Spatial and Temporal Analyses of Sulphadoxine Pyrimethamine Resistance in African *Plasmodium falciparum* malaria

Inbarani Naidoo

Thesis submitted in accordance with the requirements for the degree of

**Doctor of Philosophy of the University of London**

**March 2015**

Department of Pathogen Molecular Biology
Faculty of Infectious and Tropical Diseases

**LONDON SCHOOL OF HYGIENE & TROPICAL MEDICINE**

Funded by Gates Malaria Partnership
Declaration

I, Inbarani Naidoo confirm that the work presented in this thesis is my own, completed under the supervision of Dr Cally Roper. Where information has been derived from other sources, I confirm that this has been indicated in the thesis.

Signed ___________________________ Date ____________

[Signature]
Abstract

Sulphadoxine pyrimethamine resistance (SPR) emerged soon after SP was introduced as first line therapy for malaria in Africa during the 1990s. This thesis presents the first attempt to describe the spatio-temporal distribution of SPR in sub-Saharan Africa. Molecular and in vivo SPR data were gathered from primarily published sources onto a handwritten and digital template, recording the precise geo-location of every data point. The most commonly used methods were the World Health Organisation 1973 and 1996 protocols to measure in vivo efficacy and PCR-RFLP for molecular studies. Consistent data gaps with both SPR measures were found in Botswana, Burundi, Cape Verde, Eritrea, Somalia, Togo and Mauritius (chapter two). Three broad categories of molecular resistance emerged, reflecting the distribution patterns of dhfr and dhps point mutation prevalence: (1) scarce partially resistant mutations namely dhps 436A and 613S which are not increasing significantly over time, (2) emerging resistance mutations such as dhfr 164L and dhps 581G which are rare but increasing within distinct geographical foci and (3) major resistance mutations namely dhps 437G and 540E, which are widespread with significantly increased prevalence over time and regional differences (chapter three). Statistical analysis showed that the sensitive dhfr allelic haplotype frequencies decreased over time in all regions whilst the triple mutant increased. The models developed for dhps haplotypes showed a clear East-West divide with the double mutant dominating East Africa and the single mutant occurring mainly in West Africa (chapter four). Standardised in vivo efficacy data were matched with dhfr and dhps haplotypes using the nearest geographic location and study year and then modelled. These analyses showed for the first time, a clear effect of the frequency of dhps allelic haplotypes on SP efficacy (chapter five). These findings are important in the context of continued use of SP in sub-Saharan Africa for intermittent preventive treatment and support the use of molecular surveillance to inform policy on SP use.
Table of Contents

Declaration ............................................................................................................................ 2
Abstract ................................................................................................................................. 3
List of Figures ........................................................................................................................ 8
List of Abbreviations ............................................................................................................ 10
Acknowledgments ............................................................................................................... 14
Chapter 1 Introduction and scope of thesis ........................................................................ 15
  The current malaria situation in sub-Saharan Africa ................................................................. 15
  A chronology of antimalarial drug use and spatial context of antimalarial resistance in Africa ................................................................. 19
  Sulphadoxine Pyrimethamine use in the present day ............................................................... 23
  Antimalarial Drug Resistance ............................................................................................ 25
  The Emergence and Spread of Drug Resistance ..................................................................... 27
  Monitoring Antimalarial Drug Efficacy .................................................................................. 32
  Scope of thesis .................................................................................................................... 34
Chapter 2 Review of Sulphadoxine Pyrimethamine resistance studies in Africa and documenting data on in vivo efficacy and molecular markers in a relational database ................................................................................................. 36
  2.1 Introduction ....................................................................................................................... 36
  The in vivo therapeutic response for measuring and monitoring SP resistance ................................................................. 37
  The evolution of in vivo efficacy test systems ........................................................................ 37
  Molecular markers to measure and monitor SP resistance ................................................... 45
  2.2 Aims and Objectives ......................................................................................................... 47
  2.3 Methods ............................................................................................................................ 47
  Search strategy ...................................................................................................................... 47
  Data collection tools ............................................................................................................ 49
  The Proforma ....................................................................................................................... 49
  Data Entry System .............................................................................................................. 51
  Data validation ...................................................................................................................... 51
  Mapping within a Geographic Information System ............................................................... 51
  2.4 Results ............................................................................................................................... 52
  SP in vivo Studies .................................................................................................................. 53
  Standardisation of WHO Protocols ...................................................................................... 55
  Molecular studies .................................................................................................................. 56
  Synopsis of methods used in molecular studies of SP ............................................................. 57
  Relative accuracy of genotyping methodologies .................................................................... 60
  Distribution and coverage of dhfr and dhps in sub-Saharan Africa ......................................... 62
  2.5 Discussion ........................................................................................................................... 73
  Limitations ............................................................................................................................. 76
Chapter 3 Spatial and temporal changes of individual Sulphadoxine Pyrimethamine point mutations in sub-Saharan Africa

3.1 Introduction ........................................................................................................... 77
3.2 Aims and objectives ............................................................................................. 81
3.3 Methods ............................................................................................................... 81
   Data collection ........................................................................................................ 81
   3.3.1 Spatial mapping of point mutation prevalence ............................................... 82
   3.3.2 Statistical model development ....................................................................... 83
   Multivariable regression analyses for point mutations ............................................ 83
   Further multivariable regression analyses for dhps 437G and 540E ...................... 83
   Post estimation ....................................................................................................... 84
   Assumptions and missing data ............................................................................. 85
3.4 Results ................................................................................................................. 85
   3.4.1 Spatial mapping of point mutation prevalence ............................................... 85
   3.4.2 Change in point mutation prevalence over time ............................................ 86
   Prevalence of dhfr 164L .......................................................................................... 90
   Regional prevalence of dhfr 164L point mutation over time ..................................... 91
   Prevalence of dhps 436A ........................................................................................ 94
   Regional prevalence of dhps 436A point mutation over time .................................. 96
   Prevalence of dhps 437G ....................................................................................... 100
   Regional prevalence of dhps 437G point mutation over time ................................... 104
   Prevalence of dhps 540E ....................................................................................... 108
   Regional prevalence of dhps 540E point mutation over time .................................. 108
   Prevalence of dhps 581G ....................................................................................... 113
   Regional prevalence of dhps 581G point mutation over time .................................. 114
   Prevalence of dhps 613S ....................................................................................... 117
   Regional prevalence of dhps 613S point mutation over time ................................... 118
   Prevalence of dhps 613T ....................................................................................... 121
   Regional prevalence of dhps 613T point mutation over time ................................... 122
   3.4.3 Further analysis of dhps 437G ....................................................................... 126
   Interaction between region and time ..................................................................... 128
   3.4.4 Further analysis of dhps 540E ....................................................................... 131
   Interaction between region and time ..................................................................... 133
3.5 Discussion .......................................................................................................... 135
   Limitations ............................................................................................................. 138
   Recommendation ................................................................................................. 138

Chapter 4 Spatial and Temporal Changes of Sulphadoxine Pyrimethamine allelic haplotypes in sub-Saharan Africa ................................................................................. 140
### List of Figures

- Figure 1.1 Regional categorisation of Africa based on malaria epidemiology .......... 18
- Figure 1.2 Progression of confirmed *P. falciparum* chloroquine resistance .......... 20
- Figure 1.3 Time line of antimalarial resistance development .............................. 22
- Figure 1.4 Coverage of SP for IPT in pregnant women ......................................... 24
- Figure 1.5 Distribution of *P. falciparum* EIRs in 2010 ............................................. 30
- Figure 2.1 Sub-Saharan African countries .............................................................. 52
- Figure 2.2 Standardised WHO Protocols for SP monotherapy studies .................... 55
- Figure 2.3 Countries where SP *in vivo* surveys were conducted .............................. 56
- Figure 2.4 Summary of mutation screening methods used ....................................... 57
- Figure 2.5 Sites surveyed for dhfr I164L ................................................................. 62
- Figure 2.6 Sites surveyed for dhps A437G .............................................................. 63
- Figure 2.7 Sites surveyed for dhps K540E .............................................................. 64
- Figure 2.8 Sites surveyed for dhps A581G .............................................................. 65
- Figure 2.9 Sites surveyed for dhps S436A .............................................................. 66
- Figure 2.10 Sites surveyed for dhps A613S/T ............................................................ 67
- Figure 2.11 Sites surveyed for dhfr IRN ................................................................. 68
- Figure 2.12 Sites surveyed for dhfr NCS ............................................................... 69
- Figure 2.13 Sites surveyed for dhps GE ................................................................. 70
- Figure 2.14 Gap analysis: Molecular surveillance coverage of all dhfr and dhps point mutations .................................................................................. 72
- Figure 3.1 Prevalence of dhfr 164L ........................................................................... 90
- Figure 3.2 Surveys of dhfr 164L prevalence (with 95% confidence intervals) conducted during 1988-2008 and displayed according to their geographic region ...... 92
- Figure 3.3 Predicted probabilities of dhfr 164L over time ........................................ 94
- Figure 3.4 Prevalence of dhps 436A ......................................................................... 96
- Figure 3.5 Surveys of dhps 436A prevalence (with 95% confidence intervals) conducted during 1988-2008 and displayed according to their geographic region .......... 97
- Figure 3.6 Predicted probabilities of dhps 436A over time ...................................... 99
- Figure 3.7 Prevalence of dhps 437G ....................................................................... 103
- Figure 3.8 Surveys of dhps 437G prevalence (with 95% confidence intervals) conducted during 1988-2008 and displayed according to their geographic region .... 105
- Figure 3.9 Predicted probabilities of dhps 437G over time ...................................... 107
- Figure 3.10 Prevalence of dhps 540E ...................................................................... 109
- Figure 3.11 Surveys of dhps 540E prevalence (with 95% confidence intervals) conducted during 1988-2008 and displayed according to their geographic region .... 110
- Figure 3.12 Predicted probabilities of dhps 540E over time ..................................... 112
- Figure 3.13 Prevalence of dhps 581G ..................................................................... 113
- Figure 3.14 Surveys of dhps 581G prevalence (with 95% confidence intervals) conducted during 1988-2008 and displayed according to their geographic region .......... 115
- Figure 3.15 Predicted probabilities of dhps 581G over time ..................................... 117
- Figure 3.16 Prevalence of dhps 613S ..................................................................... 118
- Figure 3.17 Surveys of dhps 613S prevalence (with 95% confidence intervals) conducted during 1988-2008 and displayed according to their geographic region .......... 119
- Figure 3.18 Predicted probabilities of dhps 613S over time ..................................... 121
- Figure 3.19 Prevalence of dhps 613T ...................................................................... 122
- Figure 3.20 Surveys of dhps 613T prevalence (with 95% confidence intervals) conducted during 1988-2008 and displayed according to their geographic region .......... 123
- Figure 3.21 Predicted probabilities of dhps 613T over time .................................... 125
- Figure 3.22 Predicted probabilities of regional dhps 437G prevalence over time .... 127
Figure 3.23 Predicted probabilities of *dhps* 437G prevalence with interaction effects ...... 130
Figure 3.24 Predicted probabilities of regional *dhps* 540E prevalence over time .......... 132
Figure 3.25 Predicted probabilities of *dhps* 540E prevalence with interaction effects .... 134
Figure 4.1 Frequency of *dhfr* NCS haplotype .......................................................... 150
Figure 4.2 Surveys of *dhfr* NCS allelic haplotype frequency (with 95% confidence intervals) conducted during 1988-2008 and displayed according to their geographic region ...... 152
Figure 4.3 Estimated changes in *dhfr* NCS allelic haplotype frequency over time .......... 154
Figure 4.4 Frequency of *dhfr* IRN haplotype ............................................................. 156
Figure 4.5 Surveys of *dhfr* IRN allelic haplotype frequency (with 95% confidence intervals) conducted during 1988-2008 and displayed according to their geographic region .... 157
Figure 4.6 Predicted probabilities of *dhfr* IRN over time ......................................... 159
Figure 4.7 Frequency of *dhps* GE double mutant haplotype .................................. 162
Figure 4.8 Surveys of *dhps* GE double mutant allelic haplotype frequency (with 95% confidence intervals) conducted during 1988-2008 and displayed according to their geographic region ................................................................. 163
Figure 4.9 Predicted probabilities of *dhps* GE over time ......................................... 166
Figure 4.10 Frequency of *dhps* GK single mutant haplotype .................................. 167
Figure 4.11 Surveys of *dhps* GK single mutant allelic haplotype frequency (with 95% confidence intervals) conducted during 1988-2008 and displayed according to their geographic region ............................................................................................................ 169
Figure 4.12 Predicted probabilities of *dhps* GK over time ....................................... 171
Figure 4.13 Frequency of *dhps* AK fully sensitive haplotype .................................. 172
Figure 4.14 Surveys of *dhps* AK allelic haplotype frequency (with 95% confidence intervals) conducted during 1988-2008 and displayed according to their geographic region ............................................................................................................ 173
Figure 4.15 Predicted probabilities of *dhps* AK over time ....................................... 175
Figure 4.16 Samples tested positive for *dhps* GE, GK and AK haplotypes ............... 176
Figure 4.17 Predicted probabilities of regional *dhps* GE allelic haplotype frequencies with interaction between region and time .......................................................... 182
Figure 4.18 Predicted probabilities of regional *dhps* GK allelic haplotype frequencies with interaction between region and time .......................................................... 183
Figure 4.19 Predicted probabilities of regional *dhps* AK allelic haplotype frequencies with interaction between region and time .......................................................... 184
Figure 5.1 Regional classification of sub-Saharan African countries for matched *in vivo* and molecular studies ............................................................. 201
Figure 5.2 Sensitivity and specificity of *dhps* GE and *dhps* GK incremented by 10% frequencies, in predicting any phenotypic resistance ................................................. 216
Figure 6.1 Map of predicted SP failure rates (%) between 2004 and 2008 according to recent measures of *dhps* GE frequencies ................................................. 233
Figure 6.2 Migration communities in sub-Saharan Africa with *dhps* allelic haplotypes ... 240
List of Tables

Table 2.1a Summary of study end points for WHO protocols (1965-2009)........................ 43
Table 2.2 Molecular markers for dhfr and dhps............................................................... 46
Table 2.3 Inclusion and exclusion criteria applied to the selection of in vivo studies.......... 48
Table 2.4 Electronic search strategy results.................................................................... 53
Table 2.5 Summary of studies with SP monotherapy in vivo data .................................. 54
Table 2.6 Gap analyses: countries where no data were found for each point mutation and allelic haplotype .................................................................................. 71
Table 3.1 Regional classification of countries.................................................................. 82
Table 3.2 Coding of geographic regions to create dummy variables ............................ 84
Table 3.3 Summary of surveys available with prevalence data for point mutations ....... 87
Table 3.4 Summary of surveys available with prevalence data for point mutations by region .................................................................................................................. 88
Table 3.5 Model for dhfr 164L....................................................................................... 93
Table 3.6 Model for dhps 436A.................................................................................... 98
Table 3.7 Model for dhps 437G..................................................................................... 106
Table 3.8 Model for dhps 540E.................................................................................... 111
Table 3.9 Model for dhps 581G.................................................................................... 116
Table 3.10 Model for dhps 613S................................................................................... 120
Table 3.11 Model for dhps 613T.................................................................................. 124
Table 3.12 Further analysis of dhps 437G........................................................................ 126
Table 3.13 Model for dhps 437G with interaction between time and region............... 129
Table 3.14 Further analysis of dhps 540E........................................................................ 131
Table 3.15 Model for dhps 540E with interaction between time and region................. 133
Table 4.1 Common dhfr haplotypes in Africa............................................................... 142
Table 4.2 Common dhps haplotypes in Africa............................................................. 142
Table 4.3 Combinations of dhfr and dhps haplotypes and levels of SP resistance ...... 143
Table 4.4 Summary of surveys available for dhfr NCS and IRN allelic haplotypes ...... 148
Table 4.5 Summary of surveys available for dhps GE double mutant, GK single mutant and AK fully sensitive allelic haplotypes......................................................... 149
Table 4.6 Model for dhfr NCS .................................................................................. 153
Table 4.7 Model for dhfr IRN..................................................................................... 158
Table 4.8 Model for dhfr IRN triple mutant using study year and region as covariates .... 160
Table 4.9 Model for dhps GE.................................................................................... 165
Table 4.10 Model for dhps GK.................................................................................. 170
Table 4.11 Model for dhps AK................................................................................... 174
Table 4.12 Multivariable analyses of dhps GE, dhps GK and dhps AK......................... 179
Table 4.13 Model for dhps GE,GK and AK with interaction between time and region .... 181
Table 5.1 Inclusion criteria for selection of in vivo studies ........................................ 192
Table 5.2 Classification of in vivo treatment responses used to generate resistance profiles ...................................................................................................................... 194
Table 5.3 Roche et al., (2003) reclassification of in vivo end points.............................. 196
Table 5.4 Less common dhfr alleles ............................................................................. 200
Table 5.5 Prevalence of phenotypic response to SP treatment matched with frequency of dhfr NCS allelic haplotype ................................................................. 203
Table 5.6 Prevalence of phenotypic response to SP treatment and frequency of dhfr IRN allelic haplotype in matched surveys ......................................................... 204
Table 5.7 Observed prevalence of SP phenotype resistance at various dhfr allelic haplotype frequencies ................................................................................................. 205
Table 5.8 Observed prevalence of SP treatment success at various dhfr allelic haplotype frequencies ......................................................................................... 206
Table 5.9 Results of analysis for dhfr NCS and dhfr IRN predicting SP phenotype .......... 208
Table 5.10 Prevalence of phenotypic response to SP treatment among studies matched with dhps allelic haplotypes ......................................................................................... 209
Table 5.11 Frequencies of dhps GE, GK and AK allelic haplotypes in surveys matched with SP phenotype .................................................................................................................. 210
Table 5.12 Observed prevalence of any grade resistance (in vivo R) at various dhps allelic haplotype frequencies ........................................................................................................ 211
Table 5.13 Observed prevalence of high grade phenotypic resistance at various dhps allelic haplotype frequencies .................................................................................................................. 212
Table 5.14 Observed prevalence of SP treatment success at various dhfr allelic haplotype frequencies .......................................................................................................................... 213
Table 5.15 Results of dhps GE, GK, GE+GK and AK individually predicting in vivo outcome ........................................................................................................................................ 214
Table 5.16 Multivariable logistic regression analysis using both GE and GK as covariates for predicting in vivo outcome .......................................................................................................................... 217
Table 5.17 Regional variation in phenotypic outcome .......................................................................... 219
Table 5.18 Multivariate logistic regression analysis of SP phenotypic outcome with genotype and region as covariates in one model .................................................................................................................. 219
Table 5.19 Surveys with less common dhfr allelic haplotypes categorised by region .................................................................................................................. 221
Table 5.20 Predicted probability of any phenotype resistance based on varying dhps GE frequencies ........................................................................................................................................ 224
### List of Abbreviations

<table>
<thead>
<tr>
<th>Abbreviation</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>ACR</td>
<td>adequate clinical response</td>
</tr>
<tr>
<td>ACPR</td>
<td>adequate clinical and parasitological response</td>
</tr>
<tr>
<td>ACT</td>
<td>artemisinin combination therapy</td>
</tr>
<tr>
<td>ADS</td>
<td>African data Sampler</td>
</tr>
<tr>
<td>AIC</td>
<td>Akaike information criterion</td>
</tr>
<tr>
<td>CI</td>
<td>confidence intervals</td>
</tr>
<tr>
<td>DRC</td>
<td>Democratic Republic of Congo</td>
</tr>
<tr>
<td>DHFR</td>
<td>dihydrofolate reductase enzyme</td>
</tr>
<tr>
<td>dhfr</td>
<td>dihydrofolate reductase gene</td>
</tr>
<tr>
<td>DHPS</td>
<td>dihydropteroate synthase enzyme</td>
</tr>
<tr>
<td>dhps</td>
<td>dihydropteroate synthase gene</td>
</tr>
<tr>
<td>E8</td>
<td>elimination eight</td>
</tr>
<tr>
<td>EIR</td>
<td>entomological inoculation rate</td>
</tr>
<tr>
<td>ETF</td>
<td>early treatment failures</td>
</tr>
<tr>
<td>GIS</td>
<td>geographic information systems</td>
</tr>
<tr>
<td>GLURP</td>
<td>glutamate-rich protein</td>
</tr>
<tr>
<td>IPT</td>
<td>intermittent preventive treatment</td>
</tr>
<tr>
<td>IPTi</td>
<td>intermittent preventive treatment among infants</td>
</tr>
<tr>
<td>IPTc</td>
<td>intermittent preventive treatment among children</td>
</tr>
<tr>
<td>IPTp</td>
<td>intermittent preventive treatment among pregnant women</td>
</tr>
<tr>
<td>LCF</td>
<td>late clinical failure</td>
</tr>
<tr>
<td>LTF</td>
<td>late treatment failures</td>
</tr>
<tr>
<td>LPF</td>
<td>late parasitological failure</td>
</tr>
<tr>
<td>LSDI</td>
<td>Lubombo Spatial Development Initiative</td>
</tr>
<tr>
<td>Acronym</td>
<td>Description</td>
</tr>
<tr>
<td>---------</td>
<td>-------------</td>
</tr>
<tr>
<td>MAP</td>
<td>Malaria Atlas Project</td>
</tr>
<tr>
<td>MARA</td>
<td>Mapping Malaria Risk in Africa</td>
</tr>
<tr>
<td>MOI</td>
<td>multiplicity of infection</td>
</tr>
<tr>
<td>MSP</td>
<td>merozoite surface protein</td>
</tr>
<tr>
<td>NADPH</td>
<td>nicotinamide adenine dinucleotide phosphate</td>
</tr>
<tr>
<td>PABA</td>
<td>para aminobenzoic acid</td>
</tr>
<tr>
<td>PCR RFLP</td>
<td>polymerase chain reaction restriction fragment length polymorphism</td>
</tr>
<tr>
<td>PfCRT</td>
<td><em>Plasmodium falciparum</em> chloroquine resistance transporter</td>
</tr>
<tr>
<td>RBM</td>
<td>Roll Back Malaria</td>
</tr>
<tr>
<td>SMC</td>
<td>seasonal malaria chemoprevention</td>
</tr>
<tr>
<td>SNP</td>
<td>single nucleotide polymorphisms</td>
</tr>
<tr>
<td>SP</td>
<td>sulphadoxine pyrimethamine</td>
</tr>
<tr>
<td>WHO</td>
<td>World Health Organisation</td>
</tr>
<tr>
<td>WWARN</td>
<td>World Wide Antimalarial Resistance Network</td>
</tr>
</tbody>
</table>
Acknowledgments

I wish to acknowledge the support I received during my studies. My sincere gratitude goes to my supervisor Dr Cally Roper who mentored me and shared her office space during my visits. I wish to thank Dr Immo Kleinschmidt who guided me on the statistical methods and gave so willingly of his time. I am most grateful for his insights and timely feedback. I also thank Prof Karen Barnes who advised me during the early stages of my thesis.

I wish to thank the Gates Malaria Partnership for financial assistance which afforded me the opportunity to pursue my studies at LSHTM. This work was done under the Mapping Malaria Risk in Africa /Atlas du Risque de la Malaria en Afrique collaboration.

I also thank my colleagues at the Medical Research Council. I am particularly grateful to Tarylee Reddy for her support. My appreciation goes to the late Dr Brian Sharp, former Director of the Malaria Research Unit at the Medical Research Council, South Africa for encouraging me to pursue my studies and work independently, as well as to the current Director, Prof Rajendra Maharaj who encouraged me to finish. I acknowledge the database unit of the Malaria Research Unit and the Biostatistics unit, Medical Research Council, South Africa for database support.

I dedicate this work to my husband, Devan. Above all, I am eternally thankful to him for his resolute support and forbearance throughout this time.
Chapter 1 Introduction and scope of thesis

This introductory chapter summarises the current malaria situation in sub-Saharan Africa and presents a chronology of antimalarial drug use and its relationship to resistance. It describes the background and spatial context of antimalarial resistance and reviews the factors thought to be important to the emergence of resistance and its geographical distribution in Africa. The chapter concludes with an overview of current resistance surveillance practice and an outline of the research covered in this thesis.

The current malaria situation in sub-Saharan Africa

Malaria occurs in 47 sub-Saharan countries with diverse epidemiological settings and 30 of these carry the highest malaria risk and mortality rates attributed to *Plasmodium falciparum* based on national surveillance reports (WHO, 2008; WHO, 2011a; WHO, 2012). Renewed efforts to control malaria have resulted in many of those countries (which have sufficiently good surveillance systems) recording a reduction in confirmed malaria cases and mortality rates (WHO, 2011a). However global and Africa-wide malaria burden estimates reveal the extent of malaria risk. Approximately 1.2 billion people live in areas classified as high risk of malaria (>1 case per 1000 population) and 47% of these reside in Africa. In 2009 there were approximately 225 million malaria cases globally with 78% of these being from Africa (WHO, 2010a) and in 2010 the number of malaria cases worldwide dropped to 219 million with 81% of these presenting in Africa (WHO, 2011a, WHO, 2012). In 2009 there were approximately 781,000 malaria related deaths globally and 91% of these occurred in Africa (WHO, 2010a) whilst later in 2010 the number of malaria deaths worldwide was 660,000 and again the majority (91%) occurred in Africa (WHO, 2011a). Children under five years of age account for the largest proportion of the malaria mortality rates (WHO, 2011a).

The World Health Organisation (WHO) categorised the malarious countries within Africa into five subregions, grouping countries which share similar epidemiological settings, malaria control strategies and malaria incidence patterns (WHO, 2011a, WHO, 2012). The five subregions illustrated in Figure 1.1 are `Central Africa’, `West Africa’, `East Africa’ (this category includes high transmission countries in Southern Africa and Madagascar), `Southern Africa’ (the low transmission countries of Southern Africa) and finally the `Eastern Mediterranean’ subregion which includes North Sudan, South Sudan, Somalia and Djibouti (WHO, 2011a, WHO, 2012).
The Central Africa group representing nine Central African countries includes Angola, Burundi, Cameroon, Central African Republic, Chad, Democratic Republic of Congo (DRC), Equatorial Guinea, Gabon and Republic of Congo. The West African group consists of 17 countries namely Benin, Burkina Faso, Cape Verde, Gambia, Ghana, Guinea, Guinea-Bissau, Ivory Coast, Liberia, Mali, Mauritania, Niger, Nigeria, Sao Tome/Principe, Senegal, Sierra Leone and Togo. The East African subregion also includes high transmission countries in Southern Africa and consists of 11 countries namely Comoros, Eritrea, Ethiopia, Kenya, Madagascar, Malawi, Mozambique, Rwanda, Tanzania, Uganda and Zambia. The low transmission Southern African countries make up the fourth group consisting of five countries namely Botswana, Namibia, South Africa, Swaziland and Zimbabwe. The fifth group termed the Eastern Mediterranean area consists of four countries namely North Sudan, South Sudan, Somalia and Djibouti. Each group is described briefly below.

All nine Central African countries are classified as being at high risk of *P. falciparum* malaria where the numbers of reported cases are ≥ 1 per 1000 per year (WHO, 2011a). Malaria cases increased in the DRC and Republic of Congo whilst it remained stable in all other countries (WHO, 2011a). High malaria transmission occurs in Angola, Central African Republic, Equatorial Guinea, Gabon and Republic of Congo (WHO, 2011a). Both high and low transmission occurs in Burundi, Cameroon, Chad and DRC (WHO, 2011a). Low malaria transmission is characterised as having zero to one case per 1000 people (WHO, 2012).

In West Africa there are marked variations in the malaria burden among the 17 countries in this group (Figure 1.1). Predominately high malaria transmission occurs in Benin, Burkina Faso, Ivory Coast, Gambia, Ghana, Guinea, Guinea-Bissau, Liberia, Nigeria, Sao Tome/Principe, Sierra Leone and Togo (WHO, 2011a, WHO, 2012). Notably over 40% of the malaria mortality rates mentioned previously occurs in the DRC and Nigeria alone (WHO, 2012). Both high and low malaria transmission occurs in Mali, Niger, Senegal, Mauritania and Cape Verde, although there have been reductions in malaria cases in the Cape Verde and Sao Tome/Principe islands (WHO, 2011a). Sao Tome/Principe has successfully reduced malaria incidence by ≥ 75% during 2010-2011 (WHO, 2012, Snow et al., 2012).

The East and some Southern African countries are grouped together making up the third group of 11 countries. Within this region Malawi, Mozambique, Rwanda, Uganda and
Zambia are characterised by high malaria transmission whilst Comoros, Eritrea, Ethiopia, Kenya, Madagascar and Tanzania have a mixture of both high and low transmission foci (WHO, 2012). This group excludes the low transmission countries in Southern Africa, which make up the fourth subgroup.

With the exception of Zimbabwe in this fourth group, the other four countries have reported declining malaria incidence of ≥75% since 2000 (WHO, 2012) and efforts are underway to reduce malaria transmission to zero cases and thus eliminate malaria (Global Health Group, 2009). Eight countries in Southern Africa have embarked on the goal to eliminate malaria by 2020 (Malaria Elimination 8 Ministerial Meeting, 2009). The four front line countries of Elimination Eight (E8) include Botswana, Namibia, South Africa and Swaziland which are targeted first for malaria elimination (Malaria Elimination 8 Ministerial Meeting, 2009). The four second line E8 countries are Angola, Mozambique, Zambia and Zimbabwe which will eventually be targeted for malaria elimination (Malaria Elimination 8 Ministerial Meeting, 2009). Although local malaria transmission within these countries might be declining, malaria remains a priority due to the risk of malaria epidemics and the associated high morbidity and mortality especially among poor and rural communities.

Countries in the Eastern Mediterranean subregion have variable malaria transmission intensities (Figure 1.1). Predominately high malaria transmission occurs in South Sudan whilst primarily low malaria transmission occurs in Djibouti (Snow et al., 2012; WHO, 2012). North Sudan and Somalia are characterised by both high and low malaria transmission (WHO, 2012).
Targeted interventions such as indoor residual spraying, use of bed nets, diagnostic testing as well as preventive and curative therapies, improved case surveillance and health promotion are being used to control the malaria burden across sub-Saharan Africa. The scale-up of such interventions is central to elimination efforts and it is vital that they should be sustained in areas that have achieved high coverage and expanded in areas of suboptimal coverage (Alonso and Tanner, 2013). Early diagnosis and prompt treatment of confirmed malaria cases is the primary malaria control tool in many parts of Africa (Mbacham et al., 2010). There is concern about the public health implications of both insecticide and drug resistance, as future options for both interventions are limited and there is no vaccine for malaria as yet. The emergence and spread of *P. falciparum* resistance to antimalarial drugs contributes significantly to malaria-specific morbidity and mortality in Africa (Trape, 2001; Plowe, 2003; Bjorkman and Bhattarai, 2005; Greenwood
et al., 2005; Korenromp et al., 2003). Regional differences in malaria mortality rates within Africa were attributed to the emergence of drug resistance. Malaria mortality increased significantly from the 1980s to the 1990s within East and Southern Africa where resistance was rising but there was no absolute increase in malaria mortality over time in West Africa (Korenromp et al., 2003). A background into the emergence of antimalarial drug resistance in Africa is presented next.

A chronology of antimalarial drug use and spatial context of antimalarial resistance in Africa

Chloroquine monotherapy was first introduced as treatment for uncomplicated malaria around 1946 (Warrell, 1993; Jensen and Mehlhorn, 2009). Across Africa chloroquine became widely available in clinics, shops and private pharmacies from the 1960s and was the antimalarial drug of choice because it was cheap, safe and well tolerated (Warrell, 1993; Snow et al., 2012). Case reports of chloroquine resistance first emerged in South America and in Southeast Asia along the Cambodian and Thailand border in 1957 and resistance was confirmed in 1960 (WHO,1987; D, Payne, PhD Thesis, University of London, 1989¹; Warrell, 1993). Pailin province in Cambodia was identified as the focal point of chloroquine resistance (WHO, 1963). In Africa chloroquine resistance was reported much later than Asia, during the 1970s (Payne, 1987; WHO,1987).

After the first report of chloroquine resistance in East Africa in 1978 it spread westwards across the continent (Charmot et al., 1991). Payne (1989) and Charmot et al., (1991) illustrated the stepwise geographic and temporal spread of chloroquine resistance since its initial occurrence in East Africa and subsequent dispersal across Africa during the 1980s (Figure 1.2). A description of these geographic analyses is provided later in this chapter. Resistant P. falciparum parasites were found in all tropical African countries from 1978 to 1988 (Figure 1.2) (Trape, 2001; Charmot et al., 1991). Genetic studies later revealed that this was due to the progressive spread of the resistant form of the chloroquine resistance transporter (Pfcrt) gene (Fidock et al., 2000) and the African and Asian Pfcrt shared the same ancestry (Wootton et al., 2002). As chloroquine became ineffective against P. falciparum malaria nearly everywhere it was used, it was withdrawn as first line therapy. First line treatment is the official drug of choice for treatment of uncomplicated malaria (Malisa et al., 2010) as per the country’s treatment policy guidelines.

¹ referred to as Payne (1989) hereon
Pyrimethamine monotherapy was used for both malaria prophylaxis and treatment in the 1950s (Peters, 1970) but resistance emerged rapidly at multiple sites (Warrell, 1993) and it was only when it was later combined with sulphadoxine that it came to be a reliable antimalarial treatment. The antifolate combination sulphadoxine pyrimethamine (SP) was adopted as second line malaria treatment and came to be used increasingly as chloroquine efficacy waned. Second line malaria treatment is the official policy recommended alternative when the first line treatment fails to clear infection (Malisa et al., 2010). Eventually when chloroquine was rendered ineffective in an area, SP became the replacement first line therapy.
The SP combination was initially used as monotherapy in the 1990s in Africa. As with chloroquine resistance, the Thailand-Cambodia border was the source of the first reports of SP treatment failures during the late 1970s (Vinayak et al., 2010). Although the first case reports of SP resistance were sporadic, confirmation of clinical resistance came shortly thereafter and then resistance developed rapidly (White, 1992). In Thailand cure rates declined substantially within five years of replacing chloroquine with SP (Rieckman, Suebsaeng and Rooney, 1987).

The emergence and spread of *P. falciparum* resistance to both chloroquine and SP necessitated the use of alternative combination therapies. The aim of using combination therapies was to prolong the useful therapeutic life of each drug and delay antimalarial resistance. The WHO released treatment guidelines early in 2006, advising that uncomplicated *P. falciparum* malaria should be treated with artemisinin combination therapies (ACTs) as first line treatment and not by artemisinins alone or any other monotherapy (WHO, 2006). Recently the WHO (2012) reported that most malaria endemic countries had changed their treatment policy to ACTs.

The threat of resistance to antimalarial drugs is ever present and first reports of declining *P. falciparum* efficacy to artemisinin were reported in Thailand in 2003, along the North Western Thailand-Myanmar border and Western Cambodia during 2007-2008, as well as in Vietnam (Dondorp et al., 2009; WHO, 2012). Figure 1.3 illustrates the time lines for the development of resistance to chloroquine, SP and now decreased efficacy and resistance to artemisinin. Pailin province in Cambodia has become the epicentre of *P. falciparum* antimalarial drug resistance (Dondorp et al., 2009) given its precedent for first reports of both chloroquine and SP resistance. It took about 10 years for chloroquine resistance to be detected after its initial use and it took 17 years before it could be detected in East Africa (Figure 1.3). As mentioned earlier SP resistance was first detected in Southeast Asia in the 1970s and then it was detected in East Africa in the 1990s. East Africa was the forerunner of antimalarial resistance within the African continent.
Artemisinin resistance will have dire consequences for malaria control if it spreads across the globe. Whilst the mechanisms of artemisinin resistance have yet to be elucidated, strategies have been put in place for containing its spread (WHO, 2011b) especially as history has shown that once resistance emerges in one place it tends to spread to other places, even across continents. Efforts are underway to increase monitoring and surveillance of ACTs to detect outbreaks of artemisinin tolerance and resistance at new sites (Vijaykadga et al., 2006; Dondorp et al., 2009; WHO, 2011b).

With limited effective antimalarials available for use as ACTs, SP continues to be used in specific settings in Africa as a preventative treatment and in combination with other drugs as first line therapy. There is a clear perception that the Eastern and Southern African regions have experienced the majority of SP resistance problems. A systematic spatial summary of SP resistance measures which might inform the continued use of SP for treatment or prophylaxis in sub-Saharan African countries is urgently needed. The work in this thesis sets out the first attempt and possibly the only initiative of its kind to map the extent of SP resistance in sub-Saharan Africa.
**Sulphadoxine Pyrimethamine use in the present day**

SP continues to be used as treatment both in combination with artemisinin and as monotherapy for chemoprevention (WHO, 2012). The combination of artesunate and SP is being used for first line treatment of unconfirmed malaria as well as first line treatment of *P. falciparum* (WHO, 2011a; WHO, 2012). SP is being used in combination with artesunate for first line treatment of *P.falciparum* in three countries namely Djibouti (adopted in 2008), Somalia (adopted in 2006) and North Sudan (adopted in 2006) where there is low transmission (WHO, 2011a; WHO, 2012). In Eritrea, SP is being used with chloroquine as first line treatment for unconfirmed malaria and this policy was adopted in 2007 (WHO, 2011a; WHO, 2012). The WHO recommends three chemoprevention strategies involving intermittent preventive treatment (IPT) among pregnant women (IPTp) and infants (IPTi) as well as seasonal malaria chemoprevention (SMC). IPT is an intervention recommended for high risk groups such as pregnant women and infants living in high malaria transmission areas (WHO, 2012). IPT involves administering a scheduled therapeutic dose of an effective antimalarial to these high risk groups during an intermittent and defined period, aiming to reduce parasite carriage within foci of transmission (Greenwood, 2007; Gosling *et al*., 2010; WHO, 2012).

SP is viewed as the safest option for preventing malaria during pregnancy (Hyde, 2008). The three IPT strategies are discussed briefly here. The first chemoprevention strategy, IPTp recommends that all asymptomatic pregnant women living in moderate to high malaria transmission areas receive SP at each antenatal visit and the majority of African countries now use a two dose regimen (WHO, 2005; WHO, 2012). SP for IPT is only recommended in countries not using the SP+artesunate combination (WHO, 2009a). Although recent data show that SP for IPTp has been adopted in 37 sub-Saharan African countries (WHO, 2011a) (Figure 1.4) coverage among pregnant women who actually received the two doses of SP was low (WHO, 2012). Notably there was low SP-IPT coverage in the two highly populated countries of Nigeria and DRC (WHO, 2012).
Figure 1.4 Coverage of SP for IPT in pregnant women (after WHO, 2011a; WHO, 2012)
Darker shades refer to later years and lighter shades refer to recent years.

Chemoprevention in infants (IPTi) involves the scheduled administration of three doses of SP as part of the expanded programme on immunization in areas of moderate to high malaria transmission and low levels of SP resistance (WHO, 2010b, WHO, 2012). According to the WHO (2012) thus far only Burkina Faso has adopted this IPTi strategy as a national policy whilst other countries are planning to do the same.

The third chemoprevention strategy, intermittent preventive treatment of malaria in children (IPTc) now referred to as SMC involves administering the full treatment dose of an antimalarial drug among children aged between three months and five years, residing in areas experiencing highly seasonal malaria where there is an average of ≥60% of clinical malaria cases within a four month period (WHO, 2012). Currently a combination of SP plus amodiaquine is recommended across the Sahel including Senegal, Mauritania, Mali, Niger, Chad, North Sudan, South Sudan and Eritrea provided that both drugs are still efficacious in these countries (WHO, 2012). Due to resistance to both amodiaquine and SP, this strategy is not recommended in the Southern and East African regions (WHO, 2012). The
WHO SMC strategy is a relatively new recommendation and thus far no malaria endemic country has implemented it (WHO, 2012).

A description of antimalarial resistance is presented below with specific reference to SP.

**Antimalarial Drug Resistance**

There is a difference between treatment failure and drug resistance. When a drug fails, it implies that insufficient amounts of the drug or its active metabolite reached the parasites (WHO, 1986). This may be caused by several factors including inadequate absorption, unusual drug metabolism or excretion rates (WHO, 1965) as well as incorrect dosing, non-compliance with the dosing regimen and poor drug quality (Bloland, 2001) along with misdiagnosis and re-infections. For drug resistance to occur, the parasite must have survived the specified drug treatment dosages (WHO, 1986). Thus drug resistance can cause treatment failures, but not all treatment failures are caused by parasite resistance to the drug.

For uncomplicated malaria, a treatment drug should effectively eradicate parasites or reduce the parasite biomass to an acceptable point that the body can cope with the infection by its own immune defence (White, 1998; White, 2002). Therefore, an effective antimalarial drug treatment should cure malaria and thereby reduce morbidity associated with treatment failures (WHO, 2006). The success of any antimalarial drug depends on the varying response of the parasites and each stage of its life cycle differs in its susceptibility to different drugs (WHO, 1986). According to WHO (2005), antimalarial drug resistance is associated more with treatment failure than with prophylactic failure. When used curatively, the drug should clear malaria parasites and the clinical symptoms of infection (Molineaux, 1988) which is of public health significance. Antimalarial drugs are used for prophylaxis or treatment based on their stage specificity, safety profile and half-lives (Molineaux, 1988). These drugs are often characterised by the stage in the parasite’s life cycle they target (Warrell, 1993).

The SP combination is a blood schizonticide that inhibits folate metabolism which is an essential biochemical pathway for *P. falciparum* survival (Gregson and Plowe, 2005). Sulphadoxine mimics para-aminobenzoic acid (PABA) in competing for the active site on the dihydropteroate synthase (DHPS) enzyme. The PABA interaction with DHPS would
normally lead to folic acid production in the parasite but it is interrupted in the presence of sulphadoxine (Desjardins, Doberstyn and Wernsdorfer, 1988; Olliaro, 2001). Pyrimethamine inhibits plasmodial dihydrofolate reductase (DHFR) and thereby prevents the nicotinamide adenine dinucleotide phosphate (NADPH) dependent reduction of dihydrofolate to tetrahydrofolate (Olliaro, 2001). The capitalised notations DHPS and DHFR refer to the enzymes whilst the lower case italicised \textit{dhps} and \textit{dhfr} denote the genes described below. Pyrimethamine is classified as a blood schizontocide but also inhibits sporozoite development (WHO, 2006). The synergistic effect of sulphadoxine with pyrimethamine has been effective as a combined antimalarial monotherapy. SP is effective during the asexual developmental phase of the parasite life cycle but is ineffective against mature gametocytes within the sexual developmental phase. SP increases gametocyte carriage but also reduces the infectivity of gametocytes to mosquitoes (Hogh \textit{et al}., 1998). Resistance to SP is caused by point mutations in \textit{dhfr} and \textit{dhps} genes which normally confer susceptibility to inhibitors (White, 1998; Gregson and Plowe, 2005). This will be discussed further in chapter 2.

Antimalarial treatments are designed to sustain blood concentrations which are sufficient to restrict a parasite multiplication rate of < 1 per cycle until the last viable parasite has been eliminated from the body (White, 1997). This is known as the minimum inhibitory concentration. To confirm that there is indeed antimalarial drug resistance, the treatment regime must be proven to be inefficient despite being in the body in sufficient amounts and time periods to target a minimum of four asexual parasite life cycles (approximately six days for \textit{P. falciparum})(White, 1998). This property has been incorporated into the definition of antimalarial drug resistance. The definition of drug resistance is “the ability of the parasite strain to survive and/or multiply despite the administration and absorption of a drug in doses equal to or higher than those usually recommended but within the limits of tolerance of the subject” (WHO, 2005). Hence, any parasites remaining after an antimalarial treatment regimen was completed, would signify resistance. Sibley and Price (2012) further characterise resistance as an operational term rather an absolute term, stating that different measures of resistance derived from genetic markers and \textit{in vitro} tests contribute to our understanding of decreased drug susceptibility and ultimately drug resistance. Furthermore, these authors state that resistance is a relative term and is context dependent (Sibley and Price, 2012).

While the definition of drug resistance remained largely unchanged since its inception, the WHO protocols to monitor drug efficacy underwent many changes since 1973. These changes were made in order to account for malaria transmission in different
epidemiological settings. The WHO protocols were based on in vivo tests which is the goal standard for measuring resistance and these were designed to measure parasitological and later, clinical outcomes for assessing a treatment during a specified follow-up period (WHO, 2005). As expected, the outcome measures for treatment success or failure would depend on the specific WHO protocol applied in a particular setting. The characterisation of treatment outcomes based on which protocol was used has direct implications for estimates of drug efficacy, especially in high malaria transmission areas (Dorsey et al., 2004a). The WHO protocol changes for monitoring drug efficacy are discussed further in chapter 2.

The Emergence and Spread of Drug Resistance

An important factor for antimalarial drug resistance is the rate at which chloroquine and SP resistance developed and intensified. The emergence of resistance and subsequent spread of resistance are complex events arising from spontaneous mutations as mentioned earlier, as well as drug selection pressure. Smith et al., (2010) characterised the evolution of antimalarial drug resistance into three events: (1) appearance of mutations, (2) establishment of mutations and (3) spread of mutations. Firstly, the appearance of resistance is characterised by infections with de novo mutations whereby the sensitive parasites are initially the majority strain until treatment is administered and the proportion of resistant strains eventually outnumber the sensitive forms (Smith et al., 2010). The second phase is when resistance becomes established after its initial appearance and as mutant parasites persist and spread within the population (Smith et al., 2010). This establishment phase depends on the fitness cost of mutant parasite strains which in turn depends on its biological cost versus the benefits conferred by drug pressure (Smith et al., 2010). Parasites carrying mutations depend on sustained drug pressure and once this drug pressure is reduced, the mutant strains will eventually be outnumbered by sensitive strains. The third phase, ‘spread’ involves displacement of sensitive parasite strains with resistance strains within the population and this is affected by the malaria transmission intensity in communities, as discussed further below.

According to Talisuna, Bloland and D’Allesandro et al., (2004a) the factors responsible for the emergence and spread of resistance are not fully known but there are a multitude of contributory factors at play. The amount of drug use, drug half-life which is the time taken for their concentrations to decline by 50% (WHO, 1986), host heterogeneity, parasite biomass, malaria transmission intensity, host immunity and intra-host dynamics and population movements are other contributory factors affecting the appearance of
resistance (Hastings and Watkins, 2005). There is a brief discussion below on selected factors contributing to the development of antimalarial drug resistance.

Once random mutations appear, these mutant parasite strains are selected if the administered drug concentrations are sufficient to inhibit multiplication of the susceptible parasites but are insufficient to inhibit the mutant forms (Smith et al., 2010). Consequently, selection of resistant parasite strains occurs in the context of widespread drug use, especially with prolonged use of monotherapies (D’Allessandro and Buttiens, 2001; Smith et al., 2010). The most convincing evidence of this link came from the concomitant development of chloroquine resistance along with mass administration of the drug as monotherapy in the form of medicated salt (Payne, 1988). However, this relationship cannot always be inferred, as was observed in Uganda where SP resistance was high despite low SP use in an area of high transmission (Talisuna et al., 2004b). This could probably be attributed to population movements contributing to circulation of resistant parasites (discussed in chapter 6).

Adherence to the proper treatment dosage is important, especially for drugs that are rapidly eliminated from the body. This was recognized decades ago when it was found that the administration of inadequate and badly timed doses of antimalarial drugs to infected people increased the probability of resistance (WHO, 1965). A single adult dose of SP was usually 1500 mg sulphadoxine and 75 mg pyrimethamine (WHO, 1986). Sulphadoxine and pyrimethamine have relatively long half-lives of 100-200 hours and 96-192 hours respectively (WHO, 1996). Once treatment doses of SP are administered it offers an advantageous protective effect of up to two months (Hastings, Watkins and White, 2002) against P. falciparum, but this property also increases the risk of exposure to subtherapeutic drug concentrations which contributes to the emergence of resistance (Watkins and Mosobo, 1993).

As the occurrence of P. falciparum in the body elicits humoral and cellular responses, there is another factor which contributes to the protective effect. Repeated exposure to P. falciparum leads to the development of immunity in communities, which is acquired slowly (White, 2004). Some studies have found that people with this developed background immunity against malaria have better treatment outcomes when given antimalarial drugs compared to those with no background immunity (Mayxay et al., 2001). However, this protective effect of the immune response can mask or delay the observation of antimalarial
drug resistance in communities with repeated exposure because infections could be resolved successfully regardless of parasite drug sensitivity (White, 2004).

Apart from achieving clinical cure, the secondary purpose of administering antimalarial drugs is to prevent malaria transmission by interrupting sporogony in the mosquito or by acting on infected gametocytes (Warrell, 1993). Malaria transmission is characterised by describing malaria stability, intensity and seasonality (Omumbo et al., 1998). There are several indices which can be used to assess the intensity of malaria transmission such as the entomological inoculation rate (EIR), parasite rate, annual parasite index and spleen rate (Shaukat, Breman and McKenzie 2010). The EIR which describes the number of bites from infected mosquitoes per person per unit time, is the most commonly used index of malaria transmission intensity (Kelly-Hope and McKenzie, 2009). Kelly-Hope and McKenzie (2009) provided annual estimates of EIR in Africa. Burkina Faso, Cameroon, DRC, Equatorial Guinea, Gabon, Ghana, Ivory Coast, Republic of Congo, Sierra Leone in West Africa as well as Burundi and Tanzania in East Africa all had EIRs ≥100. Later Gething et al., (2011) provided updated estimates of malaria endemicity using the EIR and basic reproductive number. These authors showed that the highest risk of *P. falciparum* transmission still occurs in Africa where the malaria endemicity is highly variable and where an overall EIR of 10 was predicted whilst the rest of the world was predicted to have an EIR of <1 (Gething et al., 2011). Notably EIRs exceeding 100 were predicted in Northern Mozambique and the borders between Cameroon and Nigeria as well as Burkina Faso and Mali (Figure 1.5) (Gething et al., 2011).
Figure 1.5 Distribution of *P. falciparum* EIRs in 2010 after Gething et al., (2011) available at http://www.map.ox.ac.uk/browse-resources/entomological-inoculation-rate/Pf_EIR/africa-plus/

Areas shaded grey represent unstable malaria transmission or no malaria risk. Areas shaded blue represent areas where EIRs where between 0.1 and 1.0. Areas shaded yellow to light orange have EIRs >1. The areas shaded dark orange have EIRs of 100 or more.

Talisuna et al., (2004a) summarised three contrasting theories relating antimalarial drug resistance and malaria transmission intensity: (1) low transmission intensity increases the spread of resistance due to the dominance of monoclonal infections, (2) resistant parasites spread faster where malaria transmission is high compared to the rate of spread in low
transmission areas and (3) transmission intensity does not play a role in the early stages of parasite resistance development.

The first theory is supported by observations that chloroquine resistance was first detected in foci of low transmission (White, 2004). Hastings and Watkins (2005) stated that acquired immunity is affected by transmission intensity, suggesting that transmission intensity has an indirect rather than a direct effect on the development of drug resistance. In low transmission settings, patients are symptomatic of malaria with little or no acquired immunity and it is possible to contain the infectious reservoir through prompt and effective treatment. In this context, the rate at which immunity is acquired is low (Hastings and Watkins, 2005), patients are generally treated with a specific antimalarial and selection of resistance develops in the context of a relatively large parasite biomass that encounter the drug within each patient (White, 2004). SP resistance evolved more rapidly in foci of low transmission than in foci of high transmission (White, 2004).

Studies of molecular markers indicated that there was rapid development of SP resistance in mesoendemic areas in Uganda (Sendagire et al., 2005) which are areas characterised as having varying transmission intensities depending on local conditions (Gilles, 1993a). In holoendemic areas malaria transmission occurs all year long and its intensity confers a considerable immune protective advantage (Gilles, 1993a). Using mathematical models to estimate the impact of IPT among adults and infants on antimalarial drug resistance, O’Meara, Smith and McKenzie (2006) predicted that partial resistance was more likely to emerge and spread in areas of low malaria transmission whereas high levels of resistance was more likely to emerge in areas of high malaria transmission.

In high transmission settings, asymptomatic people make up the bulk of the infectious reservoir and as they are continuously exposed to malaria they develop partial immunity from childhood. In these areas effective drug treatment, together with other interventions is recommended such as use of insecticide treated nets and indoor residual spraying (WHO, 2006). White (2004) explained that in foci of high malaria transmission it is only during the first few years of life that large amounts of parasites are exposed to the drug, resulting in reduced probabilities of de novo selection and transmission of resistant parasites in these areas compared to low transmission areas. Given the significance of antimalarial drug resistance to malaria control efforts, various methods have been developed to measure and monitor resistance development. These are discussed next.
Monitoring Antimalarial Drug Efficacy

There are four methods of assessing and monitoring antimalarial drug efficacy among communities. These include (1) *in vivo* studies which describe clinical and parasitological therapeutic outcomes, (2) *in vitro* studies to determine drug susceptibility, (3) clinical pharmacology tests which measure drug concentrations levels and (4) molecular genotyping to identify mutant genes (Sibley *et al.*, 2008; WHO, 2010a). Advances in computing and web-based technology have made it possible to link networks across borders to collate and mine data representing different antimalarial resistance measures to support malaria control efforts.

The descriptions of the spatial context of chloroquine resistance as documented by Payne (1987, 1989) and Charmot (1991) illustrate the use of clinical reports of resistance to track the emergence and spread of chloroquine resistance. These authors did not include data on molecular markers of resistance because these technologies did not exist at the time. Furthermore the digital age and geographic information systems (GIS) for mapping advanced only during the last 30 years (Moore and Carpenter, 1999). In the present day it has become increasingly important to have comprehensive and readily available information on antimalarial resistance to guide decisions about resistance containment and treatment changes.

