The temporal and spatial distribution of malaria in Africa, with emphasis on southern Africa

INAUGURAL-DISSERTATION

zur

Erlangung der Würde eines Doktors der Philosophie

vorgelegt der

Philosophisch-Naturwissenschaftlichen Fakultät der

Universität Basel

von

Marlies H. Craig

aus

Durban, Südafrika

Südafrika, 2009

Genehmigt von der Philosophisch-Naturwissenschaftlichen Fakultät auf Antrag von Prof. Dr. M. Tanner, Prof. Dr. T. Smith, Dr. J. Cox.

Basel, den 11. Dezember 2007

Prof. Dr. Hans-Peter Hauri

Dekan

To:

Andrew,
faithful husband, friend and brother;

Miriam, Philip and Simon,
who light up my life;

and Brian Sharp,
sorely missed,
leader, mentor and friend.

Table of Contents

Page

Acknowledgements

First, I would like to thank my supervisor Prof. Tom Smith for his support during the years that this thesis was in progress, for his professional guidance, scientific insights and generous help, particularly on my last working trip to the STI. I consider myself very fortunate to have benefited from his expertise. Many thanks to him and his wife Julie for hosting me so kindly in Basel on more than one occasion. Likewise I sincerely appreciate the kind efforts of Dr Immo Kleinschmidt and Dr Penelope Vounatsou, who patiently taught me so much of what I know about statistics. Most of the statistical work in this thesis would have been quite impossible without their guidance. Sincere thanks to Dr Musa Mabaso, helpful critic of numerous manuscripts and co-worker in the MARA project. It is always a pleasure working with him. Thanks Musa, we have come a long way together.

I sincerely wish to acknowledge Dr Brian Sharp, recently deceased. I am extremely grateful for the privilege of having experienced his leadership. He was and remains a role model to me, a mentor, a guide. With sound motives and the bigger picture in mind, he was neither petty nor negative, but always saw the potential in people and situations. He did all in his power to support working mothers. I valued his opinion enormously and miss him very much. I would also like to acknowledge Dr David le Sueur, also deceased, who gave me valuable guidance during the early days of my career. He set me on a course of modelling malaria distribution, by sending me for training in Idrisi in the USA, guiding the development of the malaria distribution model, and then employing me as GIS coordinator in the MARA project.

Sincere thanks go to the other MARA collaborators, in particular Prof. Christian Lengeler and Prof. Don de Savigny, for their availability, support and commitment from the start of the

project until today. They were a major source of practical help, good advice and inspiration, and remain valued friends and allies. Prof. Lengeler was also the principal instigator who urged me to consider registering for a PhD through the STI. Andrew's chocolate factory didn't quite pan out, but the rest did. Thank you also for hosting me in Basel so kindly.

My gratitude also goes to Prof. Marcel Tanner for his enthusiastic support, and for giving me the opportunity of enrolling for doctoral studies through the STI, and to the STI and the Rudolf Geigy Stiftung zu Gunsten des Schweizerischen Tropeninstituts for supporting this study financially. I also value the support of the South African Medical Research Council, which grants its employees extended leave of absence so that we may pursue further study overseas.

I greatly appreciate the many other colleagues and fellow-students, both at the Medical Research Council and the Swiss Tropical Institute, in particular Laura and Dominic Gosoniu, Amanda Ross, Nicholas Maire, Claudia Sauerborn, Sohini Banerjee, for, above all, your friendship, for many stimulating discussions (not always about work) and for all the good times. I never could have done this work in isolation.

Thank you Andrew for your loving support and encouragement, especially when I felt like giving up. Thank you for accompanying me to Switzerland for 10 months, and for running the show back home during various shorter trips to Basel. Thank you Miriam and Philip for looking after dad when I was away. You are fantastic kids.

Sincere thanks to my aunt Marianne Kassier who kindly translated the summary into German, and to Michael Bretscher for editing the "Fachdeutsch". I really appreciate your help. Finally, many thanks to Dr Jon Cox for agreeing to be the external examiner.

Summary

The three-way relationship between the *Plasmodium* parasite, the *Anopheles* mosquito vector and the human host determines the incidence of malaria disease. The three life cycles, the interactions respectively between human and parasite, human and mosquito, and mosquito and parasite, and the ultimate transmission cycle, vary in time and space. Environmental, genetic and behavioural factors influence the three life cycles and the interactions. These factors also vary in time and space. At every level the variation itself, whether random or cyclical, is not uniform but varies in frequency and magnitude. Explaining, and particularly predicting, malaria transmission rates in time and space thus becomes a difficult undertaking.

Knowing and understanding some of this variation, and its causes, is important for well-timed and well-targeted malaria interventions. In the fringe areas of malaria in Africa, which are prone to epidemics, some forewarning of unusually high incidence periods would be valuable to malaria control and management services.

This thesis investigated the temporal and spatial effects on malaria transmission of various environmental factors, particularly climate, and of non-climatic factors, particularly those relating to malaria control. Different data sets and methodological approaches were applied in seven separate studies, and malaria distribution in time and space was investigated at different scales.

At the continental scale, the distribution of malaria in Africa was modelled as a factor of climate using raster GIS techniques.

At the national scale, using prevalence data from Botswana, spatial variation in prevalence was modelled as a factor of environmental determinants, prior to comprehensive malaria control. The spatial and inter-annual variation in prevalence, in the presence of intense control, was also modelled as a factor of climate.

At the sub-national level South Africa was used as an example. Inter-annual variation in malaria incidence in the highest-risk province was explored for possible links with climatic and non-climatic factors. Finally, inter-annual and spatial variation in sub-provincial level incidence data for South Africa, were analysed with respect to climatic and non-climatic determinants, for which data were available.

The two study areas (Botswana and South Africa) both lie at the fringe of malaria distribution, experience strongly seasonal transmission and epidemics, and both benefit from intensive malaria control. The two study areas represent two slightly different scenarios: in Botswana the analysis period covered the steady introduction of comprehensive control, while in South Africa the study period covered a time when effective control was being threatened by the spread of insecticide- and drug resistance, and the general health of the population was increasingly affected by the HIV pandemic.

The main findings were the following:

- It was possible to estimate the distribution of malaria in Africa fairly successfully from long term mean climate data via simple GIS methods. The model compared well with contemporary malaria data and historical ‘expert opinion’ maps, excepting small-scale ecological anomalies. The model provided a numerical basis for further refinement and prediction of the impact of climate change on transmission. Together with

population, morbidity and mortality data, it has provided a fundamental tool for strategic control of malaria.

- In Botswana the spatial variation in childhood malaria prevalence, prior to intense comprehensive control, was significantly associated with underlying environmental factors. It could be predicted and mapped using only three environmental predictors, namely summer rainfall, mean annual temperature and altitude. After starting with a long list of candidate variables, this parsimonious model was achieved by applying a systematic and repeatable staged variable exclusion procedure that included a spatial analysis. All this was accomplished using general-purpose statistical software.
- In the presence of intense control, the spatial and temporal variability in childhood malaria prevalence in Botswana could no longer be explained by variation in climate. The effects of malaria control and good access to treatment seem to have replaced climate as the main determinant of prevalence. This also suggests that prevalence, a less direct measure of transmission rate, is more prone to non-climatic effects than incidence rate.
- Total population malaria incidence in KwaZulu-Natal, the highest risk province of South Africa, remained significantly influenced by climate over a 30 year period, even in the presence of intense control. The inter-annual variation in case numbers were significantly associated with several climate variables, mainly mean annual daily temperatures and summer rainfall. However, climate factors did not explain the longer term total incidence rates.
- The longer term trends in total malaria incidence in KwaZulu-Natal province, over the same 30 years period, were significantly associated with the spread of anti-malarial drug resistance and HIV prevalence. Cross-border movements of people, agricultural activities and emergence of insecticide resistance also affected the level of malaria

transmission at certain periods and to some degree, but this could not be formally quantified.

- When considering malaria incidence in three malarious provinces of South Africa at a sub-provincial level, the observed temporal and spatial variation could largely be explained by available weather and drug-resistance data. However, much of the region-specific temporal trends remained unexplained. Temporal forecasts, based on 18 years of data, predicted for six years for six regions, were not very accurate and lacked precision. It seems that the interplay of climatic and non-climatic factors in the South African context is too complex to allow forecasts that are suitable for decision-making at the provincial level.
- The findings of this thesis emphasize that in addition to shorter-term variation, which seems to be driven by climate in many cases, malaria transmission is largely determined by non-climatic factors in southern Africa. This appears to be particularly true where the natural malaria endemicity has been modified by control interventions. As the drive to control malaria in Africa continues and intensifies, the need for long-term surveillance of not merely malaria transmission, but also of the coverage and effectiveness of control interventions, will grow.

Zusammenfassung

Die Verbreitung der Malaria wird von den Beziehungen zwischen dem *Plasmodium* Parasit, der *Anopheles* Mücke, und dem menschlichen Wirt bestimmt. Die Lebenszyklen der drei Spezies, und folglich ihre Interaktionen, sind grossen räumlichen und zeitlichen Veränderungen unterworfen. Auch Risikofaktoren und andere Determinanten, die sich auf diese Beziehungen auswirken, ändern sich je nach Ort und Zeit. Daher ist die mathematische Beschreibung und die Voraussage lokaler Infektionsraten ein schwieriges Unterfangen. Für eine zeitlich abgestimmte und gezielte Bekämpfung der Malaria ist jedoch ein gutes Verständnis dieser Variabilität und ihrer Ursachen äußerst wichtig. In den Randzonen der Malariagebiete Afrikas, welche hauptsächlich wiederkehrenden Epidemien ausgesetzt sind, wäre es für die verantwortlichen Instanzen von grossem Nutzen, wenn Zeiten erhöhter Infektionsgefahr vorausgesagt werden könnten.

Diese Dissertation befasst sich mit den zeitlichen und räumlichen Auswirkungen diverser Umgebungsfaktoren auf die Malariaübertragung. Es handelt sich um klimatische und außerklimatische Faktoren, wobei es sich bei letzteren in erster Linie um Malariabekämpfungsmassnamen handelt. In sechs getrennten Studien werden mehrere Datensätze mit Hilfe verschiedener methodischer Ansätze ausgewertet, um die räumlichen und zeitlichen Aspekte der Malariaverbreitung auf verschiedenen Ebenen zu untersuchen.

Auf kontinentaler Ebene wird die Malariaverbreitung in Afrika als Funktion von Klimafaktoren im Raster-GIS-Verfahren modelliert.

Auf nationaler Ebene werden ältere Daten aus Botswana verwendet, um räumliche Veränderungen der Malariaprävalenz in Abwesenheit einer umfassenden Malariakontrolle als Funktion umgebungsbedingter Determinanten zu modellieren. Zum Vergleich werden räumliche Veränderungen der Malariaprävalenz über eine Zeitspanne von 24 Jahren bei intensiver Malariakontrolle in ähnlicher Weise betrachtet.

Auf sub-nationaler Ebene wird Südafrika als Beispiel verwendet. Die jährlichen Schwankungen der Malariainzidenzrate in der Provinz mit der höchsten Infektionsgefahr werden in Bezug auf eine mögliche Beziehung zu klimatischen und außerklimatischen Faktoren untersucht. Zuletzt werden vorhandene Daten verwendet um die räumlichen Veränderungen der Inzidenzrate auf sub-provinzialer Ebene über eine Zeitspanne von mehreren Jahren zu analysieren - wiederum in Bezug auf klimatische und außerklimatische Determinanten.

Beide Studiengebiete (Botswana und Südafrika) liegen in den Randzonen der afrikanischen Malariagebiete und sind jahreszeitlichen Schwankungen der Infektionsgefahr sowie zeitweiligen Epidemien ausgesetzt. Beide Länder zeichnen sich durch intensive Anstrengungen hinsichtlich der Bekämpfung der Malaria aus. Allerdings unterscheiden sich die beiden Datensätze in einem wesentlichen Punkt: Im jeweils betrachteten Zeitraum wurden in Botswana zunehmend umfassendere Bekämpfungsmaßnahmen eingeführt, während in Südafrika die Effektivität der bestehenden Maßnahmen abnahm. Dies geschah vor allem aufgrund zunehmender Resistenzen gegen Insektizide und Malariamedikamente sowie einer erhöhten Anfälligkeit der Bevölkerung infolge der HIV-Pandemie.

Folgende Ergebnisse wurden erzielt:

- Die Malariaverbreitung in Afrika konnte anhand von im GIS-Verfahren analysierten Klimadaten erfolgreich vorausgesagt werden. Das Modell hält, abgesehen von einigen kleinen ökologischen Anomalien, einem Vergleich mit bisherigen Daten und gebräuchlichen, nach fundierten Vermutungen zusammengestellten Karten gut stand. Es bietet eine quantitative Grundlage für weitere Verbesserungen und ermöglicht eine Voraussage der Wirkung veränderter klimatischer Umstände auf die Malariaübertragung. Zusammen mit den vorhandenen Bevölkerungs-, Morbiditäts- und Sterberate-bezogenen Daten kann es deshalb zur Planung einer strategischen Malariabekämpfung benutzt werden.
- In Botswana stand vor der Einführung intensiver Malariabekämpfungsmaßnahmen die Veränderung der Malariaprävalenz bei Kindern in einer direkten Beziehung zu den herrschenden Umgebungsfaktoren. Sie konnte anhand von drei Umgebungsvariablen, nämlich dem Sommerniederschlag, der jährlichen Durchschnittstemperatur und der Höhe über dem Meeresspiegel ermittelt und kartografiert werden. Aus einer langen Liste von möglichen Variablenkombinationen konnte dieses überschaubare Modell mittels eines systematischen und wiederholt angewendeten "staged variable selection"-Verfahrens, das zuletzt eine räumliche Analyse beinhaltete, unter Verwendung normaler Statistik-Software gefunden werden.
- Nach der Einführung intensiver Malariabekämpfungsmassnamen kann die zeitliche und räumliche Veränderung der Malariaprävalenz bei Kindern in Botswana jedoch nicht mehr anhand von Klimafaktoren erklärt werden. Die Auswirkungen der Malariabekämpfung und besserer medizinischer Behandlung haben nun das Klima als wichtigste Prävalenzdeterminante ersetzt. Dies deutet auch darauf hin, dass die Prävalenz, ein indirektes Maß für die Infektionsrate, stärker von außerklimatischen Faktoren beeinflusst wird als die Inzidenzrate.

- Die Malariainzidenzrate der Bevölkerung KwaZulu-Natals, der südafrikanischen Provinz mit der höchsten Infektionsgefahr, stand trotz intensiver Malariabekämpfung über eine Zeitspanne von 30 Jahren stets stark unter dem Einfluss klimatischer Faktoren. Die jährliche Schwankungen der Zahl der Krankheitsfälle stand stets in direktem Bezug zu mehreren klimabedingten Variablen, hauptsächlich den jährlichen Tagesdurchschnittstemperaturen und dem Sommerniederschlag. Diese klimabedingten Faktoren konnten jedoch nicht die langfristigen Veränderungen der Inzidenzraten erklären.
- Die langfristige Tendenz der Malariainzidenzrate in KwaZulu-Natal im Laufe der erwähnten 30 Jahre steht in starker Beziehung zu zunehmender Resistenz gegen Malariamedikamente und der steigenden HIV-Prävalenz. Auch häufige Grenzübertritte aus und nach Mosambik und Swasiland, landwirtschaftliche Aktivitäten und Resistenz gegen Insektizide beeinflussen die Infektionsrate zu gewissen Zeiten und verschieden stark. Letzteren Einflüsse konnten allerdings nicht quantifiziert werden.
- Betrachtet man die Malariainzidenz auf sub-provinzialer Ebene in sechs verschiedenen Gebieten innerhalb drei südafrikanischer Provinzen, und über eine Zeitspanne von 24 Jahren, kann man die gemessenen zeitlichen und räumlichen Veränderungen zum großen Teil anhand des Wetters und der Resistenz gegen Malariamedikamente erklären. Viele lokale zeitliche Tendenzen bleiben allerdings unerklärt. Die Daten der ersten 18 Jahre wurden als Grundlage für eine zeitliche Malariainzidenzvoraussage von sechs Jahren benutzt. Diese Voraussagen erwiesen sich jedoch als ungenau. Im Fall von Südafrika, wo wie gesagt intensive Malariabekämpfung stattfindet, scheint das Zusammenspiel der klimatischen und außerklimatischen Faktoren so komplex zu sein, dass zeitliche Voraussagen nicht mit einer für wirkungsvolle Planung notwendigen Genauigkeit gemacht werden können.

- Diese Dissertation kommt somit zum Schluss, dass, abgesehen von klimatisch bedingten kurzfristigen Variationen, die Malariaverbreitung im südlichen Afrika grossenteils von außerklimatischen Faktoren bestimmt wird. Das scheint in besonderer Weise dort zuzutreffen, wo die natürliche Malariaübertragung durch Bekämpfungsmaßnahmen unterbrochen wird. Je stärker das Bedürfnis nach Malariabekämpfung in Afrika wird, desto stärker wird auch die Notwendigkeit einer langfristigen Überwachung nicht nur der Malariaübertragungsrate, sondern auch der Flächendeckung und Effektivität verschiedener Bekämpfungsmethoden.

Abbreviations

ACT	Artemisenin combination therapy
ADF	Augmented Dickey Fuller test
AIC	Akaike information criterion
ARIMA	Auto-regressive integrated moving average model
CMAP	Climate Prediction Centre Merged Analysis of Precipitation
CQ	Chloroquine
CRU-MARA	Climate Research Unit climate data, commissioned by MARA project
CRU-TS2	Climate Research Unit climate time series 2
DDT	Dichloro-Diphenyl-Trichloroethane
DIC	Deviance information criterion
EIR	Entomological inoculation rate
ENSO	El Nino southern oscillation
GIS	Geographic information systems
GPCP	Global Precipitation Climatology Project
HIV	Human immune-deficiency virus
ICR	Infant conversion rate
IDRC	International Development Research Centre
IRS	Indoor residual spraying
ITN	Insecticide treated nets
KZN	KwaZulu-Natal province, South Africa
LP	Limpopo province, South Africa
LSDI	Lubombo Spatial Development Initiative
MARA/ARMA	Mapping Malaria Risk in Africa project
MCMC	Markov chain Monte Carlo

MIM	International Development Research Centre
MP	Mpumalanga province, South Africa
NDVI	Normalized difference vegetation index
RBM	Roll Back Malaria
R_o	Reproductive rate of disease
SP	Sulphadoxine-pyrimethamine
STI	Swiss Tropical Institute
TB	Tuberculosis
TDR	Speical Programme for Research and Training in Tropical Diseases
WHO	World Health Organization

List of Figures

List of Tables

Chapter 1

Introduction

Background

Human malaria is caused by a protozoan parasite of the genus *Plasmodium*, and is transmitted by mosquitoes of the genus *Anopheles*. Of the four main *Plasmodium* species infecting humans, most of the disease and death, especially in Africa, is caused by *P. falciparum* (Molineaux 1988). Effective prophylactic and curative drugs and drug combinations against the parasite are available, and transmission can be interrupted through indoor spraying of residual insecticides (IRS) and use of insecticide treated nets (ITN). Despite this malaria continues to be one of the dominant diseases affecting mankind (Murray & Lopez 1997).

Estimates of the annual number of deaths and clinical episodes that occur globally diverge; but Africa is known to carry by far the bulk of the burden: in 1995 Africa saw an estimated one million deaths and around 450 million clinical episodes. Malaria can cause lasting side-effects, which affect individual development, mainly through anaemia, neurological and physiological sequelae, as well as risk of infection with the human immune-deficiency virus (HIV) following blood transfusion (Snow *et al* 1999a). Malaria also retards economical and social development through effects such as reduced working hours due to sickness or attending to the sick, income spent on financing health care (Mills 1994), which in turn lead to impacts at national level because of massive health care budgets, reduced productivity of the work force, and so on. Malaria is estimated to have cost endemic countries in Africa 3% of their economic growth every year (Sachs & Malaney 2002).

The malaria life cycle

In the human host the parasite multiplies in two stages: first in the liver (the hepatic cycle), then in the blood (the erythrocytic cycle, Figure 1.1). Parasite sporozoites are injected into a human from the salivary gland of an infectious mosquito. They first travel to the liver where they invade liver cells and undergo a tremendous initial asexual hepatic multiplication phase. One parasite can multiply 10 000-fold in *P. vivax* to 30 000-fold in *P. falciparum* (Garnham 1988). When the liver cells rupture, free-floating merozoites are released into the blood stream, where they invade red blood cells. Now called trophozoites, the parasites feed on the red blood cells as they grow. The trophozoites again divide asexually, becoming schizonts. Each *P. falciparum* schizont produces 12 - 32 new merozoites, slightly less in other species. When the red blood cell ruptures, the merozoites are released into the blood stream and invade further red blood cells. This erythrocytic reproduction phase can quickly lead to high levels of parasitaemia. Parasite densities of over 50% have been recorded in *P. falciparum* infections (2 500 000 parasites per micro litre). Densities in the other three species rarely exceed 2% in *P. vivax* and *P. ovale* and 1% in *P. malariae* (Harinasuta & Bunnag 1988).

Soon after infection of the human host, some trophozoites, instead of dividing, start transforming into gametocytes. These male and female sexual cells may then be ingested by a feeding female mosquito. In the stomach of the mosquito the male gametocyte forms several sperm cells, which fertilize the female gametocyte. The fertilized egg becomes a mobile ookinete, which invades the body cavity of the mosquito, via the stomach wall, on which it settles, turning into an oocyst. The oocyst starts dividing, forming 80 000 to 10 000 sporozoites. The sporozoites then travel through the body fluids to the mosquito's salivary gland. From here they are injected into another human when the mosquito next feeds.

These complex cycles can be summarised into two components determining the malaria transmission rate: the passage of parasite forms from humans to mosquitoes (step A in Figure 1.2), and then from mosquitoes to further humans (step B). The reproductive rate (R_o) of the infection is the product of the rates of steps A and B (in this example $3 \times 2 = 6$).

Determinants of malaria transmission

As a vector-borne disease malaria requires the presence of the human host, the anopheline mosquito vector and the plasmodial parasite. This triangular relationship is illustrated in Figure 1.3, which suggests that there are elements that relate to each organism individually, while other elements relate to each of three bi-lateral relationships, and others to the joint interaction of all three.

This picture is expanded in Figure 1.4 to reflect a range of measurements or indicators that are encountered in epidemiological literature on malaria and as they relate to the three-way relationship. Many of these form part of mathematical expressions of the transmission process (Ross 1911; Dietz 1988; Anderson & May 1991; Smith *et al* 2006).

Some important intrinsic and extrinsic determinants that affect the human, the mosquito vector and the parasite, as well as their various interaction, and ultimately the transmission of malaria, are shown in Figure 1.5 (Molineaux 1988; Mouchet *et al* 1998). Many of these have been the target of specific investigations.

Ultimately this neat three-way relationship, with all its determinants, gives rise to a complex web of cause-and-effects (Figure 1.6): new malaria infections (arguably the main point of interest) are the product of transmission which depends on parasite and vector development and interaction. Both being exotherms, their development rate depends on ambient

temperature. Mosquitoes also need breeding sites which are created by rain but also by other surface water from irrigation or perennial swamps for example. The nature of the breeding sites is further related to vegetation, soil quality, and hydrology. Rainfall and temperature determine humidity which affects vector survival. Transmission and incidence are furthermore affected by malaria control, through measures such as IRS, ITN and anti-malarial drugs, but the impact of these is modified in turn by insecticide and drug resistance. Housing, migration and other human behaviour patterns also strongly affect malaria transmission. Another important component in malaria transmission is immunity, which is affected by HIV, pregnancy, age, gender, other diseases, nutrition, and so on. In this way socio-economic factors also come into play.

Figure 1.6 is not an attempt to illustrate the exact nature of this web, but simply to highlight its minimal complexity and some of the more important determinants. Any attempts to investigate the determinants of malaria transmission need to be cognisant of the complexity involved, to avoid over-simplification and incorrect conclusions.

Apart from the causal links, the relationships illustrated in Figures 1.3 to 1.6 also point towards the temporal element of malaria transmission. The human life cycle is a matter of years, the mosquito life cycle a matter of days and weeks, while the interaction between humans and mosquitoes waxes and wanes over weeks and months. The parasite life cycle plays out over hours and days in the human, and days to weeks in the mosquito, while the interaction with the human host develops over months and years.

At no point are the interactions between human, parasite and vector, nor indeed the transmission and expression of malaria and the determinants themselves, entirely homogenous (Trape & Rogier 1996; Mouchet *et al* 1998; Hay *et al* 2000c; Mbogo *et al* 2003). Variability is

observed at every level, and itself varies in dimension, magnitude and scale. In the temporal dimension, greater or smaller changes in malaria transmission, infection and morbidity can be observed over hours and centuries and everything in-between; in the spatial dimension variability exists from the global down to the individual, even the cellular level.

Temporal variation

Malaria transmission varies greatly over time. Up until the early 20th century, indigenous malaria occurred across southern and central Europe, the Middle-East and Asia, as far North as Scandinavia and Siberia, over large areas in North and South America, most of Africa and in northern Australia (Lysenko & Semashko 1968). Since then malaria has contracted dramatically (Hay *et al* 2004), largely due to active control, industrialization, urbanization and modern medicine. Conversely, malaria resurgence has occurred, the likely result of interruption of control activities, threats to effective control, such as drug and insecticide resistance, operational problems, deforestation, large-scale migration and displacement of communities, breakdown of public health, political and industrial factors, and various other factors (Sharma 1996; Sleigh *et al* 1998; Garg *et al* 1999; Kamat 2000; Guerra *et al* 2006).

Much has also been written on the potential effects of long-term climate change and global warming on malaria (McMichael & Martens 1995; Martens *et al* 1995a; Patz & Kovats 2002; Tanser *et al* 2003). It has been argued strongly that warmer temperatures should render previously cooler latitudes and altitudes, where malaria was uncommon or absent, suitable for more frequent transmission, or that warmer and wetter conditions could increase the duration of the transmission season. However, the exact effect of climate change, or the degree to which changes in malaria can be explained by climate change, are disputed (Crabb 2002; Hay *et al* 2002a; Hay *et al* 2002b). Reiter (2000) provides a provocative discussion of malaria in Europe in the middle ages. Despite a “little ice age” in the late 16th and the 17th century,

malaria appears to have continued unabated at the time. Conversely, the greatest advances towards eradication of malaria over large parts of Eurasia and the Americas were achieved during the warmest century since medieval times.

Attempts have been made to link temporal changes in malaria observed over several years and decades, to large-scale climatological variation such as the El Niņo Southern Oscillation (ENSO) phenomenon (Kovats 2000; Kovats *et al* 2003). ENSO refers to a periodic though irregular inversion in the difference between East and West Pacific sea surface temperatures, which appears to be associated with climatic effects across the globe. Specific extreme ENSO events (El Niņo / La Niņa) have coincided with - and been blamed for - occasional floods and droughts that led to malaria epidemics in East Africa (Lindblade *et al* 1999; Kilian *et al* 1999) and southern Africa (Thomson *et al* 2005; Mabaso *et al* 2006a) and in Madagascar (Bouma 2003) for example.

An important aspect of inter-annual variation of transmission is the periodic occurrence of epidemics (Gill 1938). Epidemics can be wide-spread, as the epidemics in 1996 (Anon. 1996) across southern Africa, and in 1997 in eastern Africa (Myers *et al* 2000). These were marked by severe illness and many deaths. The cause and predictability of epidemics is of major interest for health service management (Onori & Grab 1980; Nájera *et al* 1998). Unfortunately progress in this area has been limited (Anon. 2003b), partly because of a dearth of resources, capacity and long-term data, but probably also because malaria transmission is such a multi-factorial problem that largely defies prediction.

On the intra-annual scale, seasonal periodicity in rainfall and temperature leads to seasonal fluctuation in vector populations, parasite development rates and malaria transmission. For example, major differences in man biting rates were recorded in Garki, Nigeria, between the

wet and dry seasons, which were associated with different age prevalence curves of malaria infection (Molineaux & Gramiccia 1980). Tanser and colleagues developed a seasonality model that distinguished between the extreme seasonality along the fringe of distribution and the more perennial conditions around the equator (Tanser *et al* 2003). Mabaso and colleagues recently investigated in more detail the degree of seasonality of incidence (Mabaso *et al* 2006b) and the entomological inoculation rates (EIR) (Mabaso *et al* 2007) in different parts of the African continent, with respect to climate seasonality.

Availability of breeding sites, particularly temporary ones resulting from punctuated rainfall events, can vary on a monthly and weekly scale. Agricultural practice, such as the flooding and draining of rice paddies (Dolo *et al* 1997), also play an important seasonal role in vector breeding.

On a daily and hourly scale, minimum and maximum ambient temperatures can place powerful limitations on parasite (Detinova 1962) and vector survival (Jepson *et al* 1947). Mosquito vectors show a diurnal cycle in biting habits with the main African vectors biting during the hours between midnight and dawn (Gillies 1988). The activity of parasite stages inside the human host can also be observed on an hourly basis. Peripheral parasitaemia fluctuates in 3-4 day cycles (Harinasuta & Bunnag 1988), causing periods of fever occurring every three or four days, which led to the names “tertian” and “quartern” malaria. The fluctuations in circulating parasites within the human host affects the chance of transmission to feeding vectors on an hourly and daily basis, and is one of the parameters feeding into a complex mathematical model of malaria transmission (Smith *et al* 2006).

Geographical distribution

At the same time, malaria transmission intensity varies on different spatial scales. At the global level malaria occurs predominantly in the warm humid equatorial regions, where conditions allow for the timely development of both parasite and vector. The global distribution of the principal vectors of malaria (Gwadz & Collins 1996) is an important factor in malaria endemicity as well as the success of its control. Africa, which carries the greatest burden of malaria, is also the continent where control is most difficult, due to a range of factors including several highly effective and anthropophilic vector species, a predominance of the most severe malaria species, *P. falciparum*, enormous areas that are both highly endemic and densely populated, weak and unstable economies and health structures, and so on (Coluzzi 1999).

Southern Africa (roughly 10 to 30 degrees South) is marked by strong seasonality in malaria transmission, becoming more pronounced towards the fringe of distribution. Transmission in winter is limited by a lack of rain, particularly in the dry South-western areas, as well as by temperature. The combination of high latitudes and relatively high altitudes on the southern African plateau mean that temperatures in winter are too cold to sustain vector populations and / or parasite development. The region comprises many areas of unstable transmission, as well as areas of high endemicity, but most of the region is vulnerable to malaria epidemics to some extent.

Heterogeneity in the spatial distribution of malaria transmission, at increasingly localized scales, was illustrated in South Africa (Hay *et al* 2000b). At the national level climatic effects are still important, as well as control activities within the country and in neighbouring countries. Below this scale human migration, the placement of roads and villages, etc come into play. The location of people and houses with respect to breeding sites as well as

randomness, affect the distribution of individual cases at the lowest scale. Local differences in vector density (Smith *et al* 1995), prevalence (Thompson *et al* 1997) and incidence (Trape *et al* 1993) with respect to vector breeding sites have been illustrated.

Malaria control

As much as the global distribution of malaria is affected by human anti-malarial interventions, the control of malaria also needs to take into account temporal and geographical patterns. After abandoning the eradication campaigns of the 1950's and 60's the World Health Organization (WHO) recommended particular focus on early diagnosis and prompt treatment, selective and sustainable preventive measures, detecting, containing and preventing epidemics and regular assessment of the in-country situations (Anon. 1993). In 1998 the WHO made a renewed commitment to address malaria, launching the “Roll Back Malaria” programme and assisting countries to plan and implement large-scale control, in line with previous guidelines, but aiming for major reductions in disease burden (Nabarro & Tayler 1998).

Large parts of southern Africa have benefited from extensive malaria control measures such as IRS, drug interventions and ITN. In Namibia, Botswana, Zimbabwe, South Africa and Swaziland, where malaria is already marginal, malaria control has been ongoing for decades, and malaria risk has been reduced to very low levels (Mabaso *et al* 2004). Though acknowledging the difficulty (or impossibility) of achieving eradication of malaria in the African context, several of these countries nevertheless pursue almost eradication-like control methods, aiming to detect and treat all carriers of the disease and to reduce transmission to a minimum. Further North and East, Angola, Zambia, Malawi and Mozambique, though experiencing seasonal transmission, still see high levels of endemicity and bear the full brunt of the disease. In these countries malaria is more wide-spread and control interventions have been more limited.

Since neither people, nor disease risks, nor health systems are evenly distributed, and since resources are limited, control measures have to target affected populations and high-priority areas first, to achieve maximum and equitable benefit. The timing of interventions also need to coincide with high risk periods. Furthermore, different control tools are appropriate in different transmission settings. All this requires relevant factual information (Bryce *et al* 1994; Snow *et al* 1996).

Surprisingly little information had been available and / or used to provide a rational basis for decision-making in the control and management of malaria in the past. Despite decades of malaria research in most African countries, hardly any of the results were being put to use. The MARA/ARMA (Mapping Malaria Risk in Africa) project was launched to attend to this need, by collating relevant information on malaria and, through analysis and spatial modelling, to translate it into an information tool useful for control related decision-making (Le Sueur *et al* 1997), and appropriate for low-end computer users.

Information that supports decision-making

Since malaria is unevenly distributed in time and space, you need to know where and when the risk of malaria is highest, in order to target and time interventions appropriately. The most direct measure of the risk of being infected in a certain time period is the EIR. Another is the infant conversion rate (ICR), a fairly pure measurement of incidence, or the number of infections acquired over a certain time period by a defined population (Molineaux *et al* 1988). Unfortunately the EIR is difficult to standardize and measuring ICR or incidence is time-consuming and resource-intensive, so these indicators are collected rather infrequently.

A commonly reported measure is the prevalence of patent infections. If associated with information on age, prevalence can give an indication of the level of transmission intensity

because acquired immunity results in different age prevalence curves under different infection rates. Because prevalence is easy to measure, and because methods are relatively standardized, it is frequently measured. It is for these reasons that the MARA project decided to focus its efforts on capturing historical prevalence data, as opposed to other transmission indicators (Le Sueur *et al* 1997). However, as much data as there might be, malaria risk can never be measured everywhere at all times, and a certain amount of prediction and interpolation is required, which justifies the second focus area of the MARA project on spatial modelling.

The MARA project has collated data from around 10 000 prevalence surveys carried out across sub-Saharan Africa, of which about 80% have been geo-referenced. Southern Africa has provided a large proportion of these. Several spatial models have also been produced, both theoretical (Craig *et al* 1999; Tanser *et al* 2003) and data-based. The data-based models were initially country-specific (Snow *et al* 1998; Kleinschmidt *et al* 2000), but were soon followed by a regional map for East Africa (Omumbo *et al* 2005) and West Africa (Kleinschmidt *et al* 2001a), which later included central Africa (Gemperli *et al* 2006a). These models were developed using successively more sophisticated statistical methodologies. Southern African data have been incorporated only fairly recently into the MARA database and no regional or country risk maps have yet been produced.

Aim

The overall aim of this thesis was to investigate, at various scales, the temporal and spatial effects of various environmental factors and malaria control on malaria transmission, using different methodological approaches.

Malaria distribution is investigated at the continental level, with focus on sub-Saharan Africa, and the southern part of the continent in particular. Sub-national incidence data from South

Africa and prevalence data from Botswana are examined. Both countries lie on the edge of malaria distribution, and both experience substantial temporal and spatial variation in climate and malaria.

The effect of malaria control is examined in two contrasting settings. In Botswana the analysis period covers the steady introduction of comprehensive control, while in South Africa, the thesis considers the threat to effective control presented by the spread of resistance to insecticides and anti-malarial drugs.

