain the estimate for the regression coefficient β is similar to **M7.1**;

$$L_{cox}(\beta) = \prod_{i \in \Omega} \frac{\exp(\beta' X_i)}{\sum_{j \in R_i} \exp(\beta' X_j)} \quad (\text{M7.3})$$

Here under **M7.3**, R_i is a set of all patients who have not failed or been censored but are at risk of failure at that time (i.e. all patients who are still under observation just before time t). The traditional analysis above, under **M7.1** uses only the information from the sampled cases and controls such that the nested case-control study can sometimes be considered as a biased sample from the cohort. Weighting the partial likelihood with the inverse of the inclusion probability of being sampled mitigates this problem. The weighted Cox model partial likelihood can be written as

$$L_{wcox}(\beta) = \prod_{i \in \Omega} \frac{\exp(\beta' X_i) \omega_i}{\sum_{j \in S_i} \exp(\beta' X_j) \omega_j} \quad (\text{M7.4})$$

Where $\mathcal{S}_i$ is the set of cases and controls in the nested case-control sample who are at risk at time t_i . The Cox model obtained from maximizing this likelihood (with inverse probability weights) to estimate coefficient is defined as weighted Cox regression model. The weights were defined as $\omega_j = 1/p_j$, where p_j is probability of a woman being in the nested case-control sample. The $p_j = 1$ for all the cases. In this paper, we used Kaplan Meier (KM) method inclusion probabilities. Inclusion probabilities defined by KM method are called design based since they estimate actual sampling of the controls [199]. The R package `multipleNCC`, was used to compute the weights.

We considered both conditional logistic model and weighted Cox model in our analysis. Overall, we fitted both the conditional logistic model (conditioned on risk sets) and weighted Cox regression to both the nested case-control sample with incidence density sampling and then to the nested case-control sample with path sampling. This was done separately for non-adherence and study non-completion outcomes.

In building each model, first, we fitted univariate models to assess the unadjusted effects of the covariates. Then we fitted multivariable model that included the AE exposure and IPTp treatment adjusted for similar set of confounders that were significant in the univariate analysis at 0.25. Using the likelihood ratio test, we also assessed whether inclusion of interaction between the AE occurrence and treatment arm modified the observed effect. All the point estimates from the fitted models are presented with their respective 95% confidence interval. The interpretation of all the final models was made at 0.05 significance level.

7.3 Results

The cohort where our nested case-control sample designs were derived considered 599 women who were assigned to the two arms of the trial. The women in the cohort had a median follow-up time of 116 days (IQR; 97, 130). At enrolment, these women had a mean gestational age of 22.1 (SD=2.1) weeks and a mean haemoglobin concentration of 11.7 (SD=1.3) g/dL (Table 7.1). A total of 1545 AEs occurred among 474 (79.0%) study participants over the follow-up time where women in the IPTp-CQ arm had a higher cumulative incidence of AEs (895 AEs) than the women in the IPTp-SP arm (650 AEs). Time to first AE was consistently higher in IPTp-CQ arm compared to IPTp-SP over the follow-up time (Fig 7.1) such that women in the IPTp-CQ arm were 29% more likely to experience at least one AE compared to women in the IPTp-SP arm (HR: 1.29; 95% CI: 1.07, 1.57; $p=0.008$). Further details of the AE distribution are provided in appendix 1, Tables S7.1 and S7.2.

Figure 7.1: Time to first AE occurrence among the pregnant women on IPTp up to delivery

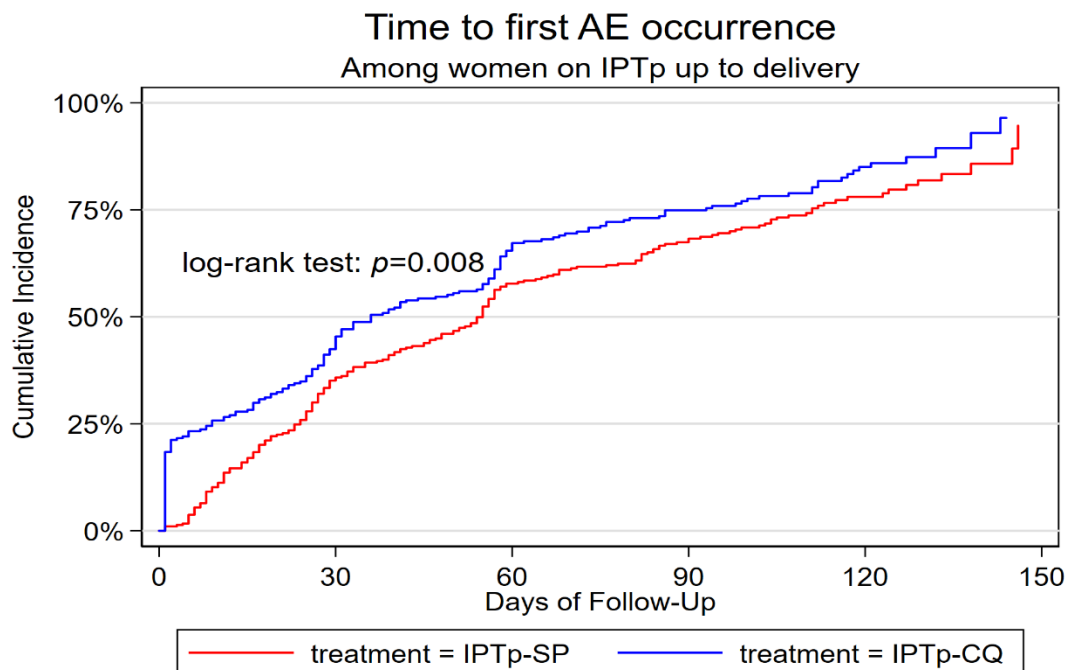


Table 7.1: Distribution of the study cohort characteristics (n=600)

Characteristic	Mean (SD) or n (%)
Maternal age(years)	21.1 (3.2)
Gestational age(weeks)	22.1 (2.1)
BMI (kg/m ²)	23.7 (3.1)
Haemoglobin(g/dl)	11.7 (1.3)
Bed net use n (%)	450 (75.0)
Non-Primigravid n (%)	258 (43.1)
AE occurrence* n (%)	474 (79.0)
Treatment n (%)	
IPTp-SP	300 (50.0)
IPTp-CQ	300 (50.0)

For the study non-completion outcome, 64 (10.7%) women did not complete the study, with 37 (12.3%) in the IPTp-SP arm and 27(9.0%) in the IPTp-CQ arm. Based on the matching criteria defined in the methods section (matching each case to four controls), we selected a total of 256 study non-completion controls. For the treatment non-adherence outcome, 106 (17.7%) women were non-adherent to trial medication, at least once during study follow-up, with 40(13.3%) in the IPTp-SP arm and 66 (22%) in the IPTp-CQ arm. After each incident case was matched to four controls, we yielded a total of 424 treatment non-adherence controls. These reported results of matching were obtained separately for the nested case-control with incidence density sampling and the nested case-control with path sampling.

7.3.1 Comparison of parameter estimates

We report the hazard ratio (HR) estimates on exponentiated scale (not logarithm scale), in Tables 7.2 and 7.3. However, the standard errors are reported on logarithm scale $\log HR$ (SE), based on the corresponding log hazard ratios. The nested case-control is abbreviated as NCC in the tables 7.2 and 7.3. Focusing on the effect of AEs on both treatment non-adherence and study non-completion, our estimates (Tables 7.2 and 7.3) consistently indicate that path sampling yielded similar hazard ratio estimates compared to those from the incidence density sampling. Comparing the weighted Cox regression model and conditional logistic regression model, the weighted Cox model yielded similar standard errors compared to those from conditional logistic regression across both study designs and both outcomes. Therefore, overall, we obtained similar estimates of the effect

of AEs on study non-completion and treatment non-adherence under both incidence density sampling and path sampling regardless of the analysis method used.

Table 7.2: Association between AE occurrence and treatment non-adherence; comparison of estimates from risk set sampling and path sampling under multivariable conditional logistic regression model and weighted Cox regression model

Characteristic	Conditional logistic regression model						Weighted Cox regression model					
	NCC with risk set sampling			NCC with Path sampling			NCC with risk set sampling			NCC with Path sampling		
	HR (95% CI)	LogHR SE	P-value	HR (95% CI)	LogHR SE	P-value	HR (95% CI)	LogHR SE	P-value	HR (95% CI)	LogHR SE	P-value
Treatment												
<i>IPTp-SP</i>	Reference											
<i>IPTp-CQ</i>	2.07 (1.32, 3.22)	0.23	0.001	2.05 (1.30, 3.23)	0.23	0.002	2.04 (1.31, 3.16)	0.22	0.001	2.08 (1.32, 3.27)	0.23	0.002
AE occurrence	0.70 (0.42, 1.17)	0.26	0.175	0.68 (0.41, 1.13)	0.26	0.137	0.67 (0.43, 1.00)	0.22	0.051	0.68 (0.44, 1.06)	0.23	0.087
Age	0.96 (0.89, 1.02)	0.03	0.190	0.94 (0.88, 1.01)	0.04	0.105	0.96 (0.90, 1.02)	0.03	0.179	0.95 (0.89, 1.02)	0.03	0.125

LogHR SE represents standard errors on logarithm scale of hazard ratio, NCC represents nested case-control

Table 7.3: Association between AE occurrence and study non-completion; comparison of estimates from risk set sampling and path sampling under multivariable Conditional logistic regression model and weighted Cox regression model

	Conditional logistic regression model						Weighted Cox regression model					
	NCC with risk set sampling			NCC with Path sampling			NCC with risk set sampling			NCC with Path sampling		
	HR (95% CI)	LogHR SE	P-value	HR (95% CI)	LogHR SE	P-value	HR (95% CI)	LogHR SE	P-value	HR (95% CI)	LogHR SE	P-value
AE occurrence	1.02 (0.56, 1.83)	0.30	0.955	0.85 (0.45, 1.60)	0.32	0.619	0.88 (0.51, 1.52)	0.28	0.648	0.82 (0.47, 1.43)	0.29	0.477
Treatment	0.96 (0.55, 1.67)	0.28	0.896	0.68 (0.39, 1.18)	0.28	0.173	0.96 (0.55, 1.69)	0.29	0.889	0.67 (0.38, 1.18)	0.29	0.162
Age	0.92 (0.83, 1.02)	0.05	0.104	0.93 (0.84, 1.01)	0.05	0.098	0.93 (0.85, 1.01)	0.04	0.095	0.94 (0.87, 1.01)	0.04	0.072

LogHR SE represents standard errors on logarithm scale of hazard ratio, NCC represents nested case-control

We observed no statistically significant effect of AE occurrence on treatment non-adherence under both incidence density sampling and path sampling regardless of the method of analysis used. Using conditional logistic regression, the effect of AE occurrence on treatment non-adherence odds ratio was 0.70 (95% CI: 0.42, 1.17; $p=0.175$) under incidence density sampling and under path sampling the hazard ratio was 0.68 (95% CI: 0.41, 1.13; $p=0.137$). Using the weighted Cox regression, we observed a negligible departure from the overall trend where the path sampling yielded a slightly higher estimate of hazard ratio (HR=0.68; 95% CI: 0.44, 1.06), compared to the incidence density sampling (HR=0.67; 95% CI: 0.43, 1.00). The negligible exception could be explained as one of the expected scenarios that can arise due to sampling variation.

Similarly, we observed no effect of AE occurrence on study non-completion under both incidence density sampling and path sampling regardless of the method of analysis used (see Table 7.3). Using conditional logistic regression, the effect of AE occurrence on study non-completion hazard ratio was 1.02 (95% CI: 0.56, 1.83; $p=0.955$) under incidence density sampling and under path sampling the hazard ratio was 0.85 (95% CI: 0.45, 1.60; $p=0.619$). The weighted Cox regression model yielded hazard ratio of 0.88 (95% CI: 0.51, 1.52, $p=0.648$) under incidence density sampling and 0.82 (95% CI: 0.47, 1.43, $p=0.477$) under path sampling.

Among the adjusted potential confounders for the treatment non-adherence outcome, only treatment was associated with treatment non-adherence (Table 7.3). We observed two times increased risk of treatment non-adherence for patient in IPTp-CQ

treatment compared those in IPTp-SP treatment. Using conditional logistic regression, the observed odds ratio was 2.07 (95% CI: 1.32, 3.22; $p=0.001$) under incidence density sampling and we observed a similar odds ratio (OR=2.05; 95% CI: 1.30, 3.23; $p=0.002$) under path sampling. Weighted Cox regression model yielded similar hazard ratios; (HR=2.04; 95% CI: 1.31,3.16; $p=0.001$) under incidence density sampling and (HR=2.08; 95% CI: 1.32,3.27; $p=0.002$) under path sampling.

7.4 Discussion

In this study, we empirically compared the efficiency of incidence density sampling and path sampling for nested case-control studies under conditional logistic regression and weighted Cox regression analysis methods. The practical application focussed on determining the impact of AEs on treatment non-adherence or discontinuation. Practically, we established that the standard errors and hazard ratios for the two designs were similar. However, it should be noted that path sampling (by design) is based on more unique controls selected implying that it naturally uses a larger sample of unique patients to achieve those similar standard errors as compared to incidence density sampling. This knowledge gained is directly useful in improving nested case-control studies that have recently gained an interest in safety studies[200, 201]. This study provides insights regarding choice of nested case-control design and the burden of AEs on treatment non-adherence and study non-completion in IPTp trials. Such insights may also facilitate understanding of expected patient safety in the post-trial period when the drug is launched [202].

The negligible gain in efficiency of weighted Cox regression model over conditional regression can be attributed to the homogeneity of the cohort that we used. Since our

cohort comes from a randomized clinical trial whose baseline characteristics are similar, breaking the match in order to use more controls could not add much information to our nested case-control sample that was already using over 50% of the cohort sample. Most of the substantial gains for the inverse probability weighting have been established under large sample properties, especially in observational studies which by design exhibit heterogeneous population unlike clinical trials.

From practical perspective, based on the example data that we used, this study did not establish evidence of a significant role of AEs on IPTp treatment non-adherence and study discontinuation. Neither, conditional logistic regression model nor the weighted Cox regression model under incidence density sampling or path sampling suggested that AE occurrence impacted treatment non-adherence or study non-completion. Our findings are contrary to the hypothesized direction of AEs effect where we postulated that AE occurrence increases the risk of IPTp non-adherence and IPTp trial discontinuation. However, our findings support what has been previously observed in IPTp context whereby frequent adverse events do not reduce adherence [203]. Our exploratory finding is useful towards ascertaining the validity of IPTp trial designs and interpretation of findings.

We observed increased risk of treatment non-adherence for the women in IPT-CQ treatment arm who were two times more likely to non-adhere compared to those in IPTp-SP arm. This is expected since the women in IPTp-CQ arm had more opportunities for non-adherence as they took four more doses compared to those on IPTp-SP arm. This agrees with previous studies that indicate that number of doses for

a given regimen is inversely associated with treatment non-adherence [204]; increasing number of days and doses for taking a particular regimen can lead to low treatment adherence [205, 206]. This has consequences on the IPTp trial designs and packaging of treatment doses.

Use of the nested case-control study design in the context of clinical trial data provided a great opportunity to assess the temporal relationship of adverse events and adherence or study non-completion [207]. This further enabled us to make causal interpretations since the time-dependent AE exposure was prospectively measured and the nested case-control sampling approach reduces selection bias. Although Cox proportional hazards model with time-dependent covariates can alternatively yield similar estimates to those obtained through nested case-control approach, the nested case control is known to be more computationally efficient [208] . Despite the fact that our work may not be generalizable to settings where there are non-repeated doses, it still offers a groundwork for improved antimalarial drug safety assessment framework (accounting for non-adherence). Such improved assessment of the relationship between non-adherence to treatment and safety outcomes is crucial in planning for improved interventions effectiveness [93].

This study had some limitations. Firstly, the differences in treatment by dose schedule between the treatment arms could affect the observed differences. Secondly, we did not collect data on patient satisfaction. This made it hard it to make a conclusive statement on the unexpected observed effect of the AEs. Since non-adherence can be affected by the patient behaviour, collection of such patient satisfaction and

perception in clinical trials would be helpful in elucidating the effect of AEs on non-adherence. Furthermore, the methodological conclusions in this study can be improved and extended by undertaking a full simulation study to understand the performance of the sampling and analysis methods under varying known scenarios.

7.5 Conclusion

Assessment of safety in relation to IPTp adherence, even for less severe AEs, is an important topic in developing solutions towards improved IPTp safety profile and risk management plan in clinical trials. Nested case-control study can be considered as a useful tool in investigating the impact of AEs on treatment non-adherence and study non-completion because it offers an opportunity for an efficient sample where cases and controls are comparable, accounting for potential differences in follow-up time. We empirically demonstrated that incidence density sampling and path sampling practically yield similar estimates whether one uses weighted Cox regression or conditional logistic regression. However, by design, path sampling yields those similar estimates at the expense of more unique controls sampled compared to the incidence density sampling. The clinical findings indicated that AE occurrence did not impact treatment non-adherence or study non-completion. The insights from this study inform future trial designs and drug safety assessment. Future extension of this work can consider evaluating the impact of AEs on treatment non-adherence and study non-adherence accounting for patient perception on IPTp benefits since treatment non-adherence is associated with patient behaviour.

Chapter 8: Discussion

This chapter provides an overall summary of the PhD study findings. Specifically, the general discussion in this section focusses on the key findings, study strengths and weakness, potential future work and recommendations.

8.1 Summary of key findings

Drug safety assessment is very important at every stage of drug development. Randomized controlled trials provide a gold standard for evaluating drug efficacy and safety such that safety data collected during premarketing drug trials provide the crucial basis for drug approval and labelling. However, analysis of the safety data is more challenging compared to efficacy data such that statistical methodology for safety data in clinical trials has received limited attention. Key challenges in analysis of safety data includes;

- a) Lack of pre-specified safety analysis plan for safety outcomes since most safety outcomes are unexpected and most clinical trials are rarely powered to detect safety effect [5]
- b) Competing risks, censoring and varying individual follow-up time
- c) Multiple safety outcomes with repeated measurements [5]
- d) Time-dependent treatment effects and unobserved heterogeneity
- e) Concomitant medication, concurrent indication and general health of the study population can confound estimates for safety data. This may induce imbalance between the treatment arms and may affect generalizability of the results since the randomization benefits may be lost.

- f) Lack of standardized guidelines of estimands for quantitative safety assessment given an indication under study [209]

This PhD work focussed on investigating optimal statistical methods for safety data analysis in clinical trials. The thesis outlines key statistical issues for analysis of safety data in clinical trials with antimalarial drug in pregnancy trials as a key example. This was followed by a set of solutions on how to deal with some of the problems such as hierarchical structure of the outcomes, multiple repeated outcomes and time-dependent effects. PREGACT trial and Chloroquine for Malaria Prevention in Pregnancy trial data were used to demonstrate the proposed set of solutions.

Specifically, this PhD study firstly systematically reviewed statistical methods applied towards analysis of safety data in malaria chemoprevention trials during pregnancy. The review focussed on the prevalence of the methods, their appropriateness and recommendations for improved analyses. Empirically, nested case control with incidence sampling was compared with nested case control with path set sampling under conditional logistic regression and weighted cox regression in investigating the burden of AEs in clinical trials. Then using a simulation study, flexible parametric shared frailty models were assessed if they could improve estimates in analysis recurrent adverse events since they account for both unobserved heterogeneity and time-dependent effects (unlike conventional shared parametric methods that do not account for time-dependent effects and are not flexible). For both the empirical and simulation-based studies above, chloroquine in pregnancy data was used as the motivating data. Finally, a novel joint model was developed aimed at improving AE

incidence estimates; the joint model included recurrent AEs, concomitant medication and repeatedly measured biochemical information outcomes adjusted for baseline characteristics[210]. PREGACT trial data was used to demonstrate the utility of the developed model.

The findings of the PhD work are discussed in the context of the objective-specific studies in the preceding paragraph.

8.1.1 Current practice in analysis of safety data

In this PhD study, a review was carried out to understand the current practice and gaps in the statistical analysis of safety data in clinical trials, focussing on the setting of drug trials for prevention of malaria in pregnancy as a key example. The review established that statistical analysis and reporting of safety data remains suboptimal. Descriptive methods (e.g. Chi-Square) and dichotomization of continuous outcomes were more prevalent. Such approaches are limited in analysis of safety in clinical trials data since they lead to loss of important information by not accounting for variability in follow-up time, censoring and correlation over time for repeatedly measured multiple outcomes and potential confounders. The findings agree with some observations of previous work in literature that has called for overall improved quantitative drug safety assessment and reporting in clinical trials [6, 70]. The choice of a statistical method(s) to use and reporting for safety data should be jointly driven by the scientific question of interest and complexity of the raw safety data. Development and application of improved statistical methods, formulation of standardized guidelines supported by drug regulatory bodies involvement can enhance improved analysis and reporting of safety data.

8.1.2 Recurrent AEs with time-dependent effects and unobserved heterogeneity

The PhD study also compared shared frailty models to establish the evidence-base for an efficient model that can account for both time-dependent effects and unobserved heterogeneity encountered in the recurrent AEs analysis. Flexible parametric modelling for shared frailty models were found appropriate in estimating unbiased hazard ratios compared to the standard shared frailty modelling approach in the presence of time-dependent effects and unobserved heterogeneity. The inverse Gaussian shared frailty model with proportional hazard assumption was found more appropriate in estimating unbiased unobserved heterogeneity. As observed in the analysis of univariate survival data, the flexible parametric survival models can provide improved treatment estimates in analysis of multivariate survival data compared to the standard survival models [146].

8.1.3 Multiple safety outcomes with repeated measurements analysis

The PhD study further introduced a framework for novel joint modelling of multivariate continuous longitudinal clinical laboratory safety outcomes, concomitant medication counts and adverse event count, exploiting the advantages of the existing extended multivariate generalised linear and non-linear mixed effects models framework[86]. The proposed joint model improved AE incidence estimates underestimated by the conventional Poisson model. Concomitant medication was strongly associated with the clinical AEs incidence, suggesting that the patients who experienced more AEs were more likely to report the more concomitant medications. The proposed joint

modelling approach efficiently harnesses multiple outcomes safety data for improved drug safety profile development. Furthermore, this modelling approach offers the basis for choosing the most appropriate biochemical markers in assessing drug safety especially in resource constrained settings.

8.1.4 Impact of AEs in clinical trials

Empirical comparison of the efficiency of incidence density sampling and path set sampling for nested case control studies under conditional logistic regression and weighted cox regression models demonstrated that incidence density sampling is similar in efficiency of estimates compared to path set sampling. These findings agree with previous work in literature that has also shown that nested case control with incidence density sample is a robust sampling approach for cohorts [188]; there is no evidence that the other alternative sampling strategies are more superior compared to the incidence density sampling. There was no difference in estimates for using either conditional logistic regression or weighted cox regression probably due to the homogeneity of the cohort since it originated from a clinical trial.

Practically, it we also found no evidence that AEs did impact non-adherence or study discontinuation of the patients. This knowledge is directly beneficial in interpreting the drug safety data and deciding on the most appropriate clinical trial design future studies.

8.2 Study strengths

The use of multifaceted study approaches ranging from systematic review, nested case control, simulation and advanced modelling approaches provides an opportunity to leverage rich multidimensional data from clinical trials in answering drug safety-related questions. This PhD study also greatly benefited from the use of both theoretical and pragmatic approaches in addressing the issues associated with drug safety data analysis. For example, the use of simulation studies enabled addressing drug safety data analysis issues (unobserved heterogeneity and time-dependent effect) under different scenarios beyond what can be answered using a single real data set. Model complexity may sometimes influence what can be considered as the most optimal approach to analysis of particular data. Use of a simulation study helped us to address this in understanding how flexible parametric shared frailty models compare with the conventional lognormal proportional shared frailty model under known scenarios. Since the data within our simulation study arose from known scenarios, it was easier to objectively determine whether model complexity was necessary based on the performance measures.

The availability of the motivating/example datasets enabled contextualising of some of the suggested solutions; this can ensure wider understanding of implementation and adapting of the solutions in future studies. The motivating datasets also consisted of both long (up to 9-12 months of follow-up time) and short (up to 2-3 months of follow-up time) follow-up trials, further enriching understanding of some issues associated with both long and short follow-up trials.

8.3 Study limitations

The systematic review that was carried out as part of this PhD work may have missed other studies published in languages other than English. Furthermore, since the review focussed only on the studies published from 2010 to 2019, it is hard to capture temporal trends that may have happened over years (considering the years outside 2010-2019) on statistical analysis and reporting of drug safety data in antimalarial drug trials in pregnancy.

Absence of data on patient satisfaction when assessing the impact of AE on non-adherence and study non-completion is a key limitation in this study. Since patient behaviour is usually associated with non-adherence, collecting data on patient satisfaction and perception about clinical trial treatment can improve analysis and interpretation of burden of AEs in clinical trials.

The proposed joint model for analysis of multiple safety outcomes was only empirically considered; no simulation study was carried out to evaluate the performance of the model under other scenarios than the motivating data setting. Although a simulation study was carried out to evaluate shared frailty models for recurrent AEs analysis, the study was limited to scenarios where censoring was assumed non-informative and constant unobserved heterogeneity.

Furthermore, this work is confined to clinical trials with traditional parallel designs. Currently, there is increased innovations in the trial designs; this provides an opportunity to explore alternative approaches to quantitative drug safety assessment

beyond what has been investigated in the current work. For example, sequential analytical methods would be more useful in quantitative drug assessment of adaptive designs unlike static evaluation methods.

Finally, the current work focussed on safety assessment based on a single trial. Plausible alternative extension (but beyond the scope of the current work) would be to borrow accumulating safety evidence across clinical trial phases using Bayesian analytical approaches. Such Bayesian approaches offer an opportunity to update the safety evidence based on accumulating data from new trials.

8.4 Future work

This body of work provides a strong ground for future work on improved drug safety data analysis in clinical trials. The developed joint model and shared models investigated in this PhD study are easily extendable to the context of multicentre clinical trials. Both the joint model and shared frailty model are a special case of hierarchical models where the observations are nested/clustered within patients. In multicentre clinical trials context, there is an additional level of clustering where the patients are also nested within individual centres. In this case, the centres are assumed independent from each other but patients from the same centre will usually have correlated outcomes. This additional level of complexity can be handled by including an additional random effect at study centre level. The inclusion of random to account for study centre increases power especially when there are few patients per centre and imbalance in treatment assignments within centres[211, 212]. However, when there are very few centres, fixed effect should be preferred to account for the

centres for the centre effects. Detailed implementation and discussion of this was beyond the scope of the current work. However, this suggested extension is an interesting future work. An interesting example would be to investigate how such extended models would perform under varying sample sizes, missing data, low AE incidence etc.

Specifically, the introduced joint model considered for analysis of multiple safety outcomes was only studied empirically. Future work can attempt to study the performance under different scenarios including varying data missingness of the biochemical parameters, varying sample sizes and varying AEs incidence. Such study can offer an opportunity for generalization of the model to several scenarios encountered in antimalarial drug safety assessment in clinical trials. In comparing shared frailty models for recurrent AEs analysis, the work was limited to the scenario where the unobserved heterogeneity was constant. However, in practice there is a potential for unobserved heterogeneity to change over the follow-up time. It will be interesting in future work to understand the performance of the shared frailty models under a complex setting where both the treatment effect and unobserved heterogeneity vary over the follow-up time and accounting for the actual medication dates. In assessing the AE burden, collection of data on patient treatment satisfaction and experience after AE occurrence can further strengthen evidence generation on the impact of AEs on treatment non-adherence and study non-completion. Furthermore, it would be interesting to undertake a simulation study in order to formally evaluate performance of the studied NCC designs under the conditional logistic regression and the weighted cox regression. The key performance measures of interest to be formally quantified would be power, bias and precision.

8.5 Conclusions

In conclusion, this thesis highlighted the importance of improved statistical methods for analysis and reporting of drug safety data in clinical trials. The work focussed on investigating methods that account for time-to-event, unobserved heterogeneity, time-dependent effects and potential correlations.

Flexible parametric shared frailty models with non-proportional hazards were found to be more advantageous in capturing time-dependent treatment effects in recurrent AE analysis. Another key lesson drawn from this work is the need to jointly model multiple correlated safety outcomes and other supporting information (e.g. biochemical laboratory parameters, concomitant medications) in order to efficiently quantify the clinical AE incidence. These findings fill some important gaps in literature for improved drug safety data methods in clinical trials. Further drug regulatory bodies commitment by developing standardized guidelines will more likely improve uptake of some of the already existing improved drug safety data analysis and reporting methods, beyond the current practice.

Empirically, this study established that occurrence of AEs does not impact patient participation in Malawi. This can probably be explained by the fact that better healthcare experienced in the clinical trial setting outweighs the AE burden such that patients opt to remain within the clinical trial [213].

8.6 Recommendations

Based on the findings in this work, we provide the following recommendations;

- a) The review conducted in this PhD study found that drug safety data analysis in clinical trials remains suboptimal;
- (i) Investing in development and application of improved statistical methods for safety data analysis and reporting, within the academia, drug regulatory bodies and pharmaceutical industry can enhance improved understanding on how to efficiently address issues encountered in drug safety data analysis.
 - (ii) Instituting of standardized guidelines (by the drug regulatory bodies) for the improved drug safety data analysis and reporting approaches, beyond the current practice and guidelines, can enhance uptake of the existing improved statistical methods.
- b) This work established that joint modelling of drug safety data efficiently improves drug safety estimates since it simultaneously accounts for the relevant multiple outcomes and correlations that can confound estimates. This calls for deliberate efforts to make sure that the important data that can confound safety estimates (e.g. concomitant medication) is collected effectively and rigorously modelled accordingly. The modelling should be driven by the scientific question(s) of interest and the complex raw structure of the data.
- c) In the presence of time-dependent effects, flexible parametric shared frailty models should be preferred over the standard shared frailty models in recurrent AEs analysis. The flexible parametric models efficiently capture the time-dependent effects and any unobserved heterogeneity.

References

1. Munsaka MS: A question-based approach to the analysis of safety data. In: *Biopharmaceutical Applied Statistics Symposium: 2018*: Springer; 2018: 193-216.
2. Chuang-Stein C, Mohberg NR, Musselman DM: Organization and analysis of safety data using a multivariate approach. *Statistics in medicine* 1992, 11(8):1075-1089.
3. Chuang-Stein C, Xia HA: The practice of pre-marketing safety assessment in drug development. *Journal of biopharmaceutical statistics* 2013, 23(1):3-25.
4. Gong Q, Tong B, Strasak A, Fang L: Analysis of safety data in clinical trials using a recurrent event approach. *Pharmaceutical statistics* 2014, 13(2):136-144.
5. Zink RC, Marchenko O, Sanchez-Kam M, Ma H, Jiang Q: Sources of Safety Data and Statistical Strategies for Design and Analysis: Clinical Trials. *Therapeutic innovation & regulatory science* 2018, 52(2):141-158.
6. Phillips R, Hazell L, Sauzet O, Cornelius V: Analysis and reporting of adverse events in randomised controlled trials: a review. *BMJ open* 2019, 9(2):e024537.
7. O'Neill RT: Regulatory perspectives on data monitoring. *Statistics in medicine* 2002, 21(19):2831-2842.
8. Vital Signs [<https://my.clevelandclinic.org/health/articles/10881-vital-signs>]
9. Understanding Laboratory Tests [<https://www.cancer.gov/about-cancer/diagnosis-staging/understanding-lab-tests-fact-sheet#q1>]
10. Investigational New Drug Application Sub Part B Sec 312.32 IND Safety Reporting [<https://www.accessdata.fda.gov/scripts/cdrh/cfdocs/cfcfr/CFRsearch.cfm?CFRPart=312>]
11. Berry SM, Berry DA: Accounting for multiplicities in assessing drug safety: a three-level hierarchical mixture model. *Biometrics* 2004, 60(2):418-426.
12. Mozzicato P: MedDRA. *Pharmaceutical Medicine* 2009, 23(2):65-75.
13. Hodel EM, Kabanywany AM, Malila A, Zanolari B, Mercier T, Beck HP, Buclin T, Olliaro P, Decosterd LA, Genton B: Residual antimalarials in malaria patients from Tanzania--

- implications on drug efficacy assessment and spread of parasite resistance. *PloS one* 2009, 4(12):e8184.
14. Allen EN, Mushi AK, Massawe IS, Vestergaard LS, Lemnge M, Staedke SG, Mehta U, Barnes KI, Chandler CI: How experiences become data: the process of eliciting adverse event, medical history and concomitant medication reports in antimalarial and antiretroviral interaction trials. *BMC medical research methodology* 2013, 13:140.
 15. McBride WG: Thalidomide and congenital anomalies. *The Lancet* 1961, 278(7216):1358.
 16. Ward SP: Thalidomide and Congenital Abnormalities. *British Medical Journal* 1962, 2(5305):646-647.
 17. Lenz W, Pfeiffer RA, Kosenow W, Hayman DJ: Thalidomide and congenital anomalies. *The Lancet*, 279(7219):45-46.
 18. Otterland A: Effect of Thalidomide Warning on Medicine prescribed for Pregnant Women: Analysis of Free Medicines Delivered to Expectant Mothers in the Västerbotten County of Sweden during 1961 and 1962. *British Journal of Preventive & Social Medicine* 1965, 19(2):94-96.
 19. Nordeng H, Ystrøm E, Einarson A: Perception of risk regarding the use of medications and other exposures during pregnancy. *European Journal of Clinical Pharmacology* 2010, 66(2):207-214.
 20. Organization WH: World malaria report 2021. 2021.
 21. Dellicour S, Tatem AJ, Guerra CA, Snow RW, ter Kuile FO: Quantifying the number of pregnancies at risk of malaria in 2007: a demographic study. *PLoS medicine* 2010, 7(1):e1000221.
 22. Rogerson SJ, Desai M, Mayor A, Sicuri E, Taylor SM, van Eijk AM: Burden, pathology, and costs of malaria in pregnancy: new developments for an old problem. *The Lancet Infectious diseases* 2018, 18(4):e107-e118.

23. Radeva-Petrova D, Kayentao K, ter Kuile FO, Sinclair D, Garner P: Drugs for preventing malaria in pregnant women in endemic areas: any drug regimen versus placebo or no treatment. *The Cochrane database of systematic reviews* 2014(10):Cd000169.
24. Nosten F, McGready R, d'Alessandro U, Bonell A, Verhoeff F, Menendez C, Mutabingwa T, Brabin B: Antimalarial drugs in pregnancy: a review. *Current drug safety* 2006, 1(1):1-15.
25. D'Alessandro U, Hill J, Tarning J, Pell C, Webster J, Gutman J, Sevene E: Treatment of uncomplicated and severe malaria during pregnancy. *The Lancet Infectious Diseases* 2018, 18(4):e133-e146.
26. Saito M, Gilder ME, Nosten F, Guérin PJ, McGready R: Methodology of assessment and reporting of safety in anti-malarial treatment efficacy studies of uncomplicated falciparum malaria in pregnancy: a systematic literature review. *Malaria journal* 2017, 16(1):491.
27. Desai M, Gutman J, Taylor SM, Wiegand RE, Khairallah C, Kayentao K, Ouma P, Coulibaly SO, Kalilani L, Mace KE *et al*: Impact of Sulfadoxine-Pyrimethamine Resistance on Effectiveness of Intermittent Preventive Therapy for Malaria in Pregnancy at Clearing Infections and Preventing Low Birth Weight. *Clinical infectious diseases : an official publication of the Infectious Diseases Society of America* 2016, 62(3):323-333.
28. Gutman J, Mwandama D, Wiegand RE, Ali D, Mathanga DP, Skarbinski J: Effectiveness of intermittent preventive treatment with sulfadoxine-pyrimethamine during pregnancy on maternal and birth outcomes in Machinga district, Malawi. *The Journal of infectious diseases* 2013, 208(6):907-916.
29. Briand V, Bottero J, Noël H, Masse V, Cordel H, Guerra J, Kossou H, Fayomi B, Ayemonna P, Fievet N: Intermittent treatment for the prevention of malaria during pregnancy in Benin: a randomized, open-label equivalence trial comparing sulfadoxine-pyrimethamine with mefloquine. *The Journal of infectious diseases* 2009, 200(6):991-1001.
30. González R, Mombo-Ngoma G, Ouédraogo S, Kakolwa MA, Abdulla S, Accrombessi M, Aponte JJ, Akerey-Diop D, Basra A, Briand V: Intermittent preventive treatment of malaria in

- pregnancy with mefloquine in HIV-negative women: a multicentre randomized controlled trial. *PLoS medicine* 2014, 11(9):e1001733.
31. Kimani J, Phiri K, Kamiza S, Duparc S, Ayoub A, Rojo R, Robbins J, Orrico R, Vandenbroucke P: Efficacy and Safety of Azithromycin-Chloroquine versus Sulfadoxine-Pyrimethamine for Intermittent Preventive Treatment of Plasmodium falciparum Malaria Infection in Pregnant Women in Africa: An Open-Label, Randomized Trial. *PloS one* 2016, 11(6):e0157045.
 32. Luntamo M, Kulmala T, Cheung YB, Maleta K, Ashorn P: The effect of antenatal monthly sulphadoxine–pyrimethamine, alone or with azithromycin, on foetal and neonatal growth faltering in Malawi: a randomised controlled trial. *Tropical Medicine & International Health* 2013, 18(4):386-397.
 33. Luntamo M, Kulmala T, Mbewe B, Cheung YB, Maleta K, Ashorn P: Effect of repeated treatment of pregnant women with sulfadoxine-pyrimethamine and azithromycin on preterm delivery in Malawi: a randomized controlled trial. *The American Journal of Tropical Medicine and Hygiene* 2010, 83(6):1212-1220.
 34. Van den Broek NR, White SA, Goodall M, Ntonya C, Kayira E, Kafulafula G, Neilson JP: The APLe study: a randomized, community-based, placebo-controlled trial of azithromycin for the prevention of preterm birth, with meta-analysis. *PLoS medicine* 2009, 6(12):e1000191.
 35. Clerk CA, Bruce J, Affipunguh PK, Mensah N, Hodgson A, Greenwood B, Chandramohan D: A randomized, controlled trial of intermittent preventive treatment with sulfadoxine-pyrimethamine, amodiaquine, or the combination in pregnant women in Ghana. *Journal of Infectious Diseases* 2008, 198(8):1202-1211.
 36. Hengelbrock J, Gillhaus J, Kloss S, Leverkus F: Safety data from randomized controlled trials: applying models for recurrent events. *Pharmaceutical statistics* 2016, 15(4):315-323.
 37. Allignol A, Beyersmann J, Schmoor C: Statistical issues in the analysis of adverse events in time-to-event data. *Pharmaceutical statistics* 2016, 15(4):297-305.

38. Bender R, Beckmann L, Lange S: Biometrical issues in the analysis of adverse events within the benefit assessment of drugs. *Pharmaceutical statistics* 2016, 15(4):292-296.
39. Megari K: Quality of Life in Chronic Disease Patients. *Health Psychology Research* 2013, 1(3):e27.
40. Dabhade S, Pandit P, Tiwari S, Pandit D, Chopra A: Reasons for patient's withdrawal from clinical trials of rheumatoid arthritis in India. *Journal of Orthopaedics and Allied Sciences* 2015, 3(1):8-11.
41. Little RJ, D'Agostino R, Cohen ML, Dickersin K, Emerson SS, Farrar JT, Frangakis C, Hogan JW, Molenberghs G, Murphy SA *et al*: The Prevention and Treatment of Missing Data in Clinical Trials. *The New England journal of medicine* 2012, 367(14):1355-1360.
42. Moore RA, Derry S, McQuay HJ: Discontinuation rates in clinical trials in musculoskeletal pain: meta-analysis from etoricoxib clinical trial reports. *Arthritis Research & Therapy* 2008, 10(3):R53-R53.
43. Foulkes MA: Safety Assessment Versus Efficacy Assessment. In: *Advances in Statistical Methods for the Health Sciences: Applications to Cancer and AIDS Studies, Genome Sequence Analysis, and Survival Analysis*. edn. Edited by Auget J-L, Balakrishnan N, Mesbah M, Molenberghs G. Boston, MA: Birkhäuser Boston; 2007: 323-334.
44. Onakpoya IJ, Heneghan CJ, Aronson JK: Post-marketing withdrawal of 462 medicinal products because of adverse drug reactions: a systematic review of the world literature. *BMC Medicine* 2016, 14(1):10.
45. Chuang-Stein C: Laboratory data in clinical trials: a statistician's perspective. *Controlled clinical trials* 1998, 19(2):167-177.
46. Siddiqui O: Statistical methods to analyze adverse events data of randomized clinical trials. *Journal of biopharmaceutical statistics* 2009, 19(5):889-899.
47. Prins TJ, Rollemans C, van Roon E, de Vries T: Improved quality of reporting safety data of medication in paediatric randomised controlled trials. *Archives of Disease in Childhood* 2021.