There has been a drive to create an open access web based database, recognizing the need for an accessible resource containing information on the levels of antimalarial drug resistance (Sibley and Ringwald, 2006). The WHO established a global database on antimalarial drug efficacy in 2000 and has since been collecting global efficacy data (WHO, 2010c). The World Wide Antimalarial Resistance Network (WWARN) embarked on an initiative in 2007 to collect and make available surveillance data emanating from different methods of detecting antimalarial resistance in countries across the globe (Sibley *et al.*, 2008). One of the main goals of the network was to set up a system to standardise the varying data outputs associated with drug efficacy studies, especially those related to the different WHO protocols such as consolidating results obtained from 14 day trials versus those done over a 28 day follow-up period. Furthermore, Sibley and Ringwald (2006) highlighted the importance of gathering associated geographic and temporal information to enhance our understanding of the spread of resistance, which is one of the goals of this thesis.
The main purpose of establishing a digital warehouse of antimalarial drug efficacy studies is to streamline in vivo data and make it accessible to decision makers. The database can also be regularly updated and verified as information becomes available. The repository can be queried and analysed easily, which will facilitate meta-analysis of data from different sources i.e. published and unpublished reports. Sibley and Ringwald (2006) reviewed the need for a database of antimalarial drug resistance studies on a global scale. Meta-analyses of drug efficacy data are rare, probably due to the difficulty in drawing together data from different sources and standardising data and analytical methodologies in pooled analysis.

Good examples of efforts to collate malaria data on a large scale are the Mapping Malaria Risk in Africa (MARA) database (http://www.mara.org.za), the Malaria Atlas Project (www.map.ox.ac.uk), the WHO database on antimalarial drug efficacy (http://globalatlas.who.int/) and the “Chronological atlas of the recorded response of P. falciparum to chloroquine” (Payne, 1989). A discussion of the latter two databases follows as these relate specifically to antimalarial drug resistance.

In his seminal work to track the spread of chloroquine resistance, Payne (1989) collected data from 76 countries across the globe and recorded the first appearance of chloroquine resistance in each country. The data were gathered from scientific journals, WHO and Pan American Health Organisation publications, WHO/Global Monitoring Programme reports which were test reports on the response of plasmodia to antimalarial drugs, submitted to the WHO global monitoring programme, as well as unpublished reports. Confirmed reports of resistance were categorised into three groups including in vitro studies, in vivo studies and clinical reports backed by laboratory investigations. Sentinel (usually non-immune expatriates) and indigenous cases were also recorded. It must be noted that White (1992) cautioned against the use of sentinel reports of drug efficacy as they do not represent indigenous communities.

Payne (1989) used these data to draft time series maps of resistance using symbols to illustrate reports of efficacy studies done in each year. These data were then used to document spatially referenced point data on maps to infer the path and spread of chloroquine resistance as illustrated in Figure 1.2. Payne (1989) found three main foci of chloroquine resistance and concluded that there was spontaneous but independent emergence of chloroquine resistance in South America and in Southeast Asia in the late 1950s. These maps showed that chloroquine resistance emerged later in East Africa.
(during the 1970s) compared with the American and Asian continents. Payne (1989) went further to suggest that chloroquine resistance spread gradually from these initial foci. He also mentioned a possibility of a fourth focus in West Africa. This work clearly illustrates the value of mining antimalarial resistance data to detect patterns in the spread of resistance.

The WHO [http://globalatlas.who.int/](http://globalatlas.who.int/) collated antimalarial drug resistance data from published articles, unpublished sources and data from surveillance studies. The purpose of this database was to support treatment policy change decisions in Africa. Data summaries were available online in the form of graphs and map illustrations depicting trends in chloroquine and SP failure rates in individual African countries. Whilst these data provided a good overview of the resistance pattern within a country, it offered limited scope for further analysis because it did not provide access to the individual patient data (Sibley and Ringwald, 2006), which could assist the assessment of the impact of resistance on vulnerable populations. As Bjorkman and Bhattarai (2005) point out, such estimates are important to provide an indication of the magnitude of the problem to help with improving policies and control. Since SP continues to be used both as treatment and intermittent prevention in some countries described previously, this suggests that the drug combination is still effective in specific areas and that there is value in understanding the geographic variation in the distribution of resistance.

**Scope of thesis**

The goal of this study was to gather and compile the large literature of *in vivo* and molecular measures of SP resistance in sub-Saharan Africa and collate these in a geographical and temporal framework. The various indices of resistance when summarised together, allow the African geography of SP resistance to be examined in a new way. This work seeks to address fundamental questions relating to the distribution of SP resistance in sub-Saharan Africa by applying statistical tests to better understand the spatial and temporal characteristics of molecular markers of resistance and *in vivo* measures of SP resistance. Each chapter is structured in a similar manner with its own background, discussion, limitations and recommendations. A description of each chapter follows.

Chapter 2 contains an overview of SP resistance studies in sub-Saharan Africa gathered from the literature. Efficacy studies measured *in vivo* and studies of molecular markers
were included in the review. Collation and geo-referencing of these data in a relational database made it possible to map all the various indices of resistance. In this chapter the data collection and data recording methods are described, together with methodological trends in molecular surveillance and the most frequently used WHO protocols for \textit{in vivo} efficacy tests. Finally the mapped measures of SP resistance are used to identify country-specific gaps in surveillance data.

In chapter 3 the spatial and temporal variation in prevalence of seven individual SP point mutations in the \textit{dhfr} and \textit{dhps} genes are analysed. The geographical distributions of point mutations are mapped and their temporal trends examined by first categorising prevalence data according to region and analysing the major trends over time. Statistical analyses are conducted to assess the effect of regional and temporal risk factors on the prevalence of each point mutation.

Chapter 4 similarly describes spatial and temporal changes in molecular markers, but this time examining the combinations of point mutations known as allelic haplotypes, in sub-Saharan Africa. There are two major allelic haplotypes at \textit{dhfr} namely the sensitive allelic haplotype (51N/59C/108S) and the \textit{dhfr} triple mutant allelic haplotype (51I/59R/108N) and three major \textit{dhps} allelic haplotypes namely 437G/540E (GE), 437G/540K (GK) and 437A/540K (AK). The frequencies of each of these allelic haplotypes are presented graphically and mapped. Statistical analyses are performed to assess the influence of regional and temporal factors on the frequencies of these allelic haplotypes.

Chapter 5 assesses the association between \textit{dhfr} and \textit{dhps} haplotypes and \textit{in vivo} measures of SP efficacy. The \textit{in vivo} data are recalibrated and this chapter proposes a system of classification for these data to guide spatial and temporal analyses. The \textit{in vivo} studies from the standardised and calibrated subset are then matched with \textit{dhfr} and \textit{dhps} allelic haplotype frequencies, according to nearest geographic location and study year. Statistical analyses are performed to assess the influence of allelic haplotype frequencies, and any region differences on phenotypic outcomes of SP efficacy.

Chapter 6 concludes by giving an overview of the major findings in each chapter. The implications of this work for the broader context are discussed. Recommendations, limitations and possibilities for future work are described.
Chapter 2 Review of Sulphadoxine Pyrimethamine resistance studies in Africa and documenting data on in vivo efficacy and molecular markers in a relational database

2.1 Introduction

Whilst SP is still being used in sub-Saharan Africa for intermittent preventative therapies, it is useful to describe the spatial context of SP resistance. As shown in chapter 1 historic documentation of the geography of chloroquine resistance has illustrated the importance of mapping to improve our understanding of the emergence and progressive spread of drug resistance in P. falciparum malaria, globally (Payne, 1987; Payne, 1989), in Africa (Charmot et al., 1991) and more recently in India (Shah et al., 2011). These early studies mapped primarily clinical reports of resistance since molecular markers of resistance had not yet been described. Innovations in both molecular DNA typing techniques and newly available GIS technologies now mean that geographical mapping of the spread of resistance in time is possible and the timing of the emergence of SP resistance in Africa sometime after the emergence of chloroquine resistance meant that studies of resistance –both in vivo and molecular spanned the period of time during which SP resistance was emerging in Africa.

The purpose of collecting and tracking SP resistance data is threefold: firstly, to identify spatial and temporal patterns of resistance at a population level and secondly, to highlight gaps in the data coverage which are useful for improvement of surveillance initiatives. Thirdly since in vivo studies are no longer carried out for SP due to resistance, reliance on molecular markers has increased. By examining spatial and temporal changes in resistance using both in vivo and molecular data, the relationship between molecular markers and drug efficacy in vivo can be elucidated.

Although in vivo studies are the gold standard of resistance measures, molecular markers have an advantage over in vivo tests because numerous surveys for detecting point mutations can be rapidly completed over many sentinel sites covering foci (large areas) or hot spots (small areas) (Plowe, Kublin and Doumbo, 1998; Talisuna et al., 2004a). Therefore, molecular markers have become useful for tracking the geographic spread of antimalarial drug resistance. In this study in vivo and molecular measures of SP resistance were identified from the literature and mapped with the aim of quantifying and geographically tracking the occurrence of SP resistance in sub-Saharan Africa. This chapter describes how the studies were identified, how they were mapped and how the
resistance data were standardised and incorporated into a database. Finally it examines what coverage was achievable using the refined standardised datasets which go on to be analysed in subsequent chapters. In this chapter the in vivo data will be described first, followed by the molecular data.

The in vivo therapeutic response for measuring and monitoring SP resistance

In an in vivo test, the parasitological and therapeutic response of a group of symptomatic and parasitaemic people over a defined follow-up period is determined. According to Bloland (2001) in vivo test results most closely reflect the presiding clinical or epidemiological situation in an area and thus offers the best information about a drug’s efficacy. Clinical studies using in vivo tests of antimalarials for clinical and parasitological efficacy are conventionally the gold standard for assessing drug efficacy and have been the cornerstone of in-country sentinel site surveillance and the primary evidence used for guiding antimalarial drug policy and monitoring drug efficacy (WHO, 1996). In an in vivo test the specified follow-up period after treatment should allow adequate time to assess the parasites' responses to the drug and guidelines on standardisation of duration of follow-up have undergone significant revisions over time. The use of a standardised protocol is important for accurate evaluation of antimalarial drug efficacy (Sibley and Hunt, 2003). The WHO protocols underwent significant revisions three times between 1973 and 2001 to accommodate malaria control programme needs, changing malaria epidemiology and testing for resistance (Table 2.1a and b) (WHO, 1996; WHO, 2002; WHO, 2009b, WHO, 2010c).

The evolution of in vivo efficacy test systems

The WHO devised standardised protocols to regulate in vivo efficacy studies so that their results could be interpreted and compared for surveillance and monitoring. Tables 2.1a and 2.1b summarise the major changes incorporated into these protocols over time, starting with the WHO (1965) protocol which defined sensitive/resistant responses to treatment during three to 11 days follow-up. Some modifications and revisions were applied to the protocols in other years but the main protocols are listed in Tables 2.1a and b.

The early protocols formulated between 1964-1972 were performed mainly among non-immune patients and the goal was to assess parasitological response rather than clinical
response to drugs, which are important for developing treatment policies (WHO, 1996).
Antimalarial resistance was measured by a non-reduction in parasite density, persistent parasitaemia or an increase or reappearance of parasitaemia.

A major change occurred from the 1973 protocol to the 1996 protocol as the emphasis shifted to clinical responses as the primary criteria for treatment policy change and parasitological testing were reduced to a minimum, within safety protocols. The aim was not to disregard the importance of parasitological monitoring completely but rather to factor in clinical responses into drug efficacy studies. Study participants were selected on the basis of clinical evidence and microscopically confirmed *P. falciparum* malaria, especially in infants and young children in intense malaria transmission areas in the WHO (1996) protocol so as to guide national drug policies. The efficacy data generated through field application of this protocol were used to assess the practical efficacy of a drug regimen and to then guide decisions on whether or not an alternative drug should be used. This was achieved by re-defining the study end points for assessing antimalarial drug efficacy *in vivo* (Tables 2.1a and b).

The WHO first developed a system of grading the parasitological response of *P. falciparum* over seven, 14 or 28 day intervals after a standard dosing regimen of chloroquine was administered over three days (WHO, 1973). Response to the drug was then categorised as S, RI, RII or RIII. The response curves obtained from plotting the daily parasitaemia load in study participants became the grading system in the standard field test, which was used by in-country malaria control programmes to assess the efficacy of antimalarials. The standard field test measured parasitological response of *P. falciparum* daily in non-immune people after antimalarial treatment over seven days. Little attention was given to clinical response or immunity although the study participant’s clinical condition was of primary importance (WHO, 1973). Due to the short duration of the seven day follow-up period, the difference between an S and RI response could not be determined easily. This was addressed with the extended test. Rieckmann (1990) provided a simplified version of the standard seven day test which measured parasite density on three days (0, 2, 7) rather than every day.

With the first protocols circa 1973, observations of high parasitaemia loads within the first two days of the follow-up period indicated a high grade resistance (RIII) whilst latter day parasitaemia indicated low grade resistance (RII or RI). As is evident from this grading
system, the conventional indicator of antimalarial resistance at that time was the observed persistence of parasites during the follow-up period (WHO, 2005; WHO, 2006).

Briefly, the new 1996 test system consisted of clinical examination, measurement of axillary temperature (on days 1, 2, 3, 7 and 14) and parasitological assessment (on days 2, 3, 7 and day 4), measured haematocrit and haemoglobin as well as supervised administration of the antimalarial. The protocol also defined specific inclusion and exclusion criteria for participation in a study.

White (2002) described the major limitations of the 1996 protocol, highlighting the fact that the early treatment failures (ETFs) specified fever with parasitaemia on day three post-treatment, even though non-immune children could still harbour parasitaemia despite having a drop in their parasite loads. Conversely an adequate clinical response (ACR) could describe a person with no fever but who may still have persistent parasitaemia (White, 2002).

Of further importance, was the need to differentiate between those parasites that were recrudescent or new infections to ascertain whether there was true resistance or not. Recrudescence is caused by survival of blood stage parasites bringing on a renewed manifestation of the original infection (WHO, 1963). There is a shortened time between the administration of a first line treatment and recrudescence, as the degree of drug resistance increases. Cure of malaria infection and hence treatment success could be achieved if the antimalarial prevented parasite recrudescence (WHO, 2006).

This is different from a re-infection, which is an entirely new infection that appears after all the original infection’s parasites have been cleared in a person (WHO, 1963). During the early days of resistance detection it was not possible to differentiate between recrudescent and new infections by microscopy alone. Detection of recrudescent and new infections became feasible as polymerase chain reaction (PCR) techniques developed, allowing the identification of genetic markers of resistance. Currently *P. falciparum* merozoite (MSP1, MSP2) and the glutamate-rich protein (GLURP) are established molecular genotyping markers used to distinguish between recrudescent and new infections. However standardised definitions of recrudescent and new infections based on these markers are required as the characterisation also depends on the method of genotyping (WHO, 2002).
The WHO 1996 protocol needed updating by 2001 primarily because many countries in sub-Saharan Africa had already started using antimalarial combination therapies rather than monotherapies (WHO, 2002). The subsequent WHO (2001) protocol was ratified at the end of 2001 and published in 2002 (WHO, 2002). Within this protocol children younger than five years of age in all types of endemicity areas were targeted. That age group is more likely to have an unsatisfactory response to antimalarial drugs compared to other age groups, even in areas with acquired immunity. If there were difficulties in recruiting young children, all age groups could be enrolled but it was recommended that results be stratified by age. Parasitological response was included in the 2001 protocol as part of the clinical response (ACPR) assessment. Late treatment failures (LTF) were categorised into late clinical failure (LCF) and late parasitological failure (LPF) in the 2001 protocol which characterised clinical and parasitological response separately. In addition molecular assessment to distinguish recrudescence from re-infections were included in the 2001 protocol. The protocol was modified in 2005 to accommodate high transmission areas. The changes related to the care of late parasitological failures and duration of follow-up with 42 days being optimal and 28 days being adequate (WHO, 2005).

During the previous decade, the revisions to the WHO guidelines for conducting in vivo efficacy studies were targeted at different endemicity settings but the recommendation in 2009 used the same definitions of classification endpoints for all transmission settings, with the intention of using only effective treatments to reduce morbidity and mortality associated with treatment failure and development of severe malaria (WHO, 2009b; WHO, 2010c). In 2009 the WHO recommended that molecular markers be used simultaneously with therapeutic efficacy tests to assess antimalarial resistance and these surveys should be conducted every year (WHO, 2009b).

As mentioned previously each WHO guideline for antimalarial drug efficacy was designed to standardise the methods applied in field tests and to improve the comparability of results across borders and time periods. However protocols were also sometimes modified to accommodate local conditions at sentinel sites. Some studies reported using WHO (1994) and WHO (2003) protocols, which are not strictly WHO protocols which would have well defined outcomes. Thus these types of user specific protocols are excluded from Tables 2.1a and b. According to the WHO (1994) report the role of the classification of parasitological response is to assess whether or not drug resistance is responsible for a reduced clinical response. The WHO (1994) report defined three categories (A, B & C) of parasitological response within a seven day follow-up period rather than the traditional S-RIII system.
Thus depending on the in vivo protocol used, different classification end points are reported (Tables 2.1a and b) and these influence the direct comparability of the results generated. The term “study end points” used here refer to the response categories defined for each protocol, based on the drug efficacy in vivo. Tahar and Basco (2007) highlighted the inconsistent results that occur due to the use of various protocols and different definitions of antimalarial treatment failure. The differences arise due to the exclusive or combined use of clinical and parasitological measurements, the duration of follow-up and whether or not there was any adjustment for recrudescent and new infections.

Importantly, the use of either parasitaemiac or clinical measures as the primary indicator of efficacy had implications for malaria treatment in Africa (Sibley and Hunt, 2003). For example, if clinical response was the primary measure in high transmission areas, people might still harbour parasites which recrudesced despite resolution of clinical symptoms during the specified follow-up period. According to Machekano, Dorsey and Hubbard (2007) the success or failure of an antimalarial drug in removing parasites can be clinically undetermined whilst Rieckmann (1990), stated that parasitaemia is a more precise indicator of treatment response compared to clinical measurements alone. Indeed some studies used parasitaemia as the primary measure of response in asymptomatic or symptomatic people on the seventh day of follow-up (Tahar and Basco, 2007).

Price et al., (2007) stated that individual investigators will likely continue to use varying study designs and the challenges of comparing the treatment efficacies from different studies will persist. Thus consideration should be given as to what information might be useful to compare such studies within these constraints. As the advent of molecular genotyping proved to be a useful new tool to detect antimalarial resistance, the WHO recommended that surveillance of antimalarial resistance in countries of decreasing malaria transmission should include tests for both in vivo efficacy and molecular markers and that the latter be used to confirm the results of in vivo efficacy studies (WHO, 2009b, WHO, 2011a).

Recommendations for molecular methods to detect resistant mutations were included for the first time by a WHO working group on monitoring antimalarial drug resistance in 2001 (WHO, 2002). As mentioned above molecular markers were subsequently recommended as indicators of antimalarial resistance and this was an optional component of surveillance of antimalarial drug efficacy (WHO, 2009b). The recommendation was to collect blood spots on filter paper from people participating in regular therapeutic efficacy tests, on the
day of enrolment (day 0) and on the day of treatment failure within the follow-up period. The WHO specified that the method used to identify the point mutations should be described to allow standardisation of molecular methods among research groups (WHO, 2001; WHO 2009b). Furthermore it was recommended that the molecular marker assays should be conducted on an annual basis in these areas.
Table 2.1a Summary of study end points for WHO protocols (1965-2009)

<table>
<thead>
<tr>
<th>Protocol</th>
<th>Classification Endpoints</th>
<th>Follow-up days</th>
<th>Defining features</th>
</tr>
</thead>
<tbody>
<tr>
<td>1. WHO(1965)</td>
<td>Sensitive/Resistant Sensitive/Resistant</td>
<td>3 days up to 11 days</td>
<td>Resistance indicated by lack of reduction in parasite density or an increase in parasite density, persistence of patent parasitaemia or reappearance of asexual parasites with or without clinical symptoms within two months of first treatment.</td>
</tr>
<tr>
<td>2. WHO (1973)</td>
<td>Standard field test S/RI, RI, RII, RIII</td>
<td>7 days</td>
<td>Introduction of grading system with specific follow-up period. No clinical measures reported. Only parasitological outcomes reported.</td>
</tr>
<tr>
<td></td>
<td>Extended field test S,RI,RII,RIII</td>
<td>28 days</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Single dose/alternative test S/RI, RI, RII, RIII</td>
<td>7 days</td>
<td></td>
</tr>
<tr>
<td>3. WHO (1996)</td>
<td>adequate clinical response (ACR)</td>
<td>14 days</td>
<td>Inclusion and exclusion criteria specified. Targeted febrile infants and children aged 6-59 months in areas of high transmission. In low transmission areas, all age groups could be included. Clinical measures included and clinical failures reported.</td>
</tr>
<tr>
<td></td>
<td>early treatment failure (ETF)</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>late treatment failure (LTF)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Protocol</td>
<td>Classification Endpoints</td>
<td>Follow-up Days</td>
<td>Defining features</td>
</tr>
<tr>
<td>----------</td>
<td>-------------------------</td>
<td>----------------</td>
<td>------------------</td>
</tr>
<tr>
<td>4. WHO (2001)</td>
<td></td>
<td></td>
<td>Different protocols for high transmission and low/moderate transmission areas. Inclusion and exclusion criteria updated accordingly. Studies longer than 14 days duration should include PCR to distinguish recrudescence from re infection. Measured fever and day two blood slides for ETF. Focus on clinical cure. Any person older than five years included if it was difficult to get sufficient numbers of under- fives. History of fever could be used if reliable. Measured increase in fever on day three for ETF. Both clinical and parasitological measures used.</td>
</tr>
<tr>
<td>4.1 High transmission areas</td>
<td>adequate clinical and parasitological response (ACPR)</td>
<td>14 days and if possible 28 days</td>
<td></td>
</tr>
<tr>
<td></td>
<td>early treatment failure (ETF)</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>late clinical failure (LCF)</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>late parasitological failure (LPF)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>4.2 Low/moderate transmission areas</td>
<td>adequate clinical and parasitological response (ACPR)</td>
<td>28 days</td>
<td></td>
</tr>
<tr>
<td></td>
<td>early treatment failure (ETF)</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>late clinical failure (LCF)</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>late parasitological failure (LPF)</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
As mentioned in chapter 1, the enzymes DHPS and DHFR from *P. falciparum* are essential to the folate biosynthesis pathway and are the target of sulphadoxine and pyrimethamine respectively (Peterson, Walliker and Wellems, 1988; Triglia and Cowman, 1994). In the late 1980s and 1990s specific point mutations of the *dhfr* and *dhps* gene encoding DHFR and DHPS were implicated in SP resistance (Cowman *et al*., 1988; Peterson *et al*., 1988; Brooks *et al*., 1994; Triglia and Cowman, 1994). Mutations at seven *dhfr* codons (16, 50, 51, 59, 108, 140 and 164) and five *dhps* codons (436, 437, 540, 581 and 613) have since been isolated as molecular markers for SP resistance. The amino acid substitutions and the underlying sequence changes in the three letter code for each codon are listed in Table 2.2.
Table 2.2 Molecular markers for *dhfr* and *dhps*

<table>
<thead>
<tr>
<th>dhfr Codon</th>
<th>16</th>
<th>50</th>
<th>51</th>
<th>59</th>
<th>108</th>
<th>140</th>
<th>164</th>
</tr>
</thead>
<tbody>
<tr>
<td>Sensitive</td>
<td>GCA</td>
<td>Cys(C)</td>
<td>Asn(N)</td>
<td>Cys(C)</td>
<td>Ser(S)</td>
<td>Val(V)</td>
<td>Ile(I)</td>
</tr>
<tr>
<td></td>
<td>TGT</td>
<td></td>
<td>AAT</td>
<td>TGT</td>
<td>AGC</td>
<td>GTT</td>
<td>ATA</td>
</tr>
<tr>
<td>Resistant</td>
<td>Ser(S)</td>
<td>Arg(R)</td>
<td>Ile(I)</td>
<td>Arg(R)</td>
<td>Asn(N)</td>
<td>Leu(L)</td>
<td>Leu(L)</td>
</tr>
<tr>
<td></td>
<td>TCA</td>
<td>CGT</td>
<td>ATT</td>
<td>CGT</td>
<td>AAC</td>
<td>CTT</td>
<td>TTA</td>
</tr>
<tr>
<td></td>
<td>Val(V)</td>
<td>GTA</td>
<td></td>
<td></td>
<td>Thr(T)</td>
<td>ACC</td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>dhps Codon</th>
<th>436</th>
<th>437</th>
<th>540</th>
<th>581</th>
<th>613</th>
</tr>
</thead>
<tbody>
<tr>
<td>Sensitive</td>
<td>Ser(S)</td>
<td>Ala(A)</td>
<td>Lys(K)</td>
<td>Ala(A)</td>
<td>Ala(A)</td>
</tr>
<tr>
<td></td>
<td>TCT</td>
<td>GCT</td>
<td>AAA</td>
<td>GCG</td>
<td>GCC</td>
</tr>
<tr>
<td>Resistant</td>
<td>Ala(A)</td>
<td>Gly(G)</td>
<td>Gly(E)</td>
<td>Gly(G)</td>
<td>Ser(S)</td>
</tr>
<tr>
<td></td>
<td>GCT</td>
<td>GGT</td>
<td>GAA</td>
<td>GGG</td>
<td>TCC</td>
</tr>
<tr>
<td></td>
<td>Cys(C)</td>
<td>TGT</td>
<td></td>
<td></td>
<td>Thr(T)</td>
</tr>
<tr>
<td></td>
<td>Phe(F)</td>
<td>GTA</td>
<td></td>
<td></td>
<td>ACC</td>
</tr>
</tbody>
</table>

The first step in genotyping *dhfr* and *dhps* point mutations is to collect a blood sample by finger prick and transfer it onto commercially available treated filter papers or cards (WHO, 2007, Collins *et al.*, 2006). The air dried blood spots can then be used to extract parasite DNA. Early laboratory methods for molecular detection of *P. falciparum* malaria were formulated in the 1980s and involved hybridisation of specific probes targeting DNA (WHO, 1986; Singh, 1997, Steenketse *et al.*, 2009). Since then advances have been made in refining high throughput genotyping protocols involving PCR based assays.
It is now well documented that the presence of multiple polymorphisms correlates with declining efficacy of an antimalarial drug. Several studies have shown that as the number of mutations progressively accumulate so do the levels of resistance. Allelic haplotypes are combinations of these point mutations. Haplotypes with many mutations tend to be more highly resistant than those with few mutations. Further discussion of haplotypes is presented in chapters 4 and 5.

Large-scale surveys determining the prevalence of $\text{dhfr}$ and $\text{dhps}$ point mutations from filter paper blood spots can be achieved relatively easily and quickly especially if combined with the existing operational structures for conducting in vivo efficacy studies. This is evident from the growing popularity of molecular surveys in certain sub-Saharan African countries since the 1990s, which are being reviewed and quantified in this study. However assessment of molecular markers is still not routinely carried out across sub-Saharan Africa (Mockenhaupt et al., 2007) and for the first time regional and country-specific data gaps will be characterised by the work described in this thesis. By collating and georeferencing all published data on in vivo and molecular measures of SP resistance it is possible to summarise what is known about the appearance of SP resistance at sentinel sites and the spread of the mutations in $\text{dhfr}$ and $\text{dhps}$ that underpinned the emergence of SP resistance in Africa.

### 2.2 Aims and Objectives

The aim of chapter 2 is to describe the literature search and the in vivo and molecular data which it generated. The standardisation, georeferencing and input of these data into a relational database will be described. Coverage of the in vivo and molecular surveillance studies included in the database is described and country-specific gaps are identified.

### 2.3 Methods

#### Search strategy

Online searches for published studies containing the free text term “malaria” were conducted between 2005 and 2011. The online databases available in Pubmed, African Healthline, African Journal Online, ISI web of Science, Bioline databases and Google
Scholar were cross-referenced for malaria related studies. From these, references with the following terms were selected: sulphadoxine or sulfadoxine or pyrimethamine or Fansidar or SP or PS or antifolate or dhfr or pfdhfr, or dihydrofolate reductase or dhps or pfdhps or dihydropteroate synthase. Specific exclusion and inclusion criteria were applied to further refine the selection of suitable studies for analysis and these are detailed in Table 2.3.

Table 2.3 Inclusion and exclusion criteria applied to the selection of *in vivo* studies

<table>
<thead>
<tr>
<th>Criteria for study selection</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Inclusion criteria</strong></td>
<td>SP monotherapy only where it was administered <em>in vivo</em> as the first line drug in clinical studies on pure <em>P. falciparum</em> infections in an African country. Study participants could be symptomatic or asymptomatic having uncomplicated malaria. Studies which might include cases of malaria imported between African countries.</td>
</tr>
<tr>
<td><strong>Exclusion criteria</strong></td>
<td>Any SP therapy administered as second line treatment, after another drug was administered as first line treatment. SP in combination with another drug. Participants with complicated or severe malaria. Prophylaxis, vaccine, <em>in vitro</em> and animal studies. Studies of travellers, malaria cases imported to non-African countries.</td>
</tr>
<tr>
<td><strong>Unrestricted criteria</strong></td>
<td>language of publication, study participants’ age or gender, malaria transmission intensity, duration of SP use in each African country, method of analysis of study outcomes (e.g. per-protocol or intention to treat)</td>
</tr>
</tbody>
</table>
The full texts of suitable *in vivo* studies were read and the data abstracted. This data collection strategy for reviewing SP resistance was part of the MARA project to collect data on antimalarial resistance in Africa and follows the strategy described elsewhere (MARA Technical Report, 1998; Omumbo and Snow, 2004). The molecular marker studies identified in the literature searches were read and abstracted in a similar manner as detailed below.

Data collection tools

**Terminology**

A 'study' refers to primarily a journal article or source document from which data were abstracted. A single study can contain data for multiple surveys conducted at different sites or at different time points.

A ‘survey’ refers to *in vivo* or molecular measures of resistance carried out at one particular time in one specific site. The data contained in each survey makes a single data point. The unit of analysis is a data point, which describes a unique location and a unique time point.

After applying the inclusion and exclusion criteria, suitable studies were abstracted by hand onto a predesigned template with specific data fields, called a Proforma. Thereafter these data were double entered into an electronic relational data entry system within Microsoft Access so that any variances that occurred during the double data entry process could be reconciled. Both tools are described in the following sections.

The Proforma

The Proforma (Appendix 1) consists of four major sections listed below:

**Section 1: Data reference**: A record of the names of the author(s) and the publication details of the study

**Section 2: Survey location**: This section described the geographic location including the country in which the study was done and the name of the study site. A small proportion of studies reported the latitude and longitude of the study sites. The latitude and longitude co-ordinates for the study site were geo-coded using Geoname digital gazetteer (GDE Systems, 1995), GEOnet names server (GNS) (National Geospatial Intelligence Agency, 1994), African Data Sampler (ADS) (World Resources Institute, 1995), the MARA
prevalence survey database or Google earth. Multiple sites from a single study were grouped if surveys were conducted at the same time and if the results were pooled. Any information on the drug policy as stated in the study was also recorded.

**Section 3: Antimalarial drug resistance:** The goal was to use specific data fields that would allow systematic recording of information while accommodating the different types of data that are reported by *in vivo* efficacy studies. This section was divided further into subsections for methods and results.

**Study methodology:** This sub-section listed the names and dosages of drugs studied, recorded the follow-up period, the study’s inclusion and exclusion criteria, PCR correction if used and WHO protocol where indicated.

**Study results:** This sub-section described the study population characteristics such as study start and end dates, reported malaria transmission season, number of participants enrolled and completing the study as well as the age of enrolled participants.

**SP efficacy:** This section recorded parasite clearance times and fever clearance times. It identified the number or proportion of participants classified according to their parasitological and/or therapeutic response. For studies spanning more than one year, the start year of the study was used in the analysis. Since individual patient-level data were not reported, the resistance outcome as reported in each study was maintained. A similar method was adopted by others looking at spatial and temporal trends of chloroquine resistance (Shah *et al*., 2011). There was no restriction based on the type of analytical approach used. In the per protocol approach, which was frequently used since 2000, any participant who was lost to follow-up for any reason would be dropped from the efficacy analysis (WHO, 2010c) whilst in the intention-to-treat analysis participants were classified as failures even though they might not represent true biological failures (Price *et al*., 2007).

**Molecular markers:** This section captured individual *dhfr* and *dhps* point mutations and combinations of point mutations (haplotypes). The number of samples tested for each mutation was recorded along with the number of samples that tested positive for both sensitive and mutant forms. From this, the prevalence for each sensitive and mutant form could be calculated. The type of method used for the molecular assay was also indicated. Both mixed mutant and pure mutant infections were included in this analysis as outlined in Table 2.2. The prevalence of each mutation was determined by calculating the proportion of infections containing that mutation relative to the total number of people sampled. Where available, the number of haplotypes found and total haplotype count were recorded separately. An inclusive approach was followed to include studies with both mutation prevalence and frequency data. Pre-and post-treatment samples were recorded separately, if available. The next step was to capture the data from the proforma into the data entry system.
Data Entry System

A relational database was designed using Microsoft Access to facilitate data entry and subsequent querying of fields for export and analysis. The system was structured to accommodate one-to-many relationships and vice versa. This facility was useful for querying multi-layered data i.e. one study could have information for different locations and methods as well as different drugs studied over varying time periods, thus producing many data points.

Data validation

The data were captured through a process of double entry into two separate databases. An error checking programme was developed to identify variances between the two databases. Any anomalies between the two databases were systematically identified and error reports were generated. The anomalies were corrected by referring to the data within the Proforma or source document if necessary. This process was continued until there were no variances between the first and second data entries.

Mapping within a Geographic Information System

Survey sites were geo-coded as point data and vector maps were produced in Mapinfo Professional ver 9.5 (Pitney Bowes Software Incorporated). The continental African boundaries were obtained from the ADS (World Resources Institute, 1995).

The 48 countries in sub-Saharan Africa are based on the Roll Back Malaria (RBM) classification of malarious countries in Africa where the majority of cases occur (RBM 2008) (Figure 2.1). This was updated to mark the independence of the Democratic Republic of South Sudan from North Sudan in July 2011 and the geographic boundaries for these two new countries are indicated separately as North and South Sudan on the maps. The North African countries of Algeria, Egypt and Morocco have residual malaria and hence were excluded from this geographic analysis. The Comoros island group includes Mayotte.
2.4 Results

This section describes the \textit{in vivo} and molecular data that were extracted during the data collection phase. The review to select source documents with SP monotherapy in Africa produced 313 studies (Table 2.4).
The studies emanated from 36 of 48 malarious sub-Saharan countries with English journal articles making up the majority of the source documents (Table 2.5). The time frame for these data was 1980-2007. Studies among children under five years of age accounted for 44% (n=313) whilst 45% of studies included participants who were older than five years of age. The remaining studies did not indicate the ages of the enrolled participants. There were varying durations of follow-up with the majority of the studies using 28 days (38%) and 14 days (35%). A short follow-up period of seven days was reported by 11% of the studies and the same proportion did not indicate any follow-up period. A long follow-up period of 42 days was indicated by 5% of studies.

SP in vivo Studies

Approximately equal proportions of studies (40%-41%) reported parasitological or therapeutic outcomes and 13% reported both measures while the remainder (6%) did not report a standard measure. Only 17% of the studies with 28 day follow-up corrected for recrudescence and new infections using PCR and 11% of these studies were conducting among children below five years of age.
### Table 2.5 Summary of studies with SP monotherapy in vivo data

<table>
<thead>
<tr>
<th>Description</th>
<th>Number of studies</th>
</tr>
</thead>
<tbody>
<tr>
<td>SP monotherapy</td>
<td>313</td>
</tr>
<tr>
<td>Number of African countries</td>
<td>36</td>
</tr>
<tr>
<td>Type of source document</td>
<td></td>
</tr>
<tr>
<td>Journal</td>
<td>268</td>
</tr>
<tr>
<td>Ministry of Health Report</td>
<td>28</td>
</tr>
<tr>
<td>Other type of Report</td>
<td>17</td>
</tr>
<tr>
<td>Language</td>
<td></td>
</tr>
<tr>
<td>English</td>
<td>296</td>
</tr>
<tr>
<td>French</td>
<td>17</td>
</tr>
<tr>
<td>Time frame during which studies were done at all sentinel sites</td>
<td>1980-2007</td>
</tr>
<tr>
<td>Studies with children under 5 years</td>
<td>138</td>
</tr>
<tr>
<td>Studies with participants over 5 years</td>
<td>142</td>
</tr>
<tr>
<td>Studies with 7 day follow-up period</td>
<td>33</td>
</tr>
<tr>
<td>Studies with 14 day follow-up period</td>
<td>111</td>
</tr>
<tr>
<td>Studies with 28 day follow-up period</td>
<td>118</td>
</tr>
<tr>
<td>Studies with 42 day follow-up period</td>
<td>16</td>
</tr>
<tr>
<td>Classified treatment response using parasitological measures – S/RI/RII/RIII</td>
<td>124</td>
</tr>
<tr>
<td>Classified treatment response using therapeutic measures – ACR/ACPR/ETF/LTF</td>
<td>128</td>
</tr>
<tr>
<td>Classified treatment response using parasitological and therapeutic measures</td>
<td>42</td>
</tr>
<tr>
<td>Surveys with 28 day or more follow-up and PCR correction for recrudescence/re-infection</td>
<td>53</td>
</tr>
<tr>
<td>Surveys with 28 day or more follow-up and PCR correction for recrudescence in children under five years of age (including mean age)</td>
<td>34</td>
</tr>
</tbody>
</table>
Standardisation of WHO Protocols

The most common *in vivo* efficacy outcome measures are described here. These outcome measures were identified by the WHO protocol used to describe them, where available. Some studies explicitly stated which WHO protocol was used in the study e.g. WHO 1994 and WHO 2003 whereas others did not. For the latter, the WHO protocol was inferred by the reported classification endpoints. Figure 2.2 summarizes the different protocols that were obtained after this process. Some studies (9%) used a combination of protocols such as WHO 1973 and WHO 1996. These were added to the protocol tally as individual protocols. The WHO 1973 and 1996 protocols were the most common, accounting for 34% and 32% respectively (n=326). The WHO 2001 protocol proved less popular (9%). Approximately 10% of the studies could not be assigned a protocol. In addition, adaptations of the major protocols were also used to accommodate different field conditions and these were extracted as they were reported by the study authors. Figure 2.2 shows the major protocols, highlighted by the red boxes and gives an indication of the scope of modified protocols. These modified protocols accounted for 15% of the studies.

![Figure 2.2 Standardised WHO Protocols for SP monotherapy studies](image)

**Figure 2.2 Standardised WHO Protocols for SP monotherapy studies**

NA = not assigned

Red boxes indicate major WHO protocols.
A sub set of *in vivo* studies from 21 countries where surveys were conducted among any age group with 28 day or longer follow-up period with PCR correction (Figure 2.3) were matched by geographic site and survey year and these are described in chapter 5. No data were available for Botswana, Burundi, Cape Verde, Central African Republic, Comoros/Mayotte, Djibouti, Equatorial Guinea, Eritrea, Ethiopia, Guinea, Guinea-Bissau, Ivory Coast, Liberia, Mauritania, Mauritius, Namibia, Niger, Nigeria, Rwanda, Sao Tome/Principe, Senegal, Somalia, Swaziland, Togo and Zimbabwe.

Figure 2.3 Countries where SP *in vivo* surveys were conducted among any age group with 28 day or longer follow-up period and where PCR correction was done

**Molecular studies**

After applying the inclusion and exclusion criteria, 153 *dhfr* and 126 *dhps* source documents were identified for review (Table 2.4).
Synopsis of methods used in molecular studies of SP

There were seven different methods used for genotyping \textit{dhfr} and \textit{dhps} mutations. The PCR Restriction Fragment Length Polymorphism (PCR-RFLP) was the most commonly used methodology for genotyping \textit{dhfr} and \textit{dhps} mutations (Figure 2.4). This and the other methods listed in Figure 2.4 are described below.

![Graph showing proportion of methods used](image)

Figure 2.4 Summary of mutation screening methods used among all studies of \textit{dhfr} and \textit{dhps}

The PCR-RFLP method was first described by Duraisingh \textit{et al.} (1998) and uses a nested PCR to amplify the \textit{dhfr} or \textit{dhps} sequence, followed by digestion with a panel of restriction enzymes. Each restriction enzyme digest is designed to identify the presence of a specific sequence variant. Different restriction enzymes recognise and cut specific sequences and the enzymes used in this method are selected to recognise known sequence variants of \textit{dhfr} and \textit{dhps} associated with SP resistance. The change in fragment size resulting from the digestion can be visualised easily using agarose gel electrophoresis and reference samples for comparison. The use of reference samples as controls allows mixed samples and incomplete digestion to be differentiated from one another.
Nested PCR amplification generates sufficient PCR product even from the small amounts of DNA template typically extracted from patient finger prick blood spots. The limits of detection for minority genotypes in mixed infections depend to some degree on the efficiency of the initial PCR amplification and ultimately on the visibility of fragments from the minor variant on an agarose gel.

Ranford-Cartwright et al., (2002) found the PCR-RFLP assay had a high specificity as it was able to correctly identify alleles that were absent in mixtures. There was a small difference in the sensitivity of PCR-RFLP, 10 parasites per sample could not always be detected at the lowest (0.002%) parasite densities.

Direct sequencing was the second most common methodology used. As with the previous method, the *dhfr* or *dhps* sequence is PCR amplified first and then subjected to direct sequencing. This method on its own is capable of detecting novel mutations which have not been previously described either at the commonly genotyped known codon positions or at any other codon position in the amplified fragment. The use of this methodology can sometimes elucidate new mutations, for example a novel I431V mutation of Nigerian origin was observed through analysis by PCR and direct DNA sequencing (Sutherland et al., 2009).

Nested PCR amplification generates sufficient PCR product for sequencing even from finger prick blood collected from patients with low density parasitaemia. The limit of detection for minority genotypes in mixed infections is determined by its peak height on the sequence electropherogram. Depending on the efficiency of sequencing and clarity of signal this may vary.

Like all other commonly used genotyping methods, the third commonly used technique of dot blotting begins by nested PCR amplification of the *dhfr* or *dhps* sequence. The amplified PCR products are spotted into a membrane, fixed onto the membrane by cross linking and hybridisation with oligonucleotide probes detects the presence of variant sequences. Sequence-specific oligonucleotide probes are designed to complement the known sequence variants in *dhfr* and *dhps* and fluorescence or radioactive labelling is used to detect their hybridisation to the PCR products fixed to the membrane. Three variations of these methods have been described in Abdel-Muhsin et al., (2002), Pearce et al. (2003) and Alifrangis et al., (2005). These authors used identical probe sequences but different methods of visualisation.
As mentioned previously, the nested PCR amplification generates sufficient PCR product even from quantities of blood obtained by finger prick extraction. The limit of detection for minor genotypes in mixed infections relies upon the stringency of hybridisation conditions in each experiment. A high background level of hybridisation can obscure minor genotypes which would otherwise be detectable in high stringency conditions. Abdel-Muhsin et al., (2002) reported that their dot blotting technique was specific, sensitive and robust. They detected dhfr alleles in <100 parasites/µl blood and distinguished minority genotypes at very low proportions (1%) in mixed clonal infections.

Another PCR specific amplification method was described by Plowe et al., (1995) and Doumbo et al., (2000). As with the other methodologies nested PCR is used to amplify the variant dhfr and dhps sequences. However in this method the primers in the second round of PCR are designed to be variant specific. When the primer complements a variant sequence in the DNA sample, a reaction is primed and a PCR product created. If it does not complement a variant sequence, then no fragment is amplified.

This PCR specific amplification methodology is reliant entirely on the stringency of the PCR conditions because false positives might arise from primers annealing to non-target sites (mis-priming). Therefore the threshold of detection is determined with reference to the specificity of the nested PCR reaction. Furthermore, this method assumes the sequence for the sensitive dhfr 50C codon in their primer for the dhfr 51 codon. This practice is acceptable in Africa as currently mutations at dhfr 50R have only been found in South America (Cortese et al., 2002).

The less common genotyping techniques were nucleotide sequencing, automated scoring of single nucleotide polymorphisms, use of micro arrays and mass spectrometry. The fifth method that was reported refers to a method of dhfr and dhps single nucleotide polymorphism (SNP) genotyping by primer extension designed by Nair et al. (2002). Genotyping was performed by primer extension and the products scored on a capillary sequencer by fragment size analysis. Multiple mutations can be genotyped in a single reaction and scored concurrently, which is the main advantage of this primer extension method. Multiple infections can be relatively easily characterised by primer extension using the different migration properties of oligonucleotides. Nair et al., (2002) did not conduct a sensitivity and specificity analyses of this method in differentiating among mixed and pure infections and stated that multiple infections could be distinguished using artificial mixtures of parasites with known alleles.
The micro array technique involves nested PCR, SNP primer extension and then sequence specific hybridisation and scanning by micro array as described by Crameri et al., (2007). The method is based on parallel PCR amplification of dhfr and dhps sequences, followed by primer-extension-mediated mini sequencing. Sequence variants are then detected by sequence-specific hybridization on a microarray.

This technique is sensitive to the DNA quality used and requires a suitably large enough amount of DNA template of low parasite density cases. The authors reported that the method is sufficiently sensitive for clinical malaria cases and most asymptomatic cases, as it was able to detect approximately 15 parasites/µl (Crameri et al., 2007). Furthermore the estimated specificity of the test was 94% compared to the gold standard of sequencing.

The final method listed in Figure 2.4 is matrix assisted laser desorption ionisation-time of flight mass spectrometry-based analysis (MALDI-TOF MS), referred to as mass spectrometry and it is based upon primer extension as described in Marks et al.(2004) and Marks et al. (2005). The results of MALDI-TOF MS were compared to those obtained by DNA sequencing and this method was more sensitive to the detection of minor genotypes in mixed infections (Marks et al., 2004).

**Relative accuracy of genotyping methodologies**

Generally, the accuracy of genotyping using these various methodologies was compared to direct sequencing as a gold standard, with the exception of the microarray technique which was compared to RFLP. When comparing these methods, their ability to differentiate between single mutations and mixed infections is important. When samples contain a single genotype the results of the different genotyping methods are highly consistent. However, the results obtained from genotyping mixed infections are less repeatable due to the detection accuracy of minor sequence variants rather than the detection accuracy of the majority genotype. It is estimated that minor genotypes are likely to be out-competeted in the first few rounds of amplification if they are less than 1:10 ratio with the majority sequence (Contamin et al., 1995).

Since all the commonly used genotyping methodologies begin with PCR amplification, it may be argued that all methods are similarly constrained in the extent to which they are likely to amplify minor genotypes in mixed infections. However, one exception may be the
PCR specific amplification method which has primers specifically designed to target variant sequences even if it is significantly outnumbered. This may make it more sensitive than other methods in detecting minor sequence variants in mixed infections.

Ranford-Cartwright et al., (2002) compared PCR-RFLP, dot blotting and mutation specific PCR and recommended the dot-blot hybridisation technique for large scale genotyping surveys. The dot blot technique proved to be similarly specific and sensitive to PCR-RFLP, but it was recommended instead of the other two assays on the basis of consistency of the results generated using this method.

There are no experiments to directly compare the sensitivity of all the methods described here, specifically in their detection of minor genotypes from mixed infections. In combining molecular surveillance data from many laboratories using different methods, equipment, reagents and research questions it is necessary to be aware of potential systematic confounders.

If for example direct sequencing is less sensitive to the detection of minor genotypes in mixed infections and there has been a trend toward increasing use of this method in molecular surveillance overall, then caution should be exercised in analysing temporal trends in prevalence measures. The difference between prevalence and frequency measures are discussed in Chapter 4. Briefly, prevalence measures the rate of detection of any mutation among a number of samples and includes major and minor variant sequences detected. Frequency estimates which take only one sequence variant per sample (in mixed infections the majority sequence) are not vulnerable to this problem, and this measure is more robust to variation in methodological sensitivity, transmission intensity and age-dependent effects on multiplicity of infection in different transmission settings.

The maps below show the coverage of \textit{dhfr} and \textit{dhps} point mutations and allelic haplotypes and highlight where gaps exist in the coverage of these.
Distribution and coverage of \textit{dhfr} and \textit{dhps} in sub-Saharan Africa

For \textit{dhfr} I164L, 184 sites (Figure 2.5) were surveyed and there are gaps in 37.5\% of countries (Figure 2.5, Table 2.6).

![Map showing distribution of I164L surveys](image)

Figure 2.5 Sites surveyed for \textit{dhfr} I164L
There were 232 surveys for *dhps* A437G extending across 39 African countries with data gaps in Burundi, Chad, Eritrea, Sierra Leone, Somalia, Togo (Figures 2.6 & 2.14 and Table 2.6). The coverage of *dhps* 437 ranked second best among all codons studied.

![Distribution of A437G surveys](image)

*Figure 2.6 Sites surveyed for *dhps* A437G*
For *dhps* K540E 261 sites were surveyed (Figure 2.7). The coverage for *dhps* K540E was the best among all codons studied but there were no data at all in nine countries (Figure 2.14, Table 2.6).

[Figure 2.7 Sites surveyed for *dhps* K540E]
The *dhps* A581G point mutation was surveyed in 121 sites across Africa. Figure 2.8 and Table 2.6 shows the data gaps in 19 countries for this point mutation.

---

**Figure 2.8** Sites surveyed for *dhps* A581G
Coverage of *dhps* S436A spanned 153 study sites (Figure 2.9). Gaps were observed in 12 countries namely Benin, Botswana, Burundi, Cape Verde, Chad, Eritrea, Liberia, Mauritius, Niger, Sierra Leone, Somalia and Togo (Table 2.6).

Figure 2.9 Sites surveyed for *dhps* S436A
The *dhps* A613S/T codon was surveyed in 92 sites as illustrated in Figure 2.10. There were no *dhps* A613S in 20 countries and no A613T data in 22 countries (Figure 2.14, Table 2.6). Generally there was lesser coverage of *dhps* A613S/T compared to the other point mutations.

![Figure 2.10 Sites surveyed for dhps A613S/T](image)

*Figure 2.10 Sites surveyed for dhps A613S/T*

*Note: dhps A613T is absent in Gabon and Ghana.*
Coverage of the *dhfr* IRN and *dhps* GE/GK/AK allelic haplotypes were comparable to that of the individual point mutations. The *dhfr* triple mutant haplotype was surveyed in 197 sites (Figure 2.11). No data were found for 13 countries namely Benin, Botswana, Burundi, Cape Verde, Chad, Djibouti, Equatorial Guinea, Eritrea, Mali, Mauritius, Sierra Leone, Somalia and Togo (Figure 2.14, Table 2.6).
The *dhfr* NCS sensitive allelic haplotype was surveyed in 195 sites (Figure 2.12). No coverage was found in 13 countries namely Chad, Equatorial Guinea, Burundi, Djibouti, Eritrea, Somalia, Botswana, Mauritius, Mali, Benin, Sierra Leone, Togo, Cape Verde (Figure 2.14, Table 2.6).
The *dhps* 437 and 540 GE, GK and AK haplotypes were monitored in 152 sites (Figure 2.13). No coverage was found in 11 countries namely Cape Verde, Chad, Burundi, Djibouti, Eritrea, Somalia, Botswana, Mauritius, Zimbabwe, Sierra Leone and Togo (Figure 2.14, Table 2.6).

After reviewing the maps of molecular markers, consistent data gaps were identified in nine countries namely, Botswana, Burundi, Cape Verde, Chad, Eritrea, Mauritius, Somalia, Sierra Leone and Togo. The dataset for *dhps* A437G and K540E (GE) had the least amount of country data gaps with no data for 19% of malarious sub-Saharan countries (Figure 2.14). At the country level, surveillance of mutations at *dhfr* 164 and other *dhps* alleles were relatively poor compared with the other alleles (Table 2.6). No SP *in vivo* or molecular resistance data were found in the following seven countries: Botswana, Burundi, Cape Verde, Eritrea, Mauritius, Somalia and Togo.
Table 2.6 Gap analyses: countries where no data were found for each point mutation and allelic haplotype

<table>
<thead>
<tr>
<th>Codon</th>
<th>Number of Countries (%)</th>
<th>Countries</th>
</tr>
</thead>
<tbody>
<tr>
<td>dhfr 164</td>
<td>18 (37.5)</td>
<td>Angola, Benin, Botswana, Burundi, Cape Verde, Chad, Eritrea, Gambia, Guinea, Guinea-Bissau, Liberia, Mauritius, Namibia, Niger, Sierra Leone, Somalia, Togo, Zimbabwe</td>
</tr>
<tr>
<td>dhps 436</td>
<td>12(25)</td>
<td>Benin, Botswana, Burundi, Cape Verde, Chad, Eritrea, Liberia, Mauritius, Niger, Sierra Leone, Somalia, Togo</td>
</tr>
<tr>
<td>dhps 437</td>
<td>9 (18.8)</td>
<td>Botswana, Burundi, Cape Verde, Chad, Eritrea, Mauritius, Sierra Leone, Somalia, Togo</td>
</tr>
<tr>
<td>dhps 540</td>
<td>9 (18.8)</td>
<td>Botswana, Burundi, Cape Verde, Chad, Eritrea, Mauritius, Sierra Leone, Somalia, Togo</td>
</tr>
<tr>
<td>dhps 581</td>
<td>19 (39.6)</td>
<td>Benin, Botswana, Burkina Faso, Burundi, Cape Verde, Chad, Eritrea, Gambia, Guinea, Guinea-Bissau, Liberia, Mauritius, Namibia, Nigeria, Sierra Leone, Somalia, Togo, Zambia, Zimbabwe</td>
</tr>
<tr>
<td>dhps 613S</td>
<td>20 (41.7)</td>
<td>Benin, Botswana, Burkina Faso, Burundi, Cape Verde, Chad, Eritrea, Gambia, Guinea, Guinea-Bissau, Liberia, Mauritius, Namibia, Nigeria, Sierra Leone, Somalia, Togo, Zambia, Zimbabwe</td>
</tr>
<tr>
<td>dhps 613T</td>
<td>22 (45.8%)</td>
<td>Benin, Botswana, Burkina Faso, Burundi, Cape Verde, Chad, Eritrea, Gabon, Gambia, Ghana, Guinea, Guinea-Bissau, Liberia, Mauritius, Namibia, Niger, Nigeria, Sierra Leone, Somalia, Togo, Zambia, Zimbabwe</td>
</tr>
<tr>
<td>dhfr IRN</td>
<td>13 (27.1%)</td>
<td>Benin, Botswana, Burundi, Cape Verde, Chad, Djibouti, Equatorial Guinea, Eritrea, Mali, Mauritius, Sierra Leone, Somalia, Togo</td>
</tr>
<tr>
<td>dhfr NCS</td>
<td>13 (27.1%)</td>
<td>Cape Verde, Chad, Equatorial Guinea, Burundi, Djibouti, Eritrea, Somalia, Botswana, Mauritius, Mali, Benin, Sierra Leone, Togo</td>
</tr>
<tr>
<td>dhps GE/GK/AK</td>
<td>11 (22.9%)</td>
<td>Cape Verde, Chad, Burundi, Djibouti, Eritrea, Somalia, Botswana, Mauritius, Zimbabwe, Sierra Leone, Togo</td>
</tr>
</tbody>
</table>
Figure 2.14 Gap analysis: Molecular surveillance coverage of all dhfr and dhps point mutations from studies in sub-Saharan Africa.
2.5 Discussion

Previous attempts at geographic mapping of antimalarial resistance have used different measures of resistance in a variety of presentations. As mentioned in chapter 1 the seminal work mapping the global spread of chloroquine resistance was completed in 1989 and contained data for 33 African countries from 79 sites in a sequence of hand drawn maps (Payne, 1989). Much later the WHO released online global maps containing summaries of antimalarial drugs including chloroquine and SP efficacy results, expressed as treatment failure rates (\%) (WHO, 2010c). These maps are available at http://apps.who.int/globalatlas/. However, no synopsis of antimalarial efficacy data and molecular markers has been done at a continental level. In this chapter a review of available in vivo efficacy data and molecular markers of SP are presented and gaps in the data are highlighted.