Objectives

The specific objectives of this thesis are:

1. to review the spatial and temporal effects of various determinants on the malaria transmission cycle, at different scales (Chapter 1);
2. at the continental scale, to model the distribution of malaria in sub-Saharan Africa as a factor of climate using raster GIS techniques, in order to describe the mean spatial distribution of endemic malaria, based on the theoretical suitability of long-term mean climate for malaria transmission, using fuzzy logic (Chapter 2)
3. at the national scale, to analyse point-referenced childhood prevalence data from Botswana to
    a. model the spatial variation in prevalence as a factor of environmental determinants, prior to comprehensive malaria control (Chapter 3);
    b. model the spatial and inter-annual variation in prevalence as a factor of climate, in the presence of intense control (Chapter 4);
4. at the sub-national scale, to analyse province and district-level total-population malaria incidence data, to

a. explore the inter-annual variation in malaria incidence in KwaZulu-Natal province over a 30 year period as a factor of climate (Chapter 5);

b. explore the inter-annual variation in malaria incidence in KwaZulu-Natal province over a 30 year period considering non-climatic factors (Chapter 6);

c. model the spatial and inter-annual variation in incidence in South Africa, at the sub-provincial level, based on climatic and non-climatic determinants (Chapter 7); and

5. to discuss the overall findings of this thesis, focussing on the various spatial and temporal aspects of the various determinants of malaria transmission, with respect to different scales, methodologies and applications (Chapter 8).

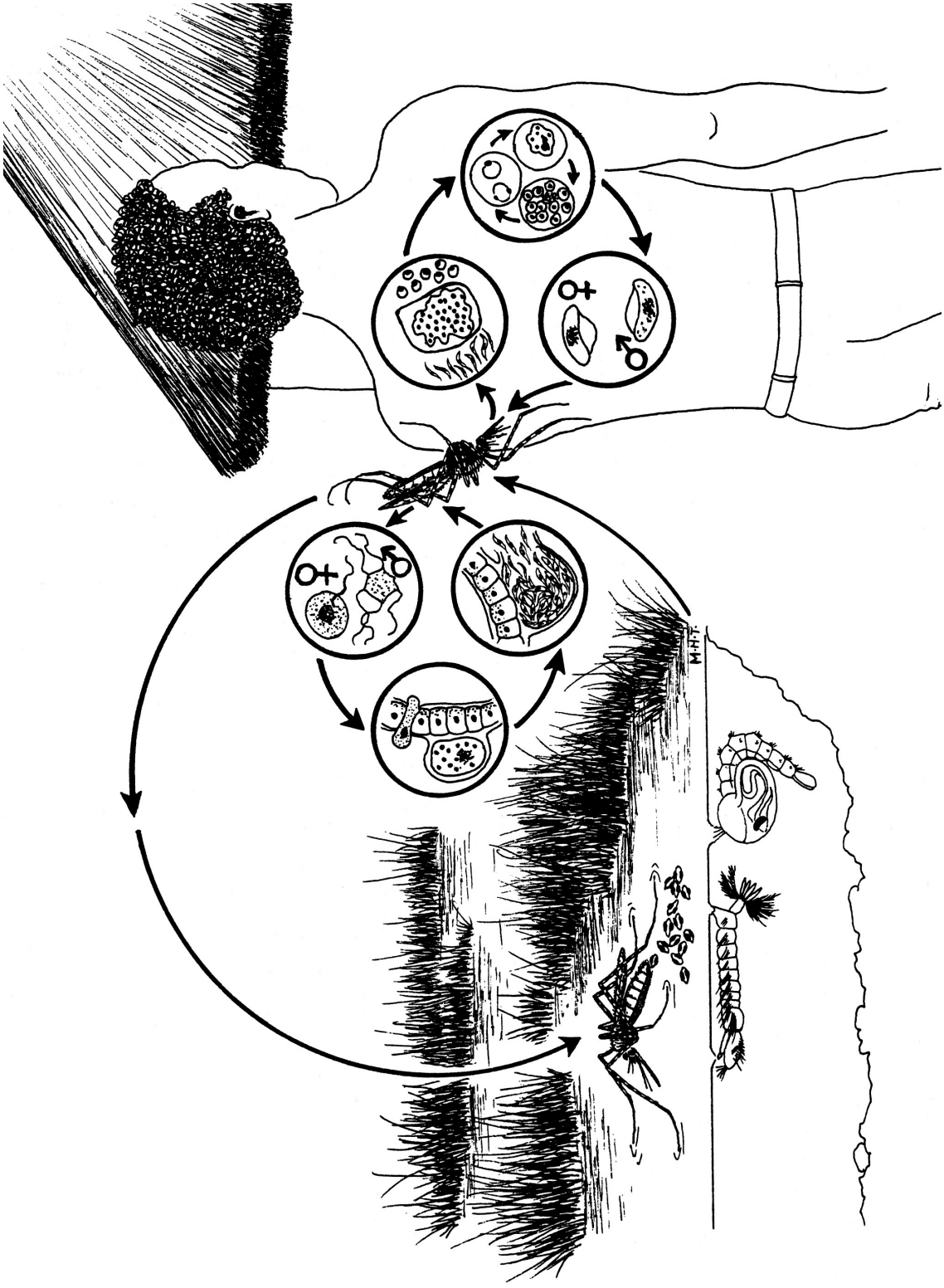


Figure 1.1 The *Plasmodium falciparum* life cycle (M. Craig, in Appleton *et al* 1995).

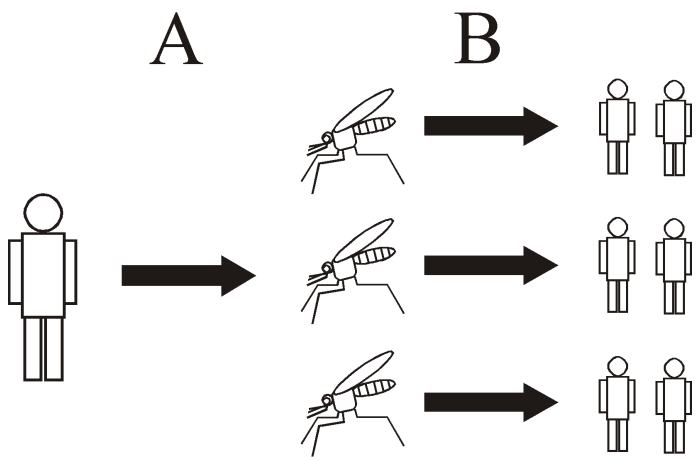


Figure 1.2 Malaria transmission from man to mosquito (A) and from mosquito to man (B).

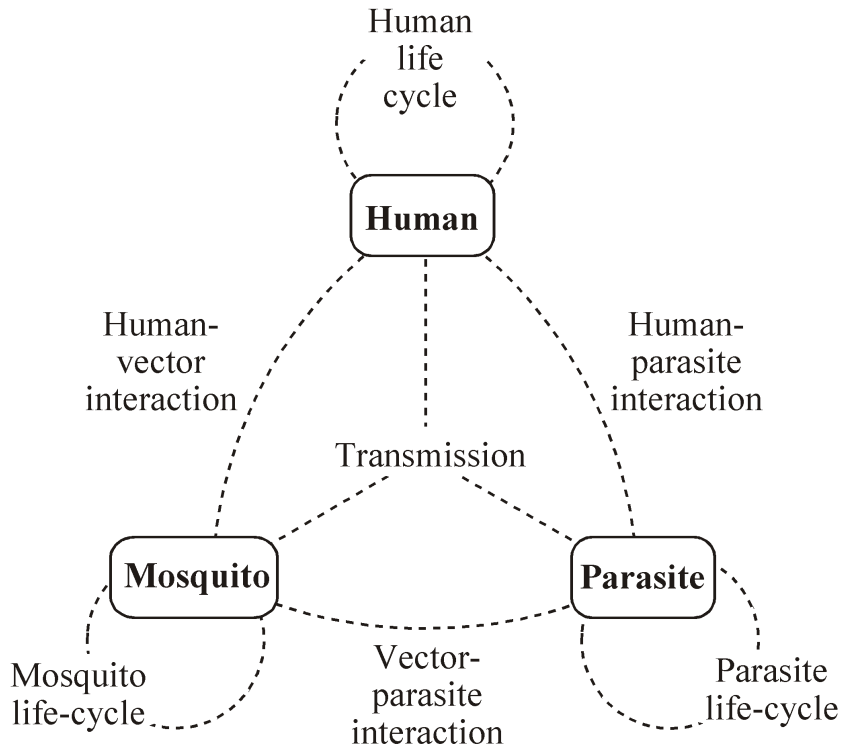


Figure 1.3 The three-way relationship between *Plasmodium*, the anopheline mosquito vector and the human host.

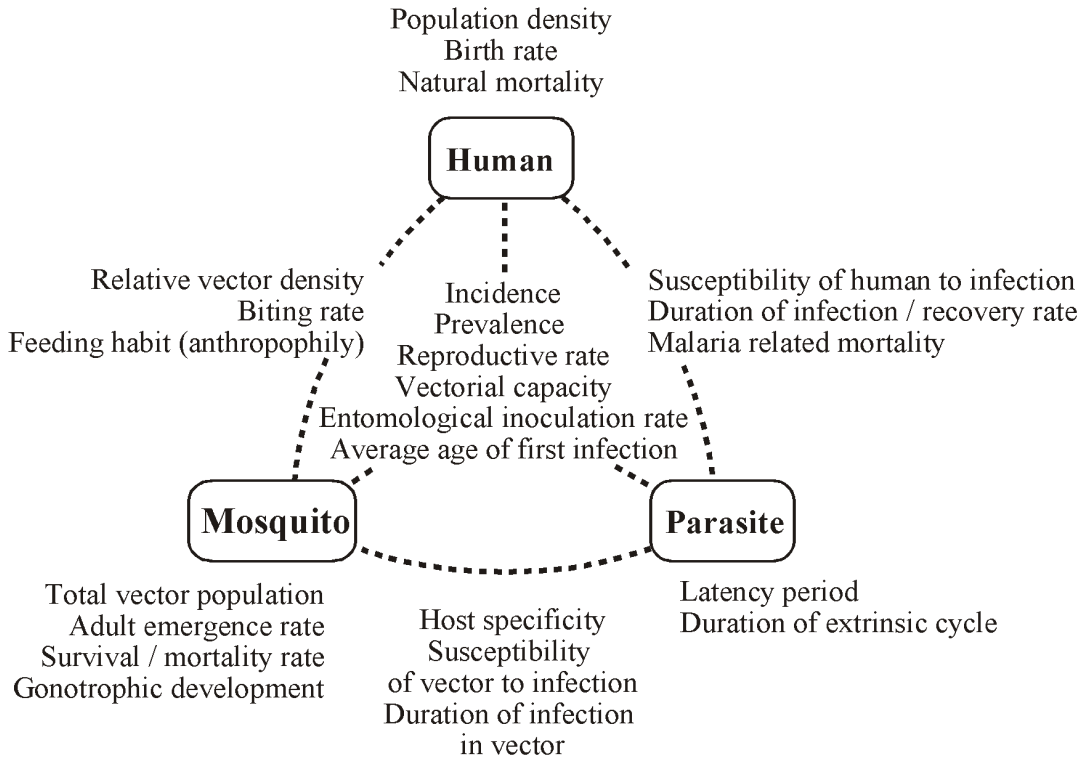


Figure 1.4 Epidemiological measurements with respect to the three-way relationship between *Plasmodium*, the mosquito vector and the human host.

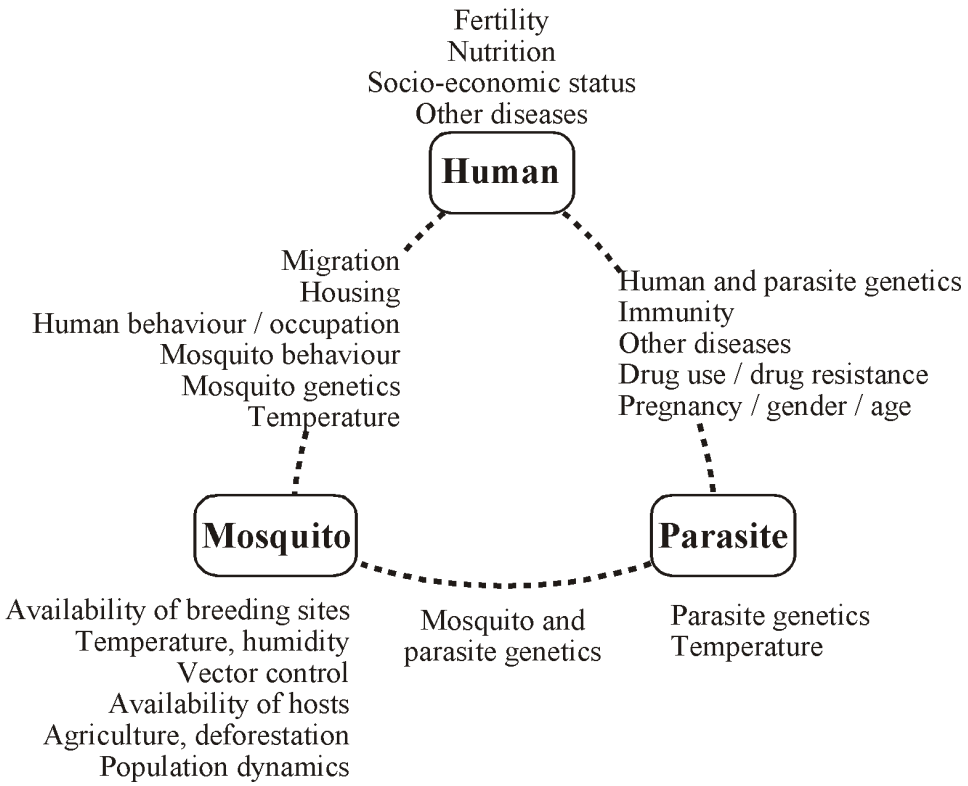


Figure 1.5 Various determinants and risk factors of malaria transmission, with respect to different interactions in the three-way relationship between *Plasmodium*, the mosquito vector and the human host.

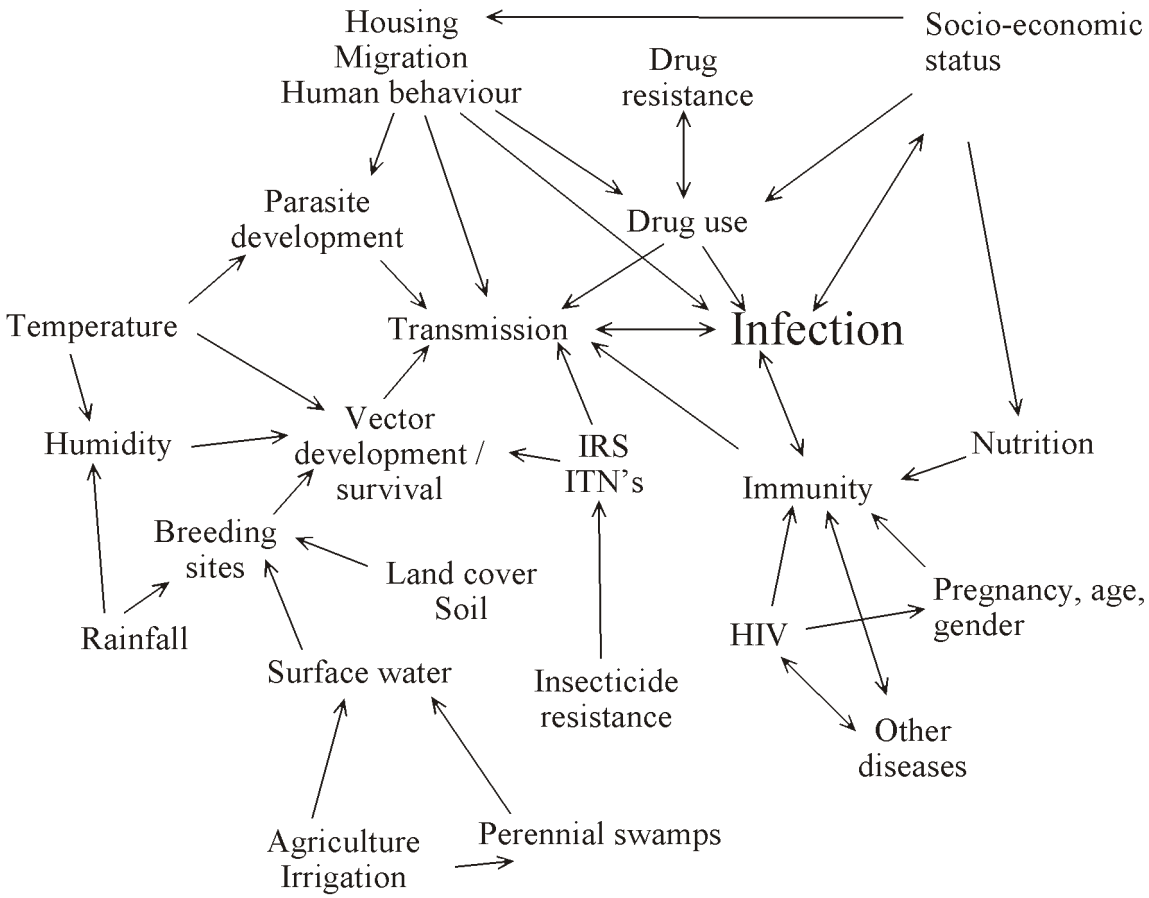


Figure 1.6 Interactions and causal links between important determinants of malaria transmission.

Chapter 2

A climate-based distribution model of malaria transmission in sub-Saharan Africa.

M.H. Craig[1] , R.W. Snow[2] and D. leSueur[1]

Affiliations:

[1] National Malaria Research Programme, South African Medical Research Council, PO Box 17120, Congella, 4013, South Africa. Tel: +27 31 251481, Fax: +27 31 251498, e-mail: craigm@mrc.ac.za

[2] Kenya Medical Research Institute / Wellcome Trust Collaborative Programme, PO Box 43640, Nairobi, Kenya.

Key words:

Malaria, *Plasmodium falciparum*, malaria distribution, geographic information systems, spatial modelling, climate, *Anopheles*

Publication status:

Parasitology Today ***15*** (3), 1999: 105-111.

Abstract

Malaria remains the single largest threat to child survival in sub-Saharan Africa and warrants long-term investment for control. Previous malaria distribution maps have been vague and arbitrary. Marlies Craig, Bob Snow and David le Sueur here describe a simple numerical approach to defining distribution of malaria transmission, based upon biological constraints of climate on parasite and vector development. The model compared well with contemporary field data and historical 'expert opinion' maps, excepting small-scale ecological anomalies. The model provides a numerical basis for further refinement and prediction of the impact of climate change on transmission. Together with population, morbidity and mortality data, the model provides a fundamental tool for strategic control of malaria.

Background

There have been several attempts to define the global and national distributions of malaria (Boyd 1949; Macdonald 1957; Lysenko & Semashko 1968). Common to all previous attempts at mapping malaria risk is that they derive from a combination of expert opinion, limited data and the use of crude geographical and climate iso-lines. None has a clear and reproducible numerical definition: consequently, their comparative value becomes limited.

Recently, the tools for the spatial representation of events have improved with the availability of affordable geographical information systems (GIS) software and large global data sets including climate, population, satellite imagery and topography. Consequently, the mapping of environmentally determined diseases is receiving a renewed interest (Gesler 1986; Sharp *et al* 1988; Smith *et al* 1995; Kitron *et al* 1996; Thomson *et al* 1996; Macé *et al* 1997; Malone *et al* 1997; Martens 1997; Hay *et al* 1998). It is into this milieus that the MARA/ARMA (Mapping Malaria Risk in Africa / Atlas du Risque de la Malaria en Afrique) (Le Sueur *et al* 1997)

project was born. One of the first objectives in MARA/ARMA was to find the limits of distribution of stable malaria transmission.

Transmission and distribution of vector-borne diseases are greatly influenced by environmental and climatic factors. An indicator of malaria stability is the reproduction rate (R_o) of the disease: where R_o is less than one, malaria is unstable, with a potential to die out, where R_o is greater than one, malaria is stable and likely to continue indefinitely. Vectorial capacity (Macdonald 1957; Garrett-Jones & Grab 1964), the main component in R_o, is strongly determined by climate. In this paper, the authors propose a fuzzy logic model of the distribution of stable malaria transmission in sub-Saharan Africa. The model is based on the effect of mean rainfall and temperature on the biology of malaria transmission. Even though the relationships between transmission potential and disease outcome are ill defined (Snow & Marsh 1998), Snow, Craig, Deichmann and le Sueur attempt, in the adjoining paper (Snow *et al* 1999b), to project burdens of malaria mortality for sub-Saharan Africa, using the model described here, in conjunction with selected mortality data.

Fuzzy logic

Defining the precise edges of distribution of malaria is difficult due to small-scale ecological variability and temporal changes in transmission risk. In reality there is a gradual, ill-defined transition from perennial to seasonal to epidemic to malaria-free regions, as well as from high to low transmission intensity. Malaria distribution is not definable either in space, since the edge of distribution is indistinct, or in time, since both intensity and distribution wax and wane with the natural periodicity of events. Predicting, for each point in space, the probability of transmission occurring or not occurring, is not possible, because many contributing factors, such as mosquito density, human activities, human and vector genetics, etc. are not measurable or available at the continental scale. Of the available data surfaces, we consider

climate to be the most important in limiting transmission and distribution of malaria at a large scale. Climate could be considered as either able or unable to sustain transmission. This would be a boolean situation, where climate is suitable (one) or unsuitable (zero). Defining boolean thresholds, above which the temperature-rainfall combination is considered suitable and where malaria is expected to occur, or below which malaria is expected not to occur, would be ignoring natural gradients and inherent uncertainty.

Fuzzy logic (Zadeh 1965) is an extension of boolean logic that deals with the concept of partial truth, or put differently, the extent to which a statement is true (fractions between zero and one): climate is completely suitable, completely unsuitable, or in-between, semi-suitable. While probability sets are fuzzy i.e. non-boolean, fuzzy sets are not probabilities, because they do not necessarily add up to one, as do probabilities. Any 0-1 curve, considered appropriate for the subject, may be applied. The type of curve chosen depends mostly on what and how much is known about the suitability gradient.

Continental climate

Continental monthly temperature and rainfall surfaces (Hutchinson *et al* 1995), essentially interpolated weather station data, were used to provide the climate data. They represent long-term mean monthly profiles, i.e. monthly means in the average year. Conceptually, regions can be defined as: (1) perennial - where conditions are always suitable for transmission; (2) seasonal - where conditions become suitable for a short season every year; (3) epidemic - where long-term variation in climate renders conditions suitable for transmission on an irregular basis (with a potential of epidemic malaria); and (4) malaria-free - where conditions are always unsuitable. Since inter-annual variation is not reflected in long-term mean climate data, epidemic zones are not detectable. Using this data set to predict regions of annual transmission would lead to an exclusion, at the fringe, of rare epidemic zones, but inclusion of

frequent epidemic zones. More finite data - in space and time - is required to define the epidemic zones and this is being addressed presently.

Temperature effects on transmission

The effects of temperature on the transmission cycle of the malaria parasite, *Plasmodium falciparum*, are manifold, but its specific effects on sporogonic duration (*n*) and mosquito survival (*p*) are the most important (Onori & Grab 1980; Molineaux 1988). The mathematical relationships are shown in Box 1.

The lower limit of temperature suitability is determined by the number of mosquitoes surviving the incubation period (p^n): while parasite development only ceases at 16°C, transmission below 18°C is unlikely because few adult mosquitoes survive the 56 days required for sporogony at that temperature, and because mosquito abundance is limited by long larval duration. At 22°C sporogony is completed in less than three weeks and mosquito survival is sufficiently high (15%) for the transmission cycle to be completed. Thus temperature below 18°C was considered unsuitable, and above 22°C, suitable for stable transmission.

The upper limit of temperature suitability is determined by vector survival, since sporogony takes less than a week. Temperatures of above 32°C have been reported to cause high vector population turnover, weak individuals and high mortality (Le Sueur 1991; Maharaj 1995). Thermal death for mosquitoes occurs around 40-42°C (Haddow 1943; Jepson *et al* 1947) and daily survival is zero at 40°C (Martens 1997).

In addition to average temperature, Leeson (1931) found that in Zimbabwe *An. gambiae s.l.* disappeared when absolute minimum air temperature in winter fell below 5°C, and de Meillon

(1934) found that in the old Transvaal province, South Africa (now Mpumalanga, Gauteng, North-West and Northern Province) vector distribution discontinued where areas experienced frost. Stuckenberg (1969) plotted effective temperature (an indicator that emphasizes the importance of summer temperature and length in terms of biological activity) against frost incidence in 84 weather stations. The highest effective temperature with at least one day of frost per annum was 16.4°C. In southern Africa the 16.4°C effective temperature iso-line compared well with the 5°C minimum temperature iso-line, the main differences occurring in parts of the Zimbabwean highlands and along a wide band across central Botswana. The 5°C minimum temperature iso-line was used here, bracketed on both sides by one degree, to account for uncertainty, so that 6°C and above was suitable, 4°C and below unsuitable.

Rainfall effects on transmission

The relationship between mosquito abundance and rainfall is complex and best studied when temperature is not limiting. Studies have demonstrated the association between *Anopheles gambiae s.l.* abundance and rainfall (Molineaux & Gramiccia 1980; Le Sueur 1991; Charlwood *et al* 1995) but a direct, predictable relationship does not exist. *Anopheles gambiae s.l.* are observed to breed more prolifically in temporary and turbid water bodies such as ones formed by rain (Gillies & de Meillon 1968; Le Sueur & Sharp 1988) while in permanent bodies predation becomes important (Christie 1958). *Anopheles funestus* in contrast prefer more permanent water bodies (Gillies & de Meillon 1968). However, both temporary and permanent water bodies are dependent on rain. Rain is also related to humidity and saturation deficit: factors that affect mosquito survival (Molineaux & Gramiccia 1980). There is good reason for using rainfall to indicate the probable presence of vectors, their survival and the potential for malaria transmission. Although it is known that flooding often causes destruction of breeding sites (Jepson *et al* 1947) and a temporary reduction of vectors, it never eliminates the vector, so that very high rainfall was still considered optimal for transmission. The amount

of monthly rain required was examined by extracting the climate patterns in regions where the status of malaria was known.

Diagnostic climate patterns

To examine the pattern of mean climate, as it relates to different epidemiological settings, monthly rainfall and temperature values were extracted from the climate data surfaces (Hutchinson *et al* 1995) for 20 different sites where malaria transmission has traditionally been regarded as perennial (annual, for more than six months), seasonal (annual, for less than six months), epidemic (transmission not recorded every year), and malaria-free (malaria never recorded). The most diagnostic examples are displayed in Box 2.

The examples confirm that the approximate temperature cut-off between epidemic and no-malaria zones is indeed around 18°C, and that 22°C allows stable transmission, while the difference between regions ‘c’ and ‘e’ (Box 2; Figures I and II, c and e) indicates a rainfall requirement for stable transmission of around 80 mm for at least five months.

The duration of the rainfall season is also important. In regions where temperature is high but rainfall limiting, such as at the fringes of the north African deserts, mosquito populations increase rapidly at the onset of rain, because of short developmental cycles. Consequently, three months of rain may be sufficient to constitute one transmission season. However, where temperature is limiting during the colder season, as is the case in large parts of southern Africa and highland areas, mosquito populations increase slowly at the onset of rain, with gradually rising temperatures, due to long developmental cycles. Parasite and vector development is slow, and favourable conditions need to last longer to provide a window of transmission. This is also illustrated by the extracted climate patterns: in Mali (Box 2; Figures I and II, d), where temperatures are always high, a three month window of high rainfall is sufficient for

transmission, whereas in southern and eastern Africa (Box 2; Figures I and II, a–c), suitable conditions need to persist for at least five months.

Constructing a fuzzy distribution model

The GIS raster software IDRISI and its FUZZY function was used to convert the climate data to climate suitability maps of fractions between zero (condition unsuitable = *U*) and one (conditions suitable = *S*). Initially a simple sigmoidal fuzzy membership curve was used, defined in IDRISI as:

$$y = \cos^2\left[\frac{x - U}{S - U} \times \frac{\pi}{2}\right] \quad \text{(4)}$$

where *y* is the fuzzy suitability of climate value *x*. In the decreasing curve, fuzzy membership is equal to *y*, in the increasing curve it is (1 - *y*). As outlined in the previous sections, for rainfall, *U* = 0, *S* = 80 mm per month; for average temperature *U* = 18, *S* = 22°C for the increasing curve and *S* = 32, *U* = 40°C for the decreasing curve. For winter minimum temperature (mean daily minimum of coldest month) *U* = 4, *S* = 6°C.

Since favourable temperature and rainfall conditions have to coincide temporally for transmission to occur, the 12 monthly fuzzy rain and temperature images were overlaid month by month. The minimum suitability rating was calculated at each point, according to whichever - rain or temperature - was more limiting. Furthermore, suitable conditions have to occur for a certain time window, constituting a transmission season, long enough for vector populations to increase and for the transmission cycle to be completed. In North Africa (>8° North) the highest value spanning any three-, and in the rest of Africa any five consecutive months, was calculated. To adjust the model for the effect of frost (Leeson 1931; de Meillon 1934), the fuzzy minimum winter temperature was overlaid, again calculating the minimum fuzzy value. The resulting model (Figure 2.1) shows the distribution of conditions

more or less suitable for stable malaria transmission, lasting for at least five consecutive months, or three in North Africa, in the average year.

Does the model agree with available data?

Comparing the model with historical maps and malaria case data in southern Africa (Figure 2.2) and in Kenya and Tanzania (Figure 2.3) the resemblance is striking. In southern Africa the edge of malaria distribution is well represented. The malaria-free East African highland regions (Figure 2.3) are also clearly reflected in the model. In Kenya the coastal and south-western endemic zones agree, as do the malaria near water regions, too dry to register as suitable in the model. Minor discrepancies are discussed in the figure captions.

It is remarkable how well a simple model such as this, driven by an understanding of the situation on the ground, approximates the edge of malaria distribution across the continent. Because we are looking at the *distribution* of *stable* malaria transmission, the edge of the suitable zone must be regarded as the lowest level of endemic malaria (hypo-endemic and / or strongly seasonal) where we expect to find substantial - not necessarily high - levels of transmission occurring every year. The situation within the suitable zone (fuzzy value 1) may vary from low to high transmission intensity, but this is not reflected in a distribution model. The situation outside the suitable zone (fuzzy values 0.9 to zero) reflects the gradient from stable to increasingly unstable transmission with lower and lower transmission intensity, until, at the outermost fringes, malaria becomes a sporadic, unpredictable event, subject to the chance influx of parasites in rare wet or warm years.

In Botswana 13 years of incidence data (Anon. 1998a) show that districts in the same fuzzy zone behave similarly from year to year in terms of actual numbers of cases recorded. Reported cases clearly decline from the three endemic districts in the North, to extremely low

numbers in the central district, where in four out of 13 years no cases were recorded at all. In a further five districts, malaria cases are reported in extremely rare years (pers. comm., David Rumisha, Ministry of Health, Botswana). The outlook of this model for public health applications is dealt with in the adjoining paper (Snow *et al* 1999b).

Around the equator, rainfall patterns are slightly to strongly bimodal, some regions receiving rain in two short, distinct seasons. The model described above required five *consecutive* months with a rainfall above 80 mm. We ran the model again, with the same fuzzy definitions, but instead of looking for *consecutive* suitable months, calculated the maximum fuzzy values persisting for five months in *total*. The difference between the two models was zero or negligible in most of Africa, except for parts of central, south-eastern and northern Kenya and with very small differences in Ethiopia, Somalia, southern Cameroun and along the northern Angolan coast. In all other areas a bimodal rainfall pattern did not affect the outcome of the model, and even the affected areas in Kenya are mostly dry and unstable, indicating that two short distinct rainy seasons are after all not sufficient for endemic malaria.

To refine the shape of the fuzzy curves, and the suitability cut-offs *S* and *U*, it may be necessary to distinguish between the north, where the limiting factor is rainfall only, from the rest of Africa, where the effect is a combination of rainfall and temperature. Equatorial regions, where diurnal and annual temperature range is low, and where temperature is limited by altitude, may also need to be differentiated from the South, where temperature range is great and minimum temperature plays an important part, and where temperature is limited largely by latitude. It is worth noting here that no true gold standard is available. Historical maps and limited long-term malaria records have to suffice for comparative purposes.

Modelling at different spatial levels

We have demonstrated that a simple climate-based model can be used to define the crude distribution of malaria transmission in Africa. This model functions at the continental level, a scale for which we believe the data sets and the methodological approach to be appropriate, but which will not take into account small scale anomalies which may affect distribution, such as rivers and flood-plains in areas of low rainfall, agricultural practise, deforestation, etc. It reflects a conservative estimate of distribution. The inclusion of other smaller scale data sets (hydrology, human activity, etc.) may allow more detailed predictions, but requires a different approach.

Thus we view the modelling of malaria in Africa as a four-tier approach: (1) the first level, at the continental scale, defines the broad distribution of disease based on climatic conditions in an average year; (2) the second level, at a sub-continental scale, refines the distribution at the periphery using annual data sets for higher temporal resolution, and takes into account differences between major malaria ecological zones; (3) the third level, at a regional or national scale, would involve relating parasite ratios to climate and other factors and to define the transmission intensity within a given zone of transmission ecology, such as perennial, seasonal or bi-seasonal transmission; and (4) the fourth level, at a scale of 30 km^2 and below, is a process which operates below the second administrative unit and seeks to define variation in transmission on a local scale. The lower one goes in scale, the more one is forced to consider whether the input required justifies the scale at which one is working and the meaning which one is drawing from the product.

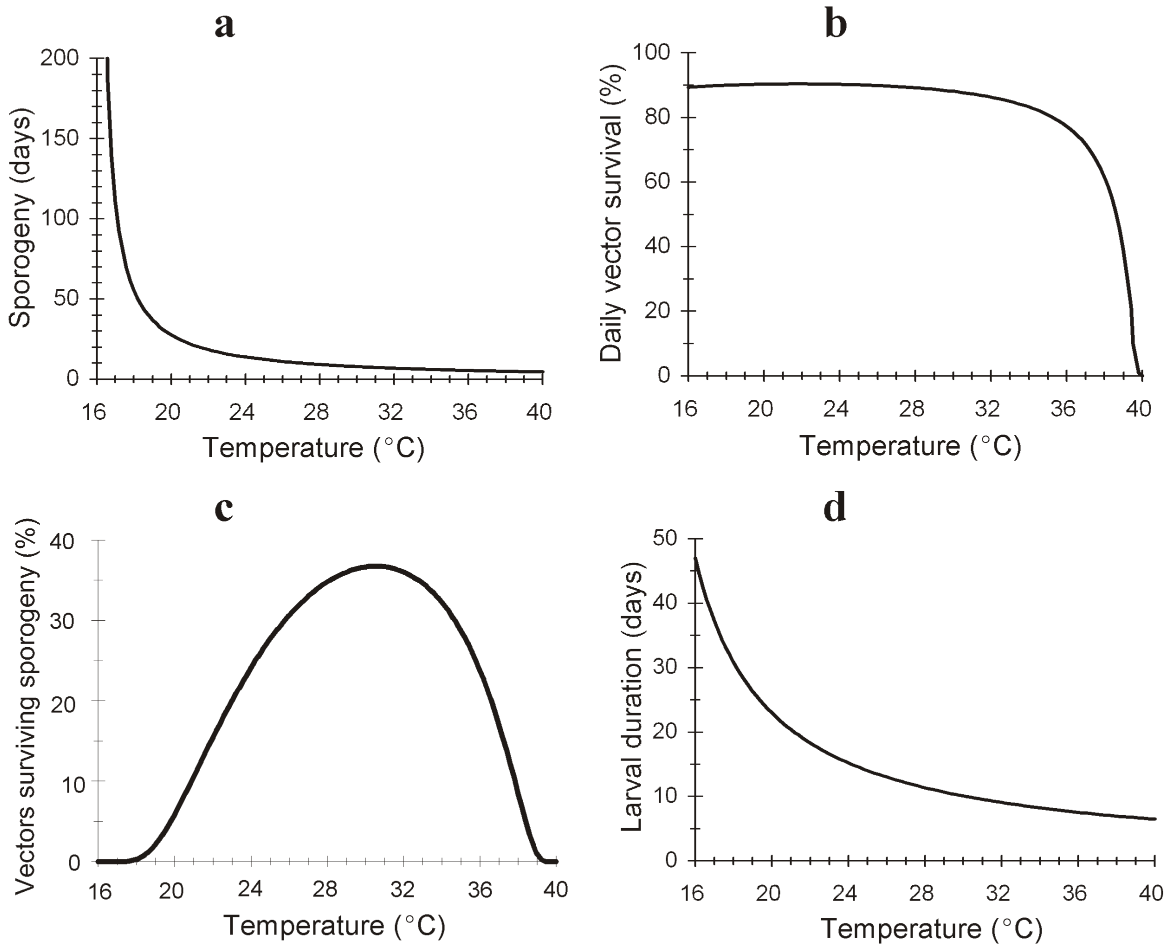
The model presented here, at the first level, introduces a new approach to numeric definition of continental malaria distribution. The main benefit lies in the fact that it can be repeated, evaluated and refined over time and can be mathematically manipulated in combination with

other data sets such as population (Deichmann 1997) to provide improved estimates of people at risk which is essential for prioritising health services (Snow *et al* 1999b). Such a model provides a baseline against which climate change scenarios (eg. global warming) can be evaluated in the long term. We are moving from the hypothetical to the quantifiable.