48. Ioannidis JP, Evans SJ, Gøtzsche PC, O'Neill RT, Altman DG, Schulz K, Moher D: Better reporting of harms in randomized trials: an extension of the CONSORT statement. *Annals of internal medicine* 2004, 141(10):781-788.
49. Crowe BJ, Xia HA, Berlin JA, Watson DJ, Shi H, Lin SL, Kuebler J, Schriver RC, Santanello NC, Rochester G: Recommendations for safety planning, data collection, evaluation and reporting during drug, biologic and vaccine development: a report of the safety planning, evaluation, and reporting team. *Clinical Trials* 2009, 6(5):430-440.
50. Commission E: Communication from the Commission—Detailed guidance on the collection, verification and presentation of adverse event/reaction reports arising from clinical trials on medicinal products for human use ('CT-3'). *Off J Eur Union* 2011, 172:1-13.
51. Food U, Administration D: Guidance for industry: safety assessment for IND safety reporting (draft). In.: December; 2015.
52. Wang W, Whalen E, Munsaka M, Li J, Fries M, Kracht K, Sanchez-Kam M, Singh K, Zhou K: On quantitative methods for clinical safety monitoring in drug development. *Statistics in Biopharmaceutical Research* 2018, 10(2):85-97.
53. Herson J: Safety monitoring. In: *Statistical Methods for Evaluating Safety in Medical Product Development*. edn.; 2014: 293-318.
54. Phillips R, Sauzet O, Cornelius V: Statistical methods for the analysis of adverse event data in randomised controlled trials: a scoping review and taxonomy. *BMC medical research methodology* 2020, 20(1):288.
55. Cornelius V, Cro S, Phillips R: Advantages of visualisations to evaluate and communicate adverse event information in randomised controlled trials. *Trials* 2020, 21(1):1-12.
56. Amit O, Heiberger RM, Lane PW: Graphical approaches to the analysis of safety data from clinical trials. *Pharmaceutical statistics* 2008, 7(1):20-35.

57. Agresti A, Klingenberg B: Multivariate tests comparing binomial probabilities, with application to safety studies for drugs. *Journal of the Royal Statistical Society: Series C (Applied Statistics)* 2005, 54(4):691-706.
58. Jiang Q, Xia HA: Quantitative evaluation of safety in drug development: Design, analysis and reporting: CRC Press; 2014.
59. Amy Xia H, Ma H, Carlin BP: Bayesian Hierarchical Modeling for Detecting Safety Signals in Clinical Trials. *Journal of biopharmaceutical statistics* 2011, 21(5):1006-1029.
60. Schildcrout JS, Jenkins CA, Ostroff JH, Gillen DL, Harrell FE, Trost DC: Analysis of longitudinal laboratory data in the presence of common selection mechanisms: a view toward greater emphasis on pre-marketing pharmaceutical safety. *Statistics in medicine* 2008, 27(12):2248-2266.
61. Southworth H, Heffernan JE: Multivariate extreme value modelling of laboratory safety data from clinical studies. *Pharmaceutical statistics* 2012, 11(5):367-372.
62. Ying Z, Chunlei K, Qi J, Seta S, Steven S: Choosing Appropriate Metrics to Evaluate Adverse Events in Safety Evaluation. *Therapeutic innovation & regulatory science* 2015, 49(3):398-404.
63. Hammad TA, Pinheiro SP, Neyarapally GA: Secondary use of randomized controlled trials to evaluate drug safety: a review of methodological considerations. *Clinical trials (London, England)* 2011, 8(5):559-570.
64. Le Noury J, Nardo JM, Healy D, Jureidini J, Raven M, Tufanaru C, Abi-Jaoude E: Restoring Study 329: efficacy and harms of paroxetine and imipramine in treatment of major depression in adolescence. *BMJ* 2015, 351.
65. Mathanga DP, Walker ED, Wilson ML, Ali D, Taylor TE, Laufer MK: Malaria control in Malawi: current status and directions for the future. *Acta tropica* 2012, 121(3):212-217.
66. Keller MB, Ryan ND, Strober M, Klein RG, Kutcher SP, Birmaher B, Hagino OR, Koplewicz H, Carlson GA, Clarke GN *et al*: Efficacy of paroxetine in the treatment of adolescent major

- depression: a randomized, controlled trial. *Journal of the American Academy of Child and Adolescent Psychiatry* 2001, 40(7):762-772.
67. Little RJA, Rubin DB: Statistical Analysis With Missing Data: Wiley; 1987.
 68. Myers WR: Handling Missing Data in Clinical Trials: An Overview. *Drug Information Journal* 2000, 34(2):525-533.
 69. Gupta SK: Intention-to-treat concept: A review. *Perspectives in Clinical Research* 2011, 2(3):109-112.
 70. Singh S, Loke YK: Drug safety assessment in clinical trials: methodological challenges and opportunities. *Trials* 2012, 13(1):138.
 71. Sutton AJ, Cooper NJ, Lambert PC, Jones DR, Abrams KR, Sweeting MJ: Meta-analysis of rare and adverse event data. *Expert review of pharmacoeconomics & outcomes research* 2002, 2(4):367-379.
 72. Ambler G, Seaman S, Omar RZ: An evaluation of penalised survival methods for developing prognostic models with rare events. *Statistics in medicine* 2012, 31(11-12):1150-1161.
 73. Gibbons RD, Amatya A: Statistical Methods for Drug Safety: CRC Press; 2015.
 74. Jones CE, Munoz FM, Spiegel HM, Heininger U, Zuber PL, Edwards KM, Lambach P, Neels P, Kohl KS, Gidudu J: Guideline for collection, analysis and presentation of safety data in clinical trials of vaccines in pregnant women. *Vaccine* 2016, 34(49):5998-6006.
 75. Saito M, Gilder ME, Nosten F, Guérin PJ, McGready R: Methodology of assessment and reporting of safety in anti-malarial treatment efficacy studies of uncomplicated falciparum malaria in pregnancy: a systematic literature review. *Malaria journal* 2017, 16(1):1-10.
 76. Singh S, Loke YK: Drug safety assessment in clinical trials: methodological challenges and opportunities. *Trials* 2012, 13:138-138.
 77. Keogh RH, Cox DR: Case-Control Studies. Cambridge: Cambridge University Press; 2014.

78. Cabarro B, Gomez-Roca C, Viala M, Rabeau A, Paulon R, Loirat D, Munsch N, Delord J-P, Filleron T: Modernizing adverse events analysis in oncology clinical trials using alternative approaches: rationale and design of the MOTIVATE trial. *Investigational new drugs* 2020.
79. Hengelbrock J, Gillhaus J, Kloss S, Leverkus F: Safety data from randomized controlled trials: applying models for recurrent events. *Pharmaceutical Statistics* 2016, 15(4):315-323.
80. Divala TH, Mungwira RG, Mawindo PM, Nyirenda OM, Kanjala M, Ndaferankhande M, Tsirizani LE, Masonga R, Muwalo F, Potter GE *et al*: Chloroquine as weekly chemoprophylaxis or intermittent treatment to prevent malaria in pregnancy in Malawi: a randomised controlled trial. *The Lancet Infectious Diseases* 2018, 18(10):1097-1107.
81. Pekyi D, Ampromfi AA, Tinto H, Traoré-Coulibaly M, Tahita MC, Valéa I, Mwapasa V, Kalilani Phiri L, Kalanda G, Madanitsa M *et al*: Four Artemisinin-Based Treatments in African Pregnant Women with Malaria. *N Engl J Med* 2016, 374(10):913-927.
82. Colditz GA, Emerson JD: The statistical content of published medical research: some implications for biomedical education. *Medical education* 1985, 19(3):248-255.
83. Emerson JD, Colditz GA: Use of statistical analysis in the New England Journal of Medicine. *N Engl J Med* 1983, 309(12):709-713.
84. Otworld KN, Petzold M, Martinson N, Chirwa T: A review of the study designs and statistical methods used in the determination of predictors of all-cause mortality in HIV-infected cohorts: 2002-2011. *PloS one* 2014, 9(2):e87356.
85. Crowther MJ, Lambert PC: Simulating biologically plausible complex survival data. *Statistics in medicine* 2013, 32(23):4118-4134.
86. Crowther MJ: Extended multivariate generalised linear and non-linear mixed effects models. *arXiv preprint arXiv:171002223* 2017.
87. Langholz B, Thomas DC: Efficiency of cohort sampling designs: some surprising results. *Biometrics* 1991, 47(4):1563-1571.

88. Tamminga C, Kavanaugh M, Fedders C, Maiolatesi S, Abraham N, Bonhoeffer J, Heininger U, Vasquez CS, Moorthy VS, Epstein JE *et al*: A systematic review of safety data reporting in clinical trials of vaccines against malaria, tuberculosis, and human immunodeficiency virus. *Vaccine* 2013, 31(35):3628-3635.
89. Moher D, Hopewell S, Schulz KF, Montori V, Gøtzsche PC, Devereaux PJ, Elbourne D, Egger M, Altman DG: CONSORT 2010 Explanation and Elaboration: updated guidelines for reporting parallel group randomised trials. *BMJ* 2010, 340:c869.
90. Lewis JA: Statistical principles for clinical trials (ICH E9): an introductory note on an international guideline. *Statistics in medicine* 1999, 18(15):1903-1942.
91. ICH Harmonised Tripartite Guideline. Statistical principles for clinical trials. International Conference on Harmonisation E9 Expert Working Group. *Statistics in medicine* 1999, 18(15):1905-1942.
92. Munsaka MS: A Question-Based Approach to the Analysis of Safety Data. In: *Biopharmaceutical Applied Statistics Symposium : Volume 2 Biostatistical Analysis of Clinical Trials*. edn. Edited by Peace KE, Chen D-G, Menon S. Singapore: Springer Singapore; 2018: 193-216.
93. Leporini C, De Sarro G, Russo E: Adherence to therapy and adverse drug reactions: is there a link? *Expert opinion on drug safety* 2014, 13 Suppl 1:S41-55.
94. Campbell Ross T, Willox Gage P, Jhund Pardeep S, Hawkins Nathaniel M, Huang F, Petrie Mark C, McMurray John JV: Reporting of Lost to Follow-Up and Treatment Discontinuation in Pharmacotherapy and Device Trials in Chronic Heart Failure. *Circulation: Heart Failure* 2016, 9(5):e002842.
95. Updated WHO policy recommendation: intermittent preventive treatment of malaria in pregnancy using sulfadoxine-pyrimethamine (IPTp-SP)
[https://www.who.int/malaria/publications/atoz/who_iptp_sp_policy_recommendation/en/

]

96. Allen EN, Chandler CIR, Mandimika N, Pace C, Mehta U, Barnes KI: Evaluating harm associated with anti-malarial drugs: a survey of methods used by clinical researchers to elicit, assess and record participant-reported adverse events and related data. *Malaria journal* 2013, 12:325-325.
97. Liberati A, Altman DG, Tetzlaff J, Mulrow C, Gotzsche PC, Ioannidis JP, Clarke M, Devereaux PJ, Kleijnen J, Moher D: The PRISMA statement for reporting systematic reviews and meta-analyses of studies that evaluate health care interventions: explanation and elaboration. *Ann Intern Med* 2009, 151(4):W65-94.
98. Schulz KF, Altman DG, Moher D: CONSORT 2010 Statement: updated guidelines for reporting parallel group randomised trials. *BMJ* 2010, 340.
99. Lesaffre E: Superiority, equivalence, and non-inferiority trials. *Bulletin of the NYU hospital for joint diseases* 2008, 66(2):150-154.
100. Colditz GA, Emerson JD: The statistical content of published medical research: some implications for biomedical education. *Medical Education* 1985, 19(3):248-255.
101. Emerson JD, Colditz GA: Use of Statistical Analysis in The New England Journal of Medicine. *New England Journal of Medicine* 1983, 309(12):709-713.
102. Luntamo M, Kulmala T, Mbewe B, Cheung YB, Maleta K, Ashorn P: Effect of repeated treatment of pregnant women with sulfadoxine-pyrimethamine and azithromycin on preterm delivery in Malawi: a randomized controlled trial. *The American journal of tropical medicine and hygiene* 2010, 83(6):1212-1220.
103. Valea I, Tinto H, Drabo MK, Huybregts L, Henry MC, Roberfroid D, Guiguemde RT, Kolsteren P, D'Alessandro U: Intermittent preventive treatment of malaria with sulphadoxine-pyrimethamine during pregnancy in Burkina Faso: effect of adding a third dose to the standard two-dose regimen on low birth weight, anaemia and pregnancy outcomes. *Malaria journal* 2010, 9:324.

104. Diakite OS, Maiga OM, Kayentao K, Traoré BT, Djimde A, Traoré B, Diallo M, Traoré M, Ongoiba A, Doumtabé D *et al*: Superiority of 3 over 2 doses of intermittent preventive treatment with sulfadoxine-pyrimethamine for the prevention of malaria during pregnancy in mali: a randomized controlled trial. *Clinical infectious diseases* 2011, 53(3):215-223.
105. Ndyomugenyi R, Clarke SE, Hutchison CL, Hansen KS, Magnussen P: Efficacy of malaria prevention during pregnancy in an area of low and unstable transmission: An individually-randomised placebo-controlled trial using intermittent preventive treatment and insecticide-treated nets in the Kabale Highlands, southwestern Uganda. *Trans R Soc Trop Med Hyg* 2011, 105(11):607-616.
106. Wini L, Appleyard B, Bobogare A, Pikacha J, Seke J, Tuni M, Hou L, Hii J, McCarthy J, van Eijk AM: Intermittent preventive treatment with sulfadoxine-pyrimethamine versus weekly chloroquine prophylaxis for malaria in pregnancy in Honiara, Solomon Islands: a randomised trial. *Malaria World Journal* 2013, 4(12).
107. Denoeud-Ndam L, Zannou DM, Fourcade C, Taron-Brocard C, Porcher R, Atadokpede F, Komongui DG, Dossou-Gbete L, Afangnihoun A, Ndam NT *et al*: Cotrimoxazole prophylaxis versus mefloquine intermittent preventive treatment to prevent malaria in HIV-infected pregnant women: two randomized controlled trials. *Journal of acquired immune deficiency syndromes (1999)* 2014, 65(2):198-206.
108. González R, Mombo-Ngoma G, Ouédraogo S, Kakolwa MA, Abdulla S, Accrombessi M, Aponte JJ, Akerey-Diop D, Basra A, Briand V *et al*: Intermittent preventive treatment of malaria in pregnancy with mefloquine in HIV-negative women: a multicentre randomized controlled trial. *PLoS medicine* 2014, 11(9):e1001733.
109. Gonzalez R, Desai M, Macete E, Ouma P, Kakolwa MA, Abdulla S, Aponte JJ, Bulu H, Kabanywany AM, Katana A *et al*: Intermittent Preventive Treatment of Malaria in Pregnancy with Mefloquine in HIV-Infected Women Receiving Cotrimoxazole Prophylaxis: a Multicenter Randomized Placebo-Controlled Trial. *PLoS medicine* 2014, 11(9) (no pagination).

110. Klement E, Pitché P, Kendjo E, Singo A, D'Almeida S, Akouete F, Akpaloo Y, Tossa K, Prince-Agbodjan S, Patassi A *et al*: Effectiveness of co-trimoxazole to prevent plasmodium falciparum malaria in HIV-positive pregnant women in sub-saharan Africa: An open-label, randomized controlled trial. *Clinical Infectious Diseases* 2014, 58(5):651-659.
111. Manyando C, Njunju EM, Mwakazanga D, Chongwe G, Mkandawire R, Champo D, Mulenga M, De Crop M, Claeys Y, Ravinetto RM *et al*: Safety of daily Co-trimoxazole in pregnancy in an area of changing malaria epidemiology: A phase 3b randomized controlled clinical trial. *PloS one* 2014, 9(5).
112. Desai M, Gutman J, L'Lanziva A, Otieno K, Juma E, Kariuki S, Ouma P, Were V, Laserson K, Katana A *et al*: Intermittent screening and treatment or intermittent preventive treatment with dihydroartemisinin-piperaquine versus intermittent preventive treatment with sulfadoxine-pyrimethamine for the control of malaria during pregnancy in western Kenya: an open-label, three-group, randomised controlled superiority trial. *Lancet (london, england)* 2015, 386(10012):2507-2519.
113. Unger HW, Ome-Kaius M, Wangnapi RA, Umbers AJ, Hanieh S, Suen CS, Robinson LJ, Rosanas-Urgell A, Wapling J, Lufele E *et al*: Sulphadoxine-pyrimethamine plus azithromycin for the prevention of low birthweight in Papua New Guinea: a randomised controlled trial. *BMC medicine* 2015, 13:9.
114. Kakuru A, Jagannathan P, Muhindo MK, Natureeba P, Awori P, Nakalembe M, Opira B, Olwoch P, Ategeka J, Nayebare P *et al*: Dihydroartemisinin-Piperaquine for the Prevention of Malaria in Pregnancy. *New England journal of medicine* 2016, 374(10):928-939.
115. Natureeba P, Kakuru A, Muhindo M, Ochieng T, Ategeka J, Koss CA, Plenty A, Charlebois ED, Clark TD, Nzarubara B *et al*: Intermittent Preventive Treatment With Dihydroartemisinin-Piperaquine for the Prevention of Malaria Among HIV-Infected Pregnant Women. *The Journal of infectious diseases* 2017, 216(1):29-35.

116. Akinyotu O, Bello F, Abdus-Salam R, Arowojolu A: Comparative study of mefloquine and sulphadoxine-pyrimethamine for malaria prevention among pregnant women with HIV in southwest Nigeria. *International journal of gynaecology and obstetrics: the official organ of the International Federation of Gynaecology and Obstetrics* 2018, 142(2):194-200.
117. Kajubi R, Ochieng T, Kakuru A, Jagannathan P, Nakalembe M, Ruel T, Opira B, Ochokoru H, Ategeka J, Nayebare P *et al*: Monthly sulfadoxine-pyrimethamine versus dihydroartemisinin-piperaquine for intermittent preventive treatment of malaria in pregnancy: a double-blind, randomised, controlled, superiority trial. *Lancet* 2019, 393(10179):1428-1439.
118. Verbeke G, Fieuws S, Molenberghs G, Davidian M: The analysis of multivariate longitudinal data: a review. *Stat Methods Med Res* 2014, 23(1):42-59.
119. Rosenkranz GK: Modeling laboratory data from clinical trials. *Computational Statistics & Data Analysis* 2009, 53(3):812-819.
120. Gould AL: Statistical Methods for Evaluating Safety in Medical Product Development: Wiley; 2015.
121. Kahan BC, Jairath V, Doré CJ, Morris TP: The risks and rewards of covariate adjustment in randomized trials: an assessment of 12 outcomes from 8 studies. *Trials* 2014, 15(1):139.
122. Gam CMB, Tanniou J, Keiding N, Løkkegaard EL: A model for the distribution of daily number of births in obstetric clinics based on a descriptive retrospective study. *BMJ open* 2013, 3(8):e002920.
123. Lawless JF, Nadeau C: Some Simple Robust Methods for the Analysis of Recurrent Events. *Technometrics* 1995, 37(2):158-168.
124. Rosenkranz GK: An Approach to Integrated Safety Analyses from Clinical Studies. *Drug information journal : DIJ / Drug Information Association* 2010, 44(6):649-657.
125. Analysis of Adverse Events in Drug Safety: A Multivariate Approach Using Stratified Quasi-least Squares [<http://biostats.bepress.com/upennbiostat/papers/art29>]

126. Dodd S, White IR, Williamson P: A framework for the design, conduct and interpretation of randomised controlled trials in the presence of treatment changes. *Trials* 2017, 18(1):498.
127. Robins JM: Correcting for non-compliance in randomized trials using structural nested mean models. *Communications in Statistics - Theory and Methods* 1994, 23(8):2379-2412.
128. Frangakis C, Rubin D: Addressing complications of intention-to-treat analysis in the combined presence of all-or-none treatment-noncompliance and subsequent missing outcomes. *Biometrika* 1999, 86(2):365-379.
129. Nich C, Carroll KM: Intention-to-treat meets missing data: implications of alternate strategies for analyzing clinical trials data. *Drug and alcohol dependence* 2002, 68(2):121-130.
130. Ye C, Beyene J, Browne G, Thabane L: Estimating treatment effects in randomised controlled trials with non-compliance: a simulation study. *BMJ open* 2014, 4(6):e005362.
131. Amit O, Heiberger RM, Lane PW: Graphical approaches to the analysis of safety data from clinical trials. *Pharmaceutical statistics* 2008, 7(1):20-35.
132. Price KL, Amy Xia H, Lakshminarayanan M, Madigan D, Manner D, Scott J, Stamey JD, Thompson L: Bayesian methods for design and analysis of safety trials. *Pharmaceutical statistics* 2014, 13(1):13-24.
133. Friede T, Posch M, Zohar S, Alberti C, Benda N, Comets E, Day S, Dmitrienko A, Graf A, Günhan BK: Recent advances in methodology for clinical trials in small populations: the InSPiRe project. *Orphanet journal of rare diseases* 2018, 13(1):186.
134. Hilgers R-D, Bogdan M, Burman C-F, Dette H, Karlsson M, König F, Male C, Mentré F, Molenberghs G, Senn S: Lessons learned from IDeAl—33 recommendations from the IDeAl-net about design and analysis of small population clinical trials. *Orphanet journal of rare diseases* 2018, 13(1):77.
135. Mitroiu M, Rengerink KO, Pontes C, Sancho A, Vives R, Pesiou S, Fontanet JM, Torres F, Nikolakopoulos S, Pateras K: Applicability and added value of novel methods to improve drug development in rare diseases. *Orphanet journal of rare diseases* 2018, 13(1):200.

136. Lineberry N, Berlin JA, Mansi B, Glasser S, Berkwits M, Klem C, Bhattacharya A, Citrome L, Enck R, Fletcher J *et al*: Recommendations to improve adverse event reporting in clinical trial publications: a joint pharmaceutical industry/journal editor perspective. *BMJ* 2016, 355:i5078.
137. Munsaka MS: A Question-Based Approach to the Analysis of Safety Data. In: *Biopharmaceutical Applied Statistics Symposium*. edn.; 2018: 193-216.
138. Patson N, Mukaka M, Otworld KN, Kazembe L, Mathanga DP, Mwapasa V, Kabaghe AN, Eijkemans MJC, Laufer MK, Chirwa T: Systematic review of statistical methods for safety data in malaria chemoprevention in pregnancy trials. *Malaria journal* 2020, 19(1):119.
139. Phillips R, Cornelius V: Understanding current practice, identifying barriers and exploring priorities for adverse event analysis in randomised controlled trials: an online, cross-sectional survey of statisticians from academia and industry. *BMJ open* 2020, 10(6):e036875.
140. Colopy MW, Gordon R, Ahmad F, Wang WW, Duke SP, Ball G: Statistical Practices of Safety Monitoring: An Industry Survey. *Therapeutic innovation & regulatory science* 2019, 53(3):293-300.
141. Box-Steffensmeier JM, De Boef S: Repeated events survival models: the conditional frailty model. *Statistics in medicine* 2006, 25(20):3518-3533.
142. Hougaard P: Analysis of multivariate survival data: Springer Science & Business Media; 2012.
143. Duchateau L, Janssen P: The frailty model: Springer Science & Business Media; 2007.
144. Amorim LD, Cai J: Modelling recurrent events: a tutorial for analysis in epidemiology. *International journal of epidemiology* 2015, 44(1):324-333.
145. Balan TA, Putter H: Nonproportional hazards and unobserved heterogeneity in clustered survival data: When can we tell the difference? *Statistics in medicine* 2019, 38(18):3405-3420.
146. Bower H, Crowther MJ, Rutherford MJ, Andersson TM-L, Clements M, Liu X-R, Dickman PW, Lambert PC: Capturing simple and complex time-dependent effects using flexible parametric

- survival models: A simulation study. *Communications in Statistics-Simulation and Computation* 2019;1-17.
147. Rutherford MJ, Crowther MJ, Lambert PC: The use of restricted cubic splines to approximate complex hazard functions in the analysis of time-to-event data: a simulation study. *Journal of Statistical Computation and Simulation* 2015, 85(4):777-793.
 148. Gasparini A, Clements MS, Abrams KR, Crowther MJ: Impact of model misspecification in shared frailty survival models. *Statistics in Medicine* 2019, 38(23):4477-4502.
 149. Cook RJ, Lawless J: The statistical analysis of recurrent events: Springer Science & Business Media; 2007.
 150. Hanagal DD: Modeling Survival Data Using Frailty Models: Second Edition: Springer Singapore; 2019.
 151. Andersen PK, Gill RD: Cox's regression model for counting processes: a large sample study. *The annals of statistics* 1982:1100-1120.
 152. Kalbfleisch JD, Prentice RL: The Statistical Analysis of Failure Time Data: Wiley; 2011.
 153. Lawless JF: Statistical Models and Methods for Lifetime Data: Wiley; 2011.
 154. Kelly PJ, Lim LLY: Survival analysis for recurrent event data: an application to childhood infectious diseases. *Statistics in medicine* 2000, 19(1):13-33.
 155. Babiker A, Cuzick J: A simple frailty model for family studies with covariates. *Statistics in medicine* 1994, 13(16):1679-1692.
 156. Pickles A, Crouchley R: A comparison of frailty models for multivariate survival data. *Statistics in medicine* 1995, 14(13):1447-1461.
 157. Morris TP, White IR, Crowther MJ: Using simulation studies to evaluate statistical methods. *Statistics in medicine* 2019, 38(11):2074-2102.
 158. Crowther MJ: merlin-a unified modelling framework for data analysis and methods development in Stata. *arXiv preprint arXiv:180601615* 2018.

159. Crowther MJ, Lambert PC: Simulating biologically plausible complex survival data. *Statistics in Medicine* 2013, 32(23):4118-4134.
160. Crowther MJ, Look MP, Riley RD: Multilevel mixed effects parametric survival models using adaptive Gauss–Hermite quadrature with application to recurrent events and individual participant data meta-analysis. *Statistics in medicine* 2014, 33(22):3844-3858.
161. Divala TH, Mungwira RG, Mawindo PM, Nyirenda OM, Kanjala M, Ndaferankhande M, Tsirizani LE, Masonga R, Muwalo F, Boudová S *et al*: Chloroquine as weekly chemoprophylaxis or intermittent treatment to prevent malaria in pregnancy in Malawi: a randomised controlled trial. *The Lancet Infectious diseases* 2018, 18(10):1097-1107.
162. Hougaard P: Life table methods for heterogeneous populations: Distributions describing the heterogeneity. *Biometrika* 1984, 71(1):75-83.
163. Merz M, Lee KR, Kullak-Ublick GA, Brueckner A, Watkins PB: Methodology to assess clinical liver safety data. *Drug safety* 2014, 37(1):33-45.
164. Klepper MJ, Cobert B: Drug safety data: how to analyze, summarize and interpret to determine risk: Jones & Bartlett Publishers; 2010.
165. Barker C: Exploratory method for summarizing concomitant medication data—the mean cumulative function. *Pharmaceutical statistics* 2010, 9(4):331-336.
166. Therneau TM, Grambsch PM: The Cox model. In: *Modeling survival data: extending the Cox model*. edn.: Springer; 2000: 39-77.
167. Fu H, Luo J, Qu Y: Hypoglycemic events analysis via recurrent time-to-event (HEART) models. *Journal of biopharmaceutical statistics* 2016, 26(2):280-298.
168. Luo J, Qu Y: Analysis of hypoglycemic events using negative binomial models. *Pharmaceutical statistics* 2013, 12(4):233-242.
169. Hickey G, Philipson P, Jorgensen A, Kolamunnage-Dona R: A comparison of joint models for longitudinal and competing risks data, with application to an epilepsy drug randomized

- controlled trial. *Journal of the Royal Statistical Society: Series A (Statistics in Society)* 2018, 181(4):1105-1123.
170. Hickey GL, Philipson P, Jorgensen A, Kolamunnage-Dona R: Joint modelling of time-to-event and multivariate longitudinal outcomes: recent developments and issues. *BMC medical research methodology* 2016, 16(1):117.
 171. Syrjälä E, Nevalainen J, Peltonen J, Takkinen H-M, Hakola L, Åkerlund M, Veijola R, Ilonen J, Toppari J, Knip M: A joint modeling approach for childhood meat, fish and egg consumption and the risk of advanced islet autoimmunity. *Scientific reports* 2019, 9(1):1-10.
 172. Abbassi-Ghanavati M, Greer LG, Cunningham FG: Pregnancy and laboratory studies: a reference table for clinicians. *Obstetrics and gynecology* 2009, 114(6):1326-1331.
 173. Kovacs SD, van Eijk AM, Sevene E, Dellicour S, Weiss NS, Emerson S, Steketee R, Ter Kuile FO, Stergachis A: The Safety of Artemisinin Derivatives for the Treatment of Malaria in the 2nd or 3rd Trimester of Pregnancy: A Systematic Review and Meta-Analysis. *PloS one* 2016, 11(11):e0164963-e0164963.
 174. Henderson R, Diggle P, Dobson A: Joint modelling of longitudinal measurements and event time data. *Biostatistics* 2000, 1(4):465-480.
 175. Ibrahim JG, Chu H, Chen LM: Basic concepts and methods for joint models of longitudinal and survival data. *Journal of Clinical Oncology* 2010, 28(16):2796.
 176. Long JD, Mills JA: Joint modeling of multivariate longitudinal data and survival data in several observational studies of Huntington's disease. *BMC medical research methodology* 2018, 18(1):1-15.
 177. Chi YY, Ibrahim JG: Joint models for multivariate longitudinal and multivariate survival data. *Biometrics* 2006, 62(2):432-445.
 178. Andrinopoulou E-R, Clancy JP, Szczesniak R: Multivariate joint modeling to identify markers of growth and lung function decline that predict cystic fibrosis pulmonary exacerbation onset. *BMC pulmonary medicine* 2020, 20:1-11.

179. Buu A, Li R, Tan X, Zucker RA: Statistical models for longitudinal zero-inflated count data with applications to the substance abuse field. *Statistics in medicine* 2012, 31(29):4074-4086.
180. Li H, Zhang Y, Carroll RJ, Keadle SK, Sampson JN, Matthews CE: A joint modeling and estimation method for multivariate longitudinal data with mixed types of responses to analyze physical activity data generated by accelerometers. *Statistics in medicine* 2017, 36(25):4028-4040.
181. Yang Y, Kang J: Joint analysis of mixed Poisson and continuous longitudinal data with nonignorable missing values. *Computational statistics & data analysis* 2010, 54(1):193-207.
182. Vrijens B, De Geest S, Hughes DA, Przemyslaw K, Demonceau J, Ruppar T, Dobbels F, Fargher E, Morrison V, Lewek P *et al*: A new taxonomy for describing and defining adherence to medications. *British journal of clinical pharmacology* 2012, 73(5):691-705.
183. Gonzalez R, Mombo-Ngoma G, Ouedraogo S, Kakolwa MA, Abdulla S, Accrombessi M: Intermittent preventive treatment of malaria in pregnancy with mefloquine in HIV-negative women: a multicentre randomized controlled trial. *PLoS medicine* 2014, 11.
184. Clerk CA, Bruce J, Affipunguh PK, Mensah N, Hodgson A, Greenwood B, Chandramohan D: A Randomized, Controlled Trial of Intermittent Preventive Treatment with Sulfadoxine-Pyrimethamine, Amodiaquine, or the Combination in Pregnant Women in Ghana. *The Journal of infectious diseases* 2008, 198(8):1202-1211.
185. Lange T, Keiding N: Skin cancer as a marker of sun exposure: a case of serious immortality bias. *International journal of epidemiology* 2014, 43(3):971-971.
186. Beyersmann J, Wolkewitz M, Schumacher M: The impact of time-dependent bias in proportional hazards modelling. *Statistics in medicine* 2008, 27(30):6439-6454.
187. Pearce N: What does the odds ratio estimate in a case-control study? *International journal of epidemiology* 1993, 22(6):1189-1192.

188. Nuño MM, Gillen DL: Chapter 5 - Alternative Sampling Designs for Time-to-Event Data With Applications to Biomarker Discovery in Alzheimer's Disease. In: *Handbook of Statistics. Volume 36*, edn. Edited by Srinivasa Rao ASR, Pyne S, Rao CR: Elsevier; 2017: 105-166.
189. Kim RS, Kaplan RC: Analysis of secondary outcomes in nested case-control study designs. *Statistics in medicine* 2014, 33(24):4215-4226.
190. Støer NC, Samuelsen SO: Inverse probability weighting in nested case-control studies with additional matching--a simulation study. *Statistics in medicine* 2013, 32(30):5328-5339.
191. Kim RS: A new comparison of nested case-control and case-cohort designs and methods. *European journal of epidemiology* 2015, 30(3):197-207.
192. Borgan Ø, Keogh R: Nested case-control studies: should one break the matching? *Lifetime data analysis* 2015, 21(4):517-541.
193. Samuelsen SO: A pseudolikelihood approach to analysis of nested case-control studies. *Biometrika* 1997, 84(2):379-394.
194. Borgan Ø, Breslow N, Chatterjee N, Gail MH, Scott A, Wild CJ: Handbook of statistical methods for case-control studies: CRC Press; 2018.
195. Cai T, Zheng Y: Evaluating prognostic accuracy of biomarkers in nested case-control studies. *Biostatistics (Oxford, England)* 2012, 13(1):89-100.
196. Taylor JMG: Choosing the number of controls in a matched case-control study, some sample size, power and efficiency considerations. *Statistics in medicine* 1986, 5(1):29-36.
197. Gail M, Williams R, Byar DP, Brown C: How many controls? *Journal of chronic diseases* 1976, 29(11):723-731.
198. Prentice R, Breslow N: Retrospective studies and failure time models. *Biometrika* 1978, 65(1):153-158.
199. Samuelsen SO, Støer N: Inverse probability weighting in nested case-control studies. *Handbook of Statistical Methods for Case-Control Studies* 2018:351-372.
200. Ann Q, Tham R: Incidence Density Sampling for Nested Case-Control Study Designs. 2019.

201. Kiri VA: Programming challenges of sampling controls to cases from the dynamic risk sets in nested case–control studies. *Pharmaceutical Programming* 2012, 5(1-2):29-33.
202. Smith DL: Patient Nonadherence in Clinical Trials: Could There Be a Link to Postmarketing Patient Safety? *Drug Information Journal* 2012, 46(1):27-34.
203. Denoeud-Ndam L, Clément MC, Briand V, Akakpo J, Agossou VK, Atadokpédé F, Dossou-Gbété L, Komongui DG, Afangnihoun A, Girard PM *et al*: Tolerability of mefloquine intermittent preventive treatment for malaria in HIV-infected pregnant women in Benin. *Journal of acquired immune deficiency syndromes (1999)* 2012, 61(1):64-72.
204. Claxton AJ, Cramer J, Pierce C: A systematic review of the associations between dose regimens and medication compliance. *Clinical therapeutics* 2001, 23(8):1296-1310.
205. Desai M, Hill J, Fernandes S, Walker P, Pell C, Gutman J, Kayentao K, Gonzalez R, Webster J, Greenwood B *et al*: Prevention of malaria in pregnancy. *The Lancet Infectious diseases* 2018, 18(4):e119-e132.
206. Hill J, Hoyt J, Achieng F, Ouma P, L'Lanziva A, Kariuki S, Desai M, Webster J: User and Provider Acceptability of Intermittent Screening and Treatment and Intermittent Preventive Treatment with Dihydroartemisinin-Piperaquine to Prevent Malaria in Pregnancy in Western Kenya. *PloS one* 2016, 11(3):e0150259-e0150259.
207. Etminan M: Pharmacoepidemiology II: The Nested Case-Control Study—A Novel Approach in Pharmacoepidemiologic Research. *Pharmacotherapy: The Journal of Human Pharmacology and Drug Therapy* 2004, 24(9):1105-1109.
208. Essebag V, Platt RW, Abrahamowicz M, Pilote L: Comparison of nested case-control and survival analysis methodologies for analysis of time-dependent exposure. *BMC medical research methodology* 2005, 5(1):5.
209. Unkel S, Amiri M, Benda N, Beyersmann J, Knoerzer D, Kupas K, Langer F, Leverkus F, Loos A, Ose C: On estimands and the analysis of adverse events in the presence of varying follow-up

- times within the benefit assessment of therapies. *Pharmaceutical statistics* 2019, 18(2):166-183.
210. Patson N, Mukaka M, D'Alessandro U, Chapotera G, Mwapasa V, Mathanga D, Kazembe L, Laufer MK, Chirwa T: Joint modelling of multivariate longitudinal clinical laboratory safety outcomes, concomitant medication and clinical adverse events: application to artemisinin-based treatment during pregnancy clinical trial. *BMC medical research methodology* 2021, 21(1):208.
 211. Kahan BC: Accounting for centre-effects in multicentre trials with a binary outcome – when, why, and how? *BMC medical research methodology* 2014, 14(1):20.
 212. Kahan BC, Morris TP: Analysis of multicentre trials with continuous outcomes: when and how should we account for centre effects? *Statistics in medicine* 2013, 32(7):1136-1149.
 213. Divala TH, Mungwira RG, Laufer MK: Moving targets: The challenges of studying infectious diseases among pregnant women in resource limited settings. *Vaccine* 2015, 33(47):6401-6405.

Appendices

Appendix 1: Supplementary tables and figures

Table s4.1: PRISMA CHECKLIST



PRISMA 2009 Checklist

Section/topic	#	Checklist item	Reported on page #
TITLE : Systematic review of statistical methods for safety data in malaria chemoprevention in pregnancy trials			
Title	1	Identify the report as a systematic review, meta-analysis, or both.	72
ABSTRACT			
Structured summary	2	Provide a structured summary including, as applicable: background; objectives; data sources; study eligibility criteria, participants, and interventions; study appraisal and synthesis methods; results; limitations; conclusions and implications of key findings; systematic review registration number.	74-75
INTRODUCTION			
Rationale	3	Describe the rationale for the review in the context of what is already known.	76-78
Objectives	4	Provide an explicit statement of questions being addressed with reference to participants, interventions, comparisons, outcomes, and study design (PICOS).	78
METHODS			
Protocol and registration	5	Indicate if a review protocol exists, if and where it can be accessed (e.g., Web address), and, if available, provide registration information including registration number.	79-81
Eligibility criteria	6	Specify study characteristics (e.g., PICOS, length of follow-up) and report characteristics (e.g., years considered, language, publication status) used as criteria for eligibility, giving rationale.	79-81
Information sources	7	Describe all information sources (e.g., databases with dates of coverage, contact with study authors to identify additional studies) in the search and date last searched.	7-81
Search	8	Present full electronic search strategy for at least one database, including any limits used, such that it could be repeated.	79-81
Study selection	9	State the process for selecting studies (i.e., screening, eligibility, included in systematic review, and, if applicable, included in the meta-analysis).	79-81
Data collection process	10	Describe method of data extraction from reports (e.g., piloted forms, independently, in duplicate) and any processes for obtaining and confirming data from investigators.	81-82
Data items	11	List and define all variables for which data were sought (e.g., PICOS, funding sources) and any assumptions and simplifications made.	81-82
Risk of bias in individual studies	12	Describe methods used for assessing risk of bias of individual studies (including specification of whether this was done at the study or outcome level), and how this information is to be used in any data synthesis.	N/A
Summary measures	13	State the principal summary measures (e.g., risk ratio, difference in means). -percentage of statistical methods used	82



PRISMA 2009 Checklist

Synthesis of results	14	Describe the methods of handling data and combining results of studies, if done, including measures of consistency (e.g., I^2) for each meta-analysis.	N/A
----------------------	----	---	-----

Page 1 of 2

Section/topic	#	Checklist item	Reported on page #
Risk of bias across studies	15	Specify any assessment of risk of bias that may affect the cumulative evidence (e.g., publication bias, selective reporting within studies).	N/A
Additional analyses	16	Describe methods of additional analyses (e.g., sensitivity or subgroup analyses, meta-regression), if done, indicating which were pre-specified.	N/A
RESULTS			
Study selection	17	Give numbers of studies screened, assessed for eligibility, and included in the review, with reasons for exclusions at each stage, ideally with a flow diagram.	83, Figure 4.1
Study characteristics	18	For each study, present characteristics for which data were extracted (e.g., study size, PICOS, follow-up period) and provide the citations.	83-85, Table 4.1
Risk of bias within studies	19	Present data on risk of bias of each study and, if available, any outcome level assessment (see item 12).	N/A
Results of individual studies	20	For all outcomes considered (benefits or harms), present, for each study: (a) simple summary data for each intervention group (b) effect estimates and confidence intervals, ideally with a forest plot.	N/A
Synthesis of results	21	Present results of each meta-analysis done, including confidence intervals and measures of consistency.	N/A
Risk of bias across studies	22	Present results of any assessment of risk of bias across studies (see Item 15).	N/A
Additional analysis	23	Give results of additional analyses, if done (e.g., sensitivity or subgroup analyses, meta-regression [see Item 16]).	N/A
DISCUSSION			
Summary of evidence	24	Summarize the main findings including the strength of evidence for each main outcome; consider their relevance to key groups (e.g., healthcare providers, users, and policy makers).	91-97
Limitations	25	Discuss limitations at study and outcome level (e.g., risk of bias), and at review-level (e.g., incomplete retrieval of identified research, reporting bias).	97
Conclusions	26	Provide a general interpretation of the results in the context of other evidence, and implications for future research.	97
FUNDING			
Funding	27	Describe sources of funding for the systematic review and other support (e.g., supply of data); role of funders for the systematic review.	4567

From: Moher D, Liberati A, Tetzlaff J, Altman DG, The PRISMA Group (2009). Preferred Reporting Items for Systematic Reviews and Meta-Analyses: The PRISMA Statement. PLoS Med 6(7): e1000097.