Different approaches might be used when reviewing and reporting data from different studies for comparison purposes. Systematic reviews can be performed with or without meta-analysis depending on the study objectives and type of information required. Shah et al., (2011) conducted a systematic review without a meta-analysis with the aim of showing overall trends in chloroquine resistance in India. Strict inclusion criteria would have precluded many in vivo studies and this can result in very poor data coverage in mapping studies. There is therefore a trade-off between data coverage and quality which must be resolved to optimise coverage while also creating a clean and well-standardised dataset. In this study a general review of studies was undertaken in the first instance and more specific inclusion and exclusion criteria were applied to select a subset of these studies for further specific analysis. In reviewing studies of SP and chloroquine for the purpose of identifying general geographic trends, Talisuna et al., (2004a) also did not apply strict inclusion and exclusion for study selection as prescribed by standard meta-analyses methods for randomised studies. These authors stated that if such strict criteria were to be applied it would preclude many in vivo studies.

The 118 SP monotherapy studies with 28 day follow-up that were identified in this review of primary published work is comparable to the 90 reported for Africa from the WHO global database on antimalarial drug efficacy (WHO, 2010c). The WHO collated data from studies done between 1996 and June 2010 for SP with a minimum 28 day follow-up. In this thesis the review of dhfr and dhps point mutations and allelic haplotypes produced in excess of 120 journal articles. This amount is comparable to the 100 articles on dhfr and dhps found from 31 African countries prior to 2009 by Sridaran et al., (2010). When taking into consideration both in vivo and molecular data for SP resistance, gaps were identified
consistently in the following seven countries: Botswana, Burundi, Cape Verde, Eritrea, Somalia, Togo and Mauritius. These gaps could represent a genuine absence of resistance surveillance activity in these countries or perhaps just a lack of availability of surveillance data in the public domain.

The WHO standard efficacy test provides the minimum essential information for deciding on a malaria treatment policy (WHO, 2009b; WHO, 2005; Vestergaard and Ringwald, 2007). Studies with the basic design formed the basis of surveillance for monitoring changes in drug efficacy over time and were useful if they were conducted in many appropriately selected sentinel sites. As described earlier in this chapter, the standard efficacy test protocols changed over the years to accommodate the different antimalarial drugs and the different therapeutic response categories also adapted to reflect the growing knowledge about antimalarial treatment failure and resistance.

As the goal of these tests were to standardise the classification endpoints to assess therapeutic failure or success, it is useful to quantify the proportion of studies that adhered to a specific WHO protocol as described here.

A fair degree of variation was observed in the utilisation of WHO protocols for SP in vivo studies and in some cases combinations of protocols were used in the same study. This finding is consistent with outcomes of other reviews, for instance, after completing a review of 108 chloroquine studies in Africa over a seven year period (1996-2002) Talisuna et al., (2004a) also observed marked variability in the application of WHO protocols for antimalarial in vivo efficacy tests. This variability is indicative of the varying operational conditions in sub-Saharan African countries. In a similar review in India Shah et al. (2011) found variation in the classification endpoints of in vivo efficacy studies and they resolved this by maintaining the endpoints described as they were reported and by forming a composite definition of treatment failure using per protocol for their spatial and temporal analysis. This composite definition of treatment failure allowed comparison of results from different protocols.

The WHO mentioned the difficulty in comparing study results from the WHO 1973 protocol in high transmission areas (WHO, 2005). This protocol used S/RI/RII/RIII parasitological responses. It was recommended only in low or moderate transmission areas where the parasitological response categories (RI+RII+RIII) from the WHO 1973 protocol could be related to (ETF+LCF+LPF) from the WHO 1996 and 2001 protocols (WHO, 2005, WHO
2010c). As mentioned previously White (2002) discussed the limitations of the WHO 1996 14 day test, stating that this protocol was actually applied both in high and low transmission areas. This illustrates the variability in the application of these tests and provides support for the use of a composite definition of treatment failure which will be presented in chapter 5, to optimise the available data for analysis of resistance patterns.

The data obtained through conducting in vivo efficacy tests within sub-Saharan Africa have been used as the primary evidence which has informed decisions on drug policy and they remain an important component in defining the antimalarial resistance profile in each country. However, with the advent of tools to measure the genetic basis of resistance, surveillance of molecular markers has become increasingly important in policy decision making. The mapped coverage of molecular markers indicated in this chapter shows that monitoring of molecular markers to detect resistance in Africa has gained momentum in Africa. However many gaps exist. There is a dearth of SP molecular data in Burundi, Chad, Sierra Leone, Somalia and Togo. The Southeast African region representing Kenya and Tanzania amongst others, has the best overall coverage of SP molecular data. The review of the dhfr triple and dhps double mutants in Africa by Sridaran et al. (2010) also found a high proportion of studies done in East African countries like Kenya and Tanzania.

Unlike the standardised protocols for in vivo efficacy studies, there are no equivalent standardised protocols for analysis and reporting of molecular markers. Whilst a variety of molecular methods are used, PCR-RFLP emerged as the most commonly used method in this review, confirming the view of Mockenhaupt et al., 2007. The PCR assay using a RFLP probe (PCR-RFLP) is a relatively simple and rapid method (de Pecoulas et al., 1995). Menard et al., (2006) working in Central African Republic stated that PCR-RFLP is a robust method which has been widely used in other African countries whilst Mockenhaupt et al., (2007) also indicated that PCR-RFLP was the most common technique for dhfr typing.

There is no reason to suppose that different methods would generate contradictory results, although there might be varying test sensitivities and specificities. The complicating factor is mixed infections (where one person is simultaneously infected with multiple genotypes), and there is still a need for a standardised framework for reporting molecular data. Two metrics to standardise point mutation and allelic haplotype data across many sites will be described and employed in the spatio-temporal analysis presented in chapter 3. The WWARN collaboration (http:www.wwarn.org) has created a platform for sharing experiences, protocols and data with the aim of assisting under resourced laboratories to
conduct molecular genotyping studies. It is envisaged that this initiative will contribute to streamlining such data. Importantly it is necessary to clarify the types of measures of molecular data which are meaningful for policy makers. This will be discussed further in chapter 4. Standardisation of the molecular measures will alleviate some of the comparability issues experienced with in vivo efficacy data.

If there are large local differences in the prevalence of molecular markers, such as that observed in the DRC, to adequately assess the antimalarial drug resistance profile in such a country would require data generated from many molecular studies at various sites. It was proposed that existing drug resistance networks carry out population-based surveys among 200–250 randomly selected participants at various geographically dispersed sites with different drug use practices and malaria transmission rates (Talisuna et al., 2004b).

Generally it is difficult to gain access to patient level data at a country level. The WWARN collaboration has thus far obtained SP data for a few studies from Angola, Chad, DRC, Mali, Mozambique, Sierra Leone, Uganda and South Africa and efficacy rates are available online only in a summarised form (http://www.wwarn.org/resistance/explorer).

Limitations

There are several limitations to using summarised data with the primary one being the diversity in the methods including varying follow-up periods, use of asymptomatic or symptomatic participants and outcome measures reported from in vivo efficacy studies. In extracting data from these in vivo efficacy studies, it was assumed that the summarised treatment outcomes were identified correctly as they were collated from primarily published studies, which would have passed through peer-review. Furthermore summarised data from various different time frames and spatial scales are restrictive in the types of analyses that can be performed and requires caution when making inferences about these data or when interpolating sparse data in regions. Within these constraints, this thesis aims to characterise the spatial and temporal trends of SP resistance, beginning with chapter 3 where the analyses of point mutations are presented.
3.1 Introduction

*P. falciparum* parasites are genetically complex single celled organisms which are haploid for most of their development and diploid only during the zygote stage in the mosquito (Gilles, 1993b). The haploid blood stage of *P. falciparum* has genetic markers. Single nucleotide polymorphisms (SNPs) or polymorphic microsatellites are useful for genetic studies showing how drug resistance evolves, emerges and spreads. The appearance and accumulation of mutant SNPs in *dhfr* and *dhps* genes have been associated with tolerance to SP treatment. Although SP was once the standard treatment for clinical *P. falciparum* malaria, due to the spread of SP resistance which systematically eliminated sensitive genotypes, SP is no longer recommended as treatment. However, SP continues to be used for IPTp and it is recommended for infants as IPTi where the drug is still efficacious (WHO, 2009a).

Assessment of the prevalence of *dhfr* and *dhps* mutant SNPs among infected individuals sampled at specific sites provides an indication of the local SP resistance rate. Point mutations must be transmitted at a faster rate than the sensitive forms, in order for them to displace the sensitive forms and spread successfully in a population, increasing their prevalence and ultimately reaching fixation where only the mutated form remains (Malisa et al., 2010). Molecular studies using the same methodologies, conducted at similar times but at different sites have found varying point mutation prevalence rates (Mugittu et al., 2004; Alker et al., 2008; Mbacham et al., 2010). Predictably there have been observations of heterogeneity in the point mutation prevalence at different sentinel sites.

Possible biological explanations for the emergence, spread and subsequent heterogeneity in point mutation prevalence have been put forward. One theory is that point mutations could have emerged *de novo* due to varying degrees of drug selection pressure and transmission intensities at those sites, whilst another theory is that the point mutations could have arrived from other sites at different times through gene flow (Hastings, 2004; Anderson et al., 2003; Nair et al., 2003; Roper et al., 2003; Roper et al., 2004). Furthermore, ethnographic reasons influence the movement of people across porous national and international borders, thereby contributing to the mixing of parasite populations and dispersal of resistant parasites.
Generally, the degree of SP resistance is proportional to the number of point mutations found in *dhfr* and *dhps* genes (Gregson and Plowe, 2005). Thus, point mutations confer varying degrees of resistance to SP either individually or when found in combination with other point mutations. A brief synopsis of each of the following point mutations is described next: *dhfr* 164L, *dhps* 436A, *dhps* 437G, *dhps* 540E, *dhps* 581G and *dhps* 613S/T.

The *dhfr* 164L mutation was identified in Southeast Asia during the late 1980s (Sirawaraporn *et al*., 1997). Although it has subsequently emerged at some sites in Africa, it is still relatively rare (McCollum *et al*., 2006; Hyde, 2008; Ochong *et al*., 2008; Gesase *et al*., 2009). Nzila *et al*., (2005) proposed two possible explanations for the rarity of *dhfr* 164L in Africa: (1) the use of the standard PCR-RFLP protocol which is not sensitive enough to detect rare point mutations especially in polyclonal *P. falciparum* infections and (2) the possibility that a fitness cost of the 164L mutation makes it less sustainable in African *P. falciparum*.

As shown in chapter 2, PCR-RFLP was identified as the most common method used in genotyping studies and further work is required to assess its sensitivity for detecting rare mutations. The *dhfr* 164L mutation is highly resistant when found together with the *dhfr* 51I+59R+108N as the quadruple mutant (Foote, Galatis and Cowman, 1990; Peterson, Milhous and Wellems, 1990) and this combination compromises the therapeutic usefulness of pyrimethamine, chlorproguanil and dapsone (Lapdap) and its combination with artesunate (Ochong *et al*., 2008).

Resistant forms of *dhps* emerged circa 1993 and foresaw the arrival of SP treatment failure in Africa (Naidoo and Roper, 2010). SP resistance is magnified when *dhps* mutations occur together with *dhfr* mutations (Anderson *et al*., 2003). The *dhps* point mutations are discussed next.

Opinions are divided as to whether or not the *dhps* 436A point mutation is associated with *in vivo* resistance. Some studies did correlate the presence of *dhps* 436A with SP treatment failure in areas of low endemicity, where there was low acquired immunity (Khalil *et al*., 2002; Alifrangis *et al*., 2003). It was found to be common in areas where there was low SP use or minimal SP resistance *in vivo* (Gregson and Plowe, 2005; Plowe *et al*., 1997, Khalil *et al*., 2002), suggesting that it is not important in SP resistance and rather that it may be ancestral and an alternative of the sensitive *dhps* 436S sensitive form (Pearce *et al*., 2009). Another suggestion is that the occurrence of *dhps* 436A precludes
the progressive accumulation of other point mutations (Gregson and Plowe, 2005; Plowe et al., 1997), but in fact under increased drug pressure, it is equally probable that the 436A like the 436S is displaced by other point mutations.

Similar to \textit{dhps} 436A, the \textit{dhps} 437G point mutation was not linked substantially to resistance after initial \textit{in vitro} studies and it was the least resistant of all the mutations after testing using enzyme kinetic studies (Gregson and Plowe, 2005). The \textit{dhps} 437G mutant is thought to be among the first point mutations to be selected under sulphadoxine drug pressure and is thus a precursor for \textit{in vivo} resistance (Triglia et al., 1997; Gregson and Plowe, 2005). Other studies observed that the \textit{dhps} 437G single point mutation, combined with the \textit{dhfr} triple mutant showed some association with treatment failure (Kun et al., 1999; Dunyo et al., 2006; Ndounga et al., 2007). The 437G mutation was common among field isolates and found either on its own or in combination with the 540E (Sibley et al., 2001). The 437G single mutant confers a lesser degree of SP tolerance than that conferred by the combination of 437G and 540E when observed \textit{in vitro} (Brooks et al., 1994; Triglia et al., 1997).

Thus whilst \textit{dhps} 437G is commonly found alone in many sites in this state it is not predictive of SP \textit{in vivo} treatment failure on its own (Marks et al., 2005; Mobula et al., 2009). However in many populations it is commonly found together in combination with \textit{dhps} 540E as a double mutant (Anderson et al., 2003; Pearce et al., 2009). The double mutant allele is most common in East Africa (Naidoo and Roper, 2010) and when occurring with the \textit{dhfr} triple mutant combination as the quintuple mutant, has a strong association with clinical SP treatment failure (Omar, Adagu and Warhurst, 2001; Staedke et al., 2004; Kublin et al., 2002). In these populations the \textit{dhps} 540E is a good resistance marker, predicting the occurrence of the \textit{dhps} double mutant (437G+540E) and the quintuple mutant (\textit{dhfr} 51I+59R+108N+\textit{dhps} 437G+540E (Kublin et al., 2002; Anderson et al., 2003). Tahar and Basco (2007) suggested that monitoring of the 540E point mutation may serve as an early warning indicator for reduced SP efficacy. Indeed the WHO recently recommended that a \textit{dhps} 540E prevalence of 50% is the threshold that should be used to determine whether or not SP-IPTi is feasible for implementation in Africa (WHO, 2009a; Naidoo and Roper, 2011).

An additional point mutation, the \textit{dhps} 581G is newly emerging in East Africa and when found together with the \textit{dhps} double mutant and \textit{dhfr} triple mutant described above it confers even higher levels of SP tolerance. When it occurs in isolation the 581G has not been associated with resistance, but occurring where SP resistance is well established
and in combination with 437G or 540E it is now recognised as a significant threat to SP efficacy (Gregson and Plowe, 1995; Plowe et al., 1997). Examples of dhps 581G increasing in prevalence in East Africa include a study done in Korogwe District, Tanzania where it increased from 11.7% to 55.6% during 2003-2007 (Alifrangis et al., 2009). In Muheza, Tanga region in Tanzania, the efficacy of IPTp was compromised among pregnant women carrying the dhps 437G+540E+581G haplotype (Harrington et al., 2009). In the Hale, Tanga region of Tanzania the same haplotype was found among 55.2% (n=87) of children under five years of age in 2006 and the 581G was found on its own in 54.5% (n=11) of the symptomatic children at enrolment (Gesase et al., 2009).

The dhps 613S/T point mutations are still rare in Africa. It has been shown to confer resistance against sulphadoxine despite initial doubt about its role (Plowe et al., 1997). Plowe and others (1997) stated that the 613S/T was correlated with *in vitro* resistance but not *in vivo* resistance.

Surveillance of the *dhfr* and *dhps* point mutations has been carried out by numerous workers at many sites in sub-Saharan Africa, for the purposes of quantifying their prevalence, frequency and/or relating these measures to clinical SP resistance. Prior to the current study, no syntheses of these point mutation data were conducted. This chapter chronicles the prevalence and rate of spread of seven different SP point mutations on the African continent, including island populations, with a view to determining whether or not any trends exist at a population level. It is known that SP resistance spread quickly across the continent but a comprehensive study of the changing patterns of *dhfr* and *dhps* point mutations over any length of time has not been conducted. Differences in the distribution patterns of *dhfr* and *dhps* point mutations have been documented within African countries and some regional trends in the prevalence of *dhps* point mutations have been studied (Alker et al., 2008; Basco et al., 2006; Pearce et al., 2009; Mbacham et al., 2010). However no descriptions or comparisons between the dispersal patterns of *dhfr* and *dhps* point mutations have been made on a regional or continental level. To better understand the emergence and spread of SP mutations, data were compiled from many records of SP resistance from multiple sites and study years in an ecological study using summarised prevalence data. The scope of the data was described in chapter 2.

This data set affords the opportunity to observe trends in the occurrence of the different point mutations over time, at a continental scale. The point mutation prevalence data collected for this chapter were grouped to test whether or not there were any regional trends as per the regions described by Pearce et al., (2009). In their study Pearce et al.,
(2009) characterised lineages of \textit{dhps} allelic haplotypes among multiple sites in 20 sub-Saharan African countries and reported that parasite circulation occurred in five distinct mainland African regions. However, whilst using allelic haplotype data, these authors did not go further into analysing any other spatial or temporal trends. The \textit{dhfr} and \textit{dhps} point mutation data are a constructive starting point for trend analyses in sub-Saharan Africa. The data used in this chapter covers a 20 year period from 1988 to 2008, with a geographic coverage in excess of 170 unique sites, within 30 sub-Saharan African countries, including three island populations.

3.2 Aims and objectives

This chapter describes the prevalence of seven point mutations namely, \textit{dhfr} 164L, \textit{dhps} 436, \textit{dhps} 437, \textit{dhps} 540, \textit{dhps} 581, \textit{dhps} 613S and \textit{dhps} 613T. Given the importance of the 437G and 540E to SP treatment efficacy, these point mutations were analysed further statistically and models were developed to characterise the joint effect of time and region on their prevalence.

3.3 Methods

Data collection

Data collection methods were described in chapter 2. The Microsoft Access database was queried to extract data points for prevalence of \textit{dhfr} 164L, \textit{dhps} 436A, \textit{dhps} 437G, \textit{dhps} 540E, \textit{dhps} 581G, \textit{dhps} 613S and \textit{dhps} 613S/T point mutations. These studies included surveys of any age group as well asymptomatic or symptomatic people.

Countries with \textit{dhfr} and \textit{dhps} data were categorised into five regions as described in Pearce \textit{et al.}, (2009) (Table 3.1) wherein five distinct regional clusters were described based on shared \textit{dhps} allelic haplotypes. A sixth island cluster was added in this study to assess the heterogeneity of point mutations and allelic haplotypes among island populations. Further discussion on the coverage of \textit{dhps} and \textit{dhfr} data for each region are presented here and in chapter 4.
The DRC was allocated to both Southeast and Southwest Africa for the regional analyses, due to the significant geographic differences in SP efficacy in this large country, which spans more than two million square kilometres (Bonnet et al., 2009). The spatial mapping and statistical model development are discussed next.

### 3.3.1 Spatial mapping of point mutation prevalence

For each mutation, the prevalence was calculated using equation one:

**Equation 1:**

\[
y = \left[ \frac{\text{number of samples positive for point mutations}}{\text{total number of samples tested}} \right] \times 100
\]

Where \( y \) = prevalence of point mutations

Vector thematic maps were constructed in Mapinfo showing the prevalence of each point mutation in sub-Saharan Africa. The maps included studies with no study year.
3.3.2 Statistical model development

Graphs were constructed in Stata Intercooled Version 12.0 (StataCorp LP, College Station, USA) showing the prevalence of each point mutation with binomial 95% confidence intervals, for each study year and region. The countries were classified into the five regions as described in Table 3.1. Surveys that had a missing study year were excluded from the spatial and temporal analysis.

Multivariable regression analyses for point mutations

Logistic models were constructed to assess whether or not there was any association between the outcome variable (prevalence of point mutation) and the risk factors study year (time) and region. The construction of the model is described in equation four. The log odds of the outcome (point mutation present = 1) was used. Odds ratios, confidence intervals and cumulative p value are reported for all the model estimates. The odds ratio of each risk factor describes its relative contribution to the prevalence outcome, whilst controlling for the influence of other risk factors (Bagley, White and Golomb, 2001). Dummy variables were created for each region category and coded as listed in Table 3.2.

The Southeast region was used as the reference region because it had good coverage of data points over the 20 year period (Table 3.2). Island populations were included in these analyses. After running the logistic regression, the likelihood ratio test was performed to compare the models with and without point mutation prevalence changing over time and allowed an assessment of the strength of the evidence for any observed temporal change.

Further multivariable regression analyses for dhps 437G and 540E

The aim was to assess whether or not there was any association between the outcome variable (prevalence of point mutation) and the risk factors study year (time) and region. In this study logistic regression analysis with robust standard errors allowing for clustering by country, was performed for each point mutation as described in equation four. This was done to model the effect of time on prevalence, using odds ratios. Thereafter the goal was to assess if any observed trend differed between regions. This was demonstrated by fitting a logistic model with an interaction term between the study year and region, as shown in equation five. Island populations were excluded in these analyses because it had the least
amount of data points and these populations differ widely from each other. The Southeast region was used as the reference region throughout.

Equation 4: \( \log\left(\frac{p}{1-p}\right) = \beta_0 + \beta_1 x_1 + \beta_2 x_2 \)

where

\( p = \) probability of point mutation being present

\( 1-p = \) probability of point mutation being absent

\( x_1 = \) study year

\( x_2 = \) region

\( x_3 = \) region x study year

\( \beta_0 = \) constant

Equation 5: \( \log\left(\frac{p}{1-p}\right) = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \beta_3 x_3 \)

Table 3.2 Coding of geographic regions to create dummy variables

<table>
<thead>
<tr>
<th>Region</th>
<th>Code</th>
</tr>
</thead>
<tbody>
<tr>
<td>Central</td>
<td>1</td>
</tr>
<tr>
<td>Northeast</td>
<td>2</td>
</tr>
<tr>
<td>Southeast</td>
<td>3</td>
</tr>
<tr>
<td>Southwest</td>
<td>4</td>
</tr>
<tr>
<td>West</td>
<td>5</td>
</tr>
<tr>
<td>Island</td>
<td>6</td>
</tr>
</tbody>
</table>

Post estimation

After running the logistic regression, the Wald test was performed to assess the strength of evidence for differences in prevalence of the \textit{dhps} 437G and 540E point mutations among regions. This test tests the null hypothesis that the prevalence is the same in all regions,
apart from sampling variation. Akaike information criterion (AIC) was generated to estimate
the goodness of fit of each model and to compare the models. Lower AIC values indicate
the better fitted model.

Linear combinations

Estimating the linear combinations of the odd ratios sequentially for each region as in
equation six below, provided the difference in the effect of time on the likelihood of the
point mutation occurring, between each region and the reference Southeast region.

Equation 6: \[ \log\left(\frac{p}{1-p}\right) \text{ or } x\beta = \beta_1x_1 + \beta_2nx_2 + \beta_3x_3 \]

Where \( n = \) each region

In Stata, graphs were constructed using the model estimates of the predicted probabilities
of positivity (\( Y=1 \)) in each region.

Assumptions and missing data

Each \( dhps \) point mutation was independently sampled within each country. Between-
country variation in responses was modelled using Stata’s option for robust estimates.
Surveys with missing data were excluded, for example where the study year was missing.

3.4 Results

3.4.1 Spatial mapping of point mutation prevalence

This section contains a series of seven maps illustrating the country-level prevalence of
each point mutation in sub-Saharan Africa.

Table 3.3 outlines the data represented in the seven maps. Countries without point
mutation data are the gaps reported previously in chapter 2. Some sites on the maps
indicate both presence (red dots) and absence (blue squares) of point mutations. This
occurs where different studies were conducted at the same site but at different times. The
percentage in brackets in Table 3.3 indicates the proportion of samples that were positive
for point mutations from the samples tested. In the following sections ‘n’ represents the
number of samples tested in a particular survey. In the map legends, the numbers in
brackets refer to the number of study sites. The references used for each mutation are provided in appendices two through eight.

3.4.2 Change in point mutation prevalence over time

The coverage of surveys available for statistical analyses is presented in Tables 3.4a and 3.4b. The prevalence for each point mutation is presented in a series of graphs with confidence intervals based on the proportion positive for each survey. These data are aggregated for each of the six regions. The results of the logistic regression analyses are presented thereafter. Generally the models showed evidence of a temporal change among all point mutations over time. The prevalence of dhfr 164, dhps 437G, dhps 540E and dhps 581G increased over time whilst that of dhps 436A and dhps 613S/T decreased over time.
Table 3.3 Summary of surveys available with prevalence data for point mutations

<table>
<thead>
<tr>
<th>Point mutation</th>
<th>Surveys</th>
<th>Samples tested</th>
<th>Samples mutant (%)*</th>
<th>Samples Sensitive Sites #</th>
<th>Unique Countries with point mutation</th>
<th>Countries without point mutation</th>
<th>Study year range</th>
<th>Number of surveys with no study years</th>
</tr>
</thead>
<tbody>
<tr>
<td>dhfr 164</td>
<td>184</td>
<td>19923</td>
<td>130 (0.7)</td>
<td>18958</td>
<td>116</td>
<td>7</td>
<td>21</td>
<td>1988-2008</td>
</tr>
<tr>
<td>dhps 436</td>
<td>153</td>
<td>19033</td>
<td>2929 (15.4)</td>
<td>15988</td>
<td>107</td>
<td>33</td>
<td>2</td>
<td>1993-2008</td>
</tr>
<tr>
<td>dhps 437</td>
<td>232</td>
<td>28851</td>
<td>14705 (51)</td>
<td>13222</td>
<td>152</td>
<td>38</td>
<td>0</td>
<td>1988-2008</td>
</tr>
<tr>
<td>dhps 540</td>
<td>261</td>
<td>30291</td>
<td>10346 (67.2)</td>
<td>18811</td>
<td>172</td>
<td>28</td>
<td>10</td>
<td>1988-2008</td>
</tr>
<tr>
<td>dhps 581</td>
<td>121</td>
<td>15398</td>
<td>782 (5.1)</td>
<td>14526</td>
<td>80</td>
<td>13</td>
<td>15</td>
<td>1993-2008</td>
</tr>
<tr>
<td>dhps 613S</td>
<td>91</td>
<td>10312</td>
<td>134 (1.3)</td>
<td>10180</td>
<td>75</td>
<td>10</td>
<td>17</td>
<td>1993-2007</td>
</tr>
<tr>
<td>dhps 613T</td>
<td>88</td>
<td>10152</td>
<td>11 (0.1)</td>
<td>10035</td>
<td>64</td>
<td>3</td>
<td>22</td>
<td>1988-2008</td>
</tr>
</tbody>
</table>

* % mutant = [number of samples positive for point mutation /number of samples tested]*100

# incorporates studies where sensitive alleles were reported or could be deduced from data provided
Table 3.4a Summary of surveys available with prevalence data for point mutations by region

<table>
<thead>
<tr>
<th>Regions</th>
<th>Southeast</th>
<th>Northeast</th>
<th>West</th>
<th>Southwest</th>
<th>Central</th>
<th>Island</th>
<th>All Regions</th>
</tr>
</thead>
<tbody>
<tr>
<td>Surveys of I164L</td>
<td>100</td>
<td>13</td>
<td>33</td>
<td>5</td>
<td>21</td>
<td>6</td>
<td>178</td>
</tr>
<tr>
<td>Samples tested for I164L</td>
<td>13232</td>
<td>1099</td>
<td>2125</td>
<td>395</td>
<td>1513</td>
<td>1368</td>
<td>19732</td>
</tr>
<tr>
<td>Samples positive for 164L</td>
<td>78</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>51</td>
<td>130</td>
</tr>
<tr>
<td>Samples positive for I164</td>
<td>12468</td>
<td>1099</td>
<td>2125</td>
<td>367</td>
<td>1390</td>
<td>1318</td>
<td>18767</td>
</tr>
<tr>
<td>Surveys unresolved I164</td>
<td>3</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>2</td>
<td>0</td>
<td>5</td>
</tr>
<tr>
<td>Unique sites with I164L</td>
<td>58</td>
<td>9</td>
<td>18</td>
<td>5</td>
<td>15</td>
<td>4</td>
<td>109</td>
</tr>
<tr>
<td>Countries with I164L</td>
<td>10</td>
<td>4</td>
<td>7</td>
<td>3</td>
<td>2</td>
<td>3</td>
<td>29</td>
</tr>
<tr>
<td>Surveys of S436A</td>
<td>72</td>
<td>13</td>
<td>25</td>
<td>11</td>
<td>11</td>
<td>5</td>
<td>137</td>
</tr>
<tr>
<td>Samples tested for S436A</td>
<td>10540</td>
<td>904</td>
<td>2223</td>
<td>1020</td>
<td>1487</td>
<td>1071</td>
<td>17245</td>
</tr>
<tr>
<td>Samples positive for 436A</td>
<td>826</td>
<td>59</td>
<td>969</td>
<td>254</td>
<td>741</td>
<td>52</td>
<td>2901</td>
</tr>
<tr>
<td>Samples positive for S436</td>
<td>9659</td>
<td>846</td>
<td>1207</td>
<td>823</td>
<td>831</td>
<td>993</td>
<td>14359</td>
</tr>
<tr>
<td>Surveys unresolved S436</td>
<td>2</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>2</td>
</tr>
<tr>
<td>Unique sites with S436A</td>
<td>46</td>
<td>10</td>
<td>22</td>
<td>8</td>
<td>4</td>
<td>4</td>
<td>94</td>
</tr>
<tr>
<td>Countries with 436A</td>
<td>11</td>
<td>4</td>
<td>10</td>
<td>4</td>
<td>2</td>
<td>3</td>
<td>34</td>
</tr>
<tr>
<td>Surveys of A437G</td>
<td>121</td>
<td>27</td>
<td>39</td>
<td>13</td>
<td>17</td>
<td>9</td>
<td>226</td>
</tr>
<tr>
<td>Samples tested for A437G</td>
<td>17860</td>
<td>1531</td>
<td>4450</td>
<td>1201</td>
<td>2269</td>
<td>1294</td>
<td>28605</td>
</tr>
<tr>
<td>Samples positive for A437G</td>
<td>9109</td>
<td>644</td>
<td>1990</td>
<td>881</td>
<td>1276</td>
<td>599</td>
<td>14499</td>
</tr>
<tr>
<td>Samples positive for A437</td>
<td>8931</td>
<td>886</td>
<td>1074</td>
<td>410</td>
<td>1104</td>
<td>729</td>
<td>13134</td>
</tr>
<tr>
<td>Surveys unresolved A437</td>
<td>12</td>
<td>0</td>
<td>9</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>21</td>
</tr>
<tr>
<td>Unique sites with A437G</td>
<td>65</td>
<td>13</td>
<td>32</td>
<td>9</td>
<td>9</td>
<td>8</td>
<td>136</td>
</tr>
<tr>
<td>Countries with A437G</td>
<td>11</td>
<td>4</td>
<td>13</td>
<td>5</td>
<td>2</td>
<td>3</td>
<td>38</td>
</tr>
<tr>
<td>Surveys of K540E</td>
<td>127</td>
<td>27</td>
<td>42</td>
<td>20</td>
<td>21</td>
<td>17</td>
<td>254</td>
</tr>
<tr>
<td>Samples tested for K540E</td>
<td>18602</td>
<td>1537</td>
<td>4543</td>
<td>1755</td>
<td>2109</td>
<td>1278</td>
<td>29824</td>
</tr>
<tr>
<td>Samples positive for 540E</td>
<td>9408</td>
<td>609</td>
<td>65</td>
<td>69</td>
<td>9</td>
<td>30</td>
<td>10190</td>
</tr>
<tr>
<td>Samples positive for K540</td>
<td>9604</td>
<td>928</td>
<td>2820</td>
<td>1719</td>
<td>2092</td>
<td>1254</td>
<td>18417</td>
</tr>
<tr>
<td>Samples unresolved K540</td>
<td>10</td>
<td>0</td>
<td>4</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>14</td>
</tr>
<tr>
<td>Unique sites with K540E</td>
<td>82</td>
<td>13</td>
<td>32</td>
<td>15</td>
<td>11</td>
<td>12</td>
<td>165</td>
</tr>
<tr>
<td>Countries with K540E</td>
<td>11</td>
<td>4</td>
<td>13</td>
<td>5</td>
<td>2</td>
<td>3</td>
<td>38</td>
</tr>
</tbody>
</table>
Table 3.4b Summary of surveys available with prevalence data for point mutations by region

<table>
<thead>
<tr>
<th>Regions</th>
<th>Southeast</th>
<th>Northeast</th>
<th>West</th>
<th>Southwest</th>
<th>Central</th>
<th>Island</th>
<th>All Regions</th>
</tr>
</thead>
<tbody>
<tr>
<td>Surveys of 581G</td>
<td>69</td>
<td>12</td>
<td>15</td>
<td>5</td>
<td>9</td>
<td>5</td>
<td>115</td>
</tr>
<tr>
<td>Samples tested for A581G</td>
<td>10846</td>
<td>835</td>
<td>859</td>
<td>275</td>
<td>1220</td>
<td>1071</td>
<td>15106</td>
</tr>
<tr>
<td>Samples positive for 581G</td>
<td>657</td>
<td>34</td>
<td>26</td>
<td>0</td>
<td>14</td>
<td>1</td>
<td>732</td>
</tr>
<tr>
<td>Samples positive for A581</td>
<td>10185</td>
<td>801</td>
<td>726</td>
<td>275</td>
<td>1208</td>
<td>1070</td>
<td>14265</td>
</tr>
<tr>
<td>Surveys unresolved 581G</td>
<td>1</td>
<td>0</td>
<td>2</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>3</td>
</tr>
<tr>
<td>Unique sites with 581G</td>
<td>43</td>
<td>9</td>
<td>11</td>
<td>4</td>
<td>4</td>
<td>3</td>
<td>78</td>
</tr>
<tr>
<td>Countries with 581G</td>
<td>9</td>
<td>4</td>
<td>6</td>
<td>3</td>
<td>2</td>
<td>3</td>
<td>27</td>
</tr>
<tr>
<td>Surveys of S613A</td>
<td>49</td>
<td>11</td>
<td>13</td>
<td>3</td>
<td>6</td>
<td>4</td>
<td>86</td>
</tr>
<tr>
<td>Samples tested for S613A</td>
<td>6893</td>
<td>754</td>
<td>735</td>
<td>184</td>
<td>798</td>
<td>778</td>
<td>10142</td>
</tr>
<tr>
<td>Samples positive for S613A</td>
<td>23</td>
<td>1</td>
<td>88</td>
<td>0</td>
<td>20</td>
<td>1</td>
<td>133</td>
</tr>
<tr>
<td>Samples positive for S613</td>
<td>6871</td>
<td>753</td>
<td>649</td>
<td>184</td>
<td>779</td>
<td>775</td>
<td>10011</td>
</tr>
<tr>
<td>Surveys unresolved S613A</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>Unique sites with S613A</td>
<td>33</td>
<td>8</td>
<td>10</td>
<td>3</td>
<td>2</td>
<td>4</td>
<td>60</td>
</tr>
<tr>
<td>Countries with S613A</td>
<td>9</td>
<td>4</td>
<td>5</td>
<td>3</td>
<td>2</td>
<td>3</td>
<td>26</td>
</tr>
<tr>
<td>Surveys of S613T</td>
<td>49</td>
<td>10</td>
<td>12</td>
<td>2</td>
<td>5</td>
<td>5</td>
<td>83</td>
</tr>
<tr>
<td>Samples tested for S613T</td>
<td>6893</td>
<td>717</td>
<td>609</td>
<td>146</td>
<td>546</td>
<td>1071</td>
<td>9982</td>
</tr>
<tr>
<td>Samples positive for S613T</td>
<td>1</td>
<td>0</td>
<td>7</td>
<td>0</td>
<td>0</td>
<td>3</td>
<td>11</td>
</tr>
<tr>
<td>Samples positive for S613</td>
<td>6871</td>
<td>716</td>
<td>531</td>
<td>146</td>
<td>535</td>
<td>1067</td>
<td>9866</td>
</tr>
<tr>
<td>Samples unresolved S613T</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>Unique sites with S613T</td>
<td>33</td>
<td>10</td>
<td>8</td>
<td>2</td>
<td>2</td>
<td>3</td>
<td>58</td>
</tr>
<tr>
<td>Countries with S613T</td>
<td>9</td>
<td>4</td>
<td>4</td>
<td>2</td>
<td>2</td>
<td>3</td>
<td>24</td>
</tr>
</tbody>
</table>
Prevalence of *dhfr* 164L

The *dhfr* 164L point mutation was found to be absent in 21 countries (Naidoo and Roper, 2011). It occurred in the following seven countries: Central African Republic (0.6%, n=167) (Menard *et al*., 2006), Comoros (1.1%, n=269) (Andriantsoanirina *et al*., 2009); Kenya (3.4%, n=58) (Oesterholt *et al*., 2009), Madagascar (8.8%, n=273) (Andriantsoanirina *et al*., 2009), Malawi (4.7%, n=85) (Alker *et al*., 2005), Rwanda (11.8%, n=391) (Karema *et al*., 2010) and Uganda (13.7%, n=51) (Lynch *et al*., 2008), where it occurred in four different sites, surveyed in 1999, 2004 and 2005 (Figure 3.1). In Madagascar four surveys were done at various sites across the island between 2006 and 2008, where 164L occurred at a prevalence ranging between 0.9% to 8.8% (Menard *et al*., 2008; Andriantsoanirina *et al*., 2009). The cumulative continental prevalence of mutant *dhfr* 164L was 0.7% (130/19923) between 1998 and 2008 (Table 3.3).

![Figure 3.1 Prevalence of dhfr 164L](image)
Regional prevalence of *dhfr* 164L point mutation over time

The *dhfr* 164L point mutation occurred only in the Southeast, Central and Island regions and was not found in the Northeast, West and Southwest regions (Table 3.4a, Figure 3.2). Out of 100 surveys in the Southeast region, the overall prevalence was 0.59% (n=13232) from eight surveys where *dhfr* 164L was reported. The earliest report of 1.2% (n=82) prevalence occurred in 1999 in Kampala, Uganda (Staedke *et al*., 2004). In the Islands, the most recent prevalence was reported in Madagascar 2008 (Andriantsoanirina *et al*., 2009).

The logistic model predicted an overall higher likelihood in *dhfr* 164L prevalence over time. For every year, the likelihood of a person having a *dhfr* 164L point mutation was 1.4 times higher (CI 1.3-1.6)(Table 3.5). The likelihood of *dhfr* 164L occurring among the Island populations was greater than it was for the Southeast region (OR 2.1, CI 1.3-3.2) and comparatively lower in the Central region (OR 0.2, CI 0.0-1.6) (Figure 3.3). The low p value (p =0.001) for the Island region suggested there was strong evidence for this effect. The likelihood ratio test comparing models with and without *dhfr* 164L prevalence changes over time provided strong evidence to suggest that the regional prevalence of *dhfr* 164L was affected by time (p< 0.001) (Table 3.5). Coverage for *dhfr* 164L was better in the Southeast and Central regions compared with the Islands. Notably there were only six surveys available within the island populations, with a single survey done in 2002 and five surveys between 2006-2008.
Figure 3.2 Surveys of dhfr 164L prevalence (with 95% confidence intervals) conducted during 1988-2008 and displayed according to their geographic region.
<table>
<thead>
<tr>
<th>Covariate</th>
<th>Proportion (%) of samples positive [95% confidence interval]</th>
<th>Odds ratio [95% confidence interval]</th>
<th>P value</th>
<th>Odds ratio [95% confidence interval]</th>
<th>P value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Time (per study year)</td>
<td>0.9 [0.6-0.8]</td>
<td>1.4 [1.3-1.6]</td>
<td>&lt;0.001</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Region</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Southeast</td>
<td>0.6 [0.5-0.7]</td>
<td>1</td>
<td>1</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Central</td>
<td>0.1 [0.0-0.4]</td>
<td>0.2 [0.0-1.6]</td>
<td>0.128</td>
<td>0.1 [0.0-0.8]</td>
<td>0.029</td>
</tr>
<tr>
<td>Island</td>
<td>3.7 [2.8-4.9]</td>
<td>2.1 [1.3-3.2]</td>
<td>0.001</td>
<td>6.5 [4.6-9.3]</td>
<td>&lt;0.001</td>
</tr>
</tbody>
</table>
The *dhps* 436A mutation occurred in 32 countries after sampling from 107 unique sites (Figure 3.4, Table 3.3). Only Ethiopia and Swaziland did not report this mutation (Gebru-Woldearegai *et al.*, 2005; Schunk *et al.*, 2006). The highest prevalence of 436A was 80% and this occurred in Central African Republic and Ghana (Menard *et al.*, 2006; Pearce *et al.*, 2009). Angola, Burkina Faso, Cameroon, Gabon, Guinea and Mali recorded ≥ 50% prevalence (Plowe *et al.*, 1997; Wang *et al.*, 1997; Basco and Ringwald, 1998; Basco, Tahar and Ringwald, 1998; Diourte *et al.*, 1999; Basco *et al.*, 2000; Mawili-Mboumba *et al.*, 2001; Bonnet *et al.*, 2007; Tahar and Basco, 2007; Democratic Republic of Congo (east), Kenya, Malawi, Mozambique, Rwanda, South Africa, Swaziland, Tanzania, Uganda, Zambia. Camero(236,794),(970,828)
Tinto et al., 2007; Menegon et al., 2009; Pearce et al., 2009; Mbacham et al., 2010). Lower prevalence than this occurred in Rwanda (0.5%-5.5%) (Karema et al., 2010).

On the Comoros island the prevalence was 1.1% (n= 269) (Andriantsoanirina et al., 2009). In Djibouti it was 5.3-6.5% (Rogier et al., 2005), in the DRC it ranged between 3.7%-24.3% from two studies done in 2002-2004 (Swarthout et al., 2006; Alker et al., 2008), in Equatorial Guinea it was 16.7% (n=12) (Berzosa, Puente and Benito, 2005) and in Kenya the prevalence ranged between 3%-23.8% from seven studies conducted between 1993-1999 (Wang et al., 1997; Plowe et al., 1997; Nzila et al., 2000a; Nzila et al., 2000b; Omar et al., 2001; Mberu et al., 2002; Zhong et al., 2008; Pearce et al., 2009; Oesterholt et al., 2009).

By 2006 in Madagascar 436A was 0.6% (Andriantsoanirina et al., 2009), in Malawi it was 7% (Plowe et al., 1997) and in Namibia it was 12% (Pearce et al., 2009). In Mozambique it was <8% after 12 surveys were conducted between 2000 and 2004 (Alifrangis et al., 2003; Enosse et al., 2008; Pearce et al., 2009; Raman et al., 2008). In the Republic of Congo it was <18% between 1999-2004 (Nsima et al., 2005; Ndounga et al., 2007; Pearce et al., 2009).

Two studies done in South Africa reported 2.4% and 5.8% prevalence between 1999 and 2000 (Roper et al., 2003; Pearce et al., 2009). In Uganda it was 16% (Jelinek et al., 1999). In Zambia two surveys done separately in 2004 reported <2% prevalence (Pearce et al., 2009). Finally, the dhps 436A prevalence was <13% (n=5) in Zimbabwe (Mlambo et al., 2007). Overall, 15.4% of the samples contained dhps 436A (n=19033) between 1993 and 2008 (Table 3.3).
Regional prevalence of \textit{dhps} 436A point mutation over time

The \textit{dhps} 436A point mutation was found in all six regions (Table 3.3, Figure 3.5). The prevalence was \(\geq 50\%\) in the West, Southwest and Central regions during 1994-2005. In the Southeast region the highest prevalence of 38.2\% \((n=68)\) was recorded in Kibaha Hospital, 30 km from Dar es Salaam in 1988 (Alifrangis \textit{et al.}, 2003). In the Northeast region the highest prevalence was 44\% \((n=75)\) in Akuem, Yargot Payam district, Bahr El Gazal, South Sudan in 2002 (Anderson \textit{et al.}, 2003). Among the Islands the prevalence ranged between 0–42.3\% from five surveys (Andriantsoanirina \textit{et al.}, 2009; Tahar \textit{et al.}, 2007; Cravo, 2004). Overall, the Central region had the highest regional prevalence of 49.8\% \((n=1487)\) (Table 3.4a).
Figure 3.5 Surveys of dhps 436A prevalence (with 95% confidence intervals) conducted during 1988-2008 and displayed according to their geographic region.
The logistic regression model predicted an overall lower likelihood in *dhps* 436A over time. For every year, the likelihood of a person having a *dhps* 436A point mutation was 0.9 times lower (CI 0.9-0.9). Compared to the Southeast region, the likelihood of *dhps* 436A occurring in the West (OR 9.0, CI 8.0-1.0), Southwest (OR 3.8, CI 3.3-4.5) and Central (OR 11.9, CI 10.5-13.5) regions was greater and the evidence for this was strong (p<0.001) (Table 3.6). Figure 3.6 shows the declining rates of *dhps* 436A prevalence over time in all regions. The likelihood of *dhps* 436A occurring in the Northeast (OR 0.8, CI 0.6-1.0) and Island (OR 0.8, CI 0.6-1.1) regions was lower than it was for the Southeast region, although the high p values suggested there was weak evidence for this (Table 3.6). The likelihood ratio test comparing models with and without the effect of time provided strong evidence to suggest that the regional prevalence of *dhps* 436A was affected by time (p< 0.001).

Table 3.6 Model for *dhps* 436A

<table>
<thead>
<tr>
<th>Covariate</th>
<th>Proportion (%) of samples positive [95% confidence interval]</th>
<th>Odds ratio [95% confidence interval]</th>
<th>P value</th>
<th>Odds ratio [95% confidence interval]</th>
<th>P value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Time (per study year)</td>
<td>16.8[16.3-17.4]</td>
<td>0.9[0.9-0.9]</td>
<td>&lt;0.001</td>
<td>1</td>
<td></td>
</tr>
<tr>
<td>Region</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Southeast</td>
<td>7.8 [7.3-8.4]</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td></td>
</tr>
<tr>
<td>Northeast</td>
<td>6.5[5.0-8.3]</td>
<td>0.8[0.6-1.0]</td>
<td>0.064</td>
<td>0.8[0.6-1.1]</td>
<td>0.158</td>
</tr>
<tr>
<td>West</td>
<td>43.6[41.5-45.7]</td>
<td>9.0[8.0-1.0]</td>
<td>&lt;0.001</td>
<td>9.1[8.1-10.1]</td>
<td>&lt;0.001</td>
</tr>
<tr>
<td>Southwest</td>
<td>24.9[22.3-27.7]</td>
<td>3.8[3.3-4.5]</td>
<td>&lt;0.001</td>
<td>3.9[3.3-4.6]</td>
<td>&lt;0.001</td>
</tr>
<tr>
<td>Central</td>
<td>49.8[47.3-52.4]</td>
<td>11.9[10.5-13.5]</td>
<td>&lt;0.001</td>
<td>11.7[10.3-13.2]</td>
<td>&lt;0.001</td>
</tr>
<tr>
<td>Island</td>
<td>4.9[3.6-6.3]</td>
<td>0.8[0.6-1.1]</td>
<td>0.163</td>
<td>0.6[0.5-0.8]</td>
<td>0.001</td>
</tr>
</tbody>
</table>
Figure 3.6 Predicted probabilities of dhps 436A over time

Democratic Republic of Congo (east),
Kenya, Malawi,
Mozambique,
Rwanda, South Africa,
Swaziland, Tanzania,
Uganda, Zambia,
Zimbabwe

Djibouti, Ethiopia,
North Sudan, South Sudan

Burkina Faso, Ivory Coast, Gambia, Ghana,
Guinea, Guinea-Bissau,
Mali, Mauritania,
Nigeria, Senegal

Angola, Gabon,
Namibia, Republic of Congo

Cameroon, Central African Republic

Comoros,
Madagascar,
Sao Tome/Principe
High prevalence of *dhps* 437G occurred heterogeneously across Africa and was found in all 37 countries surveyed among 152 unique sites (Figure. 3.7, Table 3.3). Studies done in Ethiopia (2004) (Gebru-Woldearegai *et al*., 2005), Kenya (2006) (Bonizzoni *et al*., 2009), Sao Tome/Principe (2004) (Salgueiro *et al*., 2010), South Sudan (2003) (Pearce *et al*., 2009) and Uganda (2003-2006) (Malamba *et al*., 2010, Lynch *et al*., 2008) all reported 100% *dhps* 437G.

In Kilifi, Kenya 76.3% (n=76) *dhps* 437G prevalence was reported by Ogutu *et al*. (2005) whilst Nzila *et al*. (2000b) reported 48% (n=52) prevalence during 1993-1995. In Blantyre, Malawi the prevalence was consistently high (≥ 90) between 2001 and 2005 (Alker *et al*., 2005; Nkhoma, Molyneux and Ward, 2007). In Maputo province, Mozambique the prevalence was found to increase over time. During 2004-2006, the prevalence was <50% and increased to 64.5%(n=141) in 2007 (Raman *et al*., 2010) whilst in 2008 it was 75% (n=76) (Raman *et al*., 2010).

In Tanzania 34 surveys were obtained and 437G prevalence was heterogeneous. For instance, in Kilombero and Ulanga it was measured at 13% (n=365) in 2000, 10% (n=519 and 214) in 2001 and 27% (n=603) in 2002 (Malisa *et al*., 2010). Later in 2004 it was reported at 57.3% (n=561) (Pearce *et al*., 2009). In Korogwe it was consistently high during 2004-2007 (85%-97%) (Alifrangis *et al*., 2009). Similarly in Magoda the prevalence was 78.5%-91% during 1995-1999 (Jelinek *et al*., 1998; Alifrangis *et al*., 2009). In Muheza, the prevalence of 437G was 86% (n=50) in 1995 (Curtis, Duraisingh and Warhurst, 1998) and 96% (n=28) during 1998-1999 (Mutabingwa *et al*., 2001). However in Rufigi, it was just 8.6% (n=417) in 2000 but it was 25% (n=596) when assessed in 2002 (Malisa *et al*., 2010).

In Kampala, Uganda the prevalence was 58% (n=81) in 1999 (Staedke *et al*., 2004) and 89% (n=118) in 2000 (Sendagire *et al*., 2005). In Kasangati,Uganda the prevalence rose from 58.8% (n=97) in 2002 to 86.3% (n=183) in 2003 (Sendagire *et al*., 2005; Maiga *et al*., 2007). In Tororo, Uganda the prevalence was 62.5% (n=251) in 2001-2002 (Malamba *et al*., 2006), 98.5% (n=333) in 2002-2004 (Francis *et al*., 2006), reaching 100% in another study done during 2003-2006 (Malamba *et al*., 2010). All 17 sites from Uganda had prevalence in excess of 57%.
In Macha, Zambia the prevalence was 4.5% (n=110) in 1998 and rose to 40.8% (n=49) in 2000 but declined to 32% in 2003 (Mkulama et al., 2008). In Chibombo, Chipata and Isoka the prevalence was 66.6% (n=15), 70.3% (n=54) and 75% (n=12) respectively (Pearce et al., 2009). Lower prevalence than those described above were found in Djibouti (< 38.7%) since 1998 (Rogier et al., 2005) and in Nigeria where it was < 47% (Pearce et al., 2009; Happi et al., 2005). Similarly in South Africa, it was found to reach 48% (n=73) between 1996-2002 (Roper et al., 2003; Barnes et al., 2008; Pearce et al., 2009). In 2001 in Guinea-Bissau it was 28.8% (n=97) (Kofoed et al., 2004). In Central African Republic in 2004 it was reported to be 18.6% (n=183) (Menard et al., 2006). In Mauritania in 1998 it was <22% (Eberl et al., 2001). In Zimbabwe, 437G was assessed in 2003 in three different sites and all reported <17% prevalence (Mlambo et al., 2007).