Acknowledgements

This publication is a product of the international MARA/ARMA (Mapping Malaria Risk in Africa / Atlas du Risque de la Malaria en Afrique) collaboration. We thank the South African Medical Research Council, the International Development Research Centre Canada (IDRC) and Wellcome Trust, UK for their financial support, and the Idrisi Project, Clark University, USA, in particular Nick Haan, for their input.

Box 1. Relationships between Temperature and Sporogonic Duration (*n*), Mosquito Survival (*p*) and Larval Duration



The effect of temperature on duration of the sporogonic cycle (*n*) in days is defined as (Macdonald 1957; Detinova 1962) (Figure a, above):

$$n = \frac{DD}{T - T_{\min}} \quad (1)$$

where *DD* is the total degree days for parasite development (111 for *P. falciparum*), *T* is the mean temperature in °C and T_{min} is the temperature at which parasite development ceases (16°C for *P. falciparum*). High temperature speeds up mosquito development (Jepson *et al* 1947) and decreases the interval between blood meals, leading to more frequent host-vector contact (Gillies & de Meillon 1968), but also reduces mosquito survival (Le Sueur 1991; Maharaj 1995). Daily mosquito survival (*p*) is defined by Martens (1997) as:

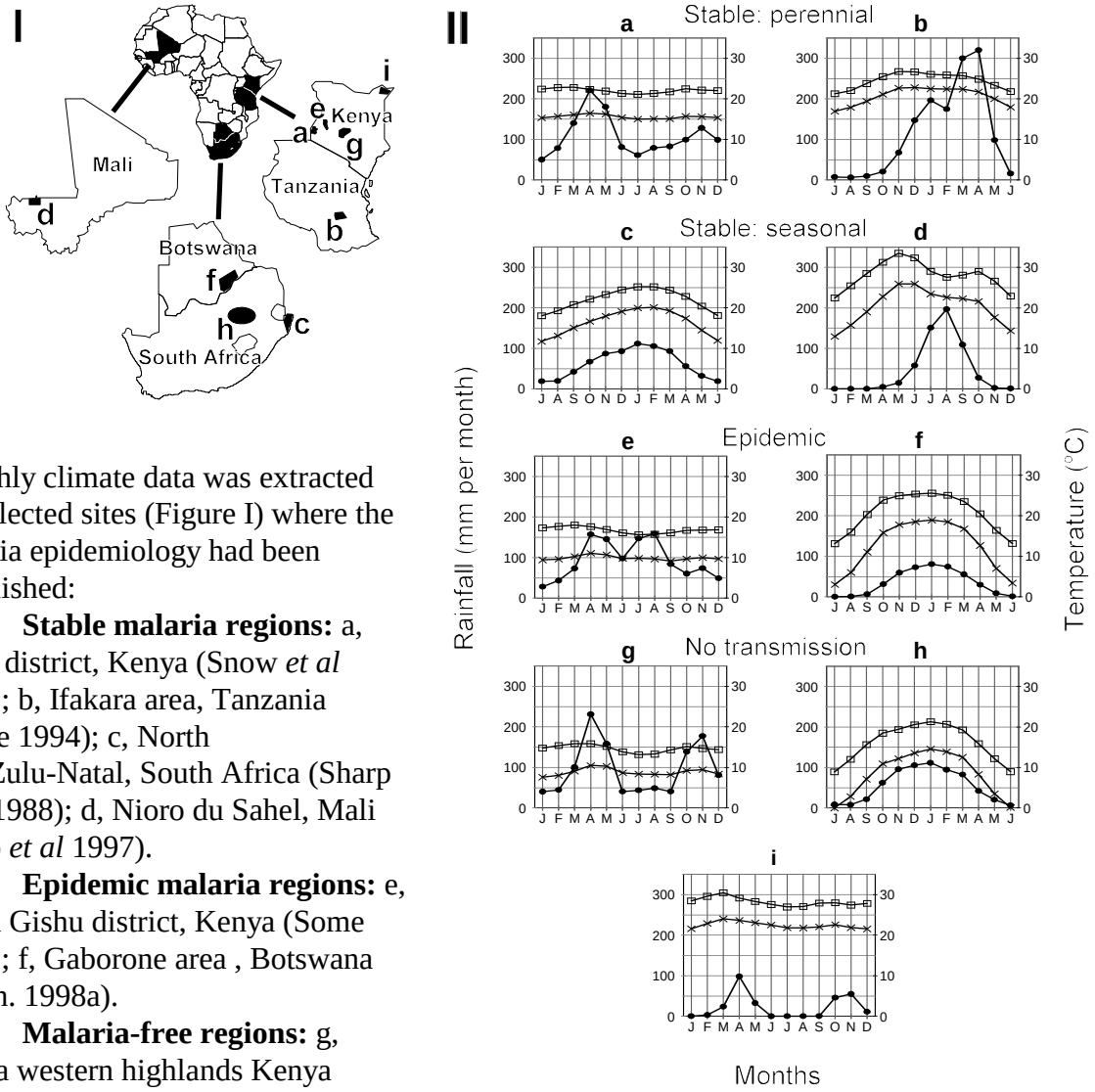
$$p = e^{-1/\left(-4.4+1.31T-0.03T^2\right)} \quad (2)$$

assuming constant humidity (b). Thus the combined effect of *n* and *p* (p^n) indicates the percentage of a vector cohort that survives after the full period required for completion of sporogony (c). Another effect of temperature, namely on larval duration (*ld*) in days, can be expressed as:

$$ld = 1 / (0.00554T - 0.06737) \quad (3)$$

and is shown in (d). The formula is derived from data published by le Sueur (1991).

Box 2: Temperature and rainfall profiles of selected regions



Monthly climate data was extracted for selected sites (Figure I) where the malaria epidemiology had been established:

Stable malaria regions: a, Siaya district, Kenya (Snow *et al* 1997); b, Ifakara area, Tanzania (Some 1994); c, North KwaZulu-Natal, South Africa (Sharp *et al* 1988); d, Nioro du Sahel, Mali (Dolo *et al* 1997).

Epidemic malaria regions: e, Uasin Gishu district, Kenya (Some 1994); f, Gaborone area , Botswana (Anon. 1998a).

Malaria-free regions: g, Kenya western highlands Kenya (Anon. 1994); h, Johannesburg area, South Africa (Anon. 1998c); i, North-East Kenya (Anon. 1994).

Graphs (Figure II) of long-term mean temperature (open squares), minimum temperature (crosses) and rainfall (closed circles) profiles by month.

The effect of mean temperature is illustrated in Figures (IIa), (IIe) and (IIg) where rainfall is high all year round: a constant temperature of 22°C in (IIa) is sufficient for perennial transmission, 18°C all year in (IIe) is too cold but epidemics occur in warmer years, while in (IIg), where mean temperature remains around 15°C, transmission never occurs. Similarly, mean temperatures in (IIc) and (IIh), which have the same seasonal rainfall pattern, suggest that seven months above 22°C allows seasonal transmission, while 6 months above 18°C does not.

In terms of rainfall, the difference between (IIc) and (IIf), which have similar mean temperature patterns, indicate that five months above 80mm rain is sufficient, but five months above 60mm is not. In (IIf) there is the added limiting effect of low minimum temperatures in winter, but rare epidemics do occur in particularly wet (Nelson 1959). It is further apparent from areas (IId) and (IIi) that where temperatures are high, one month of rain above 80mm is not sufficient for a transmission season, but that three months above 80mm is.

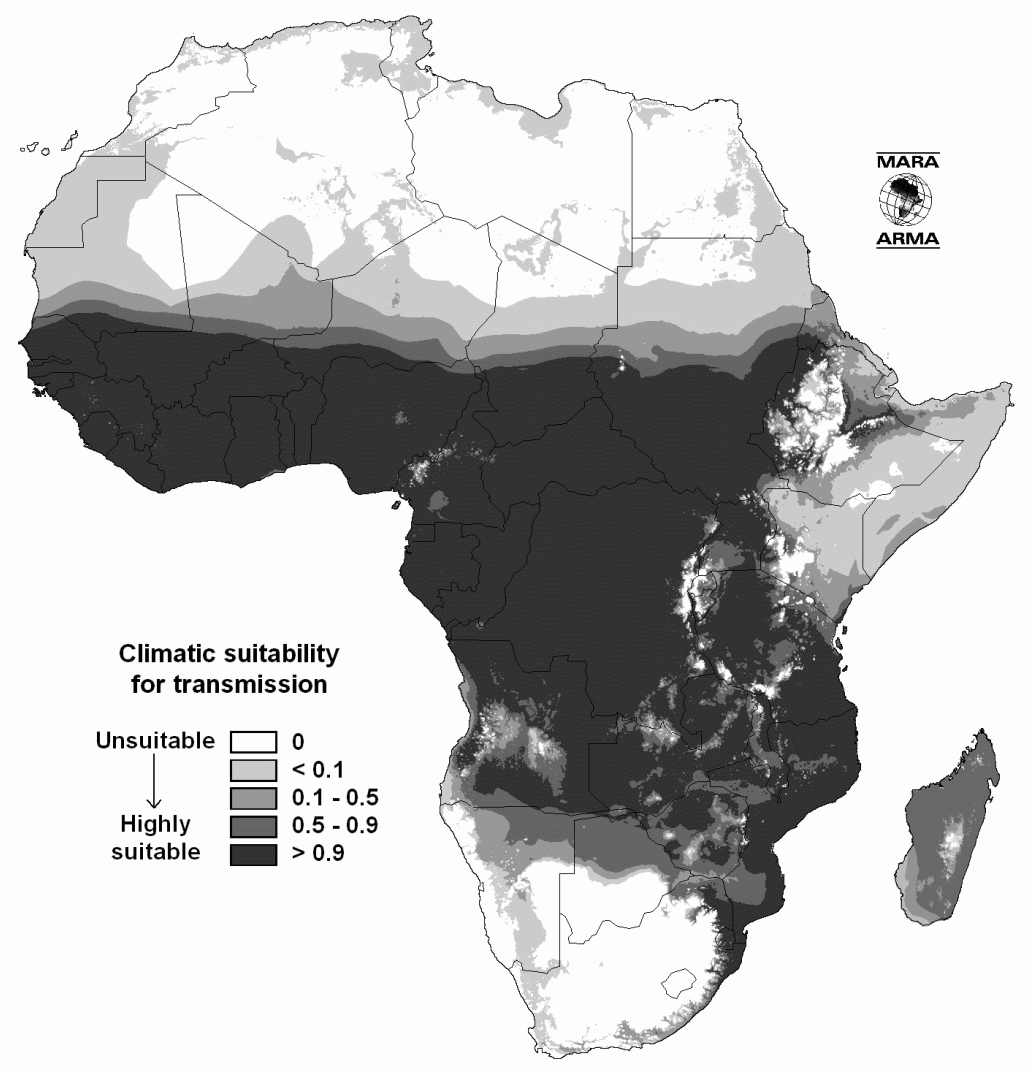


Figure 2.1 Fuzzy model for sub-Saharan Africa, showing the suitability of temperature and rainfall conditions for malaria transmission for any three consecutive months in north Africa and any five consecutive months in the rest of Africa. A value of 1 means that conditions in the average year are suitable, hence one could expect to find endemic malaria transmission (seasonal or perennial); a value of 0 means conditions are unsuitable in the average year, hence transmission should be absent or occur in rare epidemic episodes. Fractions from 0 to 1 indicate increasingly suitable climate, hence increased risk of regular transmission.

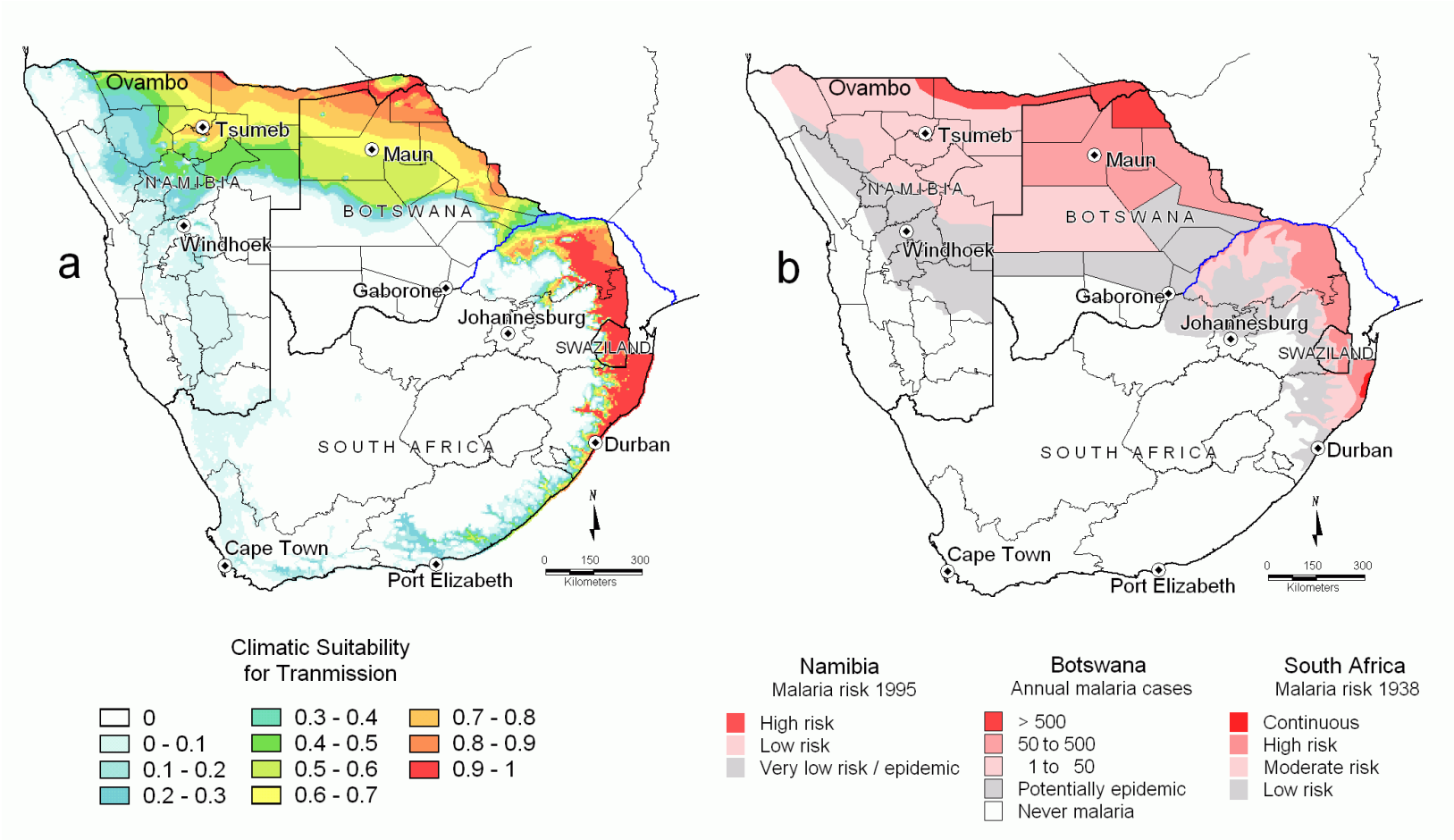


Figure 2.2 Comparison of the model with southern African distribution data. The climatic model: 0, unsuitable; 1, suitable (a). Malaria maps show malaria risk in 1995 in Namibia (Richard Kamwi, Ministry of Health and Social Services, Namibia, pers. commun.), 1938 in South Africa (Sharp *et al* 1988) and annual malaria case numbers per district in Botswana (b). The Namibia risk map is an expert opinion map, based on case data. The Botswana map is based on microscope-confirmed case data collected at district level from 1982 to 1994. Malaria case incidence in South Africa (not shown here) between 1987 and 1993 has been above 1% just north and east of Swaziland, and 1% or less elsewhere, but because malaria control has considerably reduced malaria in South Africa (Sharp *et al* 1988; Le Sueur *et al* 1993May) it is necessary to look at the historical map. Although the units in the maps of the three countries differ, agreement with the model is evident.

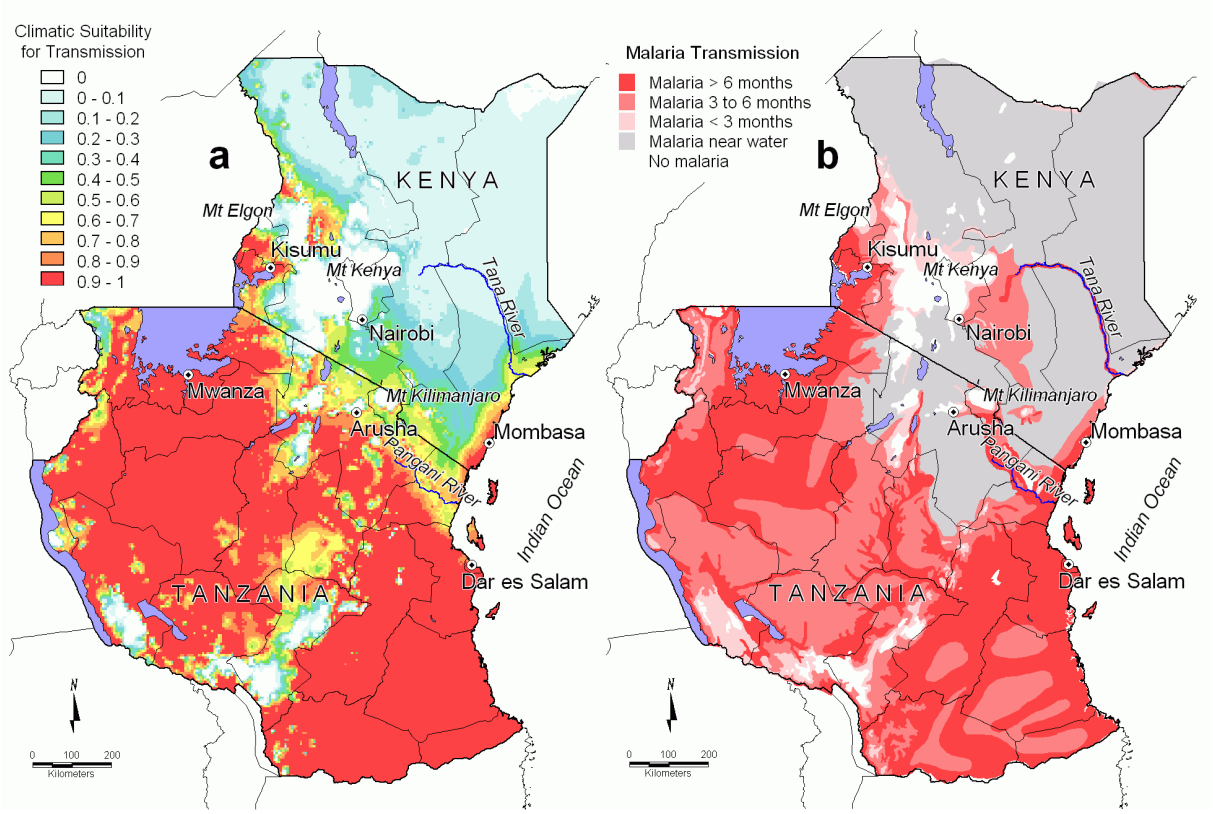


Figure 2.3 Comparison of the model with Kenyan and Tanzanian malaria maps. The climatic model: 0, unsuitable; 1, suitable (a). Malaria maps of Kenya (Nelson 1959) and Tanzania (Wilson 1956) are shown in (b). Agreement between the model and the historical maps is good. The area southeast of Mount Kenya and Nairobi was historically recorded malarious for three to six months, whereas the model predicts low climatic suitability. On closer inspection, this area is found to be flat, low-lying country, which may receive additional run-off water from the adjoining highlands; a high normalized difference vegetation index (NDVI, which is a measure of the amount of photosynthesis taking place, and hence relates to the moisture availability, saturation deficit, soil properties and humidity) indicates an abundance of water. Nevertheless, empirical data from this region (Omumbo *et al* 1998) suggest that malaria transmission is low and sporadic, and we have to question the accuracy of the historical map. The discrepancies in the Tana and Pangani (a and b above) river valleys, as well as the Limpopo river (Figure 2.2), are a result of the model using only rainfall to predict the presence of vectors so that, although rainfall may be low, breeding sites are available and humidity is high along banks and flood plains of major rivers.

Chapter 3

Developing a spatial-statistical model and map of historical malaria prevalence in Botswana using a staged variable selection procedure

Marlies H Craig [1,2], Brian L Sharp[1], Musawenkosi LH Mabaso[1,2] and Immo Kleinschmidt[1,3]

Affiliations:

[1] Malaria Research Programme, Medical Research Council, 491 Ridge Road, Overport, Durban, 4091, South Africa

[2] Swiss Tropical Institute, 57 Socinstrasse, Basel, BS 4002, Switzerland

[3] London School of Hygiene and Tropical Medicine, Keppel Street, London WC1E 7HT, United Kingdom

Keywords:

malaria; prevalence; maps; regression analysis; spatial analysis; models, statistical; environment; Bayes theorem

Publication status:

International Journal of Health Geographics **6**, 2007: 44

Abstract

Background

Several malaria risk maps have been developed in recent years, many from the prevalence of infection data collated by the MARA (Mapping Malaria Risk in Africa) project, and using various environmental data sets as predictors. Variable selection is a major obstacle due to analytical problems caused by over-fitting, confounding and non-independence in the data. Testing and comparing every combination of explanatory variables in a Bayesian spatial framework remains unfeasible for most researchers. The aim of this study was to develop a malaria risk map using a systematic and practicable variable selection process for spatial analysis and mapping of historical malaria risk in Botswana.

Results

Of 50 potential explanatory variables from eight environmental data themes, 42 were significantly associated with malaria prevalence in univariate logistic regression and were ranked by the Akaike Information Criterion. Those correlated with higher-ranking relatives of the same environmental theme, were temporarily excluded. The remaining 14 candidates were ranked by selection frequency after running automated step-wise selection procedures on 1000 bootstrap samples drawn from the data. A non-spatial multiple-variable model was developed through step-wise inclusion in order of selection frequency. Previously excluded variables were then re-evaluated for inclusion, using further step-wise bootstrap procedures, resulting in the exclusion of another variable. Finally a Bayesian geo-statistical model using Markov chain Monte Carlo simulation was fitted to the data, resulting in a final model of three predictor variables, namely summer rainfall, mean annual temperature and altitude. Each was independently and significantly associated with malaria prevalence after allowing for spatial correlation. This model was used to predict malaria prevalence at unobserved locations, producing a smooth risk map for the whole country.

Conclusions

We have produced a highly plausible and parsimonious model of historical malaria risk for Botswana from point-referenced data from a 1961/62 prevalence survey of malaria infection in 1-14 year old children. After starting with a list of 50 potential variables we ended with three highly plausible predictors, by applying a systematic and repeatable staged variable selection procedure that included a spatial analysis. All this was accomplished using general-purpose statistical software.

Background

Recent years have seen widespread application of geographic information systems and spatial statistical methods in modelling and mapping the distribution of vector borne diseases, including malaria. In sub-Saharan Africa the Mapping Malaria Risk in Africa (MARA) project has been working towards a malaria risk atlas for rational and targeted control of the disease (Snow *et al* 1996). To this end historical and current survey data have been collated of the prevalence of infection with human *Plasmodium* parasites.

A number of malaria risk maps, at country and regional level, have been produced by analysing geo-referenced prevalence data against environmental data to predict prevalence at localities where it was not recorded (Snow *et al* 1998; Kleinschmidt *et al* 2000; Kleinschmidt *et al* 2001a; Omumbo *et al* 2005; Gemperli *et al* 2006b). Different analytical approaches of varying sophistication have been explored. Multiple variable logistic regression analysis, commonly used to assess the odds of infection against potential risk factors, has been employed, and the spatial dependence in the response data has been modelled most successfully using Bayesian spatial modelling. One outstanding issue, which can greatly affect the predictions, remains the variable selection procedure, particularly when there are a large number of potential risk factors.

In regression analysis and predictive / prognostic statistics, model validity is an important aspect (Justice *et al* 1999), both the internal validity, or accuracy, i.e. the model explains the observed data well, and external validity, or generalizability, i.e. the model predicts new data well. In this context we furthermore aim for parsimony (model contains a few strong predictors that are easily interpretable) and plausibility, both of the co-variates (association with the disease are etiologically explainable) and of the predictions (believable in view of what is generally known). Taking account of the spatial correlation structure in the data is important for "geographic transportability", i.e. when predicting malaria prevalence to unobserved locations (Justice *et al* 1999).

Selecting a few predictors for spatial modelling from among a large number of potential candidates is a major challenge and can easily become arbitrary. Ideally every possible combination of variables would be tested and compared in a Bayesian spatial framework. However, this would be extremely computing-intensive and unfeasible, if not impossible, for most users. The most practical route is to reduce the list of potential explanatory variables using non-spatial selection methods, before moving to a spatial context.

Neither manual nor automated stepwise selection procedures are advised, because of frequent over-fitting, and because of the resulting "phantom degrees of freedom" (Babyak 2004, pg 416): testing and rejecting many variables increases the probability of finding a significant predictor by chance, but since this sifting remains undeclared, standard errors in the final model are underestimated. Babyak (2004), citing Harrell (2001) and others, recommend shorter lists of candidate predictor variables, which are not strongly correlated, as well as bootstrapping, as a form of simulation. Austin and Tu (2004), working on heart attack data, developed their model by running repeated step-wise selection procedures on bootstrap samples of their data, to identify the most consistent predictors.

The aim of this study was to develop a map of historical malaria risk for Botswana by analysing malaria prevalence data against a number of environmental variables from different data themes, using a systematic and repeatable staged process of variable elimination, including the stepwise bootstrap method described by Austin and Tu (2004). The resulting small subset of variables, each independently associated with the response, but possibly spurious because the condition of spatial independence was not satisfied, was tested in a Bayesian geo-statistical model. We used the spatial model derived from the observed locations, to predict prevalence of malaria infection in children 1-14 years old at unobserved map locations across the whole country.

Methods

Study area

Botswana is semi-arid to arid with few permanent water bodies. The country is flat, mostly between 900 and 1200m altitude. The rainy season is from November to March. Vegetation ranges from desert scrub-land in the South-West, where annual rainfall is <300mm, through grassland, to wooded savannah in the North, which receives >500mm rain annually. Mean annual temperatures are between 18 and 23°C. Botswana today has a total population of about 1 .6 million; population density over two thirds of the country being <1 per square km (Deichmann 1997). The population according to the 1971 census was 630 379 with an approximate 3.1% annual increase (Chayabejara *et al* 1975) which if extrapolated back in time translates to around 470 000 in 1961/62. In 1975 80% of the population lived in the eastern part of the country,

Malaria risk is highest in the tropical North (Figure 2.1). Indoor residual spraying was introduced in 1946 on a limited scale. Coverage was gradually improved culminating in a comprehensive vector control program in the 1980's (Mabaso *et al* 2004), but even by 1953

ndoor residual spraying for mosquito control was a “regular feature” in risk areas, apparently mainly in towns, along rivers and apparently excluding rural areas “remote from regular medical supervision”, but with good results (Freedman 1953). Larval control was also implemented when mosquito breeding was detected. Malaria prevalence decreased markedly after 1944, again between 1961/62 and 1974, and further thereafter (Mabaso *et al* 2004). By 1960 no prevalence above 70% was measured, suggesting meso- to hypo-endemic conditions. Further South, transmission is hypo-endemic and epidemic, and over large areas entirely absent. Incidence, like the climate, is strongly seasonal, peaking around March / April (Thomson *et al* 2005). The gradient in malaria broadly follows the environmental gradients described before.

Malaria data

Archived malaria prevalence data were collated within the MARA project, as described by Omumbo *et al* (1998). In Botswana geographical coordinates could be obtained for 613 out of a total of 1063 age-specific prevalence surveys. Of these, 20 did not report sample sizes and were excluded. Here we used only the 1961/62 national survey (Figure 3.1) to develop a historical malaria risk map. For the 1-14 year age group, 122 prevalence results were available, for 118 unique locations across the country, progressively from August 1961 to May 1962. Surveys in different regions were carried out during different months (Figure 3.2). The total number examined was 17 149; the mean sample size was 141 per survey (range 2-831). The design effect was calculated in Stata (Anon. 2001b).

Environmental data

Forty-nine variables representing different summaries and transformations of the eight environmental data themes (see Table 3.1), were included in the study: elevation (Anon. 1998b), surface water (Anon. 1995), land cover (Anderson *et al* 1976), long-term monthly

mean rainfall, temperature (Hutchinson *et al* 1995), vapour pressure (Mitchell *et al* 2003), and normalized difference vegetation index (NDVI) at 8km (Anon. 2001a) and 1km (Anon. 2007b) resolution.

Themes with monthly values (rainfall, temperature, NDVI and vapour pressure) were plotted against logit-transformed malaria prevalence, logit(p). Based on observed temporal patterns in the scatter plots, months were aggregated for "summer" (December to March) and "winter" (April to October). Different annual summary indices were also calculated for each theme. Calculations of some of the variables are shown in the appendix (pg. 55).

Distance from water bodies was calculated by projecting maps of perennial and non-perennial water bodies onto a 200x200m grid and calculating for each grid cell the Euclidian distance to the nearest water body. Values were transformed by adding a value of 100m to each pixel and deriving the natural logarithm.

For land cover, the thirteen United States Geological Survey land cover classes occurring in Botswana were re-grouped into two categories, broadly corresponding to drier and moister land cover types. Most data points were found in "grassland" and "savannah" with only isolated surveys in the other land cover types. Prevalence was generally higher in "savannah" than in "grassland" areas. Other obviously drier and lower risk land cover types ("barren or sparsely vegetated", "shrub land", "urban or built-up") were therefore included with "grassland" in a "low risk" category, while other clearly moister classes ("herbaceous wetland", "water bodies", "evergreen broadleaf forest") were included with the higher-risk "savannah" category. Other minor land cover types were included in the category alongside which they mostly commonly occurred ("grassland / crop land mosaic" was mainly found

scattered among “grassland”; “dryland crop land and pasture” and “mixed” among “savannah”).

Values were extracted from the data grids for each geographical location where a malaria survey result was available.

Variable selection and model development

We carried out a staged approach during model formulation. A flow chart of the variable selection procedure is shown in Figure 3.3.

Stage 1. The malaria prevalence database was split randomly into derivation (n=81) and validation (n=41) sub-sets. To identify the best univariate predictors, univariate logistic regression analysis against the derivation data was carried out on all 50 potential predictors. We allowed for clustering by survey location using the Hubert-White sandwich estimator in Stata (Anon. 2001b).

Stage 2. To reduce confounding arising from correlated variables, and also to reduce the variables to data ratio, we ranked the variables significant in univariate analysis by the Akaike Information Criterion (Akaike 1973) (AIC), and excluded those that were strongly correlated (Spearman’s $r > 0.85$) with a higher-ranking variable belonging to the same environmental theme. Scatter plots against logit(p) were prepared of the remaining variables (Figure 3.4).

Stage 3. Following the approach of Austin and Tu (2004), we drew 1000 bootstrap samples from the derivation data, and ran automated backward exclusion procedures on each sample. Since it was not possible in Stata to allow for clustering within the stepwise procedure, which resulted in the explanatory power of variables being over-estimated, we used stringent entry

and removal thresholds (p = 0.02 and 0.05 respectively). We recorded the co-efficients and the number of times each candidate variable was selected in the 1000 models.

Stage 4. A non-spatial multiple-variable model was derived in a manual step-wise fashion, starting with the most frequently selected variable, and adding further variables in order of selection frequency, as long as all entered variables remained significant at the 5% probability level. If a previously entered variable became non-significant with the addition of another, we retained the one more frequently selected in Stage 3 in favour of the other.

Stage 5. Back in Stage 2, variables had been excluded based on their univariate predictive power. To identify the best representative(s) of a theme in a multiple variable context, correlated variables excluded in Stage 2 were allowed to compete against each other for entry into the model in further stepwise-bootstrap procedures. The variables in the Stage 4 model constituted the basic candidate list. Working theme-by-theme, we re-introduced into the candidate list also those variables that had been excluded in Stage 2 on account of their high correlation with any variable of the same theme that had survived to Stage 4. Each time we ran a stepwise-bootstrap procedure as described above, recording which of the competitors was most frequently selected. This variable then replaced the original variable in the model. Details, in the form of an example, are provided in an annotation to Table 3.2. Using the modified model, prevalence was predicted for all 122 observations. The accuracy of the predictions for both derivation and validation data was assessed using the concordance correlation coefficient (Lin 1989; Lin 2000).

Stage 6. To account for spatial correlation in the survey data, a generalized geo-statistical spatial model using Markov chain Monte Carlo (MCMC) simulation was fitted on all 122 observed prevalence data points (Diggle *et al* 1998; Christensen & Ribeiro 2002; Gemperli &

Vounatsou 20032; Gemperli *et al* 2004). The co-variates of the Stage 5 model were included as potential explanatory variables. Spatial modelling was carried out using the package geoRglm in the statistical software system R (Christensen & Ribeiro 2002). Detailed methods are included in the appendix. For each model parameter the median and 2.5 and 97.5 percentiles were calculated from the MCMC simulations. Prevalence and its 95% CI was predicted and mapped for a grid of 2300 locations based on the co-variates and the spatial structure in the data.

Results

The design effect in the data was 52 before adjusting for co-variates. The clustered survey data thus only had the same power as 330 (17149 / 52) individuals randomly sampled over the entire country.

Of the 50 potential explanatory variables, 42 were significantly associated with malaria prevalence in univariate logistic regression in Stage 1 (Table 3.1). Scatter plots of logit(p) against the 14 variables that were selected for further analysis in Stage 2, are shown in Figure 3.4.

The selection frequency of the 14 candidate variables in the 1000 stepwise-bootstrap models of Stage 3, are shown in Table 3.2. Figure 3.5 shows the frequency distribution of coefficients for each variable. Some variables were unstable, having positive coefficients in some models and negative coefficients in others. Five variables were selected into the Stage 4 model, namely annual maximum rainfall, winter mean temperature, proportional SD temperature, elevation and land cover (marked in Table 3.2).

The results of the additional three stepwise-bootstrap procedures of Stage 5 are shown in Table 3.2. In the rainfall theme, annual maximum was outperformed and replaced by the summer total. For temperature theme, annual mean outperformed winter mean. With annual mean in the model, standard deviation became non-significant. Since standard deviation ranked lower in Stage 3 than the winter mean, it was removed, reducing the number of variables in the Stage 5 model to four. Results of the Stage 5 model are shown in Table 3.3.