Table S4.2: Search Strategy based on PICOTS framework

Components	Characteristics
Population	Pregnant women, Any geographical location
Intervention	Exposed to antimalarial drug for malaria prevention
Control	Exposed to placebo or standard antimalarial drug for prevention
Outcomes	Safety outcomes e.g. Adverse Events, Lost to follow-up, drop-outs or poor adherence. Safety assessment statistical methods
Timing	January 2010 to July 2019
Setting	Limited to English language

Table S4.3: Description of reported continuous laboratory safety outcomes collected and statistical analysis

Author, Year, (reference)	Continuous safety outcome	Dichotomized continuous outcome	Outcome: Time points collected	Statistical method (outcome)
Luntamo. 2010 (32)	haemoglobin	yes	Haemoglobin: 4 week interval, from 28 weeks of gestation age to delivery	None
Valea 2010 (35)	vital signs, haemoglobin	Yes	Haemoglobin: First ANC visit, before SP administration, whenever a malaria infection was suspected and at delivery Vital signs: At every antenatal clinic visit	None
Diakite. 2011(36)	Haemoglobin, white blood count,	Yes	Haemoglobin: At enrolment, non-scheduled visit and delivery	Multiple linear regression (haemoglobin)
Ndyomugenyi 2011 (37)	haemoglobin	yes	Haemoglobin: At enrolment, 36 weeks of gestation age, delivery	Student's t-test/Analysis of variance (haemoglobin).
Wini 2013 (38)	Haemoglobin, white blood cell	yes	Haemoglobin: At enrolment, 36 weeks of gestation age White blood cells: Not stated	Student's t-test/Analysis of variance (Haemoglobin)
Denocud-Ndam 2014 (40)	Haemoglobin, full blood count, cd4 count,	Yes	Haemoglobin, full blood count, cd4 count: At enrolment (16-28 weeks), one month interval and delivery	Wilcoxon test (haemoglobin)
González 2014 (41)	haemoglobin	yes	Haemoglobin: enrolment, unscheduled visit and delivery	None
González 2014 (42)	haemoglobin	Yes	Haemoglobin: At enrolment, monthly antenatal clinic visits and unscheduled visits and delivery	None
Klement 2014 (43)	Leukocyte count, haemoglobin	yes	Haemoglobin, leukocyte count: At enrolment, monthly antenatal clinic visits and delivery	None
Manyando 2014 (44)	Haemoglobin, glucose, proteins	yes	Haemoglobin: At enrolment, once in third trimester Glucose, proteins: not specified	None
Desai 2015 (45)	Haemoglobin, leucocyte, proteinuria	yes	Haemoglobin: At enrolment, monthly antenatal clinic visits, unscheduled visit and delivery Leucocyte and proteinuria: Not specified	Linear regression (haemoglobin)
Unger 2015 (48)	Haemoglobin	yes	Haemoglobin: at enrolment, unscheduled visit and delivery	Mean at enrolment (haemoglobin)
Kakuru 2016 (49)	white-cell count, neutrophil count, platelet count, haemoglobin, alanine aminotransferase level	yes	white-cell count, neutrophil count, platelet count alanine aminotransferase level haemoglobin: at enrolment, 8 week interval and delivery	Mean at enrolment (white-cell count, neutrophil count, platelet count, haemoglobin, alanine aminotransferase level)
Kimani 2016 (50)	Haemoglobin, vital signs, glycosuria, proteinuria	yes	Haemoglobin: Enrolment, 36 weeks of gestation age Vital signs, glycosuria, proteinuria: unscheduled visits	t-test/Analysis of variance (haemoglobin)
Natureeba 2017 (53)	Haemoglobin, full blood count, alanine aminotransferase, QT interval	yes	Haemoglobin, full blood count, alanine aminotransferase: at enrolment, every 8 weeks, delivery QT interval: prior to the first daily dose and 3–4 hours after third daily dose after reaching 28 weeks of gestation age (confined to a subsample of 55 participants)	t-test/analysis of variance (QT interval)
Divala 2018 (56)	haemoglobin	Yes	Haemoglobin: At enrolment, every 4 weeks, delivery, unscheduled visit	None
Akinyotu 2018	CD4 count	Yes	CD4 count: At enrolment	None
Kajubi 2019	Haemoglobin, QTc, alanine aminotransferase	yes	Haemoglobin, alanine aminotransferase: At enrolment, every 8 weeks, at delivery QT interval: prior to the first daily dose of study drugs and 3–4 h after the third daily dose, at 20, 28, and 36 weeks gestation	t-test/analysis of variance (QT interval), mean (haemoglobin concentration at enrolment)

Table S4.4: Detailed list of categorical safety outcomes collected

1 st Author, Year, (reference)	Statistical method(s) used (outcome)
Luntamo. 2010 (32)	Epidemiological statistic (neonatal or perinatal mortality rate) Fishers/chi-square/contingency table (abortions, stillbirths, severe anaemia and infant deaths)
Valea 2010 (35)	Descriptive statistics (Neonatal death, stillbirth, preterm birth, anaemia, abortion) Poisson (Neonatal death, stillbirth, preterm birth, anaemia, abortion)
Diakite. 2011(36)	Descriptive (Icterus, congenital anomaly, fever, vomiting, Pregnancy loss, perinatal or neonatal death, skin rash, jaundice) Poisson (Pregnancy loss, perinatal or neonatal death, anaemia)
Ndyomugenyi 2011 (37)	Fishers exact/chi-square (Icterus, congenital anomaly, fever, vomiting) Descriptive (abortion, stillbirth, neonatal death, congenital anomaly, haemoglobin, anaemia) Fishers exact/chi-square (abortion, stillbirth, neonatal death, perinatal death, maternal death, congenital anomaly)
Wini 2013 (38)	Descriptive (pins and needles, vomiting, Stillbirth, skin rash, neonatal death, anaemia, mouth blisters, shortness of breath weakness, drowsiness, numbness)
Denoeud-Ndam 2014 (40)	descriptive (abortions, stillbirths, and congenital anomaly, Vomiting, nausea, dizziness, cutaneous reaction and fatigue, neonatal deaths, neonatal infection, Fatigue, preterm birth, fever, cough, rash, pain, anaemia, leukopenia, seizure, psychosis, thrombocytopenia, Gastro enteritis and other adverse events) Fishers exact/chi-square (Vomiting, nausea, dizziness, cutaneous reaction and fatigue (abortions, stillbirths, and congenital anomaly, neonatal deaths, neonatal infection, Fatigue, preterm birth, fever, cough, rash, pain, anaemia, leukopenia, seizure, psychosis, Thrombocytopenia, Gastro enteritis and other adverse events) Descriptive (dizziness, vomiting, headache, nausea, asthenia, maternal and neonatal deaths, miscarriages, congenital malformations, anaemia, Haemoglobin, AE by system organ class) Fishers exact/chi-square (dizziness, vomiting, headache, nausea, asthenia, maternal and neonatal deaths, miscarriages, congenital malformations, anaemia, AE by system organ class)
González 2014 (41)	Descriptive (miscarriages, stillbirths, premature births, and congenital malformations, headache, dizziness, vomiting, nausea) Fishers exact/chi-square (miscarriages, stillbirths, premature births, and congenital malformations, headache, dizziness, vomiting, nausea)
González 2014 (42)	Descriptive (cutaneous reaction, Anaemia, Opportunistic infections, leukopenia, haemoglobin, leucocyte count, stillbirth, preterm birth) Fishers exact/chi-square (cutaneous reaction, Anaemia, Opportunistic infections, leukopenia, stillbirth, preterm birth)
Klement 2014 (43)	Descriptive (Infections and infestations, Nervous system and other disorders, stillbirth, preterm birth, abortion, tendinitis, deaths Septicaemia with empetigo, pre-eclampsia.)
Manyando 2014 (44)	Descriptive (abortion or still birth, neonatal death, perinatal death, infant death; Pregnancy, puerperium, and perinatal disorders, Infections and infestations, Gastrointestinal disorders, Surgical and medical procedures, Blood and lymphatic system disorders, Musculoskeletal and connective tissue disorders, Nervous system disorders, Injury, poisoning, and procedural complications, Respiratory, thoracic, and mediastinal disorders, General disorders and administration site conditions, Immune system disorders, Vascular disorders, Congenital, familial, and genetic disorders, Pregnancy, puerperium, and perinatal disorders, Infections and infestations, General disorders and administration site conditions, Respiratory, thoracic, and mediastinal disorders, Gastrointestinal disorders) Fisher's exact/chi-square (abortion or still birth, neonatal death, perinatal death, infant death) Poisson (Pregnancy, puerperium, and perinatal disorders, Infections and infestations, Gastrointestinal disorders, Surgical and medical procedures, Blood and lymphatic system disorders, Musculoskeletal and connective tissue disorders, Nervous system disorders, Injury, poisoning, and procedural complications, Respiratory, thoracic, and mediastinal disorders, General disorders and administration site conditions, Immune system disorders, Vascular disorders, Congenital, familial, and genetic disorders, Pregnancy, puerperium, and perinatal disorders, Infections and infestations, General disorders and administration site conditions, Respiratory, thoracic, and mediastinal disorders, Gastrointestinal disorders)
Desai 2015 (45)	Descriptive (anaemia, maternal death, abortion, still birth, emergency caesarean section, hypertension and other maternal SAEs, preterm birth, congenital anomaly, jaundice, infection, birth asphyxia, meconium aspiration syndrome, cephalohaematoma, dizziness, weakness, vomiting, nausea, pruritus, abdominal pain, headache, diarrhoea, facial swelling, loss of appetite, feeling hot, dyspepsia, other)
Unger 2015 (48)	Descriptive (anaemia, maternal death, abortion, still birth, emergency caesarean section, hypertension and other maternal SAEs, preterm birth, congenital anomaly, jaundice, infection, birth asphyxia, meconium aspiration syndrome, cephalohaematoma, dizziness, weakness, vomiting, nausea, pruritus, abdominal pain, headache, diarrhoea, facial swelling, loss of appetite, feeling hot, dyspepsia, other)
	Fishers exact/Chi-square (anaemia, maternal death, abortion, still birth, emergency caesarean section, hypertension and other maternal SAEs, preterm birth, congenital anomaly, jaundice, infection, birth asphyxia, meconium aspiration syndrome, cephalohaematoma, dizziness, weakness, vomiting, nausea, pruritus, abdominal pain, headache, diarrhoea, facial swelling, loss of appetite, feeling hot, dyspepsia, other)
Kakuru 2016 (49)	Descriptive (vomiting, abdominal pain, cough, headache, chills, diarrhoea, malaise, dysphagia, nausea, infection, anorexia, anaemia, congenital anomaly, stillbirth, thrombocytopenia , vaginal bleeding, adverse birth outcomes, white-cell count, neutrophil count, platelet count, haemoglobin, alanine aminotransferase level) Epidemiologic statistic , incidence rate (adverse event of any severity)
Kimani 2016 (50)	Descriptive (deaths, stillbirths, Infections and infestations, Pregnancy, puerperium, and perinatal conditions, Congenital, familial, and genetic disorders, Respiratory, thoracic, and mediastinal disorders, concomitant treatment, vital signs) Descriptive (QT interval, abortion, stillbirth, congenital anomaly, anaemia, vomiting, abdominal pain, cough, headache, anaemia, malaise, diarrhoea, chills, thrombocytopenia, anorexia, congenital anomaly, stillbirth) Fishers exact/chi-square (abortion, stillbirth, congenital anomaly, anaemia, vomiting, abdominal pain, cough, headache, anaemia, malaise, diarrhoea, chills, thrombocytopenia, anorexia, congenital anomaly, stillbirth, altered mental status, elevated ALT level)
Natureeba 2017 (53)	Descriptive (dizziness, vomiting, palpitations, headache, nausea, abdominal pain) Fishers exact/ chi-square (dizziness, vomiting, palpitations, headache, nausea, abdominal pain)
Divala 2018 (56)	Fishers exact/chi-square (Apgar scores, preterm delivery, nausea, vomiting, headache, gastric pain, dizziness)
Akinyotu 2018	Fishers exact/chi-square (Fetal or neonatal death, anaemia, QTc prolongation, vomiting, abdominal pain, cough, headache, pyuria, diarrhoea, malaise, congenital anomaly, Thrombocytopenia, altered mental status, elevated alanine aminotransferase, respiratory distress, haemorrhage) Negative binomial regression (Fetal or neonatal death, anaemia)
Kajubi 2019	

Fig S7.1: Treatment schedule over a follow-up time for women enrolled receive malaria chemoprevention treatment in Malawi

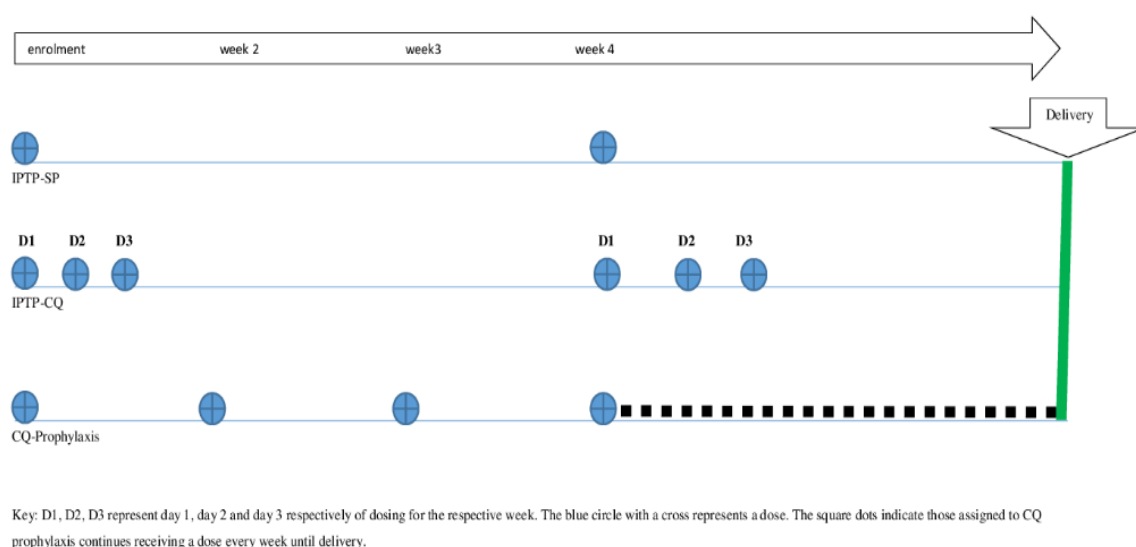


Table s7.1: Distribution of AEs by system organ class across chemoprevention treatments

Table S1: Distribution of SOC of AEs before delivery by treatment group

System Organ Class of the AE	IPTp-SP n (%)	IPTp-CQ n (%)	Total
Blood and lymphatic system disorders	9 (1.38)	11 (1.23)	20 (1.29)
Cardiac disorders	8 (1.23)	20 (2.23)	28 (1.81)
Eye disorders	7 (1.08)	12 (1.34)	19 (1.23)
Gastrointestinal disorders	109 (16.77)	204 (22.79)	313 (20.26)
General disorders and administration site conditions	44 (6.77)	74 (8.27)	118 (7.64)
Infections and infestations	192 (29.54)	165 (18.44)	357 (23.11)
Injury, poisoning and procedural complications	6 (0.92)	8 (0.89)	14 (0.91)
Investigations	3 (0.46)	6 (0.67)	9 (0.58)
Metabolism and nutrition disorders	0(0.00)	3 (0.34)	3 (0.19)
Musculoskeletal and connective tissue disorders	75 (11.54)	98 (10.95)	173 (11.20)
Neoplasms benign, malignant and unspecified (including cysts and polyps)	1 (0.15)	0 (0.00)	1 (0.06)
Nervous system disorders	83 (12.77)	149 (16.65)	232 (15.02)
Pregnancy, puerperium and perinatal conditions	51 (7.85)	56 (6.26)	107 (6.93)
Psychiatric disorders	0 (0.00)	1 (0.11)	1 (0.06)
Renal and urinary disorders	17 (2.62)	27 (3.02)	44 (2.85)
Reproductive system and breast disorders	14 (2.15)	10 (1.12)	24 (1.55)
Respiratory, thoracic and mediastinal disorders	11 (1.69)	11 (1.23)	22 (1.42)
Skin and subcutaneous tissue disorders	5 (0.77)	16 (1.79)	21 (1.36)
Vascular disorders	15 (2.31)	24 (2.68)	39 (2.52)
Total	650 (100)	895 (100)	1545 (100)

Table S7.2: Distribution of AEs by among pregnant women on IPTp

Table S2: Distribution of adverse events by amongst the pregnant women on IPTp treatment

AE characteristic	IPTp-SP	IPTp-CQ
Patient AE experience % (n)	226 (75.3)	248 (82.7)
Median AE frequency (IQR)	2 (1, 3)	2 (1, 4)
*AE SEVERITY % (n)		
1	21.9 (142)	32.3 (289)
2	74.9 (487)	65.3 (584)
≥ 3	3.2 (21)	2.5 (22)
*Medication-related AEs % (n)	0.77 (5)	23.4 (209)

*AE severity and medication related AEs is based patients who experienced AEs (n=474)

Appendix 2: Original articles published

Paper 1: Published in *Malaria Journal*

Patson et al. *Malar J* (2020) 19:119
<https://doi.org/10.1186/s12936-020-03190-z>

Malaria Journal

RESEARCH

Open Access



Systematic review of statistical methods for safety data in malaria chemoprevention in pregnancy trials

Noel Patson^{1,2}, Mavuto Mukaka^{3,4}, Kennedy N. Otworld^{1,5}, Lawrence Kazembe⁶, Don P. Mathanga², Victor Mwapasa², Alinune N. Kabaghe², Marinus J. C. Eijkemans⁷, Miriam K. Laufer^{8*} and Tobias Chirwa¹

Abstract

Background: Drug safety assessments in clinical trials present unique analytical challenges. Some of these include adjusting for individual follow-up time, repeated measurements of multiple outcomes and missing data among others. Furthermore, pre-specifying appropriate analysis becomes difficult as some safety endpoints are unexpected. Although existing guidelines such as CONSORT encourage thorough reporting of adverse events (AEs) in clinical trials, they provide limited details for safety data analysis. The limited guidelines may influence suboptimal analysis by failing to account for some analysis challenges above. A typical example where such challenges exist are trials of anti-malarial drugs for malaria prevention during pregnancy. Lack of proper standardized evaluation of the safety of antimalarial drugs has limited the ability to draw conclusions about safety. Therefore, a systematic review was conducted to establish the current practice in statistical analysis for preventive antimalarial drug safety in pregnancy.

Methods: The search included five databases (PubMed, Embase, Scopus, Malaria in Pregnancy Library and Cochrane Central Register of Controlled Trials) to identify original English articles reporting Phase III randomized controlled trials (RCTs) on anti-malarial drugs for malaria prevention in pregnancy published from January 2010 to July 2019.

Results: Eighteen trials were included in this review that collected multiple longitudinal safety outcomes including AEs. Statistical analysis and reporting of the safety outcomes in all the trials used descriptive statistics; proportions/counts ($n = 18$, 100%) and mean/median ($n = 2$, 11.1%). Results presentation included tabular ($n = 16$, 88.9%) and text description ($n = 2$, 11.1%). Univariate inferential methods were reported in most trials ($n = 16$, 88.9%); including Chi square/Fisher's exact test ($n = 12$, 66.7%), t test ($n = 2$, 11.1%) and Mann-Whitney/Wilcoxon test ($n = 1$, 5.6%). Multi-variable methods, including Poisson and negative binomial were reported in few trials ($n = 3$, 16.7%). Assessment of a potential link between missing efficacy data and safety outcomes was not reported in any of the trials that reported efficacy missing data ($n = 7$, 38.9%).

Conclusion: The review demonstrated that statistical analysis of safety data in anti-malarial drugs for malarial chemoprevention in pregnancy RCTs is inadequate. The analyses insufficiently account for multiple safety outcomes potential dependence, follow-up time and informative missing data which can compromise anti-malarial drug safety evidence development, based on the available data.

Keywords: Pregnancy, Malaria, Safety, Prevention, Clinical trials, Statistical methods

*Correspondence: mlaufer@som.umaryland.edu

⁸ Center for Vaccine Development and Global Health, University of Maryland, School of Medicine, 685 W. Baltimore St., HSF-1 Room 480, Baltimore, MD 21201, USA

Full list of author information is available at the end of the article

Background

Drug safety assessment in randomized controlled trials (RCTs) is integral in development of comprehensive drug safety profile. Recently, there is enhanced quality



© The Author(s) 2020. This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>. The Creative Commons Public Domain Dedication waiver (<http://creativecommons.org/publicdomain/zero/1.0/>) applies to the data made available in this article, unless otherwise stated in a credit line to the data.

guidance on clinical trial reporting of safety outcomes through adherence to the Consolidated Standards of Reporting Trials (CONSORT) guidelines [1, 2]. However, there is scant literature on standardized ways to statistically analyse the safety outcomes in clinical trials. Although there exist some general regulatory guidelines on safety data analysis, such as International Conference on Harmonization which recommend descriptive statistical methods supplemented by confidence intervals [3, 4], the proposed statistical methods rarely account for the complexity of the collected safety data, e.g., recurrent adverse events (AEs). Effective solutions to statistical analysis of safety data in clinical trials may need to be tailored to specific indications (set of diseases with similar characteristics) since safety data collected are also influenced by the medical condition under study. Absence of standardized guidelines for safety data analysis in specific settings may limit the ability to draw rich conclusions about the safety of the investigational product, based on collected data. Standardized guidelines can simplify integration of safety information from multiple outcomes across RCTs [5] and would ensure optimal use of data in developing the safety profile of the investigational product.

Statistical analysis of safety data in clinical trials is characterized by a challenge of multiple and related endpoints measured over time. The safety endpoints may include clinical and laboratory defined AEs. Laboratory-based AEs are defined based on standard cut-off points for measures such as vital signs (e.g., body temperature), hepato-toxicity measures (e.g., bilirubin level), cardio-toxicity measures (e.g., electrocardiograms), and other tests relevant to the medical indication being studied [5]. The safety endpoints may be correlated within patients and over time such that failure to account for this in an analysis may yield biased estimates and false inference. Furthermore, time to occurrence of the safety endpoint may be very informative in profiling the drug safety. Such data present statistical analysis and interpretation challenges due to the complexity in structure [6]. For instance, in the case of multiple, repeatedly measured, safety outcomes, false positives may arise from multiple statistical testing if appropriate longitudinal or time to event methods and/or multiplicity adjustments are not considered.

In clinical trials, AEs may impact compliance and study participation which may further affect treatment efficacy estimates [7, 8]. Occurrence of (even mild) AEs due to a drug would lead to non-adherence, leading to informative censoring. The dropping of the patients from the study generates missing data that may lead to biased results if poorly accounted for. Therefore,

safety data analysis accounting for missing data is useful to facilitate identification and characterization of the safety profile of the drug as early as possible. Other analysis challenges include lack of adequate ascertainment and classification of AEs, and limited generalizability of results [9] since some AEs cannot be pre-specified at study design stage.

There are many populations where drug safety assessment is complex. One of the special settings in safety data assessment is the use of drugs to prevent adverse outcomes in pregnancy, currently referred to as intermittent preventive treatment of malaria in pregnancy (IPTp). For example, the World Health Organization recommends that pregnant women receive routine treatment with anti-malarial drugs to clear any malaria infection that is present and also to prevent infection in the weeks after administration [10]. Recent review indicates that methodological issues in studying anti-malarial drugs in pregnancy have prevented firm conclusions on the safety of new anti-malarial drugs in pregnancy [11]. Previous efforts have attempted to standardize safety assessment methodology for anti-malarial drug trials in pregnancy, including study designs and data collection [12, 13]. However, literature remains limited in describing the standard practice in the statistical analysis of safety data that are collected on anti-malarial drugs during pregnancy trials.

The current review focusses on safety assessment in anti-malarial drugs for chemoprevention in pregnancy trials. Since anti-malarial drug for malaria chemoprevention is given repeatedly to healthy pregnant women, it is critical to improve safety assessment in this vulnerable population. Specifically, appropriate statistical analysis of safety outcomes can improve development of anti-malarial drug safety profile. This can be achieved through sufficient use of the data generated during the RCT which provides a comprehensive drug safety insight. This review, therefore, aims at identifying applied statistical methods and their appropriateness in the analysis of safety data in anti-malarial drugs for malaria prevention during pregnancy clinical trials.

Methods

The systematic review was conducted according to Preferred Reporting Items for Systematic Review and Meta-Analyses (PRISMA) statement [14] which outlines minimum standards for reporting systematic reviews and meta-analysis (Additional file 1: Table S1). The protocol for this review was registered and published with PROSPERO (CRD42019120916). The study population is pregnant women on any anti-malarial drug for malaria chemoprevention.

Search strategy

Inclusion criteria

Primary original articles published in English from Phase III RCTs were considered for inclusion. The articles were from RCTs assessing the efficacy and safety of malaria chemoprevention in pregnancy. This review focused on Phase III RCTs, because they have the largest sample size among pre-marketing trials and accommodates multidisciplinary support in safety evaluation. Further, the data are systematically collected and have the benefit of being randomized, which aids a fair comparison of treatment groups.

Exclusion criteria

Observational studies, case reports, letters to the editor, narrative reviews, systematic reviews and trials in Phase I or Phase II or Phase IV were excluded from this review. This review did not include clinical trials on malaria prevention in pregnancy using intermittent screening and treatment (ISTp) as an alternative to IPTp. ISTp refers to intermittent rapid diagnostic testing (RDT) for malaria in pregnant women followed by treatment of RDT-positive cases with an effective artemisinin-based combination therapy, and IPTp is given to pregnant women regardless of their malaria status. Hence, ISTp and IPTp consider different populations which may confound the practice in safety assessment methods (i.e., ISTp considers symptomatic population and IPTp considers both symptomatic and asymptomatic population). Non-English publications were excluded.

Selection of studies

Studies published between 1 January 2010 and 31 July, 2019 were searched from five databases (PubMed, Embase, Scopus, Malaria in Pregnancy Library (MiPL) and Cochrane Central Register of Controlled Trials (CENTRAL). The MiPL is an excellent scholarly source of articles on malaria in pregnancy that enabled the review to capture both indexed and non-indexed articles beyond the searched databases. Additional searches included reference lists of the identified trials and relevant reviews to identify trials potentially missed in the database search. The year 2010 was selected since it is when CONSORT guideline updated and emphasized on appropriate statistical analysis and reporting of clinical trials [15]. Conference proceedings were not included because they usually contain abstracts that do not give detailed analysis of the presented results and they are not rigorously peer-reviewed. The review focussed on published studies only so no experts or abstract publication authors were contacted for unpublished data. The key search items included: *malaria, anti-malarial drug, pregnancy,*

efficacy safety or tolerability. The detailed search strategy is presented in Additional file 2: Table S2. The search was customized per database. Based on PRISMA procedure, after removing duplicates, two reviewers (NP and ANK) independently screened titles and abstracts initially before arriving at a final list of eligible articles. Based on the eligible studies list, full text articles were retrieved. The references were managed using Endnote X7.1 (Thomson Reuters). If there were disagreements, the two reviewers discussed the paper to reach a consensus and reasons for exclusion were provided for ineligible publications/studies

Data extraction

The data extraction file created in Microsoft Excel was used to record all key variables from the selected articles. Some of the collected variables such as mode of safety data collection, participant withdrawal due to AE and handling of continuous measures were based on CONSORT guideline. The following key variables were extracted from the papers: main author, publication date, study design, study location, main efficacy outcome, sample size, list of safety parameters collected (including laboratory data), nature of safety data collection (i.e., passive or active), list of statistical methods used for respective safety outcomes, how the results were presented, retention rate at the end of the follow up and how missing safety or efficacy data were handled. The primary hypothesis type (as superiority, non-inferiority or equivalence) was defined based on what was reported in the actual manuscript or inferred by the lead author (NP), based on how the study framed the primary hypothesis. Superiority hypotheses aim to show whether treatment is better than control, non-inferiority hypotheses intend to show that one treatment is not worse than the other and equivalence hypotheses intend to show that a given treatment is similar to another for defined characteristics [16]. The statistical methods were classified as descriptive or inferential and univariate or multivariate depending on the purpose and nature of the statistical methods based on previous similar reviews [17, 18], reviewing statistical methods.

Data synthesis

The extracted quantitative data were reported as percentages in tables. The commonly reported safety parameters, suitability of the used statistical methods and other findings were also summarized narratively.

Results

The search identified 1103 articles. After removing duplicates, 722 unique articles were identified and considered for possible inclusion in the review. The duplicates (i.e.

repeated citations) were the same articles but identified in multiple search databases. Figure 1 presents details of the selection process. During screening, a total of 637 articles were excluded based on relevance of their titles and abstracts. The remaining 85 full text articles were assessed for possible inclusion and 18 articles satisfied the inclusion criteria, and were included in this review as shown in Table 1. Reasons for exclusion are shown in Fig. 1.

Characteristics of the trials

The trials included in this review were conducted in Oceania (2 trials, 11.1%) and sub-Saharan Africa (16 trials, 88.9%) regions. The 18 RCTs reviewed recruited 26,281 pregnant women with a median sample size of 374 (interquartile range (IQR): 173, 648) women per treatment arm in a trial. Thirteen trials (72.2%) recruited more than 200

patients per arm. As expected, all trials (18 trials, 100%) computed sample size based on the efficacy outcome(s). The majority of the trials (11 trials, 65.4%) had two treatment arms and the rest had three treatment arms. All 18 trials had an active comparator and IPTp-SP was studied as a standard malaria chemoprevention in the majority of trials (14 trials, 77.8%). Although the review focussed on published trials from 2010 to 2019, the trials were conducted between 2003 and 2017. Based on the primary hypothesis tested, superiority design RCTs were the most common (15 trials, 83.3%) and the other trials had a non-inferiority hypothesis. Over half of the trials (10 trials, 55.6%) were open-label; one trial did not state the blinding status but the other 7 trials (38.9%) were blinded. Majority of the trials (14 trials, 77.9%) reported that they had a Data Safety and Monitoring Board (DSMB).

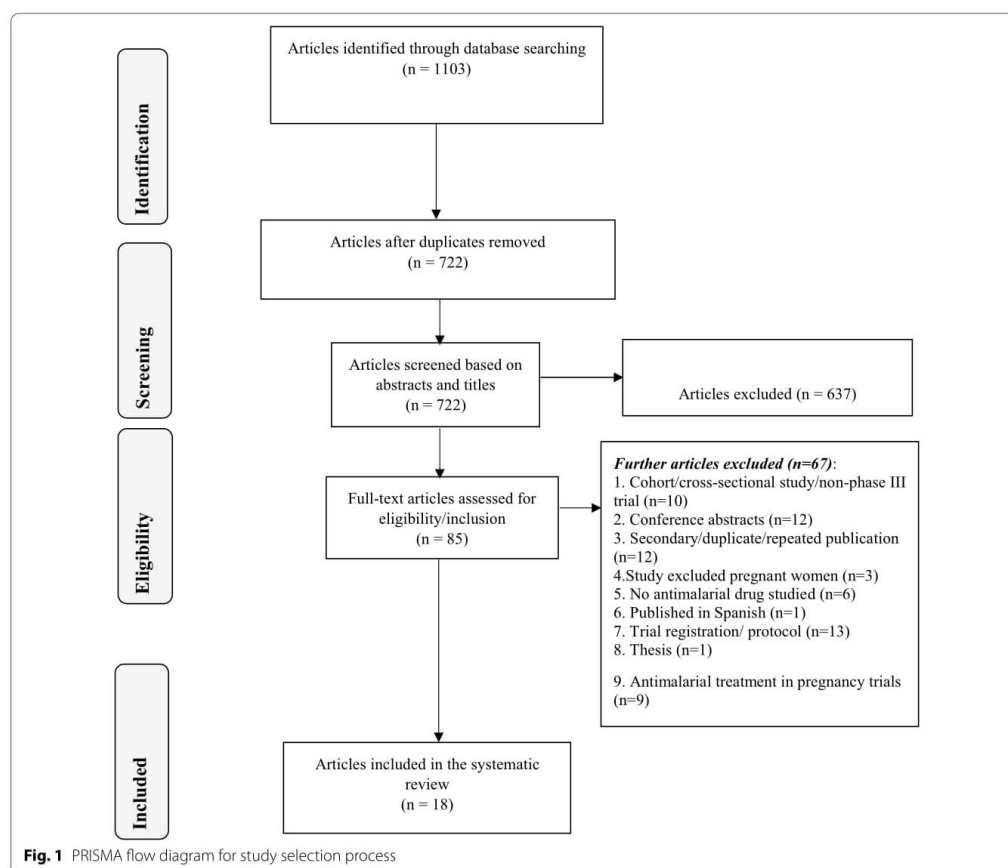


Table 1 Overview of randomized clinical trials included in the systematic review

1st Author, year (reference)	Study site	DSMB	Hypothesis	Masking	Comparator drug(s)	Experimental drug(s)
Luntamo 2010 [39]	Malawi	Yes	Superiority	Single-blinded	Two-dose IPTp-SP	AZ-SP Monthly SP
Valea 2010 [40]	Burkina Faso	No	Superiority	Not stated	Two-dose IPTp	Three IPTp-SP
Diakite 2010 [41]	Mali	Yes	Superiority	Open-label	Two-dose IPTp	Three IPTp-SP
Ndyomugenyi 2011 [42]	Uganda	Yes	Superiority	Double blinded	IPTp-SPITN + placebo	IPTp-SP + ITN
Wini 2013 [43]	Solomon Islands	Yes	Superiority	Open-label	IPTp-SP	-CQ prophylaxis
Denoeud-Ndam 2014 [44]	Benin	Yes	Noninferiority	Open-label	IPTp-MQ, (IPTp-MQ) and CTX (low CD4)	CTX (low CD4)/CTX (high CD4)
González 2014 [45]	Benin, Gabon, Mozambique, and Tanzania	Yes	Superiority	Open-label	IPTp-SP	Two-dose MQ Split-dose MQ
González 2014 [46]	Kenya, Mozambique, and Tanzania	Yes	Superiority		CTX and Placebo	CTX and (IPTp-MQ)
Klement 2014 [47]	Togo	Yes	Noninferiority	Open-label	IPTp-SP	CTX
Manyando 2014 [48]	Zambia	Yes	Noninferiority	Open-label	IPTp-SP	CTX
Desai 2015 [49]	Kenya	No	Superiority	Open-label	IPTp-SP	IPTp-DP IST-DP
Unger 2015 [50]	Papua New Guinea	Yes	Superiority	Single blinded	IPTp-AZCQ	one-dose SP-CQ
Kakuru 2016 [51]	Uganda	No	Superiority	Double blinded	IPTp-SP,	IPTp-DP IPTp-IPTp-DP
Kimani 2016 [52]	Benin, Kenya, Malawi, Tanzania, and Uganda	Yes	Superiority	Open-label	IPTp-SP	IPTp-AZCQ
Natureeba 2017 [53]	Uganda	No	Superiority	Double blinded	TMP-SMX	TMP-SMX + DP
Divala 2018 [54]	Malawi	Yes	Superiority	Open-label	IPTp-SP	IPTp-CQ CQ prophylaxis
Akinyotu 2018 [55]	Nigeria	No	Superiority	Single-blind	IPTp-SP	IPTp-MQ
Kajubi 2019 [56]	Uganda	No	Superiority	Double-blind	IPTpSP	IPTpDP

AL: Artemether-lumefantrine; CTX: Cotrimoxazole; Q: Quinine; AS: Artesunate; AQ: Amodiaquine; MQ: Mefloquine; CQ: Chloroquine; DP: Dihydroartemisinin-piperaquine; TMP-SMX: Trimethoprim-sulfamethoxazole; PQ: Piperaquine

Characteristics of the reported safety data

Over half of the trials (10 trials, 55.6%) reported that they collected safety data using a combination of scheduled and non-scheduled visits (Table 2), while a third of the trials (6 trials, 33.3%) did not specify the safety data collection approach used. The median retention rate (based on the defined efficacy outcome reported for respective trials) was 89.4% (IQR: 82.5%, 92.4%) and 10 trials (55.6%) had a retention rate below 90%.

All the reviewed trials indicated that they had collected multiple longitudinal safety endpoints. As expected, almost all the trials (17 trials, 94.4%) reported obstetric safety outcomes such as foetal loss. Table S3 and S4 in Additional file 2 provide a detailed list of safety outcomes and respective statistical method reported in each reviewed trial. Despite the reported occurrence of multiple AEs, none of trials seemingly reported recurrence of AEs during pregnancy. In total, 12 trials (66.7%) reported adverse events with different severity levels, e.g., mild, moderate and severe. All trials reported occurrence of AEs by treatment arm. Almost all trials (17 trials, 94.4%) reported laboratory data in their safety assessment

Table 2 Characteristics and structure of reported safety data (n = 18)

Characteristic	Number of articles (%)
Data Monitoring and Safety Board	14 (77.8)
Data collection approach	
Scheduled visits	1 (5.6)
Scheduled and non-scheduled visits	10 (55.6)
Non-scheduled visits	1 (5.6)
Not specified	6 (33.3)
Reported AE severity levels	13 (72.2)
Assessed AE vs adherence	3 (16.7)
Discontinue treatment due to AE	18 (100)
Reported continuous outcome dichotomization	18 (100)
Reported laboratory data	18 (100)
Multiple safety endpoints	18 (100)
Longitudinal safety endpoints	18 (100)
Reported recurrent AEs	0 (0.0)
Retention rate (IQR)	89.4% (IQR: 82.5%, 92.4%)

The retention rate is based on primary outcome reported in trial flow diagram
IQR: interquartile range

of the drug and 16 trials of these (88.9) dichotomized at least a single continuous safety outcome (e.g., haemoglobin) based on standard cut-off points, to define an AE.

Statistical analysis for safety data analysis

Analysis population and missing data

The safety analysis approach (based on treatment allocation and adherence) was specified and reported in 11 trials (61.1%). Per protocol (PP) and intention to treat (ITT) analysis approaches were used in 5 trials (27.7%) and 4 trials (22.2%), respectively. Two trials indicated that they used both PP and ITT to analyse the safety data. Although all the reviewed trials had at least one patient lost to follow-up, only 7 trials (38.9%) reported missing efficacy data and 2 of the 7 trials indicated that the missingness was ignorable after exploring data missingness patterns (Table 3). None of the reviewed trials conducted an advanced sensitivity analysis on the relationship between missing data and drug safety. For example, none of the studies assessed the safety outcomes (e.g., AEs) in relation to missing efficacy outcomes. This review found that most trials (16 trials, 88.9%) had at least one participant who experienced an AE leading to discontinuation from the trial although the studies did not formally investigate/quantify the relationship between the AEs and trial completion.

Reported statistical methods

All the trials included reviewed used descriptive statistics as one of the methods to summarize AEs (Table 3). Proportions or counts were the descriptive statistics used in all of the studies to report safety data. Definition safety data was dependent on respective trials as shown in the Additional file 2: Table S3 and S4). Incidence rates were reported in 6 trials (33.3%). Most trials (16 trials, 88.9%) reported univariate inferential statistical methods; these included Chi square or Fisher's exact test (12 trials, 66.7%), t-test ($n=4$, 22.2%). Only 3 trials reported multivariate statistical methods. The multivariable methods were Poisson regression ($n=3$, 16.7%), and negative binomial regression ($n=1$, 5.6%). Usage of at least two inferential statistical methods to compare safety outcomes was reported in 5 trials (27.8%). Although all studies reported multiple safety outcomes, none reported adjustment for multiplicity during analysis.