In Namibia the prevalence was relatively high at 86.8% (n=76) as reported by Pearce et al. (2009). In the Republic of Congo the prevalence ranged from 68.3%-85% between 1999-2004 (Nsimba et al., 2005; Ndounga et al., 2007; Pearce et al., 2009). Karembe et al. (2010) reported high 437G prevalence of 80%-96% in Rwanda during 2005-2006. In Sao Tome/Principe consistently high prevalence of >75% was reported since 2000 by Cravo et al. (2004) and others (Tahar et al., 2007; Salguiero et al., 2010).

In Sotuba and Bancoumana, Mali, a study with small sample sizes (n=10; n=8) also reported 100% 437G (Wang et al., 1997). Among the eight other studies done in Mali, the highest prevalence was found in Bandiagara in 2000 where 59% (n=66) was recorded by Thera and others (2005). In Pikine, Senegal 437 prevalence ranged from 17% (n=23) in 2000 to 40% (n=15) in 2003 (Ndiaye et al., 2005). In Niakhar, Senegal the prevalence was 91.1% (n=45) in 2004 (Sokhna et al., 2008). In Sudan, there were 17 surveys done during 1996-2003 with a wide prevalence range of 4.4%-100%. In Asar, Daraweesh, Kajara and Gedaref the reported prevalence was >70% in 2003 (A-Elbasit et al., 2007; Al-Saai et al., 2009; Pearce et al., 2009).

In Lambarene, Gabon the prevalence was 27.7% in 1995-1996 (n=36) (Kun et al., 1999) and 75% by 2007 (n=64) (Pearce et al., 2009). Similarly in Haut Ogooue, Gabon the prevalence increased from 36.8% in 1998 (n=36) (Mawili-Mboumba et al., 2001) to 63.63% (n=110) in 2000 (Aubouy et al., 2003). In Farafenni, Gambia the prevalence was 48.8% (n=86) during 2001-2004 (Dunyo et al., 2006) whilst Pearce et al., (2009) reported it as 46.45% in 2004 (n=127). In Ghana six surveys
reported prevalence ≥ 74% since 2001 (Marks et al., 2005; Mockenhaupt et al., 2005; Owusu-Agyei et al., 2009; Pearce et al., 2009).

In Angola two studies done in 2004 by Pearce et al., (2009) and Menegon et al., (2009) reported 92.5% (n=40) and 92% (n=66) 437G respectively. In a study done in Benin during 2003-2005, Nahum et al. (2009) reported 88% (n=65) prevalence. In Burkina Faso three surveys reported 58%-78% prevalence (Dokomajilar et al., 2006; Tinto et al., 2007; Pearce et al., 2009).

Two studies done in Mutengene, Cameroon reported 84.1% (n=202) prevalence in 2004 (Pearce et al., 2009) and 85.5% (n=200) between 2004-2006 (Mbacham et al., 2010). In Yaounde, Cameroon seven surveys found 437G prevalence from 16.6% in 1995 (n=6) (Basco et al., 1998) to 65% during 2001-2005 (n=252) (McCollum et al., 2008) and 62% in both 2004 (n=143) (Pearce et al., 2009) and during 2004-2006 (n=145) (Mbacham et al., 2010) (Basco and Ringwald, 1998; Basco et al., 2000, Tahar and Basco, 2007).

In the DRC, the prevalence ranged from 24.4% (n=135) in Kilwa (2003-2004) (Cohuet et al., 2006), to 93.1% (n=102) in Kinshasa (2008) (Mobula et al., 2009). In the Ivory Coast two surveys in Abidjan (2006) and Yopougan (2001) found 437G at 56.6% (n=106) (Djaman et al., 2010) and 51.7% (n=118) respectively (Djaman, Mazabraud and Basco, 2007).

In Swaziland it was 7.4% (n=27) in 1999 (Dlamini, Beshir and Sutherland, 2009). Even lower prevalence than this were reported in Comoros in 2006 (6.3%, n=269) (Andriantsoanirina et al., 2009). Overall, of the 28851 samples tested for dhps 437G between 1998 and 2008, 51% were positive (Table 3.3).
Figure 3.7 Prevalence of *dhps* 437G
Regional prevalence of *dhps* 437G point mutation over time

The *dhps* 437G point mutation prevalence tended to increase over time, as ≥50% prevalence was reported at sites within each region. The 437G prevalence was 100% in the Southeast and Northeast regions in 2003 and on the Sao Tome/Principe Islands in 2004 (Figure 3.8) (Pearce *et al.*, 2003; Malamba *et al.*, 2010; Salgueiro *et al.*, 2010). Overall, the Southwest region had the highest prevalence (73.3%; n=1201) among the regions (Table 3.4a).

This observation was confirmed by the logistic regression model which predicted an overall higher likelihood of *dhps* 437G over time. For every year, the likelihood of a person having a *dhps* 437G point mutation was 1.1 times higher (CI 1.1-1.1) and there was strong evidence for this (p<0.001) (Table 3.7). Figure 3.9 shows the increasing rates of *dhps* 437G prevalence over time in all regions.

The likelihood of *dhps* 437G occurring in the Southwest (OR 2.5, CI 2.2-2.9) and Central (OR 1.3, CI 1.2-1.4) regions was greater than the likelihood of it occurring in the Southeast region and the evidence for this was strong (p<0.001) (Table 3.7). The likelihood of *dhps* 437G occurring in the Northeast (OR 0.9, CI 0.8-1.0), West (OR 0.6, CI 0.6-0.7) and Island (OR 0.5, CI 0.5-0.6) regions was lower than it was for the Southeast region and the low p values (< 0.001) suggested there was strong evidence for this only in the Western and Island regions (Table 3.7). The likelihood ratio test comparing models with and without the effect of time provided strong evidence to suggest that the regional prevalence of *dhps* 437G was affected by time (p< 0.001).
Figure 3.8 Surveys of *dhps* 437G prevalence (with 95% confidence intervals) conducted during 1988-2008 and displayed according to their geographic region.
The \textit{dhps} 437G prevalence data were analysed further, without the Island populations in section 3.4.3. The Island populations were excluded in subsequent analysis of \textit{dhps} 437G because this region had the least amount of surveys over time ie nine surveys since 2000.
Figure 3.9 Predicted probabilities of *dhps* 437G over time
Prevalence of *dhps* 540E

As indicated in chapter 2 and Table 3.4 the best data coverage of all point mutations was for *dhps* K540E. The *dhps* 540E dataset was updated in 2011 with one additional study and 67.2% of the samples tested positive for 540E from 30291 samples tested since 1998 to 2008 (Table 3.3). The *dhps* 540E mutation was found in 27 countries and was absent in 10 countries (Figure 3.10). Generally high prevalence was observed in East Africa and relatively low prevalence in Central and West Africa. For instance in Ibadan, Nigeria the prevalence was 23.8% (n=109) (Happi et al., 2005). In Angola the prevalence ranged from 1.9% (n=54) to 38.9% (n=18) in 2007 (Fortes et al., 2011). Within the western parts of the DRC the prevalence ranged from 0.8% in 2003-2004 (n=112) in Basankusa (Cohuet et al., 2006) to 9.5% (n=105) in 2008 in Kinshasa (Mobula et al., 2009) whilst in the eastern parts of the country Alker et al., (2008) found 67% in Rutshuru, near the Rwandan border in 2002. In Ghana the prevalence was 0.79% to 2.6% between 2001 and 2005 (Mockenhaupt et al., 2005; Marks et al., 2005; Pearce et al., 2009). In Mali, the prevalence was 1.0% in two surveys conducted among 902 and 578 people in Kolokani during 2007 (Dicko et al., 2010).

Regional prevalence of *dhps* 540E point mutation over time

The *dhps* 540E point mutation was highly prevalent in the Southeast and Northeast regions where it was ≥50% at 74 sites (Figure 3.10). Overall the Southeast region had the highest regional prevalence of 51% (n=18602) (Table 3.4a). It reached 100% at three sites in the Southeast region and at one site in the Northeast region (Gebru-Woldearegai et al., 2005; Nkhoma et al., 2007; Lynch et al., 2008; Naidoo and Roper, 2011). The prevalence was consistently <50% in the West, Southwest, Central and Island regions over time (Figure 3.11).
Prevalence in the Islands ranged from 0.4% in 2006 at six sites in the Comoros (Andriantsoanirina et al., 2009) to 19.4% in 2004 in Angolares, Sao Tome/Principe (Salgueiro et al., 2010).
Figure 3.11 Surveys of *dhps* 540E prevalence (with 95% confidence intervals) conducted during 1988-2008 and displayed according to their geographic region.
The logistic regression model predicted increasing trends in \textit{dhps} 540E prevalence over time (Table 3.8, Figure 3.12). The likelihood of the \textit{dhps} 540E point mutation occurring in all five regions was lower than it was for the Southeast region, with low odds ratios (<0.934) (Table 3.8). There was strong evidence for this (p<0.001) in all but the Northeast region where p = 0.234 (Table 3.8). The likelihood ratio test provided strong evidence to suggest that the regional prevalence of \textit{dhps} 540E was influenced by time (p< 0.001). Further analyses of the \textit{dhps} 540E point mutation are presented in section 3.4.4 where the Island regions were excluded. The Island populations were excluded in that analysis as it had the least amount of surveys over time ie 17 surveys since 2000.

Table 3.8 Model for \textit{dhps} 540E

<table>
<thead>
<tr>
<th>Covariate</th>
<th>Proportion (%) of samples positive [95% confidence interval]</th>
<th>Odds ratio [95% confidence interval]</th>
<th>P value</th>
<th>Model to estimate effect of region on prevalence of \textit{dhps} 540E</th>
</tr>
</thead>
<tbody>
<tr>
<td>Time (per study year)</td>
<td>34.2[33.6-34.7]</td>
<td>1.3 [1.2-1.30]</td>
<td>&lt;0.001</td>
<td></td>
</tr>
<tr>
<td>Region</td>
<td>Proportion (%) of samples positive [95% confidence interval]</td>
<td>Odds ratio [95% confidence interval]</td>
<td>P value</td>
<td>Odds ratio [95% confidence interval]</td>
</tr>
<tr>
<td>Southeast</td>
<td>50.6[49.9-51.3]</td>
<td>1</td>
<td>1</td>
<td></td>
</tr>
<tr>
<td>Northeast</td>
<td>39.6[37.2-42.1]</td>
<td>0.934 [0.834-1.045]</td>
<td>0.234</td>
<td>0.641 [0.577-0.713]</td>
</tr>
<tr>
<td>West</td>
<td>1.4[1.1-1.8]</td>
<td>0.008 [0.006-0.011]</td>
<td>&lt;0.001</td>
<td>0.014 [0.011-0.018]</td>
</tr>
<tr>
<td>Southwest</td>
<td>3.9[3.1-4.9]</td>
<td>0.022 [0.170-0.028]</td>
<td>&lt;0.001</td>
<td>0.040 [0.031-0.0510]</td>
</tr>
<tr>
<td>Central</td>
<td>0.4[0.2-0.8]</td>
<td>0.004 [0.002-0.007]</td>
<td>&lt;0.001</td>
<td>0.004 [0.002-0.008]</td>
</tr>
<tr>
<td>Island</td>
<td>2.3[1.6-3.3]</td>
<td>0.011 [0.007-0.015]</td>
<td>&lt;0.001</td>
<td>0.023 [0.016-0.034]</td>
</tr>
</tbody>
</table>
Figure 3.12 Predicted probabilities of *dhps* 540E over time
The 581G mutation occurred in 13 countries and was absent in 15 countries between 1988 and 2008 (Figure 3.13; Table 3.3). In Sotuba and Bancoumana, Mali Wang and others (1997) found 100% and 87.5% dhps 581G although sample sizes were small (<10). In Hale and Korogwe, Tanzania the prevalence was 54%-55.5% as observed by Gesase et al., (2009) and Alifrangis et al., (2009). The prevalence of dhps 581G at seven other sites was low i.e. <38% (Curtis et al., 2002; Mbugi et al., 2006; Schonfeld, et al., 2007; Wang et al., 1997). The cumulative prevalence of 581G was only 5.1% (14526/15398) between 1993 and 2008.

Figure 3.13 Prevalence of dhps 581G
Regional prevalence of \textit{dhps} 581G point mutation over time

The \textit{dhps} 581G point mutation was absent only in the Southwest region and overall the prevalence was the highest in the Southeast region at 6.1\% (n=10846) (Table 3.4b, Naidoo and Roper, 2011). Generally the prevalence was <50\% in 24 of 27 sites where \textit{dhps} 581G was found (Figure 3.14). This synopsis excludes a single study for which no study year was available, where the prevalence was 100\% among a small sample size of 10 people as mentioned previously (Wang \textit{et al.}, 1997).

The logistic regression model predicted increasing trends in \textit{dhps} 581G prevalence over time (OR 1.5, CI 1.5-1.6) (Table 3.9, Figure 3.15). The likelihood of the \textit{dhps} 581G point mutation occurring in the Northeast region was similar to the Southeast region (OR 1.1) with weak evidence for this (p=0.760) (Table 3.9). For the other four regions the likelihood of the \textit{dhps} 581G occurring there was lower than it was for the Southeast region and there was strong evidence for this (p<0.001) in the Central and Island regions (Table 3.9). The likelihood ratio test comparing models with and without temporal variation provided strong evidence to suggest that the regional prevalence of \textit{dhps} 581G changed over time (p<0.001).
Figure 3.14 Surveys of dhps 581G prevalence (with 95% confidence intervals) conducted during 1988-2008 and displayed according to their geographic region.
Table 3.9 Model for *dhps* 581G

<table>
<thead>
<tr>
<th>Covariate</th>
<th>Proportion (% of samples positive [95% confidence interval])</th>
<th>Odds ratio [95% confidence interval]</th>
<th>P value</th>
<th>Odds ratio [95% confidence interval]</th>
<th>P value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Time (per study year)</td>
<td>4.8 [4.5-2.5]</td>
<td>1.542 [1.481-1.605]</td>
<td>&lt;0.001</td>
<td>1</td>
<td></td>
</tr>
<tr>
<td>Region</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Southeast</td>
<td>6.1 [5.6-6.5]</td>
<td>1</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Northeast</td>
<td>4.1 [2.8-5.6]</td>
<td>1.058 [0.738-1.516]</td>
<td>0.760</td>
<td>0.658 [0.463-0.936]</td>
<td>0.020</td>
</tr>
<tr>
<td>West</td>
<td>3.0 [2.0-4.4]</td>
<td>0.786 [0.521-1.184]</td>
<td>0.250</td>
<td>0.484 [0.325-0.721]</td>
<td>&lt;0.001</td>
</tr>
<tr>
<td>Central</td>
<td>1.1 [0.6-1.9]</td>
<td>0.283 [0.165-0.484]</td>
<td>&lt;0.001</td>
<td>0.180 [0.106-0.307]</td>
<td>&lt;0.001</td>
</tr>
<tr>
<td>Island</td>
<td>0.1 [0.0-0.5]</td>
<td>0.004 [0.001-0.032]</td>
<td>&lt;0.001</td>
<td>0.014 [0.002-0.103]</td>
<td>&lt;0.001</td>
</tr>
</tbody>
</table>
Coverage of \textit{dhps} 613S was generally poorer than the other mutations with 91 surveys available (Table 3.3).

\textbf{Prevalence of \textit{dhps} 613S}

The \textit{dhps} 613S mutation occurred in 10 countries and was absent in 16 countries (Figure 3.16; Table 3.3). Four surveys conducted in Yaounde, Cameroon in 1994, 1999 and 2001 found \leq3\% 613S (Basco and Ringwald, 1998; Basco \textit{et al.}, 2000; Tahar and Basco, 2007; McCollum \textit{et al.}, 2008). In the Ivory Coast, the prevalence was higher than that found in
Cameroon (Figure 3.16). Two studies done in Ivory Coast reported prevalence of 25.4% (n=118) during 2000-2001 (Djaman et al., 2007) and 22.6% (n=100) in 2006 (Djaman et al., 2010). In 1995, Mali the prevalence of 613S was reported at 11.8% (n=59) (Plowe et al., 1997) and 16.7% (n=48) (Diourte et al., 1999) in Bamako and Soluba (5 km east of Bamako) respectively. In 2005, in Hodh el Gharbi district of Mauritania the prevalence was 6.7% (n=59) in 1998 (Eberl et al., 2001). Generally, the prevalence of dhps 613S in Africa was low at 1.3% (134/10312) (Table 3.3).

Regional prevalence of dhps 613S point mutation over time

The dhps 613S point mutation was absent only in the Southwest region (Table 3.4b) and only 1.4% (n=73) samples were positive in the Islands when surveyed in 2000 (Cravo et al., 2004). A similar proportion of 1.3% (n=10142) samples tested positive overall whilst the Western region had the highest regional prevalence of 12% (n=735) (Table 3.4b). The prevalence of dhps 613S was <33% from 1993 to 2006 (Figure 3.17).
Generally, there was strong evidence for a lower likelihood of *dhps* 613S over time (OR 0.9, CI 0.8-0.9, p<0.001)(Table 3.10, Figure 3.18). The likelihood of the *dhps* 613S point mutation occurring in the Northeast (OR 0.4, CI 0.0-2.6) and Island (OR 0.6, CI 0.1-4.7) regions was lower than it was for the Southeast region with
weak evidence for this as indicated by high p values (Table 3.10). Among the
Central and Island regions, the likelihood of dhps 613S occurring was higher than
it was for the Southeast region with strong evidence for this as p<0.001 (Table
3.10). The likelihood ratio test comparing models with and without temporal
variation provided strong evidence to suggest that the regional prevalence of dhps
613S varied over time (p< 0.001).

Table 3.10 Model for dhps 613S

<table>
<thead>
<tr>
<th>Covariate</th>
<th>Proportion (%) of samples positive [95% confidence interval]</th>
<th>Odds ratio [95% confidence interval]</th>
<th>P value</th>
<th>Model to estimate effect of region on prevalence of dhps 613S</th>
</tr>
</thead>
<tbody>
<tr>
<td>Time (per study year)</td>
<td>1.3 [1.1-1.6]</td>
<td>0.9[0.8-0.9]</td>
<td>&lt;0.001</td>
<td>Odds ratio [95% confidence interval]</td>
</tr>
<tr>
<td>Region</td>
<td></td>
<td></td>
<td></td>
<td>P value</td>
</tr>
<tr>
<td>Southeast</td>
<td>0.3[0.2-0.5]</td>
<td>1</td>
<td></td>
<td>1</td>
</tr>
<tr>
<td>Northeast</td>
<td>0.1[0.0-0.7]</td>
<td>0.4[0.0-2.6]</td>
<td>0.314</td>
<td>0.4[0.0-2.9]</td>
</tr>
<tr>
<td>West</td>
<td>12.0[9.7-14.5]</td>
<td>33.4 [20.8-53.8]</td>
<td>&lt;0.001</td>
<td>40.6 [25.5-64.7]</td>
</tr>
<tr>
<td>Central</td>
<td>2.5[1.5-3.8]</td>
<td>5.9[3.2-10.8]</td>
<td>&lt;0.001</td>
<td>7.7[4.2-14.0]</td>
</tr>
<tr>
<td>Island</td>
<td>0.1[0.0-0.7]</td>
<td>0.6[0.1-4.7]</td>
<td>0.648</td>
<td>0.4[0.1-2.9]</td>
</tr>
</tbody>
</table>
Three countries had the \textit{dhps} 613T point mutation and it was absent in 21 countries (Figure 3.19; Table 3.3). The mutation was reported in Bamako, Mali (11.8%, n=59) (Plowe \textit{et al.}, 1997), Blantyre, Malawi (1.1%, n=89) (Alker \textit{et al.}, 2005) and at 12 sites in Madagascar with pooled data ($\leq$0.5%) (Andriantsoanirina \textit{et al.}, 2009). The cumulative prevalence of \textit{dhps} 613T in Africa was the lowest of all mutations at 0.1% (11/10152) between 1988 and 2008 (Table 3.3).
Regional prevalence of \textit{dhps} 613T point mutation over time

The \textit{dhps} 613T point mutation was observed in the Southeast, West and Island regions and 0.11\% (n=9982) samples tested positive overall (Table 3.4b). The Western region had the highest regional prevalence of 1.1\% (n=609) (Figure 3.20).

There was strong evidence for a decreasing temporal trend in \textit{dhps} 613T prevalence (OR 0.7, CI 0.6-0.9,\(p=0.002\)) (Table 3.11, Figure 3.21). The likelihood of \textit{dhps} 613T occurring in the West and Island regions were much higher than that of the Southeast region (Table 3.11). However the relatively high odds ratios and wide confidence intervals must be interpreted with caution given the low prevalence of \textit{dhps} 613T in all three regions and relatively limited number of surveys available. The results of further multivariate analyses for \textit{dhps} 437G and 540E are presented next.
Figure 3.20 Surveys of *dhps* 613T prevalence (with 95% confidence intervals) conducted during 1988-2008 and displayed according to their geographic region.
Table 3.11 Model for *dhps* 613T

<table>
<thead>
<tr>
<th>Covariate</th>
<th>Proportion (%) of samples positive [95% confidence interval]</th>
<th>Odds ratio [95% confidence interval]</th>
<th>P value</th>
<th>Odds ratio [95% confidence interval]</th>
<th>P value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Time (per study year)</td>
<td>0.1[0.1-0.2]</td>
<td>0.7[0.6-0.9]</td>
<td>0.002</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Region</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Southeast</td>
<td>0.01[0-0.1]</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td></td>
</tr>
<tr>
<td>West</td>
<td>1.1[0.5-2.4]</td>
<td>41.4 [5.0-341.3]</td>
<td>0.001</td>
<td>80.1[9.8-652.4]</td>
<td>&lt;0.001</td>
</tr>
<tr>
<td>Island</td>
<td>0.3[0.1-0.8]</td>
<td>103.1 [7.6-1402.6]</td>
<td>&lt;0.001</td>
<td>19.4[2.0-186.3]</td>
<td>0.010</td>
</tr>
</tbody>
</table>
Figure 3.21 Predicted probabilities of \textit{dhps} 613T over time
3.4.3 Further analysis of dhps 437G

The logistic regression estimated the joint effects of time and region and assumed there was no interaction between the effect of study year and region on the mutation prevalence. The results of the regression model for the effect of time (study year, 1988-2008) and region on dhps 437G prevalence are tabulated in Table 3.12. Overall the prevalence of dhps 437G was 50.9%, (n=28605) (Tables 3.4a and 3.12).

Table 3.12 Further analysis of dhps 437G

<table>
<thead>
<tr>
<th>Covariate</th>
<th>Proportion (%) of samples positive [95% confidence interval]</th>
<th>Odds ratio [95% confidence interval]</th>
<th>P value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Time (per study year 1988-2008)</td>
<td>50.9 [50.3-51.5]</td>
<td>1.2 [1.0-1.3]</td>
<td>0.030</td>
</tr>
<tr>
<td>Region</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Southeast</td>
<td>51.0 [50.3-51.7]</td>
<td>1</td>
<td></td>
</tr>
<tr>
<td>Northeast</td>
<td>42.1 [39.6-44.6]</td>
<td>0.9 [0.3-3.0]</td>
<td>0.869</td>
</tr>
<tr>
<td>West</td>
<td>44.7 [43.3-46.2]</td>
<td>0.6 [0.1-2.4]</td>
<td>0.473</td>
</tr>
<tr>
<td>Southwest</td>
<td>73.4 [70.8-75.8]</td>
<td>2.5 [1.0-6.1]</td>
<td>0.047</td>
</tr>
<tr>
<td>Central</td>
<td>56.2 [54.2-58.3]</td>
<td>1.3 [0.5-3.3]</td>
<td>0.616</td>
</tr>
</tbody>
</table>

The effect of time

The model predicted that for every year, the likelihood of a person having a dhps 437G mutant was 1.2 times higher (CI 1.0-1.3, p=0.030, Table 3.5). This suggests that there is a steady increase in the point mutation prevalence with time.

Regions as risk factors

The Wald test provided evidence to suggest that there were regional differences in the prevalence of dhps 437G overall (p=0.0034). For the Central and Southwest regions the likelihood of the 437G mutant occurring there was greater than it was for the Southeast region (OR 1.3, CI 0.5-3.3 and OR 2.5, CI 1.0-6.1 respectively; Figure 3.22). Conversely, the likelihood of the dhps 437G mutation occurring in the Northeast (OR 0.9, CI 0.3-3.0) and Western region (OR 0.6, 0.1-2.4) was lower than it was for the Southeast region.
(Figure 3.22). The relatively high p values indicated that the strength of the evidence for this was weak (Table 3.12).

---

**Figure 3.22** Predicted probabilities of regional *dhps* 437G prevalence over time

- **Southeast**
  - Democratic Republic of Congo (east), Kenya, Malawi, Mozambique, Rwanda, South Africa, Swaziland, Tanzania, Uganda, Zambia, Zimbabwe

- **Northeast**
  - Djibouti, Ethiopia, North Sudan, South Sudan

- **West**
  - Benin, Burkina Faso, Ivory Coast, Gambia, Ghana, Guinea, Guinea-Bissau, Liberia, Mali, Mauritania, Niger, Nigeria, Senegal

- **Southwest**
  - Angola, Democratic Republic of Congo (west), Gabon, Namibia, Republic of Congo

- **Central**
  - Cameroon, Central African Republic

---

**predicted probability logistic model**

**prevalence of dhps 437**
Interaction between region and time

Given the observed increasing trend of *dhps* 437G prevalence over time, an interaction term between time (study year) and region was introduced into a logistic model to assess whether or not the relationship between the 437G prevalence and time differed among the five regions. The stratum–specific odds ratios for each region are presented in Table 3.13 quantifying the extent to which the odds ratios for time (study year) and each regional group, differs from those in the reference Southeast region.

The odds ratios differed between regional strata, providing evidence to suggest that there was an interaction between time (study year) and region and the likelihood of *dhps* 437G prevalence i.e. higher temporal trend in prevalence differed depending on the regional categorisation and there was strong evidence for this (p <0.001), based on chi² and 4 degrees of freedom (Table 3.13). The AIC for the model without interaction was 1.337 whilst the model with interaction had an AIC of 1.287 providing support for the model with interaction between region and time.

The main difference between the two models was in the Western region where the temporal trend was lower, although there was weak evidence for this (OR 0.9, CI 0.8-1.0, p=0.097) (Figure 2.23). There was strong evidence for increasing prevalence of *dhps* 437G in the Northeast and Southwest regions (p<0.001) (Table 3.13).
Table 3.13 Model for \( dhps \) 437G with interaction between time and region

<table>
<thead>
<tr>
<th>Region</th>
<th>Difference in slope (per year) for each region</th>
<th>For each year compared to Southeast region. Study year X region</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Odds ratio [95% confidence interval]</td>
<td>Odds ratio [95% confidence interval]</td>
</tr>
<tr>
<td>Southeast</td>
<td>1.2 [1.0-1.4]</td>
<td>1</td>
</tr>
<tr>
<td>Northeast</td>
<td>2.4 [1.8-3.2]</td>
<td>&lt;0.001 2.0 [1.4-2.8]</td>
</tr>
<tr>
<td>West</td>
<td>0.9 [0.8-1.0]</td>
<td>0.097 0.7 [0.6-0.9]</td>
</tr>
<tr>
<td>Southwest</td>
<td>1.2 [1.1-1.3]</td>
<td>&lt;0.001 1.0 [0.8-1.2]</td>
</tr>
<tr>
<td>Central</td>
<td>1.0 [0.9-1.2]</td>
<td>0.471 0.9 [0.7-1.0]</td>
</tr>
</tbody>
</table>
Figure 3.23 Predicted probabilities of dhps 437G prevalence with interaction effects
3.4.4 Further analysis of \textit{dhps} 540E

The effect of time

The model predicted that for every year, the likelihood of a person having a \textit{dhps} 540E mutant was 1.3 times higher (CI 1.1-1.5, \( p=0.004 \)) (Table 3.14). This provided evidence to suggest that there was an increase in the point mutation prevalence with time, as was observed with the \textit{dhps} 437G mutant. The overall prevalence of \textit{dhps} 540E was 34.2\% (\( n=29824 \)) (Tables 3.4a and 3.14).

Regions as risk factors

The likelihood of the \textit{dhps} 540E in the Central, Southwest and Western regions was lower than it was for the Southeast region, and this likelihood difference was large as the ORs were low (Figure 3.24, Table 3.14). This evidence was strong for the Central, Southwest and Western regions (\( p<0.001 \), Table 3.14). The Wald test provided evidence to suggest that there were regional differences in the prevalence of \textit{dhps} 540E overall (\( p<0.001 \)). Hence, the interaction between time and region was analysed next.

Table 3.14 Further analysis of \textit{dhps} 540E

<table>
<thead>
<tr>
<th>Covariate</th>
<th>Proportion of samples positive (95% confidence interval)</th>
<th>Multivariable analysis for \textit{dhps} 540E prevalence</th>
<th>Odds ratio [95% confidence interval]</th>
<th>( P ) value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Time (per study year) (1988-2008)</td>
<td>34.2[33.6-34.7]</td>
<td>1.258 [1.1-1.473]</td>
<td>0.004</td>
<td></td>
</tr>
<tr>
<td>Region</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Southeast</td>
<td>50.6[49.9-51.3]</td>
<td>1</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Northeast</td>
<td>39.6[37.2-42.1]</td>
<td>0.940 [0.292-3.024]</td>
<td>0.917</td>
<td></td>
</tr>
<tr>
<td>West</td>
<td>1.4[1.1-1.8]</td>
<td>0.008 [0.002-0.044]</td>
<td>&lt;0.001</td>
<td></td>
</tr>
<tr>
<td>Southwest</td>
<td>3.9[3.1-4.9]</td>
<td>0.0216 [0.006-0.079]</td>
<td>&lt;0.001</td>
<td></td>
</tr>
<tr>
<td>Central</td>
<td>0.43[0.2-0.8]</td>
<td>0.004 [0.001-0.020]</td>
<td>&lt;0.001</td>
<td></td>
</tr>
</tbody>
</table>
Figure 3.24 Predicted probabilities of regional *dhps* 540E prevalence over time

Democratic Republic of Congo (east), Kenya, Malawi, Mozambique, Rwanda, South Africa, Swaziland, Tanzania, Uganda, Zambia, Zimbabwe

Djibouti, Ethiopia, North Sudan, South Sudan

Benin, Burkina Faso, Ivory Coast, Gambia, Ghana, Guinea, Guinea-Bissau, Liberia, Mali, Mauritania, Niger, Nigeria, Senegal

Angola, Democratic Republic of Congo (west), Gabon, Namibia, Republic of Congo

Cameroon, Central African Republic
There was strong evidence of a temporal trend which differed among the regions (p <0.001, based on chi² and 4 degrees of freedom). The AIC for the model without interaction was 0.962 whilst the model with interaction had an AIC of 0.942 providing support for the model with interaction.

**Interaction between region and time**

The likelihood of the *dhps* 540E occurring under conditions of an interaction between time and region, compared with the Southeast regions was higher, in all but the Western region. There was strong evidence for an increased likelihood of 540E in the Northeast and Southwest regions (p< 0.001, Table 3.15, Figure 3.25).

<table>
<thead>
<tr>
<th>Region</th>
<th>Covariate</th>
<th>Difference in slope (per year) for each region</th>
<th>Odds ratio [95% confidence interval]</th>
<th>P value</th>
<th>Odds ratio [95% confidence interval]</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Southeast</td>
<td>1.2 [1.0-1.4]</td>
<td>0.014</td>
<td>1</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Northeast</td>
<td>2.7 [1.8-3.9]</td>
<td>&lt;0.001</td>
<td>2.2 [1.5-3.3]</td>
<td></td>
</tr>
<tr>
<td></td>
<td>West</td>
<td>1.1[0.9-1.2]</td>
<td>0.374</td>
<td>0.9 [0.7-1.0]</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Southwest</td>
<td>1.7[1.3-2.3]</td>
<td>0.001</td>
<td>1.4 [1.0-2.0]</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Central</td>
<td>2.9 [1.1-7.6]</td>
<td>0.031</td>
<td>2.4 [0.9-6.4]</td>
<td></td>
</tr>
</tbody>
</table>
Figure 3.25 Predicted probabilities of *dhps* 540E prevalence with interaction effects
This chapter describes regional temporal and spatial variation in \textit{dhfr} and \textit{dhps} point mutation prevalence. It shows that the \textit{dhfr} 164L and \textit{dhps} 613S/T are rare on the continent, whereas the \textit{dhps} 437G and 540E mutants have systematically increased in prevalence through time across sub-Saharan Africa. Using the observed distributions of these point mutations they can be characterised into three broad categories which reflect the degree of SP resistance conferred: (1) scarce or not fully resistant (\textit{dhps} 436A, \textit{dhps} 613S), (2) emerging foci of resistance (\textit{dhfr} 164L, \textit{dhps} 581G) and (3) widespread, major resistance (\textit{dhps} 540E, \textit{dhps} 437G). The point mutations that fall into each category will be discussed in turn and thereafter some limitations of this study and recommendations for future work will be presented.

The first group describes those point mutations that are neither abundant nor fully resistant. The \textit{dhps} 436A point mutation was reported across sub-Saharan Africa with the highest prevalence reported among Western African countries. Closer inspection of the \textit{dhps} 436A distribution as well as the results of the logistic model, show that it has not increased in prevalence, but rather it has decreased over time and is possibly being displaced by the 436S sensitive forms or the other mutant alleles which dominate in these regions. This lends support to the suggestion that it confers no resistance to SP and that it is an ancestral variant which may have emerged or been lost many times over (Pearce et al., 2009). Pearce et al., (2009) found that the 436S/437A/540K allelic haplotype was no more resistant than the 436A/437A/540K allelic haplotype. The 436A occurs predominately in West African countries and is less prevalent in the Eastern region. The 613S/T point mutations also belong to this group, due them to being absent in the majority of sub-Saharan African countries with sporadic foci in the Southeast and Western regions.

The second group describes point mutations with distinctly emerging foci of resistance namely the \textit{dhfr} 164L and \textit{dhps} 581G point mutations. As described in chapter 2, there was reasonably good coverage of studies that looked at the \textit{dhfr} 164L point mutation and it is apparent from its low prevalence that it is different from the other \textit{dhfr} point mutations. Generally the 164L has a restricted distribution in sub-Saharan Africa, but it has emerged in seven countries primarily in East and Central Africa. Although it is still relatively rare across Africa, it is associated with high levels of pyrimethamine resistance when occurring with the \textit{dhfr} triple mutant, which is cause for concern for the IPT strategies in these regions. Indeed the logistic model presented here predicted an increase in \textit{dhfr} 164L over time. Hamel et al., (2008) did not find evidence for an association between \textit{dhfr} 164L and
treatment failure in their study in Kisumu, Kenya during 2002-2003 unlike their Asian counterparts (Hyde et al., 2008). It is possible that the effect of 164L in the quadruple mutant is complex and the ecological context under which this haplotype combination occurs differs from the situation in Asia (Hyde et al., 2008, McCollum et al., 2008).

The dhps 581G point mutations are newly emerging in some foci but are absent in most of sub-Saharan Africa. The dhps 581G when found on its own is not as powerful an indicator of resistance as when it is found in conjunction with the dhps double mutant (437G+540E) and dhfr triple mutant (51I+59R+108N). The dhps 581G is increasing in East Africa and was found at low prevalence (<40%) in West Africa within Mali, Niger and Ghana. In East Africa there is an association between dhps 581G and SP resistance but it is unclear whether or not such an association occurs in West Africa.

The third group describes the dhps 540E and dhps 437G point mutations both of which are widespread and confer major resistance when found together. These point mutations were analysed further in this chapter because the distribution of dhps 540E is more regionally distinct than that of dhps 437G and dhps 437G is generally found together with dhps 540E.

Generally dhps 437G and 540E point mutations were found predominantly in the Southeast regions and were less prevalent in the Western regions. This was supported by the finding that the predicted probabilities of dhps 437G and 540E in the Western and Southwestern regions were generally less than that found in the Southeast region. Thus they appear at a slower rate in the Western region compared with the Eastern region. Furthermore, the analyses showed that dhps 437G and 540E prevalence are generally increasing over time within sub-Saharan Africa, but there are apparent differences in the prevalence rates between the Eastern and Western countries.

Countries in Central Africa represent a watershed for the distribution of antimalarial resistance (Taylor et al., 2013). The DRC was allocated to both Southeast and Southwest Africa in this study. This is supported by chloroquine and SP efficacy studies in the DRC showing that the highest proportion of treatment failures occurred in the Eastern areas (Kazadi et al., 2003) where SP was used since 1995 (Alker et al., 2008). The dhps 437G/540E double mutant occurred at higher proportions in sites in Eastern DRC (Kisangani, Kilwa) compared with Western DRC (Basankusa, Boende) (Cohuet et al., 2006). Furthermore, in Rutshuru, which is in Eastern DRC near the Rwandan border, point
mutations at \textit{dhps} 437G and 540E were strongly associated with SP treatment failure (Alker \textit{et al.}, 2008).

Generally, relatively higher prevalence of \textit{dhps} 437G was observed since 2000. Although there was high prevalence of \textit{dhps} 437G (>80%) within the West African region, the model estimates of the predicted probabilities of \textit{dhps} 437G prevalence with interaction of region and time showed an anomalous declining trend in the Western region. This finding could be attributed to several factors including the following: (1) outliers of relatively low \textit{dhps} 437G prevalence ie 20.6\% (n=911) and 18.8\% (n=648) in Kolokani, Mali (Dicko \textit{et al.}, 2010) and 37.5\% (n=40) in Ndoffane, Senegal (Cisse \textit{et al.}, 2009) which contributed to the decreasing trend in prevalence since 2006, (2) the effects of sampling variation among the studies, (3) the region encompasses a diverse spectrum of communities as it represents 13 countries, covering an area of approximately six million square kilometres and (4) less drug pressure in these West African countries due to low SP use (Frosch, Venkatesan and Laufer, 2011) compared to the other regions.

The likelihood of \textit{dhps} 437G were the same in the Southeast and Southwest regions (p< 0.001), suggesting that similar temporal trends occur in both these regions. This finding was consistent among the models with and without the interaction effect of time and regional location. For \textit{dhps} 437G the Southern African countries were similar to each other and different from the Western, Central and Northeast regions.

As stated above \textit{dhps} 437G is generally found where \textit{dhps} 540E occurs and this pattern was observed distinctly in the Southeast and Northeast regions but not in the West, Southwest, Central and Island regions. There are sufficient longitudinal studies measuring the prevalence of these point mutations in the Southeast and Northeast regions, suggesting that these analyses do reflect changes in prevalence over time. However, further systematic sampling of these point mutations is needed among West African countries, which would allow close inspection of the changes in the prevalence of these point mutations over time.

Generally the logistic models provided evidence of increasing \textit{dhfr} 164L, \textit{dhps} 437G, \textit{dhps} 540E \textit{dhps} 581G prevalence over time and decreasing \textit{dhps} 436A, \textit{dhps} 613S/T prevalence over time at a population level and that relationship is conditional depending on the region. The models showing the interaction between time and region provided a better
fit compared to the models without interaction. This regional effect could be due to variation in drug exposure among the regions. Frosch et al., (2011) reported that SP use was more prolific in East African countries compared to West African countries and this frequency of SP use was related to the seasonality of malaria transmission being different among these regions.

Notably the prevalence of each point mutation was different among the Islands compared with the other regions. For example, in the Island region the 164L mutant was found in a unique combination of point mutations that has not been reported in mainland Africa to date (Naidoo and Roper, 2011). Island populations were excluded from the statistical model development because dhps point mutations evolved independently from mainland Africa (Naidoo and Roper, 2011).

Limitations

The differences in prevalence of the point mutations reflect the duration and magnitude of SP use and to some extent they predict population levels of in vivo SP resistance. The correlation between molecular markers and in vivo (phenotypic) resistance is discussed further and investigated in chapter 5. Prevalence measures of individual point mutations are a crude estimate of resistance levels and alternative means of quantifying resistance mutations in a population are employed in chapters 4 and 5 to characterise the frequencies of both individual and combinations of point mutations. Frequencies have the effect of smoothing mutation data in highly endemic and high transmission areas. Allelic haplotypes describe combinations of point mutations and reflect the joint effect of individual mutations which occur together. Using frequency and haplotype data may offer further insight into the spatial and temporal changes of resistance markers.

Recommendation

There is a need for further epidemiological surveillance of dhfr and dhps point mutations across Africa, specifically dhfr 164L, dhps 540E and dhps 581G. These mutations are of the greatest significance for policy in the context of continued use of SP for IPTp and the potential implementation of SP-IPTi in Africa. The prevalence of resistance markers amongst pregnant women participating in IPTp programmes can differ from background levels, hence there is a need for regular regional surveillance of the prevalence of SP point
mutations in high risk groups such as pregnant women, infants and children where SP will be employed.
4.1 Introduction

People who are bitten by many *P. falciparum* infected mosquitoes often harbour super-infections with multiple genotypes co-infecting individuals. Blood stage parasites are haploid and separate genotypes within the same infection are referred to as clones. These haploid parasite clones are a precursor for the diploid zygotic stage of the parasite lifecycle which occurs in the mosquito when gametocytes fuse and genetic material exchange through meiosis in the oocyst. Within any given *P. falciparum* infection sampled at the blood stage (usually by finger prick, sometimes by venipuncture) multiple haploid clones/genotypes can be detected. There are records of as many as seven distinct clones or genotypes in one infection (Hastings, Paget-McNicol and Saul, 2004) and this number is referred to as the multiplicity of infection (MOI).

Within a mixed infection there can be sensitive forms alone, mutant forms alone or mixtures of both mutant and sensitive forms. The prevalence data for individual point mutations that were used in chapter 3 recorded the proportion of infections in which a specific mutation was detected regardless of the underlying MOI of those infections. Hastings, Nsanzabana and Smith (2010) pointed out that genotype prevalence - defined as the proportion of blood samples (people) that carry a particular marker - does not distinguish whether or not one or more clones in the sample carry that marker. Prevalence of individual point mutations is the most commonly reported measure in genotyping studies, but it can be an imperfect way of assessing the level of resistance in a parasite population because it does not take any account of MOI and crucially it is the combination of point mutations that determines the level of drug sensitivity of any given parasite genotype.

Only a subset of the studies identified in the review conducted in this study, actually recorded the combination of point mutations at all codon positions in the *dhfr* or *dhps* genes in individual infections. These studies can be used to identify which specific combinations of mutations are commonly circulating among the parasite populations in Africa. Point mutations have different properties when found in such combinations. These combinations of point mutations are referred to hereon as allelic haplotypes in this thesis. An allelic haplotype describes a combination of point mutations which occur in the same gene. Allelic haplotypes can only be verified in non-mixed infections and cannot be
deduced from mixed infections due to the unknown ratio of mutant to sensitive forms in neighbouring codon positions. Analyses of haplotypes assist in assessing the intensity of SP resistance.

Researchers would generally exclude mixed infections from their analysis in order to provide definitive summaries of the combinations of point mutations that are in circulation. To compare levels of SP resistance in different populations, the number of non-mixed infections with each combination of point mutations is counted and these are used as a measure of the relative frequency of each allelic haplotype in a given population. Hastings et al., (2010) defined frequency in this context as the proportion of the parasite population which carry the marker. The frequency is measured by counting the numbers of individual parasite clones that have the marker and in so doing, the different resistance rates that arise from populations residing in different transmission intensities are standardised. Thus, the frequency measure is important for comparing SP resistance levels among populations.

In any study site, the number of mixed infections depends on the intensity of malaria transmission in that area, so the exclusion of mixed infections from calculations of allelic haplotype frequencies is most extreme in surveys conducted in highly endemic areas. Thus this is particularly an African issue where malaria endemicity tends to be high in foci and hotspots. In this chapter haplotype frequency data are analysed in order to better understand the spatio-temporal changes observed among the single point mutation prevalence data which were examined in chapter 3.

As mentioned in chapter 2, the most common resistance haplotype of dhfr in Africa is the 51I/59R/108N triple mutant but there are also mildly resistant haplotypes which have fewer mutations, as well as the sensitive allelic haplotype which are shown in Table 4.1. The amino acids are abbreviated into three and single letter codes in Table 4.1. Combinations of point mutations are indicated by shading.
Table 4.1 Common dhfr haplotypes in Africa

<table>
<thead>
<tr>
<th>Codon</th>
<th>50</th>
<th>51</th>
<th>59</th>
<th>108</th>
<th>140</th>
<th>164</th>
</tr>
</thead>
<tbody>
<tr>
<td>Sensitive</td>
<td>Cys (C)</td>
<td>Asn (N)</td>
<td>Cys (C)</td>
<td>Ser (S)</td>
<td>Val (V)</td>
<td>Ile (I)</td>
</tr>
<tr>
<td>Single mutant</td>
<td>Cys (C)</td>
<td>Asn (N)</td>
<td>Cys (C)</td>
<td>Asn (N)</td>
<td>Val (V)</td>
<td>Ile (I)</td>
</tr>
<tr>
<td>Double mutant</td>
<td>Cys (C)</td>
<td>Ile(I)</td>
<td>Cys (C)</td>
<td>Asn (N)</td>
<td>Val (V)</td>
<td>Ile (I)</td>
</tr>
<tr>
<td>Double mutant</td>
<td>Cys (C)</td>
<td>Asn (N)</td>
<td>Arg (R)</td>
<td>Asn (N)</td>
<td>Val (V)</td>
<td>Ile (I)</td>
</tr>
<tr>
<td>Triple mutant</td>
<td>Cys (C)</td>
<td>Ile(I)</td>
<td>Arg (R)</td>
<td>Asn (N)</td>
<td>Val (V)</td>
<td>Ile (I)</td>
</tr>
</tbody>
</table>

The two most common dhps resistant haplotypes are the single mutant 437G and the double mutant 437G/540E which are shaded in Table 4.2.

Table 4.2 Common dhps haplotypes in Africa

<table>
<thead>
<tr>
<th>Codon</th>
<th>436</th>
<th>437</th>
<th>540</th>
<th>581</th>
<th>613</th>
</tr>
</thead>
<tbody>
<tr>
<td>Sensitive</td>
<td>Ser (S or A)</td>
<td>Ala (A)</td>
<td>Lys (K)</td>
<td>Ala (A)</td>
<td>Ala (A)</td>
</tr>
<tr>
<td>Single mutant</td>
<td>Ser (S or A)</td>
<td>Gly (G)</td>
<td>Lys (K)</td>
<td>Ala (A)</td>
<td>Ala (A)</td>
</tr>
<tr>
<td>Double mutant</td>
<td>Ser (S or A)</td>
<td>Gly (G)</td>
<td>Glu (E)</td>
<td>Ala (A)</td>
<td>Ala (A)</td>
</tr>
</tbody>
</table>

The combinations of dhfr and dhps allelic haplotypes differ in the resistance levels they confer. Table 4.3 shows the increasing trend in SP resistance as the number of mutations accumulate with the highest resistance levels associated with the quintuple mutant (dhfr triple mutant plus dhps double mutant). For instance, based on in vitro studies, the dhps GK allelic haplotype is less resistant than the dhps GE double mutant due to the presence of the dhps 540K sensitive form in this combination. Although the dhfr IRN triple mutant
was a good general indicator of emerging SP resistance in areas where SP was newly introduced, it is its combination with the *dhps* double mutant (Table 4.2), known as the quintuple mutant mentioned above, that is predictive of *in vivo* treatment failure (Tables 4.2 and 4.3) (Nzila et al., 2000a; Kublin et al., 2002; Bwijo et al., 2003) and this is associated with full SP resistance.

Table 4.3 Combinations of *dhfr* and *dhps* haplotypes and levels of SP resistance

<table>
<thead>
<tr>
<th>dhfr</th>
<th>Sensitive NCS</th>
<th>Single mutant NCN</th>
<th>Double mutant NRN</th>
<th>Double mutant ICN</th>
<th>Triple Mutant IRN</th>
</tr>
</thead>
<tbody>
<tr>
<td>Sensitive AK</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Single Mutant GK</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>West and Central Africa</td>
</tr>
<tr>
<td>Double Mutant GE</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>East Africa</td>
</tr>
</tbody>
</table>

Chapter 3 showed that *dhps* 540E predominates in East Africa and is comparatively less common in West and Central Africa. The *dhps* 437G single mutant haplotype in combination with the *dhfr* IRN triple mutant is found throughout West and Central Africa but this quadruple genotype is not as fully resistant as the quintuple genotype that is found in East Africa (Table 4.3).

In this chapter, haplotype data for the *dhfr* sensitive allelic haplotype (51N/59C/108S), *dhfr* triple mutant allelic haplotype (51I/59R/108N), *dhps* GE double mutant, *dhps* GK single mutant and *dhps* AK sensitive alleles are summarised and spatio-temporal trends in their distribution are analysed.
4.2 Aims and Objectives

1. To gather data on the frequency of the \textit{dhfr} sensitive allelic haplotype (51N/59C/108S), \textit{dhfr} triple mutant allelic haplotype (51I/59R/108N) and the three \textit{dhps} haplotype alleles 437G/540E (GE), 437G/540K (GK) and 437A/540K (AK) within sub-Saharan Africa.

2. To investigate the influence of temporal and regional factors on the frequencies of these haplotypes.

4.3 Methods

Data processing

The data collection methods and systematic data abstraction methods were described in chapter 2. Data points containing the allelic haplotypes \textit{dhfr} NCS, \textit{dhfr} IRN as well as \textit{dhps} GE, GK and AK were collated, then extracted from the database and imported into Stata. Each data point was assigned into five regions namely Southeast, Northeast, West, Southwest and Central. Countries which fall into each region are listed in Table 3.1 in chapter 3.

Only studies which reported haplotypes could be included in these analyses. Samples with mixed infections (where haplotypes could not be determined) were excluded from the analyses. No further exclusions were applied based on age and whether or not patients were symptomatic for malaria.

The data were first mapped to observe the spatial distribution of the allelic haplotypes. Thereafter multivariable regression analyses were performed to further assess the relative contributions of the risk factors (time and region) on the allelic haplotype frequencies. The methods are discussed next.
4.3.1 Data Analysis

Spatial mapping of allelic haplotype frequency

For each survey, the frequency was calculated as the number of samples with the allelic haplotype/total number of samples analysed, after excluding mixed infections as described in equation one.

Equation 1:

\[ y = \left( \frac{\text{number of samples with the allelic haplotype}}{\text{total number of samples tested for which an allelic haplotype could be determined}} \right) \times 100 \]

Where \( y \) = frequency of allelic haplotype

Vector maps were created in Mapinfo to illustrate the spatial distribution of haplotypes. The maps included studies where the study year was not available.

Multivariable regression analyses for allelic haplotypes

The same approach was followed as in chapter 3. The data were converted into binary format (0/1) and expanded using the total number of people with haplotype determined, so that logistic regression analyses could be performed in Stata. A logistic model was constructed to assess whether or not there was any association between the outcome variable (frequency of allelic haplotype) and the risk factors study year (time) and region.

The log odds of the outcome (allelic haplotype present = 1) was used for logistic regression analysis with robust standard errors allowing for with and without clustering by country to model the effect of time on each allelic haplotype frequency (equation four). Odds ratios, 95% confidence intervals and cumulative p values are reported for the model estimates for each allelic haplotype. Island populations were included in the preliminary analyses. After running the logistic regression, the likelihood ratio test was used to compare the models with and without allelic haplotype frequency changing over time.

Thereafter further multivariate logistic analyses were performed on the \( \text{dhps GE}, \text{GK and AK} \) allelic haplotypes using an interaction term, which multiplied the effect of time (study year) and region as in equation five. For these models the Southeast region was consistently used as the reference region because it had the best coverage of data points for each haplotype over the 20 year period (Tables 4.4 and 4.5). The Island populations
were excluded in these analyses because it had the least amount of data points and these were generally conducted after 2002.

Equation 4: \( \log(p/(1-p)) = \beta_1 x_1 + \beta_2 x_2 \)

Equation 5: \( \log(p/(1-p)) = \beta_1 x_1 + \beta_2 x_2 + \beta_3 x_3 \)

where

\( p = \) probability of allelic haplotype being present

\( 1-p = \) probability of allelic haplotype being absent

\( x_1 = \) study year

\( x_2 = \) region

\( x_3 = \) region \( \times \) study year

**Post estimation**

The Wald test was used to assess the strength of evidence for differences in frequencies of allelic haplotypes among the regions from the model with interactions. This test tests the null hypothesis that the frequency is the same in all regions, given the variation in sampling. The goodness of fit of each model was estimated using AIC (see post estimation in chapter 3). Estimating the linear combinations of the odd ratios sequentially for each region quantified how time (study year) affected the frequency among the regions. The Southeast region was used as the reference against which the other regions were compared (equation six)

Equation 6: \( \log(p/(1-p)) = \beta_1 x_1 + \beta_2 nx_2 + \beta_3 x_3 \)

Where \( n = \) each region

After running the model, graphs were constructed in Stata using the model estimates of the predicted probabilities of positivity \((Y=1)\) in each region.

**4.3.2 Assumptions and missing data**
Assumptions were made as described in chapter 3. For the plots of allelic haplotypes exclusion criteria were surveys that did not have study years and surveys that did not report haplotypes.

The results for each allelic haplotype are presented sequentially in section 4.4.