Figure 3.6 shows the scatter plot of observed *vs* predicted logit(p), for the derivation and validation data of the non-spatial Stage 5 model. The concordance correlation coefficient (ρ_C) (Lin 1989; Lin 2000) for the derivation data, weighted by sample size, was 0.851, n (individuals examined) = 11182 in 66 non-zero prevalence surveys, the 95% confidence interval (CI) = 0.846 to 0.856. The unweighted ρ_C= 0.834, n = 66, CI = 0.760 to 0.908. For the validation data weighted ρ_C = 0.835, n = 4467, CI = 0.826 to 0.843; unweighted ρ_C = 0.776, n = 30, CI = 0.635 to 0.917. The difference between observed and predicted logit(p) did not vary with prevalence.

After adjusting for spatial random effects, only three co-variates remained significant. Land cover (median = -0.515; 95% CI = -1.099 and 0.059) was removed. The predictions (median and CI) from the spatial Stage 6 model are also shown in Figure 3.6. It contained three co-variates namely summer rainfall, annual mean temperature and elevation, each independently significantly associated with prevalence of infection after allowing for spatial correlation in the data (Table 3.4).

Discussion

This study was concerned with finding the best predictors of malaria prevalence in terms of plausibility, parsimony and reliability. One important question was how to summarize the

environmental data in a meaningful way. We determined to explore a range of alternative summaries of the monthly climate data, believing one appropriate summary indicator to be better for prediction than individual months (Snow *et al* 1998), quarterly aggregates (Kleinschmidt *et al* 2001a), or principal components (Omumbo *et al* 2005), the last of which are difficult to interpret. However, as more and more variables are tested against a certain data set, the risk increases that some will explain the data merely by chance, but will fail to explain new data.

In an initial attempt to derive a well-fitting and plausible model through automated step-wise variable selection (results not shown), arbitrary factors such as entry and removal threshold settings, how many variables were included in the list of candidates, and which data-subset was used for model derivation, affected which variables got selected. The best-fitting models did not produce the most plausible risk maps, and *vv*. The majority of maps resulting from these models strongly contradicted expert opinion. A more systematic selection procedure was called for.

Identification of consistent predictors is compromised by correlation among predictors. A strong, reliable predictor may ultimately be selected less frequently than a weaker predictor, if several strongly correlated alternatives compete for entry into the model so that each has a low selection frequency (Austin & Tu 2004). For this reason it was important to include in the candidate list only little-correlated variables. This was ensured in Stage 2, where the candidate list was reduced from 42 to 14.

Reliable predictors would not only explain a particular data set, but would be associated consistently with the response. The bootstrapping of Stage 3 helped identify such predictors, because those that consistently explain different sub-sets of the data, are more likely to

explain new data. In the step-wise bootstrap procedures, variables that explained the most observations would be selected most frequently while those that explained only some of the observations, would be selected only when these observations appeared in the bootstrap sample. The effect of individual observations on variable selection, especially of outliers, was thus reduced.

In the process of uni-variate ranking (Stage 1 and 2) we became guilty of "data peeking" (Babyak 2004). Using our data to assemble a candidate list of predictors set up the analysis for success. Such undeclared testing and discarding of variables may lead to illegitimately high model fit. Another problem of Stage 2 was that variables were excluded on the grounds of low uni-variate correlation with the response, while their predictive power may be quite different once other variables are accounted for. Stage 5 was an attempt to redress both these problems at once, by giving each variable excluded in Stage 2, whose relative had survived up to Stage 4, a fair chance to out-perform and supplant its competitors in a multiple-variable context, at the same time, through the bootstrap sub-sampling, to reduce the influence of the data set on this process.

A further benefit of the Stage 3 bootstrap-stepwise procedures was the information provided by the frequency distributions of coefficients in the 1000 stepwise models (Figure 3.5). A variable that has a widely varying coefficient, or one that is sometimes positive and sometimes negative, is clearly not reliable and should be considered with suspicion (Concato *et al* 1993). An example was summer vapour pressure, the strongest uni-variate predictor, but selected least frequently in multiple-variable regression (Figure 3.5J). Altitude on the other hand, a weak uni-variate predictor, became an important predictor in a multiple-variable context, with a stable positive coefficient (Figure 3.5G). In fact, the most frequently selected variables (Table 3.2) had stable coefficients (Figure 3.5), whereas the most unstable

coefficients were found among the least frequently selected variables, confirming the relative importance of predictors.

The strong association found between malaria prevalence and selected environmental data (Figure 3.7) is biologically plausible since high malaria infections have been shown to coincide with conditions that favour vector and parasite development in a given location (Kleinschmidt *et al* 2001a). However, over small distances environmental conditions vary only slightly due to the relatively simple flat Botswanan topography, while malaria prevalence showed substantial local variation, for example contemporal measures of 67% (n=48, Maun) *vs* 24% (n=557, Maun suburb), or 3% (n=219) *vs* 17% (n=116) in Matangwane. Such local variation is perhaps partly caused by the distribution of small breeding sites. Yet in studies where detailed breeding site information was available, much of the variation in incidence (Van Der Hoek W. *et al* 2003), prevalence and entomologic inoculation rate (Hightower *et al* 1998) nevertheless remained unexplained. Localized factors, such as individual, household and village characteristics, as well as the effect of sampling procedure and size, may further contribute to the unexplained variability in prevalence.

Summer rainfall and annual mean temperature, retained in the final multiple-variable model, were highly plausible predictors. The same variables - summer rain and mean temperature over the preceding year - were also found to explain inter-seasonal variation in malaria incidence in KwaZulu-Natal (Craig *et al* 2004b). Summer rainfall also explained much of the variation in inter-annual variation in malaria incidence in Botswana (Thomson *et al* 2005). High rainfall during the hot summer months allows rapid breeding and population expansion of the mosquito vectors, while high mean temperatures maximize the maturation rate of the parasite in its exothermic arthropod host (Molineaux 1988). Warmer winters reduce the die-back of mosquitoes and parasites, thereby increasing the reservoir for the following season.

The strong positive association of elevation with malaria prevalence (an increase in logit(p) of 1 every 160m, Table 3.4) was surprising, as prevalence on its own, as it usually tends to be, was higher in low-lying areas (Figure 3.4G). This positive association was difficult to explain, but may be connected with the malaria control that was ongoing at the time. It appears from early reports (Freedman 1953) that vector control operations were wide-spread and intensive along rivers and the main populated areas.

The non-spatial model of Stage 5 predicted the data fairly well but the predictions achieved by the spatial model of Stage 6 were more accurate (Figure 3.6). The map corresponding to the Stage 5 model (Figure 3.7A) had an implausible discontinuity, caused by the negative co-efficient of land-cover. Land-cover was the most frequently selected variable in the bootstrap procedures, but was not significant in the spatial model. This binary variable may simply have approximated the spatial division between high and low prevalence areas, which was ultimately described more correctly through the geo-spatial approach of Stage 6 (Figure 3.7B).

A good number of locations with observed zero prevalence had predicted prevalence of 5%, i.e. logit(p) of -3, and above (Figure 3.6). In these cases sampling error may have played an important role, as large sample sizes are needed to measure very low prevalence rates confidently. Conversely, non-zero observations were more often lower than the predictions based on environmental factors. By 1961/62 malaria prevalence in the North of Botswana was already much below the level measured in 1944 (Mabaso *et al* 2004), probably due to the limited use of indoor residual spraying which had been ongoing since the 1940's. This highlights the fact that not only environmental, but also anthropogenic factors, especially malaria control need to be considered. This furthermore highlights the need to monitor control coverage and effectiveness, as well as other potential cofactors, in order to understand the situation more accurately.

Evidence from elsewhere in Africa suggests that prevalence rates in the dry / low transmission season may differ substantially from those in the wet / high transmission season (Molineaux & Gramiccia 1980; Lindsay *et al* 1991). In this study month of survey was a significant predictor of prevalence in a univariate setting only, but not while accounting for other variables. Prevalence by month (Figure 3.4N) was confounded by where surveys were carried out when, and thus did not reflect the seasonality of malaria risk. The highest incidence months for example (March to May) would not be the lowest prevalence months, as Figure 3.4N suggests. Rather, surveys were carried out during these months in the low-risk South (Figure 3.2). To measure intra-annual variation in prevalence we would have required data from the same localities in different months.

The spatial risk map (Figure 3.7B) presents a smoothed picture of malaria risk in Botswana prior to intensive malaria control, which was highly plausible based on expert opinion and the mean incidence at district level (Craig *et al* 1999). The wide CI (Figure 3.7C) in predicted prevalence highlights the uncertainty remaining after accounting for all explained variation in the data. The confidence level needs to be taken into account when using the map for planning and evaluating control interventions, to avoid over-interpretation of the map.

Conclusion

A continuous map of malaria risk is more useful than point-prevalence rates for several reasons. First, the variability in individual observations may hide underlying patterns that have epidemiological importance. Further, it is not possible to deduce from a point-referenced map what prevalence you may expect to see in areas that have not been sampled, whereas a model such as the one developed here gives a likely range of prevalence for the entire region. A continuous prevalence map can also be combined with underlying population data to estimate the number of people at risk of - or infected with - malaria. Finally, the spatial statistical

methods employed here distinguish between the correlation among observations that can be ascribed to their spatial proximity (neighbouring villages affecting each other), and that which can be explained by environmental factors (thereby avoiding overestimating the explanatory power of the co-variates).

Though malaria risk has been reduced substantially through intense malaria control, a malaria risk map nevertheless remains highly useful from the control perspective in knowing historical prevalence levels. We have furthermore demonstrated a systematic procedure for variable selection and model formulation in developing a geo-statistical risk model from point-referenced malaria prevalence data, which has relevance to a broad range of environmentally determined infectious diseases. The failure take account of spatial correlations during the entire variable selection procedure remained a major weakness. As computing power increases and statistical software packages are further developed, variable selection within a spatial framework may end up being within the means of the average researcher.

The staged process of variable elimination employed here proved to be practical, though not necessarily the optimal solution. Stepwise variable selection on multiple bootstrap samples drawn from the data allowed us to identify the most consistent and stable explanatory variables. Selection frequency provided an objective rationale for choosing one variable above another, and to choose between similar and strongly correlated indicators. Spatial analysis was the final stage in the variable elimination process, after which we remained with a parsimonious, highly plausible model that produced a smooth, plausible map of malaria risk.

Acknowledgements

We thank the Botswana Ministry of Health for contributing their data to the efforts of the MARA project. We acknowledge Tom Smith and Penelope Vounatsou of the Swiss Tropical Institute for revising the manuscript critically for important intellectual content and thank them for their valuable comments. We thank the South African Medical Research Council and the Rudolf Geigy Stiftung zu Gunsten des Schweizerischen Tropeninstituts for supporting this study, as well as the various funders of the larger MARA project, particularly MIM / TDR and RBM.

Appendix 1

Standard deviation (SD) =

$$\sqrt{\sum_{m=1}^{12}(\hat{y}-y_m)^2}$$

where y_m = monthly value and $\hat{y}$ = mean of all y_m.

Proportional SD (based on monthly proportions) =

$$\sqrt{\sum_{m=1}^{12}(0.0833-py_m)^2}$$

where $py_m = y_m/y_{tot}$; $y_{tot} = \sum y_m$, and 0.0833 is the mean of all py_m (=1/12)

Effective temperature (Stuckenberg 1969) =

[8 * annual mean + 14 * annual range] / [8 + annual range]

Concentration of rainfall

Monthly rainfall is expressed as a vector (r_m,θ_m), rainfall being the magnitude (r) of the vector and the month its angle (θ) expressed in units of arc:

$\theta_m = m2\pi / 12$

where m is the month, so that January = 1 and December = 12.

The twelve monthly vectors are added to calculate the total vector (r_t , θ_t):

$$r_t = \sqrt{\left(\sum_{m=1}^{12} r_m \cos\theta_m\right)^2 + \left(\sum_{m=1}^{12} r_m \sin\theta_m\right)^2}$$

$$\theta_t = \tan^{-1}\left(\frac{\sum_{m=1}^{12} r_m \sin\theta_m}{\sum_{m=1}^{12} r_m \cos\theta_m}\right)$$

The concentration index C is calculated as:

$C = 100r_t$ / annual total

Concentration is 100% if all the rain falls in one month and 0% if all months have equal amount of rain.

θ_t is the mean peak month around which rainfall is concentrated.

Generalized spatial logistic regression analysis

Bayesian geostatistical model formulation has been described by a number of authors (Diggle *et al* 1998; Christensen & Ribeiro 2002; Gemperli & Vounatsou 20032; Gemperli *et al* 2004). Following these authors, the model is specified as follows:

Y_{ji} represents the binary response corresponding to the infection status of child *j* at site *i* (the survey site) taking value 1 if the child tested positive and 0 otherwise. The Y_{ji} are conditionally independent Bernoulli variables with infection probability p_i at location *i* .
The p_i are defined via a generalised linear mixed model, to take account of spatial dependence:

$\text{logit}(p_i)=\mathbf{X}_i\boldsymbol{\beta}+S(\ell_i)$

where $\boldsymbol{\beta}$ represents the regression coefficients for a set of known co-variates $\mathbf{X}$ at all locations ℓ_i of the study area;

$S=(S(\ell_1), \ldots., S(\ell_n))^T$ denotes the values of the (unobserved) Gaussian spatial process $S(\cdot)$ at sample locations ℓ_i;

$\sigma^2 = \text{Var}\{S(\ell)\}$, and Φ is a parameter of the correlation function $\rho(d_{ij}, \Phi)$, in our case $\exp(-d_{ij} / \Phi)$, where d_{ij} is the distance between locations ℓ_i and ℓ_j.

For $\boldsymbol{\beta}$ flat priors were specified respectively (defaults in geoRglm) and for σ^2 a Scaled-Inverse chi-square distribution(χ^2_{Scl}) with five degrees of freedom and a mean of 0.5. For Φ a discrete

exponential prior with mean of 0.04 and 1000 discretisation points in the interval 0.0001 to 2 was specified.

Convergence was assessed by inspecting plots of traces of simulations for individual parameters. The first 50,000 iterations were discarded; thereafter simulations were run for 250,000 iterations. Every 50th sample was retained. For each model parameter the median and 2.5 and 97.5 percentiles were calculated from the 5,000 MCMC simulations.

Models were compared by calculating the deviance information criterion (DIC) for each model (Spiegelhalter *et al* 2002). Spatial prediction using Bayesian krieging was carried out for a grid of 2300 locations which correspond to the entire surface of Botswana. For each prediction location a posterior sample of MCMC simulations was generated taking account of the estimates of regression coefficients and the spatial effects at each location, and of the uncertainty of each parameter. This process is described in detail elsewhere (Diggle *et al* 1998; Gemperli & Vounatsou 20032; Gemperli *et al* 2004), and was carried out using geoR (Christensen & Ribeiro 2002).

Table 3.1 Odds ratios (AIC in parentheses) from univariate logistic regression analysis in Stage 1, of 50 environmental variables tested against malaria prevalence. P-values were non-significant (n.s.), <0.05 (*), <0.01 (**) or <0.0005 (***), n=122. The equation was logit(prevalence) = coefficient × co-variate + constant.

	Environmental data theme					
Variable	**Rain-fall (mm)**	**Tempe-rature (°C)**	**Vapour pressure (hPa)**	**NDVI, 8km resolution §**	**NDVI, 1km resolution §**	**Other**
Annual mean (total for rainfall)	1.0085 (27.6)**	4.22 (13.6)***	1.094 (12.8)***	1.091 (28.9)*	1.07 (31.3) n.s.	
Annual maximum (highest monthly value)	1.045 (20.8)***	3.034 (23.3)***	1.067 (11.7)***	1.090 (25.7)***	10.4 (32.2) n.s.	
Annual minimum (lowest monthly value)		3.29 (13.9)***	1.11 (17.1)***	1.1048 (29.8)*	1.06 (32.7) n.s.	
Annual range (highest minus lowest month)		0.52 (27.1)**	1.12 (15.8)***	1.14 (30.8)*	1.03 (32.7) n.s.	
Standard deviation (Appendix)	1.03 (21.9)***	0.54 (25.0)***	0.54 (14.7)**	1.073 (26.6)***	1.03 (32.8) n.s.	
Proportional standard deviation (Appendix)‡	61.8 (13.0)***	-214 (17.3)***	0.004 (33.4) n.s.	0.1 (26.8)***	43.3 (32.9) n.s.	
Summer mean (total for rainfall) Dec-Mar	1.012 (22.9)***	2.59 (27.1)***	1.065 (11.6)***	1.078 (28.9)*		
Winter mean (total for rainfall) Apr-Oct	0.88 (14.8)***	3.22 (12.0)***	1.11 (16.0)***	1.097 (28.6)**		
Concentration (see Appendix)	1.39 (13.3)***					
Number of months >80mm (>60 & >40mm n.s.)	1.81 (26.6)**					
Number of months >16°C		2.72 (18.9)***				
Number of months >165 (other cut-offs were n.s.)				1.13 (31.5) n.s.		
Total in months with more than 80mm	1.0059 (24.0)***					
Total degree months above 16°C		1.050 (15.7)***				
Effective temperature (Appendix)		21.8 (12.6)***				
Mean daily minimum of coldest month		2.29 (21.4)***				
Elevation						0.997 (29.7)**
Log distance to perennial water (m)						0.56 (21.6)***
Log distance to perennial / non-perennial water (m)						0.72 (30.5)**
Land cover (binary; moist *vs* dry areas)						4.76 (25.5)***
Month of survey (binary; peak season April / May *vs* rest of year)						8.67 (29.4)***

NDVI = normalized difference vegetation index .

‡ The co-efficients, not the odds ratios, are shown, as the unit is a fraction, and the Odds Ratio near zero (=exp(co-efficient)).

§ Radiance units for NDVI (fractions from 0 to 1) are translated to a byte-compatible scale from 1 to 256.

Table 3.2 Results of bootstrap step-wise procedures; variables included in the candidate lists of Stage 3 and Stage 5, and their selection frequency (fq), in four separate automated stepwise backward variable exclusion procedures, each time against 1000 bootstrap samples of the malaria prevalence data.

Theme	Stage 3		Stage 5					
	Candidate variable list	fq	Candidate variable list 1‡	fq	Candidate variable list 2	fq	Candidate variable list 3	fq
Rainfall								
	annual maximum*	904	annual maximum	560	annual maximum	533	annual maximum	914
			summer total †	821				
			months >80mm	760				
			SD	726				
			total in months >80mm	716				
			annual total	612				
	winter total	749						
	proportional SD	642						
Temperature								
	winter mean *	885	winter mean	993	winter mean	878	winter mean	665
					annual mean †	914		
					summer mean	885		
					months >16°C	681		
					mean in months >16°C	670		
					annual maximum	665		
					winter minimum	627		
					effective	615		
					annual minimum	558		
	proportional SD*	754	proportional SD	897	proportional SD	544	proportional SD	624
							SD	786
							annual range	537
	annual maximum	660						
Vapour pressure								
	SD	495						
	summer mean	441						
NDVI	annual maximum	567						
	SD	469						
Elevation *†		874	elevation	988	elevation	819	elevation	994
Log distance to perennial water		616						
Land cover *†		988	land cover	996	land cover	997	land cover	996
Month of survey		527						

NDVI - normalized difference vegetation index; SD - standard deviation

* Variables selected into Stage 4 model

† Variables selected into Stage 5 model

‡ Example: Five alternative rainfall indicators, listed in candidate list 1 under Stage 5, were strongly correlated with - and had been excluded in favour of - the annual maximum in Stage 2. In Stage 5, all six competing rainfall indicators were included in the candidate list, along with the other variables of the Stage 4 model. Of the six competitors the most frequently selected was summer total. In Stage 5 summer total therefore replaced annual maximum rainfall.

Table 3.3 Results of the Stage 5 non-spatial model: odds ratios, z-scores, and confidence interval estimated from non-spatial regression against four variables, fitted on derivation data only (n = 81, AIC = 8.06).

Variable	Odds Ratio	z	p(z)	95% confidence interval	
				lower	upper
rainfall summer total (per 100mm)	2.33	6.94	<0.0005	1.84	2.99
temperature annual mean (per °C)	8.85	9.05	<0.0005	5.53	14.15
elevation (per 100m)	1.68	3.8	<0.0005	1.28	2.2
high risk land cover	0.188	-5	<0.0005	0.098	0.361

Table 3.4 Results of the Stage 6 spatial model: odds ratios and confidence interval estimated from Stage 6 spatial model, fitted on all prevalence data (n = 122).

Variable	Odds Ratio	95% confidence interval	
		lower	upper
rainfall summer total (per 100mm)	2.01	1.49	2.7
temperature annual mean (per °C)	5.75	4.14	8.08
elevation (per 100m)	1.82	1.49	2.22

$\Phi = 0.003$, 95% CI = 0, 0.0174, $\sigma^2 = 0.77$, 95% credible interval (0.53, 1.14)

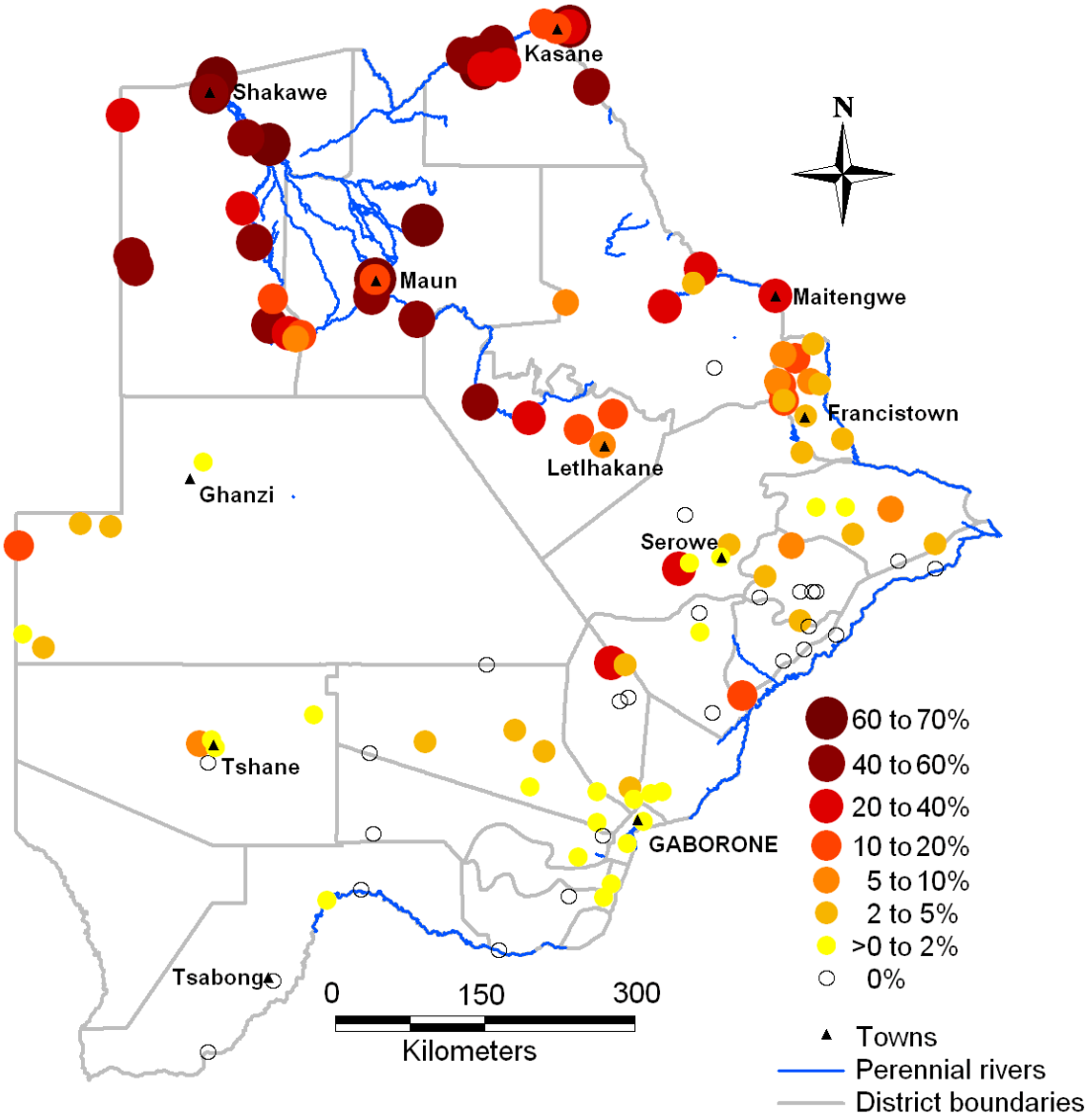


Figure 3.1 Malaria prevalence of infection in 1 to 14 year old children, in Botswana, during the 1961/62 national survey.

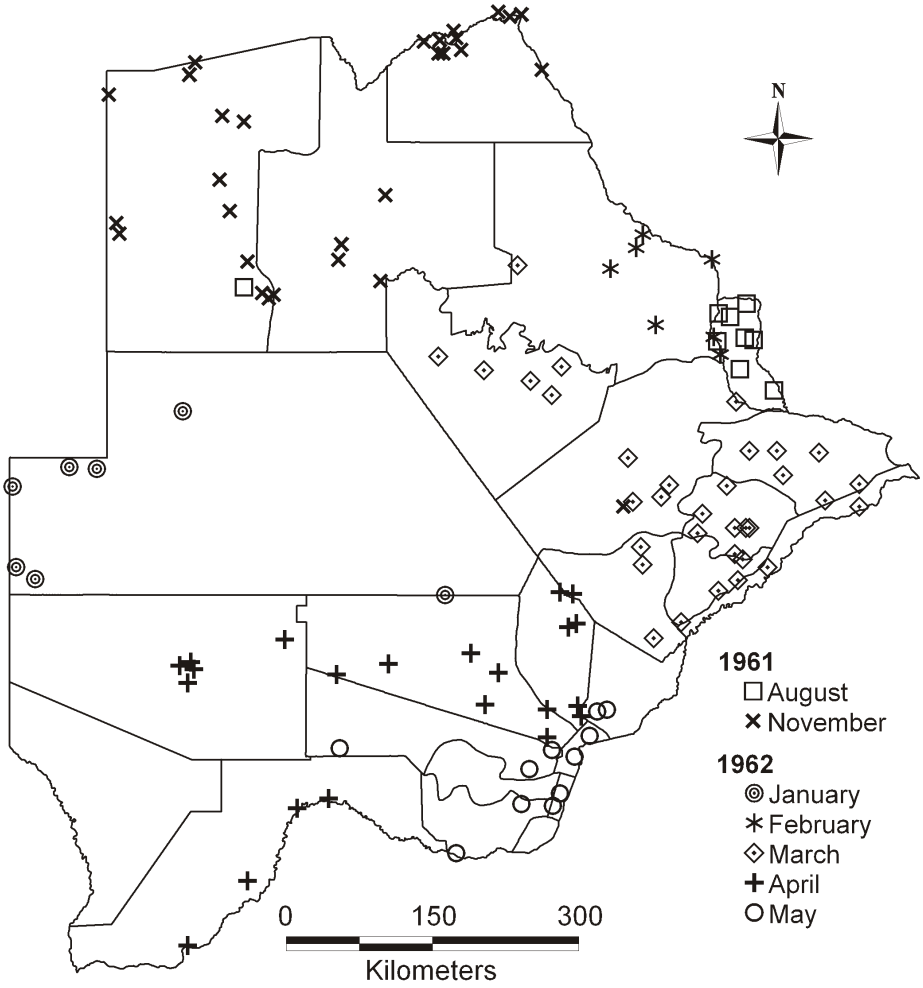


Figure 3.2 Month of survey during the 1961/62 Botswana national malaria survey.

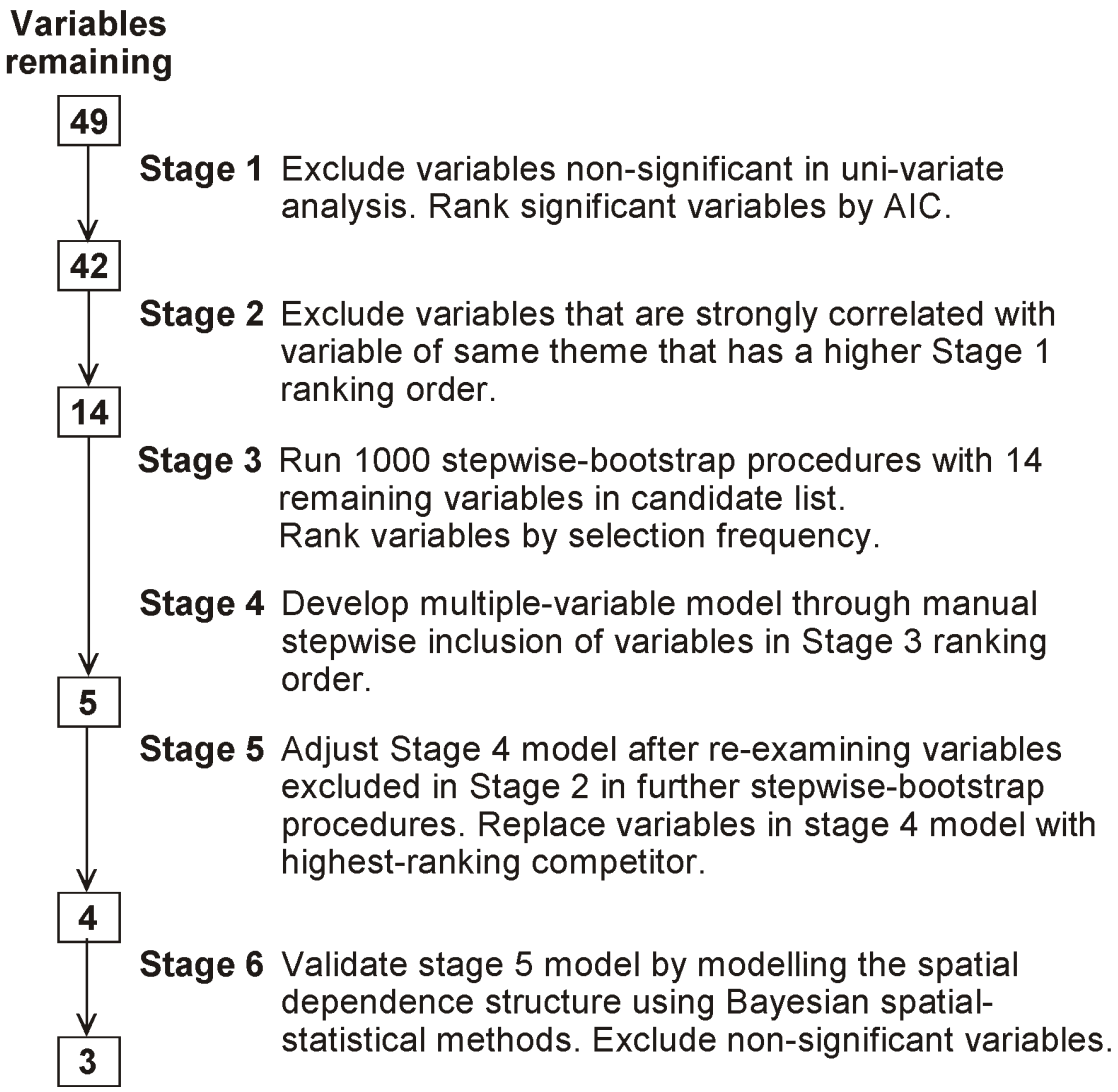


Figure 3.3 Flow diagram of staged variable selection procedure.

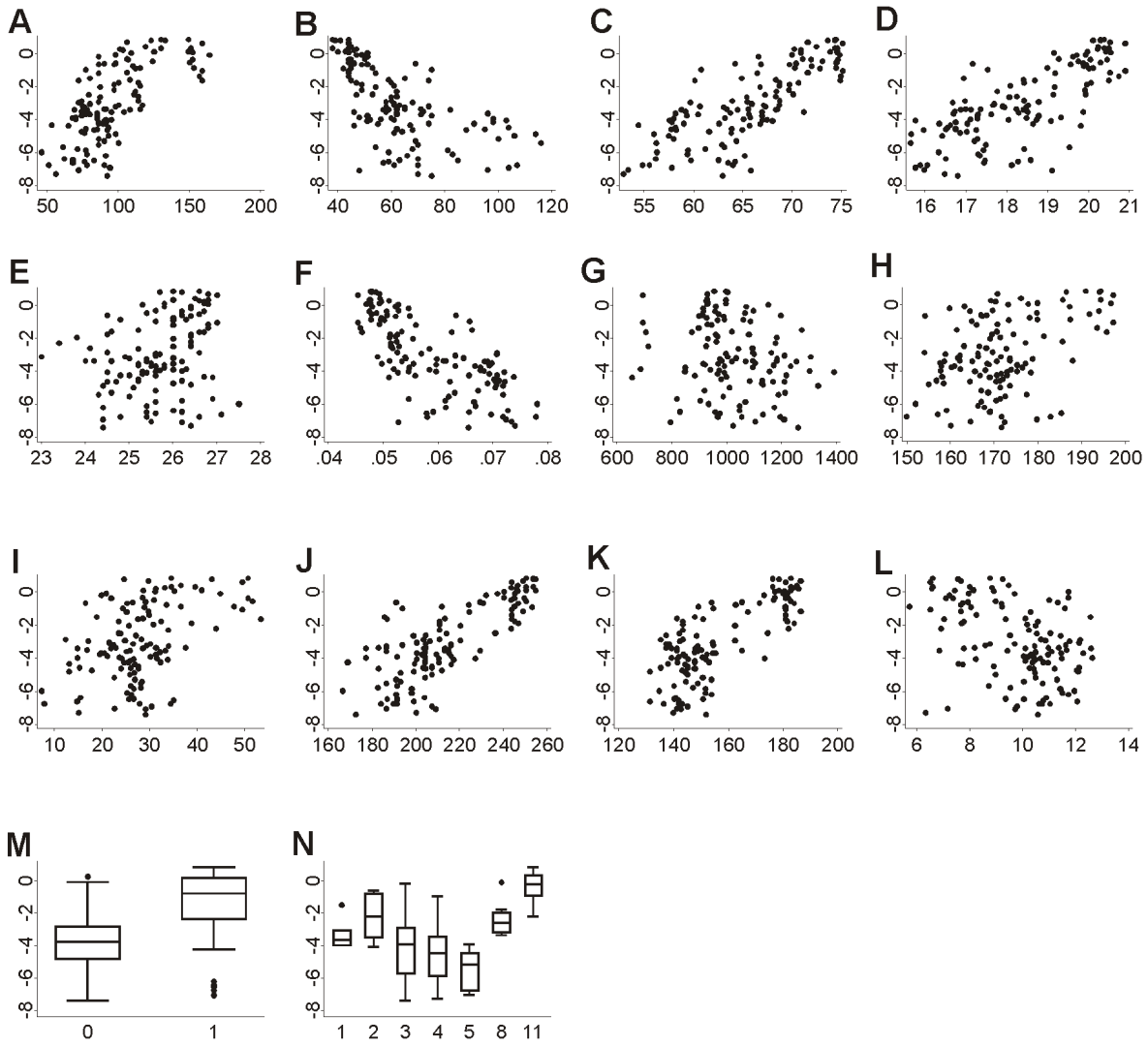


Figure 3.4 Scatter and box plots of candidate environmental explanatory variables used in the step-wise procedures. Malaria prevalence in 1 to 14 year old children, Botswana, 1961/62, is shown on the Y axis on a logit scale. (A) annual maximum rainfall (mm); (B) winter (April - October) total rainfall (mm); (C) rainfall concentration (%); (D) winter (April - October) mean temperature (°C); (E) annual maximum temperature (°C); (F) temperature proportional standard deviation (°C); (G) elevation (m); (H) annual maximum NDVI; (I) NDVI standard deviation; (J) summer (December-March) mean vapour pressure (hPa); (K) vapour pressure standard deviation (hPa); (L) log distance to permanent water (m); (M) land cover: dry / low risk, moist / high risk areas; (N) start month of survey (January, 1 to November, 11).