The review showed that at least a single optimal statistical methods was reported in 3 trials (16.7%) that considered multivariable modelling. Even though univariable analysis comparing arms in an RCT were appropriately used, further inferential statistical methods reported in the rest of the trials were suboptimal for the type of data being collected. For further details, Additional file 2:

Table 3 Statistical approaches for safety data analysis

Statistical method/approach	Number of articles (%)
Descriptive statistics	18 (100)
Proportion/count	18 (100)
Mean/median	1 (5.6)
Incidence rate	6 (33.3)
Inferential methods	
Univariate methods	
Fisher's exact test/Chi square	12 (66.7)
t-test	4 (22.2)
Linear regression	1 (5.6)
Mann-Whitney/Wilcoxon	1 (5.6)
Multivariable modelling	
Poisson regression	3 (16.7)
Negative binomial	1 (5.6)
Number of inferential methods	
None	3 (16.7)
One method	10 (55.6)
More than one method	5 (27.8)
Analysis approach	
Reporting missing data	
Missing efficacy data reported	7 (38.9)
Handling missing efficacy data	
Imputation	1 (5.6)
Analysis population definition	
ITT	4 (22.2)
PP	5 (27.8)
ITT + PP	2 (11.1)
Not specified	7 (38.9)
Multiplicity adjustment for treatment	0 (0.0)
Results presentation	
Graphs	0 (0.0)
Tables	16 (88.9)
Point estimate and confidence interval	10 (55.6)
P value	16 (88.9)

Entries in this table are mutually exclusive and may not add up to the total 26 articles since some articles used more than one approach

ITT: intention to treat; PP: per protocol

Tables S4 provide a detailed list of reported statistical methods with their respective safety outcome(s).

Presentation of safety outcomes estimates

In terms of presentation of results, none of the trials presented AEs in a graph. Only 2 trials (11.1%) narratively presented the safety results; the other 16 trials (88.9%) presented the results in tabular format. A total of 14 trials reported p-values after comparing treatments and there

were only 10 trials (55.6%) that reported point estimates with their respective confidence intervals.

Discussion

This review sought to provide a detailed overview of the actual practice of the statistical analysis of safety data in the unique setting of drug trials for the preventions of malaria in pregnancy as reflected published literature. The results demonstrate that there is limited reporting of statistical analyses of safety data, at the end of the trial, in these published reports. The findings are useful to advance the development of standardized guidelines for safety data statistical analysis in analysis in anti-malarial drugs in pregnancy trials and related fields. Such guidelines will not replace but rather complement the CONSORT guidelines that are general (i.e., not providing specific statistical methods in analysing harms in RCTs). Based on the authors' knowledge of the available literature, this is the first paper to review statistical methods for safety data in anti-malarial drugs in pregnancy.

Descriptive methods were commonly used to summarize safety data. This review found that each clinical trial used at least one descriptive method to summarize safety data. Univariate statistical methods such as Chi square or Fishers exact tests were used in two-thirds of the articles reviewed. Such descriptive statistics and univariate statistical inference ignored useful information such as variability in follow-up time, missing data and correlation (for those trials which had their multiple safety outcomes repeatedly measured). Hence there was inefficient data use during analysis that may lead to a loss of useful information for improved and informative conclusions. Although a third of the reviewed trials attempted to use crude incidence, the analyses failed to adequately account for individual patient follow-up-time and potential confounders.

All trials dichotomized at least a single continuous clinical laboratory safety outcome (i.e., where AE was defined based on standard cut-off points for adult toxicity). Although this aids in providing time-specific drug safety status and easy interpretation, the dichotomized outcome may miss some information on the magnitude of the temporal changes, overtime during the trial. The information loss may lead to reduction in statistical power to detect safety signal if it exists. Valid longitudinal methods (used without restriction on cut-off points) can address the information loss by exploiting potential within-subject correlations for the repeated clinical laboratory measurements [19–21]. Furthermore, the longitudinal methods can provide the basis for developing improved cut-off points tailored to pregnant women in malaria-specific settings. To ensure improved uptake of such methods, future work needs to strive towards

making the results from the longitudinal methods feasibly interpretable to the medical practitioners.

Only three studies appropriately used multivariable statistical methods. Adjusting for known prognostic covariates is useful to control for confounding that can be introduced due to imbalance when assessing if treatment is independently associated with safety outcome(s). Of secondary interest, covariate adjustment also preserves type I error [22]. Such adjustment for potential confounders (e.g., age) in safety data analysis are suitable in clinical trials with at least moderate sample size, unlike small sample sizes that lead to unstable estimates. Of specific interest in this review, the Poisson model was more suitable in the context of rare AEs which usually have low event rates [21, 23]. Since Poisson regression assumes a constant rate of occurrence of a rare event, it is not ideal for other multiple transient AEs that were common or recurred and would vary in occurrence overtime [24]. Alternatively, mixed effects models could be considered to characterize the safety events over time since they capture patient-specific effects [25, 26]. Whenever time to AE occurrence information is available, survival analysis models may also be preferred to characterize the time to AE occurrence. For recurrent safety events, that may induce dependence, methods that extend the Cox regression model may be preferred; such models include survival mixed effects models (e.g., frailty models) [4].

Almost half of the reviewed trials did not explicitly define the population on which the safety analysis was based. If per protocol analysis is used to address non-adherence there is potential selection bias since it destroys the balance due to randomization. Although CONSORT recommends ITT, as an alternative for analysis of safety endpoints, non-adherence cannot be explicitly addressed with ITT approach since it ignores dropouts, withdrawal or loss-to-follow up for various reasons including safety concerns; ITT-based inference ignores causal effect of the actual treatment received [27]. Patient withdrawal or dropout due to AEs can induce informative censoring useful in quantifying anti-malarial drug safety. For example, if a patient withdraws due to vomiting after taking an anti-malarial drug, their obstetric efficacy outcomes such as birth weight may appear as missing data. In the context of anti-malarial drug for malaria prevention, even mild AE can lead to drug non-adherence. Since the patient has no disease symptoms, they would judge it less costly for them to discontinue the drug than continue experiencing AEs. Hence, inclusion of information on treatment/trial completion status in relation to anti-malarial safety would enrich development of the safety profile of anti-malarial drug in pregnancy. Although study completion status, anti-malarial drug safety and missing data may be interlinked, missing

data received limited attention such that the few trials that considered efficacy missing data did not explicitly explore the potential link. Studying such complex associations requires statistical methods that can appropriately estimate the pathway from the anti-malarial chemoprevention to study completion. Advantageously, methods based on causal inference framework, such as mediation analysis [28–31] could be adapted/extended to assess the influence of the AEs on non-adherence in RCTs.

Despite about three-quarters of the trials reporting p-values after comparing safety outcomes by treatment arms, only about half of the reviewed trials adhered to International Harmonisation Conference Guideline E9 in reporting of confidence intervals in quantifying the safety effect size [3, 4]. Use of confidence interval aids in interpretation of results by providing a measure of precision. Furthermore, graphical displaying of safety data to aid in summarizing of safety data was inadequate. Graphs on safety data have a greater ability to convey insight about patterns, trends, or anomalies that may signal potential safety issues compared to presentation of such data in tabular form only [32]. For example, the graphs could help to visualize frequency and changes in AEs over time by treatment arm. The graphs could further help in assessing assumptions for some statistical methods.

Over three-quarters of the reviewed trials were designed as superiority trials based on efficacy outcomes. Although the statistical approach for safety assessment was mainly on superiority hypotheses (for both the superiority and non-superiority trials), clinical and statistical justification of assessing safety based on superiority hypotheses may be invalid. Superiority hypotheses concentrate on the absence of difference in drug safety effect/risk between or across the treatment arms which may be challenging [16]. For example, when comparing high AE incidences, non-significant difference (when using a superiority hypothesis) would not necessarily translate to a conclusion that a drug is safe and well-tolerated since sometimes all compared treatment arms may have high AE incidence. Perhaps, drug safety evaluation should strive to prove that there is no risk beyond a protocol-defined/hypothesized priori clinical safety margin (i.e., no excessive safety risk). Based on the findings in this work, researchers are encouraged to consider defining safety margins in safety assessment of anti-malarial drugs. Since safety is mostly a secondary outcome, it is not straightforward on how to define a non-inferiority margin and the appropriate analysis population. Currently, it is still unclear and debatable how to implement this, such that further research is needed [5].

Interestingly, over half of the trials were open-label which may influence physician clinical safety

assessment on a patient and patient reporting of AEs based on their expectations since they know the treatment assigned. Appropriate reporting of the AEs would be guided by data safety and monitoring boards (DSMB) from early stages of the trial. However, availability of DSMBs in over three-quarters of the trials did not translate to improved reporting and analysis. Therefore, DSMB members should advocate for improved analysis approaches for AEs.

Table 4 summarizes recommendations to consider on best practices for safety analyses. This provides a general framework for statistical analysis of safety in malaria chemoprevention in pregnancy trials. As highlighted above, the recommendations assume a context where sample size is moderate or large. For rare events, Bayesian approaches can be considered since they do not depend on asymptotic properties when handling rare events and can incorporate prior/external information [33]. Future research work can further consider adapting/extending recently developed statistical methods for rare disease or small population clinical trials towards analysis of rare safety outcomes in IPTp trials [34–36].

This review agrees with other similar publications focusing on drug safety assessment in clinical trials that have noted the need for further improvement in the statistical analysis of the safety data [9, 37]. This review concurs with a recent review that has noted that inappropriate handling of multiple test is prevalent, although their review focussed on four high impact journals, AE in general and a short time of review period [38]. Issues raised in this review include time-dependence of AEs, informative censoring due to discontinuation of treatment because of AEs, safety graphs, and repeated occurrence of AEs and multivariate longitudinal structure of laboratory data that yields complex correlation. This is an ongoing work whereby further analysis will be explored to address the identified statistical issues above.

The application of the systematic review protocol in describing the current practice is highly reliable and objective since it exhaustively identified the published anti-malarial drug clinical trials in pregnancy for studied period. However, this review covered only the last decade of publications and may have missed studies published in other languages or that did not appear in during the literature search. Because the trials reported in the publications spanned for a decade, it was difficult to assess temporal trends. This review represents the most comprehensive review of safety data analysis practice for this important indication.

Table 4 Recommendations for analyses in malaria chemoprevention in pregnancy trials based on key safety outcomes from the systematic review

Safety outcome type after treatment	Examples	Optimal statistical method(s)
Time to event	Time to infection/infestation occurrence, time to abortion	Survival methods (e.g. Kaplan–Meier plots, log rank test followed by cox regression model)
Longitudinal continuous outcomes	QTc prolongation, haemoglobin, vital signs	Mixed effects linear/nonlinear models
Recurrent events	recurrent hospitalizations, recurrent anaemia, recurrent opportunistic infections	Recurrent event models (e.g. Poisson model, negative binomial, Andersen Gill model)
Safety outcomes with informative missing data,	Treatment switching, Dropout/withdrawal due to AE (i.e. AE-induced non-adherence)	Causal inference methods (e.g. mediation analysis)
Multivariate outcomes	Abortion, miscarriage Haemoglobin, QTc, <i>alanine aminotransferase</i> level	Multivariate methods: methods for multiple outcomes (e.g. multivariate regression models based on outcome type), methods for repeated outcomes (e.g. multivariate longitudinal analysis, recurrent events models), advanced graphs
Laboratory measurements with temporal variations	Vital signs, haemoglobin	Advanced graphical methods, multivariate methods
Count	Seizure episodes, dizziness episodes	Incidence rate followed by Models for count data (Poisson regression, negative binomial regression)
Outcomes with zero inflated counts	Seizure episodes, dizziness episodes, nausea episodes	Models for counts (Poisson zero inflated models, negative binomial zero-inflated models)

Conclusion

Although useful safety data are collected in malaria chemoprevention in pregnancy clinical trials, the analysis remains sub-optimal and this hinders definitive conclusions about drug safety in this setting. Descriptive statistical methods and dichotomization of continuous outcomes are predominant which may lead to loss of useful information. The definition of analysis population and informative presentation of results are not standardized. Overall, the results suggest that choice of a statistical method(s) to use should be jointly driven by the scientific question of interest, epidemiological/clinical plausibility of the method and structure of the raw data. Further work in addressing the highlighted gaps can enhance drug safety decisions and conclusions.

Supplementary information

Supplementary information accompanies this paper at <https://doi.org/10.1186/s12936-020-03190-z>.

Additional file 1. PRISMA 2009 checklist for the systematic review.

Additional file 2. Overview of search strategy, safety outcomes and statistical analysis approaches in the reviewed trials.

Abbreviations

CONSORT: Consolidated Standards of Reporting Trials; AE: Adverse event; RCT: Randomized clinical trial; ECG: Electro-cardiograph; ICH: International Conference on Harmonization; IPTp-SP: Intermittent preventive treatment of malaria in pregnancy using sulfadoxine-pyrimethamine; WHO: World Health Organization; PRISMA: Preferred Reporting Items for Systematic Review and Meta-Analyses; PROSPERO: Prospective Register of Systematic Reviews; CTX: Cotrimoxazole; MiPL: Malaria in Pregnancy Library; CENTRAL: Cochrane Central

Register of Controlled Trials; DSMB: Data Safety and Monitoring Board; PP: Per protocol; ITT: Intention to treat; ISTp: Intermittent screening and treatment in pregnancy.

Acknowledgements

We are grateful to Steady Chasimpha, for assisting with the literature search.

Authors' contributions

NP, MKL and TC conceived the study idea and NP, ANK and TC collected, analysed and interpreted the data; NP, MKL and TC led the manuscript writing; LK, MM, KNO, ANK, VM, MJCE and DPM contributed towards the study design, data interpretation and rigorous manuscript review. All authors read and approved the final manuscript.

Funding

This work was supported by the US National Institutes of Health (Grant numbers (D43TW010075) (Fogarty International Center) to MKL and DPM and K24AI114996 (National Institute of Allergy and Infectious Diseases) to MKL).

Availability of data and materials

Not applicable.

Ethics approval and consent to participate

Not applicable.

Consent for publication

Not applicable.

Competing interests

The authors declare that they have no competing interests.

Author details

¹ School of Public Health, University of the Witwatersrand, Johannesburg, South Africa. ² University of Malawi, College of Medicine, Blantyre, Malawi. ³ Mahidol Oxford Tropical Medicine Research Unit (MORU), Bangkok, Thailand. ⁴ Centre for Tropical Medicine, Nuffield Department of Medicine, University of Oxford, Oxford, UK. ⁵ Perinatal HIV Research Unit, Faculty of Health Sciences, University of the Witwatersrand, Johannesburg, South Africa. ⁶ Department of Biostatistics, University of Namibia, Windhoek, Namibia. ⁷ Julius Center for Health Sciences and Primary Care, University Medical Center Utrecht,

Utrecht University, Utrecht, The Netherlands. ⁸ Center for Vaccine Development and Global Health, University of Maryland, School of Medicine, 685 W. Baltimore St., HSF-1 Room 480, Baltimore, MD 21201, USA.

Received: 14 December 2019 Accepted: 12 March 2020

Published online: 20 March 2020

References

1. Tammimga C, Kavanaugh M, Fedders C, Maiolatesi S, Abraham N, Bonhoeffer J, et al. A systematic review of safety data reporting in clinical trials of vaccines against malaria, tuberculosis, and human immunodeficiency virus. *Vaccine*. 2013;31:3628–35.
2. Moher D, Hopewell S, Schulz KF, Montori V, Gøtzsche PC, Devereaux PJ, et al. CONSORT 2010 Explanation and Elaboration: updated guidelines for reporting parallel group randomised trials. *BMJ*. 2010;340:c869.
3. Lewis JA. Statistical principles for clinical trials (ICH E9): an introductory note on an international guideline. *Stat Med*. 1999;18:1903–42.
4. ICH Harmonised Tripartite Guideline. Statistical principles for clinical trials. International Conference on Harmonisation E9 Expert Working Group. *Stat Med*. 1999;18:1905–42.
5. Zink RC, Marchenko Q, Sanchez-Kam M, Ma H, Jiang Q. Sources of safety data and statistical strategies for design and analysis: clinical trials. *Ther Innov Regul Sci*. 2018;52:141–58.
6. Munsaka MS. A question-based approach to the analysis of safety data. In: Peace KE, Chen D-G, Menon S, editors. *Biopharmaceutical Applied Statistics Symposium. Biostatistical Analysis of Clinical Trials*, vol. 2. Singapore: Springer; 2018. p. 193–216.
7. Leporini C, De Sarro G, Russo E. Adherence to therapy and adverse drug reactions: is there a link? *Expert Opin Drug Saf*. 2014;13(Suppl 1):S41–55.
8. Campbell RT, Willox GP, Jhund PS, Hawkins NM, Huang F, Petrie MC, et al. Reporting of lost to follow-up and treatment discontinuation in pharmacotherapy and device trials in chronic heart failure. *Circ Heart Fail*. 2016;9:e002842.
9. Singh S, Loke YK. Drug safety assessment in clinical trials: methodological challenges and opportunities. *Trials*. 2012;13:138.
10. WHO. Updated WHO policy recommendation: intermittent preventive treatment of malaria in pregnancy using sulfadoxine-pyrimethamine (IPTp-SP). Geneva, World Health Organization, 2012. https://www.who.int/malaria/publications/atoz/who_iptp_sp_policy_recommendation/en/. Accessed 3 Mar 2020.
11. D'Alessandro U, Hill J, Tarning J, Pell C, Webster J, Gutman J, et al. Treatment of uncomplicated and severe malaria during pregnancy. *Lancet Infect Dis*. 2018;18:e133–46.
12. Saito M, Gilder ME, Nosten F, Guerin PJ, McGready R. Methodology of assessment and reporting of safety in anti-malarial treatment efficacy studies of uncomplicated falciparum malaria in pregnancy: a systematic literature review. *Malar J*. 2017;16:491.
13. Allen EN, Chandler CIR, Mandimika N, Pace C, Mehta U, Barnes KI. Evaluating harm associated with anti-malarial drugs: a survey of methods used by clinical researchers to elicit, assess and record participant-reported adverse events and related data. *Malar J*. 2013;12:325.
14. Liberati A, Altman DG, Tetzlaff J, Mulrow C, Gotzsche PC, Ioannidis JP, et al. The PRISMA statement for reporting systematic reviews and meta-analyses of studies that evaluate health care interventions: explanation and elaboration. *Ann Intern Med*. 2009;151:W65–94.
15. Schulz KF, Altman DG, Moher D. CONSORT 2010 Statement: updated guidelines for reporting parallel group randomised trials. *BMJ*. 2010;340:c332.
16. Lesaffre E. Superiority, equivalence, and non-inferiority trials. *Bull NYU Hosp Jt Dis*. 2008;66:150–4.
17. Colditz GA, Emerson JD. The statistical content of published medical research: some implications for biomedical education. *Med Educ*. 1985;19:248–55.
18. Emerson JD, Colditz GA. Use of statistical analysis in the New England Journal of Medicine. *N Engl J Med*. 1983;309:709–13.
19. Verbeke G, Fieuws S, Molenberghs G, Davidian M. The analysis of multivariate longitudinal data: a review. *Stat Methods Med Res*. 2014;23:42–59.
20. Rosenkranz GK. Modeling laboratory data from clinical trials. *Computat Stat Data Anal*. 2009;53:812–9.
21. Gould AL. Statistical methods for evaluating safety in medical product development. Hoboken: Wiley; 2015.
22. Kahan BC, Jairath V, Doré CJ, Morris TP. The risks and rewards of covariate adjustment in randomized trials: an assessment of 12 outcomes from 8 studies. *Trials*. 2014;15:139.
23. Gam CMB, Tanniou J, Keiding N, Løkkegaard EL. A model for the distribution of daily number of births in obstetric clinics based on a descriptive retrospective study. *BMJ Open*. 2013;3:e002920.
24. Lawless JF, Nadeau C. Some simple robust methods for the analysis of recurrent events. *Technometrics*. 1995;37:158–68.
25. Rosenkranz GK. An approach to integrated safety analyses from clinical studies. *Drug Inform J*. 2010;44:649–57.
26. Kim H, Shults J, Patterson S, Goldberg-Alberts R. (2008) Analysis of adverse events in drug safety: a multivariate approach using stratified quasi-least squares. <http://biostats.bepress.com/upennbiostat/papers/art29>. Accessed 3 Mar 2020.
27. Dodd S, White IR, Williamson P. A framework for the design, conduct and interpretation of randomised controlled trials in the presence of treatment changes. *Trials*. 2017;18:498.
28. Robins JM. Correcting for non-compliance in randomized trials using structural nested mean models. *Commun Stat Theory Methods*. 1994;23:79–412.
29. Frangakis C, Rubin D. Addressing complications of intention-to-treat analysis in the combined presence of all-or-none treatment-noncompliance and subsequent missing outcomes. *Biometrika*. 1999;86:365–79.
30. Nich C, Carroll KM. Intention-to-treat meets missing data: implications of alternate strategies for analyzing clinical trials data. *Drug Alcohol Depend*. 2002;68:121–30.
31. Ye C, Beyene J, Browne G, Thabane L. Estimating treatment effects in randomised controlled trials with non-compliance: a simulation study. *BMJ Open*. 2014;4:e005362.
32. Amit O, Heiberger RM, Lane PW. Graphical approaches to the analysis of safety data from clinical trials. *Pharm Stat*. 2008;7:20–35.
33. Price KL, Amy Xia H, Lakshminarayanan M, Madigan D, Manner D, Scott J, et al. Bayesian methods for design and analysis of safety trials. *Pharm Stat*. 2014;13:13–24.
34. Friede T, Posch M, Zohar S, Alberti C, Benda N, Comets E, et al. Recent advances in methodology for clinical trials in small populations: the INSPiRe project. *Orphanet J Rare Dis*. 2018;13:186.
35. Hilgers R-D, Bogdan M, Burman C-F, Dette H, Karlsson M, König F, et al. Lessons learned from IDeA—33 recommendations from the IDeA-net about design and analysis of small population clinical trials. *Orphanet J Rare Dis*. 2018;13:77.
36. Mitroiu M, Rengerink KO, Pontes C, Sancho A, Vives R, Pesiou S, et al. Applicability and added value of novel methods to improve drug development in rare diseases. *Orphanet J Rare Dis*. 2018;13:200.
37. Lineberry N, Berlin JA, Mansi B, Glasser S, Berkswits M, Klem C, et al. Recommendations to improve adverse event reporting in clinical trial publications: a joint pharmaceutical industry/journal editor perspective. *BMJ*. 2016;355:i5078.
38. Phillips R, Hazell L, Sauzet O, Cornelius V. Analysis and reporting of adverse events in randomised controlled trials: a review. *BMJ Open*. 2019;9:e024537.
39. Luntamo M, Kulmala T, Mbewe B, Cheung YB, Maleta K, Ashorn P. Effect of repeated treatment of pregnant women with sulfadoxine-pyrimethamine and azithromycin on preterm delivery in Malawi: a randomized controlled trial. *Am J Trop Med Hyg*. 2010;83:1212–20.
40. Valea I, Tinto H, Drabo MK, Huybregts L, Henry MC, Roberfroid D, et al. Intermittent preventive treatment of malaria with sulphadoxine-pyrimethamine during pregnancy in Burkina Faso: effect of adding a third dose to the standard two-dose regimen on low birth weight, anaemia and pregnancy outcomes. *Malar J*. 2010;9:324.
41. Diakite OS, Maiga OM, Kayentao K, Traoré BT, Djimde A, Traoré B, et al. Superiority of 3 over 2 doses of intermittent preventive treatment with sulfadoxine-pyrimethamine for the prevention of malaria during pregnancy in Mali: a randomized controlled trial. *Clin Infect Dis*. 2011;53:215–23.
42. Ndyomugenyi R, Clarke SE, Hutchison CL, Hansen KS, Magnussen P. Efficacy of malaria prevention during pregnancy in an area of low and unstable transmission: an individually-randomised placebo-controlled trial using intermittent preventive treatment and insecticide-treated nets

- in the Kabale Highlands, southwestern Uganda. *Trans R Soc Trop Med Hyg.* 2011;105:607–16.
43. Wini L, Appleyard B, Bobogare A, Pikacha J, Seke J, Tunj M, et al. Intermittent preventive treatment with sulfadoxine-pyrimethamine versus weekly chloroquine prophylaxis for malaria in pregnancy in Honiara, Solomon Islands: a randomised trial. *Malar World J.* 2013;4:12.
 44. Denoed-Ndam L, Zannou DM, Fourcade C, Taron-Brocard C, Porcher R, Atadokpede F, et al. Cotrimoxazole prophylaxis versus mefloquine intermittent preventive treatment to prevent malaria in HIV-infected pregnant women: two randomized controlled trials. *J Acquir Immune Defic Syndr.* 2014;65:198–206.
 45. González R, Mombo-Ngoma G, Ouédraogo S, Kakolwa MA, Abdulla S, Accrombessi M, et al. Intermittent preventive treatment of malaria in pregnancy with mefloquine in HIV-negative women: a multicentre randomized controlled trial. *PLoS Med.* 2014;11:e1001733.
 46. Gonzalez R, Desai M, Macete E, Ouma P, Kakolwa MA, Abdulla S, et al. Intermittent preventive treatment of malaria in pregnancy with mefloquine in HIV-infected women receiving cotrimoxazole prophylaxis: a multicenter randomized placebo-controlled trial. *PLoS Med.* 2014;11:e1001735.
 47. Klement E, Pitché P, Kendjo E, Singo A, D'Almeida S, Akouete F, et al. Effectiveness of co-trimoxazole to prevent *Plasmodium falciparum* malaria in HIV-positive pregnant women in sub-Saharan Africa: an open-label, randomized controlled trial. *Clin Infect Dis.* 2014;58:651–9.
 48. Manyando C, Njunju EM, Mwakazanga D, Chongwe G, Mkandawire R, Champo D, et al. Safety of daily Co-trimoxazole in pregnancy in an area of changing malaria epidemiology: a phase 3b randomized controlled clinical trial. *PLoS ONE.* 2014;9:e96017.
 49. Desai M, Gutman J, L'Anziva A, Otieno K, Juma E, Kariuki S, et al. Intermittent screening and treatment or intermittent preventive treatment with dihydroartemisinin-piperaquine versus intermittent preventive treatment with sulfadoxine-pyrimethamine for the control of malaria during pregnancy in western Kenya: an open-label, three-group, randomised controlled superiority trial. *Lancet.* 2015;386:2507–19.
 50. Unger HW, Ome-Kaius M, Wangnapi RA, Umbers AJ, Hanieh S, Suen CS, et al. Sulphadoxine-pyrimethamine plus azithromycin for the prevention of low birthweight in Papua New Guinea: a randomised controlled trial. *BMC Med.* 2015;13:9.
 51. Kakuru A, Jagannathan P, Muhindo MK, Natureeba P, Awori P, Nakalembe M, et al. dihydroartemisinin-piperaquine for the prevention of malaria in pregnancy. *N Engl J Med.* 2016;374:928–39.
 52. Kimani J, Phiri K, Kamiza S, Duparc S, Ayoub A, Rojo R, et al. Efficacy and safety of azithromycin-chloroquine versus sulfadoxine-pyrimethamine for intermittent preventive treatment of *Plasmodium falciparum* malaria infection in pregnant women in Africa: an open-label, randomized trial. *PLoS ONE.* 2016;11:e0157045.
 53. Natureeba P, Kakuru A, Muhindo M, Ochieng T, Ategeka J, Koss CA, et al. Intermittent preventive treatment with dihydroartemisinin-piperaquine for the prevention of malaria among HIV-infected pregnant women. *J Infect Dis.* 2017;216:29–35.
 54. Divala TH, Mungwira RG, Mawindo PM, Nyirenda OM, Kanjala M, Ndaferankhande M, et al. Chloroquine as weekly chemoprophylaxis or intermittent treatment to prevent malaria in pregnancy in Malawi: a randomised controlled trial. *Lancet Infect Dis.* 2018;18:1097–107.
 55. Akinyotu O, Bello F, Abdus-Salam R, Arowojolu A. Comparative study of mefloquine and sulphadoxine-pyrimethamine for malaria prevention among pregnant women with HIV in southwest Nigeria. *Int J Gynaecol Obstet.* 2018;142:194–200.
 56. Kajubi R, Ochieng T, Kakuru A, Jagannathan P, Nakalembe M, Ruel T, et al. Monthly sulfadoxine-pyrimethamine versus dihydroartemisinin-piperaquine for intermittent preventive treatment of malaria in pregnancy: a double-blind, randomised, controlled, superiority trial. *Lancet.* 2019;393:1428–39.

Publisher's Note

Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Ready to submit your research? Choose BMC and benefit from:

- fast, convenient online submission
- thorough peer review by experienced researchers in your field
- rapid publication on acceptance
- support for research data, including large and complex data types
- gold Open Access which fosters wider collaboration and increased citations
- maximum visibility for your research: over 100M website views per year

At BMC, research is always in progress.

Learn more biomedcentral.com/submissions



RESEARCH

Open Access



Comparison of statistical methods for the analysis of recurrent adverse events in the presence of non-proportional hazards and unobserved heterogeneity: a simulation study

Noel Patson^{1,2*}, Mavuto Mukaka^{3,4}, Lawrence Kazembe⁵, Marinus J. C. Eijkemans⁶, Don Mathanga², Miriam K. Laufer⁷ and Tobias Chirwa¹

Abstract

Background: In preventive drug trials such as intermittent preventive treatment for malaria prevention during pregnancy (IPTp), where there is repeated treatment administration, recurrence of adverse events (AEs) is expected. Challenges in modelling the risk of the AEs include accounting for time-to-AE and within-patient-correlation, beyond the conventional methods. The correlation comes from two sources; (a) individual patient unobserved heterogeneity (i.e. frailty) and (b) the dependence between AEs characterised by time-dependent treatment effects. Potential AE-dependence can be modelled via time-dependent treatment effects, event-specific baseline and event-specific random effect, while heterogeneity can be modelled via subject-specific random effect. Methods that can improve the estimation of both the unobserved heterogeneity and treatment effects can be useful in understanding the evolution of risk of AEs, especially in preventive trials where time-dependent treatment effect is expected.

Methods: Using both a simulation study and the *Chloroquine for Malaria in Pregnancy* (NCT01443130) trial data to demonstrate the application of the models, we investigated whether the lognormal shared frailty models with restricted cubic splines and non-proportional hazards (LSF-NPH) assumption can improve estimates for both frailty variance and treatment effect compared to the conventional inverse Gaussian shared frailty model with proportional hazard (ISF-PH), in the presence of time-dependent treatment effects and unobserved patient heterogeneity. We assessed the bias, precision gain and coverage probability of 95% confidence interval of the frailty variance estimates for the models under varying known unobserved heterogeneity, sample sizes and time-dependent effects.

Results: The ISF-PH model provided a better coverage probability of 95% confidence interval, less bias and less precise frailty variance estimates compared to the LSF-NPH models. The LSF-NPH models yielded unbiased hazard ratio estimates at the expense of imprecision and high mean square error compared to the ISF-PH model.

Conclusion: The choice of the shared frailty model for the recurrent AEs analysis should be driven by the study objective. Using the LSF-NPH models is appropriate if unbiased hazard ratio estimation is of primary interest in the

*Correspondence: noelpatson@gmail.com

² School of Global and Public Health, Kamuzu University of Health Sciences, Blantyre, Malawi

Full list of author information is available at the end of the article



© The Author(s) 2022. **Open Access** This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>. The Creative Commons Public Domain Dedication waiver (<http://creativecommons.org/publicdomain/zero/1.0/>) applies to the data made available in this article, unless otherwise stated in a credit line to the data.

presence of time-dependent treatment effects. However, ISF-PH model is appropriate if unbiased frailty variance estimation is of primary interest.

Trial registration: ClinicalTrials.gov; NCT01443130

Keywords: Recurrent adverse events, Randomised controlled trials, Non-proportional hazards, Unobserved heterogeneity

Background

The development of a comprehensive drug safety profile in randomized controlled trials (RCTs) requires statistical methods that can adequately characterize the adverse event (AE) occurrence process (e.g. time of onset, recurrence and duration) [1]. Traditionally, analysis of AEs in IPTp trials, like all clinical trials, is predominantly descriptive and non-parametric such that advanced methods (e.g. parametric and semi-parametric) are rarely used [2–4]. The use of descriptive and non-parametric methods can lead to poor estimates of the effect of treatment on AE recurrence. The descriptive and non-parametric methods may not account for essential aspects of the data structure such as censoring, potential time-varying treatment effect and the AE-dependence. For example, mean cumulative function is the simple and popular non-parametric method used to facilitate the development of trajectories of AEs profiles over the study period, where a patient may experience more than one AEs [5]. However, the method fails to explicitly account for the potential dependence of the AEs or individual patient heterogeneity. The utility of the existing recurrent event methods towards the AEs analysis has been recently considered [6]. However, adopting and interpreting such recurrent events methods results in drug safety assessment setting remains a challenge. This motivates a need for further investigations on the clinical value and practical applicability of the advanced methods [7]. In the framework of benefit-risk assessment, two models (the Andersen Gill model and Prentice, Williams and Peterson model) have been demonstrated to be useful in providing both direct and indirect effects of treatment on AE recurrence [6]. However, these models yield unbiased estimates if the AEs are uncorrelated and do not efficiently capture (i.e. account for) both potential time-dependent treatment effects and unobserved heterogeneity. This motivates the need to consider using flexible parametric shared frailty models that can optimally capture both the time-dependent effects and unobserved heterogeneity, to improve drug safety estimates.

In IPTp clinical trials, the potential of recurrent AEs is high because the preventive antimalarial drugs are repeatedly administered. Recurrent AEs may induce within-patient correlation that originates from two sources: the individual patient unobserved heterogeneity

and the dependence between events [8]. The individual patient heterogeneity cannot be explained by the observed/known factors [9, 10], while the between-AE-dependence can be explained based on the observed AEs. The unobserved heterogeneity arises from unmeasured information that partly explains the treatment effect. For example, in IPTp trials, unobserved heterogeneity can arise from the unmeasured inherent characteristics (e.g. physiological changes associated with foetus development). The underlying individual patient-specific unobserved heterogeneity can influence the susceptibility of patients to AE recurrence. In IPTp trials, as the case with some drug trials, the treatment effect depends on follow-up time. Such treatment effect that is a function of follow-up time is called time-varying or time-dependent treatment effect. In the presence of the time-dependent treatment effect, the relative effect of treatment changes over the follow-up time such that the proportional assumption is violated (on log hazard scale). Ignoring the time-dependent effects can lead to biased estimates. Therefore, accounting for the potential time-dependent treatment effects in the analysis of recurrent events is necessary. However, in practice, identifying a model that can account for both unobserved heterogeneity and time-dependent effects can present a challenge. Since the recurrence of AEs can be a reflection of both time-dependent treatment effects and underlying individual patient unobserved heterogeneity. Therefore, it is vital to consider models that account for both issues.

Currently, shared frailty is one of the prevalent models for the analysis of recurrent events. This model can account for both the observed and unobserved heterogeneity. However, the common current applications of the shared frailty model assume that the treatment effect estimate (i.e. hazard ratio) is constant over the follow-up time [11]. In preventive drug trial settings, where there is repeated treatment administration and a long patient follow-up time, the proportional hazard assumption can be invalid when treatment hazard rates, between/across arms, for recurrent events can vary over the follow-up time. For example, in the context of IPTp drug safety assessment, more recurrent AEs are expected close to the day(s) of taking the drugs. The potential presence of time-dependent treatment effects can pose a challenge in applying the shared frailty model to analyse recurrent

events since it becomes complicated to separate time-dependent effects from unobserved heterogeneity [12].

Recent developments in flexible parametric survival models can effectively capture both time-dependent effects and unobserved heterogeneity [13]. Researchers have established, mainly in univariate survival data context, that flexible parametric survival models with restricted cubic splines can accurately model time-dependent effects [14, 15]. However, there is limited literature investigating the efficiency and utility of such methods in the context of multivariate survival data and drug safety assessment. Within multivariate survival data setting, Gasparini et al. recently compared the standard and spline-based shared frailty models under varying heterogeneity and baseline hazard specification assuming proportional hazards [16]. The shared parametric models with splines emerged to be more robust to model misspecification than the standard approaches. However, their work is limited to the context of the proportional hazards assumption. As highlighted above, a challenge arises in applying the shared frailty model when there are time-dependent effects. Using the standard semi-parametric shared frailty models, Balan and Putter demonstrated that the bias of frailty variance estimate, in the presence of time-dependent effects, can be mitigated as the cluster size (i.e. number of recurrent events within an individual) increases [12]. Based on the available literature, it is not clear whether the use of flexible parametric shared frailty with splines models can improve the frailty estimates compared to the standard shared frailty models in the presence of both time-dependent treatment and unobserved heterogeneity.

Our work aims at investigating whether the flexible parametric shared frailty model with non-proportional hazards and restricted cubic splines can improve both the frailty variance and hazard ratio estimates compared to the standard parametric shared frailty models, in the presence of both unobserved heterogeneity and time-dependent treatment effects. The optimal shared frailty model can efficiently capture both unobserved heterogeneity and time-dependent effects. We evaluated the following models; (a) inverse Gaussian shared frailty (ISF-PH) model with proportional hazards and Weibull baseline hazard (b) Lognormal shared frailty model with non-proportional hazard and restricted cubic splines where we model the baseline hazard function with 1 degree of freedom and model the time-dependent treatment effect with 1 degree of freedom (LSF-NPHS1) and (c) Lognormal shared frailty model with non-proportional hazard and restricted cubic splines where we model the baseline hazard function with 3 degrees of freedom and model the time-dependent treatment effect with 1 degree of freedom (LSF-NPHS3). We compare

the models using a simulation study assuming an application area of recurrent AEs in IP/Tp trials. Understanding the statistical properties would enhance the informed choice of statistical methods for the analysis of recurrent AEs. The simulation offers an opportunity to effectively understand the properties of the statistical models since the true data distribution is known, unlike in a practical clinical trial setting. We present the remaining sections of the paper as follows: In the methods section, we provide a detailed review of the models under study, outline the simulation study design and describe the analysis approach. In the results section, we present the findings of our simulation study and apply the models to the analysis of recurrent AEs observed in *Chloroquine for Malaria in pregnancy* trial (NCT01443130). In the discussion section, we discuss the key findings of our work and provide a conclusion.

Methods

This section outlines notation and provides an overview of the three models that were compared in our simulation study. Specifically, we briefly introduce and discuss the general inferential methodological framework of the models, focussing on model specification, parameter estimation and respective assumptions. The presentation of this section is customized to the context of drug safety assessment in clinical trials. For readers interested in detailed concepts, on recurrent events analysis, they can consult excellent introductory books on recurrent events analysis and frailty models methodology [10, 17, 18], written for a wider audience.

Notation

We assume that the observation begins at enrolment, denoted as time $t=0$. Let T_{ik} be the total time of the k^{th} event for the i^{th} patient where $i=1, 2, \dots, n$, $k=0, 1, 2, \dots, p$ and we assume $T_{i1} < T_{i2} < \dots < T_{ik}$. The AEs are assumed to be recorded over a period of time interval $[0, \tau]$ such that the total number of AEs recorded per patient i over that time is $= \sum_{k=0}^{\infty} I(T_{ik} \leq t)$. The history of the recorded number of events over given time t , denoted as $N(t)$, $0 \leq t$, is defined as a counting process; this forms the basis for easy-to-implement recurrent events analysis. The frailty is defined as ϕ_i and interpreted as an unobserved covariate common for all recurrent AEs within the patient. Covariate and regression coefficients vectors are defined as X_{ik} and β respectively. The covariates from the covariates vector are considered to be fixed. Under the simulation study design described below, one covariate is considered (i.e. treatment). For the non-proportional hazard models, the treatment effect is allowed to be dependent on time (i.e. $\beta(t)$). These are further discussed in context in the section below.