4.4 Results

4.4.1 Frequency maps of mutations and allelic haplotypes

A series of five maps are presented, illustrating the frequency of each allelic haplotype, as well as five sets of graphs showing the regional frequency with 95% confidence intervals. Tables 4.4 and 4.5 outlines the data represented in the five maps which shows that there were in excess of 195 surveys available for \(dhfr\) IRN and \(dhfr\) NCS in 34 countries and 152 surveys for \(dhps\) GE/GK/AK in 37 countries. A breakdown of sample sizes by region is given in Tables 4.4 and 4.5. The results for \(dhfr\) NCS and IRN are presented first, followed by the results for \(dhps\) GE, GK and AK. The references used for each haplotype are provided in appendices nine through 11.
Table 4.4 Summary of surveys available for *dhfr* NCS and IRN allelic haplotypes

<table>
<thead>
<tr>
<th>Regions</th>
<th>Southeast</th>
<th>Northeast</th>
<th>West</th>
<th>Southwest</th>
<th>Central</th>
<th>Islands</th>
<th>All Regions</th>
</tr>
</thead>
<tbody>
<tr>
<td>Surveys of <em>dhfr</em> NCS</td>
<td>97</td>
<td>11</td>
<td>35</td>
<td>13</td>
<td>28</td>
<td>11</td>
<td>195</td>
</tr>
<tr>
<td>Samples tested for <em>dhfr</em> NCS</td>
<td>8022</td>
<td>1022</td>
<td>3466</td>
<td>805</td>
<td>2661</td>
<td>2397</td>
<td>18373</td>
</tr>
<tr>
<td>Samples positive for <em>dhfr</em> NCS</td>
<td>1204</td>
<td>179</td>
<td>1170</td>
<td>83</td>
<td>658</td>
<td>782</td>
<td>4076</td>
</tr>
<tr>
<td>Unique sites with <em>dhfr</em> NCS</td>
<td>52</td>
<td>10</td>
<td>20</td>
<td>12</td>
<td>20</td>
<td>8</td>
<td>122</td>
</tr>
<tr>
<td>Countries with <em>dhfr</em> NCS</td>
<td>11</td>
<td>3</td>
<td>11</td>
<td>4</td>
<td>2</td>
<td>3</td>
<td>34</td>
</tr>
<tr>
<td>Surveys of <em>dhfr</em> IRN</td>
<td>106</td>
<td>8</td>
<td>31</td>
<td>12</td>
<td>29</td>
<td>11</td>
<td>197</td>
</tr>
<tr>
<td>Samples tested for <em>dhfr</em> IRN</td>
<td>9682</td>
<td>853</td>
<td>3404</td>
<td>745</td>
<td>2828</td>
<td>1757</td>
<td>19269</td>
</tr>
<tr>
<td>Samples positive for <em>dhfr</em> IRN</td>
<td>5739</td>
<td>146</td>
<td>1351</td>
<td>319</td>
<td>1774</td>
<td>573</td>
<td>9902</td>
</tr>
<tr>
<td>Unique sites with <em>dhps</em> IRN</td>
<td>54</td>
<td>8</td>
<td>20</td>
<td>11</td>
<td>21</td>
<td>8</td>
<td>120</td>
</tr>
<tr>
<td>Countries with <em>dhfr</em> IRN</td>
<td>11</td>
<td>3</td>
<td>11</td>
<td>4</td>
<td>2</td>
<td>3</td>
<td>34</td>
</tr>
</tbody>
</table>
Table 4.5 Summary of surveys available for dhps GE double mutant, GK single mutant and AK fully sensitive allelic haplotypes

<table>
<thead>
<tr>
<th>Regions</th>
<th>Southeast</th>
<th>Northeast</th>
<th>West</th>
<th>Southwest</th>
<th>Central</th>
<th>Islands</th>
<th>All Regions</th>
</tr>
</thead>
<tbody>
<tr>
<td>Surveys of dhps</td>
<td>80</td>
<td>11</td>
<td>24</td>
<td>16</td>
<td>13</td>
<td>8</td>
<td>152</td>
</tr>
<tr>
<td>GE/GK/AK</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Unique sites</td>
<td>52</td>
<td>10</td>
<td>21</td>
<td>14</td>
<td>8</td>
<td>7</td>
<td>112</td>
</tr>
<tr>
<td>Countries</td>
<td>10</td>
<td>3</td>
<td>14</td>
<td>5</td>
<td>2</td>
<td>3</td>
<td>37</td>
</tr>
<tr>
<td>Samples tested</td>
<td>12102</td>
<td>884</td>
<td>1980</td>
<td>1386</td>
<td>1554</td>
<td>1201</td>
<td>19107</td>
</tr>
<tr>
<td>Samples positive for dhps GE</td>
<td>4829</td>
<td>464</td>
<td>13</td>
<td>30</td>
<td>4</td>
<td>24</td>
<td>5364</td>
</tr>
<tr>
<td>Samples positive for dhps GK</td>
<td>198</td>
<td>30</td>
<td>1224</td>
<td>1037</td>
<td>845</td>
<td>504</td>
<td>3838</td>
</tr>
<tr>
<td>Samples positive for dhps AK</td>
<td>7075</td>
<td>390</td>
<td>743</td>
<td>319</td>
<td>705</td>
<td>673</td>
<td>9905</td>
</tr>
</tbody>
</table>

Figure 4.1 shows that high frequencies of dhfr NCS are found in East and West Africa. However, closer inspection of the longitudinal studies conducted at specific sites within this data set shows a gradual decline in frequencies over time. For example in Kilifi, Kenya dhfr NCS was 100% (n=5) in 1987 and by 2006 it was 2.6% (n=26) (Certain et al., 2008). In Kilombero and Ulanga in Tanzania dhfr NCS frequencies were measured at 51.3% (n=376) in 2000, it was 47.5% in 2001 (n = 238) and 27.8% (n=489) in 2002 with similar declining trends observed in Rufiji (Malisa et al., 2010). In Dielmo, Senegal dhfr NCS was 100% in 1990 (n=10) and by 1999 it was 57.1% (n=14) (Noranate et al., 2007). NCS was found to be absent in 43 sites and this will be discussed further in the next section.
In the Southeast region, apart from the survey done in the late 1980s in Kilifi, Kenya all frequencies were 51% or less since 1988 (Figure 4.2). In Shabunda, South Kivu Province which is the mid-eastern part of the DRC, dhfr NCS was 7.6% (n=158) in 2004 (Swarthout et al., 2006) whilst in Rutshuru it was 50% in 2002 (n=2) (Wilson et al., 2005). In Matola, 15km north of Maputo, dhfr NCS was 20% (n=89) in 2000 (Alifangis et al., 2003) and in Maputo itself it was 3% (n=72) in 2004 (Fernandes, Cravo and Do Rosario, 2007).

In the Northeast region, all eleven surveys were done in Ethiopia, North and South Sudan between 1996 and 2004 (Table 4.4, Figure 4.2). In Lankien, Upper Nile District in South Sudan dhfr NCS was 81.8% (n=44) in 2001 (van Den Broek et al., 2003) and 60% (n=78) in 2002 in Akuem, Yargot Payam district Bahr El Gazal (Anderson et al., 2003). All seven sites in North Sudan was less than 20% when surveyed either in 1996 or in 2003 (Al-Saai et al., 2009; A-Elbasit et al., 2007; Khalil et al., 2005; Khalil et al., 2003; Khalil et al., 2005, Hamour et al., 2005, Yusuf, Omar and Ngure, 2010). In Dilla, Southern Ethiopia (n=69)
(Schunk et al., 2006) and Jimma, Southwest Ethiopia (n=124) (Gebru-Woldearagai et al., 2005) NCS was absent in 2004.

In the Western region frequencies below 50% were reported since 1998 in 19 sites. There were a few exceptions where NCS was reported at frequencies above 50% recently (Figure 4.2). These were Yopounon and Adjame districts in the Ivory Coast where 73% (n=144) NCS was found in 2006 (Djaman et al., 2010) whilst Roper (pers. comm) found 70% (n=20) NCS in Abuja, Nigeria in 2005. In the Southwestern region seven out of 13 surveys had a NCS frequency below 50% between 1995 and 2006 (Kun et al., 1999; Mawili-Mboumba et al., 2001; Aubuoy et al., 2003; Ndounga et al., 2007; Menegon et al., 2009; Mita et al., 2009, Roper pers. comm). This allelic haplotype was found to be absent from six sites (Figure 4.2), all from Angola and surveyed in 2007 (Gama et al., 2011; Fortes et al., 2011).

In the Central region NCS was below 50% in 19 sites between 1994 and 2004 (Figure 4.2) (Basco et al., 1998; Basco et al., 2000, Basco et al., 2002; Menard et al., 2006; Tahar and Basco, 2006; McCollum et al., 2008; Mbacham et al., 2010). It was found to be above 50% only in Cameroon in seven sites surveyed between 2000 and 2003 (Tahar and Basco, 2006; Mbacham et al., 2009) (Figures 4.2 and 4.3). Also dhfr NCS was found to be absent in Douala and Manjo, Cameroon in 1999 (Basco et al., 2002) and 2002 (Tahar and Basco, 2006). Generally there was no overall change over time in the Central region (Figure 4.3) which was influenced by these large variations in frequency.

There were 10 surveys from the Island region between 2002 and 2008 (Table 4.4, Figure 4.2) with frequencies in Madagascar ranging from 1.0% (n=114) surveyed at various sites in 2006 (Menard et al., 2008) to 69% (n=292) surveyed at 12 sites also in 2006 (Andriantsoanirina et al., 2009). In Sao Tome/Principe island dhfr NCS was 4.8% (n=42) in 2002 (Tahar et al., 2007) but it was absent in Rua dos Trabalhadores in 2004 (n=6) (Salgueiro et al., 2010).
Figure 4.2 Surveys of \( dhfr \) NCS allelic haplotype frequency (with 95% confidence intervals) conducted during 1988-2008 and displayed according to their geographic region.
Multivariable logistic regression analyses for \textit{dhfr} NCS

The logistic model predicted that for every year, the likelihood of a person having a \textit{dhfr} NCS haplotype was 0.9 times lower providing evidence to suggest that there was a decline in the NCS haplotype frequencies over time (CI 0.9-0.9, p<0.001) (Table 4.6, Figure 4.3). The overall frequency of \textit{dhfr} NCS was 22.2\% (Table 4.6). The likelihood of \textit{dhfr} NCS occurring in the Southwest was lower than it was for the Southeast region over time, with strong evidence for this (p=0.019) (Table 4.6). There was strong evidence for higher likelihoods of \textit{dhfr} NCS over time in the other four regions relative to the Southeast region (p<0.001). The model with regions only as a predictor of \textit{dhfr} NCS haplotype frequency produced similar likelihood estimates. The likelihood ratio test to compare the models with and without the effect of time provided strong evidence to suggest that the regional frequency of \textit{dhfr} NCS was affected by time (p<0.001).

Table 4.6 Model for \textit{dhfr} NCS

<table>
<thead>
<tr>
<th>Covariate</th>
<th>Proportion (%) of samples positive [95% confidence interval]</th>
<th>Odds ratio [95% confidence interval]</th>
<th>P value</th>
<th>Odds ratio [95% confidence interval]</th>
<th>P value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Time (per study year)</td>
<td>22.2[21.6-22.8]</td>
<td>0.9 [0.9-0.9]</td>
<td>&lt;0.001</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Region</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Southeast</td>
<td>15.0[14.2-15.8]</td>
<td>1</td>
<td></td>
<td>1</td>
<td></td>
</tr>
<tr>
<td>Northeast</td>
<td>17.5[15.2-20.0]</td>
<td>1.3 [1.1-1.5]</td>
<td>&lt;0.001</td>
<td>1.2 [1.0-1.4]</td>
<td>0.036</td>
</tr>
<tr>
<td>West</td>
<td>33.8[32.2-35.4]</td>
<td>2.6 [2.4-2.9]</td>
<td>&lt;0.001</td>
<td>2.9 [2.6-3.2]</td>
<td>&lt;0.001</td>
</tr>
<tr>
<td>Southwest</td>
<td>10.3[8.3-12.6]</td>
<td>0.8 [0.6-1.0]</td>
<td>0.019</td>
<td>0.7 [0.5-0.8]</td>
<td>&lt;0.001</td>
</tr>
<tr>
<td>Central</td>
<td>24.7[23.1-26.4]</td>
<td>1.9 [1.7-2.1]</td>
<td>&lt;0.001</td>
<td>1.8 [1.7-2.1]</td>
<td>&lt;0.001</td>
</tr>
<tr>
<td>Island</td>
<td>32.6[30.7-34.5]</td>
<td>5.6 [4.9-6.4]</td>
<td>&lt;0.001</td>
<td>2.7 [2.5-3.0]</td>
<td>&lt;0.001</td>
</tr>
</tbody>
</table>
Figure 4.3 Estimated changes in *dhfr* NCS allelic haplotype frequency over time
The results for the *dhfr* IRN triple mutant will be discussed next. The *dhfr* IRN haplotype was found in 34 countries (Table 4.4, Figure 4.4). High frequencies, in excess of 80% were observed in 51 sites across sub-Saharan Africa from 1995 onwards (Figure 4.4).

From 106 surveys in the Southeast region, 67 presented with IRN frequencies in excess of 50% (Figure 4.5). Among the remaining 39 sites the frequencies ranged from 2.8% (n=72) in Rukungiri, Tanzania in 2005 (Lynch et al., 2008) to 46.7% in both Kariba, Zimbabwe (n=45) in 2003 (Mlambo et al., 2007) and in Rufigi, Tanzania (n=420) in 2001 (Malisa et al., 2010). There were fewer surveys (n=8) in the Northeast region compared with the Southeast region, and they spanned from 1996 to 2004 with the highest frequency of 87% (n=64) occurring in Dilla, Southern Ethiopia in 2004 (Schunk et al., 2006). The other seven sites were below 54% (Figure 4.5).

In the Western region, *dhfr* IRN frequencies in excess of 50% occurred in 13 sites between 2000 and 2006 (Figure 4.5). The highest IRN frequency was 85.1% (n=148) reported by Bonnet et al., (2007) in 2004 in a survey conducted in Laine, Liberia which borders with the Ivory Coast. Among the remaining 19 sites, IRN frequencies were between 12% (n=453) in 2002 in Ziniare and Bousse, situated north of Ouagadougou in Burkina Faso (Diallo et al., 2007) to 40% (n=15) in 2003 in Thies, Senegal (Ndiaye et al., 2005).

In the Southwest region, all eight studies from Angola, Gabon and Namibia recorded IRN frequencies of less than 48% between 1995 and 2007 (Figure 4.5) (Kun et al., 1999; Mawili-Mboumba et al., 2001; Fortes et al., 2011; Menegon et al., 2009; Roper, pers comm).

Similarly all nine surveys in the Central region reported less than 46% IRN from 1999 to 2004 (Figure 4.5). In 1999 it was 43.1% (n=51) in Bertoua, Cameroon (Basco et al., 2002) whilst in Nkambe, situated in northwest Cameroon it was 46% (n=103) in 2003 (Mbacham et al., 2009).

Finally, among the island populations a wide frequency range of 1.1% to 45% (n=269) was observed following surveys that were done at six sites in 2006 in Comoros/Mayotte (Andriantsoanirina et al., 2009). In Madagascar IRN frequencies ranged from 27.7% (n=292) in 2006 to 34.1 (n=361) in 2007 among 12 sites (Andriantsoanirina et al., 2009).
Figure 4.4 Frequency of \textit{dhfr} IRN haplotype

Namibia and Nigeria data obtained from C. Roper pers comm
Figure 4.5 Surveys of dhfr IRN allelic haplotype frequency (with 95% confidence intervals) conducted during 1988-2008 and displayed according to their geographic region.
Multivariable logistic regression analyses for \textit{dhfr} IRN

The logistic model predicted an overall increase in \textit{dhfr} IRN over time. For every year, the likelihood of a person having a \textit{dhfr} IRN haplotype was 1.1 times higher and there was strong evidence for this (CI 1.1-1.1, p<0.001) (Table 4.7, Figure 4.6). The overall frequency of \textit{dhfr} IRN was 51.4\% (Table 4.7). Generally there was a lesser likelihood of \textit{dhfr} IRN occurring in the Northeast, West, Southwest and Island regions compared to the Southeast region, with strong evidence for this reflected by p<0.001 in both models with and without the temporal effect (Table 4.7). The likelihood ratio test to compare the models with and without temporal variation provided strong evidence to suggest that the \textit{dhfr} IRN haplotype frequency varied over time (p< 0.001). The \textit{dhfr} IRN were analysed further in the next section.

Table 4.7 Model for \textit{dhfr} IRN

<table>
<thead>
<tr>
<th>Covariate</th>
<th>Proportion (%) of samples positive [95% confidence interval]</th>
<th>Odds ratio [95% confidence interval]</th>
<th>P value</th>
<th>Odds ratio [95% confidence interval]</th>
<th>P value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Time (per study year)</td>
<td>51.4[50.7-52.1]</td>
<td>1.1[1.1-1.1]</td>
<td>&lt;0.001</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Region</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Southeast</td>
<td>59.3[58.3-60.3]</td>
<td>1</td>
<td>1</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Northeast</td>
<td>17.1[14.6-19.8]</td>
<td>0.1[0.1-0.2]</td>
<td>&lt;0.001</td>
<td>0.1[0.1-0.2]</td>
<td>&lt;0.001</td>
</tr>
<tr>
<td>West</td>
<td>39.7[38.0-.41.4]</td>
<td>0.5[0.5-0.5]</td>
<td>&lt;0.001</td>
<td>0.5[0.4-0.5]</td>
<td>&lt;0.001</td>
</tr>
<tr>
<td>Southwest</td>
<td>42.8[39.2-46.5]</td>
<td>0.4[0.4-0.5]</td>
<td>&lt;0.001</td>
<td>0.5[0.4-0.6]</td>
<td>&lt;0.001</td>
</tr>
<tr>
<td>Central</td>
<td>62.7[60.9-64.5]</td>
<td>1.2[1.1-1.3]</td>
<td>&lt;0.001</td>
<td>1.2[1.1-1.3]</td>
<td>&lt;0.001</td>
</tr>
<tr>
<td>Island</td>
<td>32.6[30.4-34.9]</td>
<td>0.2[0.2-0.2]</td>
<td>&lt;0.001</td>
<td>0.3[0.3-0.4]</td>
<td>&lt;0.001</td>
</tr>
</tbody>
</table>
Figure 4.6 Predicted probabilities of *dhfr* IRN over time
4.4.2 Further Analyses of *dhfr* IRN triple mutant

Given the observed effects in the previous model, further multivariable analysis of the *dhfr* IRN triple mutant haplotype was conducted, factoring in clustering by country and excluding the Island populations. The results of the multivariable regression model showing the effect of time (study year, 1988-2008) and region on frequency are listed in Table 4.8. The overall frequency of *dhfr* IRN was 53.3% (n= 19269; Tables 4.4 and 4.8). The model predicted that for every year, the odds of a person having the *dhfr* IRN allelic haplotype was 1.1 times higher (CI 1.0-1.3, p=0.005) suggesting a general increasing trend in its frequency over time, although there was now weaker evidence (p=0.005) for this compared to the model shown in Table 4.7 (Table 4.6, Figure 4.8).

<table>
<thead>
<tr>
<th>Covariate</th>
<th>Proportion of samples positive [95% confidence interval]</th>
<th>Odds ratio [95% confidence interval]</th>
<th>P value</th>
</tr>
</thead>
<tbody>
<tr>
<td>study year (1988-2008) per year</td>
<td>53.3[52.5-54.0]</td>
<td>1.1[1.0-1.2]</td>
<td>0.005</td>
</tr>
<tr>
<td>Region</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Southeast</td>
<td>59.3[58.3-60.3]</td>
<td>1</td>
<td></td>
</tr>
<tr>
<td>Northeast</td>
<td>17.1[14.6-19.8]</td>
<td>0.1[0.0-1.1]</td>
<td>0.048</td>
</tr>
<tr>
<td>West</td>
<td>39.7[38.0-41.4]</td>
<td>0.5[0.2-1.1]</td>
<td>0.076</td>
</tr>
<tr>
<td>Southwest</td>
<td>42.8[39.2-46.5]</td>
<td>0.4[0.1-1.4]</td>
<td>0.166</td>
</tr>
<tr>
<td>Central</td>
<td>62.7[60.9-64.5]</td>
<td>1.2[0.6-2.2]</td>
<td>0.027</td>
</tr>
</tbody>
</table>
Compared with the Southeast region, only the Central region had a greater likelihood of \textit{dhfr} IRN occurring (OR 1.2, CI 0.6-2.2)(Table 4.7) whilst the other regions had a lesser likelihood of it occurring there, as all their odds ratios were less than one. These odds ratios were consistent with those reported in the previous logistic model. However, the observed relatively high p values in this model indicate that the strength of the evidence for these effects is weak (Table 4.8). The results of the Wald test also provided weak evidence of an overall regional effect in the frequency of \textit{dhfr} IRN (p=0.0569). Given these findings, no further modelling was done on \textit{dhfr} IRN in this chapter.

An important consideration is that the regional distribution of \textit{dhfr} NCS does not mirror that of \textit{dhfr} IRN. There were some sites where both low NCS and IRN were reported in the same survey. For example in Kilifi, Kenya NCS was 27.8% (n=36) whilst IRN was 30.6% (n=36) in 1997 (Nzila \textit{et al}, 2000a). Similarly in Khartoum, North Sudan during 1996 NCS was 15.9% (n=44) and IRN 11.4% (n=44) (Khallil \textit{et al}, 2005). In these studies the proportion of NCS and IRN were similar rather than reciprocal.

In Daraweesh and Kajara, which are approximately 15km from Gedaref in the eastern part of North Sudan, NCS was 2.6% (n=153) and IRN was 0.7% (n=153) in 2003 (A-elbasit \textit{et al}, 2007). In Thies, Senegal NCS was 33% (n=15) whilst IRN was 40% (n=15) in 2003 (Ndiaye \textit{et al}, 2005). On the other hand, generally the distributions of the \textit{dhps} haplotypes complement one other and these results are presented next.

The coverage of \textit{dhps} GE/GK/AK data extended to 37 countries, of which 23 countries had \textit{dhps} GE (Figure 4.7) which will be discussed first. High frequencies of ≥50 % \textit{dhps} GE were found predominately in East Africa where 26 sites had frequencies ranging between 80-100%. On the other hand, the \textit{dhps} GE double mutant allele is largely absent in parts of West Africa (Figure 4.7).
Evidence of the high *dhps* GE frequencies in East Africa is shown in Figures 4.8 and 4.9 where the regional frequencies are plotted against time. The Southeast region had the largest subset of data compared with the other regions and the majority of surveys in this region were carried out between 1997 and 2008 (Table 4.5, Figure 4.8).
Figure 4.8 Surveys of \textit{dhps GE} double mutant allelic haplotype frequency (with 95\% confidence intervals) conducted during 1988-2008 and displayed according to their geographic region.
Frequencies in excess of 80% were observed since 1997 and 100% dhps GE was found during 2005 in Malawi where both Mzimba in the North (n=46) and Dedza (n=48) had high fixed GE frequencies (Nkhoma et al., 2007). A study done in 2005 in Kabale, Uganda also reported dhps GE fixation at 100% (n=58) (Lynch et al., 2008). In the Northeast region, frequencies of 97%-100% were found in Ethiopia in 2004 at three sites namely Dilla, Humera and Jimma (Gebru-woldearegai et al., 2005; Schunk et al., 2006; Pearce et al., 2009) and the frequencies increased over time at these sites (Figure 4.8). In the Southwest, West, Central and Island regions, the frequencies never exceeded 20% (Figure 4.8).

The logistic model predicted an overall increase in dhps GE over time. The likelihood of a dhps GE haplotype occurring was 1.417 times higher per year and there was strong evidence for this (CI 1.393-1.441, p<0.001) (Table 4.9, Figure 4.9). The overall frequency of dhps GE was 28.1% (Table 4.9).

Generally there was a greater likelihood of dhps GE occurring in the Northeast compared to the Southeast region, with strong evidence for this (Table 4.9) in both models with and without the temporal effect (Table 4.7). Given the low frequency of dhps GE in the other five regions as described above, the model predicted small odds ratios for these regions.

The likelihood ratio test to compare the models with and without temporal variation provided strong evidence to suggest that the dhps GE haplotype frequency was affected by time (p< 0.001). Further multivariable analyses of dhps GE is presented in section 4.4.3.
<table>
<thead>
<tr>
<th>Covariate</th>
<th>Proportion (% of samples positive [95% confidence interval])</th>
<th>Odds ratio [95% confidence interval]</th>
<th>P value</th>
<th>Odds ratio [95% confidence interval]</th>
<th>P value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Time (per study year)</td>
<td>28.1 [27.4-28.7]</td>
<td>1.417 [1.393-1.441]</td>
<td>0.001</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Region</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Southeast</td>
<td>39.9 [39.0-40.8]</td>
<td>1</td>
<td></td>
<td>1</td>
<td></td>
</tr>
<tr>
<td>Northeast</td>
<td>52.5 [49.1-55.8]</td>
<td>2.276 [1.958-2.646]</td>
<td>0.001</td>
<td>1.664 [1.451-1.908]</td>
<td>&lt;0.001</td>
</tr>
<tr>
<td>West</td>
<td>0.7 [0.4-1.1]</td>
<td>0.009 [0.005-0.016]</td>
<td>0.016</td>
<td>0.010 [0.006-0.017]</td>
<td>&lt;0.001</td>
</tr>
<tr>
<td>Southwest</td>
<td>2.2 [1.5-3.1]</td>
<td>0.016 [0.011-0.023]</td>
<td>0.033</td>
<td>0.033 [0.023-0.048]</td>
<td>&lt;0.001</td>
</tr>
<tr>
<td>Central</td>
<td>0.3 [0.1-0.7]</td>
<td>0.005 [0.002-0.012]</td>
<td>0.004</td>
<td>0.004 [0.001-0.010]</td>
<td>&lt;0.001</td>
</tr>
<tr>
<td>Island</td>
<td>2.0 [1.3-3.0]</td>
<td>0.010 [0.007-0.015]</td>
<td></td>
<td>0.031 [0.021-0.046]</td>
<td>&lt;0.001</td>
</tr>
</tbody>
</table>
The distribution of \textit{dhps} GK allelic haplotype differs from that observed with \textit{dhps} GE, in relation to the differential frequencies observed in East and West Africa. High frequencies \(\geq 50\%\) \textit{dhps} GK were found predominately in West Africa and 18 sites had frequencies ranging between 80-100\% (Figure 4.10). The \textit{dhps} GK allelic haplotype is absent in more sites in East Africa compared with West Africa where it was absent only in Zinderou and Banizoumba, Niger between 2003-2006 (n=12) (Ibrahim \textit{et al.}, 2009).
In the Southeast region 40 surveys did not observe the dhps GK single mutant allelic haplotype and the highest frequency was 40.3% in Kyela, Tanzania in 2000 (Mugittu et al., 2004) (Figure 4.11). In the Northeast region, the frequency ranged from 1.1% (n=87) during 2004 in Humera, Ethiopia (Pearce et al., 2009) to 14.2% (n=148) in 2003 in Daraweesh and Kajara, North Sudan (A-Elbasit et al., 2007). The remaining seven surveys found no dhps GK.

In the Central region a site in Mutengene, Cameroon surveyed during 2004-2006 had a dhps GK frequency of 86.7% (n=196) (Mbacham et al., 2010). The remaining 12 sites across the Central region had dhps GK below 65.1% (Figure 4.11). Among the island populations, four sites in Sao Tome/Principe had dhps GK above 80% since 2002 (83%, n=47) and the three others were 84% (n=90) in Neves, 85.7%(n=21) in Ribeira Afonso and 87.5%(n=8) in Rua dos Trabalhadores, all were done during 2004 (Salgueiro et al., 2010).
In Bamako, Mali dhps GK was 100% (n=32) in 1995 (Plowe et al., 1997) and there were six studies in the Western region were it occurred in excess of 80%. These were three sites in Ghana surveyed in 2001 (93.4%, n=76), 2003 (88%, n=75) in the Bodomase, Ashanti region (Marks et al., 2005) and in 2003 in Navrongo (85%, n=99) and 2005 in Hoehoe (86.9%, n=99) (Pearce et al., 2009). In the Southwest region, six sites observed dhps GK in excess of 85%. In the Tenrikyo district of Makelekele, Brazzaville, Republic of Congo dhps GK was 80% (n=80) in 2003 (Ndounga et al., 2007) and the other sites surveyed were in Angola during in 2004 and 2007 (Menegon et al., 2009; Fortes et al., 2011).
Figure 4.11 Surveys of \textit{dhps} GK single mutant allelic haplotype frequency (with 95\% confidence intervals) conducted during 1988-2008 and displayed according to their geographic region.
The logistic model predicted an overall increase in *dhps* GK over time. The likelihood of a person having a *dhps* GK haplotype was 1.071 times per year with strong evidence for this (CI 1.050-1.1093, p<0.001) (Table 4.10, Figure 4.12). The overall frequency of *dhps* GK was 20.1% (Table 4.9).

Generally there was a higher likelihood of *dhps* GK in all five regions compared to the Southeast region, with strong evidence for this (Table 4.10) in both models with and without the effect of time as a covariate (Table 4.7). The likelihood ratio test to compare the models with and without temporal variation provided strong evidence to suggest that the *dhps* GK haplotype frequency was affected by time (p< 0.001). Further multivariable analyses of *dhps* GK is presented in section 4.4.

<table>
<thead>
<tr>
<th>Covariate</th>
<th>Proportion (%) of samples positive [95% confidence interval]</th>
<th>Odds ratio [95% confidence interval]</th>
<th>P value</th>
<th>Region</th>
<th>Proportion (%) of samples positive [95% confidence interval]</th>
<th>Odds ratio [95% confidence interval]</th>
<th>P value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Time (per study year)</td>
<td>20.1[19.5-20.7]</td>
<td>1.071 [1.050-1.093]</td>
<td>&lt;0.001</td>
<td>Southeast</td>
<td>1.6[1.4-1.9]</td>
<td>1</td>
<td></td>
</tr>
<tr>
<td>Northeast</td>
<td>3.4[2.3-4.8]</td>
<td>2.232 [1.510-3.299]</td>
<td>&lt;0.001</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>West</td>
<td>61.8[59.6-64.0]</td>
<td>100.420 [84.895-118.785]</td>
<td>&lt;0.001</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Southwest</td>
<td>74.8[82.4-77.1]</td>
<td>166.752 [138.393-200.923]</td>
<td>&lt;0.001</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Central</td>
<td>54.4[51.9-56.9]</td>
<td>78.140 [65.580-93.106]</td>
<td>&lt;0.001</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Island</td>
<td>42.0[39.2-44.8]</td>
<td>35.476 [29.331-42.908]</td>
<td>&lt;0.001</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Table 4.10 Model for *dhps* GK
The third allelic haplotype \textit{dhps} AK is fully sensitive and did not have any clear pattern of distribution when mapped across Africa (Figure 4.13). Unlike \textit{dhfr} IRN and \textit{dhfr} NCS which co-exist with double and single mutant allelic haplotypes, the three \textit{dhps} allelic haplotypes GE/GK/AK add up to the total number tested for these alleles. Frequencies of $\geq 50\%$ were observed in 69 sites across all six regions and the majority (61\%) of these occurred in the Southeast region (Figure 4.14).
Figure 4.13 Frequency of \textit{dhps} AK fully sensitive haplotype
Figure 4.14 Surveys of dhps AK allelic haplotype frequency (with 95% confidence intervals) conducted during 1988-2008 and displayed according to their geographic region.
The logistic model predicted a lower likelihood of \textit{dhps} AK over time. The likelihood of a person having a \textit{dhps} AK haplotype was 0.8 times lower per year with strong evidence for this (CI 0.8-0.8, p<0.001) (Table 4.11, Figure 4.15). The overall frequency of \textit{dhps} AK was 51.8% (Table 4.11).

Generally there was a lower likelihood of \textit{dhps} AK occurring in four regions except the Island region, compared to the Southeast region and there was strong evidence for this (Table 4.11). However this effect was only observed in the model including time. In the model excluding the effect of time, \textit{dhps} AK had a lower likelihood in all five regions compared with the Southeast region (Table 4.11). However the strength of the evidence for this was weak in the Island region (p=0.104). The likelihood ratio test to compare the models with and without temporal variation provided strong evidence to suggest that the \textit{dhps} AK haplotype frequency was affected by time (p< 0.001). Further multivariable analyses of \textit{dhps} AK is presented in section 4.4.5.

Table 4.11 Model for \textit{dhps} AK

<table>
<thead>
<tr>
<th>Covariate</th>
<th>Proportion (%) of samples positive [95% confidence interval]</th>
<th>Odds ratio [95% confidence interval]</th>
<th>P value</th>
<th>Odds ratio [95% confidence interval]</th>
<th>P value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Time (per study year)</td>
<td>51.8[53.2-58.9]</td>
<td>0.8[0.8-0.8]</td>
<td>&lt;0.001</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Region</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Southeast</td>
<td>58.5[57.6-59.3]</td>
<td>1</td>
<td>1</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Northeast</td>
<td>44.1[40.8-47.5]</td>
<td>0.4[0.4-0.5]</td>
<td>&lt;0.001</td>
<td>0.6[0.5-0.6]</td>
<td>&lt;0.001</td>
</tr>
<tr>
<td>West</td>
<td>37.5[35.4-39.7]</td>
<td>0.4[0.3-0.4]</td>
<td>&lt;0.001</td>
<td>0.4[0.4-0.5]</td>
<td>&lt;0.001</td>
</tr>
<tr>
<td>Southwest</td>
<td>23.0[20.8-25.3]</td>
<td>0.3[0.2-0.3]</td>
<td>&lt;0.001</td>
<td>0.2[0.2-0.2]</td>
<td>&lt;0.001</td>
</tr>
<tr>
<td>Central</td>
<td>45.4[42.9-47.9]</td>
<td>0.4[0.4-0.5]</td>
<td>&lt;0.001</td>
<td>0.6[0.5-0.7]</td>
<td>&lt;0.001</td>
</tr>
<tr>
<td>Island</td>
<td>56.0[53.2-58.9]</td>
<td>2.0[1.8-2.3]</td>
<td>&lt;0.001</td>
<td>0.9[0.8-1.0]</td>
<td>0.104</td>
</tr>
</tbody>
</table>
Figure 4.15 Predicted probabilities of dhps AK over time

The relative proportions of samples tested positive for dhps GE/GK/AK allelic haplotypes are illustrated in Figure 4.16. The map of pie charts reflects the relative proportions of samples which tested positive for dhps GE/GK/AK allelic haplotypes together (Figure 4.16) and shows a clear pattern of high GE frequencies in East Africa and high GK frequencies in West Africa, whilst there is no clear pattern of AK due to its scattered distribution.
Figure 4.16 Samples tested positive for \textit{dhps} GE, GK and AK haplotypes

The comparison of the logistic models constructed above provided evidence to show that there were regional trends in allelic haplotype frequencies over time. Following these exploratory analyses, further multivariable analyses were performed to further explore the relationship between \textit{dhps} GE/GK/AK and the covariates, region and time whilst accounting for clustering by country and excluding the Island populations. The results for each allelic haplotype are presented next in a sequential manner.
4.4.3 Further Analyses of dhps GE double mutant

The results of the multivariable regression model showing the effect of time and region on dhps GE frequency are listed in Table 4.12. The model predicted that for every year, the likelihood of a dhps GE double mutant haplotype allele occurring was 1.422 times higher (95% CI 1.151-1.758), p<0.001) (Table 4.12). This suggests that there is a higher likelihood of dhps GE double mutant frequency over time.

The Wald test assessed if there were any differences in dhps GE frequencies among the regions. Overall the Wald test showed that there was strong evidence of between-region differences in the frequency of dhps GE (p<0.001).

In the Central (OR 0.005, CI 0.002-0.011), Southwest (OR 0.016, CI 0.004-0.063) and Western (OR 0.009, CI 0.002-0.044) regions the likelihood of the dhps GE double mutant allele being present was lower than it was for the Southeast region, with strong evidence for this as p<0.001 (Table 4.12). In the Northeast region the likelihood of dhps GE being present was higher than it was for the Southeast region but this may be due to the effect of sampling variation as seen from the large width of the 95% confidence interval and weak strength of evidence indicated by the high p value (OR 2.282, CI 0.618-8.425, p=0.215).

4.4.4 Further Analyses of dhps GK single mutant

The model predicted that for every year, the likelihood of a person having dhps GK was 1.102 times higher (CI 1.007-1.207, p=0.035) (Table 4.12). As with dhps GE described above, this suggests a higher likelihood of the dhps GK allele frequency per year. However, the strength of the evidence for this effect was weak with p=0.035. The Wald test provided strong evidence to suggest that the frequency of dhps GK was different among the regions (p<0.001).

For all regions the likelihood of the dhps GK allele occurring was greater than it was for the Southeast region. Large odds ratios were observed in the Central, Southwest and Western regions, suggesting very large differences between these regions and the Southeast reference region. The strength of the evidence for this observation was good among the three regions where p< 0.001 (Table 4.12) with the exception of the Northeast region.
4.4.5 Further Analyses of \textit{dhps} AK fully sensitive allele

The model predicted that the likelihood of the \textit{dhps} AK sensitive haplotype allele was 0.758 times lower per year (95\% CI 0.662-0.862) and there was strong evidence for this downward trend in frequency (p<0.001).

The Wald test indicated there was evidence for regional variation in the frequencies of \textit{dhps} AK (p=0.0242). The likelihood of the \textit{dhps} AK allele occurring in the Central, Northeast, Southwest and Western regions was less than the likelihood of it occurring in the Southeast region although the strength of evidence for this was good for the Southwest region only where p<0.001 (Table 4.12).

Overall, the multivariable analyses for \textit{dhps} GE, GK and AK showed evidence of trends over time and the likelihood of their occurrence differed among the regions which justified inclusion of the regional categorisation in the models. Given these observations, an interaction term between time (study year) and location (region) was factored into the logistic models to assess whether or not the relationship between the \textit{dhps} GE/GK/AK frequencies, differed among the five regions with time. This test would indicate if any particular region's influence over these frequencies was significant over time or not.
Table 4.12 Multivariable analyses of *dhps* GE, *dhps* GK and *dhps* AK

<table>
<thead>
<tr>
<th>Covariate</th>
<th>Proportion (%) of samples positive [95% confidence interval]</th>
<th>Odds ratio [95% confidence interval]</th>
<th>P value</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Model for <em>dhps</em> GE</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>study year (1988-2008) per year</td>
<td>29.8[29.2-30.5]</td>
<td>1.422 [1.151-1.758]</td>
<td>&lt;0.001</td>
</tr>
<tr>
<td>Region</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Southeast</td>
<td>39.9[39.0-40.8]</td>
<td>1</td>
<td></td>
</tr>
<tr>
<td>Northeast</td>
<td>52.5[49.1-55.8]</td>
<td>2.282[0.618-8.425]</td>
<td>0.215</td>
</tr>
<tr>
<td>West</td>
<td>0.7[0.4-1.1]</td>
<td>0.009[0.002-0.044]</td>
<td>&lt;0.001</td>
</tr>
<tr>
<td>Southwest</td>
<td>2.2[1.5-3.1]</td>
<td>0.016[0.004-0.063]</td>
<td>&lt;0.001</td>
</tr>
<tr>
<td>Central</td>
<td>0.3[0.1-0.7]</td>
<td>0.005[0.002-0.011]</td>
<td>&lt;0.001</td>
</tr>
<tr>
<td><strong>Model for <em>dhps</em> GK</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>study year (1988-2008) per year</td>
<td>18.7[18.1-19.2]</td>
<td>1.102[1.007-1.207]</td>
<td>0.035</td>
</tr>
<tr>
<td>Region</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Southeast</td>
<td>1.6[1.4-1.9]</td>
<td>1</td>
<td></td>
</tr>
<tr>
<td>Northeast</td>
<td>3.4[2.3-1.9]</td>
<td>2.284[0.584-8.934]</td>
<td>0.235</td>
</tr>
<tr>
<td>West</td>
<td>61.8[59.6-64.0]</td>
<td>102.734[45.112-233.955]</td>
<td>&lt;0.001</td>
</tr>
<tr>
<td>Southwest</td>
<td>74.8[72.4-77.1]</td>
<td>164.125[73.854-364.734]</td>
<td>&lt;0.001</td>
</tr>
<tr>
<td>Central</td>
<td>54.4[51.9-56.9]</td>
<td>81.720[42.708-156.70]</td>
<td>&lt;0.001</td>
</tr>
<tr>
<td><strong>Model for <em>dhps</em> AK</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>study year (1988-2008) per year</td>
<td>51.6[50.8-52.3]</td>
<td>0.758[0.662-0.862]</td>
<td>&lt;0.001</td>
</tr>
<tr>
<td>Region</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Southeast</td>
<td>58.5[57.6-59.3]</td>
<td>1</td>
<td></td>
</tr>
<tr>
<td>Northeast</td>
<td>44.1[40.8-59.3]</td>
<td>0.428[0.120-1.533]</td>
<td>0.193</td>
</tr>
<tr>
<td>West</td>
<td>37.5[35.4-39.7]</td>
<td>0.373[0.148-0.935]</td>
<td>0.036</td>
</tr>
<tr>
<td>Southwest</td>
<td>23.0[20.8-25.3]</td>
<td>0.259[0.114-0.587]</td>
<td>0.001</td>
</tr>
<tr>
<td>Central</td>
<td>45.4[42.9-47.9]</td>
<td>0.425[0.206-0.879]</td>
<td>0.021</td>
</tr>
</tbody>
</table>
4.4.6 Interaction between region and time

Table 4.13 shows the extent to which the odds ratios for time (study year) and each regional group, differs from those in the reference Southeast region. The results for each *dhps* allelic haplotype are discussed below.

**Model for *dhps* GE double mutant**

The odds ratios differed among regional strata for *dhps* GE, providing evidence to suggest that there was an interaction between time (study year) and region and the likelihood of *dhps* GE occurring. Hence the increasing temporal trend in *dhps* GE frequency differed depending on the regional categorisation and there was strong evidence for this (p <0.001, based on chi² and 4 degrees of freedom) (Table 4.13)(Figure 4.17).

As mentioned in chapter 3 section 3.3.2, the lower of the two AIC values indicate the better fitted model. The AIC for the model without interaction was 0.894 whilst the model with interaction had a lower AIC of 0.876 providing some support for the model with interaction between region and time. The Southwest region did not show any association between the frequency of *dhps* GE and the interaction term (OR 1.0, CI 0.7-1.4), although there was weak evidence for this with p=0.952. The two models for *dhps* GE (with and without the interaction) showed increasing trends of *dhps* GE frequency in the Southeast and Northeast regions only.

**Model for *dhps* GK single mutant**

The odds ratios also differed among regional strata for *dhps* GK suggesting that the increasing temporal trend in frequency differed depending on the regional categorisation and there was strong evidence for this (p <0.001, based on chi² and 4 degrees of freedom) (Table 4.13)(Figure 4.18). The AIC for the model without interaction was 0.478 whilst the model with interaction had an AIC of 0.471 indicating that the latter interaction model had a better fit. Both models (with and without the interaction term) showed increasing trends of *dhps* GK frequency in the Central, Southwest and Western regions, although this evidence was strong only in the Southwest region where p<0.001 (Table 4.13).
Model for *dhps* AK fully sensitive allele

There was a decreasing likelihood of *dhps* AK among all five regions with strong evidence for this ($p < 0.030$, based on chi² and 4 degrees of freedom) (Table 4.13)(Figure 4.19). The likelihood of the *dhps* AK occurring in the Central, Southwest and Western regions was marginally better than it was for the Southeast region. The strength of the evidence for this was better in the Southwest region where $p < 0.001$ compared with the other regions (Table 4.13). The AIC for the model without interaction was 1.232 whilst the model with interaction had an AIC of 1.213 providing some evidence of a better fit for the latter model.

<table>
<thead>
<tr>
<th>Covariate</th>
<th>Difference in slope (per year) for each region</th>
<th>Multivariable analysis using allelic haplotype frequency with interaction for study year x region</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Odds ratio [95% confidence interval]</td>
<td>P value</td>
</tr>
<tr>
<td><strong>dhps GE</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Region</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Southeast</td>
<td>1.4[1.1-1.7]</td>
<td>0.001</td>
</tr>
<tr>
<td>Central</td>
<td>2.0[1.6-2.6]</td>
<td>&lt;0.001</td>
</tr>
<tr>
<td>Northeast</td>
<td>26.2[14.0-49.2]</td>
<td>&lt;0.001</td>
</tr>
<tr>
<td>Southwest</td>
<td>1.4[1.1-1.8]</td>
<td>0.011</td>
</tr>
<tr>
<td>West</td>
<td>1.3[0.9-2.0]</td>
<td>0.180</td>
</tr>
<tr>
<td><strong>dhps GK</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Region</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Southeast</td>
<td>0.9[0.8-1.0]</td>
<td>0.020</td>
</tr>
<tr>
<td>Central</td>
<td>1.1[1.0-1.2]</td>
<td>0.064</td>
</tr>
<tr>
<td>Northeast</td>
<td>0.9[0.8-1.2]</td>
<td>0.560</td>
</tr>
<tr>
<td>Southwest</td>
<td>1.3[1.2-1.4]</td>
<td>&lt;0.001</td>
</tr>
<tr>
<td>West</td>
<td>1.2[0.9-1.4]</td>
<td>0.202</td>
</tr>
<tr>
<td><strong>dhps AK</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Region</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Southeast</td>
<td>0.7[0.6-0.9]</td>
<td>0.002</td>
</tr>
<tr>
<td>Central</td>
<td>0.9[0.8-1.0]</td>
<td>0.051</td>
</tr>
<tr>
<td>Northeast</td>
<td>0.2[0.0-2.0]</td>
<td>0.184</td>
</tr>
<tr>
<td>Southwest</td>
<td>0.8[0.7-0.8]</td>
<td>&lt;0.001</td>
</tr>
<tr>
<td>West</td>
<td>0.9[0.7-1.1]</td>
<td>0.188</td>
</tr>
</tbody>
</table>
Figure 4.17 Predicted probabilities of regional dhps GE allelic haplotype frequencies with interaction between region and time.
Figure 4.18 Predicted probabilities of regional \( dhps \) GK allelic haplotype frequencies with interaction between region and time.
Figure 4.19 Predicted probabilities of regional dhps AK allelic haplotype frequencies with interaction between region and time.
4.5 Discussion

The main finding of the ecological analyses of \textit{dhfr} IRN, \textit{dhfr} NCS and \textit{dhps} GE/GK/AK allelic haplotype frequencies presented in this chapter is that their distributions are different within sub-Saharan Africa, suggesting that SP efficacy also varies across the continent. These differences are described below first for the \textit{dhfr} IRN triple mutant and \textit{dhfr} NCS fully sensitive allelic haplotype and then for the three \textit{dhps} allelic haplotypes.

Limitations and recommendations emanating from the work presented in this chapter are discussed. Generally, the frequencies of the sensitive alleles \textit{dhfr} NCS and \textit{dhps} AK was lower per year, whilst the frequencies of the mutant alleles \textit{dhfr} IRN, \textit{dhps} GE and \textit{dhps} GK was higher per year.

The logistic models showed that there was a lower likelihood of \textit{dhfr} NCS occurring in the Southeast region compared to the Northeast, West, Central and Island regions. Conversely, there was a higher likelihood of \textit{dhfr} IRN triple mutant in the Southeast region compared to the Northeast, West, Southwest and Island regions. It was shown that the \textit{dhfr} NCS sensitive allelic haplotype is generally decreasing, whilst the \textit{dhfr} IRN triple mutant has dispersed and increased over time to fixation in many sites across Africa. Thus, it is no longer possible to trace the stepwise dispersal pattern of \textit{dhfr} IRN. As Roper \textit{et al.}, (2003) showed the triple mutant \textit{dhfr} IRN haplotype was imported into Africa from Southeast Asia. Early reports of resistance to pyrimethamine monotherapy after its introduction as both malaria prophylaxis and treatment in the 1950s (Peters, 1970) are probably attributable to local mutations which emerged soon after the introduction of pyrimethamine in malaria control (Peters, 1969). The emergence of the triple mutant \textit{dhfr} IRN haplotype coincided with the emergence of resistance much later when pyrimethamine was used in combination with sulphadoxine. Evidence of this comes from a study conducted by Certain \textit{et al.}, (2008) who found that the Asian \textit{dhfr} IRN was present in Kilifi, Kenya around 1988 and there was a marked increase in \textit{dhfr} IRN during 1987-2006 after SP was used in this area.

The distribution of the sensitive alleles \textit{dhfr} NCS generally does not mirror that of IRN largely because of the occurrence of the \textit{dhfr} 51I/108N or 59R/108N double mutants and the \textit{dhfr} 108N single mutant shown in Table 4.1.

The maps and multivariate regression models used data from 152 sites and 34 countries and showed that there are clear differences in the geographies of the \textit{dhps} GE double
mutant, GK single mutant and AK fully sensitive allelic haplotypes. A distinct pattern of dispersal for these \textit{dhps} allelic haplotypes can be characterised. For \textit{dhps} GE, there is an East Africa-West Africa divide with high frequencies occurring in the Eastern countries and low frequencies occurring in West Africa. Conversely, there are high \textit{dhps} GK frequencies primarily in West Africa and low \textit{dhps} GK frequencies in East Africa.

The results for \textit{dhps} GE and \textit{dhps} GK showing an East-West divide presented in this chapter are consistent with the work done by Pearce \textit{et al.}, (2009) who used a relatively smaller dataset of 50 sites among 27 countries and showed that the \textit{dhps} GE double mutant was prevalent in East Africa whilst the \textit{dhps} GK single mutant was prevalent in West Africa. Since the research in this chapter was carried out, new \textit{dhps} allelic haplotype frequency data from the DRC were published by Taylor \textit{et al.}, (2013) and they showed the same East-West divide within the country with foci of the \textit{dhps} GE double mutant occurring primarily in the eastern parts of the country and dominant foci of the \textit{dhps} GK single mutant occurring in the western parts of the country.

This study found a decreasing trend in \textit{dhps} AK across all regions with some evidence for regional differences in its occurrence as the likelihood of this allele occurring in Central and West Africa was marginally greater than it was for East Africa when the effect of time and region were considered together. This finding is consistent with the observation that these sensitive alleles are being displaced by resistant alleles under selection by high SP use in East Africa especially.

These differences in the geographies of the \textit{dhps} GE double mutant, GK single mutant and AK fully sensitive allelic haplotypes can be explained by the differential SP drug pressure in West Africa and East Africa. There has been large-scale use of SP as first line monotherapy in East Africa (WHO, 2001) whilst SP use has been low in West Africa (Frosch \textit{et al.}, 2011) due to the varying malaria transmission intensities in these regions. As Pearce \textit{et al.}, (2009) noted, in Africa once \textit{dhps} mutations arose independently at multiple times and became established, they circulated among five distinct mainland African regions, with strong selection in foci of high SP use.

The timing of the introduction and duration of SP use is a contributory factor to the observed differential \textit{dhfr} and \textit{dhps} frequencies over time. Possible explanations for these observed differential gradients are as follows: (1) some regions used less SP compared to
others as mentioned previously, (2) SP was introduced at a later time in one region compared to other regions for operational reasons such as scheduling of drug policy changes, (3) a different allele is being selected for in that region or (4) a combination of all these factors.

Limitations

As mentioned above, patterns of SP use contribute to the differential gradients of dhfr and dhps frequencies over time but these were not factored into the models in this study because trends were examined by region. Frosch et al. (2011) correlated chloroquine use at a country level, with chloroquine resistance in five individual countries where data were available. To do a similar analysis, clear data regarding the duration of SP use as monotherapy and combination therapy with other drugs are needed for each sub-Saharan African country. There are some challenges in identifying the specific year in which drug policies changed to any SP therapy per country, noting that there is often a time lag between the actual operational use of SP and drug policy change and the sometimes slow uptake of artemisin combinations at sentinel sites. This information is accessible for IPT strategies as the WHO (2011a) has summarised the year in which IPT was adopted in each country.

Recommendations

The maps presented in section 4.4 afford the opportunity to highlight gaps in the dhps allelic haplotype data. Generally there was better coverage of the data in the Southeast regions than the other regions. Further monitoring of dhps allelic haplotypes across Africa is crucial and up-scaled surveillance using longitudinal studies which are repeated at the same sentinel sites over different time periods, specifically in Central and West Africa are important in the context of IPT strategies for malaria control.

Whilst it is clear that SP efficacy varies across sub-Saharan Africa, the observed regional patterns in the frequency of allelic haplotypes shows that neighbouring malaria control programmes could benefit from joint regional initiatives such as the Lubombo Spatial Development Initiative (LSDI) (Sharp et al., 2007).
As mentioned in this chapter, frequency data are useful when comparing SP resistance across populations. However, there are no standard formats for collating and reporting frequency versus prevalence data. The results presented in this chapter on allelic haplotype frequencies show the value of isolating mixed infections from analyses of SP resistance markers. The WWARN initiative (www.wwarn.org) which was described previously will assist in standardising how these data types are collated and reported.

In addition, there is a need to assess the current distribution of the quintuple mutation, which represents fully resistant SP parasites using the more stringent measure of frequency rather than prevalence. Mobula et al., (2009) gathered published data on the prevalence of the quintuple mutant since 2000 in the following 17 countries: Burkina Faso, Cameroon, Republic of Congo, DRC, Ethiopia, Gabon, Ghana, Guinea, Guinea-Bissau, Malawi, Mozambique, Nigeria, Senegal, Sudan, Tanzania, Uganda and Zambia. These authors illustrated that there were foci of high quintuple mutant frequencies in East Africa and relatively low frequencies in West African foci.