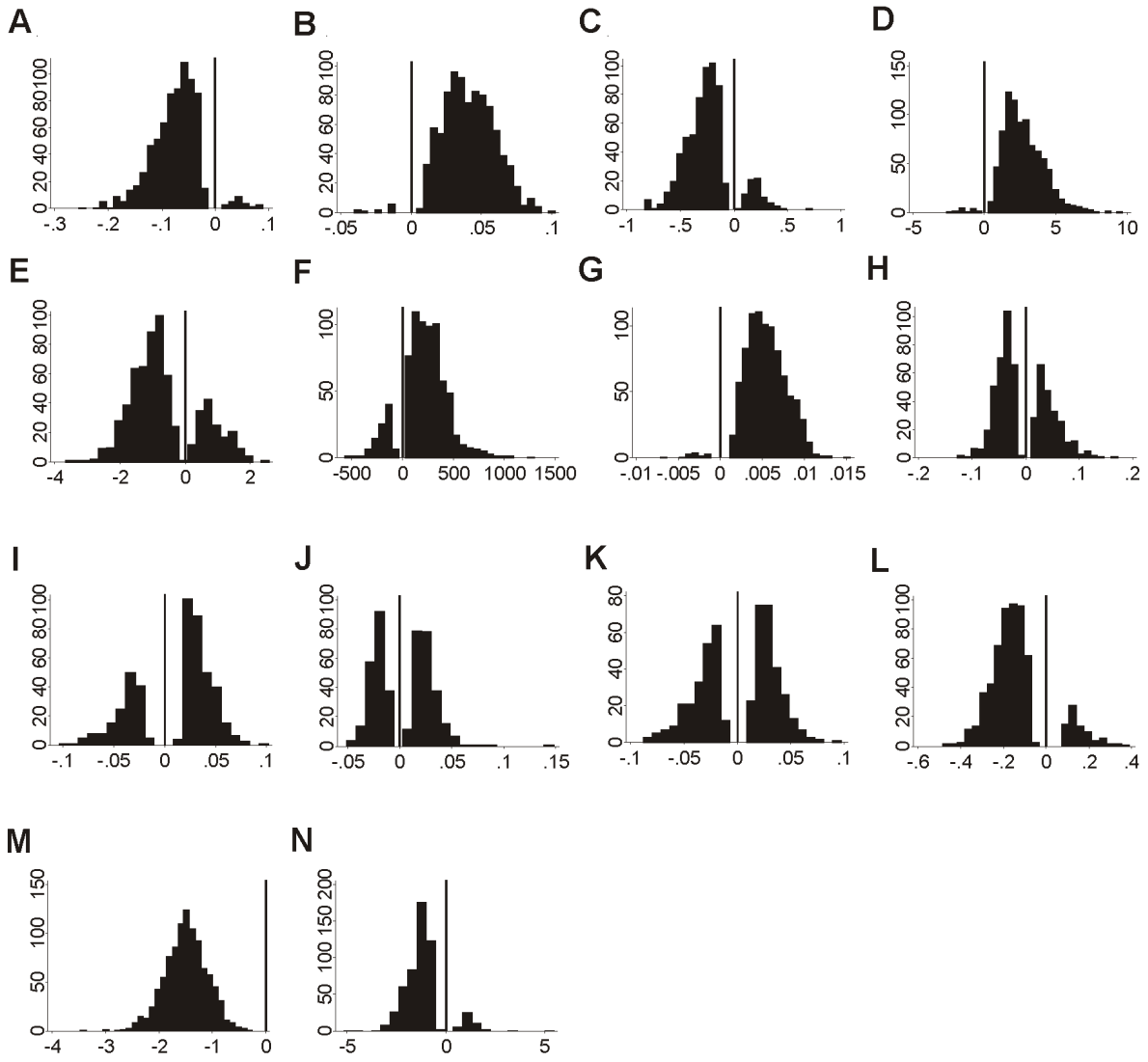


Figure 3.5 Frequency histograms of coefficients obtained in automated backward stepwise exclusion regression analysis against 1000 bootstrap samples of the malaria prevalence data in Stage 3. In each case the vertical black line indicates coefficient = 0. (A) annual maximum rainfall (mm); (B) winter (April - October) total rainfall (mm); (C) rainfall concentration (%); (D) winter (April - October) mean temperature (°C); (E) annual maximum temperature (°C); (F) temperature proportional standard deviation (°C); (G) elevation (m); (H) annual maximum NDVI; (I) NDVI standard deviation; (J) summer (December-March) mean vapour pressure (hPa); (K) vapour pressure standard deviation (hPa); (L) log distance to permanent water (m); (M) land cover: dry / low risk, moist / high risk areas; (N) start month of survey: main season (April-May).

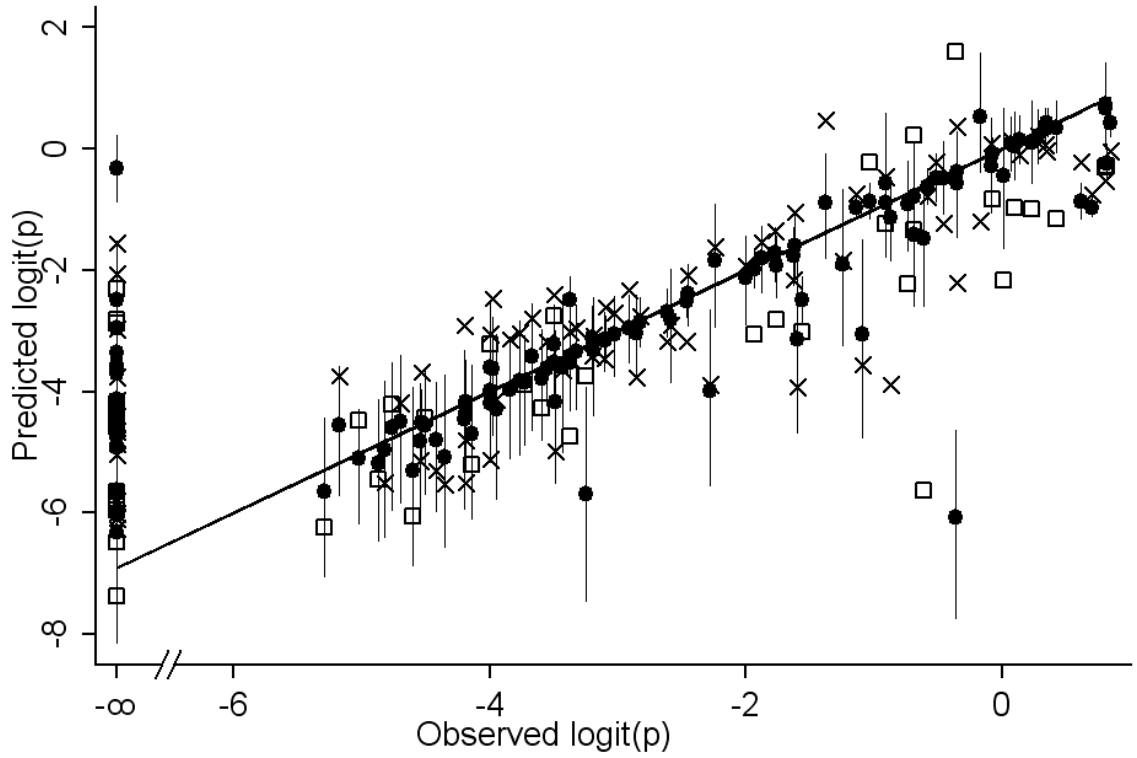


Figure 3.6 Predicted *vs* observed prevalence on a logit scale, for the derivation (crosses) and validation (squares) data of the Stage 5 non-spatial model, and for the median (closed circles) and upper / lower confidence interval (spikes) of the Stage 6 spatial model.

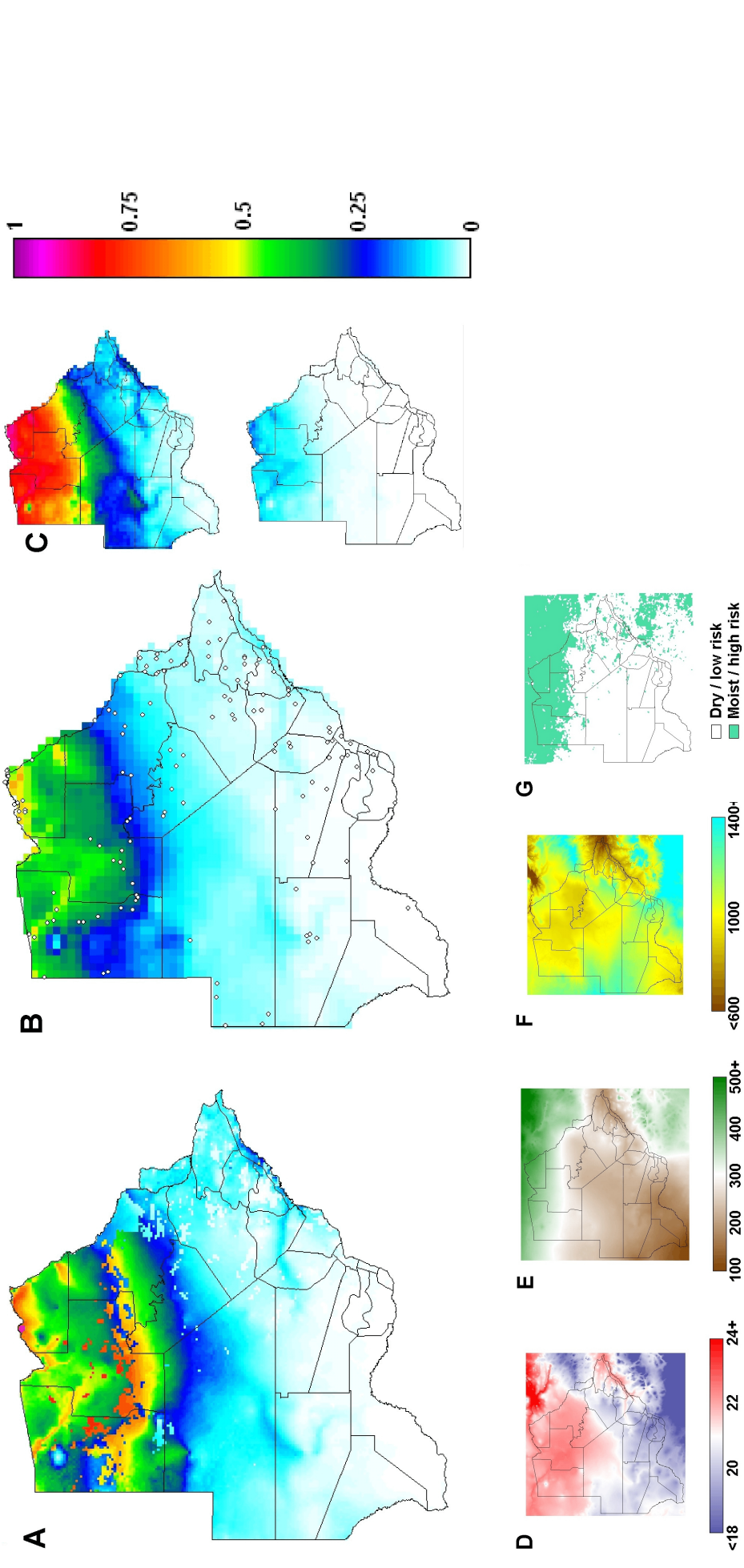


Figure 3.7 Predicted pre-control childhood malaria prevalence maps for Botswana, resulting from (A) the stage 5 non-spatial model and (B) the stage 6 spatial model; 118 survey sites are shown; (C) the upper and lower 95% CI of the spatial model. Co-variates used in the models: (D) annual mean temperature, °C; (E) summer total rainfall, mm; (F) elevation, m; (G) land cover categories, high-risk / low-risk. Lines represent district boundaries.

Chapter 4

Time-space analysis of malaria prevalence data in Botswana

M.H. Craig[1,2], I. Kleinschmidt[1,3], M.L.H. Mabaso[1,2], P. Vounatsou[2], T. Smith[2]

Affiliations:

[1] Malaria Research Programme, Medical Research Council, PO Box 17120, Congella, 4013 Durban, South Africa. Tel: +27-31-2043653, Fax: +27-31-2051498, email: craigm@mrc.ac.za

[2] Swiss Tropical Institute, 57 Socinstrasse, Basel, BS 4002, Switzerland

[3] London School of Hygiene and Tropical Medicine, Keppel Street, London WC1E 7HT, United Kingdom

Keywords:

malaria; prevalence; maps; regression analysis; temporal analysis; spatial analysis; models, statistical; climate; Bayes theorem

Publication status:

Not submitted for publication.

Abstract

Introduction

Malaria transmission and its temporal and spatial distribution are affected by environmental factors, particularly climate. Malaria prevalence, which reflects the balance of the acquisition and loss of infections, is more prone to non-climatic effects than incidence, and this study aimed to examine whether spatial and temporal patterns in malaria prevalence in Botswana could be largely explained by climate alone, as has been shown to be the case with incidence.

Methods

Retrospective prevalence survey data from Botswana, collated by the MARA project, were extracted. For the 5/6 to 9/10 age group 327 prevalence rates, for 1974 to 1997 and from 87 unique locations, were analysed via logistic regression against annual mean temperature, summer rainfall and elevation, while allowing for spacial and inter-annual correlation and a possible linear time trend. The model was fitted in a Bayesian framework, using a Markov chain Monte Carlo algorithm. Fifty random observations were set aside for validation. Two models were fitted using different rainfall data series. The one that better predicted the validation data was fitted on all data.

Results

Inter-annual variation in the prevalence data could not be explained by inter-annual variation in the environmental data with a meaningful level of confidence. In both models less than a third of the validation points were correct, even at the 95% credibility level. Temporal variation in prevalence, when aggregated for the country, could also not be explained by the variation in climate.

Conclusion

The results suggest that available monthly climate time series are not useful in predicting inter-annual changes in prevalence. While incidence rates appear to remain prone to variations in climate, even in the presence of intensive malaria control operations, this study suggests that the effects of vector control and good access to treatment over-ride the effects of climate on malaria prevalence.

Introduction

Malaria is an environmental disease insofar as its transmission rate is largely determined by the impact of environmental factors on the *Plasmodium* parasite and *Anopheles* vector populations. This link affects the temporal and spatial distribution of the disease (Molineaux 1988). On the fringes of malaria distribution transmission is not only less intense, but more prone to temporal variation, both intra- and inter-annually, particularly following climatic patterns. Inter-annual variations particularly relate to the prediction, detection and management of epidemics (Najera 1974).

Incidence is a more immediate indicator of transmission intensity, in that it reflects the rate at which new infections are being acquired in a population. Point prevalence of infection, the proportion of people infected at a particular time point, represents the net balance between the acquisition of new infections and the loss of infections due to recovery. Recovery rate, in areas with reasonable access to treatment, is affected by treatment policies, drug efficacy and health-seeking behaviour. Prevalence is therefore more prone than incidence to non-climatic effects, which tend to be monitored less regularly. While temporal changes in incidence have been clearly linked to temporal changes in climate, a temporal link between climate and prevalence has not been demonstrated.

The aim of this study was to analyse the effects of environmental factors on malaria prevalence in children over a 24 year period in Botswana, in order to examine whether temporal climatic variation is important against a background of fairly intensive and effective control.

Methods

Study area

Botswana lies at the fringe of malaria distribution in southern Africa. The North of the country is fairly moist and humid, even tropical around the Okavango delta and the bounding rivers in the North, but increasingly cool and dry towards the South. Malaria is concentrated in the North, tapering off towards the South and West. Botswana is sparsely populated, its inhabitant mainly concentrated in the South-East of the country (Figure 4.1). A more detailed description of the study area was given in Chapter 3.

History of malaria control

Indoor residual spraying was introduced in 1946 (Mabaso *et al* 2004) and by 1974 most villages in the North of the country were covered with indoor spraying of residual insecticides (Chayabejara *et al* 1975). Coverage was fairly complete, as the report mentions that at the time 60 000 people lived in Chobe and Ngamiland, and that around 20 000 houses were sprayed in 1973, though apparently logistic and equipment problems were experienced. By 1980 a comprehensive malaria control programme was fully operational. The period covered by this study (1974-1997) thus represents a period of more or less uniform intervention, but which was probably failing gradually due to the spread of chloroquine resistance until it was replaced with sulphadoxine-pyremethamine in 1997 (Thomson *et al* 2005).

Surveillance and malaria data

A national survey of malaria prevalence was carried out in 1961/62. After this, prevalence data are available for 1974, and from 1979 to 1995 for almost every year, but for the northern, malaria-affected half of the country only. The 1974 survey presented data in standard age groups: 0, 1, 2-4, 5-9, 10-14, 15+, but from 1979 onwards, with a few exceptions, only the 5/6 to 9/10 age group was sampled. The original surveys are simple unpublished ministry reports, and nothing further is known about the sampling schemes for these surveys.

For this study we extracted 376 prevalence rates from 1974 to 1997, for the 5/6 to 9/10 age group, representing 37 481 individuals, and covering 18 years over a 24 year period with two gaps. Of these surveys, 327 could be geo-referenced, and represented 87 unique locations. Mean sample size was 115 (range 3 - 516). Surveillance of clinical cases - unconfirmed and confirmed - was going on at the same time (Thomson *et al* 2005) but these data are not included in the present study.

Environmental co-variates

Only a limited number of sources of climate data going back to the 1970's are available with sufficient temporal and spatial resolution to allow a full spatial-temporal analysis. Monthly temperature and rainfall surfaces were sourced from the Climate Research Unit at the East Anglia University (Mitchell *et al* 2003), at half-degree spatial resolution (CRU-TS2), and from a similar data series from the same authors (New & Hulme 1997), commissioned by the MARA project (CRU-MARA), which essentially represent a higher resolution of an older version of the CRU data. Two further rainfall data series were sourced from CMAP (Anon. 2007a) and GPCP (Adler *et al* 2003). Both of these combine station and satellite data to derive their estimates. Table 4.1 shows the four rainfall data sources along with other relevant information.

In an analysis of the Botswana 1961/62 national malaria prevalence survey (Chapter 3), summer rainfall, mean annual temperature and elevation were found to be significant spatial predictors of malaria. For the current analysis we calculated the mean monthly rainfall during the summer months (December to March) corresponding to the beginning of each year for which we had malaria data, and the mean annual temperature. The coldest and driest months in Botswana are June and July, while the peak malaria incidence is observed in March and April (Thomson *et al* 2005). We thus aggregated the climate data year according to the climatic season (July of year *t*-1 to June of year *t*), rather than by calendar year, linking this to the malaria prevalence of year *t*. We extracted climate and altitude (Anon. 1998b) data for each survey location, and for each season for which malaria prevalence had been measured.

Statistical analysis

Binomial regression was used to analyse risk factors for malaria prevalence. To allow for temporal correlation, first order auto-regressive terms were included in the model. The year was also included as a risk factor, to control for a possible long-term linear trend. To allow for spatial correlation a generalized geo-statistical spatial model (Diggle *et al* 1998) fitted with a Bayesian framework, using Markov chain Monte Carlo (MCMC) simulation in Winbugs version 1.4 (Anon. 2004b).

To determine which of the four potential rainfall data sources best explained inter-annual variation in malaria prevalence, malaria prevalence and rainfall data were extracted for the time period covered by all four rainfall data series (1979 to 1995, 281 observations) and regressed against the malaria prevalence data. In each case the time trend was included as co-variate.

Fifty random observations were set aside for validation. The remaining 277 data points were then analysed against summer mean monthly rainfall, mean annual temperature and elevation, in a multiple variable binomial regression analysis. Two models were developed. In the first (model A) the CRU-TS2 data series was used, in the second (model B) the GPCP rainfall data was used.

Models A and B were compared using the deviance information criterion (DIC) statistic and by comparing the proportion of validation points that were predicted within the correct confidence quantile. The better-fitting model (based on the fraction of validation data correctly predicted) was then run again, this time fitting on all data (model C).

Results

The median sample size was 100; in 67 surveys sample sizes were below 50 and in only five below 10, while seven large surveys had sampled over 300 children.

The prevalence data from 1974 onwards have been summarized in Figure 4.2. Data were available for 18 separate years, over a 24 year period. The number of surveys carried out per year varied from five in 1979 and 1988 to 44 in 1987. The highest prevalence was observed in 1974, 1996, and 1997. Figure 4.3 shows the location of all 327 data points in northern Botswana. Figure 4.4 illustrates inter-annual variation, by displaying prevalence in 17 locations where 8 or more surveys were carried out over the study period.

In Figure 4.5 the four rainfall data sources are plotted against each other, using summer rainfall for the analysis period and for the prevalence survey locations as indicator. There was a fairly marked discrepancy between summer rainfall estimates obtained from the different

sources, even between the related CRU-MARA and CRU-TS2. Mean annual temperature was more consistent than rainfall between CRU-MARA and CRU-TS2 (Figure 4.6).

The results from the bi-variate regression of the four rainfall data sources against prevalence are shown in Table 4.2. Only the GPCP rainfall data was significantly associated with prevalence after accounting for temporal and spatial correlation and for a possible linear time trend.

Inter-annual variation in the prevalence data could not be explained by inter-annual variation in the environmental data with a meaningful level of confidence (Table 4.3). The DIC statistic of model B was slightly lower than that of model A. Also both year and elevation were significant in model B rather than only elevation in model A. Nevertheless model A predicted the validation data slightly better: in 6, 13, 15 and 15 of the 50 validation points the observed prevalence fell within the 50, 80, 90 and 95% credible intervals respectively, whereas in model B the corresponding numbers were 5, 11, 12 and 15.

Model C, fitted on all data, and using the CRU-TS2 rainfall series (Table 4.3 and Figure 4.7), also had a bad overall fit and did not predict the data well. Of the 327 prevalence points, only 10, 20, 24 and 28% were correct within the 50, 80, 90 and 95% credible intervals respectively, which was even lower than the proportion of the 50 validation data predicted after fitting on only 277 observations.

Since national level incidence data in Botswana were found to be associated with climate (Thomson *et al* 2005), we calculated the mean prevalence for the country, on an annual basis, weighted by sample size. Mean summer rainfall and mean annual temperature were calculated by year from all pixels corresponding to the survey sites, also weighted by sample size. In a

binomial regression analysis, as described above, elevation, a linear time trend and temporal correlation were accounted for. In this case the spatial correlation was not considered. Again, the temporal variation in prevalence could not be explained by the variation in climate.

Discussion

The national malaria control programme was launched in 1974 although some level of indoor residual spraying was going on since the 1940's. By 1961/62 malaria prevalence in the North of Botswana was already much below the level measured in 1944 (about 40 *vs* 70%), and by 1974 was only around 10% (Mabaso *et al* 2004). In 1997 the national control programme changed the first-line drug from chloroquine to sulphadoxine-pyrimethamine, improved access to treatment for severe malaria, replaced DDT with pyrethroids for indoor residual spraying and extended the areas covered (Thomson *et al* 2005). The period covered by this study (1974-1997) thus represents a period of more or less uniform, but possibly gradually failing, intervention. This is also suggested by the increasing incidence rates from the early 1980's to the late 1990's (Thomson *et al* 2005).

Malaria prevalence in the childhood population of Botswana from 1974 to 1997 remained at with meso- and hypo-endemic levels with only seven local prevalence estimates above 50%. Inter-annual variability is apparent from the range of prevalence rated measured in locations where surveys were carried out repeatedly over time (Figures 4.3 and 4.4). Some years appear to have seen relatively widespread increase (eg 1981, 1985, 1993 and 1996) or decrease (eg 1982, 1984, 1991 and 1994) in prevalence, however the temporal pattern was not consistent. This inter-annual variation and the high frequency of zero prevalence in known malarious areas (Figure 4.3) also points to the epidemic nature of malaria in this country.

Annual case notifications since 1990 have been mostly below 50 per 1000 population (Anon. 2003c). With its small population of around 1.8 million (Anon. 2004a) and fairly high GNP, Botswana has the highest per-capita government spending on health in the Africa WHO region. Treatment is free, and no stock-outs were recorded over a three-month surveillance period (Anon. 2003c). This speaks of a situation that is probably as good as it can be, short of complete elimination. In such a controlled situation it would have been important to incorporate into the analysis indicators of control and its effectiveness, as was possible in Chapters 6 and 7, especially considering that prevalence is more prone to non-climatic factors than incidence. However such data were not available.

As in South Africa (Chapter 7) aggregated inter-annual variation in incidence in Botswana could be linked to climatic factors, such as rainfall, temperature, vapour pressure and NDVI (Thomson *et al* 2005). The spatial and temporal distribution of malaria incidence in Zimbabwe (Mabaso *et al* 2006b) and South Africa (Chapter 7) was also associated with climate, but a similar association could not be demonstrated in these prevalence data from Botswana.

Both models A and B predicted the validation points extremely badly. Only 30% were correctly predicted at the 95% confidence level. The predictions were slightly better in model A, though model B appeared to be the better model based on other criteria. Almost all associations were negative. Elevation was a significant predictor in both models - clearly only of the spatial and not of the temporal variation in prevalence, the negative effect pointing towards lower prevalence in higher-altitude areas and *vv.*, which was not surprising. A negative year trend was also understandable, given the effect of control. But a negative, if non-significant, impact of temperature and rainfall was counter-intuitive. Higher summer rainfall and mean temperatures were not associated with an increase in prevalence in multiple

variable regression analysis. In model C, which was fitted on all data, rainfall was positively associated with prevalence, but the effect was not statistically significant.

The limited explanatory power of the climate data could be interpreted in various ways: it could confirm that prevalence is not a suitable indicator of transmission intensity in the presence of control, or it may be due to a lack of accuracy in the climate data series, or else it may be due to a bad temporal match between the malaria and climate data.

The first of these explanations, that prevalence is not a suitable indicator of transmission intensity in the presence of control, is not unlikely. At low levels prevalence can vary markedly, both in time and space, and more than is useful for interpretation, so that large sample sizes are required to get a reliable estimate of the true prevalence in the population (Molineaux *et al* 1988). Roughly, sample sizes below 100 are only sufficient to demonstrate a significant difference between proportions of 0.25 to 0.3 in any particular location. On the other hand, based on the number of survey points available for each year for the entire country, the sample sizes were large enough to demonstrate a significant change at an aggregated level (at the 95% level, and with 90% power) smaller than the observed change in 13 out of the 17 observed time steps. So broadly speaking the data were able to detect inter-annual change in prevalence in the country as a whole, but not on a location-specific basis. With such great uncertainty in the spatial-temporal malaria prevalence data, it would not be surprising if much or most of the variation could not be explained.

The second possible explanation, that the bad results are due to a lack of accuracy in the climate data series, is equally likely. The problem does not seem to have been the low spatial resolution, as the high-resolution CRU-MARA data did not explain the prevalence data better than the low-resolution data series (Table 4.2). The CRU climate coverages are based on

weather station data. Rainfall pixels were calculated based on all weather stations within a 450km radius, while temperature range was calculated from stations within 750km, and mean temperature within 1200km (Tim Mitchell, pers. comm.). The purpose of the CRU-TS2 data series is the long-term assessment of global-scale climate trends, and not precise statistical analysis against disease outcomes. In Botswana the weather station network utilized in the production of this climate series must have been extremely uneven and generally very sparse. Although the actual locations of stations are not available from the authors, images showing the number of stations considered in the calculation of each pixel show a very low average density in Botswana. In the GPCP data the density of weather stations was also extremely low. This means, while the large-scale climate trends may be reflected, perhaps even quite accurately, that much of the local variation in climate gets lost.

The third possible explanation, namely a bad temporal match between the malaria and climate data, could theoretically be assessed by limiting the analysis to survey results in which the month is known and regressing survey prevalence rates against mean climate in say the three preceding months. However, this would mean losing 140 (almost half) of the observations. Together with the lack of precision in the climate data series, further investigation along this line did not seem warranted.

That many of the data had no record of month of survey remains a potential problem, as other studies have shown that prevalence can vary substantially between the wet and dry season (Molineaux & Gramiccia 1980; Lindsay *et al* 1991). However, over 60% of surveys, for which the month was known, were carried out in the high-transmission month of March, and over 90% in the first half of the year, so that not much extra information would be gained by taking account of the month of survey. This also suggests that most of the remaining surveys

were likely to have been carried out during the first half of the calendar year as well, which corresponds to the second half of the season by which the climate data were aggregated.

The most important issue here is the fact that overall malaria prevalence post 1970 was very much below the original malaria risk, presumably due to intensive control. Areas where the vectorial capacity to transmit malaria is already marginal, responded extremely well to vector control. This is evidenced by the significant reductions in malaria transmission over the past decades observed in several southern African countries (Mabaso *et al* 2004). Even hyper- and holo-endemic areas can respond well to comprehensive and integrated control (Sharp *et al* 2007). Malaria control, and the efficacy of control, thus complicate the association between malaria and climate.

A study in South Africa (Craig *et al* 2004a) found that in the presence of intense control, with transmission well below pre-control levels, overall levels and longer-term trends in annual malaria incidence were associated with non-climatic factors, predominantly the effectiveness of control. Nevertheless inter-annual variation in malaria incidence could be explained largely by inter-annual trends in rainfall and temperature (Craig *et al* 2004b). Other studies in the region have also found temporal links between climate and malaria incidence in Zimbabwe (Freeman & Bradley 1996; Mabaso *et al* 2006b) and Botswana (Thomson *et al* 2005).

While incidence rates appear to remain prone to variations in climate, even in the presence of intensive malaria control operations, this study suggests that the effects of vector control and good access to treatment over-ride the effects of climate on malaria prevalence.

Conclusion

In this study we examined childhood prevalence in Botswana over a period during which malaria control was intense, to determine whether inter-annual variation in prevalence, like incidence, was associated with inter-annual variation in climate. The results showed that available monthly climate time series were not useful in predicting inter-annual changes in prevalence.

Acknowledgements

We thank the Botswana Ministry of Health for contributing their data to the efforts of the MARA project. We thank the South African Medical Research Council and the Rudolf Geigy Stiftung zu Gunsten des Schweizerischen Tropeninstituts for supporting this study, as well as the various funders of the larger MARA project (particularly MIM / TDR and RBM).

Table 4.1 Four long-term spatial monthly rainfall data sources included in this study.

	Spatial resolution	**Time period**	**Stations used**
CRU-TS2	0.5°	1901-2000 (all data, 327 surveys)	Number of stations used for calculating each pixel (within 600km radius), and not the actual distribution of stations, are provided.
CRU-MARA	0.05°	1951-1995 (1996/97 excluded, 281 surveys left)	Stations per pixel can be calculated from station locations
GPCP	0.5°	1951-present (all data)	Stations per pixel available
CMAP	2.5°	1979-2006 (1974 excluded, 303 surveys left)	No information

CRU-TS2 = Climate Research Unit Time Series 2 (Mitchell *et al* 2003)
CRU-MARA = Climate Research Unit climate data, commissioned by MARA project (New & Hulme 1997)
GPCP = Global Precipitation Climatology Project (Adler *et al* 2003)
CMAP = Climate Prediction Centre Merged Analysis of Precipitation (Anon. 2007a)

Table 4.2 Co-efficients in bi-variate spatio-temporal analysis of different co-variates against malaria prevalence in 5 to 10 year old children in Botswana, 1979 to 1995, n=281. The 95% credible interval is shown in parentheses, significance at the 95% level is indicated with an asterisk.

	Year effect	**Co-efficient**	**DIC**
Elevation (per m)	-0.05 (-0.39 to 0.18)	-0.004 (-0.01 to 0.002)	1840
Summer rainfall (per mm)			
CRU-TS2	-0.009 (-0.7 to 0.4)	0.003 (-0.001 to 0.006)	1839
CRU-MARA	0.002 (-0.5 to 0.3)	-0.003 (-0.008 to 0.002)	1840
GPCP	-0.03 (-0.4 to 0.4)	-0.005* (-0.009 to -0.0005)	1836
CMAP	-0.03 (-0.7 to 0.25)	0.007 (-0.0001 to 0.01)	1836
Mean temperature (per °C)			
CRU-TS2	-0.07 (-0.5 to 0.2)	0.6* (0.01 to 1.1)	1831
CRU-MARA	-0.1 (-0.7 to 0.2)	1.3* (0.7 to 1.6)	1782

Table 4.3 Median incidence rate ratio and the 95% credible interval of co-variates fitted on malaria prevalence in 5 to 10 year old children in Botswana, 1974 to 1997, for three different multivariate spatial-temporal models. Significance at the 95% level is indicated with an asterisk.

Variable	Median Incidence Rate Ratio	95% credible interval	
		lower	upper
Model A: results using CRU-TS2 rainfall series fitted on derivation data only (n = 277, DIC = 2296)			
rainfall summer mean (per mm)	0.999	0.995	1.002
temperature annual mean (per °C)	0.799	0.6	1.047
elevation (per m)	0.996*	0.993	0.999
year trend	0.883	0.741	1.131
Model B: results using GPCP rainfall series, fitted on derivation data only (n = 277; DIC = 2293)			
rainfall summer mean (per mm)	0.998	0.995	1.001
temperature annual mean (per °C)	0.795	0.635	1.066
elevation (per m)	0.995*	0.993	0.999
year trend	0.871*	0.683	0.976
Model C: results using CRU-TS2 rainfall series, fitted on all data (n =327; DIC = 2729)			
rainfall summer mean (per mm)	1.001	0.998	1.004
temperature annual mean (per °C)	0.876	0.643	1.078
elevation (per m)	0.996*	0.992	0.999
year trend	0.997	0.763	1.176

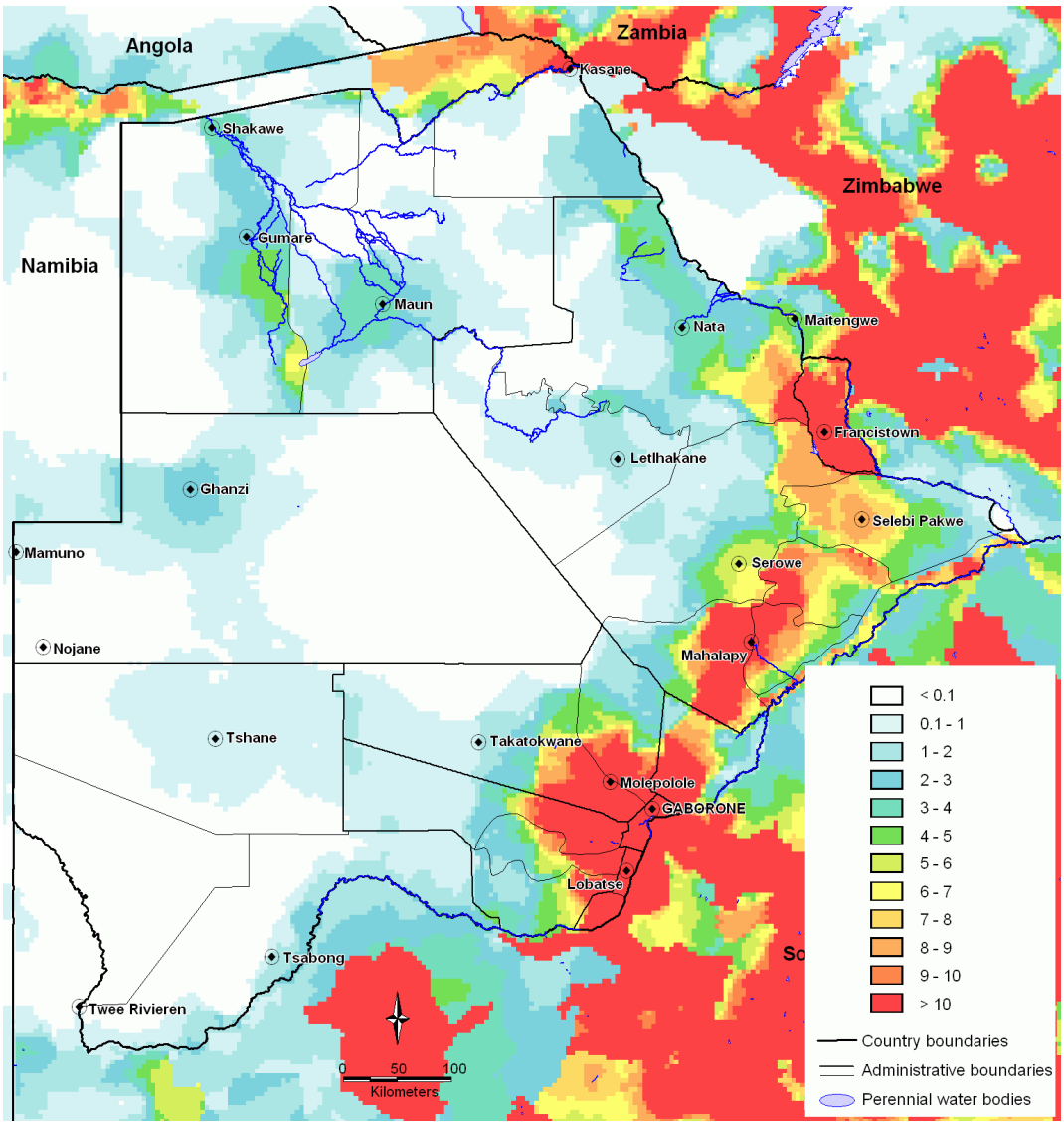


Figure 4.1 Total population density in Botswana per square km, 1995 (Deichmann, 1997).