Statistical methods for recurrent events

Recurrent events modelling is popularly based on three main approaches; event counts, event intensity function and joint distribution of gap times between successive events. Event counts-based models are suitable when there are frequent events and event occurrence does not impact the occurrence of the subsequent events [17, 19]. In the context of AEs, some mild AEs may not alter the event process (e.g. mild headache). When events are infrequent and the prediction of the next event is of interest, gap times methods are appropriate [20, 21]. To ensure that the event process history is appropriately incorporated, the intensity function is used; the function defines the instantaneous probability of an event occurring at time t given the event process history [17].

$$\lambda(t|H(t)) = \lim_{\Delta t \rightarrow 0} \frac{\Pr(\Delta N(t) = 1 | H(t))}{\Delta t} \quad (M1)$$

Where $\Delta N(t)$ is the number of events in the time interval $[t, t + \Delta t)$, $H(t)$ is the history of the event process at time t . Since event counts (i.e. Poisson process) modelling approach assumes that the events are independent such that the event history process at t , $H(t)$, does not affect the instantaneous probability of an event at time t , the intensity function is reduced to the rate function $\rho(t)$ below;

$$\lambda(t|H(t)) = \lim_{\Delta t \rightarrow 0} \frac{\Pr(\Delta N(t) = 1)}{\Delta t} = \rho(t) \quad (M2)$$

This implies; $\Delta t \rho(t) = E(\Delta N(t))$. Hence if we let the expected cumulative rate at time t to be $\mu(t)$, we can derive the expected cumulative rate by integrating the rate function over the interval $[0, t]$.

$$\mu(t) = \int_0^t \rho(s) ds \quad (M3)$$

The gap/waiting times method is another alternative approach that focusses on modelling time between consecutive events. Methods based on gap time are suitable mainly in setting where the occurrence of the events is considered relatively infrequent [17]. The work in this paper is confined to multiplicative intensity-based models because it permits deriving and interpretation of failure intensity relative to the treatment under investigation. Shared frailty models are considered as important intensity-based models for the analysis of recurrent events that can account for both measured and unmeasured heterogeneity.

The shared frailty models

In practice, the common approach to account for unobserved heterogeneity in recurrent events analysis is to use

a robust variance (i.e. robust sandwich covariance matrix is used in estimating the log hazard ratio). Adjusting for within-subject unobserved heterogeneity using robust variance has been shown to be inadequate in some settings [22]. One of the solutions to this is explicitly modelling the correlation using patient-specific random effect. Such a model is defined as a frailty model. Frailty models are hazard models having a multiplicative frailty factor [10]; these are conditional models. Overall, the model has frailty, linear predictor and the baseline hazard function as the multiplicative factors. The random effect introduced in frailty models describes the excess risk that cannot be explained by the measured/observed variables [10]. The unmeasured observations may introduce heterogeneity between the analysis units e.g. for recurrent AEs, observations within a patient may be more similar compared to different patients. Hence frailty model attempts to account for the within-patient correlation. Ignoring the potential heterogeneity in modelling recurrent events can lead to underestimation of treatment effect [23, 24].

For recurrent events, when the frailty is modelled as a common random effect for all events within a patient, it is defined as a shared frailty model (see model M4). The dependence accounted for in the shared frailty model is only due to the unobserved individual heterogeneity, ϕ_i . This measure of the extent of heterogeneity among the patients can follow several distributions [18]. Gamma and inverse Gaussian distributions are the common frailty distribution specifications, possibly due to their mathematical tractability and software availability. Across all the patients, the respective random effects are assumed to be independent and identically distributed.

$$\lambda_{ik}(t_{ij}|\phi_i, X_{ik}) = \phi_i \lambda_0(t_{ij}) \exp(X_{ik}^T(t) \beta) \quad (M4)$$

If the baseline hazard, $\lambda_0(t_{ij})$, is unspecified, the model is called a semi-parametric frailty model. Specifying the baseline hazard function ensures that an appropriate shape of the hazard function is captured. However, miss-specification of the baseline hazard distribution may induce bias on regression coefficients [16]. Therefore, caution needs to be exercised on the choice of the baseline hazard distribution by paying attention to the most clinically plausible hazard function. Traditionally, the hazards are assumed proportional over the whole follow-up time. Within the flexible parametric survival methods framework, the proportional hazards assumption can be relaxed to accommodate time-dependent treatment effects (i.e. non-proportional hazards [13]). This can be implemented by interacting treatment with splines of log time as part of the linear predictor. Fitting such flexible non-proportional shared frailty

model on log hazard scale is helpful since it can effectively accommodate interpretation and presentation of results. The shared frailty model with non-proportional hazards and restricted cubic splines can be written as;

$$\begin{aligned} \log \{ \lambda_{ik}(t_{ij} | \varphi_i, X_{ik}) \} = & s\{\log(t) | \gamma, k_0\} \\ & + X_{ik}^T \beta + \log(\varphi_i) \\ & + \sum_{f=1}^F s\{\log(t) | \delta_f, k_f\} x_{ikf} \end{aligned} \quad (M5)$$

The treatment x_{ik} is interacted with restricted cubic function of log time, $s\{\log(t) | \gamma, k_0\}$, such that γ is parameter vector and k_0 is knot vector. Similarly, $s\{\log(t) | \delta_f, k_f\}$ represents a spline function of log time for the f^{th} time-dependent effect with parameter vector δ_f and knots vector k_f . Since the number of restricted cubic splines can determine the most appropriate model to use we considered to evaluate two flexible parametric shared frailty models as follows;

- (i) **LSF-NPHS1:** We considered a shared frailty model assuming lognormal frailty distribution and non-proportional hazards with restricted cubic splines where the baseline hazard function is modelled using 1 degree of freedom and the time-dependent treatment effect is modelled using 1 degree of freedom.
- (ii) **LSF-NPHS3:** We also considered a shared frailty model was fitted assuming lognormal frailty distribution and non-proportional hazards with restricted cubic splines where the baseline hazard function is modelled using 3 degrees of freedom and the time-dependent treatment effect is modelled using 1 degree of freedom.

Generally, under shared frailty models, depending on the model specification, parameter estimation can be done using expectation-maximization algorithm, maximum likelihood penalized partial likelihood and Markov chain Monte-Carlo methods [18]. Since in this paper we focus on parametric shared frailty models from a Frequentist perspective, the parameter estimation is based on maximum likelihood estimation. It should be noted that in drug clinical trials not all patients may experience the AEs (i.e. some patients are censored). Censoring in this paper is considered independent of the AE occurrence. Specifically, the model parameters can be estimated by maximizing marginal log-likelihood as shown below; Let Z be the vector containing the observed patient-specific information such that

$$Z = ((y_{ij}, \xi_{ij}, x_{ij}) | j = 0, 1, \dots, k)$$

where y_{ij} , ξ_{ij} and x_{ij} are time to AE occurrence, AE occurrence indicator (1 for AE occurrence and 0 otherwise) and fixed covariate for patient i and j^{th} AE occurrence, respectively. Let U be a vector containing the latent information (the frailty terms) for n patients.

$$U = (u_1, u_2, \dots, u_n),$$

therefore, the marginal log-likelihood of the observed data Z can be written as:

$$\begin{aligned} l_{\text{marg}}(\xi, \beta, Z | \tau) = & \sum_{i=1}^n \left\{ \left[\sum_{j=0}^k \xi_{ij} (\log(h_0 y_{ij})) + (X_{ij}^T \beta) \right] \right. \\ & + \log[(-1)^{d_i} \left(\mathcal{L}^{(d_i)} \left(\sum_{j=0}^k H_0(y_{ij}) \exp(X_{ij}^T \beta) \right) \right)] \\ & \left. - \log \left[\mathcal{L} \left(\sum_{j=0}^k H_0(\tau_{ij}) \exp(X_{ij}^T \beta) \right) \right] \right\} \end{aligned}$$

where d_i represents the total number of events for patient and $\mathcal{L}^{(v)}(\cdot)$ represents the v^{th} derivative of the Laplace transform for the derivative. Availability of the Laplace transform simplifies the maximization of the log-likelihood in order to compute the desired model parameters. Under the time-dependent treatment effect, the fixed treatment effect β can be replaced by the time-dependent treatment effect $\beta(t)$.

Simulation study

This simulation study was conducted and reported based on "Aims, Data generating process, Method of analysis, Estimands and Performance" approach [25] since it provides a scientifically coherent structured framework for designing, interpreting and reporting simulation studies.

Aim

There is a dearth of literature highlighting the performance of recurrent statistical methods when both AE-dependence and heterogeneity is accounted for in analysis of recurrent AEs in clinical trials. The aim of the simulation study was to evaluate whether the use of Lognormal shared frailty model with restricted splines and non-proportional hazards improves both the frailty variance and hazard ratio estimates in the presence of time-dependent effects and unobserved heterogeneity under varying unobserved heterogeneity levels and sample sizes. The performance of this flexible model was compared against

the standard proportional hazards-based inverse Gaussian shared frailty model with proportional hazard and Weibull baseline hazard distribution assumptions.

Data generating mechanisms

The data generating mechanism (DGM) in this simulation study focussed on the specification and estimation of frailty variance and hazard ratio in the presence of time-dependent treatment effects and unobserved heterogeneity (Table 1). For each scenario, we considered 1000 simulations, lognormal frailty distribution, Weibull baseline hazard, fixed effect treatment effect of log hazard ratio 0.5 and up to a maximum of 4 AEs experienced by a patient. The choice of possible maximum of 4 AEs per patient was aimed at mimicking a setting for moderately high occurrence of recurrent AEs, in IPTp trials. The choice of lognormal frailty distribution was driven by the mathematical tractability of the distribution, availability of software for implementation and interpretability of the results in the context of AEs. The DGMs varied sample size ($n = 300$, $n = 500$), frailty distribution variance (0.25, 0.50, 0.75) and time-dependent effect (0.03, -0.03). The choice of the sample size was intended to reflect sample sizes that are commonly employed in most IPTp trials. Choice of frailty variance (i.e. a measure of the strength of the unobserved heterogeneity within an individual) intended to represent a setting 0.25, 0.5 and 0.75 represented weak variance, moderate variance and heavy variance respectively. The time-dependent treatment effect allowed the fixed effect treatment effect of log hazard ratio 0.5 to decay over time (when set at -0.03) or increase over time (when set at 0.03) or remain constant when set at 0.0). All the DGMs were parametrically based, to accommodate the investigation of several scenarios unlike resampling that confines investigation to data with properties only for the application data under study [25]. Hence, we simulated recurrent AE data similar to common scenarios that are practically and theoretically expected in IPTp trials. In all the simulation scenarios, the data generation mimicked a randomized controlled trial with two treatment arms. In all scenarios under study, we assumed that there were no substantial competing risks observed during the trial such that censoring was considered independent. The individual follow-up times were simulated based on uniform distribution and maximum follow-up time of 12 months (i.e. administrative censoring was set at 1 year), using *Survsim* package in stata. The choice of 12 months of follow-up time was motivated by the fact that most IPTp trials follow-up the mother for 12 months from enrolment to beyond delivery.

The hazard functions were simulated using inversion method, under Weibull baseline hazard function, accommodating both time-varying treatment effect and

unobserved heterogeneity with lognormal frailty distribution. The inversion method enabled sampling from a given distribution using uniform distribution and quantile function [26] where uniform distribution was assumed to have minimum 0 and maximum of 1. An overview of the investigated scenarios appears in Table 1 below.

During the data generation process, convergence of the models was also continuously monitored. After the data generation descriptive summary statistics were also done to check for any abnormal observations. For each scenario, graphical exploration of the distribution of the estimates was done to ascertain the assumed underlying distribution of the generated data.

Simulation study targets

Of primary interest in this study was shared frailty models' ability to effectively capture the unobserved heterogeneity. Therefore, the primary estimand of interest was log frailty variance ϕ_i for each of the investigated models. Since our second interest was to assess the ability of the models in capturing time-dependent treatment effect, the log hazard ratio β computed at 3 months of follow-up time was considered as a secondary estimand. We opted for log hazard ratio at the specific follow-up time in order to ensure that the estimates from both the proportional and non-proportional hazard models are comparable.

Methods for data analysis

Each simulated data set was analysed using the three methods under investigation: ISF-PH, LSF-NPHS1 and LSF-NPHS3. Based on our simulation scenarios, the ISF-PH model miss-specified the proportionality of the hazards although it correctly specified the baseline hazard function (i.e. the baseline hazard function for the ISF-PH was modelled using Weibull distribution). The LSF-NPHS1 and LSF-NPHS3 models correctly specified the proportionality of the hazards although the baseline hazard function is approximated by the restricted cubic splines. When there are non-proportional hazards, reporting of a single average HR is considered not practically useful. Therefore, in order to ensure that the performance measures for the HR estimates are comparable across the three models, we computed hazard ratio at 3 months of follow-up time since enrolment (i.e. LSF-NPHS1 and LSF-NPHS3 are non-proportional hazard models and ISF-PH is a proportional hazard model). The choice of the time at which the HR was calculated was motivated by our practical experience where most patients take approximately half of their scheduled IPTp doses by third month of follow-up time.

The three shared frailty models (ISF-PH, LSF-NPHS1 and LSF-NPHS3) are also used in the analysis

Table 1 Frailty variance estimates bias, coverage probability across shared frailty models in the presence of unobserved heterogeneity and non-proportional hazards

Data Generating Mechanisms				ISF-PH Model	LSF-NPHS1 Model	LSF-NPHS3 Model	% gain in precision relative to ISF-PH model	
Scenario ^a	Sample size	Frailty variance	TDE	Bias (MCSE), Coverage (MCSE), MSE (MCSE)	Bias (MCSE), Coverage (MCSE), MSE (MCSE)	Bias (MCSE), Coverage (MCSE), MSE (MCSE)	LSF-NPHS1% gain in precision (MCSE)	LSF-NPHS3% gain in precision (MCSE)
1	500	0.25	0.03	−0.0389 (0.0064), 96.40 (0.5891), 0.0423 (0.0021)	−0.0556 (0.0060), 95.90 (0.6270), 0.0386 (0.0020)	−0.0263 (0.0061), 96.20 (0.6046), 0.0374 (0.0019)	14.6985 (0.7269)	11.1644 (1.6768)
2	500	0.50	0.03	−0.0431 (0.0051), 95.00 (0.6892), 0.0278 (0.0012)	−0.1329 (0.0045), 87.10 (1.0600), 0.0375 (0.0015)	−0.1098 (0.0045), 90.20 (0.9402), 0.0323 (0.0014)	30.9883 (0.9472)	27.8901 (1.8680)
3	500	0.75	0.03	−0.0689 (0.0048), 92.80 (0.8174), 0.0280 (0.0013)	−0.2151 (0.0041), 61.80 (1.5365), 0.0628 (0.0020)	−0.1939 (0.0041), 69.10 (1.4612), 0.0544 (0.0019)	40.7088 (1.0645)	38.4618 (1.9530)
4	300	0.25	0.03	−0.0363 (0.0079), 97.20 (0.5217), 0.0637 (0.0029)	−0.0542 (0.0073), 97.50 (0.4937), 0.0569 (0.0027)	−0.0248 (0.0075), 96.70 (0.5649), 0.0561 (0.0027)	12.5407 (1.8701)	12.5407 (1.8701)
5	300	0.50	0.03	−0.0557 (0.0069), 94.90 (0.6957), 0.0504 (0.0025)	−0.1448 (0.0061), 90.60 (0.9228), 0.0575 (0.0027)	−0.1221 (0.0061), 92.50 (0.8329), 0.0518 (0.0025)	29.3473 (0.9645)	28.3377 (1.8182)
6	300	0.75	0.03	−0.0674 (0.0062), 94.50 (0.7209), 0.0431 (0.0020)	−0.2141 (0.0052), 76.60 (1.3388), 0.0729 (0.0026)	−0.1950 (0.0053), 80.90 (1.2431), 0.0660 (0.0025)	42.0668 (1.1767)	37.7659 (2.0100)
7	500	0.25	−0.03	−0.0212 (0.0063), 96.10 (0.6122), 0.0405 (0.0019)	−0.0694 (0.0060), 95.20 (0.6760), 0.0413 (0.0021)	−0.0348 (0.0061), 95.90 (0.6270), 0.0384 (0.0020)	9.7913 (0.6757)	7.8448 (1.5502)
8	500	0.50	−0.03	−0.0317 (0.0051), 95.30 (0.6693), 0.0268 (0.0012)	−0.1425 (0.0045), 85.60 (1.1102), 0.0405 (0.0016)	−0.1163 (0.0045), 89.00 (0.9894), 0.0341 (0.0014)	27.3902 (0.9049)	25.1537 (1.7685)
9	500	0.75	−0.03	−0.0605 (0.0048), 93.30 (0.7906), 0.0269 (0.0013)	−0.2236 (0.0041), 59.20 (1.5541), 0.0668 (0.0021)	−0.1998 (0.0041), 67.70 (1.4788), 0.0569 (0.0019)	38.1980 (1.0547)	37.1299 (1.8588)
10	300	0.25	−0.03	−0.0190 (0.0079), 97.00 (0.5394), 0.0621 (0.0027)	−0.0680 (0.0075), 97.70 (0.4740), 0.0602 (0.0028)	−0.0335 (0.0075), 97.00 (0.5394), 0.0577 (0.002)	11.1739 (0.7791)	9.1266 (1.7207)
11	300	0.50	−0.03	−0.0454 (0.0069), 94.80 (0.7021), 0.0493 (0.0024)	−0.1549 (0.0061), 89.50 (0.9694), 0.0614 (0.0029)	−0.1286 (0.0061), 92.50 (0.8329), 0.0537 (0.0026)	26.3505 (0.9108)	26.8512 (1.7308)
12	300	0.75	−0.03	−0.0588 (0.0062), 94.80 (0.7021), 0.0418 (0.0019)	−0.2219 (0.0053), 75.70 (1.3563), 0.0768 (0.0027)	−0.2005 (0.0053), 79.90 (1.2673), 0.0684 (0.0026)	39.0878 (1.1318)	36.0944 (1.9042)
13	500	0.25	0	−0.0296 (0.0064), 96.50 (0.5812), 0.0412 (0.0020)	−0.0619 (0.0060), 95.70 (0.6415), 0.0398 (0.0020)	−0.0300 (0.0061), 95.90 (0.6270), 0.0378 (0.0019)	12.1870 (0.6812)	9.4918 (1.6201)
14	500	0.50	0	−0.0373 (0.0051), 95.00 (0.6892), 0.0272 (0.0012)	−0.1373 (0.0045), 86.80 (1.0704), 0.0389 (0.0016)	−0.1127 (0.0045), 89.70 (0.9612), 0.0331 (0.0014)	29.0623 (0.8985)	26.5314 (1.8048)

Table 1 (continued)

Data Generating Mechanisms				ISF-PH Model	LSF-NPHS1 Model	LSF-NPHS3 Model	% gain in precision relative to ISF-PH model	
Scenario ^a	Sample size	Frailty variance	TDE	Bias (MCSE), Coverage (MCSE), MSE (MCSE)	Bias (MCSE), Coverage (MCSE), MSE (MCSE)	Bias (MCSE), Coverage (MCSE), MSE (MCSE)	LSF-NPHS1% gain in precision (MCSE)	LSF-NPHS3% gain in precision (MCSE)
15	500	0.75	0	−0.0644 (0.0048), 93.00 (0.8068), 0.0274 (0.0013)	−0.2188 (0.0041), 61.30 (1.5402), 0.0646 (0.0020)	−0.1964 (0.0041), 68.80 (1.4651), 0.0554 (0.0019)	39.4216 (1.0400)	37.9288 (1.8940)
16	300	0.25	0	−0.0273 (0.0079), 97.10 (0.5307), 0.0627 (0.0028)	−0.0605 (0.0074), 97.70 (0.4740), 0.0583 (0.0028)	−0.0283 (0.0075), 96.70 (0.5649), 0.0567 (0.0027)	13.2530 (0.7841)	10.8368 (1.7814)
17	300	0.50	0	−0.0500 (0.0069), 94.80 (0.7021), 0.0495 (0.0024)	−0.1491 (0.0061), 90.20 (0.9402), 0.0591 (0.0028)	−0.1249 (0.0061), 92.60 (0.8278), 0.0526 (0.0026)	27.7522 (0.9164)	27.0827 (1.7637)
18	300	0.75	0	−0.0627 (0.0062), 94.50 (0.7209), 0.0423 (0.0020)	−0.2174 (0.0052), 76.20 (1.3467), 0.0745 (0.0027)	−0.1972 (0.0053), 80.60 (1.250), 0.0668 (0.0026)	40.6654 (1.1293)	37.0383 (1.9578)

TDE time dependent effect

^a each scenario is based on 1000 simulations

of recurrent adverse events that were observed in the *Chloroquine for Malaria Prevention in Pregnancy* trial (NCT01443130) data. The motivating data description, data analysis and results appear in the application section of this paper, after simulation results presentation.

Performance measures

In this paper, our model performance measures of interest were defined as described in Morris et al [25]. Bias and MSE were main performance measures for our estimands. For frailty variance estimand, we assessed the model performance based on precision, bias, MSE and coverage probability of 95% confidence interval of the frailty variance estimates. For the hazard ratio estimand, we assessed the model performance based on precision, bias and MSE. Point estimates of performance are reported including Monte Carlo standard errors in order to highlight the certainty of our point estimates.

We used bias in order to quantify whether the estimator of interest targeted the true value of interest. Bias is defined as the difference between the expected value of an estimator and the true parameter value. Secondly, in order to integrate bias and variance of an estimator into one performance measure, we computed the mean square error (MSE). MSE is defined as sum of the squared bias and variance of the estimator. In defining relative precision gain, ISF-PH model was considered as a reference model. Therefore, precision gain is defined as the inverse squared ratio of the empirical SE of the respective flexible parametric model (i.e. LSF-NPHS1 or

LSF-NPHS3) to the empirical SE of the ISF-PH model. Additionally, coverage was also computed, defined as proportion of confidence intervals for an estimator that contain the true parameter value.

Software and computer hardware specification

The simulation study was conducted using user-written *survsim* and *merlin* packages in Stata version 15.1 [27–29]. The analyses of the simulated and application data were also done using *streg* built-in function in Stata for the ISF-PH model.

The simulation study was performed using the Stata version 15.1 that was installed in a purposely-built Threadripper 3990X GT 1030 workstation, equipped with a Threadripper 3990X (288MB Cache, 64x Cores, 4.3GHz) turbo processor, MSI TRX40 PRO 10G AMD Ryzen Threadripper Motherboard, Nvidia GeForce RTX 2070 SUPER OC Ed. 8GB GDDR6 GPU, Corsair Vengeance PRO RGB 3200MHz 64GB High-Performance Gaming, Hikvision E2000 1TB M.2 SSD that reads up to 3.5GB/s and a 4TB High-Performance Hard Disk Drive (HDD).

Ethical considerations

The *Chloroquine for malaria in pregnancy* trial was conducted in accordance with the Declaration of Helsinki and Good Clinical Practice guidelines. The trial obtained ethical approval from the Institutional Review Board at the University of Maryland, the College of Medicine Research and Ethics Committee at the University of

Malawi and the Malawi Pharmacy Medicines and Poisons Board. The current study was approved by the University of the Witwatersrand's Human Research Ethics Committee.

Results

Frailty variance estimates coverage probability, bias and MSE across the shared frailty models

Based on 1000 simulations for each scenario, we observed a 100% convergence rate of all the models that were fitted. Across all the scenarios, the magnitude of bias for the frailty variance estimates increased as the known unobserved heterogeneity variance increased from 0.25 to 0.75 whether the sample size was 300 or 500 (Table 1). The observed bias across all the scenarios was negative suggesting underestimation of the unobserved heterogeneity. The magnitude of bias for the frailty variance estimates were slightly higher when the sample size was 300 compared to the sample size of 500. Overall, the coverage probability of the frailty variance estimates consistently decreased as the known unobserved heterogeneity variance increased from 0.25 to 0.75. The coverage probability across all the models was slightly above 95% for scenarios where the unobserved heterogeneity variance was 0.25, suggesting some negligible over-coverage. For example, for scenario 1, coverage (MCSE) was 96.40 (0.5891), 95.90 (0.6270) and 96.20 (0.6046) for ISF-PH, LSF-NPHS1 and LSF-NPHS3 models respectively. Overall, substantial under-coverage was observed for scenarios where the unobserved heterogeneity variance was 0.50 or 0.75 among the flexible parametric models (LSF-NPHS1 and LSF-NPHS3). The under-coverage worsened as the sample size increased from 300 to 500. Such under-coverage suggested inaccuracy of the frailty variance estimates obtained using the LSF-NPHS1 and LSF-NPHS3 models.

As expected, overall, among the non-proportional hazard models with restricted cubic splines, the LSF-NPHS3 had lower bias and higher coverage of the frailty variance estimates compared to the LSF-NPHS1. This result suggests that for the non-proportional hazard models with restricted cubic splines, increasing the number of degrees of freedom modelling the baseline hazard function from 1 to 3 improves the frailty variance estimates. For the scenarios where the known unobserved heterogeneity was 0.50 or 0.75, we observed a higher magnitude of bias of frailty variance estimates for the non-proportional hazard models with restricted cubic splines models compared to the ISF-PH model. For the scenarios where the known unobserved heterogeneity variance was 0.25, we observed the lowest magnitude of bias for the frailty variance estimates in LSF-NPHS3 model followed by ISF-PH model then LSF-NPHS1 model. For example, under

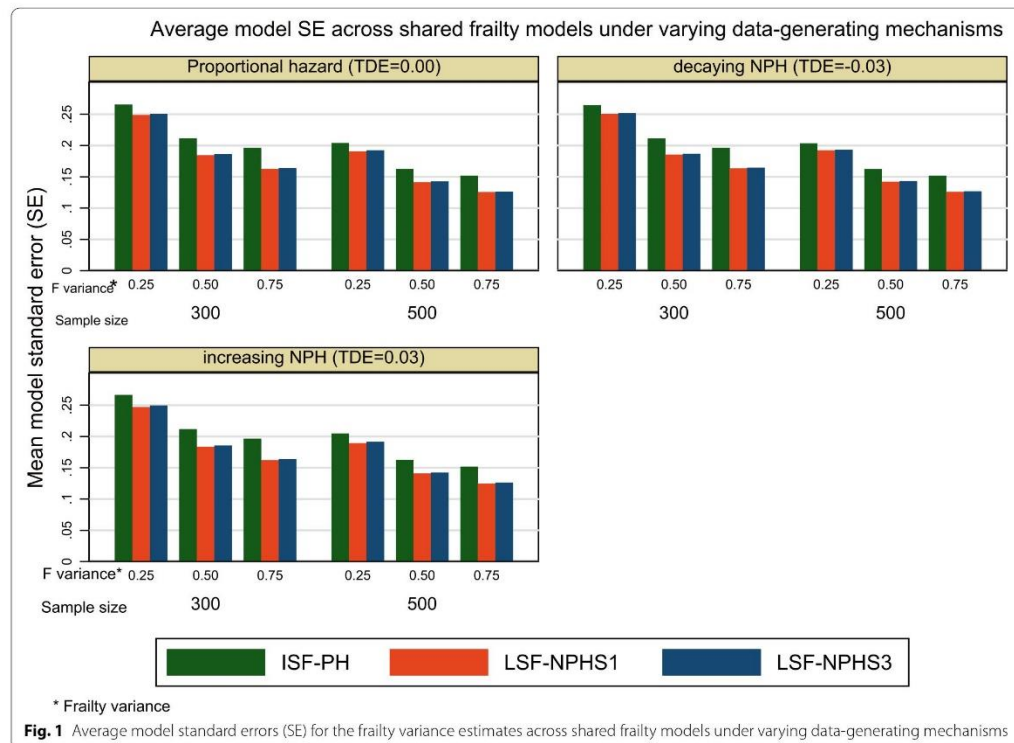
scenario 1, the frailty variance estimate bias (MCSE) was -0.0389 (0.0064), -0.0556 (0.0060), -0.0263 (0.0061) for ISF-PH, LSF-NPHS1 and LSF-NPHS3 models respectively. Pertaining to coverage, although the coverage probability decreased with increase in unobserved heterogeneity variance, decreasing the sample size from 500 to 300 yielded results with higher coverage. The decrease in coverage probability for the frailty variance estimates with increased unobserved heterogeneity was more drastic among the non-proportional hazards with restricted cubic splines models (LSF-NPHS1 or LSF-NPHS3) compared to the ISF-PH model. Overall, the ISF-PH model consistently yielded coverage probability closer to the nominal level of 95% compared to the non-proportional hazards with restricted cubic splines models (LSF-NPHS1 or LSF-NPHS3).

Overall, the MSE for the ISF-PH model frailty variance estimates were lower compared to the LSF-NPHS1 and LSF-NPHS3 models, except when the known unobserved heterogeneity variance was 0.25. Across all the 18 scenarios studied, we observed that the MSE estimates for the ISF-PH model decreased as the unobserved heterogeneity variance increased. However, the MSE estimates for the LSF-NPHS1 and LSF-NPHS3 models increased as the unobserved heterogeneity variance increased. Notably, the MSE for the LSF-NPHS3 model yielded lower MSE estimates compared to the LSF-NPHS1 model. As expected, increasing the sample size from 300 to 500 reduced the MSE estimates for the corresponding scenarios.

Our investigation showed that LSF-NPHS1 and LSF-NPHS3 models were more efficient than the ISF-PH model as demonstrated by the lower mean of model-based standard errors (Fig. 1). The difference in model-based standard errors between that LSF-NPHS1 and LSF-NPHS3 models was negligible. As expected, increasing sample size from 300 to 500 improved the efficiency of the shared frailty models under study such that the model-based standard errors for the frailty variance estimates decreased.

Percentage gain in precision for frailty variance estimates of non-proportional hazards shared frailty models with restricted cubic splines relative to ISF-PH model

Overall, as the unobserved heterogeneity increased, percentage gain in precision for frailty variance estimates of non-proportional hazards shared frailty models with restricted cubic splines relative ISF-PH model also increased (Table 1). The gain in precision was higher when time-dependent treatment effects were increasing (i.e. TDE = 0.03) compared to the decaying time-dependent treatment effects (TDE = -0.03). Across the models, the gain in precision was higher in the LSF-NPHS1



model compared to LSF-NPHS3 model. The observed gain in precision for LSF-NPHS1 and LSF-NPHS3 models compared to ISF-PH model under proportional hazard scenarios (i.e. $TDE=0$) was similar to the scenarios when time-dependent treatment effects were increasing (i.e. $TDE=0.03$).

Bias and MSE for the estimated log hazard ratio estimates across the shared frailty models under varying unobserved heterogeneity and sample size

Overall, across all the three models under study, the MSE for the HR estimates increased as the unobserved heterogeneity variance increased (Table 2). The ISF-PH model consistently yielded lowest MSE estimates for the HRs across the three models regardless of the nature of the time-dependent treatment effect. Increasing the sample size from 300 to 500 also led to reduction in the estimated MSE for the HRs. We observed a loss in precision of HR estimates for the ISF-PH model compared to the LSF-NPHS1 and LSF-NPHS3 model. The precision loss ranged from 9 to 20%. The highest precision loss was observed when the unobserved heterogeneity was low

(i.e. frailty variance equal to 0.25). The precision loss estimates were similar for both LSF-NPHS1 and LSF-NPHS3 models. There was low precision loss in the absence of time-dependent treatment effects ($TDE=0$). In the presence of time-dependent treatment effects (i.e. $TDE=0.03$ and $TDE=-0.03$), we observed higher magnitude of bias in the ISF-PH model compared to the LSF-NPHS1 and LSF-NPHS3 models. In the absence of the time-dependent treatment effects, the bias for the HR estimates was lower in the ISF-PH model compared to the LSF-NPHS1 and LSF-NPHS3 models. However, in the absence of the time-dependent treatment effects, the bias across all the three models was relatively lower (i.e. under 0.01) compared to scenarios when time-dependent effects are present.

Example data application: Chloroquine for malaria in pregnancy trial

We illustrated the application of the three investigated shared frailty models above (ISF-PH, LSF-NPHS1 and LSF-NPHS3) to data on recurrent AEs collected from a clinical trial evaluating antimalarial drugs to prevent

Table 2 Bias and mean square error (MSE) for log hazard ratio estimates, estimated at 3 months, across shared frailty models in the presence of both unobserved heterogeneity and non-proportional hazards

Data Generating Mechanisms				ISF-PH Model	LSF-NPHS1 Model	LSF-NPHS3 Model	% gain in precision relative to ISF-PH model	
Scenario ^a	Sample size	Frailty variance	TDE	Bias (MCSE), MSE (MCSE)	Bias (MCSE), MSE (MCSE)	Bias (MCSE), MSE (MCSE)	LSF-NPHS1% gain in precision (MCSE)	LSF-NPHS3% gain in precision (MCSE)
1	500	0.25	0.03	−0.0284 (0.0021), 0.0050 (0.0002)	−0.0033 (0.0023), 0.0051 (0.0002)	0.0021 (0.0023), 0.0052 (0.0002)	−17.4231 (2.1208)	−18.7145 (2.0875)
2	500	0.50	0.03	−0.0285 (0.0024), 0.0068 (0.0003)	−0.0023 (0.0026), 0.0067 (0.0003)	0.0031 (0.0026), 0.0068 (0.0003)	−11.1578 (1.9985)	−12.3537 (1.9639)
3	500	0.75	0.03	−0.0233 (0.0027), 0.0076 (0.0004)	0.0035 (0.0029), 0.0081 (0.0004)	0.0093 (0.0029), 0.0083 (0.0004)	−12.7599 (1.9452)	−14.1747 (1.8756)
4	300	0.25	0.03	−0.0259 (0.0028), 0.0085 (0.0004)	−0.0015 (0.0030), 0.0093 (0.0004)	0.0041 (0.0031), 0.0096 (0.0004)	−15.4391 (2.0856)	−18.0137 (2.0061)
5	300	0.50	0.03	−0.0215 (0.0032), 0.0105 (0.0005)	0.0047 (0.0033), 0.0111 (0.0005)	0.0103 (0.0034), 0.0114 (0.0005)	−9.4606 (2.1449)	−11.4682 (2.0577)
6	300	0.75	0.03	−0.0252 (0.0035), 0.0126 (0.0006)	0.0019 (0.0037), 0.0134 (0.0006)	0.0069 (0.0037), 0.0137 (0.0006)	−10.3721 (1.9615)	−12.1647 (1.8949)
7	500	0.25	−0.03	0.0287 (0.0021), 0.0051 (0.0002)	−0.0075 (0.0023), 0.0053 (0.0002)	0.0018 (0.0023), 0.0053 (0.0002)	−18.3401 (2.1438)	−19.5151 (2.1016)
8	500	0.50	−0.03	0.0289 (0.0024), 0.0068 (0.0003)	−0.0072 (0.0026), 0.0069 (0.0003)	0.0021 (0.0026), 0.0069 (0.0003)	−12.1807 (2.0392)	−13.6435 (1.9931)
9	500	0.75	−0.03	0.0341 (0.0027), 0.0083 (0.0004)	−0.0011 (0.0029), 0.0083 (0.0004)	0.0086 (0.0029), 0.0085 (0.0004)	−14.1771 (1.9844)	−15.7796 (1.9005)
10	300	0.25	−0.03	0.0311 (0.0028), 0.0088 (0.0004)	−0.0059 (0.0031), 0.0095 (0.0004)	0.0035 (0.0031), 0.0098 (0.0004)	−16.7676 (2.1182)	−19.2818 (2.0272)
11	300	0.50	−0.03	0.0361 (0.0032), 0.0114 (0.0005)	0.0004 (0.0033), 0.0112 (0.0005)	0.0096 (0.0034), 0.0116 (0.0005)	−10.0998 (2.1977)	−12.4175 (2.0889)
12	300	0.75	−0.03	0.0316 (0.0035), 0.0130 (0.0006)	−0.0032 (0.0037), 0.0136 (0.0006)	0.0052 (0.0037), 0.0139 (0.0007)	−11.6808 (2.0013)	−13.7017 (1.9235)
13	500	0.25	0	−0.0009 (0.0021), 0.0042 (0.0002)	−0.0051 (0.0023), 0.0052 (0.0002)	0.0020 (0.0023), 0.0053 (0.0002)	−17.8005 (2.1303)	−19.0529 (2.0932)
14	500	0.50	0	−0.0009 (0.0024), 0.0059 (0.0003)	−0.0044 (0.0026), 0.0067 (0.0003)	0.0027 (0.0026), 0.0068 (0.0003)	−11.4262 (2.0238)	−12.6627 (1.9860)
15	500	0.75	0	0.0044 (0.0027), 0.0071 (0.0003)	0.0016 (0.0029), 0.0082 (0.0004)	0.0091 (0.0029), 0.0084 (0.0004)	−13.4164 (1.9621)	−14.9156 (1.8853)
16	300	0.25	0	0.0017 (0.0028), 0.0079 (0.0004)	−0.0032 (0.0031), 0.0094 (0.0004)	0.0041 (0.0031), 0.0097 (0.0004)	−15.8830 (2.0985)	−18.3944 (2.0120)
17	300	0.50	0	0.0061 (0.0032), 0.0101 (0.0005)	0.0027 (0.0033), 0.0112 (0.0005)	0.0097 (0.0034), 0.0115 (0.0005)	−9.7878 (2.1647)	−11.9642 (2.0672)
18	300	0.75	0	0.0021 (0.0035), 0.0120 (0.0006)	−0.0004 (0.0037), 0.0134 (0.0006)	0.0062 (0.0037), 0.0138 (0.0006)	−11.0009 (1.9807)	−12.9017 (1.9108)

malaria in pregnancy. The trial is called *Chloroquine for malaria in pregnancy*.

The current modelling focussed on investigating the effect of treatment on recurrent AEs. We included a population of 600 pregnant women in their first or second trimesters who were randomized to receive IPTp of (a) sulfadoxine-pyrimethamine (SP, two doses, 4 weeks apart), (b) chloroquine (CQ, 600mg on day 1, 600mg on day 2, and 300mg on day 3 two days apart). Although the original trial had three treatment arms, in the current analysis, we dropped the arm of CQ prophylaxis (to focus on IPTp arms) since its treatment schedule was radically

different from the IPTp arms. Full details of the trial are published elsewhere [30].

Of the 600 pregnant women analysed in the current analysis, 474 (79%) experienced at least one AE. Cumulatively, more AEs were experienced in the CQ arm (895 AEs) than those in the SP arm (650 AEs). We fitted ISF-PH model followed by the LSF-NPHS1 and LSF-NPHS3 models where IPTp treatment was an exposure of interest. The SP treatment was considered as a reference treatment in the interpretation of the results. ISF-PH assumed the baseline hazard function followed Weibull distribution, unobserved heterogeneity followed inverse

Gaussian distribution and the hazard rates between the CQ and SP treatments were proportional over the follow-up time. The non-proportional hazard models with restricted models (NPHS1 and LSF-NPHS3) assumed that unobserved heterogeneity followed lognormal distribution and time-dependent effects could be modelled using one degree of freedom. NPHS1 and LSF-NPHS3 approximated the baseline hazard function using 1 and 2 degrees of freedom respectively. We fitted adjusted and unadjusted model for all the three model under investigation. The adjusted model included the following covariates: treatment, maternal age, maternal body mass index and gravidity (i.e. whether the pregnant woman had previous pregnancy or not). We adjusted for the same covariates for the three models to ensure that the estimates were comparable. Since the results from the adjusted and unadjusted models were similar, our results reported in Table 3 below are based the unadjusted models since we considered them more parsimonious. The reported hazard ratios across the three models are based on marginal hazard ratio, predicted hazard ratio at 14 days and 60 days of follow-up. Additionally, we plotted the hazard ratio estimates over the follow-up time.

In this clinical example for recurrent AEs data analysis, the estimated marginal log hazard ratios were 0.2281 (0.0499) 0.2328 (SE: 0.0517), 0.3090 (0.0769), 0.3360 (SE: 0.0850) and 0.3024 (0.0746) for ISF-PH, LSF-NPHS1, LSF-NPHS2, LSF-NPHS3 and LSF-NPHS4 models respectively, indicating an increased risk of AEs among the women who took the CQ treatment arm compared to those in the SP treatment arm (Table 3). At 14 days, the Log HR across the non-proportional hazard models were similar ranging between 0.43 to 0.47 and were higher compared to the ISF-PH model. However, at 60 days only LSF-NPHS2, LSF-NPHS3 and yielded similar log HR estimates approximately 0.20 and were lower compared to the ISF-PH model estimates. The frailty variance estimates generally higher among the non-proportional hazard models compared to the ISF-PH model. Among the non-proportional hazards models with restricted cubic splines, the AIC estimates show that the model fit was

good when the baseline hazard function was approximated with minimum of 2 degrees.