This chapter showed the large regional differences in the frequencies among dhps allelic haplotypes. Further analyses linking SP in vivo treatment failures to allelic haplotype data will provide insight into whether or not molecular measures of SP resistance correlate with in vivo resistance. This is investigated in chapter 5.
Chapter 5 The association between *dhfr* and *dhps* haplotypes and *in vivo* measures of Sulphadoxine Pyrimethamine resistance at population level

5.1 Introduction

The mapping presented in previous chapters has shown that different mutations are common in East and West Africa, yet the implications of this for SP efficacy are unclear. The relationship between *dhfr* and *dhps* mutations respectively and treatment outcome has most commonly been examined at an individual patient level but to identify a significant regional effect across the continent, it may be more appropriate to explore this using summarised population level data rather than individual patient level data. This chapter explores the relationship between population level *in vivo* measures of SP efficacy and summarised *dhfr* and *dhps* molecular data which were matched using the study year and study location and assesses the effects of genotype frequencies occurring in East and West Africa.

The first significant obstacle to this type of approach is the standardisation of summarised *in vivo* data. Previous classifications of antimalarial treatment efficacy from *in vivo* studies have not been standardised which rendered it difficult to compare outcomes. This study uses secondary data collated from studies that assessed the *in vivo* response of SP and proposes a system of classification for these data to guide spatial and temporal analyses.

As described in chapter 2, *in vivo* studies assess the treatment outcome of people who received proper doses of SP over a defined follow-up period. Initially when SP was first introduced after treatment failures with chloroquine, there were more treatment successes and adequate clinical responses compared with inadequate clinical responses and treatment failures. With extensive use of the drug, efficacy eventually waned and treatment failures occurred against a backdrop of progressive development of resistance measured by clinical and parasitological indicators. Currently SP is no longer recommended as treatment for malaria and instead ACTs are recommended in sub-Saharan Africa. Therefore no further *in vivo* studies of SP treatment efficacy are being conducted. However as mentioned in chapter 1, SP is being used for IPT in 37 countries in sub-Saharan Africa (WHO, 2011a) and it is thus imperative that there is continued monitoring of point mutations and allelic haplotypes that are associated with resistance, to provide information on SP efficacy in these settings.
During the past 15-20 years, there have been several attempts to correlate the genotypes (molecular markers) with phenotypes (clinical outcomes) of SP resistance, with somewhat inconsistent results. Plowe et al., (1997) suggested that progressive intensification of SP resistance from RI, RII to RIII occur in vivo because of the progressive accumulation of \textit{dhfr} and \textit{dhps} point mutations. These authors added that the influence of factors such as immunity levels also have major influence on treatment outcome (Plowe et al., 1997). As shown in previous chapters the molecular determinants of SP resistance, in particular those mutations in \textit{dhps}, differ in East and West Africa. However, the effects of these differences have never been examined.

In East Africa, evidence of a relationship between specific point mutations and treatment failure was shown in several studies and these have been reviewed by Abdul-Ghani, Farag and Allam (2013). The first studies were reported from Tanzania where reduced efficacy of SP was associated with the occurrence of \textit{dhfr} and \textit{dhps} mutations (Curtis et al., 1998; Wang et al., 1997). However, in Tanzania, Jelinek et al., (1998) did not find any statistically significant relationship between clinical outcomes and \textit{dhfr} or \textit{dhps} point mutations in 1995 or 1996 among pre-treatment samples. In a study conducted in July 1997 and January 1999 in Kenya, Nzila et al., (2000a) showed that the \textit{dhfr} 51I/59R/108N triple mutant was associated with RI and RII resistance to SP when found with or without the \textit{dhps} point mutations. Kyabayinze et al., (2003) found that the \textit{dhfr} 59R and \textit{dhps} 540E were strongly associated with SP treatment failure in Kampala, Uganda during July 2000-August 2001 and subsequently it was shown that \textit{dhps} 540E played the primary role in this association (Dorsey et al., 2004b). In Blantyre, Malawi the \textit{dhfr} 51I/59R/108N triple mutant when found together with \textit{dhps} 437G/540E double mutant alleles as the quintuple mutant, were associated with SP treatment failure (Kublin et al., 2002).

Alifrangis et al., (2003) studied \textit{dhfr} and \textit{dhps} at four sites with different endemicities in three different countries namely Hag Yousif, Sudan; Kibaha and Magoda, Tanzania and Matola, Mozambique and found there were generally higher frequencies of mutations in the SP parasitologic or clinical failure groups compared with the groups that had treatment success.

There were similar attempts to study SP efficacy in terms of phenotype and genotype in West African countries.
In Cameroon three patients who experienced SP treatment failure also had the \textit{dhfr} IRN triple mutant (Basco \textit{et al.}, 1998). In a mesoendemic area of Mali, Diourte \textit{et al.}, (1999) could not address directly whether or not \textit{dhfr} and \textit{dhps} point mutations were predictive of the low SP parasitological failure rate observed (1.1%, \textit{n}=190) during a 14 day follow-up period. However they observed higher rates of \textit{dhfr} 51I, 59R and 108N among post-treatment samples compared to pre-treatment samples suggesting that these mutations were precursors for SP failure \textit{in vivo} in areas of unstable malaria transmission (Diourte \textit{et al.}, 1999, Djimde \textit{et al.}, 2004). Other studies done in West Africa did find evidence of an association between \textit{dhps} point mutations and \textit{in vivo} treatment failure (Kun \textit{et al.} 1999; Mockenhaupt \textit{et al.} 2005; Dunyo \textit{et al.} 2006; Ndounga \textit{et al.}, 2007).

It is necessary to calibrate \textit{in vivo} studies emanating from different protocols and some issues related to this are presented in this chapter, together with a method of standardising results from \textit{in vivo} studies using different protocols. To examine the relationship between molecular markers and SP efficacy, both \textit{in vivo} and molecular studies across different countries in sub-Saharan Africa were matched according to time and place. This unique dataset was then used to assess the predictive value of allelic haplotypes or phenotypic outcome in exploratory analyses of the relationship between these two measures of resistance in East and West Africa.

5.2 Aim

To explore the relationship between measures of \textit{in vivo} SP efficacy and the frequency of \textit{dhfr} and \textit{dhps} allelic haplotypes circulating in the parasite population at the time of the efficacy trial.

This chapter presents a simplified categorisation of summarised antimalarial \textit{in vivo} efficacy data which enables studies carried out using different WHO protocols to be amalgamated and mapped across multiple sites.

5.3 Objectives

1. To calibrate \textit{in vivo} studies according to a classification system that allows comparison across different protocols
2. To match *in vivo* studies with *dhfr* and *dhps* allelic haplotype studies based on geographic location (nearest neighbour) and study year (± two years)

3. To investigate the influence of *dhfr* and *dhps* allelic haplotype frequencies on the prevalence of *in vivo* outcomes using logistic regression

### 5.4 Methods

#### Study selection

Basic inclusion and exclusion criteria were described in chapter 2. A subset of studies with a 28 day or longer follow-up period and PCR correction were drawn from the database (Table 5.1).

<table>
<thead>
<tr>
<th>Criteria for study selection</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Inclusion criteria</strong></td>
<td>SP monotherapy</td>
</tr>
<tr>
<td></td>
<td>28 day or longer follow-up period</td>
</tr>
<tr>
<td></td>
<td>PCR correction</td>
</tr>
<tr>
<td><strong>Unrestricted criteria</strong></td>
<td>all age groups</td>
</tr>
<tr>
<td></td>
<td>study design – including randomised and other trials</td>
</tr>
<tr>
<td></td>
<td>symptomatic or asymptomatic study participants</td>
</tr>
<tr>
<td></td>
<td>malaria transmission setting</td>
</tr>
</tbody>
</table>

#### Standardised *in vivo* treatment outcomes

In this study, a classification scheme for the different outcomes measured by *in vivo* efficacy studies was adopted. It categorises outcomes into three variables: S, R1 and R2 on the basis of a combination of parasitological and therapeutic responses (Table 5.2).

Descriptions of parasitological and therapeutic responses, where provided for each SP *in vivo* efficacy study, were noted along with the WHO protocol used and any deviations

192
thereof. These notes were used to classify outcomes into the three standardised categories as described below. Descriptions of the study end points as per the WHO protocols were provided in Tables 2.1a and b in chapter 2.

**R1:** Late treatment failure (LTF) or late clinical Failure (LCF) and late parasitological failure (LPF) were combined into the category R1 as it was not always clear whether or not fever could be a reliable indicator of treatment efficacy given confounding by antipyretics, concomitant illnesses and the acquired immunity of individual study participants. These regularly used *in vivo* study endpoints were grouped together with R1 parasitological failures, which had the same definition as LPF. R1 (late) responses were classified as R1 due to the occurrence of recrudescent parasites, implying decreasing drug efficacy.

**R2:** RII and RIII parasitological responses were grouped given the variability of microscopy which could limit the accuracy of the distinction between RII and RIII responses. These were grouped together with early treatment failures (ETFs) because they all reflect failures that would have occurred by seven days of follow-up. R1 (early) responses were classified as R2 (poor response) as its definition was similar to an RII response.

The combined total of R2+R1 gives the total treatment failure rate. From a public health perspective, there is a substantial difference between the consequences associated with ETFs (R2) and LTFs (R1) and in many instances the R1 responses needed to be analysed independently from the R2 responses.

**S:** This category included adequate clinical response (ACR), adequate clinical and parasitological response (ACPR) and treatment cure or success.
Table 5.2 Classification of in vivo treatment responses used to generate resistance profiles

<table>
<thead>
<tr>
<th>Resistance profile</th>
<th>Endpoint from WHO Protocols</th>
<th>Description of treatment efficacy</th>
</tr>
</thead>
<tbody>
<tr>
<td>R2</td>
<td>RI (early), RII, RIII, ETF</td>
<td>High grade resistance reflecting a poor response to SP, requiring alternative treatment</td>
</tr>
<tr>
<td>R1</td>
<td>RI (late), RI, LTF, LCF, LPF</td>
<td>Intermediate grade resistance response, indicates decreasing SP efficacy</td>
</tr>
<tr>
<td>R</td>
<td>R1 + R2</td>
<td>Total failure rate</td>
</tr>
<tr>
<td>S</td>
<td>S, ACR, ACPR</td>
<td>No resistance, adequate response indicating cure or treatment success</td>
</tr>
</tbody>
</table>

The matrix with the resistance profile is based on a review of the descriptions and classifications used by each in vivo study. It offers a uniform method of classifying heterogeneous antimalarial efficacy study endpoints in one of three resistance categories for comparison purposes.

Assumptions and points to consider with this approach

The assumption is that all studies adhered strictly to the WHO protocol that it employed. For example, patients who responded favourably to SP treatment remained so after the last follow-up day.

Some studies use both clinical (ACR/ETF/LTF) and parasitological (RI/RII/RIII) response which implies that the data would fit into more than one category in the matrix for the same resistance profile. There are two examples below from Driessen et al., (2002) and Talisuna et al., (2004a).

In Driessen et al., (2002) the WHO (1996) therapeutic response (ACR/ETF/LTF) and the RI/RII/RIII response for parasitological response were used. In assessing policy change criteria, Driessen et al., (2002) suggested that if the level of RIII parasitological response
exceeds 5–10%, the clinical response should be carefully evaluated and if this shows that
the median duration of clinical response was less than 14 days and/or the haematological
response was suboptimal compared with an effective drug, a change in first line treatment
ought to be considered.

In the study done by Talisuna et al., (2004a) the WHO (1996) protocol and ACR/ETF/LTF
therapeutic endpoints were used. However, these authors chose the RI/RII/RIII
classification system after the 28th day of follow-up because less than 10% of recruited
patients had treatment failures at day 14. Thus they did not use the ACPR/LCF/LPF/ETF
classification at day 14 due to the low treatment failure rate.

Feasibility of comparing results from different WHO protocols, as described by
WHO

WHO (1973) and WHO (1996) or WHO (2001)

A comparison of the classification endpoints described in the 1973 and 2001 protocols
appeared in a WHO (2005) report. Two issues concerning study variability and different
transmission intensity settings were highlighted in this report and these are mentioned
below:

(1) Application of the WHO (1973) protocol in the field was prone to variability and
adaptation which might influence comparability among studies. Furthermore, different
definitions of the S/RI/RII/RIII response might have been used within these studies.

(2) In high transmission areas, the S/RI/RII/RIII endpoints measured with the WHO (1973)
protocol are not comparable with the endpoints measured with the new protocols i.e. the
sum of RI+RII+RIII treatment failures is not equivalent to the sum of ETF+LCF (WHO,
2001).

(3) In low-to-moderate transmission areas, the RI+RII+RIII endpoints measured using the
WHO (1973) protocol are equivalent to ETF+LCF+LPF (WHO, 2001).

WHO (1996) and WHO (2001)

As mentioned previously, the primary difference between the WHO (1996) protocol and
the WHO (2001) protocol was the categorisation of treatment failure into LCF and LPF and
inclusion of parasitological response in ACPR. In high transmission settings, the
possibility of re-infections is high which can confound LCF and LPF endpoints. Despite
these differences between the two protocols, WHO (2005) reported that they are still
comparable. The 2001 protocol took into account different malaria transmission settings and changes were made to the follow-up period, sampling method and classification of efficacy. The WHO (2001) protocol recommended 14 or possibly 28 day follow-up periods in high transmission settings and 28 day follow-up periods in low/moderate transmission settings (Table 2.1 in chapter 2). This can be summarised as follows: in low to moderate malaria transmission settings, the RI+RII+RIII (WHO, 1973) endpoints are equivalent to ETF+ LCF+LPF (WHO, 2001) or ETF + LTF (WHO, 1996) since LTF = LCF + LPF.

Examples of studies that re-categorised therapeutic data

Two studies done in Equatorial Guinea (1992-1999) and Cameroon (pre-1998) related therapeutic response to parasitological response and both re-categorised the response in a similar way to that proposed here. These two studies are discussed below. Firstly, Roche et al., (2003) did a study in Malabo, Equatorial Guinea using the WHO (1996) protocol (ACR/ETF/LTF) with a 14 day follow-up period. They reclassified the WHO categories as described in Table 5.3 to compare in vivo efficacy studies done pre-1996 and post-1996. Studies done pre-1996 reported S/RI/RII/RIII responses and did not consider the clinical aspects of the treatment response.

<table>
<thead>
<tr>
<th>Reclassification by Roche et al., (2003)</th>
<th>Resistance Profile used</th>
<th>From WHO Protocols</th>
</tr>
</thead>
<tbody>
<tr>
<td>ETF = RIII</td>
<td>R2</td>
<td>RI (early), RII, RIII,ETF</td>
</tr>
<tr>
<td>LTF = RII, RII</td>
<td>R1</td>
<td>S/RI (late),RI, LTF,LCF,LPF</td>
</tr>
<tr>
<td>S = ACR</td>
<td>S</td>
<td>S,ACR,ACPR</td>
</tr>
</tbody>
</table>

The primary difference between the reclassification of Roche et al., (2003) and this study’s proposed reclassification system is to do with the RII parasitological response. In the work by Roche and colleagues (2003), the RII parasitologic response was classified as LTF while this study proposes that RII is more similar to an RIII and ETF response than a RI and LTF response for the reasons described previously, which were related to the variability of assessing parasites under the microscope.
In the second study, Basco et al., (1998) used the WHO (1996) and 28 day follow-up period in Yaounde, Cameroon. The definition of ACR/ETF/LTF in the Roche et al., (2003) and Basco et al., (1998) studies were similar and fit in with the standardised matrix with one difference related to presence or absence of fever with ACR used by the latter authors. An anomaly from these two studies was that the RII parasitological response was classified as indicative of ETF or LTF in Basco et al., (1998) but as LTF only in Roche et al., (2003).

Another study used the WHO (1994) protocol, although it is not a mainstream protocol. For example in Uganda, Kamya et al., (2001) used both WHO (1996) and WHO (1994) protocols. The revised WHO (1996) protocol for areas of intense transmission was used together with the modified criteria in WHO (1994). In this study clinical failures (ETF or LTF) and parasitological resistance (RII/RIII) were reported separately. The authors considered the re-categorisation of both resistance classification schemes by introducing two modifications related to the RII/RIII and ETF responses. The reclassification took into account ETFs and their parasite densities within the first three days of follow-up. Those with ETF and increasing parasite densities were classified as RI, whilst those who were ETF but had decreasing parasite densities were classified as RII. According to the present study’s proposed standardisation matrix, both categories would fall into the R2 resistance profile.

These re-calibrated in vivo data were used for further analyses with the molecular data described in the next section. In total 53 in vivo studies passed quality control criteria and the details of these studies are listed in Appendices 12 through 16.

**Matched resistance profile from in vivo studies with dhfr and dhps haplotypes**

The selected in vivo studies were matched with molecular haplotype studies of dhfr IRN, dhfr NCS (Appendix 13) and dhps (GE/GK/AK) (Appendix 14) according to the closest study period (± two year gap) and nearest neighbour. The data were grouped broadly into East and West African countries as shown in the map in Figure 5.1. The Western region consisted of 11 countries namely Angola, Burkina Faso, Cameroon, Chad, Democratic Republic of Congo (West) Gabon, Ghana, Madagascar, Mali, Republic of Congo and Sierra Leone. The Eastern region consisted of nine countries namely Democratic Republic
of Congo (East) Kenya, Malawi, Mozambique, South Africa, Tanzania, Uganda, North Sudan and South Sudan.

The frequencies of dhfr IRN and dhfr NCS haplotypes exclude a third category of ‘other’ haplotypes (see below). The coverage of in vivo data points are illustrated in Figure 2.3 in chapter 2. The different datasets were mapped using Mapinfo and Google Earth.

For each matched survey, the allelic haplotype frequency was calculated as the number of samples with the allelic haplotype/total number of samples analysed, as in equation one.

Equation 1:

Frequency of allelic haplotype = [number of samples with the allelic haplotype/total number of samples tested for which an allelic haplotype could be determined] X 100

The prevalence of in vivo efficacy was calculated as follows:

[number of people with in vivo efficacy outcome /total number of people completing the survey] X 100

To explore trends in the raw data, allelic haplotype frequencies were ordered from lowest to highest and frequencies at quarterly intervals were selected. The corresponding phenotype prevalence was observed at these frequencies. In the event of there being more than one phenotype prevalence, at a specific allele frequency, the mean prevalence was calculated.

**Logistic regression analysis of phenotypic and genotypic responses**

The aim was to assess whether or not there was any association between in vivo studies describing SP phenotypic outcome and genotypic measures of resistance. The data were derived by matching studies with the nearest geographic neighbour and the study year.

The phenotypic outcome was the number of people clinically classified as having resistance to SP treatment (in vivo R) or clinical success (in vivo S). Two measures of in vivo resistance were used in this analysis namely any resistance (R) and high grade resistance (R2). To perform a logistic regression the phenotypic outcome was expanded to each individual based on the number of people that completed each survey. Genotype frequencies were rescaled by 10% and the data were treated as surveys in Stata.
A logistic regression with robust standard errors was constructed, allowing for clustering by study site and stratified across the two regional strata namely East and West as illustrated by the map in Figure 5.1 and using equation one.

Equation 1: \[ \log \left( \frac{p}{1-p} \right) = \beta_0 + \beta_1 x_1 \]

where

\[ p = \text{probability of phenotype being present} \]
\[ 1-p = \text{probability of phenotype being absent} \]
\[ x_1 = \text{frequency of } dhfr \text{ or } dhps \text{ allelic haplotype} \]
\[ \beta_0 = \text{constant} \]

Receiver operating characteristic curves (ROC) were constructed using predicted probabilities from the logistic regression analysis. The predicted probabilities of any phenotype resistance were estimated using the results from the \textit{dhps} GE and \textit{dhps} GK models. The allelic haplotype frequencies were unadjusted. No cut-offs were used to generate the ROC using the phenotype response as the classification variable. The area under the receiver operating curve (AUC) was used to compare the models generated for \textit{dhps} GE and \textit{dhps} GK. No further interpretation of the AUC was conducted.

To assess whether or not there were any regional differences in the model estimated above, another logistic regression with robust standard errors was constructed, allowing for clustering by study site and this time without stratification across regions, using Stata’s survey methodology. Genotype frequencies were rescaled by 10%. Region was modelled as a covariate with and without interaction with the \textit{dfhr} IRN, \textit{dhps} GE and \textit{dhps} GK genotype. The Eastern region was used as the reference region throughout.

\textbf{Other \textit{dhfr} alleles}

The database was queried for \textit{dhfr} studies where haplotypes were available. Combinations of allelic haplotypes were extracted that were not of the \textit{dhfr} 51I/59R/109N triple mutant or fully sensitive 51N/59C/109S allelic haplotypes. These \textit{dhfr} single or double mutant allelic haplotypes are less common than the triple mutant or fully sensitive forms. This group was termed ”other” and haplotypes included in this category are listed in Table 5.4.
### Table 5.4 Less common \textit{dhfr} alleles

<table>
<thead>
<tr>
<th>Codon</th>
<th>50</th>
<th>51</th>
<th>59</th>
<th>108</th>
<th>140</th>
<th>164</th>
</tr>
</thead>
<tbody>
<tr>
<td>Sensitive</td>
<td>Cys (C)</td>
<td>Asn (N)</td>
<td>Cys (C)</td>
<td>Ser (S)</td>
<td>Val (V)</td>
<td>Ile (I)</td>
</tr>
<tr>
<td>Triple Mutant</td>
<td>Ile (I)</td>
<td>Arg (R)</td>
<td>Asn (N)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Others</td>
<td>Ile (I)</td>
<td>Cys (C)</td>
<td>Asn (N)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Ile (I)</td>
<td>Cys (C)</td>
<td>Asn (N)</td>
<td>Leu (L)</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Ile (I)</td>
<td>Cys (C)</td>
<td>Ser (S)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Ile (I)</td>
<td>Arg (R)</td>
<td>Ser (S)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Ile (I)</td>
<td>Arg (R)</td>
<td>Thr (T)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Asn (N)</td>
<td>Cys (C)</td>
<td>Asn (N)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Asn (N)</td>
<td>Cys (C)</td>
<td>Thr (T)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Asn (N)</td>
<td>Arg (R)</td>
<td>Thr (T)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Asn (N)</td>
<td>Arg (R)</td>
<td>Asn (N)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Asn (N)</td>
<td>Arg (R)</td>
<td>Ser (S)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Asn (N)</td>
<td>Tyr (T)</td>
<td>Ser (S)</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
The studies were categorised into East or West African sites as shown in the map in Figure 5.1 and a t-test with equal variances was performed to assess whether or not the frequencies of the less common *dhfr* allelic haplotypes differed between these two regions. The results were confirmed using the Wilcoxon Mann Whitney test. No further analysis was performed on these less common *dhfr* haplotypes.

Figure 5.1 Regional classification of sub-Saharan African countries for matched *in vivo* and molecular studies
5.5 Results

There were 53 SP in vivo studies that fulfilled the inclusion criteria and of these 42 studies were matched with the \textit{dhfr} surveys and 26 studies matched with surveys of \textit{dhps} allelic haplotypes. Appendices 13 and 14 list the matched studies and the publications from which these data were obtained are listed in appendices 12, 15 and 16. The following sections summarises the available data and presents the results of the logistic regression analyses.

**Matched in vivo studies with \textit{dhfr} NCS fully sensitive allelic haplotype**

The coverage of surveys available for statistical analyses of \textit{dhfr} NCS matched with SP phenotype is presented in Table 5.5. The study years ranged from 1996 to 2006, spanning 32 unique sites in 17 countries. The proportion of people with any \textit{in vivo} resistance was 29.1\% \((n=3766)\) whilst 8.0\% of people had high grade resistance \((n=3370)\). The frequency of \textit{dhfr} NCS was similar in surveys matched to \textit{in vivo} studies with R and R2 phenotypic responses \((21.6\% \text{ and } 23.5\%)\).
Table 5.5 Prevalence of phenotypic response to SP treatment matched with frequency of dhfr NCS allelic haplotype

<table>
<thead>
<tr>
<th>Phenotype Category</th>
<th>Total number of people with clinical outcome across all sites [median, range]</th>
<th>Total number of people included across all sites</th>
<th>Proportion* with treatment outcome (%) [95% confidence interval]</th>
<th>Samples tested positive for dhfr NCS haplotype</th>
<th>Samples tested for dhfr NCS haplotype</th>
<th>Proportion with dhfr NCS (%) [95% confidence interval]</th>
</tr>
</thead>
<tbody>
<tr>
<td>Category 1:</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Any phenotypic</td>
<td>1097 [18,1-89]</td>
<td>3766</td>
<td>29.1 [27.7-30.6]</td>
<td>1011</td>
<td>4672</td>
<td>21.6 [20.5-22.8]</td>
</tr>
<tr>
<td>resistance (in vivo R)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Category 2:</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>High grade</td>
<td>271 [6.5, 0-38]</td>
<td>3370</td>
<td>8.0 [7.1-9.0]</td>
<td>938</td>
<td>3994</td>
<td>23.5 [22.2-24.8]</td>
</tr>
<tr>
<td>phenotypic</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>resistance (in vivo R2)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Category 3:</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Treatment success</td>
<td>2645 [44, 7- 345]</td>
<td>3766</td>
<td>70.2 [68.7-71.7]</td>
<td>1011</td>
<td>4672</td>
<td>21.6 [20.5-22.8]</td>
</tr>
<tr>
<td>(in vivo S)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

*Proportion (%) with treatment outcome = number with specific treatment outcome/total number of people completing the survey x 100

Matched *in vivo* studies with dhfr IRN triple mutant allelic haplotype

Table 5.6 shows the coverage of surveys for the matched dhfr IRN and SP phenotype data. There were 32 unique sites from 17 countries covering the period from 1996 to 2006. The proportion of people with any phenotype resistance was 26.6% (n=5329). These were matched with surveys in which 52.1% samples contained the dhfr IRN triple mutant allelic haplotype (n=3344). The frequency of dhfr IRN was similar for matched surveys with R and R2 phenotypic responses (52.1% and 55.1%).
Table 5.6 Prevalence of phenotypic response to SP treatment and frequency of *dhfr* IRN allelic haplotype in matched surveys

<table>
<thead>
<tr>
<th>Phenotype Category</th>
<th>Total number of people with clinical outcome across all sites [median, range]</th>
<th>Total number of people included across all sites</th>
<th>Proportion with treatment outcome (%) [95% confidence interval]</th>
<th>Samples tested positive for <em>dhfr</em> IRN haplotype (frequency)</th>
<th>Samples tested for <em>dhfr</em> IRN haplotype</th>
<th>Proportion with <em>dhfr</em> IRN (%) [95% confidence interval]</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Category 1:</strong></td>
<td>1418 [18, 0-89]</td>
<td>5329</td>
<td>26.6 [25.4-27.8]</td>
<td>3344</td>
<td>6421</td>
<td>52.1 [50.8-53.3]</td>
</tr>
<tr>
<td>Any phenotypic resistance (<em>in vivo</em> R)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>Category 2:</strong></td>
<td>346 [5,0-38]</td>
<td>4600</td>
<td>7.5 [6.8-8.3]</td>
<td>3063</td>
<td>5562</td>
<td>55.1 [53.8-56.4]</td>
</tr>
<tr>
<td>High grade phenotypic resistance (<em>in vivo</em> R2)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>Category 3:</strong></td>
<td>3862 [52, 7-345]</td>
<td>5329</td>
<td>72.5 [71.3-73.7]</td>
<td>3344</td>
<td>6421</td>
<td>52.1 [50.8-53.3]</td>
</tr>
<tr>
<td>Treatment success (<em>in vivo</em> S)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
Tables 5.7 and 5.8 describe the observed prevalence of SP phenotype resistance at various \textit{dhfr} and allelic haplotype frequencies. Relatively low \textit{dhfr} NCS frequencies were observed with specific higher phenotypic resistance levels. A clear trend of generally increasing \textit{dhfr} NCS frequencies was observed with increasing treatment success rates (Table 5.8).

Table 5.7 Observed prevalence of SP phenotype resistance at various \textit{dhfr} allelic haplotype frequencies

<table>
<thead>
<tr>
<th>Frequency (%) of \textit{dhfr} allelic haplotype</th>
<th>Observed prevalence (%) of any grade resistance (\textit{in vivo} R)</th>
</tr>
</thead>
<tbody>
<tr>
<td>NCS</td>
<td></td>
</tr>
<tr>
<td>10.6</td>
<td>41.5</td>
</tr>
<tr>
<td>25.4</td>
<td>40.2</td>
</tr>
<tr>
<td>58.3</td>
<td>16.3</td>
</tr>
<tr>
<td>74.0</td>
<td>13.8</td>
</tr>
<tr>
<td>IRN</td>
<td></td>
</tr>
<tr>
<td>11.4</td>
<td>16.1</td>
</tr>
<tr>
<td>27.5</td>
<td>26.8</td>
</tr>
<tr>
<td>51.0</td>
<td>40.2</td>
</tr>
<tr>
<td>75</td>
<td>42.4</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Frequency (%) of \textit{dhfr} allelic haplotype</th>
<th>Observed prevalence (%) of high grade resistance (R2)</th>
</tr>
</thead>
<tbody>
<tr>
<td>NCS</td>
<td></td>
</tr>
<tr>
<td>10.6</td>
<td>9.8</td>
</tr>
<tr>
<td>25.4</td>
<td>2.8</td>
</tr>
<tr>
<td>58.3</td>
<td>12.0</td>
</tr>
<tr>
<td>74.0</td>
<td>5.7</td>
</tr>
<tr>
<td>IRN</td>
<td></td>
</tr>
<tr>
<td>11.9</td>
<td>5.7</td>
</tr>
<tr>
<td>20.0</td>
<td>1.3</td>
</tr>
<tr>
<td>58.4</td>
<td>1.5</td>
</tr>
<tr>
<td>75</td>
<td>4.6</td>
</tr>
</tbody>
</table>
Increasing \textit{dhfr} IRN frequencies were observed with increasing phenotypic resistance (Table 5.7) but this was more apparent among the matched studies of any phenotypic resistance (R) rather than with high grade phenotypic resistance (R2).

Table 5.8 Observed prevalence of SP treatment success at various \textit{dhfr} allelic haplotype frequencies

<table>
<thead>
<tr>
<th>Frequency (%) of \textit{dhfr} allelic haplotype</th>
<th>Observed Prevalence (%) of treatment success (S)</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>NCS</strong></td>
<td></td>
</tr>
<tr>
<td>10.6</td>
<td>58.5</td>
</tr>
<tr>
<td>25.4</td>
<td>63.7</td>
</tr>
<tr>
<td>58.3</td>
<td>83.7</td>
</tr>
<tr>
<td>74.0</td>
<td>85.4</td>
</tr>
<tr>
<td><strong>IRN</strong></td>
<td></td>
</tr>
<tr>
<td>11.4</td>
<td>67.7</td>
</tr>
<tr>
<td>27.5</td>
<td>71.4</td>
</tr>
<tr>
<td>58.4</td>
<td>87.7</td>
</tr>
<tr>
<td>75</td>
<td>57.6</td>
</tr>
</tbody>
</table>
Univariate logistic regression analysis of SP phenotypic outcomes with *dhfr* genotypes

With every 10% increase in *dhfr* NCS frequencies, there was a lower likelihood of phenotypic resistance (R) and higher likelihood of treatment success (Table 5.9).

The odds of any phenotypic resistance was lowered 0.835 (CI 0.676-1.032) times, per 10% increase in *dhfr* NCS frequencies although there was weak evidence for this with p=0.092. Similarly, a 10% increase in *dhfr* NCS frequencies resulted in a lower likelihood of high grade phenotypic resistance (R2) (OR 0.831, CI 0.703-0.982) and the evidence for this was strong with p=0.031 (Table 5.9). Furthermore, the odds of treatment success were 1.187 times higher (CI 0.964-1.461) per 10% increase in *dhfr* NCS frequencies. However p=0.104, indicating weak evidence for this.

The model for *dhfr* IRN triple mutant showed increasing phenotypic resistance and decreasing treatment success with every 10% increment in *dhfr* IRN frequencies (Table 5.9). Specifically the odds of any phenotypic resistance was 1.180 (CI 1.040-1.339) times higher with every 10% increment in *dhfr* IRN frequencies and there was strong evidence for this with p=0.012. A 10% increase in *dhfr* IRN resulted in higher odds (OR 1.110, CI 0.980-1.258) of high grade phenotypic resistance although there was weak evidence for this (p=0.097). Conversely, the odds of treatment success resulted in 0.858 times lower odds (CI 0.764-0.964) per 10% increase in *dhfr* IRN and the evidence for this was strong with p=0.011.
Table 5.9 Results of analysis for dhfr NCS and dhfr IRN predicting SP phenotype

The genotype frequencies were rescaled by 10% and dhfr NCS and dhfr IRN was modelled individually as covariates with SP treatment outcome in different models. Two models were constructed for each phenotypic outcome category.

| Category 1: Any phenotypic resistance (in vivo R) |
|---|---|---|---|
| **Covariate** | **Prevalence of treatment outcome (%) [95% confidence interval]** | **Odds ratio of treatment outcome per 10% increase in allelic haplotype [95% confidence interval]** | **P value** |
| **Model 1: dhfr NCS** | 29.1 [27.7-30.6] | 0.835[0.676-1.032] | 0.092 |
| **Model 2: dhfr IRN** | 26.6 [25.4-27.8] | 1.180[1.040-1.339] | 0.012 |

| Category 2: High grade phenotypic resistance (in vivo R2) |
|---|---|---|---|
| **Model 1: dhfr NCS** | 8.0 [7.1-9.0] | 0.831[0.703-0.982] | 0.031 |
| **Model 2: dhfr IRN** | 7.5 [6.8-8.3] | 1.110[0.980-1.258] | 0.097 |

| Category 3: Treatment success (in vivo S) |
|---|---|---|---|
| **Model 1: dhfr NCS** | 70.2 [68.7-71.7] | 1.187[0.964-1.461] | 0.104 |
| **Model 2: dhfr IRN** | 72.5 [71.3-73.7] | 0.858[0.764-0.964] | 0.011 |

Given the abovementioned evidence that dhfr IRN was a strong predictor of phenotypic outcome, a logistic regression was performed using region as a covariate. Overall, there was strong evidence showing a higher likelihood of any SP phenotypic resistance occurring in the Eastern region compared to the Western region (OR 2.234, CI 1.119-4.458, p = 0.024).

In the model, the likelihood of any phenotypic resistance (R) was 1.158 times higher (CI 1.022-1.314) per 10% dhfr IRN and there was strong evidence for this as p=0.023. There was strong evidence indicating that any phenotypic resistance (R) was lower in the Western region compared to the Eastern region (OR 0.520, CI 0.280 – 0.967, p = 0.039).
However, when the interaction between \textit{dhfr} IRN and region was included in the model this regional effect was no longer evident (OR 1.004, CI 0.826 – 1.222, p=0.961).

**Matched \textit{in vivo} studies with \textit{dhps} allelic haplotypes**

Data were available for 29 unique sites in 18 countries where \textit{dhps} allelic haplotypes were matched with SP phenotypic response. The overall rate of treatment success was 74.2\% (Table 5.10). The total frequencies of \textit{dhps} GE, GK and AK allelic haplotypes in surveys matched with the three SP phenotypes are listed in Table 5.11. Overall, the AK frequencies were highest, followed by GE frequencies and GK frequencies (Table 5.11). Inspection of the \textit{dhps} allelic haplotype frequencies and matched prevalence of resistance was used to look for trends in variability at specific phenotypic responses (Tables 5.12, 5.13 and 5.14). There was an observed higher likelihood of high grade phenotypic response (R2) at specific increasing \textit{dhps} GE frequencies only (Table 5.13).

Table 5.10 Prevalence of phenotypic response to SP treatment among studies matched with \textit{dhps} allelic haplotypes

<table>
<thead>
<tr>
<th>Phenotype Category</th>
<th>Total number of people with clinical outcome across all sites [median, range]</th>
<th>Total number of people included across all sites</th>
<th>Proportion with treatment outcome (%) [95% confidence interval]</th>
</tr>
</thead>
<tbody>
<tr>
<td>Category 2: High grade phenotypic resistance \textit{(in vivo R2)}</td>
<td>266[5, 0-38]</td>
<td>3837</td>
<td>6.9 [0.61-7.8]</td>
</tr>
<tr>
<td>Category 3: Treatment success \textit{(in vivo S)}</td>
<td>4510 [51,7-345]</td>
<td>3350</td>
<td>74.2 [73.0-75.6]</td>
</tr>
</tbody>
</table>
Table 5.11 Frequencies of \textit{dhps} GE, GK and AK allelic haplotypes in surveys matched with SP phenotype

<table>
<thead>
<tr>
<th>Phenotype Category</th>
<th>Samples tested positive for \textit{dhps} GE [% positive]</th>
<th>Samples tested positive for \textit{dhps} GK [% positive]</th>
<th>Samples tested positive for \textit{dhps} AK [% positive]</th>
<th>Samples tested for \textit{dhps} allelic haplotype</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Category 1:</strong> Any phenotypic resistance (in vivo R)</td>
<td>1551 [22.1]</td>
<td>1410 [20.1]</td>
<td>4027 [57.5]</td>
<td>7006</td>
</tr>
<tr>
<td><strong>Category 2:</strong> High grade phenotypic resistance (in vivo R2)</td>
<td>1572 [25.8]</td>
<td>1349 [22.2]</td>
<td>3143 [51.6]</td>
<td>6090</td>
</tr>
</tbody>
</table>

Generally, the three \textit{dhps} allelic haplotypes GE/GK/AK add up to the total number tested for these alleles.
Table 5.12 Observed prevalence of any grade resistance (*in vivo R*) at various *dhps* allelic haplotype frequencies

<table>
<thead>
<tr>
<th>Frequency <em>dhps</em> haplotype (%)</th>
<th>Observed prevalence (%) of any grade resistance (<em>in vivo R</em>)</th>
</tr>
</thead>
<tbody>
<tr>
<td><em><strong>dhps GE</strong></em></td>
<td></td>
</tr>
<tr>
<td>9.8</td>
<td>10.1</td>
</tr>
<tr>
<td>30.6</td>
<td>42.4</td>
</tr>
<tr>
<td>45.2</td>
<td>22.9</td>
</tr>
<tr>
<td>98.6</td>
<td>58.2</td>
</tr>
<tr>
<td><em><strong>dhps GK</strong></em></td>
<td></td>
</tr>
<tr>
<td>10.2</td>
<td>42.2</td>
</tr>
<tr>
<td>25.0</td>
<td>7.0</td>
</tr>
<tr>
<td>46.2</td>
<td>4.4</td>
</tr>
<tr>
<td>98.3</td>
<td>13.8</td>
</tr>
<tr>
<td><em><strong>dhps AK</strong></em></td>
<td></td>
</tr>
<tr>
<td>9.2</td>
<td>41.5</td>
</tr>
<tr>
<td>25.0</td>
<td>18.5</td>
</tr>
<tr>
<td>51.2</td>
<td>4.4</td>
</tr>
<tr>
<td>100.0</td>
<td>69.9</td>
</tr>
</tbody>
</table>
Table 5.13 Observed prevalence of high grade phenotypic resistance at various *dhps* allelic haplotype frequencies

<table>
<thead>
<tr>
<th>Frequency <em>dhps</em> haplotype (%)</th>
<th>Observed prevalence (%) of high grade resistance (R2)</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>dhps GE</strong></td>
<td></td>
</tr>
<tr>
<td>9.8</td>
<td>1.3</td>
</tr>
<tr>
<td>30.6</td>
<td>4.5</td>
</tr>
<tr>
<td>45.2</td>
<td>16.7</td>
</tr>
<tr>
<td>98.6</td>
<td>24.8</td>
</tr>
<tr>
<td><strong>dhps GK</strong></td>
<td></td>
</tr>
<tr>
<td>10.2</td>
<td>4.5</td>
</tr>
<tr>
<td>25.0</td>
<td>2.0</td>
</tr>
<tr>
<td>46.2</td>
<td>1.1</td>
</tr>
<tr>
<td>98.3</td>
<td>5.7</td>
</tr>
<tr>
<td><strong>dhps AK</strong></td>
<td></td>
</tr>
<tr>
<td>9.2</td>
<td>9.8</td>
</tr>
<tr>
<td>25.0</td>
<td>4.4</td>
</tr>
<tr>
<td>51.2</td>
<td>1.1</td>
</tr>
<tr>
<td>100.0</td>
<td>5.8</td>
</tr>
</tbody>
</table>
Table 5.14 Observed prevalence of SP treatment success at various *dhfr* allelic haplotype frequencies

<table>
<thead>
<tr>
<th>Frequency <em>dhps</em> haplotype (%)</th>
<th>Observed prevalence (%) of treatment success (S)</th>
</tr>
</thead>
<tbody>
<tr>
<td><em>dhps</em> GE</td>
<td></td>
</tr>
<tr>
<td>9.8</td>
<td>93.2</td>
</tr>
<tr>
<td>30.6</td>
<td>57.6</td>
</tr>
<tr>
<td>45.2</td>
<td>77.1</td>
</tr>
<tr>
<td>98.6</td>
<td>41.8</td>
</tr>
<tr>
<td><em>dhps</em> GK</td>
<td></td>
</tr>
<tr>
<td>10.2</td>
<td>57.6</td>
</tr>
<tr>
<td>25.0</td>
<td>93.0</td>
</tr>
<tr>
<td>46.2</td>
<td>95.6</td>
</tr>
<tr>
<td>98.3</td>
<td>85.4</td>
</tr>
<tr>
<td><em>dhps</em> AK</td>
<td></td>
</tr>
<tr>
<td>9.2</td>
<td>58.5</td>
</tr>
<tr>
<td>25.0</td>
<td>80.8</td>
</tr>
<tr>
<td>51.2</td>
<td>95.6</td>
</tr>
<tr>
<td>100.0</td>
<td>30.1</td>
</tr>
</tbody>
</table>
Univariate logistic regression analysis of SP phenotypic and *dhps* genotypic responses

The results of the logistic regression analysis for *dhps* GE, *dhps* GK and *dhps* AK individually predicting *in vivo* outcome are shown in Table 5.15. A fourth logistic regression analysis assessed the combined effect of the *dhps* GE double mutant together with the *dhps* GK single mutant on phenotypic outcome. The results are discussed sequentially for each of the three phenotypic outcomes. Category one describing any phenotype resistance is discussed first below.

Table 5.15 Results of *dhps* GE, GK, GE+GK and AK individually predicting *in vivo* outcome

The genotype frequencies were rescaled by 10% and each covariate was modelled individually with phenotype outcome. Four models were constructed for each phenotypic outcome category.

<table>
<thead>
<tr>
<th>Covariate</th>
<th>Frequency of allelic haplotype (%) [95% confidence interval]</th>
<th>Odds ratio of treatment outcome per 10% increase in genotype [95% confidence interval]</th>
<th>P value</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Category 1: Any phenotypic resistance (in vivo R)</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Model 2: GK</td>
<td>20.1 [19.2-21.1]</td>
<td>0.969 [0.858-1.094]</td>
<td>0.602</td>
</tr>
<tr>
<td>Model 3: AK</td>
<td>57.5 [56.3-58.6]</td>
<td>0.851 [0.742-0.977]</td>
<td>0.023</td>
</tr>
<tr>
<td>Model 4: GE+AK</td>
<td>42.3 [41.1-43.4]</td>
<td>1.178 [1.025-1.354]</td>
<td>0.022</td>
</tr>
<tr>
<td><strong>Category 2: High grade phenotypic resistance (in vivo R2)</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Model 1: GE</td>
<td>25.8 [24.7-26.9]</td>
<td>1.148 [1.048-1.257]</td>
<td>0.004</td>
</tr>
<tr>
<td>Model 2: GK</td>
<td>22.2 [21.1-23.2]</td>
<td>0.951 [0.832-1.087]</td>
<td>0.449</td>
</tr>
<tr>
<td>Model 3: AK</td>
<td>51.6 [51.6-52.8]</td>
<td>0.847 [0.761-0.942]</td>
<td>0.003</td>
</tr>
<tr>
<td>Model 4: GE+GK</td>
<td>48.0 [46.7-49.2]</td>
<td>1.182 [1.060-1.318]</td>
<td>0.004</td>
</tr>
<tr>
<td><strong>Category 3: Treatment success (in vivo S)</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Model 1:GE</td>
<td>24.7 [23.7-25.7]</td>
<td>0.856 [0.787-0.931]</td>
<td>0.001</td>
</tr>
<tr>
<td>Model 2:GK</td>
<td>19.4 [18.5-20.4]</td>
<td>1.069 [0.947-1.205]</td>
<td>0.272</td>
</tr>
<tr>
<td>Model 3:AK</td>
<td>55.5 [54.3-56.6]</td>
<td>1.152 [1.024-1.295]</td>
<td>0.020</td>
</tr>
<tr>
<td>Model 4:GE+GK</td>
<td>44.2 [43.0-45.3]</td>
<td>0.867 [0.769-0.977]</td>
<td>0.021</td>
</tr>
</tbody>
</table>

**Category 1: Any phenotypic resistance (in vivo R)**

The model predicted a higher likelihood of phenotypic resistance with an increase of 10% *dhps* GE and *dhps* GE+ *dhps* GK. The odds of any phenotypic resistance was 1.164 times higher (CI 1.059-1.280) when *dhps* GE increased by 10% and there was strong evidence for this as indicated by p=0.002 (Table 5.15). The relationship of *dhps* GK with any phenotypic resistance (R) was weak (p=0.602) and the likelihood of any phenotypic
resistance was actually lower with every 10% increase in dhps GK (OR 0.969, CI 0.858-1.094).

When estimating the combined effect of dhps GE+ dhps GK, the odds of any phenotypic resistance was 1.178 times higher (CI 1.025-1.354) per 10% increase in these allelic haplotypes and there was strong evidence for this as indicated by p=0.022. As expected the odds of any phenotypic resistance (R) was reduced (OR 0.969, CI 0.858-1.094) with a 10% increase in dhps AK frequency. There was strong evidence for this (p = 0.23) (Table 5.15).

**Category 2: High grade phenotypic resistance (in vivo R2)**

The likelihood of high grade phenotypic resistance (R2) was higher with every 10% increment in dhps GE, (OR 1.148, CI 1.048-1.25) and there was strong evidence for this as indicated by the low p value of p=0.004 (Table 5.15). Conversely the likelihood of high grade phenotypic resistance was lowered with each 10% increment in dhps AK frequency (OR 0.951, CI 0.832-1.087). There was strong evidence for this (p = 0.003). Interestingly, as with R, the odds of high grade phenotypic resistance (R2) was reduced 0.951 times per 10% increase in dhps GK, though the strength of the evidence for this was weak with p=0.449. When the combined effect of dhps GE+ dhps GK was assessed, the odds of high grade phenotypic resistance was 1.182 times higher (CI 1.060-1.318) with p=0.004.

**Category 3: Treatment success (in vivo S)**

There was a reduction in treatment success (OR 0.856, 0.787-0.931) as dhps GE frequency increased by 10%. There was strong evidence for this (p = 0.001). Similarly the odds of treatment success was lowered (OR 0.867, CI 0.769-0.977) as the combined effect of dhps GE+ dhps GK increased by 10% and the evidence for this was strong (p = 0.021) (Table 5.15). For every 10% increase in dhps GK alone, however there was a higher likelihood of treatment success (OR 1.069, CI 0.947-1.205), although there was weak evidence for this effect (p= 0.272). The odds of treatment success were higher with every 10% increase in dhps AK (OR 1.152, CI 1.024-1.295) and there was strong evidence for this as p=0.020.

To summarise the models presented in Table 5.15, there was strong evidence of decreasing high grade phenotype resistance and any phenotype resistance with
increasing frequencies of the \textit{dhps} AK sensitive allelic haplotype. While there was strong evidence of increasing high grade phenotype resistance and any phenotype resistance with increasing \textit{dhps} GE frequencies, surprisingly there was little evidence of a relationship between \textit{dhps} GK frequencies and phenotypic resistance.

Following these results, ROC curves were constructed to compare the sensitivity and specificity of the individual models for \textit{dhps} GE and \textit{dhps} GK.

\textbf{Sensitivity and specificity of \textit{dhps} GE and \textit{dhps} GK models}

Using the results presented for category one in Table 5.15 receiver operator curves were constructed. The area under the curve was 0.627 (95\% CI 0.606-0.648) for \textit{dhps} GE and 0.515 (95\% CI 0.495-0.534) (Figure 5.2). The model for \textit{dhps} GE was better in discriminating between phenotypic resistance versus no phenotypic resistance (p<0.001) compared to the \textit{dhps} GK model.

Figure 5.2 Sensitivity and specificity of \textit{dhps} GE and \textit{dhps} GK incremented by 10\% frequencies, in predicting any phenotypic resistance

The two \textit{dhps} GE and GK allelic haplotypes were then included in the model as covariates and these results are described next.
Multivariable logistic regression analyses of SP phenotypic and \textit{dhps} genotypic responses

The results of the logistic regression analysis using \textit{dhps} GE, \textit{dhps} GK as covariates predicting \textit{in vivo} outcome are shown in Table 5.16. The results are discussed separately for each of the three phenotypic outcomes.

Table 5.16 Multivariable logistic regression analysis using both GE and GK as covariates for predicting \textit{in vivo} outcome

The genotype frequencies were rescaled by 10\% and the two covariates \textit{dhps} GE and \textit{dhps} GK were modelled together with each SP phenotype outcome. Therefore three models were constructed for each phenotypic outcome category using the two covariates.

<table>
<thead>
<tr>
<th>\textit{dhps} Covariate</th>
<th>Frequency of allelic haplotype (%) [95% confidence interval]</th>
<th>Adjusted* Odds ratio of treatment outcome per 10% increase in genotype [95% confidence interval]</th>
<th>P value</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Category 1:</strong> Any phenotypic resistance (<strong>in vivo</strong> R)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>GE</td>
<td>22.1 [21.2-23.1]</td>
<td>1.211[1.060-1.384]</td>
<td>0.006</td>
</tr>
<tr>
<td>GK</td>
<td>20.1 [19.2-21.1]</td>
<td>1.096[0.949-1.266]</td>
<td>0.206</td>
</tr>
<tr>
<td><strong>Category 2:</strong> High grade phenotypic resistance (<strong>in vivo</strong> R2)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>GE</td>
<td>25.8 [24.7-26.9]</td>
<td>1.197[1.087-1.319]</td>
<td>0.001</td>
</tr>
<tr>
<td>GK</td>
<td>22.2 [21.1-23.2]</td>
<td>1.096[0.995-1.206]</td>
<td>0.062</td>
</tr>
<tr>
<td><strong>Category 3:</strong> Treatment success (<strong>in vivo</strong> S)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>GE</td>
<td>24.7[23.7-25.7]</td>
<td>0.839[0.747-0.941]</td>
<td>0.004</td>
</tr>
<tr>
<td>GK</td>
<td>19.4[18.5-20.4]</td>
<td>0.952[0.835-1.085]</td>
<td>0.453</td>
</tr>
</tbody>
</table>

*Odds ratio with \textit{dhps} GE adjusted for \textit{dhps} GK and vice versa

**Category 1: Any phenotype resistance:**

The odds of any phenotypic resistance was higher (Adjusted OR 1.211, CI 1.060-1.384) with every 10\% increment in \textit{dhps} GE frequencies, after adjusting for \textit{dhps} GK frequencies and the evidence for this was strong with p=0.006 (Table 5.16.) Similarly, the likelihood of any phenotypic resistance was higher (Adjusted OR 1.096, CI 0.949-1.266) with every 10\% increase in \textit{dhps} GK frequencies after adjusting for \textit{dhps} GE, but the evidence for this was weak with p=0.206.
Category 2: High grade phenotypic resistance \textit{(in vivo R2)}

After adjusting for \textit{dhps} GK the odds of high grade phenotypic resistance was 1.197 times higher (CI 1.087-1.319) per 10% increase in \textit{dhps} GE frequencies and the evidence for this was strong with \( p=0.001 \) (Table 5.16.). After adjusting for \textit{dhps} GE, the odds of any phenotypic resistance was 1.096 times higher (CI 0.995-1.206) per 10% increase in \textit{dhps} GK frequencies but the evidence for this was weak with \( p=0.062 \) (Table 5.16.).

Category 3: Treatment success \textit{(in vivo S)}

The odds of treatment success was lowered with every 10% increase in both \textit{dhps} GE and \textit{dhps} GK frequencies, OR 0.839 (CI 0.747-0.941) and OR 0.952 (CI 0.835-1.085) respectively. There was strong evidence for this effect but only for the \textit{dhps} GE frequencies with \( p=0.004 \).

Comparing the models in Tables 5.15 and 5.16, there is evidence that the effect of \textit{dhps} GK to increase any phenotypic resistance is not as strong as the effect of \textit{dhps} GE on any phenotypic resistance.

The model in Table 5.16 assessed the effect of \textit{dhps} GE and \textit{dhps} GK as covariates on phenotypic outcome and found that both genotypes increased the likelihood of phenotypic resistance and decreased the likelihood of treatment success. However, it must be noted that the evidence for this was weak when assessing \textit{dhps} GK as a covariate in this model. With increasing \textit{dhps} GE frequencies there was a higher likelihood of phenotypic resistance and lower likelihood of treatment success (Table 5.15). The same trend was observed when \textit{dhps} GE frequencies were added together with \textit{dhps} GK frequencies in a logistic model.

In a subsequent analysis all three \textit{dhps} allelic haplotypes namely \textit{dhps} GE, \textit{dhps} GK and \textit{dhps} AK were used as covariates predicting \textit{in vivo} outcome. However, the results were inconsistent with the previous two models and the high \( p \) values (\( p>0.263 \)) indicate weak evidence to support this model.

Given the observed effect of \textit{dhps} GE on phenotypic resistance as described in Tables 5.15 and 5.16 and the observation that \textit{dhps} GE is predominant in East Africa, region was
introduced into the model as an explanatory variable to test whether or not the effect was due to region or to $dhps$ GE itself. The results are presented in Table 5.17.