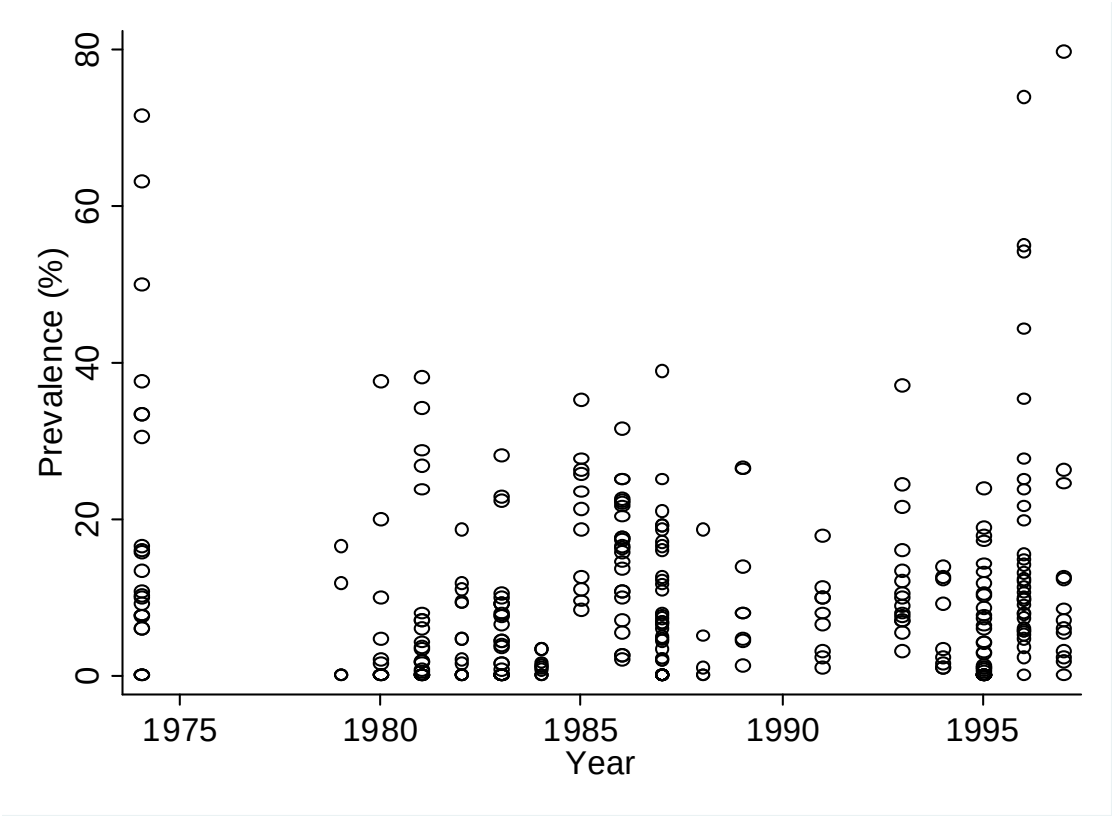


Figure 4.2 Malaria prevalence of infection in 5 to 10 year old children, in Botswana, for 327 surveys, from 1974 to 1997, over 18 separate years and 87 separate locations.

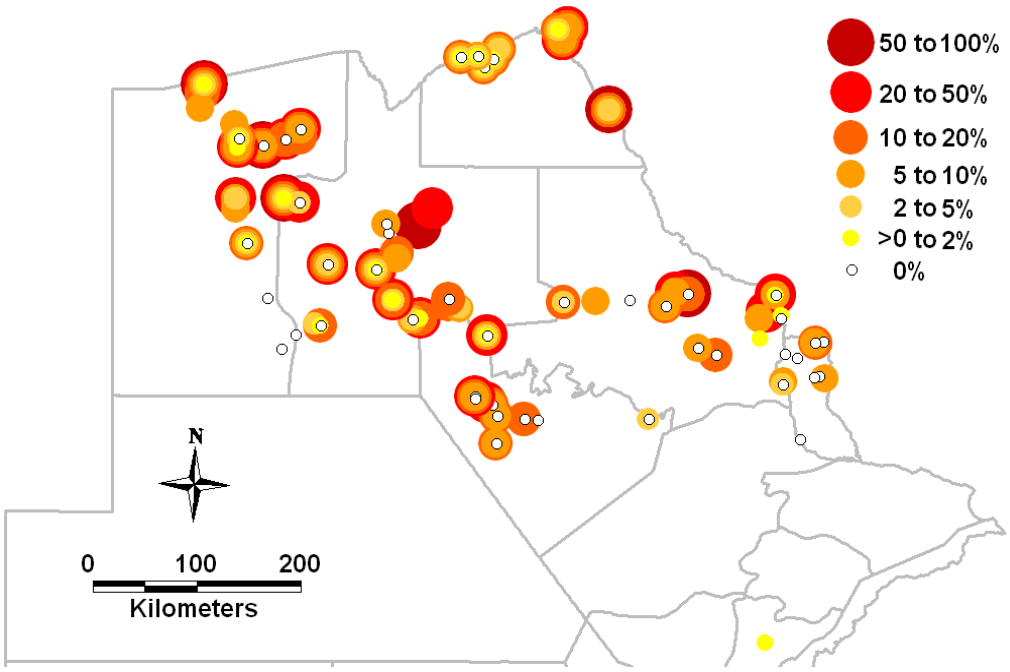


Figure 4.3 Malaria prevalence of infection in 5 to 10 year old children, in northern Botswana, 1974 to 1997. Lines represent district boundaries.

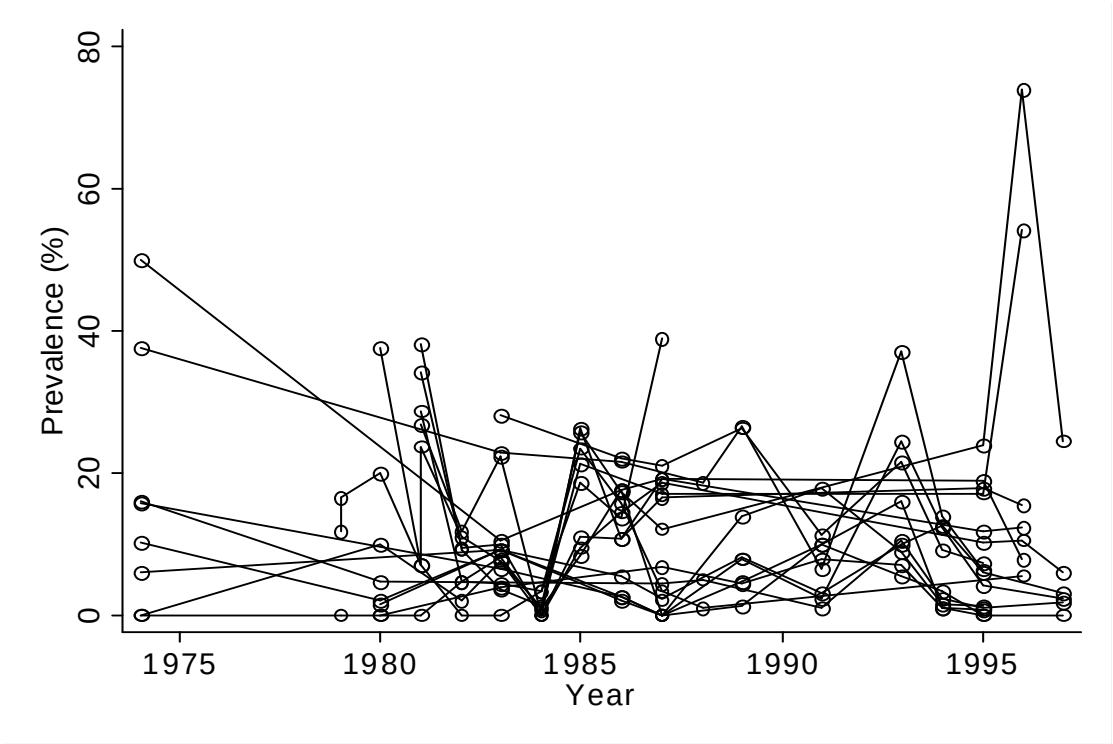


Figure 4.4 Malaria prevalence of infection in 5 to 10 year old children, in 17 locations in northern Botswana, where eight or more surveys were carried out over the period 1974 to 1997.

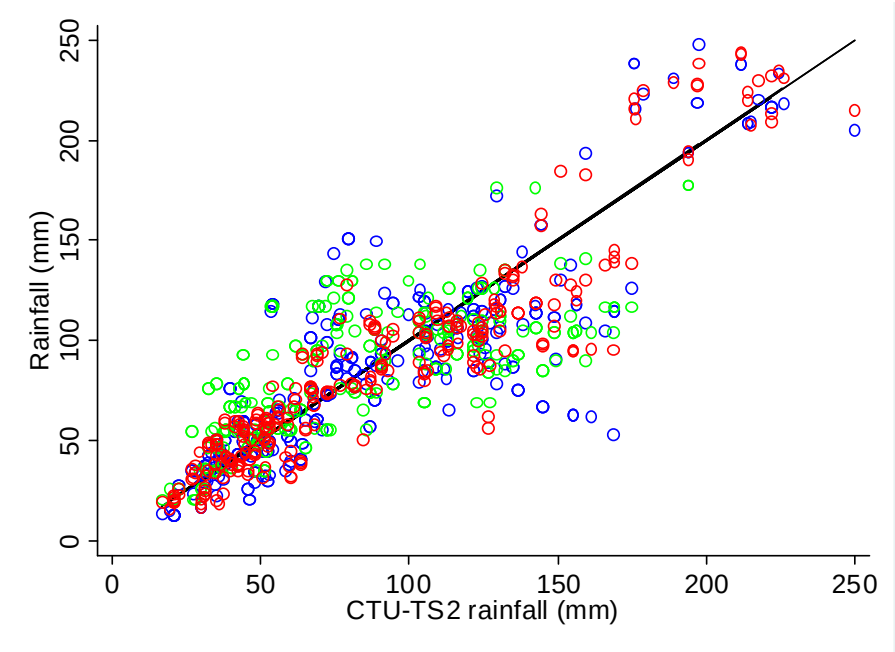


Figure 4.5 Summer (December to March) rainfall by year, for 287 surveys covered temporally by all four rainfall data sources, plotted against CRU-TS2: GPCP (blue), CMAP (green), CRU-MARA (red) and CRU-TS2 (line).

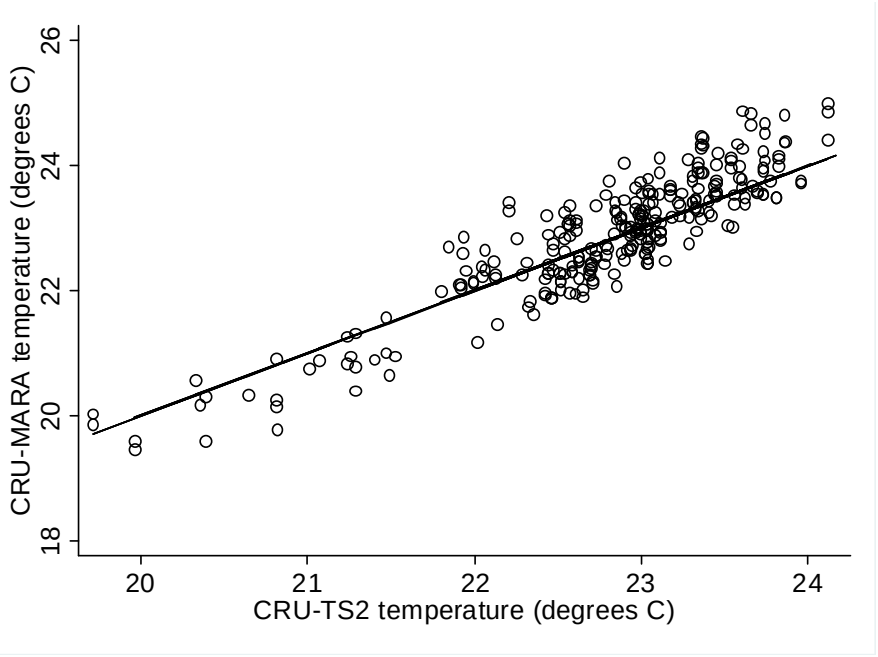


Figure 4.6 Mean annual temperature (calculated over 12 month periods starting in July, ending in June), by year and by location, for the two different temperature data sources: CRU-MARA on y-axis, CRU-TS2 on the x-axis.

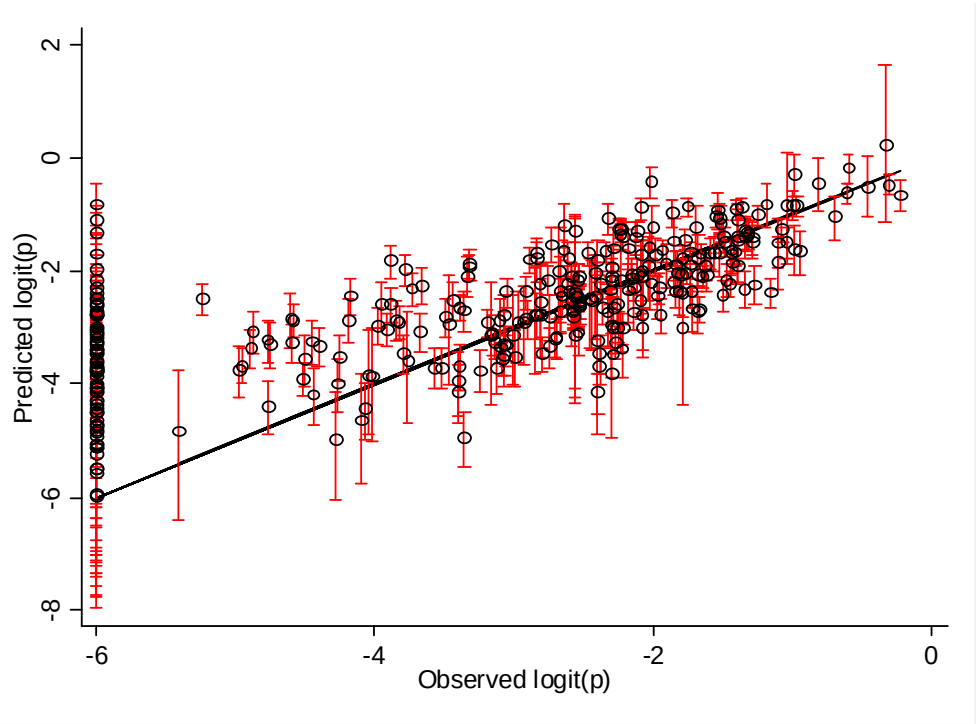


Figure 4.7 Predicted prevalence plotted against observed prevalence, on the logit scale (hollow circles), and the 95% credible interval (red vertical lines), after fitting the model on all 327 malaria prevalence data points for Botswana in children 5 to 10 years old, from 1974 to 1997.

Chapter 5

Exploring thirty years of malaria case data in KwaZulu-Natal, South Africa, Part I: the impact of climatic factors

M.H. Craig[1], I. Kleinschmidt[1], J.B. Nawn[1], D. le Sueur[1], B.L. Sharp[1]

Affiliations:

[1] Malaria Research Programme, Medical Research Council, PO Box 17120, Congella, 4013 Durban, South Africa. Tel: +27-31-2043653, Fax: +27-31-2051498, email: craigm@mrc.ac.za

Keywords:

climate, disease outbreaks, disease reservoirs, environment, incidence, malaria, risk factors, risk seasons, South Africa, transmission

Publication status:

Tropical Medicine and International Health ***9***(12): 1247–1257, 2004

Abstract

Large parts of Africa are prone to devastating malaria epidemics. Advance epidemic warning would give health services an opportunity to prepare. Because malaria transmission is largely limited by climate, climate based epidemic warning systems are a real possibility. To develop and test such a system, good long-term malaria and climate data are needed. In KwaZulu-Natal, South Africa, 30 years of confirmed malaria case data provide a unique opportunity to examine short- and long-term trends. Seasonal case totals and seasonal changes in cases (both log-transformed) were analysed against a range of climatic indicators obtained from three weather stations in the highest malaria incidence districts, using linear regression analysis. Seasonal changes in case numbers (delta log cases) were significantly associated with several climate variables. The two most significant ones were mean maximum daily temperatures from January to October of the preceding season ($n = 30$, $r^2 = 0.364$, $p = 0.0004$) and total rainfall during the current summer months of November to March ($n = 30$, $r^2 = 0.282$, $p = 0.003$). These two variables, when entered into the same regression model, together explained 49.7% of the total variation in delta log cases. We found no evidence of association between case totals and climate. In KwaZulu-Natal, where malaria control operations are intense, climate appears to drive the inter-annual variation of malaria incidence, but does not determine its overall level. The accompanying paper (Chapter 6) provides evidence that overall levels are associated with non-climatic factors such as drug resistance and HIV prevalence.

Introduction

Over large parts of Africa malaria transmission occurs in distinct seasons and epidemics, which vary in severity from year to year. Epidemics may be caused by a range of factors including movement and displacement of human populations, breakdown of control, environmental changes, and meteorological / climatic factors (Nájera *et al* 1998).

Climatic determinants are considered particularly important, since both the disease agent (*Plasmodium*) and vectors (*Anopheles* mosquitoes) are strongly affected by climate: temperature determines parasite and vector development, rainfall provides mosquito breeding sites, and humidity, together with temperature, affects mosquito survival. It is thus believed that malaria epidemics caused by meteorological factors can be predicted from climatic indicators and climate forecasts.

In climatic conditions that are marginally suitable for transmission, patent and severe epidemics may follow extreme climatic conditions. The link between climate and malaria in strongly seasonal but endemic settings is probably weaker and sometimes masked by non-climatic variation. Yet climate still plays an important role in driving inter-annual variation in transmission, and climate data my prove useful for alerting malaria control authorities to unexpected risk of epidemic conditions.

Climate-based epidemic warning systems rely on the relationship between climate and malaria incidence. To demonstrate and quantify such a relationship, good, long-term malaria data sets are required, but these are rare in sub-Saharan Africa. Thirty years of incidence data from KwaZulu-Natal province (KZN), South Africa, provide a unique perspective on temporal patterns of malaria incidence over the past three decades. In this and the accompanying paper

(Craig *et al* 2004a) we focussed on the seasonal case totals, examining possible reasons for the observed season to season and longer term changes.

Malaria incidence in KZN

During the first half of the 20th century malaria was far more widely distributed in South Africa than it is now; some areas suffered from hyper-endemic conditions (Swellengrebel 1931). Resource-intensive control measures were put in place in the late 1940's (Le Sueur *et al* 1993). The malaria control programme rendered large, previously endemic areas practically malaria-free. That malaria in South Africa is at its southern limit must have aided the situation because re-invasion of controlled areas was only possible from the North.

Malaria is a notifiable disease in South Africa. The case reporting system aims to capture every infection rather than clinical cases only, through both passive and active surveillance (Sharp *et al* 1988). Malaria transmission is distinctly seasonal, with transmission limited to the warm and rainy summer months. Case numbers are available for the KZN province from 1970 and by district from 1981. Of three malarious provinces, KZN reports on average half of the country's cases, and over last twenty years about 90% of KZN cases have been reported from its two northern-most districts, Ingwavuma and Ubombo (Figure 5.1).

Case notifications generally increase from November onwards, peak in late summer to autumn (March to May) and decline by the end of June. The average seasonal pattern in malaria incidence follows the periodicity in rainfall and temperature with a three to four month lag.

Two longer-term trends are immediately evident in this 30 year data series (Figure 5.2): firstly a steep exponential upward trend. A spatial comparison of malaria incidence between the 1988/89 and 1998/99 seasons also confirmed geographical spread, in that the steepest

increases were observed in areas with previously lowest incidence, and *vice versa* (Kleinschmidt *et al* 2002). An almost ten-fold increase in cases over five years (from 1994/95 to 1999/2000), in spite of a malaria control programme which was costing the state R82m in 1999 (at that time about US$13m, Rajendra Maharaj, then national malaria control manager, pers. comm.) requires an explanation.

Statistics South Africa (Orkin 1999) cite an average exponential population growth rate of 0.024 per annum for rural communities in KZN for the period 1991 to 1999. At this rate the population would have doubled over 30 years. The exponential growth rate of malaria on the other hand has been around 0.206 per annum (see Figure 5.2 for details). This translates into an increase, over 30 years, of 400 times. Even if population growth was higher in the early part of the period, population growth cannot account for the strong long-term trend in malaria cases.

The second obvious trend is the occurrence of ‘epidemic jumps’, such as in 1975, 78, 80 and 84, and more visibly in 1987, 93 and 96. In Figure 5.2 the degree to which cases increased or decreased compared to the previous season is shown. It was calculated as the log of (current season total / previous season total) or simply ‘delta log cases’. A high value corresponded to an ‘epidemic jump’ in incidence.

The impact of climate

Epidemiological models of malaria describe the rate at which susceptible members of a population become infected, and infected members are recovering (Dietz 1988; Anderson & May 1991). The reproductive rate (R_o) describes the rate of disease propagation. While R_o is above one the disease spreads; when it is below one the disease declines. In a seasonal setting

R_o increases as conditions for transmission improve, causing incidence to rise. At the end of the season conditions deteriorate, R_o decreases and transmission ceases.

Several factors in this system are dependant upon climate: mosquito abundance, survival and biting rate, and extrinsic parasite development. Theoretically, high-incidence seasons (and epidemics) are most likely when the transmission season is preceded by a warm and moist winter (allowing greater mosquito survival and breeding in winter, hence a larger starting population in spring); if the preceding season was very wet (a high water Table may lead to increased pooling of the first rains and thus earlier availability of breeding sites); if the previous season's climate was generally more favourable (resulting in larger parasite and vector populations); if favourable conditions persisted for longer in the previous season (resulting in a greater reservoir of infected people at the start of the transmission season); if rains start or if temperatures rise earlier than normal (increasing early growth of vector populations and earlier completion of extrinsic parasite development); if mean conditions during summer are more favourable than normal (increasing the rate of transmission); or if warm and moist conditions persist later into autumn than usual (allowing transmission to persist longer than usual). An epidemic may be least likely if these conditions are reversed.

Methods

Malaria case data for 1981 onwards were extracted for KZN province from the malaria information system housed at the Medical Research Council. These are confirmed case reports submitted by the Malaria Control Programme. For the years 1971 to 1980 provincial data from the National Department of Health were used.

Cases were aggregated by annual season, a season for a particular year being defined as the time from 1st July of the previous year to 30th June of the current year. A measure reflecting the changes in case numbers between consecutive years was defined as

$$\text{delta log cases} = \log(y_t) - \log(y_{t-1}), \text{ or } \log(y_t / y_{t-1}),$$

where y_t is the seasonal case total for season t.

Meteorological data were obtained from three weather stations in Ingwavuma and Ubombo districts (Figure 5.1). Rainfall data for the full time period were available from Makhatini and Ndumu, and temperature data from Makhatini and Mantuma. Monthly mean daily maximum temperature (Tmax) was calculated as the mean of daily maximums for each month. Monthly mean daily temperature (Tavg) was calculated as the average of daily minimums and maximums for each month.

The climate risk factors were summarised and calculated for each season as indicated in Table 5.1. Seasons were defined as follows: winter = June to August; spring = August to November; summer = November to March; autumn = March to June.

Spring indicators (*x*s) were calculated using weighted means:

$xs_t = (4x_{t,Aug}+3x_{t,Sep}+2x_{t,Oct}+x_{t,Nov}) / 10$ where $x_{t,month}$ is the mean monthly temperature (maximum or average) in °C or monthly rainfall in mm for year t. This was done to weight the mean in favour of early spring values. Similarly, yearly autumn means, xa, were calculated, weighting in favour of later conditions, i.e. $(x_{t,Mar}+2x_{t,Apr}+3x_{t,May}+4x_{t,Jun}) / 10$. The differential weighting was intended to capture situations where suitable conditions start earlier or persist longer than normal, thus extending the transmission season at both ends.

Single variable linear regression analysis of log cases and delta log cases was carried out against each of the variables shown in Table 5.1. Every significant variable was tested for random walk using the Augmented Dickey Fuller (ADF) test (Dickey & Fuller 1979), and for evidence of trend over the 30-year time period. Analysis was carried out using STATA version 7 (Anon. 2001b) and SPSS version 10 (Anon. 1999b).

Since the temperatures of the previous season, summer, autumn, and spring were all significant in single-variable regression analysis, and since there was cross-correlation between them, it was thought best to replace them with a single summary temperature variable. To obtain the best summary mean, monthly temperatures were systematically averaged unweighted over different contiguous time periods. Each combination was in turn regressed against delta log cases. Starting with the definition for "previous autumn" ($x_{t-1,Mar}$ to $x_{t-1,Jun}$), months were added consecutively onto the start of this period, up to and including previous October ($x_{t-1,Oct}$). Correlation with delta log cases increased, reaching a maximum for the Jan-Jun period ($x_{t-1,Jan}$ to $x_{t-1,Jun}$), and decreased thereafter. Next, months were added consecutively to the end of this period, up to and including current January ($x_{t,Jan}$). Again correlation increased, reached a maximum for the January-October period ($x_{t-1,Jan}$ to $x_{t,Oct}$), decreasing thereafter. The same procedure was followed using maximum monthly temperatures, but the correlation between delta log cases and mean temperature was stronger. Mean monthly temperature during the period $x_{t-1,Jan}$ to $x_{t,Oct}$ was thus established as the best summary temperature variable.

Current season rainfall was better represented by current summer rainfall and was not used in the multiple variable analysis. None of the remaining three, i.e. rainfall in previous season, preceding winter and current summer, were correlated with each other and were therefore retained.

The summary temperature variable and three rainfall variables were entered together into a multi-variate linear regression analysis to determine their combined effect. Variables that were non-significant in the combined model were removed. The final model contained summer rainfall and the summary temperature indicator. Regression residuals were tested for serial auto-correlation by constructing correlograms and computing the Box-Pierce Portmanteau Q-statistic (Box & Pierce 1970) for lags from 1 to 8 years.

For comparative purposes precipitation and temperature data was also extracted from the global CRU-TS2 time series supplied by the Climate Research Unit (Mitchell *et al* 2003). This gridded data set is available by month for 1901 to 2000, at a 0.5 degree spatial resolution (approximately 50km). The grids represent interpolations of weather station data. Six pixels overlap with the two districts of Ingwavuma and Ubombo. Precipitation, mean temperature and temperature range data were extracted for each of these six pixels. A mean was calculated for the six pixels. Mean maximum daily temperature was calculated as mean temperature + 0.5*range.

Climate indicators were derived as with the weather station data, and regression analysis was carried out. As before, a summary temperature indicator was computed and, together with summer rainfall, regressed against delta log cases in a multi-variate linear regression analysis.

Results

Several climate indicators were associated with delta log cases, but none were significantly associated with seasonal case totals (direct or log-transformed). Table 5.2 shows all variables from Table 5.1 that were significantly correlated with delta log cases at the 5% probability level. The two most strongly correlated variables are shown in Figure 5.2. The coefficients of determination (r^2) were relatively low, and even the most significant explanatory variable only

explained 28% of the total variance in the data. The summary temperature indicator (Figure 5.2) explained 36.4% of the total variation in delta log cases (n = 30, p = 0.0004).

We found significant evidence against the null hypothesis of random walk for any of the climatic variables, when applying the ADF test to the series, and no evidence of time trend (Table 5.2). In the absence of evidence for a time trend any association with malaria case data was therefore not simply an association of two quantities that possess a long term drift.

The first multiple variable regression model including four variables (three rainfall and one temperature variable) explained 57% of the variation in delta log cases. Two variables, namely previous season rainfall and previous winter rainfall, were no longer significant, and were removed. In the final model, the summary temperature variable (daily average temperatures during preceding January to October) (T) and total rainfall during the current summer (R) together explained 49.7% of the total variation: predicted delta log cases = -10.649 + 0.463T + 0.001R (n = 31, overall $P < 0.00001$; P of coefficient T = 0.002 and of R = 0.013).

There was no evidence of auto-correlation between seasons in the residuals of the final model (Table 5.3). This shows that any temporal dependence between observations is explained by the model, and standard errors of model coefficients are not likely to have been under-estimated as a result.

Regression diagnostic tests on the final model revealed no evidence of pattern in the residuals with fitted values (test for heteroskedasticity, P=0.96), confirming that regression modelling assumptions had not been violated. The two variables used in the final model and the

predicted delta log cases from this model were plotted against observed delta log cases (Figure 5.3).

The response delta log cases, as predicted from the model, was transformed back to the predicted change in cases and multiplied with the observed previous season case totals:

$$Y_t = y_{t-1}\Delta_{t,t-1}$$

where Y_t is the predicted number of cases for each season, conditional on the previous season's observed cases y_{t-1} and the ratio of current : previous year's cases $\Delta_{t,t-1}$ predicted from the model (by exponentiation of fitted delta log cases). The results (Figure 5.4) show that the variability in the data series can be reproduced remarkably well from the climate variables and case totals observed during the previous season.

In the CRU-TS2 data analysis, only average temperature during the last autumn, and maximum temperature during the previous season and summer were significantly associated with delta log cases (see Table 5.1). Correlation was higher when the summary temperature variable was used. Of the rainfall variables only current summer rainfall produced marginally significant results. Correlation with the previous season rainfall was also negative, but non-significant ($P = 0.1$). No CRU-based variable was significantly correlated with delta log cases that was not also correlated in the station-based data.

When summer rainfall and the summary temperature variable were regressed together, the combined equation could explain 30.6% of total variation; both explanatory coefficients were significant: predicted delta log cases = $-12.53 + 0.554T + 0.00093R$ ($n = 29$, overall $P < 0.0087$; $P(T) = 0.015$ and $P(R) = 0.045$).

Discussion

Climate seems to be the main limiting factor of malaria, restricting transmission both spatially and temporally on a large scale (Craig *et al* 1999; Tanser *et al* 2003). Even at the small scale, analysis of the *spatial* heterogeneity of malaria incidence in northern KZN (Kleinschmidt *et al* 2001b) showed that case incidence in 1994/95 was significantly related to average winter rainfall, average winter maximum temperature and also inversely related to the distance of the nearest mapped water body. It is therefore not surprising that climate was also associated with the general *temporal* variability of malaria transmission.

The fact that total seasonal case numbers were not associated with climate firstly needs to be viewed against the background of a highly controlled situation. For example, recently collated prevalence data from Swaziland, which borders our study area (MARA/ARMA, unpublished data) showed that prevalence dropped from around 80% to <10% when malaria vector control was implemented.

Secondly, the malaria case totals display a strong long-term trend, over two orders of magnitude during the past decade alone (582 cases in 1992, 34 364 in 2000). No such trend was found in any of the meteorological indicators. Under these conditions a lack of association between seasonal case totals and the climate variables examined, is not surprising. A study of four localities in the East African highlands, where marked increases of malaria incidence had been reported, also found no evidence that these were due to climate (Hay *et al* 2002a).

Against this background the strength of the correlation, found between climate and inter-seasonal variability of malaria as represented by the quantity delta log cases, is noteworthy. The association between delta log cases and high maximum temperatures in the

previous autumn suggests that temperature limits transmission at the end of the season and conversely, that prolonged transmission occurs during unseasonably warm autumns. The later transmission comes to a standstill, the more infections survive through winter, swelling the reservoir that seeds the next transmission season. Prolonged higher temperatures also sustain larval development, resulting in larger over-wintering mosquito populations.

Temperatures during the preceding summer and current spring were also significantly associated with log delta cases, suggesting that the effect is a combination of high rate and prolonged duration of transmission in the previous season, as well as rapid growth of parasite and / or vector populations in early spring.

The only way to explain the strong predictive value of the summary temperature indicator is that temperatures over the entire period, i.e. from the time of peak transmission in the previous season all through winter, up to early spring, play a role in determining the size of the reservoir of parasites and /or mosquito vectors surviving and seeding the following season, and consequently making an increase in cases more likely.

The dominant local vector, *Anopheles arabiensis* (White 1974) breeds in shallow sun-lit temporary pools formed by rain, by hoof prints at the edge of water bodies, or similar sites (Gillies & de Meillon 1968). Availability of breeding sites during summer, when mosquito development rate is highest, would impact more on total vector populations than rain during the rest of the year when transmission rate is lower and mosquito development slower. This may explain why summer rainfall was a better predictor of delta log cases than total seasonal rainfall.

Winter rainfall was only marginally significant, and failed to contribute significantly towards predicting delta log cases in the combined model. The anticipated effect of moist winters (enhancing mosquito survival during the dry months and increasing the chance of an epidemic surge in the following season) was therefore minor and eclipsed by the effect of a wet summer.

The negative correlation of delta log cases with the previous summer's rainfall can be explained as follows: unusually high rainfall during the previous summer would have been accompanied by a greater increase in cases relative to the season before that. But a rise in transmission is followed by a rise in immunity, which reduces the chances of another epidemic of the same magnitude following straight after. This phenomenon has been described in terms of human vulnerability to epidemics (Thomson & Connor 2001a). Signs of the temporary acquisition of immunity in malaria affected populations of KwaZulu-Natal have been previously observed (Kleinschmidt & Sharp 2001). It is interesting that these data confirm this statistically.

In Figure 5.3(c) it can be seen that in 13 out of 30 seasons, cases were correctly predicted to increase, in 10/30 seasons case numbers were correctly predicted to decrease, and in 7 seasons a decrease was predicted when an increase was observed, or *vv*. However, in years when the direction of change was predicted correctly, the predicted magnitude of the change was not necessarily accurate.

Figure 5.4 shows the outcome of applying the change in cases predicted from the model to the previous season's observed case totals. The rise in case numbers observed in 1987,1988, 1993, 1996, 1999 and 2000 for example, corresponded with increases predicted from climate

data, although the predictions frequently over- or under-estimated case numbers actually reported.

With regard to the seeming over-prediction in 1999/2000 it must be added that case reporting is known to have suffered greatly during this and the previous one or two seasons; the great influx of patients simply did not allow regular paperwork to continue as normal. It is known for example that thousands of confirmed cases from certain clinics were never incorporated into the district case register.

It needs to be emphasized that, since the rainfall data used in the analysis was data from the current summer, the “prediction” shown in Figure 5.4 is not a forecast in time. Whether actual temporal forecasts can be achieved through long range weather forecasts, will be the subject of further investigation.