Discussion

The analysis of recurrent AEs in clinical trials rarely account for potential unobserved heterogeneity and time-dependent effects [2]. Using a simulation study, we investigated the performance of three statistical methods for recurrent events that can be adapted towards the analysis of recurrent AEs in clinical trials, in the presence of both time-dependent effects and unobserved heterogeneity. Our paper mainly focused on evaluating the ability of the flexible parametric shared frailty models with non-proportional hazards and restricted cubic splines to capture both the unobserved heterogeneity and time-dependent treatment effects. We established that the inverse Gaussian shared frailty with proportional hazard model consistently yielded good coverage closer to the nominal level of 95%, lower bias and lower MSE of the frailty variance estimates at the expense of imprecision compared to the flexible shared frailty models with non-proportional hazards and restricted cubic splines. Interestingly, the frailty variance estimates coverage and bias, for the shared frailty models with non-proportional hazards and restricted cubic splines, improved upon increasing the number of degrees of freedom used to model the baseline hazard function. As expected, the flexible parametric shared frailty model with non-proportional hazards and restricted cubic splines yielded better hazard ratio estimates, reflecting the time-dependent treatment effects, compared to the inverse Gaussian shared frailty with proportional hazard model. The flexible parametric shared frailty models with non-proportional hazards and restricted cubic splines yielded unbiased estimates of HRs at the expense of imprecision and inaccuracy (i.e. high MSE) compared to the ISF-PH model. Therefore, our findings show that the more precise estimators were biased and imprecise estimators were unbiased. In drug safety assessment using methods that yield an unbiased hazard ratio would be of primary interest such that the flexible parametric shared frailty model should be

Table 3 Comparing sharing frailty models for analysis of recurrent AEs among pregnant women on IPTp treatment in Malawi

Estimate	ISF-PH	LSF-NPHS1	LSF-NPHS2 ^a	LSF-NPHS3	LSF-NPHS4 ^b
Log HR at 14 days (SE)	0.2281 (0.0499)	0.4758 (0.3572)	0.4343 (0.3724)	0.4648 (0.4033)	0.4404 (0.3710)
log HR at 60 days (SE)	0.2281 (0.0499)	0.0236 (0.1240)	0.2011 (0.2562)	0.2251 (0.2831)	0.1837 (0.2441)
Marginal log HR (SE)	0.2281 (0.0499)	0.2328 (0.0517)	0.3090 (0.0769)	0.3360 (0.0850)	0.3024 (0.0746)
Frailty variance	0.0324	0.0476	0.4591	0.6440	0.3977
AIC	6562.57	19,922.40	19,658.30	19,651.20	19,570.70

^a Lognormal shared frailty model with non-proportional hazard and restricted cubic splines where the baseline hazard is modelled with 2 degrees of freedom and the time-dependent treatment effects is modelled with 1 degree of freedom

^b Lognormal shared frailty model with non-proportional hazard and restricted cubic splines where the baseline hazard is modelled with 4 degrees of freedom and the time-dependent treatment effects is modelled with 1 degree of freedom

considered if there is the presence of both unobserved heterogeneity and time-dependent effects. However, such a decision should be made at the cost of imprecision and loss of accuracy. Hence, we emphasize that the flexible shared frailty with non-proportional hazards and restricted cubic splines should be applied only when the proportional hazards assumption is violated. If interest is in estimating frailty variance, the inverse Gaussian model with proportional hazards can be considered sufficient even in the presence of both time-dependent treatment effect and unobserved heterogeneity.

The inverse Gaussian shared frailty with proportional hazard probably effectively captured the unobserved heterogeneity because we did not mis-specify the baseline hazard (i.e. Weibull distribution). Secondly, the longer follow-up period of 12 months made the population more homogeneous with time since the unobserved heterogeneity for inverse Gaussian frailty decays over the follow-up time [31]. Thirdly, the inverse Gaussian frailty is practically close to the lognormal frailty where the simulated data arose. The structure of the inverse Gaussian frailty would be considered more appropriate in the context of recurrent AEs in antimalarial drug trials where there are more frequent AEs early in the clinical trial [18]. Unfortunately, the conditional hazard ratio derived from the inverse Gaussian shared frailty with proportional hazard is constant over the follow-up time which does not reflect the non-proportionality of the hazards in the presence of time-dependent treatment effect.

The interpretation of our results is confined to scenarios where the censoring is assumed to be non-informative. In the presence of informative censoring, it will be interesting to understand how the models perform. Failure to account for time to medication presents a limitation in all the models considered in this work. Another key limitation of our study is relying on the assumption of constant unobserved heterogeneity. Future work can consider evaluating the performance of the shared frailty models accounting for both actual time to medication and time to AE occurrence under complicated scenarios where there is both time-varying treatment effects and time-varying unobserved heterogeneity.

Generally, in drug clinical trials, understanding drug safety profile requires efficient modelling of the recurrent AEs including accounting for both unobserved heterogeneity and any non-linear effects of the investigated drug. The methods studied in this paper provide key knowledge on choosing the most appropriate model for analysing recurrent AE data presenting the unobserved heterogeneity and time-dependent effects issues. The flexible parametric shared methods accounting for both

time-dependent effects and unobserved heterogeneity are directly useful in IPTp trials since physiological changes that occur during pregnancy are mostly unobserved and can contribute to the heterogeneity of the treatment effect.

Based on our findings, we provide some practical recommendations on the studied shared frailty models to analyse the recurrent AEs when the hazards are non-proportional and the potential unobserved heterogeneity. Overall, the choice of the most appropriate model should be driven by the objective of the analysis. For instance, if interest is mainly in estimating hazard ratios, a shared frailty model with non-proportional hazards and restricted cubic splines is appropriate. When interest is in obtaining the unobserved heterogeneity estimates, inverse Gaussian with proportional hazard can serve the purpose. In the absence of time-dependent effects, using a shared frailty model with non-proportional hazards and restricted cubic splines should be avoided since it leads to biased estimates due to model overfitting.

Abbreviations

IPTp: Intermittent preventive treatment for malaria prevention during pregnancy; AEs: Adverse events; RCTs: Randomized controlled trials; ISF-PH: Gaussian shared frailty model with proportional hazard; LSF-NPH1: Lognormal Shared frailty model with non-proportional hazards and single restricted cubic spline; LSF-NPH3: Lognormal Shared frailty model with non-proportional hazards and three restricted cubic splines; DGM: Data generating mechanism; TDE: Time-dependent effects; MCSE: Monte-Carlo Standard error; CQ: Chloroquine; SP: Sulfadoxine-pyrimethamine.

Acknowledgements

We thank all the study participants and data collectors in the clinical trial.

Authors' contributions

Study idea conceptualization: NP, TC, MM, MJCE. Methodology development: NP, MM, LK, TC, MKL, DM, MJCE. Analysis and first draft development: NP. Supervision, writing revised drafts-reviewing and editing: MM, MKL, LK, TC, DM, MJCE. Enrolled patients and conducted clinical investigation: MKL. Reading and approving the final manuscript: NP, MM, MKL, LK, TC, DM, MJCE. The author(s) read and approved the final manuscript.

Funding

The clinical trial was supported by funding from NIH (grants U01AI087624 and K24AI114996 to MKL). NP was supported by the Fogarty International Center of the National Institutes of Health [Grant number D43TW010075 to MKL and DM]. NP was also supported by South African Research Chairs Initiative (SARChI) of the Department of Science and Technology and National Research Foundation (grant 64788) awarded to Dr. Ikechukwu Achilonu (SARChI Chair in Protein Biochemistry and Structural Biology), School of Molecular and Cell Biology, University of the Witwatersrand, Johannesburg, South Africa.

Availability of data and materials

The code for simulated data generated is available from the corresponding author upon reasonable request. The *Chloroquine for Malaria in Pregnancy* (NCT01443130) trial data that support the findings of this study are available from Blantyre Malaria Project, Malawi but restrictions apply to the availability of these data, which were used under license for the current study, and so are not publicly available. Data are however available from the corresponding author upon reasonable request and with permission of Blantyre Malaria Project, Blantyre, Malawi.

Declarations

Ethics approval and consent to participate

The current study has been performed in accordance with the relevant regulations and guidelines. The current study is based on computer simulations where no human subjects are used. The current study was approved by the University of the Witwatersrand's Human Research Ethics Committee. The Chloroquine for Malaria in Pregnancy trial obtained ethical approval from the Institutional Review Board at the University of Maryland, the College of Medicine Research and Ethics Committee at the University of Malawi and the Malawi Pharmacy Medicines and Poisons Board. Informed consent was obtained from all the participants.

Consent for publication

Not applicable.

Competing interests

The authors declare that they have no competing interests.

Author details

¹School of Public Health, University of the Witwatersrand, Johannesburg, South Africa. ²School of Global and Public Health, Kamuzu University of Health Sciences, Blantyre, Malawi. ³Mahidol Oxford Tropical Medicine Research unit (MORU), Bangkok, Thailand. ⁴Centre for Tropical Medicine, Nuffield Department of Medicine, University of Oxford, Oxford, UK. ⁵Department of Biostatistics, University of Namibia, Windhoek, Namibia. ⁶Julius Center for Health Sciences and Primary Care, University Medical Center Utrecht, Utrecht University, Utrecht, The Netherlands. ⁷Center for Vaccine Development and Global Health, University of Maryland School of Medicine, Baltimore, MD, USA.

Received: 1 May 2021 Accepted: 19 November 2021

Published online: 20 January 2022

References

- Munsaka MS. A question-based approach to the analysis of safety data. In: Biopharmaceutical Applied Statistics Symposium. Biostatistical analysis of clinical trials, Vol. 2. Singapore: Springer; 2018. p. 193–216.
- Patson N, Mukama M, Otumbe KN, Kazembe L, Mathanga DP, Mwapasa V, et al. Systematic review of statistical methods for safety data in malaria chemoprevention in pregnancy trials. *Malar J*. 2020;19(1):119.
- Phillips R, Cornelius V. Understanding current practice, identifying barriers and exploring priorities for adverse event analysis in randomised controlled trials: an online, cross-sectional survey of statisticians from academia and industry. *BMJ Open*. 2020;10(6):e036875.
- Colopy MW, Gordon R, Ahmad F, Wang WW, Duke SP, Ball G. Statistical practices of safety monitoring: an industry survey. *Ther Innov Reg Sci*. 2019;53(3):293–300.
- Siddiqui O. Statistical methods to analyze adverse events data of randomized clinical trials. *J Biopharm Stat*. 2009;19(5):889–99.
- Hengelbrock J, Gillhaus J, Kloss S, Leverkus F. Safety data from randomized controlled trials: applying models for recurrent events. *Pharm Stat*. 2016;15(4):315–23.
- Cabarrou B, Gomez-Roca C, Viala M, Rabeau A, Paulon R, Loirat D, et al. Modernizing adverse events analysis in oncology clinical trials using alternative approaches: rationale and design of the MOTIVATE trial. *Investig New Drugs*. 2020;38:1879–87.
- Box-Steffensmeier JM, De Boef S. Repeated events survival models: the conditional frailty model. *Stat Med*. 2006;25(20):3518–33.
- Hougaard P. Analysis of multivariate survival data. New York: Springer; 2012.
- Duchateau L, Janssen P. The frailty model. New York: Springer; 2007.
- Amorim LD, Cai J. Modelling recurrent events: a tutorial for analysis in epidemiology. *Int J Epidemiol*. 2015;44(1):324–33.
- Balan TA, Putter H. Nonproportional hazards and unobserved heterogeneity in clustered survival data: when can we tell the difference? *Stat Med*. 2019;38(18):3405–20.
- Crowther MJ. Extended multivariate generalised linear and non-linear mixed effects models. 2017. <https://arxiv.org/abs/1710.02223>. Accessed 5 Dec 2020.
- Bower H, Crowther MJ, Rutherford MJ, Andersson TM-L, Clements M, Liu X-R, et al. Capturing simple and complex time-dependent effects using flexible parametric survival models: a simulation study. *Commun Stat Simul Comput*. 2021;50(11):3777–93.
- Rutherford MJ, Crowther MJ, Lambert PC. The use of restricted cubic splines to approximate complex hazard functions in the analysis of time-to-event data: a simulation study. *J Stat Comput Simul*. 2015;85(4):777–93.
- Gasparini A, Clements MS, Abrams KR, Crowther MJ. Impact of model misspecification in shared frailty survival models. *Stat Med*. 2019;38(23):4477–502.
- Cook RJ, Lawless J. The statistical analysis of recurrent events. New York: Springer; 2007.
- Hanagal DD. Modeling survival data using frailty models. Singapore: Springer; 2019.
- Andersen PK, Gill RD. Cox's regression model for counting processes: a large sample study. *Ann Stat*. 1982;10(4):1100–20.
- Kalbfleisch JD, Prentice RL. The statistical analysis of failure time data. New York: Wiley; 2011.
- Lawless JF. Statistical models and methods for lifetime data. New York: Wiley; 2003.
- Kelly PJ, Lim LLY. Survival analysis for recurrent event data: an application to childhood infectious diseases. *Stat Med*. 2000;19(1):13–33.
- Babiker A, Cuzick J. A simple frailty model for family studies with covariates. *Stat Med*. 1994;13(16):1679–92.
- Pickles A, Crouchley R. A comparison of frailty models for multivariate survival data. *Stat Med*. 1995;14(13):1447–61.
- Morris TP, White IR, Crowther MJ. Using simulation studies to evaluate statistical methods. *Stat Med*. 2019;38(11):2074–102.
- Bender R, Augustin T, Blettner M. Generating survival times to simulate cox proportional hazards models. *Stat Med*. 2005;24(11):1713–23.
- Crowther MJ. Merlin-a unified modelling framework for data analysis and methods development in Stata. 2018. <https://arxiv.org/abs/1806.01615>. Accessed 30 Nov 2020.
- Crowther MJ, Lambert PC. Simulating biologically plausible complex survival data. *Stat Med*. 2013;32(23):4118–34.
- Crowther MJ, Look MP, Riley RD. Multilevel mixed effects parametric survival models using adaptive Gauss-Hermite quadrature with application to recurrent events and individual participant data meta-analysis. *Stat Med*. 2014;33(22):3844–58.
- Divala TH, Mungwira RG, Mawindo PM, Nyirenda OM, Kanjala M, Ndafer-anthande M, et al. Chloroquine as weekly chemoprophylaxis or intermittent treatment to prevent malaria in pregnancy in Malawi: a randomised controlled trial. *Lancet Infect Dis*. 2018;18(10):1097–107.
- Hougaard P. Life table methods for heterogeneous populations: distributions describing the heterogeneity. *Biometrika*. 1984;71(1):75–83.

Publisher's Note

Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Ready to submit your research? Choose BMC and benefit from:

- fast, convenient online submission
- thorough peer review by experienced researchers in your field
- rapid publication on acceptance
- support for research data, including large and complex data types
- gold Open Access which fosters wider collaboration and increased citations
- maximum visibility for your research: over 100M website views per year

At BMC, research is always in progress.

Learn more biomedcentral.com/submissions



RESEARCH

Open Access



Joint modelling of multivariate longitudinal clinical laboratory safety outcomes, concomitant medication and clinical adverse events: application to artemisinin-based treatment during pregnancy clinical trial

Noel Patson^{1,2*}, Mavuto Mukaka^{3,4}, Umberto D'Alessandro⁵, Gertrude Chapotera², Victor Mwapasa², Don Mathanga², Lawrence Kazembe⁶, Miriam K. Laufer⁷ and Tobias Chirwa¹

Abstract

Background: In drug trials, clinical adverse events (AEs), concomitant medication and laboratory safety outcomes are repeatedly collected to support drug safety evidence. Despite the potential correlation of these outcomes, they are typically analysed separately, potentially leading to misinformation and inefficient estimates due to partial assessment of safety data. Using joint modelling, we investigated whether clinical AEs vary by treatment and how laboratory outcomes (alanine amino-transferase, total bilirubin) and concomitant medication are associated with clinical AEs over time following artemisinin-based antimalarial therapy.

Methods: We used data from a trial of artemisinin-based treatments for malaria during pregnancy that randomized 870 women to receive artemether–lumefantrine (AL), amodiaquine–artesunate (ASAQ) and dihydroartemisinin–piperaquine (DHAPQ). We fitted a joint model containing four sub-models from four outcomes: longitudinal sub-model for alanine aminotransferase, longitudinal sub-model for total bilirubin, Poisson sub-model for concomitant medication and Poisson sub-model for clinical AEs. Since the clinical AEs was our primary outcome, the longitudinal sub-models and concomitant medication sub-model were linked to the clinical AEs sub-model via current value and random effects association structures respectively. We fitted a conventional Poisson model for clinical AEs to assess if the effect of treatment on clinical AEs (i.e. incidence rate ratio (IRR)) estimates differed between the conventional Poisson and the joint models, where AL was reference treatment.

Results: Out of the 870 women, 564 (65%) experienced at least one AE. Using joint model, AEs were associated with the concomitant medication (log IRR 1.7487; 95% CI: 1.5471, 1.9503; $p < 0.001$) but not the total bilirubin (log IRR: -0.0288; 95% CI: -0.5045, 0.4469; $p = 0.906$) and alanine aminotransferase (log IRR: 0.1153; 95% CI: -0.0889, 0.3194; $p = 0.269$). The Poisson model underestimated the effects of treatment on AE incidence such that log IRR for ASAQ was 0.2118 (95% CI: 0.0082, 0.4154; $p = 0.041$) for joint model compared to 0.1838 (95% CI: 0.0574, 0.3102; $p = 0.004$) for Poisson model.

*Correspondence: noelpatson@gmail.com

² School of Public Health and Family Medicine, College of Medicine, University of Malawi, Blantyre, Malawi

Full list of author information is available at the end of the article



© The Author(s) 2021. **Open Access** This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>. The Creative Commons Public Domain Dedication waiver (<http://creativecommons.org/publicdomain/zero/1.0/>) applies to the data made available in this article, unless otherwise stated in a credit line to the data.

Conclusion: We demonstrated that although the AEs did not vary across the treatments, the joint model yielded efficient AE incidence estimates compared to the Poisson model. The joint model showed a positive relationship between the AEs and concomitant medication but not with laboratory outcomes.

Trial registration: ClinicalTrials.gov: [NCT00852423](https://clinicaltrials.gov/ct2/show/study/NCT00852423)

Keywords: Adverse events, Randomised controlled trials, Joint model, Drug safety, Concomitant medication

Background

There has been limited analysis of safety data being conducted in drug clinical trials compared to efficacy data. During drug clinical trials, scheduled and non-scheduled visits are conducted where multiple safety outcomes are collected including longitudinal clinical laboratory outcomes and clinical AEs [1, 2]. Data on concomitant medications defined as any medication/supplement taken by the trial participant other than the investigational product, are also repeatedly collected in order to support drug safety evidence. The repeatedly measured clinical laboratory outcomes are considered as objective and proxy indicators for drug-induced human body organ dysfunction and tend to be strongly associated with incidence of clinical AEs. For example, classic liver function biomarkers such as total bilirubin and alanine aminotransferase are usually used in assessment of drug safety [3]. However, each of the clinical laboratory safety outcomes is usually analysed separately based on summary statistics such as means or proportion of patients with elevated values for the safety markers [4, 5]. The clinical AE data is also usually analysed separately using the descriptive non-parametric methods (e.g. cumulative incidence function and crude incidence rate) [6–8]. Although ignored in most safety analyses, concomitant medication taken during the trial period can also influence the magnitude of drug safety estimates since the concomitant medication can interact with the treatment under investigation. Concomitant medication is usually also analysed separately (e.g. using mean cumulative function) [9]. The standard approaches that analysed each of the clinical laboratory safety data, clinical AEs and concomitant medication separately are inefficient and may yield biased estimates of AE incidence rate (i.e. clinical AEs occurrence over follow-up time) since they ignore the correlation structure of the multiple outcomes over the follow-up time, leading to information loss. Furthermore, the separate analyses present challenges in interpreting overall drug safety profile. Integrated approach in analysis of the clinical laboratory safety data, concomitant medication and clinical AEs through the use of joint models, can offer an opportunity to efficiently harness the available information towards enriching the drug safety estimates. The joint models can efficiently quantify how the incidence rate of AEs is associated with the multivariate longitudinal process of

the multiple continuous laboratory-based measurements, accounting for baseline covariates, concomitant medications and unobserved heterogeneity that is not captured in the conventional analytical approaches of safety data, all of which can potentially confound the estimates.

Recently, there have been increased proposals for improved analysis of multivariate clinical laboratory data, concomitant medications and clinical AEs. Schil-drout et al. [10] and Rosenkranz [11] discuss the utility of longitudinal modelling of clinical laboratory data using estimating equations and maximum likelihood. However, such separate longitudinal models ignore correlation between the multiple safety outcomes. Southworth and Heffernan proposed joint modelling approach that captures the joint behaviour of clinical laboratory safety data elevated values indicative of potential liver damage [12]. However, this was limited as the model did not use all the available information due to focus on extreme values and, further, such modelling is mainly suited for large sample sizes. Merz et al. [3] also recommended accounting for both multivariate longitudinal clinical laboratory data and clinical AEs, although they do not explicitly discuss how this can be practically achieved. Barker [9] demonstrates how mean cumulative function can be used as an exploratory tool in analysis number of concomitant medications taken over the follow-up time in order to support drug safety evidence interpretation. Building on these recent developments, this paper explores whether jointly modelling multivariate longitudinal clinical laboratory, concomitant medication and clinical AEs adjusted for baseline characteristics can be efficient in estimating AE incidence rate and useful in understanding joint evolution of the potentially correlated multiple safety outcomes.

In clinical trials with frequent AE occurrence (e.g. antimalarial treatment in pregnancy trials), event count models such as Poisson or negative binomial regression are considered important in quantifying the AE rate [13–16]. However, the AEs are also highly correlated with the cumulative concomitant medications and clinical laboratory safety outcomes. Joint modelling can offer an opportunity to explicitly model the correlations and account for the follow-up time, ensuring efficient and unbiased estimates. Despite the advanced development and evident benefits of joint modelling of multivariate

longitudinal data [17–19], joint modelling has received limited attention in drug safety assessment. Recent statistical software developments have provided an opportunity for extending joint models to more different types of outcomes [20, 21], beyond the traditionally implemented joint models with single continuous longitudinal data and single survival outcome. In this paper, we present a joint model for longitudinal continuous clinical laboratory safety data (alanine amino transferase and total bilirubin), individual clinical AEs and concomitant medication. We focus on demonstrating the utility of the proposed model in assessing the association between the clinical AEs in relation to these outcomes controlling for baseline covariates. Furthermore, we assessed whether the joint model improved the estimates in log incidence rate ratio (IRR) on effect of treatment on clinical AEs compared to the conventional Poisson model. The proposed joint model is useful in providing evidence-based decisions in determining the most important clinical safety measurements to focus on for drug safety assessment especially in resource-constrained settings.

Methods

Data

Our proposed joint model was applied to data from PREGACT trial as the motivating data. PREGACT trial (ClinicalTrials.gov number, NCT00852423) was a multicentre, open-label randomized trial carried out in 4 African countries (Burkina Faso, Ghana, Malawi, and Zambia). The current work uses the data from Malawi. In Malawi, the trial was implemented between June 2010 and August 2013 and enrolled 870 pregnant women during their second or third trimester with falciparum malaria. The women were randomly allocated in a 1:1:1 ratio to be treated with artemether–lumefantrine (AL), amodiaquine–artesunate (ASAQ) or dihydroartemisinin–dipeperazine (DHAPQ). The primary outcome was polymerase-chain reaction (PCR) adjusted cure rates at day 63.

The safety outcomes collected at baseline and during follow-up were AEs, total bilirubin, alanine aminotransferase, white blood cell and red blood cell counts. The patients were directly observed on days 0–2 (after receiving the dose). The patients were then asked to return to the clinic for follow-up visits on days 3 and 7 and then weekly thereafter until day 63. For this analysis, we focus on four outcomes; two clinical laboratory safety biomarkers (total bilirubin and alanine aminotransferase), clinical AEs and concomitant medication. The total bilirubin and alanine aminotransferase data was collected at enrolment, day 7, 14, 28 and 63. We focused on bilirubin and alanine aminotransferase biochemical parameters since they are key biochemical parameters in antimalarial drug

safety assessment during pregnancy and were also measured in the PREGACT trial.

The clinical AE count, defined as cumulative number of AEs experienced by the end of follow-up time is the primary outcome of interest in the current analysis. In all the subsequent discussions, the clinical AEs that were defined as *definitely not related* (by the study physician) to the antimalarial drug treatment are not considered in developing the joint model to avoid spurious results. The concomitant medication outcome is also defined as cumulative number of reported concomitant medication use by the end of the follow up time for each patient. In the current study, we considered the reported concomitant medication use regardless of its intended use.

Ethical considerations

The PREGACT trial was conducted in accordance with the Declaration of Helsinki and Good Clinical Practice guidelines. The trial obtained ethical clearance from ethics committee at the Antwerp University Hospital in Belgium and College of Medicine Research Ethics Committee at the University of Malawi [22]. Prior to enrolment, informed consent was also sought from the mother. Ethical approval was also obtained from University of the Witwatersrand Human Research Ethics Committee, prior to access and utilization of the data for the current analysis.

Notation and joint model specification

Let each patient who was randomized and received at least a dose be denoted as $i = 1, \dots, n$. Let $V_i = (v_{i1}^T, v_{i2}^T)^T$ be a bivariate continuous clinical laboratory outcome vector. Specifically, in our context we consider a bivariate scenario; $k=1$ is alanine amino transferase and $k=2$ is total bilirubin. Each of the two continuous outcome vectors (v_{i1}, v_{i2}) are of $(n_k \times 1)$ dimension for the observed longitudinal measurements of the k -th outcome; $v_{ik} = (v_{i1k}, \dots, v_{ink})^T$. We accommodate the situations where observation times, t_{ijk} may differ between individuals and outcomes ($j = 1, \dots, n_{ik}$). For each patient, we let $C_i = (c_{i1}^T, c_{i2}^T)^T$ represent a bivariate count outcome vector where $k=1$ is for total number of AEs experienced during their follow-up time T_i and $k=2$ for total number concomitant medication used over the follow-up time T_i . A set of covariates that were collected at baseline for each individual are defined as $X_i = \{X_{i1}, X_{i2}, \dots, X_{ip}\}$.

Conventional Poisson model for clinical AEs

In drug trials, where multiple safety outcomes are repeatedly collected to support safety evidence, clinical AEs are usually of primary interest. The occurrence of the clinical AEs is typically correlated with other safety-related

outcomes such as alanine aminotransferase, total bilirubin and concomitant medication. Modelling these outcomes separately is inefficient since it insufficiently accounts for the correlations. Traditionally, modelling of clinical AEs count is done using the conventional Poisson model, where the Poisson mean response (of clinical AEs in this context), ϕ , is assumed to be equal to the variance. Since the clinical AEs count is non-negative the logarithm of ϕ can naturally be linked with the baseline variables such that the conventional Poisson model for the clinical AE count can be presented as;

$$\log(\phi) = X_{ik}^T \beta \quad (1)$$

Where ϕ represents Poisson mean response and β is a vector of coefficients (i.e. log IRR) for the corresponding vector of baseline covariates, X_{ik}^T , that included maternal age, gravidity, trimester at enrolment and treatment arm. The model M1 has log link of the mean response of clinical AEs count and assumes a Poisson distribution of the clinical AEs. The coefficients of the covariates are interpreted on logarithm scale as expected change in the log of mean clinical AEs count per unit change in the covariate. Exponentiating the coefficients (i.e. log IRR) yields IRR. However, the estimated effect of treatment on clinical AEs from model M1 may be insufficient since it does not account for other factors that can confound the estimates e.g. concomitant medication and clinical laboratory outcomes. Alternatively, one can consider incorporating this additional information (of concomitant medication and clinical laboratory outcomes) as part of covariates in model M1. This can yield model M2 below.

$$\log(\phi) = X_{ik}^T(t) \beta \quad (2)$$

Since the concomitant medication and clinical laboratory outcomes (alanine aminotransferase, total bilirubin) are added in the model as time-varying covariates (with time t_0 , representing baseline covariates in M1), the vector of covariates adjusted in model M2 is $X_{ik}^T(t)$ denoting the vector of covariates value at time t . However, both model M1 and M2 do not account for correlation overtime across the longitudinal alanine aminotransferase, longitudinal total bilirubin, concomitant medication and clinical AEs outcomes. In this context of multiple outcomes that are correlated overtime, joint modelling offers an opportunity to obtain improved and efficient estimates since it can efficiently account for the potential correlations and baseline covariates. The joint model also enables formal quantification of the relationship between correlated outcomes by estimating the strength of the association between the outcomes.

Most previously proposed joint models consist of two sub-models; longitudinal sub-model and time-to-event

sub-model (where the time to event outcome is of primary interest). Here, we propose a joint model where the primary outcome of interest is count outcome (i.e. clinical AEs count) such that time to event outcome sub-model is replaced with count outcome sub-model. Since we assumed Poisson distribution of the clinical AEs count, the count outcome sub-model was specified as Poisson-distributed. Therefore, the joint model presented in this paper consist of two longitudinal continuous laboratory safety biomarkers sub-models (one for total bilirubin and one for alanine aminotransferase) two Poisson sub-models (one for the individual clinical AEs count and one for the individual concomitant medication count). The connecting of these outcomes sub-models is very flexible such that we could use random effects or expected value of outcomes [23] as detailed in the model specification below.

Joint model formulation

As highlighted above, the joint model is formulated in such a way that clinical AE count is a primary outcome. In this section we describe the structures of the sub-models that made the joint model considered in this work. The estimates in all the models are considered on a logarithm scale.

Longitudinal sub-model

We modelled two continuous longitudinal clinical laboratory safety outcomes (i.e. total bilirubin and alanine aminotransferase) using a longitudinal sub-model based on a flexible linear mixed effects model that could accommodate nonlinear changes of the laboratory safety outcomes. The flexibility was achieved through the use of restricted cubic splines. Each continuous longitudinal clinical laboratory outcome v_{i1} , v_{i2} was assumed to be normally distributed with mean μ_k and variance δ_k^2 ($k=1, 2$) on a logarithm scale. Each continuous longitudinal clinical laboratory outcome is modelled using the sub-model below

$$v_{ik}(t) = X_{ik}^T(t) \beta + Z_i(t) b_i + \varepsilon_{ik}(t), i = 1, \dots, n; k = 1, 2 \quad (3)$$

where $\varepsilon_{ik}(t)$ is the error term observed at time t for the k th outcome from patient i . The error terms for the model are assumed to be independent, identical and normally distributed with mean 0. The mean response model $X_{ik}^T(t) \beta$, specified as a linear function of the covariates at a given time for the outcome k is flexible such that it can accommodate time-varying covariates. The $Z_i(t)$ is an indicator vector of random effects for k th outcome for patient i at time t such that it takes the value of 1 when there is a random effect and 0 otherwise. The vector of

the patient-specific shared random effects $\mathbf{b}_i$ is assumed to have a multivariate normal distribution with mean 0 and variance-covariance matrix Σ , s.t. $(\mathbf{b}_i) \sim MVN(\mathbf{0}, \Sigma)$.

Alanine aminotransferase sub-model Based on M1, considering patient-specific random intercept as our shared parameter of interest, linking the outcomes, the longitudinal sub-model for alanine aminotransferase can be written simply as M4 below;

$$v_{i1}(t) = \mathbf{X}_{i1}^T(t)\boldsymbol{\beta} + \mathbf{b}_{0i1} \quad (4)$$

where $\mathbf{b}_{0i1}$ is the patient specific random intercept corresponding to the alanine aminotransferase outcome.

Total bilirubin sub-model Similarly, the longitudinal sub-model for total bilirubin, M5, can be formulated as;

$$v_{i2}(t) = \mathbf{X}_{i2}^T(t)\boldsymbol{\beta} + \mathbf{b}_{0i2} \quad (5)$$

where $\mathbf{b}_{0i2}$ is the patient specific random intercept corresponding to total bilirubin outcome.

Count sub-models

We considered two event count sub-models; one for concomitant medication count and the other for the AE count. The two count sub-models were derived from a modified form of the generic Poisson model. The count sub-model was as follows;

Concomitant medication count sub-model Assuming that the concomitant medication count had a Poisson distribution, its sub-model was defined as

$$\log(\phi|\mathbf{b}_{03i}) = \mathbf{X}_{ik}^T(t)\boldsymbol{\beta} + \mathbf{b}_{03i}\alpha_3 \quad (6)$$

such that $\log(\phi|\mathbf{b}_{03i})$ represents the logarithm of the Poisson mean response, ϕ , conditional on the patient-specific random intercept, $\mathbf{b}_{03i}$ for concomitant medication count (i.e. the concomitant medication count sub-model is linked with the AE count sub-model in the joint model via the random intercept). The $\mathbf{X}_{ik}^T(t)\boldsymbol{\beta}$ is the linear predictor for the Poisson model has a logarithm link function and contains a vector of fixed effects $\boldsymbol{\beta}$ corresponding to the baseline covariates. The α_3 represents log incidence rate ratio per unit increase in the patient-specific deviation from the mean random intercept of the reported concomitant medication use.

Clinical AE count sub-model Considering that AE count was the primary outcome of interest and assuming a Poisson distribution for the AE count, the AE count sub-model was formulated as;

$$\log(\phi|\mathbf{b}_{01i}, \mathbf{b}_{02i}, \mathbf{b}_{03i}) = \mathbf{X}_{ik}^T(t)\boldsymbol{\beta} + \mathbf{b}_{01i}\alpha_1 + \mathbf{b}_{02i}\alpha_2 + \mathbf{b}_{03i}\alpha_3 \quad (7)$$

The $\log(\phi|\mathbf{b}_{01i}, \mathbf{b}_{02i}, \mathbf{b}_{03i})$ represents the logarithm of Poisson mean response, ϕ , conditional on the shared random intercepts for the two longitudinal clinical laboratory outcomes and the concomitant medication count. Each random intercept for the longitudinal sub-models and concomitant medication sub-model ($\mathbf{b}_{01i}, \mathbf{b}_{02i}, \mathbf{b}_{03i}$) is linked to the linear predictor $\mathbf{X}_{ik}^T(t)\boldsymbol{\beta}$ for the Poisson model. This yields random effects association structure where $\alpha_1, \alpha_2, \alpha_3$ estimate the strength of the association between the respective continuous longitudinal clinical laboratory outcome, concomitant medication count and the AE count; the $\alpha_1, \alpha_2, \alpha_3$ represent log incidence rate per unit increase in the patient-specific deviation from the mean random intercept of a respective continuous longitudinal clinical laboratory outcome or the concomitant medication count outcome. For example, α_1 represents log incidence rate ratio per unit increase in the patient-specific deviation from the mean random intercept of a respective log alanine aminotransferase.

Alternative clinically meaningful formulation of the AE count sub-model could be linked to the expected value of the respective continuous longitudinal clinical laboratory outcome with the linear predictor of the Poisson model. This yields the current value association structure where the $\alpha_1, \alpha_2, \alpha_3$ represent log incidence rate ratio per unit increase of the respective continuous longitudinal clinical laboratory outcome, at time t. Current value association structure is very important and clinically plausible when linking continuous outcomes. In our cases, given the two continuous longitudinal clinical laboratory outcomes (i.e. log alanine aminotransferase and log total bilirubin) $\mathbf{v}_1, \mathbf{v}_2$, under the current value association structure, we can modify the AE count sub-model as;

$$\log(\phi) = \mathbf{X}_{ik}^T(t)\boldsymbol{\beta} + E[\mathbf{v}_1(t)|\mu_1(t)]\alpha_1 + E[\mathbf{v}_2(t)|\mu_2(t)]\alpha_2 + \mathbf{b}_{03i}\alpha_3 \quad (8)$$

The $E[\mathbf{v}_1(t)|\mu_1(t)]$ and $E[\mathbf{v}_2(t)|\mu_2(t)]\alpha_2$ are the expected current values of log alanine aminotransferase and total bilirubin respectively. It can be noted from M6 that the concomitant medication count, $\mathbf{v}_3$, sub-model is still linked to the AE count sub-model via a random intercept $\mathbf{b}_{03i}$. We maintained the random intercept ($\mathbf{b}_{03i}\alpha_3$) parameterization since this efficiently deals with any potential over-dispersion problem encountered in Poisson models. Therefore, model M8 contains a mixture of both the expected current value and random effects association structures. Since both continuous outcomes (the log alanine aminotransferase and log total bilirubin) are assumed to be normally distributed (i.e. with identity link), the model can further be simplified as;

$$\log(\varphi) = X_{ik}^T(t)\beta + [\mu_1(t)]\alpha_1 + [\mu_2(t)]\alpha_2 + b_{03i}\alpha_3 \quad (9)$$

In this paper, we focus on reporting results from the M9 association structure since it is more clinically meaningful since it has association parameters that are directly interpretable in clinical practice.

Joint model likelihood and estimation

The estimates for the joint model are obtained by maximizing the marginal likelihood of the joint distribution of the observed data and the random effects [23]. The likelihood function for the observed data can be specified as;

$$L_i(\theta) = \prod_{i=1}^n f(V_i|b_i, \theta) f(T_i, C_i|b_i, \theta) f(b_i|\theta) db_i \quad (10)$$

where θ represents the set of parameters of interest to be estimated including both the fixed and random effects (e.g. association parameters α_1 , α_2 , α_3 , and coefficients for the baseline covariates, β). Component $f(b_i|\theta)$ is a normal density function for the random effects conditional on θ and $f(V_i|b_i, \theta)$ is a normal density function for the log total bilirubin and log alanine aminotransferase conditional on θ and b_i . The $f(T_i, C_i|b_i, \theta)$ represents Poisson density function conditional on θ and b_i ; patient-specific total follow up days is T_i and C_i represents total events recorded (for AE count sub-model and concomitant count sub-model). Since obtaining the log likelihood requires integrating out the random effects, it becomes challenging as number of random effects increases such that the traditionally used adaptive Gauss-Hermite quadrature would be limited (to handle the large number of random effects in our complex joint model). To mitigate this problem, we employed Monte Carlo integration technique since the number of draws it makes from the random effects do not need to change with increase in number of random effects [23, 24] (hence reducing the computation burden). Estimation of our models was done using merlin in Stata [21]. Only 7 patients had missing data points and the data missing mechanism was considered completely at random.

Statistical data analysis

Firstly, we computed the summary of the baseline characteristics for the women at enrolment. We presented means with respective standard deviation for the continuous variables (e.g. maternal age, gestational age) and frequencies with respective percentages for the categorical variables (e.g. trimester at enrolment, bed-net use). These summaries were also computed for the clinical AE occurrence and concomitant medication counts. The concomitant medication was also summarized based frequency of

the specific patients and the number of patients who took the medication. In order to aid visualization of the trend of the data, we plotted box plots for the two continuous longitudinal clinical laboratory data (total bilirubin and alanine aminotransferase) outcomes by study visit day.

We considered two different models to compare how they efficiently estimated effect of treatment on clinical AEs over the follow-up time. The models compared were conventional Poisson model M1 and joint model of M9 formulation above. Both the models adjusted for the same set of baseline covariates (maternal age, gravidity and trimester at enrolment). The traditional Poisson model was the first model to be fitted and we used logarithm of the total follow-up time for each patient as an offset.

Then we fitted a joint model with a mixture of expected current value of outcome and random effects association structure as shown in M9. We considered the joint model to assess the joint evolution (/association) between the clinical AEs and other three outcomes (alanine aminotransferase, total bilirubin and concomitant medication) where α_1 , α_2 , α_3 as defined in M9 are the parameters quantifying the between-outcome association. Secondly, we were also interested to investigate the efficiency of the joint model in improving AE incidence rate estimate of the treatment effect obtained from the conventional Poisson model M1. As part of sensitivity analysis we also investigated how the association structure of the joint model affects the magnitude of the treatment effect estimates. Comparing the association structures was achieved by fitting a joint model with random effects association structure only as specified in model M7 above and joint model with mixture of current value and random effects association structure of joint model M9. For both joint models, in order to flexibly model nonlinear changes of the continuous longitudinal clinical laboratory outcomes we used restricted cubic splines with 3 degrees of freedom considered sufficient to enhance flexibility of the model. We compared the fit of the two joint models (M9 versus M7) using the Akaike Information Criterion (AIC), with small values of AIC suggesting better model.

Additionally, we did an exploratory analysis to identify how the frequently reported concomitant medications are associated with the AEs. Such exploratory analysis was helpful in understanding how specific frequently used concomitant medication influence the overall impact of concomitant medication on AEs. Concomitant medication was defined as frequent if it constituted at least 10% of the reported medications. A joint model of a similar structure to M9 but where the concomitant medication count was confined to the specific frequently reported concomitant medication (i.e. paracetamol) was used during the exploratory analysis.

Results

The trial recruited 870 women who were randomly assigned to received AL, ASAQ, DHAPQ; each arm recruited 290 women. Baseline characteristics were similar across and between the trial arms (Table 1). Maternal age was skewed to the right. Overall, median maternal age was 20 years (IQR: 18.0–24.0). Overall mean haemoglobin concentration at enrolment (g/dl) was 10.1 (SD: 1.3) and distribution was also consistently similar across all the treatment arms. The overall reported bed net use in a night before enrolment was 23.2% such and it was similar across and between the study arms (Table 1).

In total 1512 AEs were observed over the whole follow-up time and 108 of these were definitely not related to the treatment. Proportion of women who experienced at least one AE in the AL, ASAQ and DHAPQ treatment arms were 64.5% (187 of 300 women), 70.7% (205 of 290 women) and 59.3% (172 of 290 women), respectively. Almost all patients took at least a concomitant medication over the follow-up such that for each arm 99% (287 out of 290) took at least a concomitant medication (Table 2). Across all the treatment arms, the reported median concomitant medications taken was 3; for AL 3 (IQR: 2–5), for ASAQ 3 (IQR: 1–4) and for DHAPQ 3 (IQR: 2–5).