**Effect of $dhps$ GE and $dhps$ GK and region on any phenotypic resistance**

Firstly, the model tested whether or not there were any regional differences in the prevalence of any resistance (Table 5.17). The odds of any resistance was lower in the Western region (OR 0.605, 0.252-1.452) compared to the Eastern region but the high p value of 0.252 indicated weak evidence for this.

Table 5.17 Regional variation in phenotypic outcome

<table>
<thead>
<tr>
<th>Covariate</th>
<th>Prevalence of treatment outcome (%) [95% confidence interval]</th>
<th>Odds ratio [95% confidence interval]</th>
<th>P value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Model 1: Any phenotypic resistance (in vivo R)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>East</td>
<td>69.3 [66.3-72.1]</td>
<td>1</td>
<td></td>
</tr>
<tr>
<td>West</td>
<td>30.7 [27.9-33.7]</td>
<td>0.605 [0.252-1.452]</td>
<td>0.252</td>
</tr>
</tbody>
</table>

Thereafter, the effect of $dhps$ GE and $dhps$ GK on any phenotypic resistance (category one) was modelled to assess whether or not they were modified by regional differences (Table 5.18).

Table 5.18 Multivariate logistic regression analysis of SP phenotypic outcome with genotype and region as covariates in one model

<table>
<thead>
<tr>
<th>Covariate</th>
<th>Prevalence of treatment outcome (%) [95% confidence interval]</th>
<th>Adjusted* Odds ratio [95% confidence interval]</th>
<th>P value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Model 1: Any phenotypic resistance (in vivo R) with $dhps$ GE and region</td>
<td>dhps GE 23.2 [21.9-24.4]</td>
<td>1.195 [1.032-1.384] per 10% increase</td>
<td>0.019</td>
</tr>
<tr>
<td>Region</td>
<td>East 69.3 [66.3-72.1]</td>
<td>1</td>
<td></td>
</tr>
<tr>
<td>Region</td>
<td>West 30.7 [27.9-33.7]</td>
<td>1.440 [0.443-4.681] per 10% increase</td>
<td>0.535</td>
</tr>
<tr>
<td>Model 2: Any phenotypic resistance (in vivo R) with $dhps$ GK and region</td>
<td>dhps GK 23.2 [21.9-24.4]</td>
<td>1.229 [1.003-1.507] per 10% increase</td>
<td>0.047</td>
</tr>
<tr>
<td>Region</td>
<td>East 69.3 [66.3-72.1]</td>
<td>1</td>
<td></td>
</tr>
<tr>
<td>Region</td>
<td>West 30.7 [27.9-33.7]</td>
<td>0.166 [0.0367-0.750] per 10% increase</td>
<td>0.021</td>
</tr>
</tbody>
</table>

*Odds ratio with $dhps$ allelic haplotype adjusted for region
The likelihood of any phenotypic resistance per 10% increase in \textit{dhps} GE was higher after adjusting for regional differences (OR 1.195, CI 1.032-1.384) and there was strong evidence for this with p=0.019 (Table 5.18). The higher likelihood of any phenotypic resistance per 10% increase in \textit{dhps} GE was consistent with the previous models presented earlier. Although the likelihood of any phenotypic resistance was higher in the Western region compared to the Eastern region (OR 1.440, CI 0.443-4.681) after adjusting for differences in \textit{dhps} GE frequencies, the evidence for this difference was weak (p=0.535) (Table 5.18).

After adjusting for the confounding effect of region, there was borderline evidence for a higher likelihood of SP phenotypic resistance associated with increasing GK (Adjusted OR 1.229, 0.047). This suggests some evidence of an independent association between SP phenotypic resistance and \textit{dhps} GK frequencies.

After adjusting for differences in the frequency of \textit{dhps} GK, the likelihood of any phenotypic resistance was lower (OR 0.166, CI 0.0367- 0.750) in the West compared to the East and there was strong evidence for this with p=0.021 (Table 5.18). To assess the regional effect of \textit{dhps} GE and \textit{dhps} GK on SP phenotypic resistance, an interaction term was introduced.

Again, there was a higher likelihood of any phenotypic resistance per 10% increase in \textit{dhps} GE (OR 1.193, CI 1.030-1.382) with p=0.020, indicating strong evidence for this. The likelihood of any phenotypic resistance was higher when the regional effect was assessed together with \textit{dhps} GE frequency as an interaction term (OR 1.572, CI 0.712-3.472] although there was weak evidence for this as p=0.255.

There was a higher likelihood of any phenotypic resistance per 10% increase in \textit{dhps} GK (OR 1.739, CI 0.0534-5.665) but there was weak evidence for this as p=0.348. There was also a lower likelihood of any phenotypic resistance when the regional effect was assessed together with \textit{dhps} GK frequency as an interaction term (OR 0.696, CI 0.210-2.304), with weak evidence for this as p=0.543.
Other *dhfr* alleles

The t-test showed that there is a statistically significant difference between the mean frequency of the less common *dhfr* allelic haplotypes among the East and West African countries (*t* = -3.4128, *p*=0.0008). The East African countries have a significantly higher mean frequency (45.5%) of the ‘other’ *dhfr* allelic haplotypes than the West African countries (29.9%) (Table 5.19).

Table 5.19 Surveys with less common *dhfr* allelic haplotypes categorised by region

<table>
<thead>
<tr>
<th>Region</th>
<th>Number of Surveys</th>
<th>Mean (%) frequency of other <em>dhfr</em> allelic haplotypes</th>
</tr>
</thead>
<tbody>
<tr>
<td>West</td>
<td>88</td>
<td>29.9[23.4-36.5]</td>
</tr>
<tr>
<td>East</td>
<td>153</td>
<td>45.5[39.8-51.2]</td>
</tr>
</tbody>
</table>

5.6 Discussion

There are other factors which affect *in vivo* response to SP treatment in addition to the influence of *dhfr* and *dhps* allelic haplotypes explored here. Not all treatment failures can be attributed to drug resistance as mentioned in chapter 2. A major factor is immunity. Among patients surveyed in Nagongera Health Centre, Tororo District in Eastern Uganda which is found along the Kenyan border, Talisuna *et al.*, (2004b) found that patients with *P. falciparum* infections responded favourably to SP treatment despite detecting relatively high prevalence of *dhfr* and *dhps* point mutations. Acquired immunity allows people to clear parasites even when point mutations are present (Khalli *et al.*, 2002). Against the backdrop of these highly variable patient-specific effects, the ecological analyses described here assessed the effect of *dhfr* and *dhps* allelic haplotypes on SP phenotypic outcome.

As expected the sensitive *dhfr* NCS allelic haplotype was associated with a lower likelihood of SP phenotypic resistance and higher likelihood of treatment success although the evidence for the latter effect was weak. There was increasing SP *in vivo* resistance and decreasing SP treatment success associated with increasing *dhfr* IRN triple mutant
frequencies. In the model with \textit{dhps} IRN and region as covariates there was strong
evidence to suggest that any SP phenotypic resistance was lower in the Western region
compared to the Eastern region. This finding is not explained by differences in \textit{dhfr} IRN
frequencies and might be due to differences in \textit{dhps} allelic haplotype frequencies or to
more sites with high levels of acquired immunity in West Africa.

Notably the population frequency of the fully sensitive 51N/59C/108S allelic haplotype is
not a mirror image of the \textit{dhfr} 51I/59R/108N triple mutant frequency as mentioned in
section 4.5, chapter 4. This is influenced by the \textit{dhfr} double mutant haplotypes 51I/108N
and 59R/108N which displaced the 51N/59C/108S sensitive allelic haplotype in many
populations prior to the appearance of the \textit{dhfr} 51I/59R/108N triple mutant. The ‘other’ \textit{dhfr}
alleles i.e. neither 51N/59C/108S nor 51I/59R/108N but which are partially resistant alleles
constitute a considerable proportion of the total number tested. The observation of weak
evidence for increasing SP treatment efficacy along with higher \textit{dhfr} NCS frequencies
might perhaps be explained by the presence of intermediate resistance alleles and/or to
the additional factor which lies with \textit{dhps} and its sulphadoxine sensitivity. There are few
minor \textit{dhps} alleles and the frequency of \textit{dhps} GE/GK/AK generally sums up to the total
number tested. Thus these \textit{dhps} frequencies are more likely to reciprocate each other.

The \textit{dhps} sensitive AK allelic haplotype was associated with a lower likelihood of high
grade SP phenotype resistance and any phenotype resistance. Furthermore there was
strong evidence of a higher likelihood of high grade phenotype resistance and any
phenotype resistance corresponding with increasing \textit{dhps} GE frequencies throughout the
analyses. Coinciding with this observation, there was strong evidence of a lower likelihood
of treatment success with increasing \textit{dhps} GE frequencies.

There was borderline evidence for a higher likelihood of SP phenotypic resistance
associated with increasing \textit{dhps} GK frequencies only after adjusting for regional effects.
This suggests some evidence of a positive association between SP phenotypic resistance
and \textit{dhps} GK frequencies. Furthermore, after adjusting for differences in \textit{dhps} GK
frequencies, there was evidence of regional differences in SP phenotypic resistance
between East and West Africa. There was a lower likelihood of SP phenotypic resistance
in the West compared to the East, which might be explained by differential drug pressure
in this region.
The work presented in chapter 4 showed that different *dhps* haplotypes prevail in East Africa and in West Africa. High *dhps* GE frequencies occur in East Africa and low frequencies occur in West Africa. The *dhps* GK allelic haplotype predominates among West African sites. However, evidence of any association between the *dhps* 437G point mutation and SP treatment efficacy has been inconsistent. Previous work in four West African countries found the *dhps* 437G point mutation to be associated with *in vivo* treatment failure namely Lambarene, Gabon (1995-1996) (Kun et al. 1999), Tamale, Ghana (2002) (Mockenhaupt et al. 2005), Farefenni, Gambia (2001) (Dunyo et al. 2006), Brazzaville, Republic of Congo (2003-2004) (Ndounga et al. 2007). However, Marks et al. (2005) found no association between the *dhps* 437G point mutation and *in vivo* outcomes in the hyperendemic malaria area of Bodomase, Ashanti Region, Ghana when surveyed in 2001 and 2003. As mentioned in chapter 4 there are differences in the use of SP in East and West Africa, with large-scale use of SP as first line monotherapy in East Africa and relatively low SP use in West Africa ((WHO, 2001, Frosch et al., 2011) due to the different malaria transmission intensities among these regions.

**Predicted *in vivo* failure rates based on *dhps* allelic haplotypes**

The matched *dhps* allelic haplotype frequencies and *in vivo* prevalence showed that *dhps* GE raised the likelihood of phenotypic resistance and lowered the likelihood of treatment success. The same trend was observed when *dhps* GE frequencies were added together with *dhps* GK frequencies as well as after adjusting for *dhps* GK and the evidence for this was strong. The higher likelihood of phenotypic resistance with increasing *dhps* GE was consistent after adjusting for region. Using the univariate logistic regression results for *dhps* GE from Table 5.15, the predicted *in vivo* resistance rates were extrapolated using the model’s regression coefficients (Table 5.20) as in the equation:

\[
\text{Predicted } \text{*in vivo* } R = \frac{(exp(-1.687951+(0.0152146 \times \text{dhps GE frequency}))}{(1+exp(-1.687951+(0.015214 \times \text{dhps GE frequency}))}
\]

The *dhps* GE frequencies were unadjusted in these estimates. At 50% *dhps* GE, the predicted phenotypic resistance was 28.3% and at 100% *dhps* GE the predicted phenotypic resistance was 45.8%. In chapter 6, these predicted *in vivo* prevalence measures were mapped and linked to the WHO (2005) defined categories for assessing clinical failures leading to a recommendation for drug policy change in high transmission areas. In areas saturated with *dhps* GE, the phenotypic resistance is also likely to be influenced by other factors which impact on treatment failure. This analysis represents the first approximation of the relationship between *dhps* GE and any SP *in vivo* resistance.
Table 5.20 Predicted probability of any phenotype resistance based on varying *dhps* GE frequencies

<table>
<thead>
<tr>
<th>Frequency <em>dhps</em> GE haplotype (%)</th>
<th>Predicted prevalence (%) of any grade phenotypic resistance (in vivo R) with 95% confidence intervals</th>
</tr>
</thead>
<tbody>
<tr>
<td>0%</td>
<td>15.6 [8.5-22.7]</td>
</tr>
<tr>
<td>10%</td>
<td>17.7 [10.8-24.6]</td>
</tr>
<tr>
<td>25%</td>
<td>21.3 [14.5-28.0]</td>
</tr>
<tr>
<td>50%</td>
<td>28.3 [20.7-36.0]</td>
</tr>
<tr>
<td>75%</td>
<td>36.7 [25.6-47.7]</td>
</tr>
<tr>
<td>100%</td>
<td>45.8 [29.9-61.8]</td>
</tr>
</tbody>
</table>

The re-calibration of the *in vivo* studies enabled comparison with the *dhfr* and *dhps* allelic haplotypes in an ecological analysis and the results correlating these two phenotypic and genotypic measures of resistance suggests that there is value in such a categorisation. This study illustrates that population measures of molecular markers can be used to predict *in vivo* efficacy at a population level.

**Limitations**

The *in vivo* efficacy data are summarized estimates rather than individual patient level estimates and limitations relating to such data, mentioned previously in chapter 2, also apply here. Furthermore, the complete dataset was reduced after matching the allelic haplotypes with SP *in vivo* efficacy based on nearest neighbour and study year. The two regional groupings used in this analysis cover 19 countries including an Island, with different malaria transmission intensities and these differences might have been too large to allow distinct regional patterns to emerge.

Being young of age is a risk factor for SP treatment failure. This study used any age group due to the paucity of suitable *in vivo* studies available to match with the molecular studies. Age group seems to be less important once all five *dhfr* and *dhps* mutations occur together and in this scenario the risk of treatment failure is the same regardless of age group.
(Talisuna et al., 2004b). This study did not take into account the immunity status of study participants and transmission intensity in the sites studied, which are important confounders and could be factored into future models.

**Recommendation**

Here the first attempt to quantify the differential effects of alternative *dhps* GE and *dhps* GK resistance alleles on SP phenotypic outcome is described. This study shows strong evidence that the *dhps* GE double mutant is consistently associated with a higher likelihood of SP phenotypic resistance. In addition there was some evidence of a positive association between SP phenotypic resistance and *dhps* GK frequencies. The threshold of 50% prevalence of *dhps* 540E is already being used as a cut-off for SP in IPTi (WHO, 2009a). These findings support the use of GE as a marker for SP resistance and indicate that continued monitoring of *dhps* GE and *dhps* GK in Africa should be informative for use of SP for IPT in specific regions.
Chapter 6 Summary and Discussion

The research conducted in this thesis aimed to summarise the pan-African geography of SP resistance using aggregate data and to apply analytical methods to characterise its emergence and rate of spread. This concluding chapter contains a summary and discussion of the major findings. It highlights the main recommendations and limitations of this work, considers possibilities for future work and outlines its implications for the broader context of artemisinin resistance containment.

Overview of major findings

After outlining the history of antimalarial drug use in Africa and exploring the various factors known to influence the emergence and spread of resistance, the scope of the research project was outlined in chapter 1. The goal of the project was to gather molecular and *in vivo* SP resistance data from the published literature to create a georeferenced and standardised database of SP resistance measures. These data could then be analysed to explore spatial and temporal patterns across sub-Saharan Africa.

The scale and coverage of the georeferenced measures of resistance which were abstracted from the literature were described in chapter 2. The mapped distributions of SP *in vivo* efficacy studies and molecular surveys for a range of different point mutations and allelic haplotypes related to SP resistance were shown. The WHO (1973) and WHO (1996) protocols were used most frequently to assess SP *in vivo* efficacy. The most commonly used method to assess the prevalence of resistance markers was PCR-RFLP followed by sequencing and dot blotting. Consistent gaps in data coverage for both *in vivo* and molecular measures of SP resistance were identified in seven countries namely Botswana, Burundi, Cape Verde, Eritrea, Somalia, Togo and Mauritius.

The mapped sub-Saharan distribution of seven individual point mutations and their prevalence in every suitable molecular survey were examined in chapter 3. Exploratory analyses were performed on prevalence data for each point mutation. To assess whether or not there were any observable trends in prevalence over time, surveys were grouped into specific geographic regions. These regions had been previously defined according to whether or not *P. falciparum* parasites from those countries shared drug resistance mutations with a common origin. The observations made from the mapping and regional analysis of temporal changes were used to categorise point mutations into three broad
categories reflecting the pattern in which they occur: (1) scarce and/or not associated with foci of SP resistance (e.g. \textit{dhps} 436A and \textit{dhps} 613S), (2) emerging mutations (e.g. \textit{dhfr} 164L and \textit{dhps} 581G) which are comparatively rare but increasing in specific geographical foci of resistance (3) major resistance mutations such as \textit{dhps} 540E and \textit{dhps} 437G which are widespread.

The first group describes point mutations that scarce or rare. The \textit{dhps} 436A mutation occurred in 32 countries and 107 unique sites covering all six regions. There was a lower likelihood of \textit{dhps} 436A occurring over time. The minimal prevalence of \textit{dhps} 436A in the Southeast suggests that it may be an ancestral variant. The likelihood of \textit{dhps} 436A occurring in the West, Southwest and Central regions were higher than that of the Southeast.

Also in this group is the \textit{dhps} 613S point mutation. It was found in 10 countries but was absent in the Southwest region, suggesting that its distribution is sporadic in the other regions. The cumulative prevalence of \textit{dhps} 613T in Africa was the lowest of all mutations at 0.1% (11/10152) between 1988 and 2008 and it was present in three countries (in the Southeast, West and Island regions) and was absent in 21 countries. Generally, there was a lower likelihood of \textit{dhps} 613S/T occurring over time, with the highest likelihoods in the West.

The second group consists of point mutations with emerging foci of resistance, \textit{dhfr} 164L and \textit{dhps} 581G. The \textit{dhfr} 164L point mutation was absent in 21 countries, occurring only in the Southeast, Central and Island regions and being absent in the Northeast, West and Southwest regions. Its cumulative continental prevalence was 0.7% (130/19923) between 1998 and 2008. Thus \textit{dhfr} 164L is still relatively rare in sub-Saharan Africa. Notably there was a higher likelihood of \textit{dhfr} 164L occurring over time with the highest likelihoods among the Island populations.

The \textit{dhps} 581G point mutation is emerging sporadically and was found in 13 countries, was absent in 15 countries between 1988 and 2008 with a cumulative prevalence of only 5.1% (14526/15398) between 1993 and 2008. Generally the prevalence of \textit{dhps} 581G was low across sub-Saharan Africa with emerging foci in Mali and Tanzania. Generally, there was a higher likelihood of \textit{dhps} 581G occurring over time, with the highest likelihoods in the Southeast and Northeast regions.
The third group includes the widespread \textit{dhps} 437G and 540E point mutations which confer major resistance where they occur together as the \textit{dhps} GE double mutant. The prevalence of the \textit{dhps} 437G and 540E point mutations increased over time but there were noteworthy regional differences in their prevalence. High prevalence of \textit{dhps} 437G occurred in all 37 countries and 152 unique sites surveyed.

The \textit{dhps} K540E had the best data coverage of all the point mutations studied and it occurred in 27 countries and was absent in 10 countries. Generally high \textit{dhps} K540E prevalence was observed in East Africa with the highest prevalence recorded in the Southeast and Northeast regions, where it was $\geq 50\%$ at 74 sites. The prevalence was consistently low (<50\%) in the West, Southwest, Central and Island regions over time.

The results of the logistic regression models for \textit{dhps} 540E and \textit{dhps} 437G suggested that there was a steady increase in the prevalence of both point mutations over time but that there were regional differences in prevalence. These point mutations were found predominantly in the Southeast regions and were less prevalent in the Western regions, where a declining prevalence trend was predicted. The predicted probabilities of \textit{dhps} 437G and 540E in the Western and Southwestern regions were generally less than that found in the Southeast region. For \textit{dhps} 437G the Southern African countries were similar to each other and different from the Western, Central and Northeast regions. The \textit{dhps} 437G point mutation generally occurred together with \textit{dhps} 540E in the Southeast and Northeast region.

Summarised frequency distributions of the \textit{dhfr} sensitive allelic haplotype (51N/59C/108S), \textit{dhfr} triple mutant allelic haplotype (51I/59R/108N) and the three \textit{dhps} haplotype alleles 437G/540E (GE), 437G/540K (GK) and 437A/540K (AK) were presented graphically and mapped in chapter 4.

High frequencies of the \textit{dhfr} NCS sensitive haplotype were found scattered across East and West Africa but cumulatively there was a gradual decline in its frequency over time in all regions. The \textit{dhfr} IRN allelic haplotype generally increased over time with strong evidence for this. However, there was only weak evidence of an overall regional effect in the frequency of \textit{dhfr} IRN. As \textit{dhfr} IRN has evidently dispersed and reached fixation in many sites throughout \textit{Africa}, regional differences are less apparent. The regional
distribution of \textit{dhfr} NCS does not mirror that of \textit{dhfr} IRN and this is because of the occurrence of the \textit{dhfr} 51I/108N or 59R/108N double mutants.

There were regional differences in the African distribution of the \textit{dhps} haplotypes. The maps and regression models showed clear differences in the distribution of the \textit{dhps} GE double mutant, GK single mutant and AK fully sensitive allelic haplotypes with some distinct regional patterns discernible. For \textit{dhps} GE, there appears to be an East Africa-West Africa divide and high frequencies occur in the Eastern countries and low frequencies occur in West Africa. The distribution of \textit{dhps} GK mirrors that of \textit{dhps} GE with a West-East divide. High \textit{dhps} GK frequencies occur primarily in West Africa and low frequencies occur in East Africa. Generally, there were higher likelihoods of \textit{dhps} GE and \textit{dhps} GK and correspondingly lower likelihoods of \textit{dhps} AK per year.

In chapter 5 the relationship between these molecular markers and \textit{in vivo} efficacy was explored at a population level using a standardised categorisation of summarised antimalarial \textit{in vivo} efficacy data and matching molecular surveys from similar times and places. The SP \textit{in vivo} studies selected for inclusion were matched with \textit{dhfr} and \textit{dhps} allelic haplotype studies based on nearest geographic location and study year and an assessment of their association was conducted. The influence of East African and West African genotypes on clinical outcome could then be explored.

Given the broad distribution of \textit{dhfr} IRN it is expected that it should be a good predictor of \textit{in vivo} resistance. Where the frequency of \textit{dhfr} IRN is high, treatment failures and resistance rates should be high, whilst the treatment success rates should be low. This was consistent with what was observed in this study as there was a higher likelihood of SP \textit{in vivo} resistance as the \textit{dhfr} IRN triple mutant frequencies increased. This finding supports the results obtained by Picot \textit{et al.}, (2009) who conducted a meta-analysis and found that the presence of the \textit{dhfr} IRN triple mutant increased the risk of SP treatment failure. Whilst there was a higher likelihood of any phenotypic resistance in the East compared to the West, this was not consistent when the interaction between region and the \textit{dhfr} IRN triple mutant was factored in the model. It is also expected that \textit{dhfr} NCS would be a good predictor of \textit{in vivo} efficacy, where its frequency is high the treatment success rate should also be high. Generally, there was a lower likelihood of phenotypic resistance and higher likelihood of treatment success with increasing \textit{dhfr} NCS frequencies but the evidence for this was inconsistent. This inconsistency may be due to the observed high frequencies of \textit{dhfr} double mutant allelic haplotypes in East Africa.
For *dhps* AK to be a good predictor of *in vivo* efficacy, its frequencies should be positively correlated with treatment success and negatively correlated with resistance rates. Indeed the likelihood of any *in vivo* resistance was lower as *dhps* AK frequencies increased. Similarly the likelihood of treatment success was higher with increasing *dhps* AK frequencies.

Since *dhps* GE and GK are resistance mutants it is expected that their frequency in populations would be good predictors of *in vivo* efficacy. Their frequencies are expected to be negatively correlated with treatment success and positively correlated with resistance rates. There was strong evidence for this with *dhps* GE but weak evidence for this with *dhps* GK when these genotypes were modelled together as covariates in a logistic regression analysis. The evidence for any regional differences upon phenotypic outcome was weak after adjusting for *dhps* GE. There was a lower likelihood of any phenotypic resistance in the West compared to the East when region was modelled with *dhps* GK, but this was not consistent when the interaction between region and *dhps* GK was factored in the model. The likelihood of any SP *in vivo* resistance on its own was higher in the East compared to the West, but there was weak evidence for this, suggesting there were no regional differences in phenotypic outcome.

**Discussion of major findings and recommendations**

Since the first protocols for measuring antimalarial drug resistance were developed, new tools such as *in vitro* tests and molecular methods have been developed. The WHO definition of antimalarial drug resistance described in chapter 1 can be simplified as an infection that survives, multiplies and gets carried over to the next host despite being treated (Sibley and Price, 2012). This definition is based on clinical responses of the patient and does not consider other measures of resistance. A combination of the results of clinical, molecular, pharmacological and *in vitro* tests to characterise the definition of resistance to SP would be useful (Sibley and Price, 2012). As Basco and Ringwald (2000) stated, the original definition of antimalarial resistance is inadequate and ought to factor in treatment failures, 50% inhibitory concentration of the drug as well as mutations associated with resistance. A revised definition of drug resistance would be timely in the context of SP being used for IPT strategies and artemisinin resistance containment.

As the use of molecular methods to detect resistance gains momentum and the WHO recommendations now include a molecular prevalence threshold for the decision on IPTi
implementation, there is a need for a standardised framework for measuring and equally importantly, reporting of molecular data in a systematic manner. This thesis described the geographic and temporal changes using measures of the prevalence of single point mutations and the frequency of allelic haplotypes. Among the genotyping studies gathered in this study, prevalence was the more commonly reported measure compared to frequencies. To carry out an analysis on the combinations of point mutations called ‘allelic haplotypes,’ reporting of mutation data needs to be carried out in such a way that haplotypes can be identified and mixed infections should be documented and clearly reported so that these can be handled in a standardised way.

Ideally patient level data would be reported but it is not always practical to do so and wherever longitudinal studies are carried out there should be adherence to standardised study designs, data collection methods, data analyses and reporting standards. With online infrastructure becoming increasingly available, data sharing is far more practical now than it was a few decades ago.

Recognising the need for support structures to collect and standardised data the WWARN partnership offers support for designing protocols and has set up regional centres for data collection, providing tools for partners to collect, clean and analyse their data and then standardise the data according to a predefined data management and statistical analysis plan. These tools are available online for easy access at http://www.wwarn.org/partnerships/study-groups/asaqal-molecular-marker-study-group. To date WWARN has obtained SP clinical resistance data from Angola, Chad, DRC, Mali, Mozambique, Sierra Leone, Uganda and South Africa but again efficacy rates are available in summarised estimates at http://www.wwarn.org/resistance/explorer. The data gathered in this thesis have been made available online at www.drugresistancemaps.org where the reference for each study is provided together with the study year, number of samples tested for a particular point mutation and number of samples tested positive. This website of African surveillance data originated a global online database (http://www.wwarn.org/resistance/surveyors/molecular) where these African data are combined with georeferenced molecular data from sites across the globe.

The data gaps identified in the surveillance coverage of in vivo and molecular markers of SP resistance in this study, point to a shortage in the West African countries of Chad, Sierra Leone and Togo as well as Somalia in the horn of Africa and Burundi in East Africa. As stated previously, these gaps could represent a genuine absence of data or simply
point to a lack of availability of such data in the public domain. Where SP is no longer used as first line treatment, molecular methods are the only practical means of quantifying resistance. Systematic surveillance for establishing molecular levels of SP resistance using longitudinal studies is especially relevant where these gaps exist, as IPT programmes are rolled out. Surveillance should include symptomatic and asymptomatic infections. Asymptomatic infections act as a reservoir and active surveillance of asymptomatic infection is more likely to reflect the true biomass of circulating parasites in the community (Malisa et al., 2010). Sampling only at health facilities where participants are symptomatic is likely to overestimate resistance levels since patients may have had prior treatment, which would have selected for resistant parasites (Malisa et al., 2010). Alker et al., (2008) suggested the collection of some clinical data to support the use of molecular markers when monitoring drug resistance. These clinical measures could include parasite density among study participants who test positive for P. falciparum.

The results presented in this thesis provided evidence for important differences in the genetic basis of SP resistance between West Africa and East Africa. High dhps GE frequencies occur in East Africa and not in West Africa. However, there was no evidence of regional differences associated with phenotypic outcome after adjusting for dhps GE using matched studies for each of these measures. There was consistent evidence that dhps GE was negatively correlated with treatment success and positively correlated with SP resistance.

To examine the implications of this for SP efficacy and monitoring dhps GE in the present day a subset of surveys carried out since 2004 were selected to extrapolate SP phenotypic outcome from dhps GE. The predicted in vivo efficacy for each survey site was calculated using the equation described in section 5.6 of chapter 5. These were used to explore the contemporary in vivo treatment failure rates in all sites where recent measures of dhps GE had been recorded.

As indicated in chapter 5 the WHO (2005) defined four milestones for clinical failures leading to a change in drug policy in high transmission areas: 0 to 4% treatment failures indicated a grace period, 5 to 14% failures indicated an alert period, 15 to 24% treatment failures indicated the action period and any treatment failures ≥25% required a change in the drug policy. The predicted failure rates are mapped in Figure 6.1, showing sites that had dhps GE frequencies between 15%-24% and 25%-100%. Notably no sites fell into the
0-4% and 5%-14% treatment failure categories, which indicate the grace and alert periods respectively, for monitoring SP efficacy (WHO, 2005).

Figure 6.1 Map of predicted SP failure rates (%) between 2004 and 2008 according to recent measures of dhps GE frequencies. The brown circles indicate the action phase and red circles indicate the need for a treatment policy change based on WHO (2005) recommendations.

Twenty two countries had predicted in vivo treatment failures between 15%-24% which were defined as the action period for monitoring efficacy (WHO, 2005). These were the Southwest (Angola, Democratic Republic of Congo (west), Gabon, Namibia, Republic of Congo), West (Benin, Burkina Faso, Gambia, Ghana, Guinea, Ivory Coast, Niger, Nigeria, Senegal) Central (Cameroon, Central African Republic), Southeast (Mozambique, Swaziland, Tanzania) and the Islands (Comoros, Madagascar, Sao Tome/Principe).

Higher SP failure rates of ≥25% were predicted by dhps GE in nine East African countries only. These were Democratic Republic of Congo (east), Ethiopia, Kenya, Malawi,
Mozambique, Rwanda, Tanzania, Uganda and Zambia (Figure 6.1). These sites would require a change in treatment policy (WHO, 2005). Based on the WHO (2005) grading system and predicted in vivo treatment failures using the \textit{dhps} GE double mutant, more sites in East Africa required a change in drug policy from SP to ACTs.

A clear theme in this work is that there is regional variability in the pattern of \textit{dhps} genotypes in sub-Saharan Africa which provides evidence to support the view that there should be different drug policies for different regions, which would be driven by the local malaria epidemiology as described by WHO (2012) and illustrated in Figure 1.1. The WHO guidelines should specify methods for measuring molecular markers of resistance as these markers become available, to confirm treatment failures in its guidelines for monitoring drug efficacy. These should be used to guide global malaria control efforts and should also advocate for malaria control based on regional groupings of countries with similar malaria epidemiology. Talisuna \textit{et al.}, (2012) drew attention to the lack of systematic surveillance of antimalarial drug resistance in Africa since ACTs were adopted in Africa. The authors advocated for regional malaria surveillance activities in Africa and recommended that they be coordinated by a central body with a regional institutional centre so that regional ownership is maintained (Talisuna \textit{et al.}, 2012). These activities would include a strong emphasis on data quality and data standardisation.

While it is uncertain whether or not the pattern of spread of artemisinin resistance will follow that of chloroquine and SP resistance, the global plan for artemisinin resistance containment advocates customising the monitoring guidelines to regional requirements (WHO, 2011b). It then becomes necessary to define these regional groupings and there is utility in categorising malaria endemic countries in sub-Saharan Africa not only according to similar malaria transmission intensities but also according to antimalarial drug resistance profiles. Interestingly there is a link between the malaria surveillance networks in Africa and migration and antimalarial resistance dispersal patterns (Talisuna \textit{et al.}, 2012). Malaria migration as described in Tatem \textit{et al.}, (2010) and dispersal of drug resistance as described in Pearce \textit{et al.}, (2003) have similar regional patterns which are also reflected in the regional malaria surveillance networks. For example the participating countries in the East African Network for Monitoring Antimalarial Treatment (EANMAT) were Burundi, Kenya, Rwanda, Tanzania and Uganda. Pearce \textit{et al.}, (2003) found that the \textit{dhps} SGE1 haplotype was distinct to Tanzania, Kenya and Uganda whilst Tatem \textit{et al.}, (2010) showed that these five countries represent common migration communities.
Limitations

The scope of this thesis required a comprehensive dataset with maximal geographic coverage which could only be sourced from summarised estimates of molecular markers and SP in vivo efficacy studies in published studies. Thus this is an ecological study which uses aggregate measures to make inferences about groups of people rather than individuals (Morgenstern, 1995). The sourcing of data from primarily published studies introduces the risk of publication bias towards positive results as mentioned in Picot et al., 2009. However, the SP resistance data collected in this thesis did cover a 20 year period which provided sufficient evidence to draw conclusions about its historical distribution over time and space.

Although Basco and Ringwald (2000) recommended that standardised methods be applied when monitoring antimalarial drug resistance, this thesis demonstrated a wide variation in the application of WHO protocols for SP in vivo efficacy studies. This confirms that investigators often deviate from the prescribed inclusion and exclusion criteria based on their study rationale (Sibley and Price, 2012). Furthermore, as mentioned in chapter 2 it is likely that investigators will continue to use varying study designs (Price et al., 2007) and any attempts at consolidating the results of these studies in a pooled analyses will always have this limitation.

Given the variation in adherence to WHO protocols for in vivo efficacy testing as well as the issues related to reporting prevalence or frequencies of molecular markers, strict inclusion criteria for studies that used standardised follow-up periods, similar outcome measures and study populations would have limited the number of studies available for analyses in this thesis. Shah et al., (2011) also used summary estimates of in vivo chloroquine resistance and opted not to use a meta-analysis framework, choosing instead a systematic review process to select studies that would describe overall trends for chloroquine resistance dispersal in India. The criteria applied in the meta-analysis conducted by Picot et al., (2009) gathered studies over a 10 year period to correlate molecular markers and treatment efficacy but this effectively restricted their analyses over time and space. Sridaran et al., (2010) noted the challenge in comparing different molecular studies for dhfr and dhps in the absence of standardised methods for measuring and reporting different mutants and very few studies became eligible for inclusion in their meta-analysis. As discussed in chapter 2, a balance has to be sought between adhering to the criteria for systematic reviews and collating sufficient data that will allow meaningful analyses to answer specific research questions, in the absence of individual patient data.
In this study SP *in vivo* efficacy data were matched with molecular measures of resistance to assess if there was any correlation between these two measures or not. The method for matching these studies was based on nearest neighbours and closest study year. Ideally the treatment outcome of each patient together with individual patient molecular data should be directly measured at the time of sampling, alleviating the need to develop a method for matching them at a later stage. Picot *et al.*, (2009) provided a checklist for the design of future studies on molecular markers and advised that clinical and molecular measures be linked to each study participant.

Thus, with secondary data collection data quality and availability are important considerations which dictate the analytical plan. There are several reasons for the analytical approach used in this thesis: (1) the variation in *in vivo* efficacy study protocols and outcomes, (2) the variations related to prevalence and frequency of molecular markers and accounting for mixed infections (3) availability of point mutation data versus allelic haplotype data and (4) the data gap analysis showing countries where no resistance data were found, represents a result itself but these data gaps also presents challenges when making inferences about these data as they represent missing data. In addition to these data gaps some studies had to be excluded on the basis of insufficient details. For instance an important variable such as missing study years despite efforts to obtain them rendered some studies unusable.

The variability in the data available for each country is another limitation. Some countries such as those grouped together in the Southeast region had good coverage of data across the 20 year study period whereas other countries had limited data with gaps in the temporal sequence (see Tables 3.4a, 3.4b, 4.4 and 4.5). This variability of data among countries leads to confounding, which was mitigated by identifying each country as a cluster in the multivariable analysis. Ideally both ecologic studies and individual patient data would be used to develop models of SP resistance over space and time. Within these data constraints and with the aim of providing insights that would be useful for decision makers, a conservative analysis plan was developed using multivariable regression.

The basic elements of epidemiology are person, place, time (Moore and Carpenter, 1999) and this thesis mapped SP resistance rates but did not attempt to develop spatially explicit models predicting its distribution where no data exist nor did it forecast future trends over time. The ecological analyses are useful for generating new hypothesis for further work,
possibly using sophisticated statistical methods such as Bayesian analysis. These and other ideas for future work are discussed next.

**Future work**

Particular attention should be paid to monitor and gather data on emerging resistance markers such as \textit{dhfr} 164L and \textit{dhps} 581G. Monitoring the emergence of \textit{dhfr} 164L in new hotspots and tracking its spread across foci are important due to its association with pyrimethamine resistance when found together with the \textit{dhfr} triple mutant. When \textit{dhps} 581G is found with the \textit{dhps} GE double mutant and the \textit{dhfr} triple mutant, it is an important resistance marker. Although there is good coverage of \textit{dhps} 437G and 540E, additional longitudinal studies of these together with the quintuple mutant are required, given their predictive value for SP efficacy.

The spatial predictions using methods such as Bayesian models mentioned above would have resulted in smooth mapped outputs where every data gap would be filled. These spatial models would need to be structured to account for spatial uncertainties in the data such as autocorrelation and be validated using appropriate methodologies, as outlined by Brooker (2007). Furthermore uncertainties in the data and the models generated from them require careful consideration if the goal is to develop spatially explicit models that have practical significance for malaria control activities (Brooker, 2007).

In this study the primary goal was to compile the large literature of \textit{in vivo} and molecular measures of SP resistance in sub-Saharan Africa in a geographical and temporal framework. The duration of follow-up and use of PCR genotyping were used as selection criteria for extracting SPR phenotype studies which were then matched with genotype studies. Thus, phenotype and genotype data were not linked and needed to be matched using their nearest geographic location and study year. Two covariates representing time and geography were factored into the models to track SPR emergence and spread. Other data were also collected such as patient’s age, malaria endemicity, malaria seasonality and country drug policy. These additional data were collected with the possibility of using these as covariates for future work outside the scope of this thesis, once their coverage and consistency are assessed.
Selected factors contributing to the development of antimalarial drug resistance were discussed in chapter 1 and advanced models could incorporate the following covariates:

(1) population density in sub-Saharan Africa (2) drug use data as described by Frosch et al., (2011) who assessed the level of chloroquine use and chloroquine resistance. The question that could be posed is whether or not drug use accounts for the regional differences among dhps GE and GK allelic haplotypes, (3) age as a surrogate for acquired immunity and a risk factor for treatment failure, (4) transmission intensity reflecting the parasite biomass in communities, (5) treatment failures attributed to HIV infection rather than antimalarial drug resistance and the use of trimethoprim sulphamethoxazole which has a similar mode of action to SP (Laufer and Plowe, 2004) and (6) migration communities that are linked by high levels of malaria infection movements e.g. Nigeria which represent a problem for containing the spread of resistance (Tatem and Smith, 2010; Lynch and Roper, 2011).

In Chapter 2, the different WHO protocols for collecting clinical data were described and children under five years of age were highlighted as a vulnerable group for unsatisfactory treatment response, even in areas with acquired immunity (WHO, 2002). Whilst young age is a risk factor for SP phenotypic failure, age seems to be less important in areas of high genotypic resistance for example where the five dhfr and dhps mutations occur together (Talisuna et al., 2004b, Staedke et al., 2004). Against this high genotypic resistance backdrop the phenotypic failure risk is the same regardless of age group (Talisuna et al., 2004b) and a possible explanation for this observation is that with increasing parasite resistance, the relative role of host immunity becomes less significant. Furthermore, few studies have directly accounted for age in genotyping studies (Staedke et al., 2004). Complete data linking phenotypic outcome with genotypic outcome, stratified by age would be valuable in assessing age as an explanatory variable for the spread of antimalarial drug resistance in future models. Age could be accounted for using an age structured population based model as in Pongtavornpinyo et al., (2008).

Similarly transmission intensity as discussed in chapter 1, could be factored in as a covariate using the transmission intensity model described in Gething et al., (2011). Other workers have modelled different severities of P. falciparum phenotypic malaria outcomes across all transmission intensities in sub-Saharan Africa (Carneiro et al., 2010). Carneiro et al., (2010) found that younger age groups were associated higher malaria burdens in areas of increasing transmission intensity.
The link between human movement and transport of antimalarial resistance warrants some discussion here as it was mentioned previously in chapters 1, 3 and 4.

Migration plays a central role in the regional dispersal of drug resistant malaria. Circulating SP drug resistant lineages reflects communities that are connected by high levels of migration (Lynch and Roper, 2011). As shown in this thesis, resistance alleles circulate regionally and the qualitative characteristics of resistance are regionally similar whilst quantitative variability in resistance levels can occur, presumably because of variable drug pressure within each region. Some patterns are evident when migration communities are mapped together with resistance markers (Figure 6.2).

The distribution of the \textit{dhps} GE double mutant indicates significant migration among countries in Southeast Africa including Malawi, Tanzania, Mozambique Swaziland and South Africa and separately among the Southwestern countries of Namibia and Angola (Figure 6.2). The path of resistance dispersal as reflected by migration communities has significant implications for the dispersal of artemisinin resistance and the tiered approach recommended for different regions (continental) by WHO for resistance containment (WHO, 2011b). There is a need to establish guidelines for best practices for malaria control along border areas in migrant populations (Lynch and Roper, 2011) as is the case in Asia (Shah \textit{et al.}, 2011).

The influence of human migration on the transport of antimalarial drug resistance between island populations and the mainland is worth investigating. Notably the island populations are different with different drug resistance markers occurring there. These differences are apparent even among islands that are situated in close proximity to each other (Menard \textit{et al.}, 2007). In Madagascar the \textit{dhfr} 164L point mutation was found among a unique combination of point mutations that have not been reported in mainland Africa to date (Andriantsoanirina \textit{et al.}, 2009, Naidoo and Roper, 2011). The \textit{dhfr} IRN triple mutant decreased among the island populations, although it was observed to have reached fixation in mainland Africa. The \textit{dhps} point mutations have evolved independently among the Island regions. The presence of \textit{dhfr} and \textit{dhps} point mutations among the islands albeit at low prevalence, illustrates the role that migration must play in resistance dispersal rather than mosquito mediated dispersal. The latter is unlikely given the distance a mosquito would need to travel between the islands and the mainlands (Salgueiro \textit{et al.}, 2010). Furthermore migration of people between the islands and mainlands is less frequent than the movement among the regions and resistance dispersal is consequently
stochastic (Naidoo and Roper, 2011). Further work is required to assess the contribution of migration to the regional dispersal of drug resistance lineages.

Figure 6.2 Migration communities in sub-Saharan Africa with dhps allelic haplotypes. The migration communities are taken from Tatem and Smith (2010).

Surveillance of SP for IPT should be carried out routinely and these data assessed to see if there are any shifts in the regional patterns observed in this thesis, particularly where there are gaps in the data and in regions where there is no clear trend in the prevalence of resistance. According to Cairns et al., (2012) the high levels of SP resistance in Southern and Eastern Africa as demonstrated in Naidoo and Roper (2012) provided evidence that a replacement for the SP plus amodiaquine combination is needed for SMC in these regions. The SP plus amodiaquine combination remains suitable for SMC in West Africa but requires close monitoring. Molecular markers such the dhps GE double mutant and
dhps GK single mutant could be useful indicators of waning SP efficacy for IPT, but the thresholds of these allelic haplotypes related to IPT efficacy needs to be assessed.

As Sibley and Price (2012) point out, once resistance foci are identified the main questions are whether or not it has arisen or spread to other hotspots or foci and what the routes of resistance dispersal are. Talisuna et al., (2012) also stated that empirical analyses of risk factors associated with the spatial emergence of resistance to chloroquine and SP would be useful.

This thesis attempted to answer these questions for SP resistance using retrospective data and several issues bear reference when reflecting on artemisinin resistance containment.

Efforts to collate artemisinin efficacy data from individual patients to pool them for analyses at different geographic scales should be supported by national malaria control programmes. Myint et al., (2004) observed an improvement in the quality of antimalarial drug trials in recent years which gave rise to better data on the newer artemisinin derivates compared to the data available for established drugs such as SP. Adherence to standardised methods of measuring tolerance/resistance as well as reporting results should continue to be encouraged through platforms like the WWARN initiative. Good quality spatially referenced antimalarial efficacy data are crucial especially if they are to be used to establish baseline levels of artemisinin resistance to inform decisions at small scales such as village or district levels, to identify hot spots and foci of resistance and to delineate buffer zones for artemisinin resistance containment especially at border posts and at the fringes of endemic malaria areas. These data can also be used to track the emergence and spread of artemisinin resistance over time across both small (village and district level) and large (country, regional) geographic scales. Such data can then be used to establish if there are any distinct regional patterns related to SP resistance and to support the recommended regional efforts to contain resistance dispersal.
References


GEOnet names server (GNS) (1994) National Geospatial Intelligence Agency, USA.


Mapinfo Professional version 9.5. Pitney Bowes Software Incorporated, USA.


Stata Intercooled Version 12.0. StataCorp LP, College Station, USA.


Appendix 1 Proforma

SECTION 1: DATA REFERENCE

ADMINISTRATIVE INFORMATION

<table>
<thead>
<tr>
<th>Date data transcribed (dd/mm/yyyy)</th>
<th>-</th>
<th>-</th>
</tr>
</thead>
<tbody>
<tr>
<td>Data transcribed</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Data checked at co-ordinating centre</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

DATA REFERENCE

Category of document

<table>
<thead>
<tr>
<th>Journal</th>
<th>MOH report</th>
</tr>
</thead>
<tbody>
<tr>
<td>Book</td>
<td>Other report</td>
</tr>
<tr>
<td>Thesis</td>
<td>Personal communication</td>
</tr>
</tbody>
</table>

Language

Reference source

<table>
<thead>
<tr>
<th>Authors</th>
</tr>
</thead>
</table>

Year of publication

<table>
<thead>
<tr>
<th>Title</th>
</tr>
</thead>
</table>

<table>
<thead>
<tr>
<th>Journal</th>
<th>Non-journal</th>
</tr>
</thead>
<tbody>
<tr>
<td>Volume</td>
<td>Issue no.</td>
</tr>
<tr>
<td>Pages from</td>
<td>to</td>
</tr>
</tbody>
</table>

How many data points are there in this report?
SECTION 2: SURVEY LOCATION

2.1 Geographic location of survey

<table>
<thead>
<tr>
<th>Name</th>
<th>Code</th>
</tr>
</thead>
<tbody>
<tr>
<td>Country</td>
<td></td>
</tr>
<tr>
<td>First level administrative unit (e.g. province)</td>
<td></td>
</tr>
<tr>
<td>Second level administrative unit (e.g district)</td>
<td></td>
</tr>
<tr>
<td>Third level (e.g. area, division)</td>
<td></td>
</tr>
<tr>
<td>Fourth level (e.g. location)</td>
<td></td>
</tr>
<tr>
<td>Nearest large town</td>
<td></td>
</tr>
<tr>
<td>Community level name (village, settlement)</td>
<td></td>
</tr>
</tbody>
</table>

2.2 Survey location georeference

N.B.  Latitude refers to North (+) or South (-)

Longitude refers to East (+) or West (-)

Decimal degrees:

<table>
<thead>
<tr>
<th>Point</th>
<th>Point name</th>
<th>Latitude</th>
<th>+/− N/S</th>
<th>Longitude</th>
<th>+/− E/W</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td></td>
<td>.</td>
<td></td>
<td>.</td>
<td></td>
</tr>
<tr>
<td>2</td>
<td></td>
<td>.</td>
<td></td>
<td>.</td>
<td></td>
</tr>
<tr>
<td>3</td>
<td></td>
<td>.</td>
<td></td>
<td>.</td>
<td></td>
</tr>
<tr>
<td>4</td>
<td></td>
<td>.</td>
<td></td>
<td>.</td>
<td></td>
</tr>
</tbody>
</table>

or,

Degrees and decimal minutes:

<table>
<thead>
<tr>
<th>Point</th>
<th>Point name</th>
<th>Latitude</th>
<th>Longitude</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td>Deg</td>
<td>Minutes</td>
</tr>
<tr>
<td>1</td>
<td></td>
<td>.</td>
<td>.</td>
</tr>
<tr>
<td>2</td>
<td></td>
<td>.</td>
<td>.</td>
</tr>
<tr>
<td>3</td>
<td></td>
<td>.</td>
<td>.</td>
</tr>
<tr>
<td>4</td>
<td></td>
<td>.</td>
<td>.</td>
</tr>
</tbody>
</table>

or,
Degrees, minutes and seconds:

<table>
<thead>
<tr>
<th>Point</th>
<th>Point name</th>
<th>Latitude</th>
<th>Longitude</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td>Deg</td>
<td>Min</td>
</tr>
<tr>
<td>1</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>2</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>3</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>4</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Comments:

Is the survey location a:

<table>
<thead>
<tr>
<th>Point</th>
<th>Polygon</th>
<th>Multiple points (with pooled results)</th>
<th>Administrative unit</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

2.4 How did you georeference?

<table>
<thead>
<tr>
<th>Source</th>
<th>Method</th>
</tr>
</thead>
<tbody>
<tr>
<td>Africa Data Sampler</td>
<td>Scanned topographic map</td>
</tr>
<tr>
<td>Africa Data Sampler</td>
<td>Topographic map, manually</td>
</tr>
<tr>
<td>Census</td>
<td>MARA Prevalence</td>
</tr>
<tr>
<td>Geoname</td>
<td>Google Earth</td>
</tr>
<tr>
<td>Geonet</td>
<td>GPS</td>
</tr>
<tr>
<td>Health Map</td>
<td>Encarta</td>
</tr>
<tr>
<td>Other digital map</td>
<td>Other (specify):</td>
</tr>
<tr>
<td>Co-ordinates given in report</td>
<td></td>
</tr>
</tbody>
</table>

If a topographic map was used, indicate its scale (e.g. 1: 5 000)

2.5 Climatological and geographical data

Is a graph, map or table available (of area, rainfall, temperature, disease profile, seasonal abundance of infection or vectors, etc.)?