Correlations of delta log cases with the CRU-TS2 climate data were weaker than with the weather station data. Many variables, including summer rainfall, were only marginally significant, or not at all. The multiple variable regression equation was significant, though it explained substantially less of the overall variation in the malaria data.

As mentioned, the CRU-TS2 data represent interpolations from weather stations, with each pixel value being calculated from available station data within a 450km radius in the case of precipitation and a 1200km radius in the case of mean temperature. In our study area pixel values draw on data from 40 to 60 weather stations within a radius that includes the Drakensberg mountains (±10°C mean annual temperature and ±1000mm total annual rainfall), the South African ‘highveld’ plateaux (15°C, 700mm), humid tropical Southern Mozambique (25°C, 700mm) and part of the dry Limpopo River valley (25°C, 400mm).

We observed considerable discrepancy between the station data and the CRU-TS2 data, particularly in precipitation, and particularly in the seasonal means. Temperature data were well correlated by month, but not by season. It is clear that a global data set such as this, prepared at a coarse resolution, would be suitable for global rather than local applications (Patz *et al* 2002).

Given these facts, it is remarkable that some of the climate variables were correlated with delta log cases, and that they were the very ones that produced the best results in the weather station data. This strengthens our confidence in the outcome of the analysis, suggesting that observed associations between delta log cases and climate indicators were real and not accidental.

Weather station data, it appears, provide better indication of local climate than the CRU-TS2 data, and should be used for local studies if available. Another obvious reason for using weather station data is that they are available immediately, whereas interpolated climate data can only be obtained in retrospect. They require special skills and resources to prepare and their accuracy depends on how many weather stations are available within a reasonable distance.

Conclusion

While the overall levels of malaria incidence in KZN could not be explained by climate, the inter-seasonal variability was correlated with several climate indicators. As with most retrospective observational studies, temporal coincidence of peaks and troughs cannot offer evidence of causal links. Specifically, in this case, the link between a meteorological measurement, such as rainfall, and malaria incidence is not direct. Rather the two are connected through a chain of biological processes, which are non-linear and still ill-defined in

practice. But the indicators were defined according to their expected role in transmission, and the significant ones were thus plausible in terms of the aetiology of the disease.

This paper, along with the accompanying paper, illustrates the risks involved with over-simplified approaches which try to establish statistical association between long term health data and climate variability and / or climate change, without looking for possible alternative explanations for observed trends.

Whether seasonal malaria incidence can be predicted in advance, with sufficient accuracy and time to help plan health care and control, remains to be seen. If the season-to-season variation is indeed driven by climate, and the overall level by non-climatic factors, then the road to malaria prediction has become a little clearer: quantifying and accounting for non-climatic determinants, which may alter or override the specific effect of climate, will be essential when working towards a climate-based predictive system.

Acknowledgements

Sincere thanks go to Prof. Linda Haines for extensive help with statistical aspects, Amanda Jackson for assistance in the malaria data summaries, Dr Simon Hay for valuable comments on the manuscripts and to the South African Medical Research Council for financial support.

Table 5.1 Climatic risk factors analysed against malaria case data.

Risk factor	Indicator
Larger reservoir of vectors / parasites	
Wet preceding season	Total rainfall during preceding season (July to June)
Warm preceding summer	Mean daily maximum / average temperature in preceding summer (November to March)
Prolonged high rainfall in preceding autumn	Weighted average of rainfall (March to June)
Prolonged high temperatures in preceding autumn	Weighted average of mean daily maximum / average temperature (March to June)
Transmission starts early	
Water Table high	Total rainfall during preceding season (July to June)
Preceding winter wet	Total rainfall during preceding winter (June to August)
Preceding winter warm	Mean daily maximum / average temperature during winter (June to August)
Early spring rain	Weighted average of rainfall (August to November)
Early rise in spring temperatures	Weighted average of mean daily maximum / average temperature (August to November)
Transmission unusually high	
Wet summer	Total rainfall in current summer (November to March)
Warm summer	Mean daily maximum / average temperature in current summer (November to March)
Wet season	Total rainfall in current season (July to June)
Transmission ends late	
Prolonged autumn rainfall	Weighted average of rainfall (March to June)
Prolonged high autumn temperatures	Weighted average of mean daily maximum / average temperature curve (March to June)

Table 5.2 Results of single variable linear regression of delta log cases against climatic explanatory variables obtained from weather station data (n=31). Only significant results are shown. Relevant results from regression against the CRU-TS2 data set are shown in parentheses (n=29).

Variable	Linear regression analysis			Random walk and trend analysis		
	Coefficient A [a]	r^2	*P*-value	ADF test statistic [b]	*t*-statistic for trend [c]	*p*-value for trend
Temperature (°C)						
Mean during previous season						
Average temperature	0.432	0.155	0.032 (0.086)	–4.914	0.81	0.425
Maximum temperature	0.376 (0.424)	0.236 (0.150)	0.007 (0.035)	–5.056 (–4.290)	–0.80 (1.68)	0.430 (0.105)
Mean during previous summer						
Maximum temperature	0.266 (0.341)	0.181 (0.144)	0.019 (0.039)	–5.223 (–4.718)	–0.86 (1.35)	0.395 (0.188)
Weighted mean, previous autumn						
Average temperature	0.370 (0.381)	0.266 (0.163)	0.004 (0.027)	–4.967 (–4.734)	1.00 (1.64)	0.327 (0.112)
Maximum temperature *	0.260	0.280	0.003 (0.61)	–5.587	0.26	0.799
Weighted mean during spring						
Average temperature	0.26	0.133	0.047 (0.46)	–5.723	0.37	0.713
Rainfall (mm)						
Total during previous season	–0.00091	0.196	0.014 (0.10)	–4.875	–0.21	0.832
Total during preceding winter	0.00812	0.133	0.048 (0.32)	–5.599	–0.21	0.838
Total during summer *	0.0014 (0.0009)	0.282 (0.128)	0.003 (0.057)	–5.93 (–4.274)	–0.32 (–0.83)	0.749 (0.416)
Total during the season	0.00089	0.181	0.019 (0.13)	–5.319	–0.12	0.902
Summary temperature variable (°C)						
Mean average temperature, previous January to October *	0.574 (0.560)	0.364 (0.187)	< 0.001 (0.019)	–5.111	0.56	0.581

[a] $y = Ax + B$; $y = \log_{10}$(cases in current season) - $\log_{10}$(cases in past season), x = climate variable. 10^A therefore is the number of additional cases per case in previous season per °C of temperature, or per mm of rain.

[b] 5% critical value = -3.580, -2.989 for the CRU-TS2 data

[c] 29 degrees of freedom

* data series shown in Figure 5.2

Table 5.3 Correlograms of the residuals of the regression model of delta log cases against summer rainfall and mean maximum daily temperatures during preceding January to October.

Lag	Auto-correlation	Partial auto-correlation	Q-statistic	Q-statistic (*P*-value)
1	-0.1312	-0.1311	0.5695	0.4505
2	-0.0529	-0.0831	0.6653	0.7170
3	-0.0834	-0.1231	0.9128	0.8223
4	-0.2284	-0.3276	2.8383	0.5852
5	0.0696	-0.0431	3.0245	0.6962
6	0.0571	0.046	3.1547	0.7892
7	-0.0809	-0.2743	3.4276	0.8428
8	0.0185	-0.1074	3.4425	0.9036

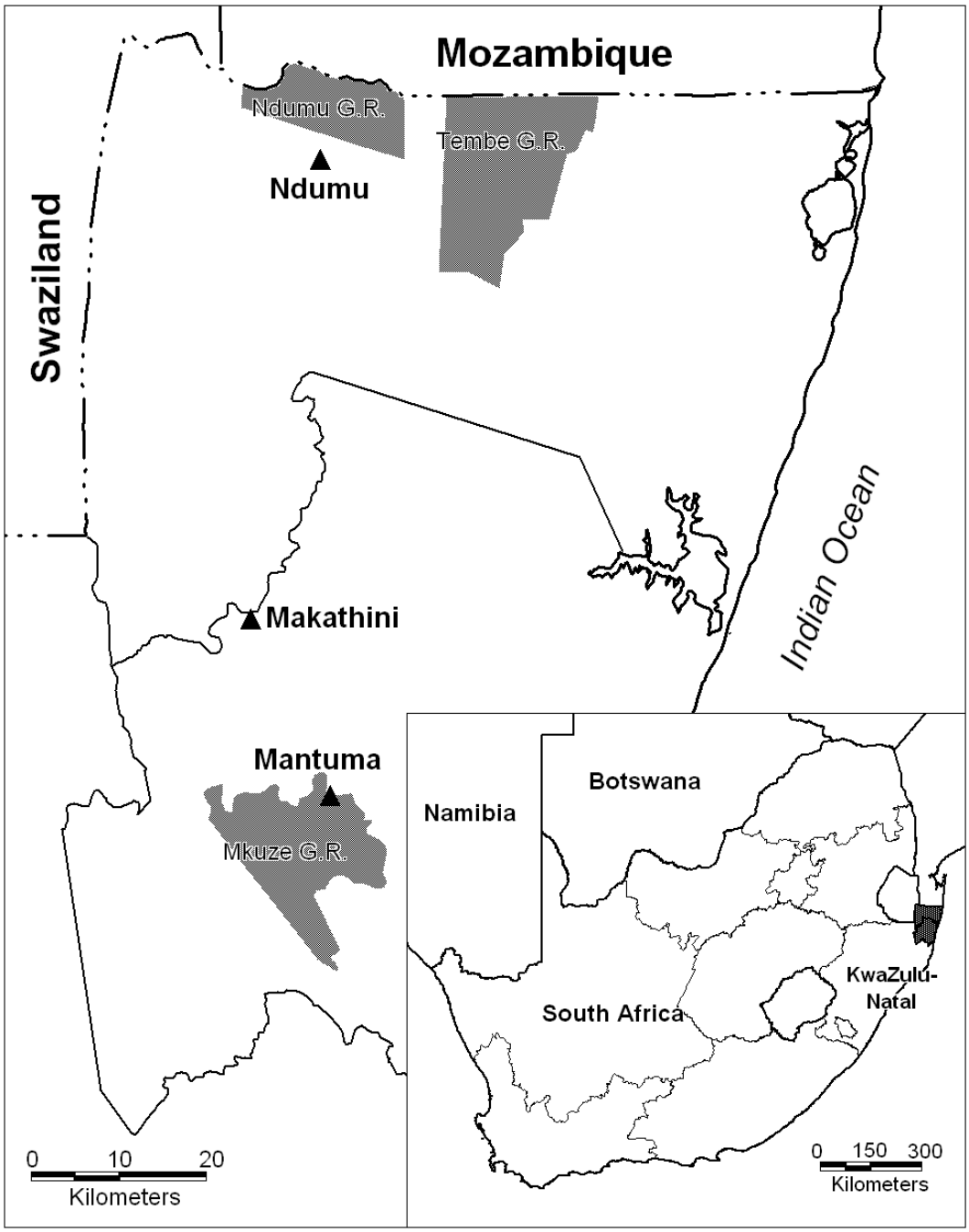


Figure 5.1 Three weather stations in Ingwavuma and Ubombo districts, northern KwaZulu-Natal. The inset shows the location of these two districts in relation to the rest of South Africa. G.R. = game reserve.

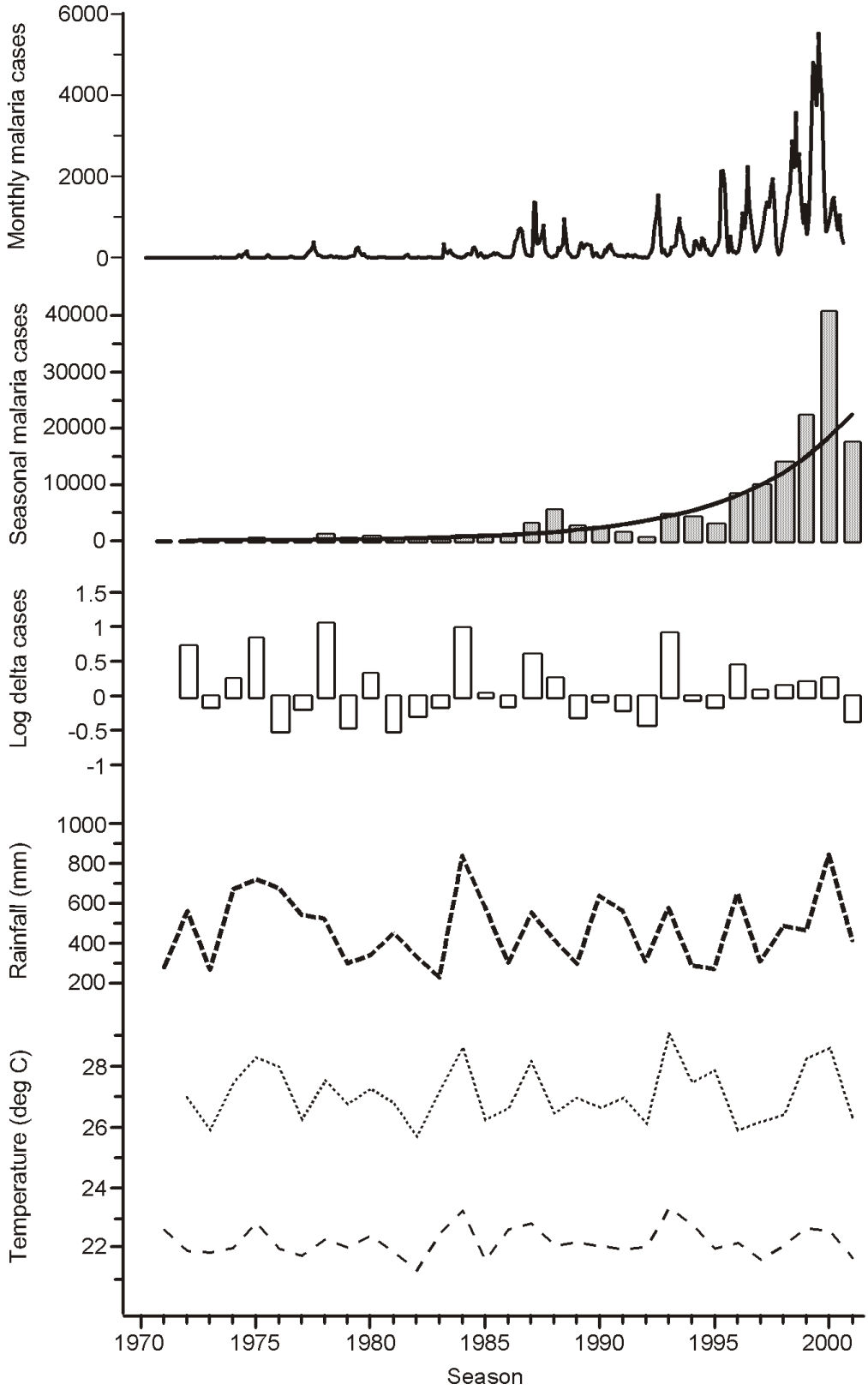


Figure 5.2 Total malaria case numbers recorded in KwaZulu-Natal province, South Africa from July 1971 to June 2001 by month (light solid line) and aggregated by season (July to June) (shaded bars); the exponential curve modelled on the seasonal data (bold solid line) where total cases = 38.0733 * exp(0.2057x) and x = the season (year) minus 1970 (r^2 = 0.828, n = 30, p = <0.0005); the seasonal change in malaria cases (open bars) calculated as log of (total case numbers in current / previous season); total summer rainfall (bold dashed line) in mm; weighted mean daily maximum temperature during the preceding autumn (dotted line) and mean daily average temperature from preceding January to October (light dashed line) in °C.

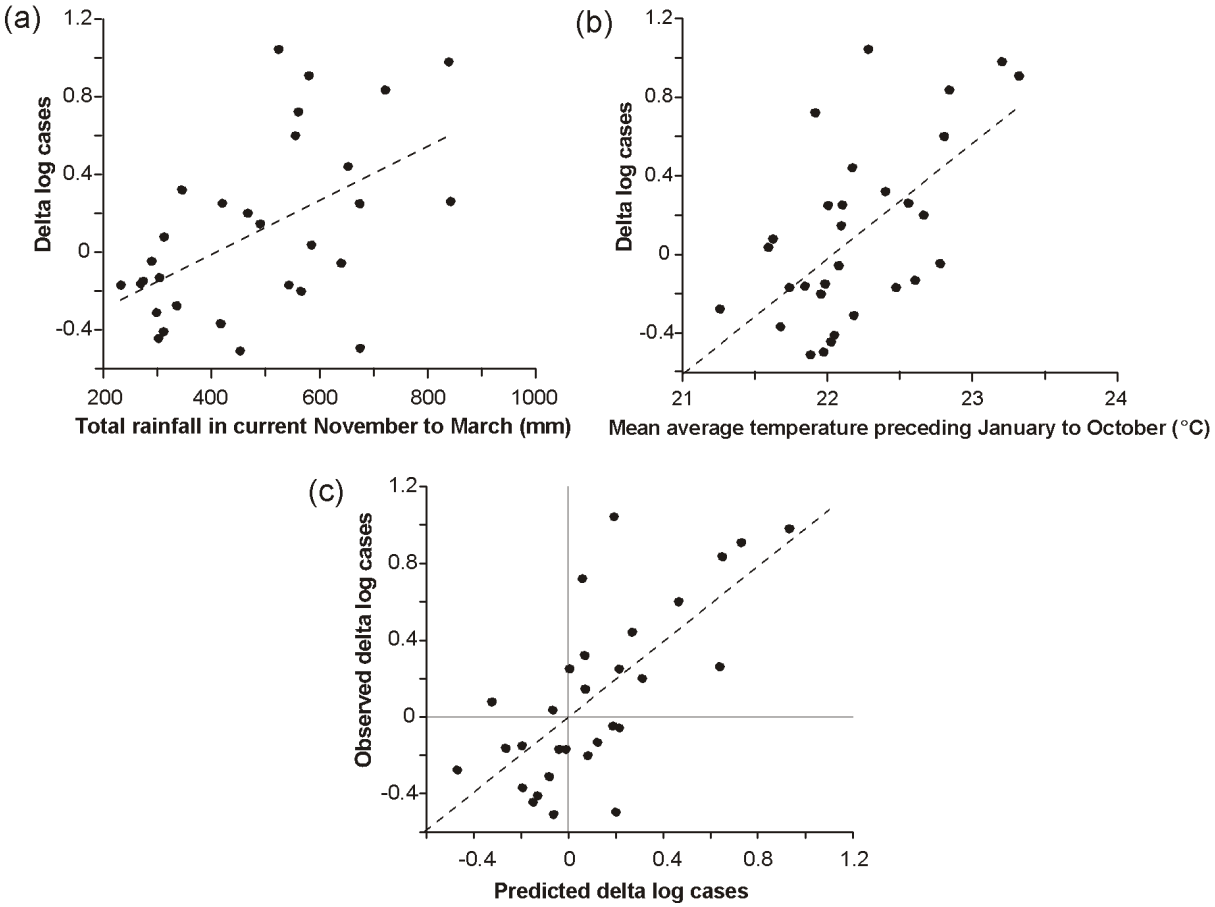


Figure 5.3 Scatter plots (a) and (b) of the two variables used in the final model: (a) total rainfall during current summer (November to March) with regression model (dashed line) where delta log cases = 0.001388 * rainfall – 0.563 (n = 30, $r^2 = 0.282$, p = 0.003); and (b) mean daily average temperature during previous season January to current season October, with the linear regression model where delta log cases = 0.574 * temperature – 12.632 (n = 30, $r^2 = 0.364$, p = 0.00004). (c) Scatter plot of predicted vs observed delta log cases where predicted delta log cases = 0.001 * rainfall indicator (a) + 0.463 * temperature indicator (b) – 10.649 ($r^2 = 0.497$, $p < 0.00001$). The position of y = x (dashed line) and zero change in cases (solid lines) are shown for visual purposes.

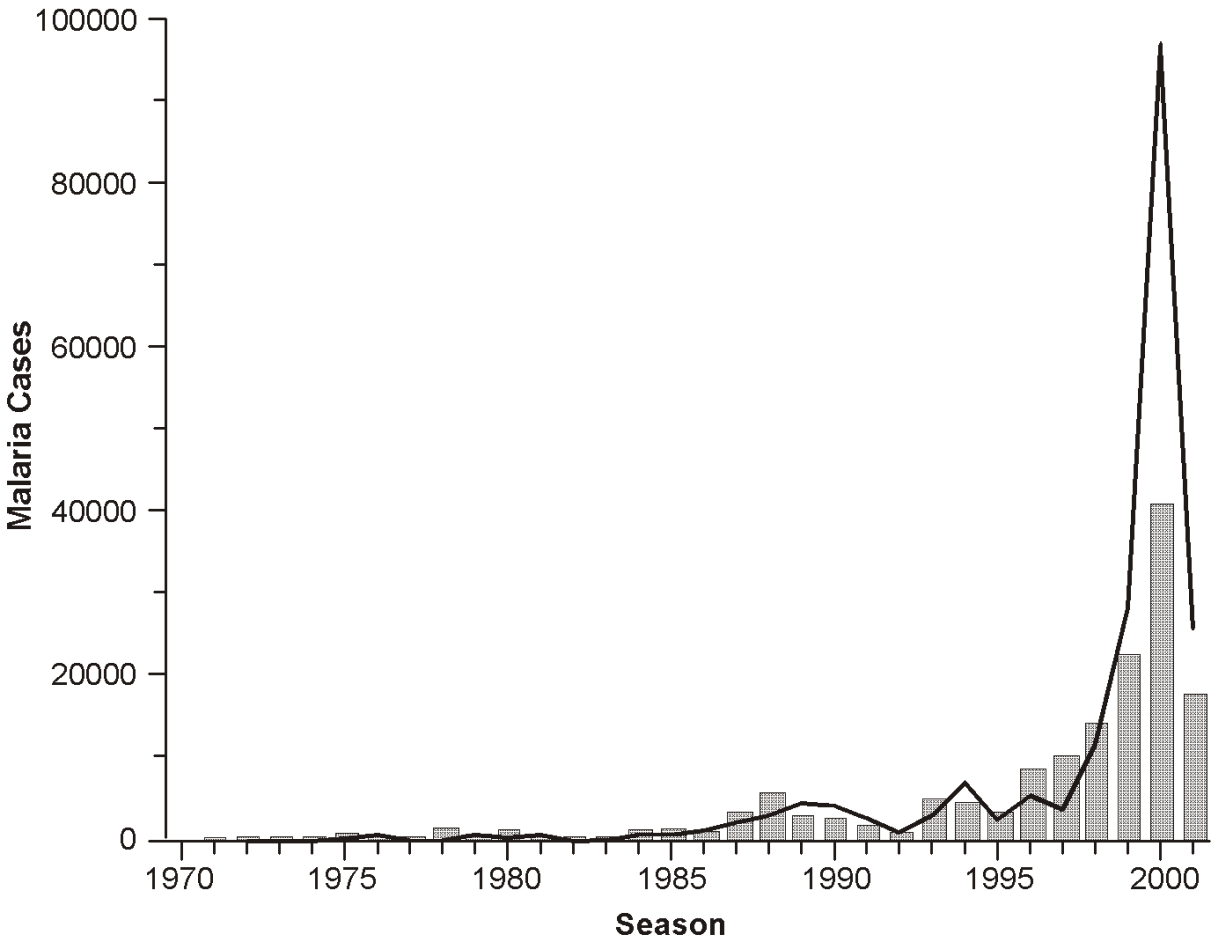


Figure 5.4 Total malaria case numbers recorded in KwaZulu-Natal province, South Africa from July 1971 to June 2001 (bars) aggregated by season (July to June), and the predicted number of cases (line), calculated though multiplying the predicted change in cases, as shown in (c), by the case totals of the previous season. The prediction is an estimate based on climate but not a forecast in time; see the discussion for more detail.

Chapter 6

Exploring thirty years of malaria case data in KwaZulu-Natal, South Africa, Part II: the impact of non-climatic factors

M.H. Craig[1], I. Kleinschmidt[1], D. le Sueur[1], B.L. Sharp[1]

Affiliations:

[1] Malaria Research Programme, Medical Research Council, PO Box 17120, Congella, 4013 Durban, South Africa. Tel: +27-31-2043653, Fax: +27-31-2051498, email: craigm@mrc.ac.za

Keywords:

Anopheles, drug resistance, epidemiological factors, human immune deficiency virus, incidence, insect control, insecticide resistance, irrigation, malaria, South Africa.

Publication status:

Tropical Medicine and International Health ***9***(12): 1258–1266, 2004

Abstract

Malaria transmission is a multi-factorial phenomenon. Climate is a major limiting factor in the spatial and temporal distribution of malaria, but many non-climatic factors may alter or override the effect of climate. Thirty years of monthly malaria incidence data from KwaZulu-Natal province, South Africa, reveal strong medium and long term trends which were not present in the climate data. This paper explores various non-climatic factors that may have contributed towards the observed trends. The development of anti-malarial drug resistance, available information on HIV prevalence, cross-border people movements, agricultural activities, emergence of insecticide resistance and the case reporting system are reviewed and their potential effect on malaria transmission examined. Single-variable linear regression analysis showed significant association between seasonal case totals (log transformed) and the measured level of drug resistance (log transformed) ($r^2 = 0.558$, $n = 10$, $p = 0.013$) as well as relative measures of HIV infection since 1990 ($r^2 = 0.846$, $n = 11$, $p = 0.001$). The other factors appear to have affected the level of malaria transmission at certain periods and to some degree. The importance of surveillance and inclusion of non-climatic variables in analysis of malaria data is illustrated.

Introduction

Since the late 1940's malaria has been strictly controlled in South Africa. Indoor residual spraying of insecticides, periodic larval control and treatment of known infections have restricted malaria to the north-eastern border regions. Infection rates today are well below those observed before systematic control was implemented (Swellengrebel 1931; Le Sueur *et al* 1993). Up until the early 1980's recorded cases in KwaZulu-Natal (KZN) province remained low, but since then case numbers have risen exponentially. The trend was particularly pronounced in the 1990's: cases increased from about 600 in 1991/92 to over 30 000 in the 1999/2000 season.

Several possible causes for this trend have been put forward: climatic determinants, agricultural developments, drug resistance, cross-border movement of people between KZN and Mozambique, insecticide resistance and changes in the malaria case reporting process (Sharp *et al* 1988; Sharp & Le Sueur 1996; Durrheim *et al* 2001). Intrinsic factors, related to host-vector population dynamics, may additionally cause periodicity in transmission over longer periods (Hay *et al* 2000a). The difficulty lies in teasing out and quantifying the contribution of various climatic and non-climatic factors (Rogers *et al* 2002).

In an effort to understand the dynamics of the disease in South Africa, a 30-year database of monthly malaria case data in KZN is being analysed. In the associated paper rainfall and temperature indicators were analysed in connection with the observed trends in malaria incidence (Craig *et al* 2004b), while this paper explores possible links between malaria case numbers and some of the non-climatic factors mentioned above.

Non-climatic determinants

Drug resistance

Resistance of *Plasmodium falciparum* to chloroquine was confirmed in all southern African countries by 1985 (Deacon *et al* 1994). In KZN a small *in vitro* drug resistance study indicated 88% resistance in 1987 (n=17) and 100% in 1988 (n=14) (Freese *et al* 1988). Drug treatment failures were increasing markedly (Hansford 1989) so that in February 1988 the first-line drug was changed to sulphadoxine / pyrimethamine (SP). SP resistance was low on introduction (Hansford 1989), but soon started rising, reaching around 80% in 2000 (Bredenkamp *et al* 2001) the national malaria advisory group recommended an alternative drug policy and co-artemether was introduced as first-line treatment in 2001.

In Mpumalanga Province (MP) chloroquine resistance developed much later. Only isolated cases of treatment failure were reported in 1988 (Hansford 1989) and only started rising in the early 1990's (Kruger *et al* 1996). Two small *in vitro* studies showed 71% and 86% resistance in 1993 and 1996 respectively (Deacon *et al* 1994; Van Nierop *et al* 1996). By 1997 *in vivo* resistance had reached 48.4% (Freese *et al* 2001). These findings again influenced policy and SP was introduced as first-line drug in 1997. A baseline study found SP resistance of 4% in 1998 (Govere *et al* 1999), which then increased to 6% in 1999 (Mabuza *et al* 2001).

In Limpopo Province (LP) *in vivo* chloroquine drug resistance was 40% in early 1997 (Freese *et al* 2001) and SP was introduced in the 1998/99 malaria season.

Evidence of *in vivo* drug resistance in South Africa has been summarised in Figure 6.1. The 'treatment failure' data published by Hansford (1989) and Kruger *et al.* (1996) are effectively non-standardised *in vivo* studies, based on post-treatment follow-up of confirmed blood smear positives, not at fixed times, but instead after 2 to 4 weeks as part of routine active surveillance. Exponential curves modelled on the resistance data are also shown.

In all three malarious provinces drug policy changes were accompanied by reduction in malaria incidence: in KZN malaria, following introduction of SP in February 1988 (Figure 6.2), seasonal case totals were halved by next season. Cases were reduced in 1997 in MP and in 1998 in LP, each time after introduction of SP. Following the introduction of co-artemether in KZN after the highest incidence malaria season in 1999/2000, cases were reduced to less than half. However, drug policy was not the only change that was introduced in 2001 (see below).

HIV / AIDS

HIV / AIDS has spread rapidly in South Africa since the early 1990's. According to annual national surveys of women attending public antenatal clinics, HIV sero-prevalence in KZN increased from 1.6% in 1990 to 36.2% in reports (Anon. 2004c; Anon. 2007d). Even though the data are (possibly biassed) estimates of HIV prevalence in a particular population group, the steep upward trend is clear (Figure 6.3). Sero-prevalence is somewhat lower in MP and much lower in LP.

Early studies failed to detect significant relationships between malaria and HIV / AIDS (Muller & Moser 1990; Greenberg *et al* 1991). Recent studies however have demonstrated clear association between HIV infection and malaria prevalence in pregnant women (Steketee *et al* 1996; Verhoeff *et al* 1999) placental malaria infection rates and newborn infection (Steketee *et al* 1996), resulting in retarded foetal growth and losses. HIV infection was also related to severe malaria, hospitalisation and need for blood transfusion in small children (Kalyesubula *et al* 1997), higher post-neonatal mortality in infants whose mothers were co-infected with malaria and HIV (Bloland *et al* 1995) and higher risk of clinical malaria in adults (Whitworth *et al* 2000). Higher parasite densities were found in HIV positive adults (Whitworth *et al* 2000) but not in children (Kalyesubula *et al* 1997). Malaria infection in turn has been found to increase HIV viral loads (Hoffman *et al* 1999).

Cross-border movement

Malaria in KZN is primarily a border problem, as can be seen from maps of the area (Sharp & Le Sueur 1996; Kleinschmidt & Sharp 2001). The persistence of malaria in the KZN border areas, in the face of intense local malaria control, has been attributed to immigrating malaria carriers. Indeed, people cross the border between South Africa and Mozambique daily for various reasons (Sharp & Le Sueur 1996) and the area running between Ndumu and Tembe

game reserves on the border with Mozambique - the “Mbangweni corridor” - regularly reports the highest malaria incidence in KZN. Across the border opposite this corridor, malaria prevalence in children was around 89% (Sharp *et al* 2003). People entering KZN at this point therefore come from or pass through a hyper-endemic area. In 1996 58% of people coming across the border were infected (J. Mthembu, Head of Malaria Control, KZN). Being mostly non-symptomatic, they do not present at clinics, remain untreated for longer and thereby contribute significantly to local transmission.

Figure 6.4 shows the number of cases reported as originating from Mozambique, or simply as ‘imported’ with no origin specified. Most of these unspecified ‘imported’ cases probably also originated from Mozambique, since cases reported as originating from other countries (including Swaziland) are essentially negligible. Under-reporting of imported malaria is probable, due to the fear of deportation in the non-South Africans.

Figure 6.4 also shows the number of cases that originated in South Africa, and the proportion of imported *vs* total cases. From 1986 to 1992, 20 to 40% of cases were imported. This period coincided with the final years of war in Mozambique. The influx of refugees was considered a major reason for the increase of cases around this time (Sharp *et al* 1988). Since the early nineties however, imported case numbers have remained around 6% of the total. This surely is an important contribution towards the reservoir of infections, but can probably not be blamed for the steep upward trend during the 1990's.

Agricultural developments

In KZN 90% of malaria cases are usually reported from the two northernmost districts, Ingwavuma and Ubombo, which were historically hyper-endemic (Swellengrebel 1931; Anon. 1938). These districts were divided into control sectors over 20 years ago and case data have

been spatially dis-aggregated since then (Sharp & Le Sueur 1996). For most of this period the Mamfene control area, (±60km South of the border) contributed on average only 3% (0 - 6.7%) to the total malaria cases in these two districts. In 1987 and 1988 it increased to 18%.

The Balamhlanga swamp, running along the centre of this area, was a wetland that dried up in winter. However, in 1984/85 water dumping from an irrigation scheme turned it into a permanent swamp, inundating the periphery and providing ideal winter breeding sites for the local vectors. The subsequent increase in malaria prompted intensive larviciding in this area, in addition to routine house spraying, and also repeat spraying of replastered houses, which brought the local epidemic under control. The situation was finally resolved in 1991 with the construction of a retaining dam which prevented further water spillage (Sharp *et al* 1993a).

Insecticide resistance

Dichloro-Diphenyl-Trichloroethane (DDT) was successfully used for indoor residual spraying in South Africa since 1948 (Sharp *et al* 1993b). For several reasons, including human breast milk contamination, community objection and international pressure (Le Sueur *et al* 1993; Sharp *et al* 1993b), DDT was replaced by synthetic pyrethroids.

In 1999 four members of the *Anopheles funestus* group, once eradicated, were found resting in sprayed houses in northern KZN. Between 11 and 50% (14 - 25% in first generation offspring) were resistant to permethrin. Though only few individuals were collected, the sporozoite rate was 5.4%. These mosquitoes were thus involved in local transmission (Hargreaves *et al* 2000). As a result vector control policy changed again and in the winter of 2000 the KZN control programme re-sprayed all houses in this region with DDT.

An. funestus, due to its habit of breeding in permanent water bodies rather than temporary puddles (Gillies & de Meillon 1968), and its strong preference for man, tended to be associated with intense and all-year transmission before large-scale control. Since its eradication in this province, transmission has been much reduced, more seasonal, and by the more generalist vector *An. arabiensis*. One may thus expect the re-emergence of *An. funestus* to be followed by increase in winter malaria. However, as the control programme responded immediately by reverting to DDT, it is not possible now to quantify the impact that the re-emergence of this species had on malaria incidence.

Case reporting

In South Africa active malaria surveys are carried out as part of the malaria control operations, to reduce transmission by treating asymptomatic carriers. Passive reporting covers only patients with clinical symptoms presenting at public sector health facilities. In a population with low immunity, such as the South African population most infections lead to clinical symptoms. Nevertheless, over the past 19 years actively detected cases have exceeded passive cases on average ratio by 2:1. That the population is at least partly immune has been previously noted (Sharp *et al* 1988; Kleinschmidt & Sharp 2001). Spatially dis-aggregated data also reveal that the proportion of passively reported cases was lowest in high-incidence areas (unpublished data).

In a low transmission environment one would expect a surge of clinical malaria during epidemics as populations with little immunity acquire many new infections. This can be observed in the KZN data (Figure 6.5). Indeed, the proportion of passive to total cases is clearly correlated with delta log cases ($R2 = 0.46$, $p = 0.0014$, $n = 19$), i.e. the more cases increase with respect to the previous season, the higher the proportion of passively detected cases. This could suggest that reporting has been reasonably complete or that resources for

active surveillance are exhausted during epidemic periods, and active surveillance therefore suffers.