We observed that iron supplement (26.3%), paracetamol (26.2%) and albendazole (10.7%) were frequently reported concomitant medications (Table 3). Almost all the patients analysed, 99.2%, took the iron supplements and approximately two-thirds, 60.3%, of the patients took paracetamol apart from the antimalarial drug under investigation. During the follow-up period, a total of 353 patients reported having used albendazole once.

Overall median of log-transformed alanine aminotransferase was 2.7 (IQR: 2.5–3.0) such that the medians for AL was 2.8 (IQR: 2.5–3.0), for ASAQ was 2.7 (IQR: 2.5–3.0) and for DHAPQ was 2.8 (IQR: 2.5–3.0). Overall median for log-transformed total bilirubin was 2.0 (IQR: 1.7–2.4). Figure 1 below shows the detailed distribution

Table 2 Summary of AE occurrence and concomitant medication use for women enrolled in PREGACT trial

Characteristic	AL	ASAQ	DHAPQ
Total AEs	475	569	468
AE drug-relatedness totals			
Definitely not	36	34	38
Possibly/probably	26	96	26
Unlikely	413	443	404
At least one AE, n (%)	187 (64.5)	205 (70.7)	172 (59.3)
Total concomitant medications	1229	1033	1042
Median concomitant medications (IQR)	3 (2, 5)	3 (1, 4)	3 (2, 5)
At least one concomitant medication, n (%)	287 (99.0)	287 (99.0)	287 (99.0)

of each of clinical laboratory safety outcomes over the follow-up days by antimalarial drug treatment arm. As shown in graph 1 within Fig. 1, the alanine amino-transferase did not vary much throughout the follow-up days. The general trend indicated that across the follow-up days, the alanine amino-transferase concentration was slightly higher among the women who received DHAPQ compared to those women who received either AL or ASAQ. After day 0 the total bilirubin concentration gradually decreased across all the treatment arms and, on average, women treated with AL had the highest bilirubin concentration across the follow-up days (see Graph 2 within Fig. 1).

Parameter estimates for the models

In this section, we report the estimates from the conventional Poisson model (where AEs count is an outcome) (Table 4) and the proposed joint model (Table 5). Our interest was to establish whether the joint model yielded better results than the conventional Poisson model and how the jointly modelled outcomes evolved over the follow-up time in relation to clinical AEs; the results from

Table 1 Baseline characteristics for women enrolled in PREGACT trial in Malawi

Characteristic	Overall	AL	ASAQ	DHAPQ
Maternal age (years), median (IQR)	20.0 (18.0–24.0)	20.0 (18.0–24.0)	20.0 (18.0–25.0)	20.0 (19.5–24.0)
Gestation age (weeks), median (IQR)	21.0 (18.0–25.0)	21.0 (18.0–25.0)	21.0 (18.0–25.0)	21.0 (18.0–24.0)
Haemoglobin (g/dl), mean (SD)	10.1 (1.3)	10.2 (1.4)	10.0 (1.3)	10.1 (1.3)
Primigravida, n (%)	444 (51.1)	144 (49.8)	144 (49.7)	156 (53.8)
Trimester				
Second, n (%)	730 (83.9)	242 (83.4)	244 (84.1)	244 (84.1)
Third, n (%)	140 (16.1)	48 (16.6)	46 (15.9)	46 (15.9)
Bed net use, n (%)	202 (23.2)	65 (22.4)	67 (23.1)	70 (24.1)

Table 3 Summary of concomitant medication reported by pregnant women in PREGACT trial

Concomitant medication	Number of times reported ^b , n(%)	Number of patients ^a , n(%)
Iron supplement	869 (26.3)	856 (99.2)
Paracetamol	864 (26.2)	520 (60.3)
Others	433 (13.1)	302 (35.0)
Albendazole	353 (10.7)	353 (40.9)
Amoxicillin	220 (6.7)	192 (22.3)
Benzathine	108 (3.3)	68 (7.9)
Pethidine	49 (1.5)	49 (5.7)
Chloramphenicol	48 (1.5)	43 (5.0)
Oral Rehydration Salts (ORS)	47 (1.4)	44 (5.1)
Dextrose	47 (1.4)	45 (5.2)
Lignocaine	39 (1.2)	39 (4.5)
Metronidazole	38 (1.2)	36 (4.3)
Quinine	35 (1.1)	26 (3.0)
Gentamycin	35 (1.1)	35 (4.1)
Sulfadoxine-Pyrimethamine	31 (0.9)	27 (3.1)
Aspirin	16 (0.5)	14 (1.6)
Pitocin	16 (0.5)	16 (1.9)
Erythromycin	13 (0.4)	12 (1.4)
Piriton	12 (0.4)	12 (1.4)
Methylidopa	12 (0.4)	9 (1.0)
Cotrimoxazole	8 (0.2)	8 (0.9)
Promethazine	7 (0.2)	7 (0.8)
Multivitamin Tabs	4 (0.1)	3 (0.4)

^a This is out of 863 patients who used concomitant medication; some patients took multiple concomitant medications

^b This is out of 3304 times of reported concomitant medication use

the AE count sub-model in the joint model (Table 5) are of main interest in our reporting.

Parameter estimates from the conventional Poisson model for the clinical AE count

Using the conventional Poisson model, M9, the clinical AE incidence rate did not differ significantly across the treatment groups ($p < 0.149$). However, the estimates suggested that the AE incidence rate was higher in the ASAQ treatment group compared to the AL treatment group (log IRR: 0.1838; 95% CI: 0.0574, 0.3102; $p = 0.004$) (Table 4). Exponentiating the point estimate of the log IRR yields IRR of 1.2018, implying that the patients treated with ASAQ were expected to have 20.18% higher rate of clinical AEs occurrence compared to those treated with AL. The AE incidence rate did not differ between DHAPQ and AL treatment groups (log IRR: -0.0255; 95% CI: -0.1587, 0.1076; $p = 0.707$). A year increase in maternal age was associated with 0.0208 increase in log IRR of the AEs (95% CI: 0.0068, 0.0347; $p = 0.004$). The AE incidence rate was significantly higher among the those who enrolled in the second trimester compared to those who enrolled in the third trimester (log IRR: 0.3768; 95% CI: 0.1943, 0.5192; $p < 0.001$).

Parameter estimates from the joint model

We report the fixed and random parameter estimates for the final fitted joint model based on mixed association structure, M9, (i.e. random effects and current value association structure) in Table 5. AE count sub-model indicated that the AE incidence rate did not vary across all the treatment groups ($p = 0.376$). However, the estimates

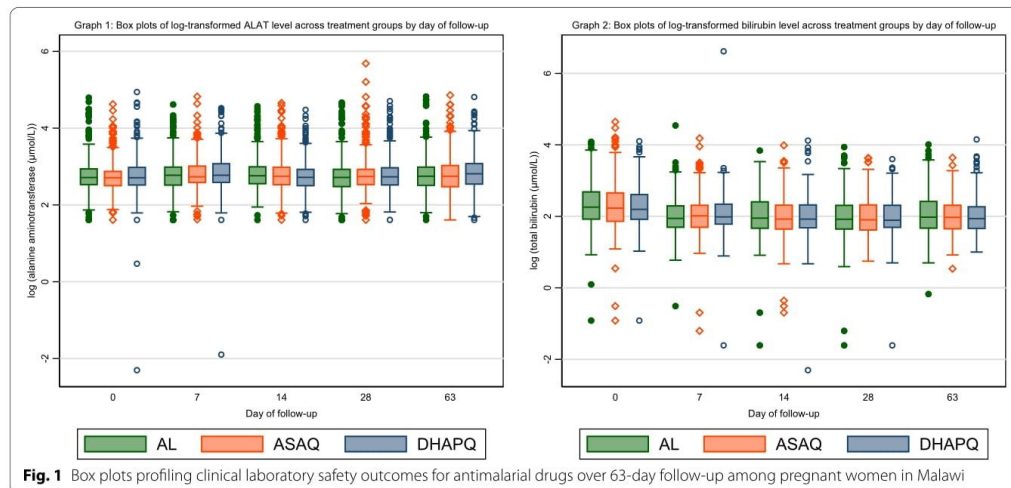


Table 4 Parameter estimates for clinical AE count incidence using conventional Poisson model

Characteristic	Coefficient	SE	95% CI	P-value
Treatment				
ASAQ	0.1838	0.0645	0.0574, 0.3102	0.004
DHAPQ	-0.0255	0.0679	-0.1587, 0.1076	0.707
Maternal age	0.0208	0.0071	0.0068, 0.0347	0.004
Primigravid	0.1208	0.0731	-0.0224, 0.2641	0.098
Second trimester	0.3768	0.0829	0.1943, 0.5192	< 0.001
Constant	-0.3901	0.2021	-0.7862, 0.0060	0.054

indicated that AE incidence rate was higher in ASAQ treatment group compared to AL (log IRR: 0.2118, 95% CI: 0.0082, 0.4154; $P=0.041$) suggesting that patients treated with ASAQ were expected to have 23.59% higher rate of clinical AEs occurrence compared to those treated with AL (i.e. $IRR = \exp(0.2118) = 1.2359$). The AE incidence rate did not significantly differ in DHAPQ treatment compared to AL (log IRR: -0.0508; 95% CI: -0.2559, 0.1542). Our results show that the magnitude of the treatment effect was higher in the joint model, M9, compared to the estimates in the conventional Poisson model, M1, above. Among the association parameters describing joint evolution between the outcomes, only concomitant medication was associated with clinical AEs. The association parameter, α_3 , linking concomitant medication and clinical AEs suggested that the clinical AEs outcome was positively and strongly associated with the concomitant medications (log IRR: 1.7487; 95% CI: 1.5471, 1.9503; $p < 0.001$). A unit increase for a patient in deviation from the mean random intercept of concomitant medication use was associated with 5-fold increase in rate of clinical AEs (i.e. $IRR = \exp(1.7487) = 5.7471$). We found that the clinical AEs were not associated the alanine aminotransferase nor total bilirubin over the follow-up time as demonstrated by their respective association parameters that were non-significant. For the association between alanine aminotransferase and clinical AEs over follow-up time the estimated log IRR was -0.0288 (95% CI: -0.5045, 0.4469; $p=0.906$) and for the association between total bilirubin and clinical AEs the log IRR was 0.1153 (95%CI: -0.0889, 0.3194; $p=0.269$). As also observed in the conventional Poisson model above, the clinical AE sub-model indicated that those women who enrolled in the second trimester had a higher AE incidence rate compared to those who enrolled in the third trimester (log IRR 0.3986; 95% CI: 0.1592, 0.6381; $p=0.001$); the women who enrolled in the second trimester had 48.97% higher rate of clinical AEs (i.e. $IRR = \exp(0.3986) = 1.4897$) compared to those who enrolled in the third trimester.

Table 5 Parameter estimates from the joint multivariate longitudinal clinical laboratory, concomitant medication and AE counts model

Variable	Coefficient (SE)	95% Confidence interval	P-value
Log alanine amino transferase longitudinal sub-model			
Maternal age	0.0006 (0.0023)	-0.00394, 0.0052	0.793
Treatment			
ASAQ	-0.0036 (0.0275)	-0.0575, 0.0504	0.897
DHAPQ	0.0031 (0.0275)	-0.0508, 0.0570	0.909
Second trimester	-0.0212 (0.0311)	-0.0820, 0.0397	0.496
Constant	2.8221 (0.0609)	2.7027, 2.9415	< 0.001
Log total bilirubin longitudinal sub-model			
Maternal age	-0.0026 (0.0035)	-0.0095, 0.0043	0.464
Treatment			
ASAQ	-0.0229 (0.0424)	-0.1059, 0.0601	0.589
DHAPQ	-0.0413 (0.0422)	-0.1240, 0.0414	0.328
Second trimester	-0.1786 (0.0493)	-0.2752, -0.0820	< 0.001
Constant	2.3103 (0.0963)	2.1216, 2.4990	< 0.001
Concomitant medication count sub-model			
Maternal age	0.0086 (0.0052)	-0.0017, 0.0189	0.101
Treatment			
ASAQ	-0.1841 (0.0619)	-0.3054, -0.0629	0.003
DHAPQ	-0.1723 (0.0610)	-0.2918, -0.0528	0.005
Second trimester	0.2861 (0.0718)	0.1454, 0.4268	< 0.001
Constant	0.9147 (0.1395)	0.6414, 1.188	< 0.001
Adverse events count sub-model			
Maternal age	0.0189 (0.0088)	0.0017, 0.0362	0.031
Treatment			
ASAQ	0.2118 (0.1039)	0.0082, 0.4154	0.041
DHAPQ	-0.0508 (0.1046)	-0.2559, 0.1542	0.627
Second trimester	0.3986 (0.1222)	0.1592, 0.6381	0.001
$\alpha_{\log(ALT)}$	-0.0288 (0.2427)	-0.5045, 0.4469	0.906
$\alpha_{\log(bilirubin)}$	0.1153 (0.1042)	-0.0889, 0.3194	0.269
α_{conmed}	1.7487 (0.1029)	1.5471, 1.9503	< 0.001
Constant	-0.8220 (0.7739)	-2.3388, 0.6949	0.288
Patient-specific random effects standard deviations			
Standard deviations			
$\delta_{\log(ALT)}$	0.2547 (0.0119)	0.2324, 0.2792	
$\delta_{\log(bilirubin)}$	0.5025 (0.0154)	0.4732, 0.5336	
$\delta_{\log(conmed)}$	0.5378 (0.0248)	0.4914, 0.5887	

ALT Alanine aminotransferase, conmed Concomitant medication

In the sub-model for log alanine aminotransferase, we observed no significant difference across all the variables included in that sub-model. For the log total bilirubin sub-model, we observed 0.1786 lower log total bilirubin among the women who enrolled in the second trimester compared to those who enrolled in the third trimester (log IRR: -0.1786; 95% CI: -0.2752, -0.0820, $p < 0.001$). The concomitant sub-model showed that

concomitant medication was associated with the treatment group. Women enrolled in ASAQ treatment group reported lower incidence rate of concomitant medication use compared to those in AL treatment group (log IRR: -0.1841; 95% CI: -0.3054, -0.0629; $p=0.003$). Women in DHAPQ treatment group also reported lower incidence rate of concomitant medication use compared to those in AL treatment group. Those women who enrolled in the second trimester reported a higher incidence rate of concomitant medication use than those who enrolled in the trial while in the third trimester (log IRR 0.2861; 95% CI: 0.1454, 0.4268; $p<0.001$).

The standard deviations for patient-specific random intercepts, across the outcomes, suggested that concomitant medication had the highest unobserved heterogeneity with standard deviation of 0.5378 (95% CI: 0.4914, 0.5887). The lowest unobserved heterogeneity was observed in log alanine aminotransferase outcome with random intercept standard deviation of 0.2547 (0.2324, 0.2792).

Sensitivity analyses

As part of sensitivity analysis, we attempted to relax the normality assumption of the random effects. We fitted a robust joint model for the mixture association structure, M9, where the random effects were assumed to have a t-distribution. The model yielded similar results as those in Table 5 and Akaike Information Criterion (AIC) of 18,230.28 which was comparable to an AIC of 18,278.53 for the model in Table 5. Secondly, fitting a joint model with random effects association structure, M5, we obtained similar estimates as those in Table 5 and with negligibly lower AIC of 18,277.87.

Following a reviewer's suggestion, we further investigated the effect of the concomitant medications with high frequencies on AEs. Although iron supplement was the most commonly reported concomitant medication, we did not consider it in assessing the impact of specific concomitant medication on AEs since it was taken by almost all the patients, 99.2%. Instead, we considered paracetamol since it was the second-highest reported concomitant medication where 60.3% of the patients reported it at least once. Assessment of the effect of treatment on clinical AEs using a joint model of a similar structure to M9 but where the concomitant medication count was confined to paracetamol yielded similar results as those obtained in Table 5 (see Table 6). The paracetamol concomitant medication was positively and strongly associated with clinical AEs (log IRR: 1.2969; 95% CI: 1.1328, 1.4609; $p<0.001$). However, this observed effect of paracetamol concomitant medication was slightly lower compared to the overall concomitant medication effect as reported in Table 4 (log IRR: 1.7487; 95% CI: 1.5471, 1.9503; $p<0.001$). No interaction

was observed between the paracetamol concomitant medication and the treatment ($p=0.078$) such that we focussed on reporting results with the main effect only model as shown in Table 6. The model with interaction terms had a negligibly lower AIC of 16,556.19 than the model without the interaction terms with an AIC of 16,557.29.

Table 6 Parameter estimates from the joint multivariate continuous longitudinal clinical laboratory, paracetamol concomitant medication count and AE count model

Variable	Coefficient (SE)	95% Confidence interval	P-value
Log alanine amino transferase longitudinal sub-model			
Maternal age	0.0006 (0.0023)	-0.0040, 0.0051	0.810
Treatment			
ASAQ	-0.0016 (0.0276)	-0.0557, 0.0525	0.953
DHAPQ	0.0031 (0.0275)	-0.0508, 0.0570	0.909
Second trimester	-0.0223 (0.0311)	-0.0833, 0.0387	0.474
Constant	2.8236 (0.0608)	2.7043, 2.9428	<0.001
Log total bilirubin longitudinal sub-model			
Maternal age	-0.0027 (0.0034)	-0.0094, 0.0039	0.418
Treatment			
ASAQ	-0.0346 (0.0419)	-0.1168, 0.0476	0.409
DHAPQ	-0.0483 (0.0416)	-0.1299, 0.0334	0.246
Second trimester	-0.1747 (0.0471)	-0.2669, -0.0824	<0.001
Constant	2.3118 (0.0911)	2.1394, 2.4966	<0.001
Concomitant medication count sub-model			
Maternal age	0.0144 (0.0088)	-0.0030, 0.0318	0.103
Treatment			
ASAQ	-0.3469 (0.1068)	-0.5563, -0.1376	0.001
DHAPQ	-0.2850 (0.1046)	-0.4901, -0.0798	0.006
Second trimester	0.3850 (0.1273)	0.1355, 0.6344	0.002
Constant	-0.6712 (0.2385)	-1.1387, -0.2038	0.005
Adverse events count sub-model			
Maternal age	0.0161 (0.0092)	0.0019, 0.0340	0.079
Treatment			
ASAQ	0.1687 (0.1099)	0.0466, 0.3840	0.125
DHAPQ	-0.0876 (0.1103)	-0.3038, 0.1286	0.427
Second trimester	0.3986 (0.1270)	0.1497, 0.6474	0.002
$a_{\log(ALT)}$	0.0016 (0.2392)	-0.4673, 0.4704	0.995
$a_{\log(bilirubin)}$	0.0218 (0.1207)	-0.2148, 0.2585	0.856
a_{concmcd}	1.2969 (0.0837)	1.1328, 1.4609	<0.001
Constant	-0.6805 (0.7793)	-2.2079, 0.8469	0.383
Patient-specific random effects standard deviations			
Standard deviations			
$\delta_{\log(ALT)}$	0.2555 (0.0119)	0.2331, 0.2800	
$\delta_{\log(bilirubin)}$	0.5028 (0.0153)	0.4736, 0.5338	
$\delta_{\log(\text{concmcd})}$	0.7931 (0.0480)	0.7043, 0.8931	

ALT Alanine aminotransferase, *concmcd* Paracetamol concomitant medication

The interest and structure of the proposed joint model could not permit us to assess the impact of albendazole on AEs. The patients who took albendazole took it once; therefore, this concomitant medication arose as a binary outcome that does not meet the joint model specification (i.e. where the concomitant medication is incorporated in the model as a count outcome). Since the objective of the exploratory analysis was to establish the impact of specific frequent concomitant medication on AEs, we did not undertake any additional analysis on the effect of the “other” concomitant medication on AEs.

Discussion

In clinical trials, multiple safety outcomes collected over the follow-up time require advanced analyses in order to develop a valid drug safety profile that makes full use of the information from the multiple safety outcomes. This paper introduced a framework for joint modelling of multivariate continuous longitudinal clinical laboratory safety outcomes, concomitant medication counts and adverse event count, exploiting the advantages of the existing extended multivariate generalised linear and non-linear mixed effects models framework [23]. Our work focussed on investigating whether clinical AEs varied by treatment arm and how the laboratory outcomes and concomitant medication were associated with clinical AEs over follow-up time in artemisinin-based treatment of malaria during pregnancy clinical trial. We found that the concomitant medication was strongly associated with the clinical AEs incidence, suggesting that the patients who experienced more AEs were more likely to report the use of more concomitant medications. If not well-accounted for such increased use of concomitant medication can distort the treatment effect on the clinical AE occurrence. From methodological perspective, we discovered that the conventional Poisson model underestimated the treatment effect on the AE incidence rate ratio; the effect of treatment on AE incidence rate estimates for the joint model were higher than those from conventional Poisson model. The difference in the models can be attributed to the fact that the joint model uses more of the available information through the explicit modelling of the observed correlation, unobserved heterogeneity and any nonlinear effects [25–27]. The joint model efficiently uses the available information, including the unobserved heterogeneity, leading to more improved estimates compared to the conventional Poisson model involving separate analysis.

Controlling for baseline covariate to address any potential confounding in AE incidence rate ratio estimates, we found that maternal age and enrolling in the second trimester were factors associated with the AE occurrence. A year increase in maternal age was associated with 2%

increased rate of AE incidence. This could be as a result that as the women become older they are more likely to report AEs out of experience than the younger women. As expected women who enrolled in the second trimesters had 49% increased rate of AE incidence compared to those who enrolled in the third trimester. This could probably be due to the fact that physiological changes that take place during pregnancy are more impactful in the second than in the third trimester. Hence, the susceptibility to AE occurrence after taking either the anti-malarial drug or the concomitant medication (that was strongly associated with increase rate of AE occurrence) since the pregnancy related body physiological changes can interact with the drug.

In assessment of the impact of concomitant medication on AEs, identifying specific concomitant medication is usually of interest. In this paper, we observed that paracetamol greatly contributed to the overall association between concomitant medication and AEs. Although the difference in the effect of overall concomitant medication on AEs and the effect of paracetamol concomitant medication on AEs was not big, the difference could not be considered clinically negligible. Therefore, undertaking both analyses (i.e. effect of overall concomitant medication and effect of specific most frequent concomitant medication on AEs) should be considered useful in understanding antimalarial drug safety profile. Alternatively, an interesting future joint model development can consider quantifying the contribution of each of the specific frequent concomitant medication on AEs occurrence. This can be helpful in identifying and handling any potential noisy concomitant medications that contribute less to the AEs occurrence.

Within the joint modelling framework, our proposed model provides additional knowledge by extending the number of outcomes that can be analysed simultaneously, including multiple longitudinal outcomes and multiple count outcomes, focussing on drug safety assessment. We acknowledge that other researchers have previously proposed similar kind of modelling but applied in the efficacy context [28–32]. Buu et al. considered a joint model of count and binary outcome [30]. Li et al. considered jointly modelling proportion, count and continuous longitudinal outcomes [31]. Yang and Kang introduced a joint model for mixed Poisson outcome and continuous outcome [32]. Our model is considered as a special case of these previously proposed models focussing on drug safety assessment. Key unique features of our proposed joint model include incorporating mixed association structure of random effects and expected value for linking the outcomes, use of restricted cubic splines to accommodate the capturing of any potential time-dependent effects, accommodating multiple safety

outcomes and novel application to drug safety assessment. A case study based paper like ours can further facilitate the adoption of improved statistical analysis of safety data in clinical trials [33].

In this paper, data missingness mechanism was ignorable. Our future work will consider investigating the impact of missing data on the model performance under varying follow-up times and model association structures. Secondly, in this paper we worked with a large sample size. An interesting further methodological work would be to assess how the sample size impacts the performance of the model that we applied. Unfortunately, the key limitation of the joint model introduced and applied in this paper is that it is computationally demanding such that it takes longer time to converge. This directly affects the scope of the simulations that can be done to investigate the highlighted potential methodological issues associated with such complicated models. An extension can also be made to the number of outcomes included the joint model depending on the availability of the data. For example, aspartate amino transferase, highly correlated with alanine amino transferase, can be added in our joint model.

Conclusion

Jointly modelling of multivariate longitudinal clinical laboratory safety data, concomitant medication and clinical AEs efficiently harnesses the safety data in order to better understand drug safety profile. This paper has demonstrated the utility of joint modelling to yield improved AE incidence rate estimates compared to the conventional Poisson model. The proposed joint model helps to quantify the association between the laboratory safety outcomes, concomitant medication, and clinical AEs over follow-up time. Key public health relevance of the joint modelling of the safety outcomes includes facilitating the choice of the most important clinical laboratory safety and other safety outcomes for profiling drug safety, especially in resource-constrained settings. In this exploratory analysis, we have also established that the AEs that were observed in the PREGACT trial in Malawi were more linked to concomitant medications that the patients took. Our findings provide an assurance that artemisinin-based treatments can be safely used in second and third trimester during pregnancy.

Abbreviations

AEs: Adverse events; AL: Artemether–lumefantrine; ASAQ: Amodiaquine–artesunate; DHAPQ: Dihydroartemisinin–piperaquine; IRR: Incidence rate ratio; CI: Confidence interval; PCR: Polymerase-chain reaction; AIC: Akaike Information Criterion.

Acknowledgements

We thank all the study participants and data collectors in the clinical trial.

Authors' contributions

Study idea conceptualization: NP, TC, MM. Methodology development: NP, MM, LK, TC, MKL, DM, VM, GC, UD. Analysis and first draft development: NP. Supervision, writing revised drafts-reviewing and editing: MM, LK, TC, MKL, DM, VM, GC, UD. Enrolled patients and conducted clinical investigation: VM, GC, UD. Reading and approving the final manuscript: NP, MM, LK, TC, MKL, DM, VM, GC, UD.

Funding

The clinical trial was supported by the European and Developing Countries Clinical Trials Partnership, the Malaria in Pregnancy Consortium (which is funded through a grant from the Bill and Melinda Gates Foundation to the Liverpool School of Tropical Medicine), the Belgian Development Cooperation Agency, the Liverpool School of Tropical Medicine, the Medical Research Council UK, the Netherlands Organization for Scientific Research, and Sanofi-Aventis NP was funded by the Fogarty International Center of the National Institutes of Health [Grant number D43TW010075 to MKL, TC and DM].

Availability of data and materials

The data that support the findings of this study are available from Institute of Tropical Medicine, Antwerp, Belgium but restrictions apply to the availability of these data, which were used under license for the current study, and so are not publicly available. Data are however available from the corresponding author upon reasonable request and with permission of Institute of Tropical Medicine, Antwerp, Belgium (<http://www.itg.be>).

Declarations

Ethics approval and consent to participate

This study was approved by the University of the Witwatersrand's Human Research Ethics Committee. The PREGACT trial received ethical clearance from ethics committee at the Antwerp University Hospital and College of Medicine Research Ethics Committee at the University of Malawi in Malawi [22]. Informed consent was obtained from all the participants.

Consent for publication

Not applicable.

Competing interests

The authors declare that they have no competing interests.

Author details

¹School of Public Health, University of the Witwatersrand, Johannesburg, South Africa. ²School of Public Health and Family Medicine, College of Medicine, University of Malawi, Blantyre, Malawi. ³Mahidol Oxford Tropical Medicine Research Unit (MORU), Bangkok, Thailand. ⁴Centre for Tropical Medicine, Nuffield Department of Medicine, University of Oxford, Oxford, UK. ⁵Medical Research Council Unit, The Gambia at the London School of Hygiene and Tropical Medicine, Fajara, Gambia. ⁶Department of Biostatistics, University of Namibia, Windhoek, Namibia. ⁷Center for Vaccine Development and Global Health, University of Maryland, School of Medicine, Baltimore, MD, USA.

Received: 22 March 2021 Accepted: 17 September 2021

Published online: 09 October 2021

References

1. Zink RC, Marchenko O, Sanchez-Kam M, Ma H, Jiang Q. Sources of safety data and statistical strategies for design and analysis: clinical trials. *Ther Innov Reg Sci*. 2018;52(2):141–58.
2. Munsaka MS. A question-based approach to the analysis of safety data. In: *Biopharmaceutical applied statistics symposium. Biostatistical analysis of clinical trials*, vol. 2. Singapore: Springer; 2018. p. 193–216.
3. Merz M, Lee KR, Kullak-Ublick GA, Brueckner A, Watkins PB. Methodology to assess clinical liver safety data. *Drug Saf*. 2014;37(1):33–45.
4. Chuang-Stein C. Laboratory data in clinical trials: a statistician's perspective. *Control Clin Trials*. 1998;19(2):167–77.

5. Klepper MJ, Cobert B. Drug safety data: how to analyze, summarize and interpret to determine risk. Sudbury: Jones & Bartlett Learning; 2010.
6. Siddiqui O. Statistical methods to analyze adverse events data of randomized clinical trials. *J Biopharm Stat*. 2009;19(5):889–99.
7. Unkel S, Amiri M, Benda N, Beyersmann J, Knoerzer D, Kupas K, et al. On estimands and the analysis of adverse events in the presence of varying follow-up times within the benefit assessment of therapies. *Pharm Stat*. 2019;18(2):166–83.
8. Patson N, Mukaka M, Otumbe KN, Kazembe L, Mathanga DP, Mwapasa V, et al. Systematic review of statistical methods for safety data in malaria chemoprevention in pregnancy trials. *Malar J*. 2020;19(1):1–11.
9. Barker C. Exploratory method for summarizing concomitant medication data—the mean cumulative function. *Pharm Stat*. 2010;9(4):331–6.
10. Schildcrout JS, Jenkins CA, Ostroff JH, Gillen DL, Harrell FE, Trost DC. Analysis of longitudinal laboratory data in the presence of common selection mechanisms: a view toward greater emphasis on pre-marketing pharmaceutical safety. *Stat Med*. 2008;27(12):2248–66.
11. Rosenkranz GK. Modeling laboratory data from clinical trials. *Computat Stat Data Anal*. 2009;53(3):812–9.
12. Southworth H, Heffernan JE. Multivariate extreme value modelling of laboratory safety data from clinical studies. *Pharm Stat*. 2012;11(5):367–72.
13. Cook RJ, Lawless J. The statistical analysis of recurrent events. New York: Springer; 2007.
14. Therneau TM, Grambsch PM. Modeling survival data: extending the cox model. New York: Springer; 2000.
15. Fu H, Luo J, Qu Y. Hypoglycemic events analysis via recurrent time-to-event (HEART) models. *J Biopharm Stat*. 2016;26(2):280–98.
16. Luo J, Qu Y. Analysis of hypoglycemic events using negative binomial models. *Pharm Stat*. 2013;12(4):233–42.
17. Hickey G, Philipson P, Jorgensen A, Kolamunnage-Dona R. A comparison of joint models for longitudinal and competing risks data, with application to an epilepsy drug randomized controlled trial. *J R Stat Soc Ser A (Stat Soc)*. 2018;181(4):1105–23.
18. Hickey GL, Philipson P, Jorgensen A, Kolamunnage-Dona R. Joint modeling of time-to-event and multivariate longitudinal outcomes: recent developments and issues. *BMC Med Res Methodol*. 2016;16(1):117.
19. Syrjäälä E, Nevalainen J, Peltonen J, Takkinen H-M, Hakola L, Åkerlund M, et al. A joint modeling approach for childhood meat, fish and egg consumption and the risk of advanced islet autoimmunity. *Sci Rep*. 2019;9(1):1–10.
20. Hickey GL, Philipson P, Jorgensen A, Kolamunnage-Dona R. joinerML: a joint model and software package for time-to-event and multivariate longitudinal outcomes. *BMC Med Res Methodol*. 2018;18(1):50.
21. Crowther MJ. merlin—a unified modeling framework for data analysis and methods development in Stata. *Stata J*. 2020;20(4):763–84.
22. Pekyi D, Ampromfi AA, Tinto H, Traore-Coulibaly M, Tahita MC, Valea I, et al. Four artemisinin-based treatments in African pregnant women with malaria. *New England J Med*. 2016;374(10):913–27.
23. Crowther MJ. Extended multivariate generalised linear and non-linear mixed effects models. 2017. <https://arxiv.org/abs/1710.02223>. Accessed 5 Dec 2020.
24. Tuerlinckx F, Rijnmen F, Verbeke G, De Boeck P. Statistical inference in generalized linear mixed models: a review. *Br J Math Stat Psychol*. 2006;59(2):225–55.
25. Henderson R, Diggle P, Dobson A. Joint modelling of longitudinal measurements and event time data. *Biostatistics*. 2000;1(4):465–80.
26. Ibrahim JG, Chu H, Chen LM. Basic concepts and methods for joint models of longitudinal and survival data. *J Clin Oncol*. 2010;28(16):2796.
27. Long JD, Mills JA. Joint modeling of multivariate longitudinal data and survival data in several observational studies of Huntington's disease. *BMC Med Res Methodol*. 2018;18(1):1–15.
28. Chi YY, Ibrahim JG. Joint models for multivariate longitudinal and multivariate survival data. *Biometrics*. 2006;62(2):432–45.
29. Andrinopoulou E-R, Clancy JP, Szczesniak R. Multivariate joint modeling to identify markers of growth and lung function decline that predict cystic fibrosis pulmonary exacerbation onset. *BMC Pulm Med*. 2020;20:1–11.
30. Buu A, Li R, Tan X, Zucker RA. Statistical models for longitudinal zero-inflated count data with applications to the substance abuse field. *Stat Med*. 2012;31(29):4074–86.
31. Li H, Zhang Y, Carroll RJ, Keadle SK, Sampson JN, Matthews CE. A joint modeling and estimation method for multivariate longitudinal data with mixed types of responses to analyze physical activity data generated by accelerometers. *Stat Med*. 2017;36(25):4028–40.
32. Yang Y, Kang J. Joint analysis of mixed Poisson and continuous longitudinal data with nonignorable missing values. *Computat Stat Data Anal*. 2010;54(1):193–207.
33. Phillips R, Cornelius V. Understanding current practice, identifying barriers and exploring priorities for adverse event analysis in randomised controlled trials: an online, cross-sectional survey of statisticians from academia and industry. *BMJ Open*. 2020;10(6):e036875.

Publisher's Note

Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Ready to submit your research? Choose BMC and benefit from:

- fast, convenient online submission
- thorough peer review by experienced researchers in your field
- rapid publication on acceptance
- support for research data, including large and complex data types
- gold Open Access which fosters wider collaboration and increased citations
- maximum visibility for your research: over 100M website views per year

At BMC, research is always in progress.

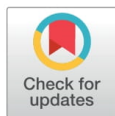
Learn more biomedcentral.com/submissions



RESEARCH ARTICLE

Effect of adverse events on non-adherence and study non-completion in malaria chemoprevention during pregnancy trial: A nested case control study

Noel Patson^{1,2*}, Mavuto Mukaka^{3,4}, Ingrid Peterson⁵, Titus Divala^{6,7}, Lawrence Kazembe⁸, Don Mathanga², Miriam K. Laufer⁵, Tobias Chirwa¹



1 School of Public Health, University of the Witwatersrand, Johannesburg, South Africa, **2** School of Public Health and Family Medicine, College of Medicine, University of Malawi, Blantyre, Malawi, **3** Mahidol Oxford Tropical Medicine Research Unit (MORU), Bangkok, Thailand, **4** Centre for Tropical Medicine, Nuffield Department of Medicine, University of Oxford, Oxford, United Kingdom, **5** Center for Vaccine Development and Global Health, University of Maryland School of Medicine, Baltimore, MD, United States of America, **6** TB Centre, London School of Hygiene and Tropical Medicine, London, United Kingdom, **7** Helse Nord Tuberculosis Initiative, College of Medicine, University of Malawi, Blantyre, Malawi, **8** Department of Biostatistics, University of Namibia, Windhoek, Namibia

* noelpatson@gmail.com

OPEN ACCESS

Citation: Patson N, Mukaka M, Peterson I, Divala T, Kazembe L, Mathanga D, et al. (2022) Effect of adverse events on non-adherence and study non-completion in malaria chemoprevention during pregnancy trial: A nested case control study. PLoS ONE 17(1): e0262797. <https://doi.org/10.1371/journal.pone.0262797>

Editor: Dafna Yahav, Rabin Medical Center, Beilinson Hospital, ISRAEL

Received: December 1, 2020

Accepted: October 18, 2021

Published: January 19, 2022

Copyright: This is an open access article, free of all copyright, and may be freely reproduced, distributed, transmitted, modified, built upon, or otherwise used by anyone for any lawful purpose. The work is made available under the [Creative Commons CC0](https://creativecommons.org/licenses/by/4.0/) public domain dedication.

Data Availability Statement: All relevant data and its supporting information files are within the manuscript. However, full access to the Chloroquine for Malaria in Pregnancy trial data is restricted and available upon request per ethical guidelines by Institutional Review Board at the University of Maryland and Malawi's College of Medicine Research and Ethics Committee. For researchers who meet the criteria for access to confidential data, application to access the data can be made to the Blantyre Malaria Project

Abstract

Background

In drug trials, adverse events (AEs) burden can induce treatment non-adherence or discontinuation. The non-adherence and discontinuation induce selection bias, affecting drug safety interpretation. Nested case-control (NCC) study can efficiently quantify the impact of the AEs, although choice of sampling approach is challenging. We investigated whether NCC study with incidence density sampling is more efficient than NCC with path sampling under conditional logistic or weighted Cox models in assessing the effect of AEs on treatment non-adherence and participation in preventive antimalarial drug during pregnancy trial.

Methods

Using data from a trial of medication to prevent malaria in pregnancy that randomized 600 women to receive chloroquine or sulfadoxine-pyrimethamine during pregnancy, we conducted a NCC study assessing the role of prospectively collected AEs, as exposure of interest, on treatment non-adherence and study non-completion. We compared estimates from NCC study with incidence density against those from NCC with path sampling under conditional logistic and weighted Cox models.

Results

Out of 599 women with the outcomes of interest, 474 (79%) experienced at least one AE before delivery. For conditional logistic model, the hazard ratio for the effect of AE occurrence on treatment non-adherence was 0.70 (95% CI: 0.42, 1.17; $p = 0.175$) under incidence density sampling and 0.68 (95% CI: 0.41, 1.13; $p = 0.137$) for path sampling. For

institutional Data Access Committee
(achavula@bmp.medcol.mw).

Funding: The clinical trial was supported by funding from NIH (grants U01AI087624 and K24AI114996 to MKL). NP was funded by the Fogarty International Center of the National Institutes of Health [Grant number D43TW010075 to MKL and DM]. The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

Competing interests: The authors have declared that no competing interests exist.

study non-completion, the hazard ratio was 1.02 (95% CI: 0.56, 1.83; $p = 0.955$) under incidence density sampling and 0.85 (95% CI: 0.45, 1.60; $p = 0.619$) under path sampling. We obtained similar hazard ratios and standard errors under incidence density sampling and path sampling whether weighted Cox or conditional logistic models were used.

Conclusion

NCC with incidence density sampling and NCC with path sampling are practically similar in efficiency whether conditional logistic or weighted Cox analytical methods although path sampling uses more unique controls to achieve the similar estimates.

Trial registration

ClinicalTrials.gov: [NCT01443130](https://clinicaltrials.gov/ct2/show/study/NCT01443130).

Introduction

Treatment non-adherence and treatment discontinuation, in which patients either do not take their medication as prescribed or do not take it at all, is a challenge in analysis and interpretation of randomized controlled trials results since it can lead to underestimation of treatment effects [1]. Adverse events (AEs) occurrence is one of the factors that influence treatment non-adherence and discontinuation. Therefore, treatment non-adherence and discontinuation can be considered as proxy outcomes for a measure of the impact of AEs on patients in clinical trials [2]. However, most trials descriptively report treatment non-adherence and discontinuation information in relation to AEs under safety and tolerability section and participant flow chart of the main study report [3–5]. Such kind of reporting and analysis is less informative since it does not provide complete estimates on how the AEs affect non-adherence over the follow-up time, after accounting for potential confounders e.g. patient age. Further, AEs that impact treatment non-adherence and study discontinuation outcomes precede these outcomes. Ignoring the fact that the AEs that impact treatment non-adherence and study non-completion outcomes occur before the observation of the treatment non-adherence or discontinuation can lead to underestimation of the effect of AEs [6, 7]. Establishing association between the AEs and treatment non-adherence or discontinuation is also both statistically and ethically challenging since the AE exposure cannot be randomized and it is difficult to pre-specify at design stage. Use of a nested case-control approach to understand the impact of AEs can be useful in this context since it derives a sample of cases and controls that are comparable, accounting for potential differences in follow-up time. However, the choice of an efficient nested case-control design and analytical method can present challenges. In order to address this challenge, empirical comparison of competing nested case-control designs is useful especially towards the choice of the most efficient nested case-control sampling and analytical approach.