<table>
<thead>
<tr>
<th>Yes</th>
<th>No</th>
</tr>
</thead>
</table>

269
2.6 General information on the survey area:

<table>
<thead>
<tr>
<th>Is malaria transmission in the area:</th>
<th>Yes</th>
<th>No</th>
<th>Not stated</th>
</tr>
</thead>
<tbody>
<tr>
<td>Epidemic/unstable/highland</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Endemic/stable</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Seasonal</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Are the population studied refugees or displaced persons?</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Is rice grown in the area?</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Is there an irrigation scheme?</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Is there a swamp/wetlands/dam etc. close by?</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Is the area:</th>
<th>Urban</th>
<th>Peri-urban</th>
<th>Rural</th>
<th>Not stated</th>
</tr>
</thead>
</table>

<table>
<thead>
<tr>
<th>Estimated contemporary entomological inoculation rate</th>
<th>EIR</th>
<th>Year estimated</th>
<th>EIR Comments</th>
</tr>
</thead>
</table>

Information on country drug policy or drug resistance

Comments:

---

270
### ANTIMALARIAL DRUG RESISTANCE

#### 5.1 INSTITUTION RESPONSIBLE FOR REPORT:

<table>
<thead>
<tr>
<th>Ministry of Health</th>
<th>Non-government organization (NGO)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Government Med Research Inst</td>
<td>WHO</td>
</tr>
<tr>
<td>Other</td>
<td>University</td>
</tr>
</tbody>
</table>

If other, specify:

#### STUDY METHODOLOGY:

**Type of Report and Study Design:**

| Case report of treatment failure | Case report of prophylaxis failure |
| Case series of treatment failure | Case series of prophylaxis failure |
| Field trial of prophylaxis | *In vitro* efficacy study |
| *In vivo* efficacy study | Other |
| Intermittent preventive treatment (IPT) in pregnant women | |

**Study Design**

| Case series | Randomised |
| Comparative | Randomised controlled trial open label |
| Cross-sectional | Randomised controlled trial |
**Drugs studied in this report:**

<table>
<thead>
<tr>
<th>Drug name</th>
<th>Dosage (mg/kg)</th>
<th>Drug name</th>
<th>Dosage (mg/kg)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Chloroquine</td>
<td></td>
<td>Quinine</td>
<td></td>
</tr>
<tr>
<td>Sulphadoxine/Pyrimetamine</td>
<td></td>
<td>Mefloquine</td>
<td></td>
</tr>
<tr>
<td>Amodiaquine</td>
<td></td>
<td>Halofantrine</td>
<td></td>
</tr>
<tr>
<td>Artesunate</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Other drug or Drug</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Combination or Dosage</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>(specify name(s) and dosage)</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

If dosage by age group or weight, specify age range or weight and dose:

<table>
<thead>
<tr>
<th>All doses observed or given by study staff?</th>
<th>Yes</th>
<th>No</th>
<th>Not stated</th>
</tr>
</thead>
</table>

<table>
<thead>
<tr>
<th>Copy of report on file?</th>
<th>Yes</th>
<th>No</th>
</tr>
</thead>
</table>

<table>
<thead>
<tr>
<th>Study Methodology described?</th>
<th>Yes</th>
<th>No</th>
</tr>
</thead>
</table>

<table>
<thead>
<tr>
<th>Study Protocol provided?</th>
<th>Yes</th>
<th>No</th>
</tr>
</thead>
</table>

Follow-up days (check all days patients were to be seen - Day 0 = initial day of enrolment):

<table>
<thead>
<tr>
<th>Day</th>
<th>Day</th>
<th>Day</th>
<th>Day</th>
<th>Day</th>
<th>Day</th>
<th>Day</th>
</tr>
</thead>
<tbody>
<tr>
<td>0</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>1</td>
<td>8</td>
<td>15</td>
<td>22</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>2</td>
<td>9</td>
<td>16</td>
<td>23</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>3</td>
<td>10</td>
<td>17</td>
<td>24</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>4</td>
<td>11</td>
<td>18</td>
<td>25</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>5</td>
<td>12</td>
<td>19</td>
<td>26</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
If specific days for any week of follow-up not provided, mark below

<table>
<thead>
<tr>
<th>Week1</th>
<th>Week 2</th>
<th>Week 3</th>
<th>Week 4</th>
<th>Week 5</th>
<th>Week 6</th>
</tr>
</thead>
</table>

Definitions of treatment success and failure used:

WHO Protocol

|------------------------------------------------|------------|------------|------------|------------------------------------------|

Definition of Parasitologic Response:

<table>
<thead>
<tr>
<th>Standard (S)/RI/RII/RIII</th>
<th>Yes</th>
<th>No</th>
</tr>
</thead>
<tbody>
<tr>
<td>RI (late)</td>
<td>Yes</td>
<td>No</td>
</tr>
</tbody>
</table>

If yes, RI (late) defined as:

<table>
<thead>
<tr>
<th>RI (early)</th>
<th>Yes</th>
<th>No</th>
</tr>
</thead>
</table>

If yes, RI (early) defined as:

Other definitions used for parasite response (specify):
Describe any deviations from the WHO standard protocol for parasitological response:

Definitions of Therapeutic Response:

<table>
<thead>
<tr>
<th>Standard WHO (1996) ETF/LTF/ACR</th>
<th>Yes</th>
<th>No</th>
</tr>
</thead>
<tbody>
<tr>
<td>Standard WHO (2001) ETF/LCF/LPF/ACPR</td>
<td>Yes</td>
<td>No</td>
</tr>
</tbody>
</table>

Other definitions for clinical response (specify):

Describe any deviations from the WHO protocol for therapeutic response:

<table>
<thead>
<tr>
<th>Adverse event outcomes recorded</th>
<th>Yes</th>
<th>No</th>
</tr>
</thead>
<tbody>
<tr>
<td>If yes, describe adverse event outcomes</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

INCLUSION CRITERIA:

Age Ranges
<table>
<thead>
<tr>
<th>Age Range</th>
<th>All ages</th>
<th>School Aged</th>
<th>Not specified</th>
<th>Adults</th>
<th>Other age range</th>
<th>Pregnant women</th>
<th>Specific age range given (specify months or years)</th>
<th>Min</th>
<th>Max</th>
</tr>
</thead>
</table>

**Parasite Information Species**

<table>
<thead>
<tr>
<th></th>
<th>P. falciparum</th>
<th>P. ovale</th>
<th>P. vivax</th>
<th>P. malariae</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Method used to determine parasite count

<table>
<thead>
<tr>
<th>Expressed as asexual parasites/μL</th>
<th>Yes</th>
<th>No</th>
</tr>
</thead>
<tbody>
<tr>
<td>Minimum parasite count</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Maximum parasite count</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Expressed as + system</th>
<th>Yes</th>
<th>No</th>
</tr>
</thead>
<tbody>
<tr>
<td>1-10 parasites per 100 thick film fields</td>
<td>+</td>
<td></td>
</tr>
<tr>
<td>11-100 parasites per 100 thick film fields</td>
<td>++</td>
<td></td>
</tr>
<tr>
<td>1-10 parasites per single thick film field</td>
<td>+++</td>
<td></td>
</tr>
<tr>
<td>&gt; 10 parasites per single thick film field</td>
<td>++++</td>
<td></td>
</tr>
</tbody>
</table>

Describe if other method used:

<table>
<thead>
<tr>
<th>Minimum parasite count</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>Maximum parasite count</td>
<td></td>
</tr>
<tr>
<td>Specification</td>
<td>Yes</td>
</tr>
<tr>
<td>---------------</td>
<td>-----</td>
</tr>
<tr>
<td>Specify frequency of parasite smears</td>
<td></td>
</tr>
<tr>
<td>Gametocyte response reported?</td>
<td>Yes</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Specification</th>
<th>Min</th>
<th>Max</th>
<th>Not stated</th>
</tr>
</thead>
<tbody>
<tr>
<td>History of fever?</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Measured fever? (°C)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Haemoglobin (g/dl)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Haematocrit (%)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Other inclusion criteria? Specify:</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

**EXCLUSION CRITERIA**

<table>
<thead>
<tr>
<th>Specification</th>
<th>Yes</th>
<th>No</th>
<th>Not stated</th>
<th>Specify drug tested for</th>
</tr>
</thead>
<tbody>
<tr>
<td>Previous treatment excluded?</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Method of urinalysis, if used</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Pregnancy excluded?</td>
<td>Yes</td>
<td>No</td>
<td>Not stated</td>
<td></td>
</tr>
<tr>
<td>Severe malaria excluded?</td>
<td>Yes</td>
<td>No</td>
<td>Not stated</td>
<td></td>
</tr>
<tr>
<td>Concurrent illness excluded?</td>
<td>Yes</td>
<td>No</td>
<td>Not stated</td>
<td></td>
</tr>
<tr>
<td>Hospitalised excluded?</td>
<td>Yes</td>
<td>No</td>
<td>Not stated</td>
<td></td>
</tr>
<tr>
<td>Anaemia excluded?</td>
<td>Yes</td>
<td>No</td>
<td>Not stated</td>
<td></td>
</tr>
</tbody>
</table>
If yes, below what level?

Other exclusion criteria? Specify:

<table>
<thead>
<tr>
<th>Data point number</th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
</table>

**STUDY RESULTS**

A. Study population characteristics:

| Beginning date (dd/mm/yyyy) | - | - | - |
| Ending date (dd/mm/yyyy) | - | - | - |

<table>
<thead>
<tr>
<th>Reported malaria transmission/ season</th>
<th>Dry</th>
<th>Short Rains</th>
<th>Rains</th>
<th>Long Rains</th>
</tr>
</thead>
</table>

Name of drug(s) tested

Number enrolled

Number completing study/evaluated at end of study?

<table>
<thead>
<tr>
<th>Age ranged of enrolled? (specify years or months)</th>
<th>Min</th>
<th>Max</th>
<th>Follow-up days</th>
</tr>
</thead>
<tbody>
<tr>
<td>Mean age of enrolled (specify years or months)</td>
<td>SD</td>
<td></td>
<td>Not stated</td>
</tr>
<tr>
<td>Mean weight (kg)</td>
<td>SD</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Gender (%)</td>
<td>M</td>
<td>F</td>
<td>Not stated</td>
</tr>
<tr>
<td>Parasite density at enrolment?</td>
<td>Geometric mean</td>
<td>SD</td>
<td>Range</td>
</tr>
<tr>
<td>-------------------------------</td>
<td>----------------</td>
<td>----</td>
<td>-------</td>
</tr>
<tr>
<td>Haemoglobin (g/dl) at enrolment</td>
<td>Mean</td>
<td>SD</td>
<td>Range</td>
</tr>
<tr>
<td>Haematocrit (%) at enrolment</td>
<td>Mean</td>
<td>SD</td>
<td>Range</td>
</tr>
<tr>
<td>Gametocytaemia (µL) at enrolment</td>
<td>Mean</td>
<td>SD</td>
<td>Range</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Parasite clearance time (specify days or hours)</th>
<th>Mean</th>
<th>SD</th>
</tr>
</thead>
<tbody>
<tr>
<td>Fever clearance time (specify days or hours)</td>
<td>Mean</td>
<td>SD</td>
</tr>
<tr>
<td>Parasite reduction ratio at 48 hours</td>
<td>Mean</td>
<td>SD</td>
</tr>
</tbody>
</table>

**Describe haematologic results:**

**Type of analyses**

<table>
<thead>
<tr>
<th>Intention to treat</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>Per protocol</td>
<td></td>
</tr>
<tr>
<td>Survival analysis Kaplan-Meier</td>
<td></td>
</tr>
<tr>
<td>Other</td>
<td></td>
</tr>
</tbody>
</table>

**Were results corrected by PCR genotyping?**

<table>
<thead>
<tr>
<th>Yes</th>
<th>No</th>
</tr>
</thead>
</table>
PCR Genotyping:

Markers used

<table>
<thead>
<tr>
<th>msp 1</th>
<th>msp 2</th>
<th>msp 1 and msp 2</th>
<th>msp 1, msp 2 &amp; glurp</th>
<th>msp 1, msp 2 &amp; others</th>
<th>Not reported</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Definition of recrudescence:

Interpretation of mixed results:
Parasitologic Results:
Standard WHO categories (as reported) – specify if other combined results reported

<table>
<thead>
<tr>
<th>Response</th>
<th>Follow-up days</th>
<th>Number</th>
<th>%</th>
<th>Resistance Profile</th>
</tr>
</thead>
<tbody>
<tr>
<td>Sensitive</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>RI</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>RII</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>RIII</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>RI + RII</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>RII + RIII</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>RI + RII + RIII</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

If other definitions used, specify results:
Therapeutic Response:

Standard WHO categories (as reported):

<table>
<thead>
<tr>
<th>Response</th>
<th>Follow-up days</th>
<th>Response Recrudescence</th>
<th>New Infection</th>
<th>Unresolved</th>
<th>95% CI</th>
<th>Resistance Profile</th>
</tr>
</thead>
<tbody>
<tr>
<td>ETF</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>LTF</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>LCF</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>LPF</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>ACR:</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>ACPR:</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Cure</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Total/Glob failure rate</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Describe if other definitions used:

Describe results for adverse events:
### MOLECULAR MARKERS

#### ADMINISTRATIVE INFORMATION:

**Country:**

<table>
<thead>
<tr>
<th>Country</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>Beginning date (mm/yyyy)</td>
<td>-</td>
</tr>
<tr>
<td>Ending date (mm/yyyy)</td>
<td>-</td>
</tr>
<tr>
<td>Comments if author(s) contacted</td>
<td></td>
</tr>
</tbody>
</table>

**Survey Site:**

<table>
<thead>
<tr>
<th>Survey Site Name</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>Beginning date (mm/yyyy)</td>
<td>-</td>
</tr>
<tr>
<td>Ending date (mm/yyyy)</td>
<td>-</td>
</tr>
<tr>
<td>Comments if author(s) contacted</td>
<td></td>
</tr>
</tbody>
</table>

Data point number - - - -
<table>
<thead>
<tr>
<th>Method</th>
<th>Reference</th>
</tr>
</thead>
<tbody>
<tr>
<td>A Sequencing (looks for everything)</td>
<td></td>
</tr>
<tr>
<td>B Dot blotting</td>
<td>Pearce <em>et al.</em>, AAC 2003</td>
</tr>
<tr>
<td></td>
<td>Abdel-Muhsin <em>et al.</em>, AJTMH 2002</td>
</tr>
<tr>
<td></td>
<td>Alifrangis <em>et al.</em>, AJTMH 2005</td>
</tr>
<tr>
<td></td>
<td>(all same probe seqs)</td>
</tr>
<tr>
<td>C PCR-RFLP</td>
<td>Duraisingh <em>et al.</em>, Exp. Parasitol. 1998</td>
</tr>
<tr>
<td>D PCR Specific Amplification</td>
<td>Plowe <em>et al.</em>, AJTMH 1995</td>
</tr>
<tr>
<td></td>
<td>Doumbo <em>et al.</em>, JID 2000</td>
</tr>
<tr>
<td></td>
<td>(this method assumes seq for wt codon 50 in their primer for 51)</td>
</tr>
<tr>
<td>E Mass Spectrometry</td>
<td>Marks <em>et al.</em>, AAC 2004</td>
</tr>
<tr>
<td>F SNP Typing by Primer Extension</td>
<td>Nair <em>et al.</em>, Int J. Parasitol 2002</td>
</tr>
<tr>
<td></td>
<td>Marks <em>et al.</em>, AAC 2005</td>
</tr>
<tr>
<td>G Other</td>
<td></td>
</tr>
<tr>
<td>H Indicate if more than one method used</td>
<td></td>
</tr>
<tr>
<td>Codon w/t</td>
<td>16A</td>
</tr>
<tr>
<td>---------</td>
<td>-----</td>
</tr>
<tr>
<td>Looked at?</td>
<td>YES</td>
</tr>
<tr>
<td>(Tick yes/no)</td>
<td>NO</td>
</tr>
<tr>
<td>What was found?</td>
<td>A(GCA)</td>
</tr>
<tr>
<td></td>
<td>S(TCA)</td>
</tr>
<tr>
<td></td>
<td>V(GTA)</td>
</tr>
</tbody>
</table>

<p>|</p>
<table>
<thead>
<tr>
<th>Prevalence</th>
<th>N</th>
<th>N</th>
<th>N</th>
<th>N</th>
<th>N</th>
<th>N</th>
<th>N</th>
<th>N</th>
<th>Other</th>
</tr>
</thead>
<tbody>
<tr>
<td>A</td>
<td>C</td>
<td>N</td>
<td>C</td>
<td>S</td>
<td>V</td>
<td>I</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>S</td>
<td>R</td>
<td>N</td>
<td>R</td>
<td>N</td>
<td>L</td>
<td>L</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>V</td>
<td>I</td>
<td>T</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Can haplotypes be deduced? : YES NO

Haplotypes Identified

<table>
<thead>
<tr>
<th>Count</th>
<th>16</th>
<th>50</th>
<th>51</th>
<th>59</th>
<th>108</th>
<th>140</th>
<th>164</th>
<th>Other</th>
</tr>
</thead>
</table>

HAPLOTYPE COUNT (TOTAL N):
DHPS:
Method by which mutations looked for:

<table>
<thead>
<tr>
<th>Method</th>
<th>Reference</th>
</tr>
</thead>
<tbody>
<tr>
<td>A Sequencing (looks for everything)</td>
<td></td>
</tr>
</tbody>
</table>
| B Dot blotting Pearce et al., AAC 2003  
Alifrangis et al., AJTMH 2005 |                               |
| C PCR-RFLP                     | Duraisingh et al., Exp. Parasitol. 1998 |
| D PCR Specific Amplification   | Plowe et al., JID 1997         |
| E Mass Spectrometry            | Marks et al., AAC 2004         |
| F SNP Typing by Primer SNP typing by primer extension | Nair et al., Int J. Parasitol 2002  
Marks et al., AAC 2005 |
<p>| G Other                        |                               |
| H Indicate if more than one method used |                               |</p>
<table>
<thead>
<tr>
<th>Codon w/t</th>
<th>436S</th>
<th>437A</th>
<th>540K</th>
<th>581A</th>
<th>613A</th>
<th>Other Mutations/Notes</th>
</tr>
</thead>
<tbody>
<tr>
<td>Looked at? (Tick yes/no)</td>
<td>YES</td>
<td>YES</td>
<td>YES</td>
<td>YES</td>
<td>YES</td>
<td>YES</td>
</tr>
<tr>
<td>NO</td>
<td>NO</td>
<td>NO</td>
<td>NO</td>
<td>NO</td>
<td>NO</td>
<td>NO</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Prevalence</th>
<th>N Tot</th>
<th>N Tot</th>
<th>N Tot</th>
<th>N Tot</th>
<th>N Tot</th>
<th>N Tot</th>
<th>Other</th>
</tr>
</thead>
<tbody>
<tr>
<td>S</td>
<td>A</td>
<td>K</td>
<td>A</td>
<td>A</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>F</td>
<td>G</td>
<td>E</td>
<td>G</td>
<td>S</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>A</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>T</td>
<td></td>
<td></td>
</tr>
<tr>
<td>C</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Can haplotypes be deduced? : YES NO

<table>
<thead>
<tr>
<th>Haplotypes Identified</th>
</tr>
</thead>
<tbody>
<tr>
<td>Count</td>
</tr>
<tr>
<td>-------</td>
</tr>
<tr>
<td></td>
</tr>
</tbody>
</table>

HAPLOTYPE COUNT (TOTAL N):   

286
Appendix 2  


unusual pfhdfr and pfmdr-1 haplotypes and the decrease of dihydroartemisinin susceptibility. *Antimicrobial Agents and Chemotherapy*, 53, 4588-4597.


Noranate, N., Durand, R., Tall, A., Marrama, L., Spiegel, A., Sokhna, C., Pradines, B.,
Cojean, S., Guillotte, M., Bischoff, E., Ekala, M. T., Bouchier, C., Fandeur, T.,
Ariey, F., Patarapotikul, J., Le Bras, J., Trape, J. F., Rogier, C. & Mercereau-

pyrimethamine and chlorcycloguanil activity *in vitro* and point mutations in the
dihydrofolate reductase domain. *Antimicrobial Agents and Chemotherapy*, 42,
164-169.

W. M. (2000) Towards an understanding of the mechanism of pyrimethamine-
sulfadoxine resistance in *Plasmodium falciparum*: genotyping of dihydrofolate
reductase and dihydropteroate synthase of Kenyan parasites. *Antimicrobial
Agents and Chemotherapy*, 44, 991-996.

(2003) Molecular monitoring of the Leu-164 mutation of dihydrofolate reductase in

Ochong, E., Bell, D. J., Johnson, D. J., D’alessandro, U., Mulenga, M., Muangnoicharoen,
(2008) *Plasmodium falciparum* strains harboring dihydrofolate reductase with the
I164L mutation are absent in Malawi and Zambia even under antifolate drug

Oesterholt, M. J., Alifrangis, M., Sutherland, C. J., Omar, S. A., Sawa, P., Howitt, C.,
gametocytes and the transmission of antifolate-resistant *Plasmodium falciparum* in

and *dhfr* point mutations in *Plasmodium falciparum* infections be used to predict
sulfadoxine-pyrimethamine treatment failure? *Transactions of the Royal Society of
Tropical Medicine and Hygiene*, 95, 315-319.

determination of point mutation haplotypes in the dihydrofolate reductase and
dihydropteroate synthase of *Plasmodium falciparum* in three districts of northern

Community pyrimethamine-sulfadoxine use and prevalence of resistant


Appendix 3 dhps 436


isolates of *Escherichia coli* from the same children in Guinea-Bissau. *Tropical Medicine and International Health*, 9, 171-177.


chloroquine and pyrimethamine in *Plasmodium falciparum* from Zimbabwe.

*Parasitology Research*, 101, 1147-1151.


*Tropical Medicine and International Health*, 10, 901-908.


Raman, J., Little, F., Roper, C., Kleinschmidt, I., Cassam, Y., Maharaj, R. & Barnes, K. I. (2010) Five years of large-scale dhfr and dhps mutation surveillance following the phased implementation of artesunate plus sulfadoxine-pyrimethamine in Maputo


Resistance to antifolates in *Plasmodium falciparum* monitored by sequence analysis of dihydropteroate synthetase and dihydrofolate reductase alleles in a large number of field samples of diverse origins. *Molecular and Biochemical Parasitology*, 89, 161-177.

Appendix 4 dhps 437


highland sites and antimalarial drug resistance in Western Kenya. Infection, Genetics and Evolution, 9, 806-812.


of uncomplicated falciparum malaria in a refugee settlement in Zambia. *Tropical Medicine and International Health*, 10, 139-145.


resistance in island populations of *Plasmodium falciparum*. BMC Infectious Diseases, 10, 163.


Appendix 5 dhps 540


highland sites and antimalarial drug resistance in Western Kenya. *Infection, Genetics and Evolution*, 9, 806-812.


dihydropteroate synthase may not be the only determinants in sulfa resistance. *Experimental Parasitology*, 101, 90-96.


resistance exerted by the long-acting antifolate Pyrimethamine/Sulfadoxine compared with the shorter-acting chlorproguanil/dapsone on Kenyan *Plasmodium falciparum*. *Journal of Infectious Diseases*, 181, 2023-2028.


Sudan. *Transactions of the Royal Society of Tropical Medicine and Hygiene*, 97, 229-235.


Appendix 6 dhps 581


Appendix 7 dhps 613S


Appendix 8 dhps 613T


Appendix 9 dhfr IRN


genotypic resistance and efficacy of the artesunate plus sulfadoxine-
pyrimethamine and artesunate plus amodiaquine combinations. Transactions of the Royal Society of Tropical Medicine and Hygiene, 99, 548-554.


*Plasmodium falciparum* à la sulfadoxine-pyriméthamine par le test moléculaire à

Tinto, H., Ouedraogo, J. B., Zongo, I., Van Overmeir, C., Van Marck, E., Guiguemde, T. R.
& D'alessandro, U. (2007) Sulfadoxine-pyrimethamine efficacy and selection of
*Plasmodium falciparum DHFR* mutations in Burkina Faso before its introduction as
intermittent preventive treatment for pregnant women. *American Journal of
Tropical Medicine and Hygiene*, 76, 608-613.

Wang, P., Lee, C. S., Bayoumi, R., Djimde, A., Doumbo, O., Swedberg, G., Dao, L. D.,
Resistance to antifolates in *Plasmodium falciparum* monitored by sequence
analysis of dihydropteroate synthetase and dihydrofolate reductase alleles in a
large number of field samples of diverse origins. *Molecular and Biochemical
Parasitology*, 89, 161-177.

Wilson, P. E., Kazadi, W., Alker, A. P. & Meshnick, S. R. (2005) Rare Congolese
*Plasmodium falciparum DHFR* alleles. *Molecular and Biochemical Parasitology*,
144, 227-229.

dihydrofolate reductase genes and multidrug resistance gene 1-86 on treatment of
falciparum malaria in Sudan. *Journal of Infection in Developing Countries*, 4, 61-69.

Unpublished data

Appendix 10 dhfr NCS


with uncomplicated malaria in Burkina Faso. *American Journal of Tropical Medicine and Hygiene*, 76, 237-244.


Noranate, N., Durand, R., Tall, A., Marrama, L., Spiegel, A., Sokhna, C., Pradines, B., Cojean, S., Guillotte, M., Bischoff, E., Ekala, M. T., Bouchier, C., Fandeur, T.,


370


Unpublished data

Appendix 11 dhps GE,GK and AK


dihydropteroate synthase may not be the only determinants in sulfa resistance. 

Experimental Parasitology, 101, 90-96.


378
reductase and dihydropteroate synthase and epidemiologic patterns of pyrimethamine-sulfadoxine use and resistance. *Journal of Infectious Diseases*, 176, 1590-1596.


Appendix 12 Suitable and matched *in vivo* studies

There were 53 suitable *in vivo* studies and of these 38 were matched with molecular markers.

Suitable studies: unmatched


Suitable studies: matched with molecular markers


Moderate effect of artemisinin-based combination therapy on transmission of *Plasmodium falciparum*. *Journal of Infectious Diseases*, 193, 1151-1159.


Menard, D., Ratsimbasona, A., Randriaranivelosia, M., Rabarjaona, L. P., Raharimala, L., Domarle, O., Randrianasolo, L., Randriamanantena, A., Jahevitra, M.,


### Appendix 13 Matched studies: SP in vivo studies with dhfr IRN and NCS

<table>
<thead>
<tr>
<th>Country</th>
<th>in vivo site</th>
<th>Longitude</th>
<th>Latitude</th>
<th>Demographic information</th>
<th>Follow-up days</th>
<th>dhfr IRN site study period Reference</th>
<th>dhfr NCS site study period Reference</th>
</tr>
</thead>
<tbody>
<tr>
<td>Democratic Republic of Congo</td>
<td>Kabalo 2004 Bonnet et al., 2009</td>
<td>26.91708</td>
<td>-6.05011</td>
<td>8 months-96 months</td>
<td>28</td>
<td>Shabunda, 330 km north of Kabalo 2004</td>
<td>Swarthout et al., 2006</td>
</tr>
<tr>
<td>Kenya</td>
<td>Mbita 2003-2004 Bousema et al., 2006</td>
<td>34.23986</td>
<td>-0.0004</td>
<td>2 years – 5.3 years</td>
<td>28</td>
<td>Kakamega, 35 km east of Mbita; Kombewa, 28 km east of 2005</td>
<td>Zhong et al., 2008</td>
</tr>
<tr>
<td>Kenya</td>
<td>Siaya 2002-2004 Shah et al., 2006</td>
<td>34.286</td>
<td>0.06</td>
<td>Adults - median 31 years</td>
<td>28</td>
<td>Kisumu, 51 km east of Siaya 2002</td>
<td>Hamel et al., 2008</td>
</tr>
<tr>
<td>Malawi</td>
<td>Chileka 2003-2005 Bell et al., 2008</td>
<td>34.43509</td>
<td>-15.97416</td>
<td>12 months – 60 months</td>
<td>42</td>
<td>Blantyre, 57.5 km north of Chileka 2005</td>
<td>Nkhoma et al., 2007</td>
</tr>
<tr>
<td>Mozambique</td>
<td>Maputo Province (Catuane, Namaache, Boane, Magude) 2003-2005 Allen et al., 2009</td>
<td>32.589166</td>
<td>-25.965277</td>
<td>5 years – 22 years</td>
<td>42</td>
<td>Maputo 2003</td>
<td>Raman et al., 2008</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Location</td>
<td>Description</td>
<td>Latitude</td>
<td>Longitude</td>
<td>Age Range</td>
<td>Study Duration</td>
<td>Location Details</td>
<td></td>
</tr>
<tr>
<td>----------------</td>
<td>--------------------------------------------------</td>
<td>----------</td>
<td>-----------</td>
<td>------------</td>
<td>----------------</td>
<td>--------------------------------------------------------------------------------</td>
<td></td>
</tr>
<tr>
<td>North Sudan</td>
<td>New Halfa (Elhara Elrula) 2001-2002 Adam et al., 2004</td>
<td>35.583</td>
<td>15.333</td>
<td>1 year – 14 years</td>
<td>28</td>
<td>Daraweesh and Kajara, 130 km south of New Halfa 2003 A-elbasit et al., 2007</td>
<td></td>
</tr>
<tr>
<td>North Sudan</td>
<td>New Halfa (Alhara Aloula) 2004 Adam et al., 2005</td>
<td>35.583</td>
<td>15.333</td>
<td>Mean 12.8 years</td>
<td>28</td>
<td>Ngorban country, 610 km southwest of New Halfa 2003 Hamour et al., 2005; Yusuf et al., 2010</td>
<td></td>
</tr>
<tr>
<td>South Africa</td>
<td>Mosvold Hospital 1996 Freese et al., 1996</td>
<td>32.00527</td>
<td>-27.13678</td>
<td>23.8 years</td>
<td>28</td>
<td>Ingwavuma (Mosvold and Ndumu) 1996 Roper et al., 2003</td>
<td></td>
</tr>
<tr>
<td>South Africa</td>
<td>Mangweni Clinic 2002 Mabuza et al., 2005</td>
<td>31.82</td>
<td>-25.74</td>
<td>2 years and above</td>
<td>42</td>
<td>Mpumalanga 2000, 2002 Barnes et al., 2008</td>
<td></td>
</tr>
</tbody>
</table>

No match
<table>
<thead>
<tr>
<th>Country</th>
<th>Location</th>
<th>Latitude</th>
<th>Longitude</th>
<th>Duration</th>
<th>Reference</th>
<th>Match</th>
</tr>
</thead>
<tbody>
<tr>
<td>South Africa</td>
<td>Naas</td>
<td>31.86</td>
<td>-25.67</td>
<td>9 years - 28 years; 10 years – 23 years; 11 years – 28 years</td>
<td>Naas and Mangweni 2000,2002 Barnes et al.,2008</td>
<td>No match</td>
</tr>
<tr>
<td>Tanzania</td>
<td>Ipinda (2003)</td>
<td>33.76491</td>
<td>-9.05015</td>
<td>6 months – 59 months</td>
<td>Mkokola and Kwamasimba, 670 km north east of Ipinda 2003 Alifrangis et al.,2009 Masasi is 33 km northeast of Ipinda 1997 Mugittu et al.,2004</td>
<td>Mkokola and Kwamasimba, 670 km north east of Ipinda 2003 Alifrangis et al.,2009 Masasi is 33 km northeast of Ipinda 1997 Mugittu et al.,2004</td>
</tr>
<tr>
<td>Tanzania</td>
<td>Eukayosi, Bagamoyo (2004)</td>
<td>38.814341</td>
<td>-6.107977</td>
<td>6 months</td>
<td>Korogwe, 123 km north of Eukayosi 2004 Alifrangis et al.,2009 Kibaha is 69 km south of Eukayosi 1998 Alifrangis et al.,2003</td>
<td>Korogwe, 123 km north of Eukayosi 2004 Alifrangis et al.,2009 Kibaha is 69 km south of Eukayosi 1998 Alifrangis et al.,2003</td>
</tr>
<tr>
<td>Country</td>
<td>Location</td>
<td>Year</td>
<td>Latitude</td>
<td>Longitude</td>
<td>Age Range</td>
<td>Sample Size</td>
</tr>
<tr>
<td>-------------</td>
<td>----------------</td>
<td>---------------</td>
<td>-----------</td>
<td>-----------</td>
<td>---------------</td>
<td>-------------</td>
</tr>
<tr>
<td>Tanzania</td>
<td>Tindiga (Kilosa)</td>
<td>2004</td>
<td>37.002</td>
<td>-6.00258</td>
<td>6 months – 60 months</td>
<td>28</td>
</tr>
<tr>
<td>Tanzania</td>
<td>Tanga</td>
<td>2006</td>
<td>38.6</td>
<td>-5.283</td>
<td>Mean 31.1 months</td>
<td>28</td>
</tr>
<tr>
<td>Uganda</td>
<td>Bundi Bugyo</td>
<td>2002</td>
<td>30.07</td>
<td>0.7</td>
<td>6 months – 59 months</td>
<td>28</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Uganda</td>
<td>Mulago</td>
<td>1999-2001</td>
<td>32.58</td>
<td>0.35</td>
<td>6 months – 10 years</td>
<td>28</td>
</tr>
<tr>
<td>Country</td>
<td>Location</td>
<td>Year</td>
<td>Cases</td>
<td>Age Range</td>
<td>Duration</td>
<td>Match</td>
</tr>
<tr>
<td>---------</td>
<td>----------</td>
<td>------</td>
<td>-------</td>
<td>-----------</td>
<td>----------</td>
<td>-------</td>
</tr>
<tr>
<td>Uganda</td>
<td>Nagongera 2001-2002 Talisuna et al., 2004</td>
<td>34.02</td>
<td>0.77</td>
<td>6 months – 59 months</td>
<td>28</td>
<td>Tororo Nagongera is 20km northeast from Tororo 2003-2006 Malamba et al., 2010</td>
</tr>
<tr>
<td>Uganda</td>
<td>Tororo 2001-2002 Talisuna et al., 2004</td>
<td>34.186</td>
<td>0.695</td>
<td>6 months – 59 months</td>
<td>28</td>
<td>Tororo 2003-2006 Malamba et al., 2010</td>
</tr>
<tr>
<td>Country</td>
<td>Site</td>
<td>Study period</td>
<td>Latitude</td>
<td>Demographic information</td>
<td>Follow-up days</td>
<td>dhfr IRN site study period</td>
</tr>
<tr>
<td>--------------</td>
<td>---------------------------</td>
<td>--------------</td>
<td>------------</td>
<td>-------------------------</td>
<td>----------------</td>
<td>---------------------------</td>
</tr>
<tr>
<td>Angola</td>
<td>Caala</td>
<td>2002</td>
<td>-12.85</td>
<td>Mean age 26.4 months</td>
<td>28</td>
<td>Uige, 630 km north of Caala</td>
</tr>
<tr>
<td>Burkina Faso</td>
<td>Ouagadougou</td>
<td>2003</td>
<td>12.35886</td>
<td>6 months – 59 months</td>
<td>28</td>
<td>Nanoro, 85km from Ouagadougou</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>-1.01939</td>
<td></td>
<td></td>
<td>Ziniare and Bousse are 19 km north of Ouagadougou</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>9.13403</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Location</td>
<td>Country</td>
<td>Year</td>
<td>Study Reference</td>
<td>Latitude</td>
<td>Longitude</td>
<td>Duration</td>
</tr>
<tr>
<td>------------------------</td>
<td>----------------------</td>
<td>---------------</td>
<td>----------------------------------</td>
<td>----------</td>
<td>-----------</td>
<td>-------------------</td>
</tr>
<tr>
<td>Chad</td>
<td></td>
<td>2002</td>
<td>Grandesso et al., 2006</td>
<td>15.3711</td>
<td>10.27591</td>
<td>6 months – 59 months</td>
</tr>
<tr>
<td>Democratic Republic of</td>
<td></td>
<td>2003-2004</td>
<td>Bonnet et al., 2009</td>
<td>21.1310</td>
<td>0.10117</td>
<td>6 months-59 months</td>
</tr>
<tr>
<td>Gabon</td>
<td></td>
<td>2000</td>
<td>Aubouy et al., 2003</td>
<td>12.3</td>
<td>-1.55</td>
<td>Mean 3.8 years Maximum 10 years</td>
</tr>
<tr>
<td>Gabon</td>
<td></td>
<td>2005-2006</td>
<td>Oyakhirome et al., 2007</td>
<td>10.212</td>
<td>-0.707</td>
<td>3 years- 14 years</td>
</tr>
<tr>
<td>Gabon</td>
<td></td>
<td>2005</td>
<td>Nsimba et al., 2008</td>
<td>11.5638</td>
<td>1.0183</td>
<td>6 months-59 months</td>
</tr>
<tr>
<td>Ghana</td>
<td></td>
<td>2002</td>
<td>Mockenhaupt et al., 2004</td>
<td>-0.834</td>
<td>9.4</td>
<td>6 months-59 months</td>
</tr>
<tr>
<td>Country</td>
<td>Location</td>
<td>Year(s)</td>
<td>Latitude</td>
<td>Longitude</td>
<td>Age Group</td>
<td>Site Details</td>
</tr>
<tr>
<td>-----------</td>
<td>---------------------</td>
<td>-------------</td>
<td>----------------</td>
<td>---------------</td>
<td>---------------------------</td>
<td>-------------------------------------------------------------------------------</td>
</tr>
<tr>
<td>Ghana</td>
<td>Navrongo</td>
<td>2003</td>
<td>-1.1</td>
<td>10.9</td>
<td>8 months-56 months</td>
<td>Tamale 150 km south of Navrongo 2002</td>
</tr>
<tr>
<td></td>
<td>Koram et al., 2005</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>Muehlen et al., 2004</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>Mockenhaupt et al., 2005</td>
</tr>
<tr>
<td>Ghana</td>
<td>Nkoranza</td>
<td>2003-2004</td>
<td>-1.03517</td>
<td>7.05108</td>
<td>Children under 5 years</td>
<td>Kintampo, Brong Ahafo Region, &lt; 1 km south of Nkoranza 2003</td>
</tr>
<tr>
<td></td>
<td>Tagbor et al., 2007</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>Owusu-agyei et al., 2009</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>Ashanti region, Bodomase 2003</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>Marks et al., 2005</td>
</tr>
<tr>
<td>Madagascar</td>
<td>Moramanga and</td>
<td>2006</td>
<td>47.7779 1</td>
<td>-20.35941 5</td>
<td>6 months – 12 months</td>
<td>12 sites 2006 Andriantsaroarina et al., 2009</td>
</tr>
<tr>
<td></td>
<td>Saharevo</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>12 sites 2006 Andriantsaroarina et al., 2009</td>
</tr>
<tr>
<td></td>
<td>Menard et al., 2007</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>12 sites 2006 Andriantsaroarina et al., 2009</td>
</tr>
<tr>
<td>Madagascar</td>
<td>Ejeda</td>
<td>2006-2007</td>
<td>44.517</td>
<td>-24.35</td>
<td>6 months – 15 years</td>
<td>various sites 2006 Menard et al., 2008</td>
</tr>
<tr>
<td></td>
<td>Menard et al., 2008</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>various sites 2006 Menard et al., 2008</td>
</tr>
<tr>
<td></td>
<td>de Radigues et al., 2006</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>Tekete et al., 2009</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>Bougoula-</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>No match</td>
</tr>
<tr>
<td>Region</td>
<td>Location</td>
<td>Latitude</td>
<td>Longitude</td>
<td>Duration</td>
<td>Location Details</td>
<td>Notes</td>
</tr>
<tr>
<td>----------------------</td>
<td>---------------------------</td>
<td>----------</td>
<td>-----------</td>
<td>----------------</td>
<td>----------------------------------------------------------------------------------</td>
<td>----------------------------------------------------------------------</td>
</tr>
<tr>
<td>Mali</td>
<td>Kolle</td>
<td>-5.281</td>
<td>11.314</td>
<td>6 months – 59 months</td>
<td>Kąle 2002 - 2003 Djimde et al., 2008</td>
<td>No match</td>
</tr>
<tr>
<td>Republic of Congo</td>
<td>Milou and Makelekele health districts 2003-2004</td>
<td>11.88</td>
<td>-4.25</td>
<td>6 months – 67 months</td>
<td>Pointe Noire, Brazzaville and Gambona. Pointe Noire is 170 km</td>
<td>Pointe Noire, Brazzaville and Gambona. Pointe Noire is 170 km</td>
</tr>
</tbody>
</table>

Hameau, 200 km west of Koumanto 2002 Djimde et al., 2008
<table>
<thead>
<tr>
<th>Study Location</th>
<th>Study Area</th>
<th>Latitude</th>
<th>Longitude</th>
<th>Duration</th>
<th>Age</th>
<th>Match Status</th>
<th>Match Status</th>
</tr>
</thead>
<tbody>
<tr>
<td>Sierra Leone</td>
<td>Freetown</td>
<td>-13.25</td>
<td>8.5</td>
<td>6 months – 59 months</td>
<td>28</td>
<td>No match</td>
<td>No match</td>
</tr>
</tbody>
</table>

Ndounga et al., 2007

north east of Makelekele 2006
Mita et al., 2009
Mita et al., 2009

Sierra Leone Freetown 2002-2003 Checci et al., 2005
### Appendix 14 Matched studies: SP *in vivo* studies with *dhps* GE, GK and AK

<table>
<thead>
<tr>
<th>East</th>
<th><em>in vivo</em> site</th>
<th>Study period</th>
<th>Longitude</th>
<th>Latitude</th>
<th>Demographic information</th>
<th>Follow-up days</th>
<th><em>dhps</em> GE/GK/AK site</th>
<th>study period</th>
<th>Reference</th>
</tr>
</thead>
<tbody>
<tr>
<td>Democratic Republic of Congo</td>
<td>Kabalo 2004</td>
<td>26.91708</td>
<td>-6.05011</td>
<td>8 months-96 months</td>
<td>28</td>
<td>Shabunda, 330 km north of Kabalo 2004 Swarthout et al., 2006</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Bonnet et al., 2009</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Kenya</td>
<td>Mbita 2003-2004</td>
<td>34.23986</td>
<td>-0.0004</td>
<td>2 years – 5.3 years</td>
<td>28</td>
<td>Kombewa, 28 km east of 2005 Zhong et al., 2008</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Bousema et al., 2006</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Kenya</td>
<td>Siaya 2002-2004</td>
<td>34.286</td>
<td>0.06</td>
<td>Adults - median 31 years</td>
<td>28</td>
<td>Kakamega, 32 km east of Siaya 2005 Zhong et al., 2008</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Shah et al., 2006</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Malawi</td>
<td>Chileka 2003-2005</td>
<td>34.43509</td>
<td>-15.97416</td>
<td>12 months – 60 months</td>
<td>42</td>
<td>Blantyre, 57.5 km north east of Chileka 2005 Nkhoma et al., 2007</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Bell et al., 2008</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Mozambique</td>
<td>Maputo Province (Catuane, Namaache, Boane, Magude) 2003-2005</td>
<td>32.58916</td>
<td>-25.96527</td>
<td>5 years – 22 years</td>
<td>42</td>
<td>Maputo 2003,2004 Raman et al., 2008</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Allen et al., 2009</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Maputo 2004,2005 Raman et al., 2010</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Location</td>
<td>Area</td>
<td>Latitude</td>
<td>Longitude</td>
<td>Age Range</td>
<td>Sample Size</td>
<td>Notes</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>-------------------</td>
<td>-----------------------------</td>
<td>----------</td>
<td>-----------</td>
<td>------------</td>
<td>-------------</td>
<td>----------------------------------------------------------------------</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>North Sudan</td>
<td>New Halfa (Elhara Elrula)</td>
<td>35.583</td>
<td>15.333</td>
<td>1 year – 14 years</td>
<td>28</td>
<td>Khartoum, 20km north of New Halfa 1996-1997 Khalil et al., 2005</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>New Halfa (Alhara Aloula)</td>
<td>35.583</td>
<td>15.333</td>
<td>Mean 12.8 years</td>
<td>28</td>
<td>Khartoum, 20km north of New Halfa 1996-1997 Khalil et al., 2005</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Daraweesh and Kajara</td>
<td>35.37</td>
<td>14.0</td>
<td>Not stated</td>
<td>28</td>
<td>Daraweesh and Kajara 2003 A-elbasit et al., 2007</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>South Africa</td>
<td>Mosvold Hospital</td>
<td>32.00527</td>
<td>-27.13678</td>
<td>23.8 years</td>
<td>28</td>
<td>Ingwavuma 1995-1996 Roper et al., 2003</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Mangweni Clinic</td>
<td>31.82</td>
<td>-25.74</td>
<td>2 years and above</td>
<td>42</td>
<td>Komatipoort 2000 Pearce et al., 2009</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Naas</td>
<td>31.86</td>
<td>-25.67</td>
<td>9 years - 28 years; 10 years – 23 years; 11</td>
<td>42</td>
<td>Komatipoort 2000 Pearce et al., 2009</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Country</td>
<td>Location</td>
<td>Latitude</td>
<td>Longitude</td>
<td>Duration</td>
<td>Notes</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>-----------</td>
<td>----------------</td>
<td>---------------</td>
<td>---------------</td>
<td>-------------------</td>
<td>----------------------------------------------------------------------</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>South Sudan</td>
<td>Kajo Keji</td>
<td>31.667</td>
<td>3.883</td>
<td>6 months – 59</td>
<td>Lankien, 500km northeast of Kajo Keji 2001 van den Broek et al., 2003</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Tanzania</td>
<td>Ipinda</td>
<td>33.76491</td>
<td>-9.05015</td>
<td>6 months – 59</td>
<td>Masasi is 33 km northeast of Ipinda 2000 Mugittu et al., 2004</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Tanzania</td>
<td>Eukayosi, Bagamoyo</td>
<td>38.81434</td>
<td>-6.107977</td>
<td>6 months</td>
<td>Korogwe, 123 km north of Eukayosi 2004 Alifrangis et al., 2009</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Tanzania</td>
<td>Tindiga(Kilosa)</td>
<td>37.002</td>
<td>-6.00258</td>
<td>6 months – 60</td>
<td>Chamwino, 120 km west of Tindiga 2005 Enevold et al., 2007</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Tanzania</td>
<td>Tanga</td>
<td>38.6</td>
<td>-5.283</td>
<td>Mean 31.1 months</td>
<td>Korogwe, 33km northwest of Tanga 2006 Alifrangis et al., 2009</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Uganda</td>
<td>Mbarara</td>
<td>30.65</td>
<td>-0.62</td>
<td>6 months – 59</td>
<td>Mbarara, 79km southeast from Rukungiri 2005-2008</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Country</td>
<td>Location</td>
<td>Year</td>
<td>Age Range</td>
<td>Average</td>
<td>Median</td>
<td>Distance</td>
<td>Notes</td>
<td></td>
<td></td>
</tr>
<tr>
<td>---------</td>
<td>----------</td>
<td>------</td>
<td>-----------</td>
<td>---------</td>
<td>--------</td>
<td>----------</td>
<td>-------</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Uganda</td>
<td>Bundi Bugyo</td>
<td>2002</td>
<td>6 months – 59 months</td>
<td>30.07</td>
<td>0.7</td>
<td>279 km east of Bundi Bugyo</td>
<td>2000-2001 Lynch et al., 2008</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Uganda</td>
<td>Mulago</td>
<td>1999-2001</td>
<td>6 months – 10 years</td>
<td>32.58</td>
<td>0.35</td>
<td>3 km south of Mulago</td>
<td>2000 Dorsey et al., 2004</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Uganda</td>
<td>Nagongera</td>
<td>2001-2002</td>
<td>6 months – 59 months</td>
<td>34.02</td>
<td>0.77</td>
<td>160 km east of Nagongera</td>
<td>2000-2001 Talisuna et al., 2004</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Uganda</td>
<td>Tororo</td>
<td>2001-2002</td>
<td>6 months – 59 months</td>
<td>34.186</td>
<td>0.695</td>
<td>175 km west of Tororo</td>
<td>2000-2001 Talisuna et al., 2004</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Uganda</td>
<td>Tororo</td>
<td>2001-2002</td>
<td>6 months – 59 months</td>
<td>34.186</td>
<td>0.695</td>
<td>175 km west of Tororo</td>
<td>2000-2001 Talisuna et al., 2004</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Uganda</td>
<td>Tororo</td>
<td>2001-2002</td>
<td>6 months – 59 months</td>
<td>34.186</td>
<td>0.695</td>
<td>175 km west of Tororo</td>
<td>2000-2001 Talisuna et al., 2004</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Uganda</td>
<td>Tororo</td>
<td>2001-2002</td>
<td>6 months – 59 months</td>
<td>34.186</td>
<td>0.695</td>
<td>175 km west of Tororo</td>
<td>2000-2001 Talisuna et al., 2004</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Uganda</td>
<td>Tororo</td>
<td>2001-2002</td>
<td>6 months – 59 months</td>
<td>34.186</td>
<td>0.695</td>
<td>175 km west of Tororo</td>
<td>2000-2001 Talisuna et al., 2004</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Country</td>
<td>Site</td>
<td>Study period</td>
<td>Reference</td>
<td>Longitude</td>
<td>Latitude</td>
<td>Demographic information</td>
<td>Follow-up days</td>
<td>dhps GE/GK/AK site study period</td>
<td>Reference</td>
</tr>
<tr>
<td>---------</td>
<td>------</td>
<td>--------------</td>
<td>-----------</td>
<td>-----------</td>
<td>----------</td>
<td>--------------------------</td>
<td>----------------</td>
<td>---------------------------------</td>
<td>-----------</td>
</tr>
<tr>
<td>Angola</td>
<td>Caala</td>
<td>2002</td>
<td>Guthmann et al., 2005</td>
<td>15.558</td>
<td>-12.85</td>
<td>Mean age 26.4 months</td>
<td>28</td>
<td>Huambo is 16km north east of Caala 2007</td>
<td>Fortes et al., 2011</td>
</tr>
<tr>
<td>Burkina Faso</td>
<td>Ouagadougou</td>
<td>2003</td>
<td>Gansane et al., 2009</td>
<td>-1.01939</td>
<td>12.35886</td>
<td>6 months – 59 months</td>
<td>28</td>
<td>Nanoro, 85km from Ouagadougou 2003</td>
<td>Tinto et al., 2007</td>
</tr>
<tr>
<td>Chad</td>
<td>Bongor</td>
<td>2002</td>
<td></td>
<td>15.3711</td>
<td>10.27591</td>
<td>6 months – 59 months</td>
<td>28</td>
<td>Garoua, 210 north west of</td>
<td></td>
</tr>
<tr>
<td>Location</td>
<td>City/Region</td>
<td>Study Period</td>
<td>Age Range</td>
<td>Mean Age ± SD</td>
<td>Maximum Age</td>
<td>Notes</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>----------------------------------</td>
<td>------------------------------</td>
<td>--------------------</td>
<td>--------------------</td>
<td>---------------</td>
<td>------------</td>
<td>------------------------------------------------------------------------------------------</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Democratic Republic of Congo</td>
<td>Bongor</td>
<td>2004-2006</td>
<td>Mbacham et al., 2010</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Gabon</td>
<td>Lambarene</td>
<td>2005-2006</td>
<td>Oyakhirome et al., 2007</td>
<td>10.21 ± 0.70</td>
<td>3 years-14 years</td>
<td>Franceville, Haut-Ogooue district in south east Gabon 1998, 120km east of Bakoumba Mawili-mboumba et al., 2001</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Gabon</td>
<td>Oyem</td>
<td>2005</td>
<td>Nsimba et al., 2008</td>
<td>11.56 ± 1.01</td>
<td>6 months-59 months</td>
<td>Lambarene, 230km northwest from Oyem 2007 Pearce et al., 2009</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Ghana</td>
<td>Tamale</td>
<td>2002</td>
<td>Mockenhaupt et al., 2005</td>
<td>-0.83 ± 0.94</td>
<td>6 months-59 months</td>
<td>Navrongo, 150km south of Tamale 2003 Pearce et al., 2009</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Country</td>
<td>Location</td>
<td>Year(s)</td>
<td>Reference(s)</td>
<td>Start Age</td>
<td>End Age</td>
<td>Duration</td>
<td>Sites</td>
<td></td>
<td></td>
</tr>
<tr>
<td>---------</td>
<td>----------</td>
<td>---------</td>
<td>--------------</td>
<td>-----------</td>
<td>---------</td>
<td>----------</td>
<td>-------</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Ghana</td>
<td>Navrongo</td>
<td>2003</td>
<td>Koram et al., 2005</td>
<td>-1.1</td>
<td>10.9</td>
<td>8 months-56 months</td>
<td>28</td>
<td>Navrongo, 2003 Pearce et al., 2009</td>
<td></td>
</tr>
<tr>
<td>Ghana</td>
<td>Nkoranza</td>
<td>2003-2004</td>
<td>Tagbor et al., 2007</td>
<td>-1.03517</td>
<td>7.05108</td>
<td>Children under 5 years</td>
<td>28</td>
<td>Ashanti region, Bodomase 2003 Marks et al., 2005</td>
<td></td>
</tr>
<tr>
<td>Madagascar</td>
<td>Moramanga and Saharevo</td>
<td>2006</td>
<td>Menard et al., 2007</td>
<td>47.7779 1</td>
<td>-20.35941 5</td>
<td>6 months – 12 months</td>
<td>28</td>
<td>12 sites 2006 Andriantsoanirina et al., 2009</td>
<td></td>
</tr>
<tr>
<td>Madagascar</td>
<td>Ejeda</td>
<td>2006-2007</td>
<td>Menard et al., 2008</td>
<td>44.517</td>
<td>-24.35</td>
<td>6 months – 15 years</td>
<td>28</td>
<td>12 sites 2007 Andriantsoanirina et al., 2009</td>
<td></td>
</tr>
<tr>
<td>Mali</td>
<td>Koumanto</td>
<td>2003-2004</td>
<td>de Radigues et al., 2006</td>
<td>-6.83333 3</td>
<td>11.41666 6</td>
<td>1 year – 3 years</td>
<td>28</td>
<td>Sotuba, 190 km north west of Koumanto 1995 Diourte et al., 1999</td>
<td></td>
</tr>
<tr>
<td>Republic of Congo</td>
<td>Location</td>
<td>Population</td>
<td>Latitude</td>
<td>Duration</td>
<td>Duration in months</td>
<td>Notes</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>------------------</td>
<td>----------</td>
<td>------------</td>
<td>----------</td>
<td>----------</td>
<td>--------------------</td>
<td>-------</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Kindamba, 18 km west of Brazzaville 2004 Pearce et al., 2009

Brazzaville is 180 km southwest of Makelekele 2004-2009 Ndounga et al., 2007
<table>
<thead>
<tr>
<th>Country</th>
<th>Location</th>
<th>Age range</th>
<th>Duration</th>
<th>Sample size</th>
<th>Match</th>
</tr>
</thead>
<tbody>
<tr>
<td>Sierra Leone</td>
<td>Freetown 2002-2003</td>
<td>-13.25</td>
<td>8.5</td>
<td>28</td>
<td>No match</td>
</tr>
<tr>
<td></td>
<td>Checci et al., 2005</td>
<td></td>
<td>6 months – 59 months</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
Appendix 15 Matched dhfr IRN and NCS with in vivo studies


Appendix 16 Matched dhps GE, GK and AK studies with in vivo studies


Appendix 17 Publications and presentations emanating from this thesis

This section describes my contribution to publications in peer-reviewed journals and conference proceedings:

   In consultation with Dr Cally Roper, I identified the research questions, performed the analyses, generated and interpreted the results and prepared the manuscript.

   In consultation with Dr Cally Roper, I identified the research questions, performed the analyses, generated and interpreted the results and prepared the manuscript.

I contributed to the following publications:

   I contributed to data collection and critical revision of the paper.

   I contributed to data collection and critical revision of the paper.

   I contributed to data collection, data analyses and critical revision of the paper.

I contributed to data collection, data analyses and critical revision of the paper.

I contributed to data collection, data analyses and critical revision of the paper.

I contributed to the development of the concept and critical revision of the paper.

I contributed to development of the concept and critical revision of the paper.

Poster presentations:

In consultation with Dr Cally Roper and Dr Immo Kleinschmidt, I identified the research questions, performed the analyses, generated and interpreted the results, prepared and presented the poster.
In consultation with Dr Cally Roper and Dr Immo Kleinschmidt, I identified the research questions, performed the analyses, generated and interpreted the results, prepared and presented the poster.

I contributed to data collection and critical review of the poster.

In consultation with Dr Cally Roper, I identified the research questions, performed the analyses, generated and interpreted the results, prepared and presented the poster.

This poster presented the results from Gesase *et al.*, 2009, *PLoS One*, 4, e4569

In consultation with Dr Cally Roper, I identified the research questions, performed the analyses, generated and interpreted the results, prepared and presented the poster.
Oral presentation


In consultation with Dr Cally Roper and Prof Rajendra Maharaj, I identified the research questions, performed the analyses, generated and interpreted the results, prepared and delivered the presentation.

Endorsed by Dr Cally Roper.