Traditionally incidence density sampling (i.e. risk-set sampling) is used in most nested case-control studies. Under a nested case-control study with incidence density sampling design, each index case is matched to a specified number of controls sampled at the time of index case occurrence. The nested case-control study with incidence density sampling design is considered as an unbiased and statistically efficient approach to cohort sampling that accommodates causal interpretation of the observed effect of an exposure on an outcome [8]. The matching on time ensures that the cases are comparable to control by minimizing confounding that would arise due to the differences in follow-up time between the cases and controls.

Hence, the obtained estimates are unbiased and reflect the original trial cohort [9]. The selection of controls is done with replacement such that a subject could be selected as a control multiple times and may later become a case. However, it has also been argued that greater efficiency can also be achieved if the set of controls for different cases are as disjoint as possible. For instance, Langholz and Thomas proposed a modified alternative sampling approach that aims at ensuring that minimum number of controls are used more than once [10]. Under this second alternative nested case-control sampling approach, firstly a path set is constructed that includes all subjects who entered the study before t_i in time interval (t_{i-1}, t_i) and exited after time t_j in time interval (t_j, t_{j+1}) . Under this second sampling approach the specified number of controls are randomly sampled from the path set without replacement until all the controls in that path set are selected and the path set can be replenished. Once the desired number of controls is achieved the path sampling is stopped. As defined elsewhere, in this paper we refer to this variation of nested case-control with incidence density sampling design as nested case-control study with *path sampling* design [11]. Based on the provided descriptions of the nested case-control with incidence density sampling and nested case-control with path sampling, path sampling samples more unique patients than the traditional incidence density sampling. In practice, it is important to choose a nested case-control sampling design that can optimally estimate the effect of AEs on treatment non-adherence and study non-completion. However, there is limited work highlighting whether inclusion of more unique controls via path sampling offers an added advantage over incidence density sampling.

Conventionally, conditional logistic regression or Cox regression stratified on case-control sets are used in analysis of most nested case-control studies [12]. Recently inverse probability weighting (IPW) Cox models have been shown to have higher power and efficiency than conditional logistic regression analysing nested case-control studies [13–16]. Under IPW Cox models, individual log-likelihood is weighted by the inverse probability of being included in the nested case-control study. The weighting is can even offer more advantages of adapting the modelling, generally, beyond proportional hazards regression models [17]. Furthermore, the weighted Cox model also breaks the match to accommodate the use of cases and controls at all-time points when they are at risk.

Most of the studies discussed above have demonstrated the utility of the nested case-control with incidence density sampling under conditional logistic regression model and IPW models such as weighted Cox regression model. However, there is limited work demonstrating the utility and comparing the efficiency of the two alternative sampling designs (i.e. incidence density sampling and path sampling) under IPW Cox models and the traditional conditional logistic regression. In this paper, we investigated whether nested case-control with path sampling is more efficient than nested case-control study with incidence density sampling under conditional logistic regression model and weighted Cox regression model. We empirically compared nested case-control with incidence density sampling design against nested case-control with path sampling design under weighted Cox model and conditional logistic regression model, in investigating the effect of AEs on treatment non-adherence or study non-completion in the context of intermittent preventive treatment of malaria in pregnancy (IPTp) trial as case study. We hypothesized that path sampling design and weighted Cox regression will yield more efficient estimates with lower standard errors and AE exposure will increase the risk of non-adherence or discontinuation.

Methods

Motivating data description

To assess the effect of AEs on treatment non-adherence and study non-completion, we conducted a secondary data analysis of data from a clinical trial of preventive treatment for

malaria in pregnancy Malawi (NCT01443130) using nested case-control study approach. The trial was a randomised, open-label, controlled clinical trial evaluating whether chloroquine for intermittent preventive treatment to prevent malaria in pregnancy (IPTp-CQ) or chloroquine chemoprophylaxis was better than IPTp using sulfadoxine-pyrimethamine (IPTp-SP). The trial included women in first or second trimester before 27 weeks of gestational age, attending first antenatal visit and willing to remain in the study are until 14 weeks after delivery. The trial excluded women who were HIV positive, on any known chronic therapy with antimalarial or anti-folate drugs and had known high risk pregnancies.

Between February 2012 and May 2014 a total of 600 pregnant women in their first or second trimester were recruited and randomized to receive (a) sulfadoxine-pyrimethamine (SP) IPTp (two doses, four weeks apart), (b) chloroquine IPTp (600 mg on day 1, 600 mg on day 2, and 300 mg on day 3 two days apart), (see S1 Fig under supplementary materials). In the current analysis, we excluded a third arm consisting of 300 women randomized to receive weekly chloroquine prophylaxis since that arm had a radically different treatment schedule compared to the IPTp arms. The study treatments were administered under directly observed therapy. Detailed procedures, participant demographics and treatment effects for the trial are described in the primary trial manuscript [18]. The clinical trial protocol was registered with [ClinicalTrials.gov](https://clinicaltrials.gov) (NCT01443130).

Safety and tolerability assessments were based on unsolicited adverse events collected from the first day of the treatment administration until the final study visit. AEs were graded based on a standardized protocol [19], (i) graded on a four-point scale as, mild, moderate, severe, or life-threatening and (ii) further classified by System Organ Class category. The clinical trial was reviewed and approved by Institutional Review Board at the University of Maryland, the College of Medicine Research and Ethics Committee at the University of Malawi and the Malawi Pharmacy Medicines and Poisons Board. From every participant, a written informed consent was obtained just before undergoing any trial procedure. During the current analysis we only considered follow-up time from enrolment up to delivery because medication was administered until delivery.

Nested case-control study designs

We considered two types of nested case-control study designs; (a) Nested case-control study with risk-set sampling (i.e. incidence density sampling) (b) Nested case-control study with path sampling. The samples for the two designs were derived from the cohort $\zeta = \{i, \dots, n\}$ of women who had taken at least their first dose of IPTp administered in the trial. Both nested case-control designs were based on two study outcomes: (i) treatment non-adherence and (ii) study non-completion. However, if the only outcome of interest (e.g. treatment non-adherence or study non-completion) occurred only at enrolment time ($t = 0$), the woman was considered ineligible for the nested case-control study since matching was done on follow-up time ($t > 0$). For both designs we identified same set of baseline characteristics as potential confounders. The potential confounders included maternal age in years, body mass index (BMI; calculated as weight in kilograms divided by height in metres squared), gravidity (primigravid or not) and, IPTp treatment arm.

Nested case-control with incidence density sampling. Definition of a case was based on experiencing an outcome of interest as defined in the preceding section. Firstly, we defined cases of *treatment non-adherence* as those missing at least one dose of treatment after the first dose. Although there was a possibility of observing more than one non-adherence events per woman, we focused on time to the first treatment non-adherence after enrolment. Secondly, we defined *study non-completion* as discontinuing participation in the study before delivery.

The controls were defined as women who had not yet experienced the outcome interest (i.e. *treatment non-adherence* or *study non-completion*) but were at risk to become cases. For each treatment non-adherence and study non-completion cases, we randomly selected four controls using risk set sampling (i.e. incidence density sampling). The selection of the controls (and corresponding analyses) was done separately for each endpoint. Each case was matched on follow-up time since enrolment by selecting controls from the set of women in the clinical trial cohort who had not experienced the outcome up until the time that the case arose; controls were selected without regard to outcomes that might occur after the time of matching. Hence a control could be sampled multiple times and could later become a case. We considered the selection of four controls per case as sufficient since there is negligible gain in statistical efficiency when more than four controls are selected per case [20–22].

The exposure in this analysis was having an AE before selection in risk-set sampling. We dichotomized the AEs as (having an AE versus not having one before the index case time). Due to the exploratory nature of our study in demonstrating the efficiency of incidence density sampling over path sampling under different analytical methods, we considered only any AE exposure regardless of the relatedness to the treatment or severity.

Nested case-control with path sampling. The definition of cases and exposure are as in the incidence density sampling above. Every aspect for the path sampling approach is similar to the standard nested case-control with incidence density sampling except that this approach (i.e. path sampling) attempted to minimize the number of times that a control could be selected more than once. This was achieved using the path set concept where we sampled a path set according to the probabilities given the path set representation in the entire risk set at each event time [10]. As highlighted in the introduction section above, Langholz and Thomas indicate that the path set Q_{ij} consist of all subjects who entered the study before t_i in time interval (t_{i-1}, t_i) and exited after time t_j in time interval (t_j, t_{j+1}) . In our context, the path set consisted of women who enrolled before t_i and exited the trial after t_j in the respective time intervals. After the path set was defined and sampled, the controls were randomly sampled from the selected path set until the whole path set was emptied when the set can be replenished (i.e. once a woman was selected as a control from the path set, she could not be selected again until all women in that path set were selected). We implemented this sampling design using the recently illustrated implementation of the design in R statistical software [11]. This sampling approach yields as unique number of patients as possible included in the nested case-control study. Therefore, path sampling attempts to use more information from the controls by sampling more unique controls from the cohort. As the case with nested case-control with incidence density sampling, the selection of the controls (and corresponding analyses) was done separately for each endpoint.

Statistical analysis

We conducted all the data management and statistical analyses in Stata version 15 (StataCorp, College Station, TX, USA) and R Statistical software packages.

Descriptive analysis. Firstly, we summarized the baseline characteristics of the study participants by case-control status. For continuous variables, we reported mean (standard deviation [SD]) or median (inter-quartile range [IQR]) and proportion for categorical variables. We also summarized the distribution of AE occurrence within respective treatment arms based on AE grade and frequency. We plotted Kaplan Meier graph to characterise time to first AE occurrence between the IPTp arms.

Modelling the effect of AEs on non-adherence and study non-completion. Let likelihood for the nested case-control for parameter β be $L_{ncc}(\beta)$. Traditionally, estimates in nested

case-control studies are derived by maximizing partial likelihood function that can be written as

$$L_{ncc}(\beta) = \prod_{i \in \Omega} \frac{\exp(\beta' X_i)}{\sum_{j \in R_i} \exp(\beta' X_j)} \quad (\text{M1})$$

Where $L_{ncc}(\beta)$ is likelihood for nested case-control for parameter β . The β represents a vector of the regression coefficients, such that β' is a transposed vector corresponding to vector(s) of covariates. The X_i and X_j are the vectors of covariates for the individuals i and j . The sampled risk set R_i consists of case i and sampled m controls matched to the case on given time t since enrolment. The Ω represents the set of all the cases. The conditional likelihood under logistic regression coincides with Cox regression partial likelihood [23]. Therefore, the analysis of nested case-control study using Cox regression model or conditional logistic stratified on case-control set yield similar results. In this study we firstly considered, conditional logistic models whose results are always the same as the traditional stratified Cox regression model (i.e. we stratified the model by case-control set indicator δ (see the equation M2 below)). The baseline hazard for the stratum at follow-up time t was considered as $h_{0\delta}(t)$.

$$h_i(t) = h_{0\delta}(t) \exp(\beta' X_i) \quad (\text{M2})$$

As already defined, the partial likelihood that is maximized to obtain the estimate for the regression coefficient β is similar to M1;

$$L_{cox}(\beta) = \prod_{i \in \Omega} \frac{\exp(\beta' X_i)}{\sum_{j \in R_i} \exp(\beta' X_j)} \quad (\text{M3})$$

Here under M3, R_i is a set of all patients who have not failed or been censored but are at risk of failure at that time (i.e. all patients who are still under observation just before time t). The traditional analysis above, under M1 uses only the information from the sampled cases and controls such that the nested case-control study can sometimes be considered as a biased sample from the cohort. Weighting the partial likelihood with the inverse of the inclusion probability of being sampled mitigates this problem. The weighted Cox model partial likelihood can be written as

$$L_{wcox}(\beta) = \prod_{i \in \Omega} \frac{\exp(\beta' X_i) \omega_i}{\sum_{j \in S_i} \exp(\beta' X_j) \omega_j} \quad (\text{M4})$$

Where S_j is the set of cases and controls in the nested case-control sample who are at risk at time t_i . The Cox model obtained from the maximizing this likelihood (with inverse probability weights) to estimate the coefficient is defined as weighted Cox regression model. The weights were defined as $\omega_j = 1/p_j$, where p_j is probability of a woman being in the nested case-control sample. The $p_j = 1$ for all the cases. In this paper, we used Kaplan Meier (KM) method inclusion probabilities. Inclusion probabilities defined by KM method are called design based since they estimate actual sampling of the controls [24]. The R package multipleNCC [25], was used to compute the weights.

We considered both conditional logistic model and weighted Cox model in our analysis. Overall, we fitted both the conditional logistic model (conditioned on risk sets) and weighted Cox regression to both the nested case-control sample with incidence density sampling and then to the nested case-control sample with path sampling This was done separately for non-adherence and study non-completion outcomes.

In building each model, Firstly, we fitted univariate models to assess the unadjusted effects of the covariates. Then we fitted multivariable model that included the AE exposure and IPTp treatment adjusted for similar set of confounders that were significant in the univariate analysis at 0.25. Using the likelihood ratio test, we also assessed whether inclusion of interaction between the AE occurrence and treatment arm modified the observed effect. All the point estimates from the fitted models are presented with their respective 95% confidence interval. The interpretation of all the final models was made at 0.05 significance level.

Results

The cohort where our nested case-control sample designs were derived considered 599 women who were assigned to the two arms of the trial. The women in the cohort had a median follow-up time of 116 days (IQR; 97, 130). At enrolment, these women had a mean gestational age of 22.1 (SD = 2.1) weeks and a mean haemoglobin concentration of 11.7 (SD = 1.3) g/dL (Table 1). A total of 1545 AEs occurred among 474 (79.0%) study participants over the follow-up time where women in the IPTp-CQ arm had a higher cumulative incidence of AEs (895 AEs) than the women in the IPTp-SP arm (650 AEs). Time to first AE was consistently higher in IPTp-CQ arm compared to IPTp-SP over the follow-up time (Fig 1) such that women in the IPTp-CQ arm were 29% more likely to experience at least one AE compared to women in the IPTp-SP arm (HR: 1.29; 95% CI: 1.07, 1.57; $p = 0.008$). Further details of the AE distribution are provided under supplementary S1 and S2 Tables.

For the study non-completion outcome, 64 (10.7%) women did not complete the study, with 37 (12.3%) in the IPTp-SP arm and 27(9.0%) in the IPTp-CQ arm. Based on the matching criteria defined in the methods section (matching each case to four controls), we found a total of 256 study non-completion controls. For the treatment non-adherence outcome, 106 (17.7%) women were non-adherent to trial medication, at least once during study follow-up, with 40(13.3%) in the IPTp-SP arm and 66 (22%) in the IPTp-CQ arm. After each incident case was matched to four controls, we yielded a total of 424 treatment non-adherence controls. These reported results of matching were obtained separately for the nested case-control with incidence density sampling and the nested case-control with path sampling.

Comparison of parameter estimates

We report the hazard ratio (HR) estimates on exponentiated scale (not logarithm scale), in Tables 2 and 3. However, the standard errors are reported on logarithm scale (logHR SE), based on the corresponding log hazard ratios. The nested nested-control is abbreviated as

Table 1. Distribution of the study cohort characteristics (n = 600).

Characteristic	Mean (SD) or n (%)
Maternal age(years)	21.1 (3.2)
Gestational age(weeks)	22.1 (2.1)
BMI (kg/m ²)	23.7 (3.1)
Haemoglobin(g/dl)	11.7 (1.3)
Bed net use n (%)	450 (75.0)
Non-Primigravid n (%)	258 (43.1)
AE occurrence* n (%)	474 (79.0)
Treatment n (%)	
IPTp-SP	300 (50.0)
IPTp-CQ	300 (50.0)

<https://doi.org/10.1371/journal.pone.0262797.t001>

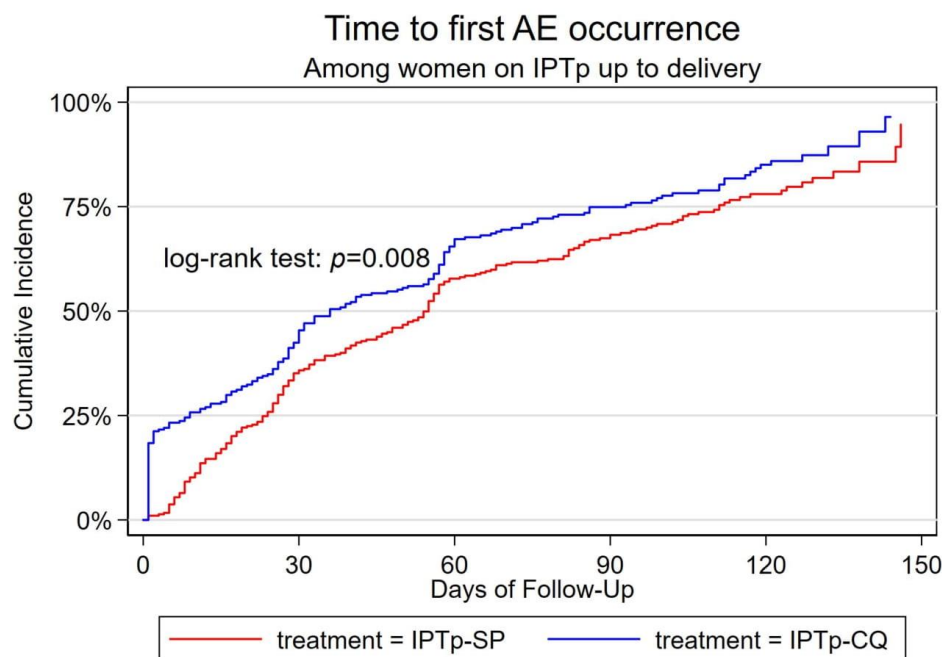


Fig 1. Time to first AE occurrence among the pregnant women on IPTp up to delivery.

<https://doi.org/10.1371/journal.pone.0262797.g001>

NCC in the Tables 2 and 3. Focusing on the effect of AEs on both treatment non-adherence and study non-completion, our estimates (Tables 2 and 3) consistently indicate that path sampling yielded similar hazard ratio estimates compared to those from the incidence density sampling. Comparing the weighted Cox regression model and conditional logistic regression model, the weighted Cox model yielded similar standard errors compared to those from

Table 2. Association between AE occurrence and treatment non-adherence; comparison of estimates from risk set sampling and path sampling under multivariable conditional logistic regression model and weighted Cox regression model.

Characteristic	Conditional logistic regression model						Weighted Cox regression model					
	NCC with risk set sampling			NCC with Path sampling			NCC with risk set sampling			NCC with Path sampling		
	HR (95% CI)	LogHR SE	P-value	HR (95% CI)	LogHR SE	P-value	HR (95% CI)	LogHR SE	P-value	HR (95% CI)	LogHR SE	P-value
Treatment	Reference											
IPTp-SP												
IPTp-CQ	2.07 (1.32, 3.22)	0.23	0.001	2.05 (1.30, 3.23)	0.23	0.002	2.04 (1.31, 3.16)	0.22	0.001	2.08 (1.32, 3.27)	0.23	0.002
AE occurrence	0.70 (0.42, 1.17)	0.26	0.175	0.68 (0.41, 1.13)	0.26	0.137	0.67 (0.43, 1.00)	0.22	0.051	0.68 (0.44, 1.06)	0.23	0.087
Age	0.96 (0.89, 1.02)	0.03	0.190	0.94 (0.88, 1.01)	0.04	0.105	0.96 (0.90, 1.02)	0.03	0.179	0.95 (0.89, 1.02)	0.03	0.125

LogHR SE represents standard errors on logarithm scale of hazard ratio, NCC represents nested case-control

<https://doi.org/10.1371/journal.pone.0262797.t002>

Table 3. Association between AE occurrence and study non-completion; comparison of estimates from risk set sampling and path sampling under multivariable conditional logistic regression model and weighted Cox regression model.

	Conditional logistic regression model						Weighted Cox regression model					
	NCC with risk set sampling			NCC with Path sampling			NCC with risk set sampling			NCC with Path sampling		
	HR (95% CI)	LogHR SE	P	HR (95% CI)	LogHR SE	P	HR (95% CI)	RobSE	P	HR (95% CI)	robSE	P
AE occurrence	1.02 (0.56, 1.83)	0.30	0.955	0.85 (0.45, 1.60)	0.32	0.619	0.88 (0.51, 1.52)	0.28	0.648	0.82 (0.47, 1.43)	0.29	0.477
Treatment	0.96 (0.55, 1.67)	0.28	0.896	0.68 (0.39, 1.18)	0.28	0.173	0.96 (0.55, 1.69)	0.29	0.889	0.67 (0.38, 1.18)	0.29	0.162
Age	0.92 (0.83, 1.02)	0.05	0.104	0.93 (0.84, 1.01)	0.05	0.098	0.93 (0.85, 1.01)	0.04	0.095	0.94 (0.87, 1.01)	0.04	0.072

LogHR SE represents standard errors on logarithm scale of hazard ratio, NCC represents nested case-control.

<https://doi.org/10.1371/journal.pone.0262797.t003>

conditional logistic regression across both study designs and both outcomes. Therefore, overall, we obtained similar estimates of the effect of AEs on study non-completion and treatment non-adherence under both incidence density sampling and path sampling regardless of the analysis method used.

We observed no statistically significant effect of AE occurrence on treatment non-adherence under both incidence density sampling and path sampling regardless of the method of analysis used. Using conditional logistic regression, the effect of AE occurrence on treatment non-adherence hazard ratio was 0.70 (95% CI: 0.42, 1.17; $p = 0.175$) under incidence density sampling and under path sampling the hazard ratio was 0.68 (95% CI: 0.41, 1.13; $p = 0.137$). Using the weighted Cox regression, we observed a negligible departure from the overall trend where the path sampling yielded a slightly higher estimate of hazard ratio (HR: 0.68; 95% CI: 0.44, 1.06), compared to the incidence density sampling (HR: 0.67; 95% CI: 0.43, 1.00). The negligible exception could be explained as one of the expected scenarios that can arise due to sampling variation.

Similarly, we observed no effect of AE occurrence on study non-completion under both incidence density sampling and path sampling regardless of the method of analysis used (see Table 3). Using conditional logistic regression, the effect of AE occurrence on study non-completion hazard ratio was 1.02 (95% CI: 0.56, 1.83; $p = 0.955$) under incidence density sampling and under path sampling the hazard ratio was 0.85 (95% CI: 0.45, 1.60; $p = 0.619$). As expected according to the overall trend, weighted Cox regression model yielded hazard ratio of 0.88 (95% CI: 0.51, 1.52, $p = 0.648$) under incidence density sampling and 0.82 (95% CI: 0.47, 1.43, $p = 0.477$) under path sampling.

Among the adjusted potential confounders for the treatment non-adherence outcome, only treatment was associated with treatment non-adherence (Table 2). We observed two times increased risk of treatment non-adherence for patient in IPTp-CQ treatment compared those in IPTp-SP treatment. Using conditional logistic regression, the observed hazard ratio was 2.07 (95% CI: 1.32, 3.22; $p = 0.001$) under incidence density sampling and we observed a similar hazard ratio (HR: 2.05; 95% CI: 1.30, 3.23; $p = 0.002$) under path sampling. Weighted Cox regression model yielded similar hazard ratios; (HR: 2.04; 95% CI: 1.31, 3.16; $p = 0.001$) under incidence density sampling and (HR: 2.08; 95% CI: 1.32, 3.27; $p = 0.002$) under path sampling.

Discussion

In this study, we empirically compared the efficiency of incidence density sampling and path sampling for nested case-control studies under conditional logistic regression and weighted Cox regression analysis methods. The practical application focussed on determining the impact of AEs on treatment non-adherence or discontinuation. Practically, we established that

the standard errors and hazard ratios for the two designs were similar. However, it should be noted that path sampling (by design) is based on more unique controls selected [11] implying that it naturally uses a larger sample of unique patients to achieve those similar standard errors as compared to incidence density sampling. This knowledge gained is directly useful in improving nested case-control studies that have recently gained an interest in safety studies [26, 27]. This study provides insights regarding choice of nested case-control design and the burden of AEs on treatment non-adherence and study non-completion in IPTp trials. Such insights may also facilitate understanding of expected patient safety in the post-trial period when the drug is launched [28].

The negligible gain in efficiency of weighted Cox regression model over conditional regression can be attributed to the homogeneity of the cohort that we used. Since our cohort comes from a randomized clinical trial whose baseline characteristics are similar, breaking the match in order to use more controls could not add much information to our nested case-control sample that was already using over 50% of the cohort sample. Most of the substantial gains for the inverse probability weighting have been established under large sample properties [24], especially in observational studies which by design exhibit heterogeneous population unlike clinical trials.

From practical perspective, based on the example data that we used, this study did not establish a significant role of AEs on IPTp treatment non-adherence and study discontinuation. Neither, conditional logistic regression model nor the weighted Cox regression model under incidence density sampling or path sampling suggested that AE occurrence impacted treatment non-adherence or study non-completion. Our findings are contrary to the hypothesized direction of AEs effect where we postulated that AE occurrence increases the risk of IPTp non-adherence and IPTp trial discontinuation. However, our findings support what has been previously observed in IPTp context whereby frequent adverse events do not reduce adherence [29]. Our exploratory finding is useful towards ascertaining the validity of IPTp trial designs and interpretation of findings.

We observed increased risk of treatment non-adherence for the women in IPT-CQ treatment arm who were two times more likely to non-adhere compared to those in IPTp-SP arm. This is expected since the women in IPTp-CQ arm had more opportunities for non-adherence as they took two more doses compared to those on IPTp-SP arm. This agrees with previous studies that indicate that number of doses for a given regimen is inversely associated with treatment non-adherence [30]; increasing number of days and doses for taking a particular regimen can lead to low treatment adherence [31, 32]. This has consequences on the IPTp trial designs and packaging of treatment doses.

Use of the nested case-control study design in the context of clinical trial data provided a great opportunity to assess the temporal relationship of adverse events and adherence or study non-completion [33]. This further enabled us to make causal interpretations since the time-dependent AE exposure was prospectively measured and the nested case-control sampling approach reduces selection bias. Although Cox proportional hazards model with time-dependent covariates can alternatively yield similar estimates to those obtained through nested case-control approach, the nested case control is known to be more computationally efficient [34]. Despite the fact that our work may not be generalizable to settings where there are non-repeated doses, it still offers a groundwork for improved antimalarial drug safety assessment framework (accounting for non-adherence). Such improved assessment of the relationship between non-adherence to treatment and safety outcomes is crucial in planning for improved interventions effectiveness [35].

This study had some limitations. Firstly, the differences in treatment by dose schedule between the treatment arms could affect the observed differences. Secondly, we did not collect

data on patient satisfaction. This made it hard it to make a conclusive statement on the unexpected observed effect of the AEs. Since non-adherence can be affected by the patient behaviour, collection of such patient satisfaction and perception in clinical trials would be helpful in elucidating the effect of AEs on non-adherence. Furthermore, the methodological conclusions in this study can be improved and extended by undertaking a full simulation study to understand the performance of the sampling and analysis methods under varying known scenarios.

Conclusion

Assessment of safety in relation to IPTp adherence, even for less severe AEs, is an important topic in developing solutions towards improved IPTp safety profile and risk management plan in clinical trials. Nested case-control study can be considered as a useful tool in investigating the impact of AEs on treatment non-adherence and study non-completion because it offers an opportunity for an efficient sample where cases and controls are comparable, accounting for potential differences in follow-up time. We empirically demonstrated that incidence density sampling and path sampling practically yield similar estimates whether one uses weighted Cox regression or conditional logistic regression. However, by design, path sampling yields those similar estimates at the expense of more unique controls sampled compared to the incidence density sampling. The clinical findings indicated that AE occurrence did not impact treatment non-adherence or study non-completion. The insights from this study inform future trial designs and drug safety assessment. Future extension of this work can consider evaluating the impact of AEs on treatment non-adherence and study non-adherence accounting for patient perception on IPTp benefits since treatment non-adherence is associated with patient behaviour.

Supporting information

S1 Table. Distribution of SOC of AEs before delivery by treatment group.
(DOCX)

S2 Table. Distribution of adverse events by amongst the pregnant women on IPTp treatment.
(DOCX)

S1 Fig. Treatment schedule over a follow-up time for women enrolled to three malaria chemoprevention treatments in Malawi.
(TIF)

Acknowledgments

We thank all the participants and data collectors in this clinical trial.

Author Contributions

Conceptualization: Noel Patson, Mavuto Mukaka, Ingrid Peterson, Titus Divala, Lawrence Kazembe, Don Mathanga, Miriam K. Laufer, Tobias Chirwa.

Data curation: Noel Patson, Titus Divala, Miriam K. Laufer.

Formal analysis: Noel Patson.

Funding acquisition: Don Mathanga, Miriam K. Laufer, Tobias Chirwa.

Investigation: Noel Patson, Titus Divala, Miriam K. Laufer.

Methodology: Noel Patson, Mavuto Mukaka, Ingrid Peterson, Titus Divala, Lawrence Kazembe, Don Mathanga, Miriam K. Laufer, Tobias Chirwa.

Project administration: Mavuto Mukaka, Ingrid Peterson.

Software: Noel Patson, Mavuto Mukaka.

Supervision: Mavuto Mukaka, Ingrid Peterson, Titus Divala, Lawrence Kazembe, Don Mathanga, Miriam K. Laufer, Tobias Chirwa.

Writing – original draft: Noel Patson.

Writing – review & editing: Noel Patson, Mavuto Mukaka, Ingrid Peterson, Titus Divala, Lawrence Kazembe, Don Mathanga, Miriam K. Laufer, Tobias Chirwa.

References

1. Vrijens B, De Geest S, Hughes DA, Przemyslaw K, Demonceau J, Ruppar T, et al. A new taxonomy for describing and defining adherence to medications. *Br J Clin Pharmacol*. 2012; 73(5):691–705. Epub 2012/04/11. <https://doi.org/10.1111/j.1365-2125.2012.04167.x> PMID: 22486599
2. Phillips R, Hazell L, Sauzet O, Cornelius V. Analysis and reporting of adverse events in randomised controlled trials: a review. *BMJ open*. 2019; 9(2):e024537. <https://doi.org/10.1136/bmjopen-2018-024537> PMID: 30826796
3. Gonzalez R, Mombongo-Ngoma G, Ouedraogo S, Kakolwa MA, Abdulla S, Accrombessi M. Intermittent preventive treatment of malaria in pregnancy with mefloquine in HIV-negative women: a multicentre randomized controlled trial. *PLoS medicine*. 2014; 11. <https://doi.org/10.1371/journal.pmed.1001733> PMID: 25247709
4. Clerk CA, Bruce J, Affingungh PK, Mensah N, Hodgson A, Greenwood B, et al. A Randomized, Controlled Trial of Intermittent Preventive Treatment with Sulfadoxine-Pyrimethamine, Amodiaquine, or the Combination in Pregnant Women in Ghana. *The Journal of infectious diseases*. 2008; 198(8):1202–11. <https://doi.org/10.1086/591944> PMID: 18752443
5. Patson N, Mukaka M, Otumweme KN, Kazembe L, Mathanga DP, Mwapasa V, et al. Systematic review of statistical methods for safety data in malaria chemoprevention in pregnancy trials. *Malaria Journal*. 2020; 19(1):119. <https://doi.org/10.1186/s12936-020-03190-z> PMID: 32197619
6. Lange T, Keiding N. Skin cancer as a marker of sun exposure: a case of serious immortality bias. *International journal of epidemiology*. 2014; 43(3):971–. <https://doi.org/10.1093/ije/dyu100> PMID: 24894096
7. Beyersmann J, Wolkewitz M, Schumacher M. The impact of time-dependent bias in proportional hazards modelling. *Statistics in medicine*. 2008; 27(30):6439–54. <https://doi.org/10.1002/sim.3437> PMID: 18837068
8. Keogh RH, Cox DR. *Case-Control Studies*. Cambridge: Cambridge University Press; 2014.
9. Pearce N. What does the odds ratio estimate in a case-control study? *International journal of epidemiology*. 1993; 22(6):1189–92. <https://doi.org/10.1093/ije/22.6.1189> PMID: 8144304
10. Langholz B, Thomas DC. Efficiency of cohort sampling designs: some surprising results. *Biometrics*. 1991; 47(4):1563–71. PMID: 1786329
11. Nuño MM, Gillen DL. Chapter 5—Alternative Sampling Designs for Time-to-Event Data With Applications to Biomarker Discovery in Alzheimer's Disease. In: Srinivasa Rao ASR, Pyne S, Rao CR, editors. *Handbook of Statistics*. 36: Elsevier; 2017. p. 105–66.
12. Kim RS, Kaplan RC. Analysis of secondary outcomes in nested case-control study designs. *Stat Med*. 2014; 33(24):4215–26. Epub 2014/06/13. <https://doi.org/10.1002/sim.6231> PMID: 24919979
13. Steyer NC, Samuelsen SO. Inverse probability weighting in nested case-control studies with additional matching—a simulation study. *Stat Med*. 2013; 32(30):5328–39. Epub 2013/10/18. <https://doi.org/10.1002/sim.6019> PMID: 24132909
14. Kim RS. A new comparison of nested case-control and case-cohort designs and methods. *European journal of epidemiology*. 2015; 30(3):197–207. Epub 2014/12/03. <https://doi.org/10.1007/s10654-014-9974-4> PMID: 25446306
15. Borgan Ø, Keogh R. Nested case-control studies: should one break the matching? *Lifetime Data Anal*. 2015; 21(4):517–41. Epub 2015/01/23. <https://doi.org/10.1007/s10985-015-9319-y> PMID: 25608704
16. Samuelsen SO. A pseudolikelihood approach to analysis of nested case-control studies. *Biometrika*. 1997; 84(2):379–94.

17. Borgan Ø, Breslow N, Chatterjee N, Gail MH, Scott A, Wild CJ. Handbook of statistical methods for case-control studies: CRC Press; 2018.
18. Divala TH, Mungwira RG, Mawindo PM, Nyirenda OM, Kanjala M, Ndaferankhande M, et al. Chloroquine as weekly chemoprophylaxis or intermittent treatment to prevent malaria in pregnancy in Malawi: a randomised controlled trial. *Lancet Infect Dis*. 2018; 18(10):1097–107. [https://doi.org/10.1016/S1473-3099\(18\)30415-8](https://doi.org/10.1016/S1473-3099(18)30415-8) PMID: 30195996
19. Division of Microbiology and Infectious Diseases (DMID) adult toxicity tables: National Institute of Allergy and Infectious Diseases; 2007 [cited 2019 October 30]. Available from: <https://www.niaid.nih.gov/sites/default/files/dmidadulttox.pdf>.
20. Cai T, Zheng Y. Evaluating prognostic accuracy of biomarkers in nested case-control studies. *Biostatistics*. 2012; 13(1):89–100. Epub 08/19. <https://doi.org/10.1093/biostatistics/kxr021> PMID: 21856652
21. Taylor JMG. Choosing the number of controls in a matched case-control study, some sample size, power and efficiency considerations. *Statistics in medicine*. 1986; 5(1):29–36. <https://doi.org/10.1002/sim.4780050106> PMID: 3961313
22. Gail M, Williams R, Byar DP, Brown C. How many controls? *Journal of chronic diseases*. 1976; 29(11):723–31. Epub 1976/11/01. [https://doi.org/10.1016/0021-9681\(76\)90073-4](https://doi.org/10.1016/0021-9681(76)90073-4) PMID: 791962
23. Prentice R, Breslow N. Retrospective studies and failure time models. *Biometrika*. 1978; 65(1):153–8.
24. Samuelsen SO, Støer N. Inverse probability weighting in nested case-control studies. *Handbook of Statistical Methods for Case-Control Studies*. 2018:351–72.
25. Støer NC, Samuelsen SO. multipleNCC: Inverse Probability Weighting of Nested Case-Control Data. *R J*. 2016; 8(2):5.
26. Ann Q, Tham R. Incidence Density Sampling for Nested Case-Control Study Designs. 2019.
27. Kiri VA. Programming challenges of sampling controls to cases from the dynamic risk sets in nested case-control studies. *Pharmaceutical Programming*. 2012; 5(1–2):29–33.
28. Smith DL. Patient Nonadherence in Clinical Trials: Could There Be a Link to Postmarketing Patient Safety? *Drug Information Journal*. 2012; 46(1):27–34. <https://doi.org/10.1177/0092861511428300>
29. Denoeud-Ndam L, Clément MC, Briand V, Akakpo J, Agossou VK, Atadokpédé F, et al. Tolerability of mefloquine intermittent preventive treatment for malaria in HIV-infected pregnant women in Benin. *Journal of acquired immune deficiency syndromes (1999)*. 2012; 61(1):64–72. Epub 2012/06/19. <https://doi.org/10.1097/QAI.0b013e3182615a58> PMID: 22706291
30. Claxton AJ, Cramer J, Pierce C. A systematic review of the associations between dose regimens and medication compliance. *Clinical therapeutics*. 2001; 23(8):1296–310. Epub 2001/09/18. [https://doi.org/10.1016/s0149-2918\(01\)80109-0](https://doi.org/10.1016/s0149-2918(01)80109-0) PMID: 11558866
31. Desai M, Hill J, Fernandes S, Walker P, Pell C, Gutman J, et al. Prevention of malaria in pregnancy. *The Lancet Infectious diseases*. 2018; 18(4):e119–e32. Epub 2018/02/06. [https://doi.org/10.1016/S1473-3099\(18\)30064-1](https://doi.org/10.1016/S1473-3099(18)30064-1) PMID: 29395997
32. Hill J, Hoyt J, Achieng F, Ouma P, L'Lanziva A, Kariuki S, et al. User and Provider Acceptability of Intermittent Screening and Treatment and Intermittent Preventive Treatment with Dihydroartemisinin-Piperaquine to Prevent Malaria in Pregnancy in Western Kenya. *PloS one*. 2016; 11(3):e0150259-e. <https://doi.org/10.1371/journal.pone.0150259> PMID: 26986471
33. Etminan M. Pharmacoepidemiology II: The Nested Case-Control Study—A Novel Approach in Pharmacoepidemiologic Research. *Pharmacotherapy: The Journal of Human Pharmacology and Drug Therapy*. 2004; 24(9):1105–9. <https://doi.org/10.1592/phco.24.13.1105.38083> PMID: 15460170
34. Essebag V, Platt RW, Abrahamowicz M, Pilote L. Comparison of nested case-control and survival analysis methodologies for analysis of time-dependent exposure. *BMC Medical Research Methodology*. 2005; 5(1):5. <https://doi.org/10.1186/1471-2288-5-5> PMID: 15670334
35. Leporini C, De Sarro G, Russo E. Adherence to therapy and adverse drug reactions: is there a link? Expert opinion on drug safety. 2014; 13 Suppl 1:S41–55. Epub 2014/08/30. <https://doi.org/10.1517/14740338.2014.947260> PMID: 25171158

Appendix 3: Ethical clearance

UNIVERSITY OF THE
WITWATERSRAND,
JOHANNESBURG



R14/49 Mr Noel Patson

HUMAN RESEARCH ETHICS COMMITTEE (MEDICAL)

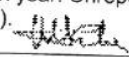
CLEARANCE CERTIFICATE NO. M180987

NAME: Mr Noel Patson
(Principal Investigator)
DEPARTMENT: School of Public Health
PROJECT TITLE: Multivariate longitudinal analysis of drug safety
in clinical trials: case of antimalarials in pregnancy in
Malawi
DATE CONSIDERED: 28//09/2018
DECISION: Approved
CONDITIONS: Providing ethics clearance from Malawi Ethics Committee.
SUPERVISOR: Prof Tobias Chirwa
APPROVED BY: 
Dr C Penny, Chairperson, HREC (Medical)
DATE OF APPROVAL: 13/11/2018

This clearance certificate is valid for 5 years from date of approval. Extension may be applied for.

DECLARATION OF INVESTIGATORS

To be completed in duplicate and **ONE COPY** returned to the Research Office Secretary in Room 301, Third floor, Faculty of Health Sciences, Phillip Tobias Building, 29 Princess of Wales Terrace, Parktown, 2193, University of the Witwatersrand. I/we fully understand the conditions under which I am/we are authorized to carry out the above-mentioned research and I/we undertake to ensure compliance with these conditions. Should any departure be contemplated, from the research protocol as approved, I/we undertake to resubmit the application to the Committee. **I agree to submit a yearly progress report.** The date for annual re-certification will be one year after the date of convened meeting where the study was initially reviewed. In this case, the study was initially reviewed September and will therefore be due in the month of September each year. Unreported changes to the application may invalidate the clearance given by the HREC (Medical).


Principal Investigator Signature

20/11/2018
Date

PLEASE QUOTE THE PROTOCOL NUMBER IN ALL ENQUIRIES