Analysis and Results

Reported malaria cases were aggregated by annual season, a season for a particular year being defined as the time from 1st July of the previous year to 30th June of the current year. A measure reflecting the changes in case numbers between consecutive years was defined as

$$\text{delta log cases} = \log(y_t) - \log(y_{t-1}), \text{ or } \log(y_t / y_{t-1}),$$

where y_t is the seasonal case total for season t.

Log case totals and delta log cases were analysed against all drug resistance data (chloroquine and SP) and HIV prevalence, using simple linear regression analysis. In years for which more than one drug resistance result was available, an average was calculated, weighted by sample size. SPSS software (Anon. 1999b) was used for the analysis.

Log delta cases were not significantly associated with drug resistance or HIV prevalence but log cases were associated with both drug resistance ($r^2 = 0.558$, $n = 10$, $P = 0.013$) and HIV prevalence ($r^2 = 0.846$, $n = 11$, $P = 0.001$). Unfortunately there were only two years when both HIV and drug resistance data were available, so combined regression analysis of HIV and drug resistance, using raw data, was not possible.

Discussion

The association of case numbers with drug resistance may be explained through the contribution to transmission of persistent infections in treated individuals. In 1984 only 1.2% of hospital-treated cases remained positive on follow-up (Hansford 1989). In 1985 9% were positive on first follow-up, 5% remained positive after a second curative dose of chloroquine.

By 1987 3% still remained positive after having been treated four times. Though these persistent infections are not counted again, they remain in circulation for a number of extra months, increasing the parasite reservoir, thus accelerating transmission.

Several authors have also observed that patients under SP treatment harbour an excess of gametocytes in their blood compared to other treatments (Robert *et al* 2000; Bredenkamp *et al* 2001; von Seidlein *et al* 2001). Where SP is used exclusively such excess gametocyte production must further accelerate transmission, particularly of resistant strains.

The strong association of the trend in malaria cases with HIV prevalence is note-worthy. It is reminiscent of the documented HIV-related tuberculosis (TB) epidemic (Wilkinson & Davies 1997). However, since immunity does not prevent infection of malaria (Molineaux 1988), why have there been more reported cases? Immunity does affect severity of disease (Molineaux 1988), yet the current data are not clinical cases, but actual infections detected through active and passive surveillance. So why would case numbers increase with HIV prevalence?

Loss of immunity reduces recovery rates (Molineaux 1988), so HIV infection and drug resistance may be working together, affecting recovery from infections with resistant strains. In the case of an immune system weakened by HIV even partial resistance in the parasite is bound to result in recrudescence, while a healthy immune system may rid the body of low levels of surviving resistant parasites after treatment. Thus as HIV prevalence increased, the effect of drug resistance may have become worse. With the additional effect of increased gametocyte production following SP treatment, these three effects may have worked together to create the exceptionally high incidence in the late 1990's. That the malaria upward trend

was strongest in KZN, which has the highest HIV prevalence in South Africa, and where SP drug resistance was probably above 50% by 1998, supports this argument to some extent.

Another possible factor is that loss of immunity due to HIV may lead to larger parasite loads, which result in greater infectiousness to feeding vectors, thereby increasing transmission. Increased parasite loads are also more easily picked up by routine microscopy, which may miss low-level infections in semi-immune carriers.

How and to what extent HIV infection impacts on vector-borne diseases certainly deserves further attention. It is reasonable to accept that both HIV infection and drug resistance contributed significantly to overall malaria incidence in KZN. Another long-term data series of malaria admissions available from Kericho (Shanks *et al* 2000), reveals a similar exponential increase in malaria case numbers during the 1990's. The authors offer the spread of drug resistance as explanation for the trend but perhaps this was exacerbated by the spread of HIV at the same time.

The impact of vector resistance to synthetic pyrethroids, and the re-invasion of the highly anthropophilic *An. funestus* definitely contributed towards transmission in 1999 and 2000 (Hargreaves *et al* 2000), but there is insufficient data to analyse the extent of the impact. It remains only an interesting observation that in 2001 case numbers were dramatically reduced immediately after indoor residual spraying was repeated with DDT, even before the new drug had been introduced.

One problem in this data set was the co-incidence of different explanatory variables. In the late 1980's imported malaria and chloroquine drug resistance peaked simultaneously and a local epidemic occurred due to agricultural practice. In the 1990's HIV infection and SP

resistance emerged simultaneously and in addition DDT was replaced with pyrethroids in 1996. Finally, in 2000/01 malaria incidence decreased substantially following re-introduction of DDT for insecticide spraying, introduction of a new effective anti-malaria drug, and implementation of large-scale vector control in Southern Mozambique (Sharp *et al* 2003) as part of the Lubombo Spatial Development Initiative (LSDI). The relative importance of each variable can only be inferred if long time series of malaria case data and explanatory variables are available, and perhaps not even then. So the problem is not a lack of possible explanations, but the abundance of highly plausible ones.

The recent decrease in malaria has been substantial. Cases reported to the Department of Health in KZN were down to about 17 500 in 2001 from over 40 700 in 2000, and have been reduced further to 3500 in the 2002 season. That such reductions were achieved in the face of high HIV prevalence is extremely encouraging. The gametocidal action of artemisinin compared to SP (von Seidlein *et al* 2001) surely helped to reduce transmission to a minimum, and forms an additional intervention for effective control.

Cases also decreased in MP, but only by about 7%. Control operations in this province continued as before. But because MP only partially borders the LSDI area, malaria control in Mozambique did not seem to impact this province. Swaziland, which borders completely on the LSDI area, saw a 65% reduction in malaria incidence, even though malaria control in this country also remained unchanged (Sharp *et al* 2003). In LP, which does not adjoin the LSDI area, cases have continued to rise. Thereby the wider and cross-border impact of large-scale control is well illustrated.

The combined evidence may help unravel to what extent each factor contributed towards reducing malaria incidence. The trend in cases beyond 2002 will give further insight, as drug

resistance in KZN will have returned close to zero, while HIV prevalence will remain at very high levels for some time even if large-scale anti-retroviral treatment becomes policy.

Regular surveillance of important variables is crucial. In Sough Africa research has been driving policies on drug use and vector control in the past, and is now also beginning to impact at a regional level. This is a healthy trend that needs to be maintained and that needs to become mainstream practice elsewhere in Africa.

Conclusion

In this study longer-term trends in malaria incidence have been found to coincide with, and - where such analysis was possible - to correlate with levels of drug resistance, HIV prevalence, trends in indoor residual spraying and insecticide resistance, among others. The effect of climate could be seen in the inter-seasonal variation in malaria cases (Craig *et al* 2004b).

The conclusion that the rate of transmission, where transmission is controlled, should be subject to the effectiveness of these control measures, is not unreasonable or surprising. In the same vein, one may expect stronger correlation between malaria variability and climate variability where, firstly, climate is marginally suitable for malaria transmission, or suitable for brief periods only, and secondly, where the natural state has not been greatly altered by malaria control.

Malaria transmission is a highly complex and dynamic system. Nevertheless it was possible to explain incidence patterns in retrospect, at least in part. Obviously one can only account for factors for which good long-term data are available. This underlines the importance of long-term surveillance, including surveillance of the coverage and effectiveness of control interventions.

Acknowledgements

Sincere thanks go to Amanda Jackson for assistance in the malaria data aggregation, to Dave Durrheim and Simon Hay for valuable comments on the manuscript and to the South African Medical Research Council for financial support.

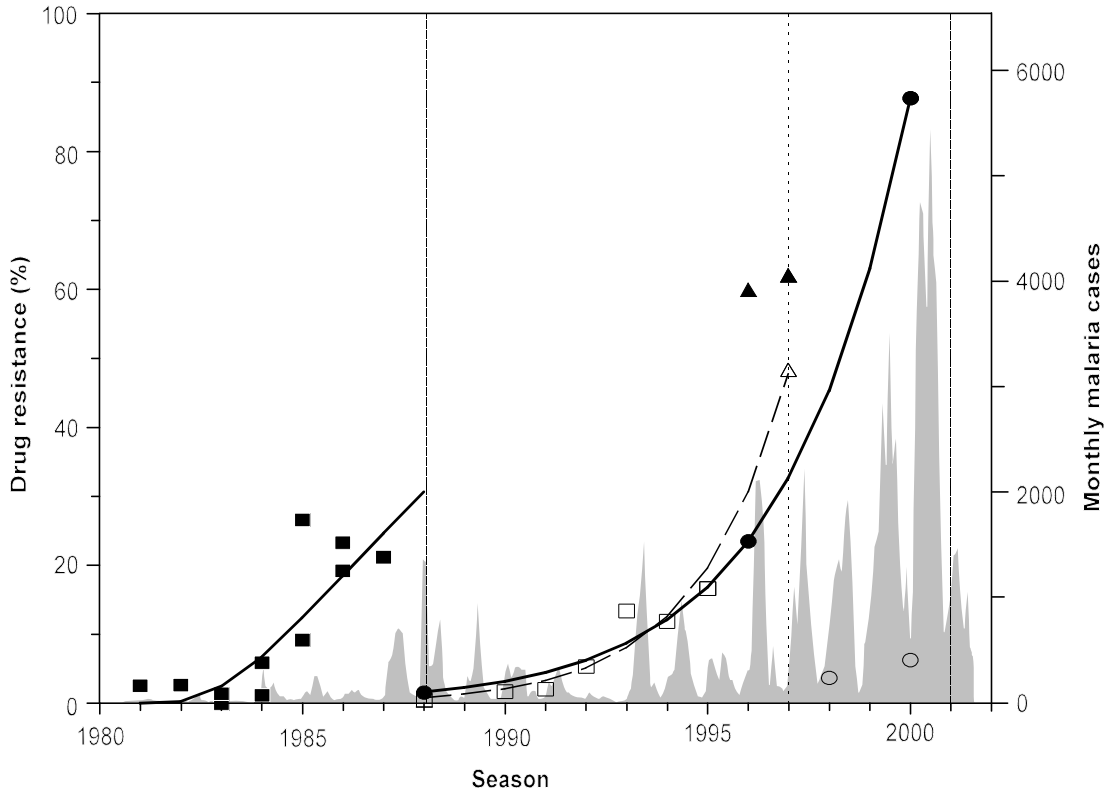


Figure 6.1 KwaZulu-Natal province: chloroquine resistance treatment failure (solid square) (Herbst *et al* 1987; Hansford 1989) and chloroquine *in vivo* resistance (solid triangle) (Freese *et al* 2000) with the modelled curve (light solid line); *in vivo* sulphadoxine-phyrimethamine (SP) resistance (solid circle) (Hansford 1989; Freese *et al* 2000; Bredenkamp *et al* 2001) with its modelled curve (heavy solid line). Mpumalanga province: chloroquine resistance treatment failure (open square) (Hansford 1989; Kruger *et al* 1996) and chloroquine *in vivo* resistance (open triangle) (Freese *et al* 2001) with the modelled curve (broken line); *in vivo* SP resistance (open circle) (Deacon *et al* 1994; Govere *et al* 1999; Mabuza *et al* 2001). The vertical dashed lines indicate drug policy changes in KZN from chloroquine to SP (1988) and from SP to co-artemether (2001), the dotted line indicates change from chloroquine to SP in MP (1997).The shaded area graph shows monthly malaria cases in KZN. The modelled curves are as follows: for chloroquine resistance in KZN $y = 1.7864 - 1.7173x + 0.7386x^2$ where x is the year minus 1980 ($n = 11$, $r^2 = 0.675$, $p = 0.004$); for chloroquine resistance in MP $y = 0.0106 * \exp(0.5041x)$ where x is year - 1980 ($n = 8$, $r^2 = 0.965$, $p < 0.0005$); for SP resistance in KZN $y = 1.163 * \exp(0.33x)$ where x is year - 1987 ($n = 3$, $r^2 = 1$, $p = 0.002$).

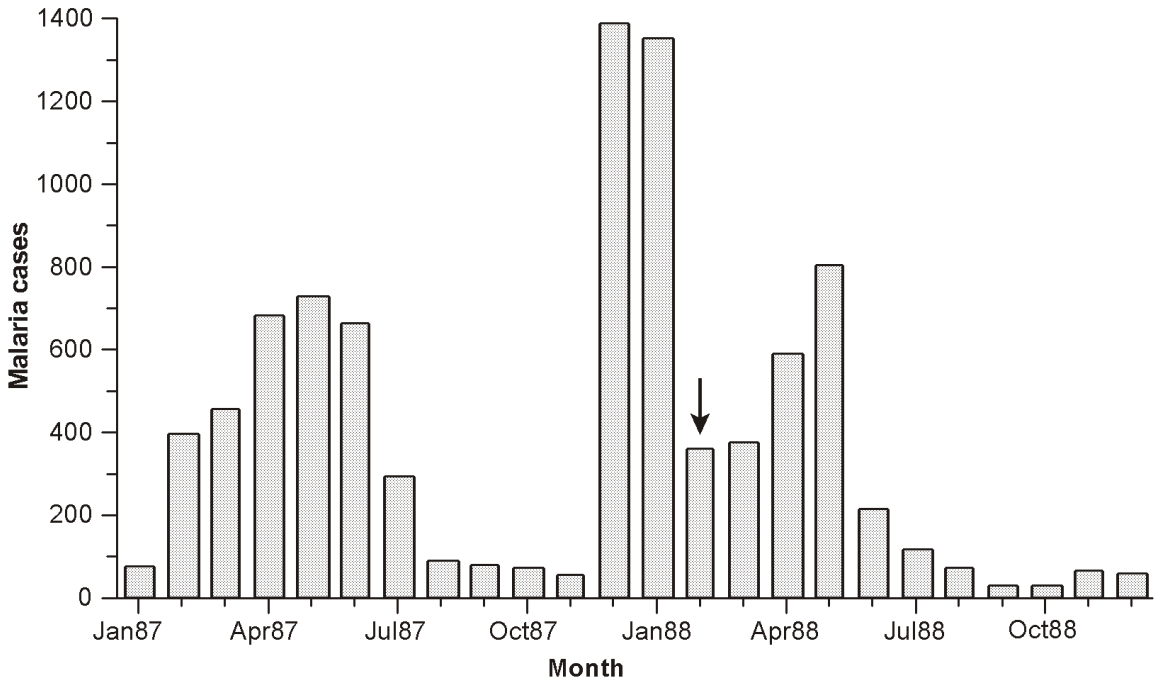


Figure 6.2 Malaria case numbers in KZN reported by month during 1987 and 1988 (bars). The arrow indicates introduction of sulphadoxine / pyrimethamine and the associated reduction in cases.

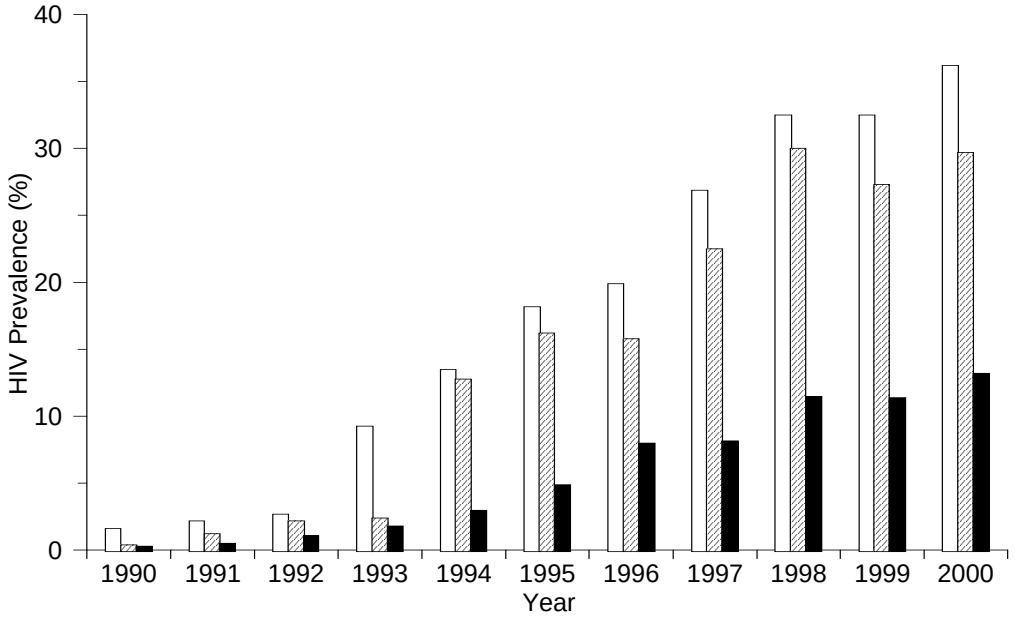


Figure 6.3 HIV sero-prevalence in women attending public antenatal clinics in KwaZulu-Natal (open bars), Mpumalanga (shaded bars) and Northern Province (solid bars), South Africa.

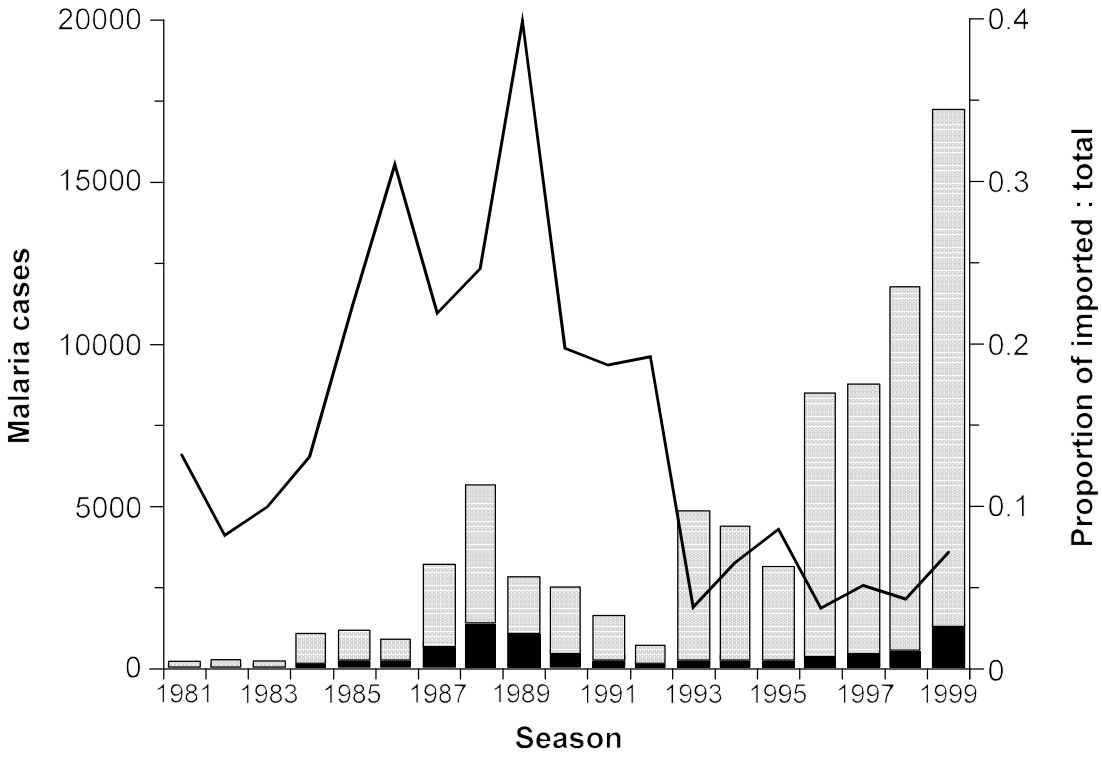


Figure 6.4 Number of malaria cases reported in KwaZulu-Natal as of Mozambique origin or 'imported' with unspecified origin (solid bar) and of 'local' or 'inconclusive' origin (shaded bar), and the ratio of all imported to total number of cases reported (line).

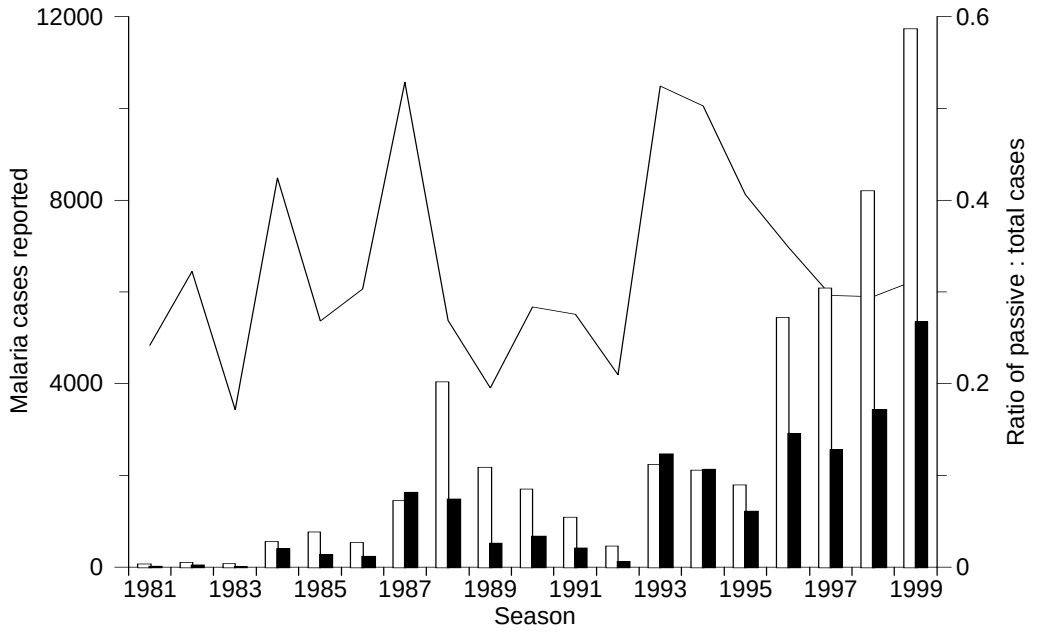


Figure 6.5 Total number of malaria cases reported in KwaZulu-Natal by season (open bar); number of cases reported from passive surveillance, i.e. patients reporting to clinics (closed bar) and the ratio of passive to total cases (line).

Chapter 7

Spatial and temporal variation in malaria incidence in South Africa

M.H. Craig[1,2], I. Kleinschmidt[1,3], M.L.H. Mabaso[1,2], P. Vounatsou[2], T. Smith[2]

Affiliations:

[1] Malaria Research Programme, Medical Research Council, PO Box 17120, Congella, 4013 Durban, South Africa. Tel: +27-31-2043653, Fax: +27-31-2051498, email: craigm@mrc.ac.za

[2] Swiss Tropical Institute, 57 Socinstrasse, Basel, BS 4002, Switzerland

[3] London School of Hygiene and Tropical Medicine, Keppel Street, London WC1E 7HT, United Kingdom

Keywords:

climate, disease outbreaks, drug resistance, environment, epidemiological factors, incidence, insecticide resistance, malaria, risk factors, South Africa, transmission.

Publication status:

Not yet submitted.

Abstract

Introduction

Inter-annual variations in malaria incidence reflect heterogeneities in both climatic and non-climatic, particularly control-related factors. South Africa is situated on the southern fringe of the distribution of malaria in Africa, where the relative importance of variations in climate, and of non-climatic factors (such as anti-malarial drug resistance) in determining spatial and inter-annual patterns in incidence have been unclear.

Methods

Monthly malaria case records, aggregated into six sub-provincial regions and by malaria season (July to June), and calibrated by total population, were analysed against four potential risk factors using negative binomial autoregressive models allowing for spatial and inter-annual correlations. An initial model (A) was fitted to data from 1981/82 to 1998/99 and validated against data for 1999/2000 to 2004/05. A second model (B) was fitted to all data.

Results

Model A did not predict well when forecasting. Predictions during the validation period were both inaccurate and uncertain. In model B summer rainfall, annual mean temperature (which had been identified as effective climatic predictors in a related study focussing on KwaZulu-Natal), and drug resistance were significant predictors of malaria incidence. Major changes in drug resistance levels occurred late in the study period, and drug resistance was non-significant in model A.

Conclusion

Temporal patterns in malaria incidence reflect variations in the effectiveness of control, which vary as a result of the emergence of drug and insecticide resistance and policy changes.

Therefore simple climate data driven malaria forecasting systems are not very useful unless they incorporate factors that are proxy for the effectiveness of malaria control.

Introduction

South Africa is situated on the fringe of the distribution of malaria in Africa, which is concentrated in the tropics (Craig *et al* 1999). Fringe areas tend to be epidemic and case incidence in South Africa shows substantial inter-annual variation. It has been suggested that inter-annual variation in malaria incidence may be linked to climatic variation on an inter-annual time scale, and possibly to the El Niņo Southern Oscillation (ENSO) (Mabaso *et al* 2006a) and climate change (Bouma *et al* 1994). This has further led to suggestions that inter-annual climate data, satellite data, and the ENSO index, could be used in epidemic-prone areas to give early warning, potentially helping health managers to plan interventions (Hay *et al* 1998; Thomson *et al* 2000; Thomson & Connor 2001b; Mabaso *et al* 2006a).

The current malaria risk areas in South Africa represent only a fraction of areas where climate is suitable for malaria, mainly as a result of intensive long-term malaria control using indoor residual spraying (Le Sueur *et al* 1993; Mabaso *et al* 2004). Currently malaria is limited to the North-East border regions, in the provinces of KwaZulu-Natal, Mpumalanga and Limpopo. Variations in malaria incidence within the remaining endemic areas are thought to reflect not only heterogeneities in climate but also effects of migration, drug and insecticide resistance, and agricultural practices (Sharp *et al* 1988; Sharp & Le Sueur 1996; Durrheim *et al* 2001).

We recently examined how these factors interact in determining patterns of malaria in KwaZulu-Natal (Craig *et al* 2004a; Craig *et al* 2004b), but so far no analysis has considered the contributions of these different factors to the risk of malaria in the whole endemic part of South Africa (that is, areas with autochthonous malaria). We have now assembled a

comprehensive data set of case incidence data for the years 1981 to 2005, for the three endemic provinces. We examined both spatial and temporal patterns in annual incidence and the extent to which these can be explained by climatic and non-climatic factors.

Methods

Study area

South Africa is a country of widely varied topography and environment. It can be bisected into a dry western half, with <500mm rain per year, and a moist eastern half, with annual rainfall of 500 to >1000mm (Schulze 1997). Most of the country forms a plateau that lies above 1000m altitude (Anon. 1998b), which is too cold for malaria at these latitudes (22 to 35° South). The plateau peaks in the centre-East along the Drakensberg mountain range, dropping off steeply towards the south-eastern coast-line, more gradually so towards the North and West. The southern and western coastal areas are too cold and dry to allow malaria transmission, so that malaria is limited to the warm, moist, partly sub-tropical and low-altitude "low-veld" and the mid-altitude "middle-veld", along the eastern seaboard and the north-eastern borders with Mozambique and Zimbabwe.

The country is divided into nine provinces, three of which are malaria-endemic. Each province is sub-divided into districts, which are the administrative units for which population and malaria control data are available. The provinces have recently been organized into political units termed municipalities with boundaries that are sometimes misaligned with those of the districts (Figure 7.1). For the purpose of this analysis, the district-based data were aggregated into six regions whose borders correspond approximately to what are known as "Type C" municipalities.

Malaria data

In South Africa malaria has been notifiable since 1956 (Sharp *et al* 1988) though early data have been lost. The health system aims to detect every infection rather than clinical cases only, through definitive diagnosis, via blood smear or rapid diagnostic test, during passive and active surveillance (Sharp *et al* 1988; Kleinschmidt *et al* 2002). There is a relatively good spread of health facilities, including hospitals, clinics and mobile or malaria clinics. Based on the 1996 census enumerator areas, and current health facility data, the proportion of the population in the analysis regions who live within 5km of a health facility respectively range from 52% in region A (arguably the most rural area) to 90% in region C. The proportion of people who live within 10km of a health facility was 93 to 99% in regions A to E and 78% in region F.

The provincial malaria control programmes employ a number of epidemiological surveillance teams consisting of a field officer and several field assistants, also known as malaria agents. These teams follow up malaria cases diagnosed by health facilities, private practitioners or malaria agents. The purpose of the follow-up is to identify the probable source of infection, to find non-symptomatic malaria carriers by taking blood smears from all people staying with the patient, neighbours and immigrants, to identify possible vector breeding places for larval control, to identify possible outbreaks and clustering of cases via special or mass surveillance, to conduct health education in the community and also to assist in the mapping of malaria cases as well as data collecting during special studies such as *in vivo* studies for monitoring drug efficacy.

Confirmed malaria cases, presenting at health facilities or detected by active surveillance, are recorded on standard report forms which are collected at one of the four provincial malaria control centres in Nelspruit (Mpumalanga), Tzaneen (Limpopo), Jozini and Richards Bay

(KwaZulu-Natal). Here they are entered into a Microsoft Access-based national malaria information system developed for data entry and partial validation. Electronic data are then sent to the Medical Research Council for further cleaning and collation, and to the national malaria control office at the Department of Health in Pretoria. Until the mid-1990's data were entered from the report forms into a variety of computer programmes. These have since been standardised and added into the malaria information system.

Malaria case records are available by district and by week (using the date of the blood test) for KwaZulu-Natal from 1992 onwards (data by month are available from 1981). Weekly malaria case data for Mpumalanga are available from 1987, and for Limpopo from July 1986. In all but 0.5% of cases the age has been reported.

In KwaZulu-Natal and Mpumalanga most cases were reported in a small number of districts, while in Limpopo cases were distributed over a wider area, though incidence was generally lower. Since 1992, 95% of all cases were reported in one of the analysis regions shown in Figure 7.1. Of all reported cases, 72% also reported the presumptive location of the source of infection. In 79% of these the presumed source district was the same as the reporting district. In the present analysis the cases are assigned to the district in which they were reported, rather than the district in which they are presumed to have originated.

Population data

The mean population throughout the study period was estimated from the data of two censuses of 10 October 1996 (Anon. 1999a) and 10 October 2001 (Anon. 2003a) for each analysis region, by assuming a constant growth rate and projecting both backwards and forwards in time. The populations assigned to each July-June period were calculated as the geometric mean of the estimated weekly populations.

Drug resistance data

Available drug resistance data for South Africa were reviewed previously (Craig *et al* 2004a). Drug policy is decided at provincial level and the study period included the first-line drug policy change from chloroquine (CQ) to sulphadoxine-syramethamine (SP), and then to artemisinin combination therapy (ACT) (Table 7.1).

Climate data

Daily rainfall, and minimum and maximum temperature data were obtained from the South African Weather Services (Anon. 2007c). Though a fairly dense network of weather stations exists, only few currently operational stations have generated both temperature and rainfall data for the entire period covering the malaria data. After careful investigation, the best available, currently operational station in each malaria region was selected to represent the region (Figure 7.2). Data gaps were filled using data from the closest available stations (filler stations). Table 7.2 shows the main, currently operating stations representing each region, the filler stations, their distances from the main stations, and the dates of the data gaps.

Rainfall, being a sporadic event, is not well correlated on a daily basis even between points in close proximity, so missing days were substituted with filler station data. This was done because some of the data gaps occurred in the main summer rainfall period, and because the rainfall gauge network is quite dense, a suitable station could usually be found nearby for the required time gap. Daily temperatures on the other hand are strongly correlated even over long distances. Consequently the temperature-recording stations are more widely spaced. Filler stations were only available between 50 and 100km distant, and auto-regressive integrated moving average (ARIMA) regression models in STATA (Anon. 2001b) were used to estimate missing data.

Most of the data gaps were minor, involving days and weeks, or a month here and there. In the rainfall data a one year gap in Messina and a ten month gap in Phalaborwa had to be filled from stations 18 and 30km away respectively. The rainfall series for region B comes from three successively operating stations around Nelspruit, a few kilometres apart. The most serious case of missing data concerned the temperature data in Nelspruit, which had to be inferred from three other stations for the entire first half of the study period, via ARIMA regression based on the second half of the period. A one year, and an 18 month gap in Messina and Phalaborwa respectively were the only other major temperature data gaps (Table 7.2).

Regression analysis

Given the over-dispersed nature of the case data, negative binomial regression was used to analyse risk factors for malaria incidence, allowing for the overall patterns in space and inter-annual variation. The recorded malaria cases were aggregated into 12 month periods starting in July and ending June, to avoid splitting the transmission season. The corresponding total population of the region was used as a rate multiplier. The logarithm of the incidence rate during the season (July to June) was then modelled as the sum of terms in each of the risk factors, comprising mean temperature, summer rainfall and predicted drug resistance prevalence (Table 7.3). These factors had emerged as important predictors of temporal variation of malaria incidence in KwaZulu-Natal (Craig *et al* 2004a; Craig *et al* 2004b).

To estimate drug resistance an exponential growth curve was assumed in the resistance rate during the period of use of each drug. Separate curves were fitted to the data for each province where sufficient data were available, using non-linear least-squares estimation (nl command) in STATA (Anon. 2001b). The drug resistance was predicted for time *t* as follows:

$$\text{predicted drug resistance} = \beta_0(\beta_1^{t-t0})$$

where β_0 and β_1 are constants specific for each drug/area curve, and t_0 is the introduction date of the drug.

Because of sparse data, CQ resistance was assumed to follow the same dynamics in Mpumalanga and Limpopo (the same β_0 and β_1 but different introduction dates were used). The SP data were combined for all three provinces. Drug resistance against ACT has been assumed to be very low; a constant value of 0.5% was used.

A hierarchical Bayesian model was fitted using a Markov chain Monte Carlo (MCMC) algorithm in Winbugs (Lunn *et al* 2000). Cases Y in region i and year t were assumed to follow a negative binomial distribution $Y_{it} \sim \text{NB}(p_{it}, r)$ with rate p_{it} and the over-dispersion factor r. The rate p is related to the average incidence rate μ of the negative binomial distribution and the dispersion parameter r as follows:

$$p_{it} = r / (r + \mu_{it})$$

The following regression model was used to estimate the mean μ_{it} for each region i and year t:

$$\log(\mu_{it}) = \log(\text{population size}_{it}) + \beta_0 + \beta_1(\text{summer rainfall})_{it} + \beta_2(\text{annual mean temperature})_{it} + \beta_3(\text{drug resistance prevalence})_{it} + \varphi_i + \omega_t$$

where $\beta_0..\beta_4$ are the regression coefficients of the co-variates, φ_i the spatial random effect for region i, and ω_t the temporal random effect for year t. Normal distributions were assumed for coefficients β.

The random spatial effects φ were modelled by a conditional autoregressive model (CAR) by assuming that φ_i's for region i are independent (conditional on the neighbours) and normally

distributed, with a mean equal to the mean spatial effects $\overline{\varphi_i}$ of all neighbours of *i*, and variance σ_φ^2 inversely proportional to the number of neighbours n_i:

$$\varphi_i \mid \varphi_{-i} \sim \text{Normal}\ (\overline{\varphi_i}\ , \sigma_\varphi^2 / n_i)$$

The random temporal effects ω were assumed to be independent for each year (conditional on the previous year) and normally distributed, with a mean $\overline{\omega_t}$ equal to the mean $\overline{\omega}_{t-1}$ of the previous year *t*-1 and the temporal correlation ρ factor:

$$\omega_t \sim \text{Normal}\ (\rho\omega_{t-1}, \sigma^2_\omega),\ t=2,\ldots T \text{ and } \omega_1 \sim N(0, \frac{\sigma^2_{\omega_0}}{1-\rho^2}),\ |\rho| < 1.$$

The variance of the first year was based on the variance estimated for the rest of the time series. A Uniform U(-1,1) prior distribution was considered for ρ and vague inverse gamma distributions were assumed for variances $\sigma^2_\varphi, \sigma^2_\omega, \sigma^2_{\omega_0}$ and for the dispersion factor *r*.

The model was initially fitted only on the data for the period of 1981/82 to 1998/99 and validated by comparing the predictions for the seasons 1999/2000 until 2004/05 with the observations for these years (Model A). Subsequently the model was also fitted on all data (Model B). Both models were run for 800 000 iterations. Convergence was confirmed using the Geweke and Heidelberger/Welch convergence tests in BOA (Smith 2005). The agreement between the predicted median log incidence from each model and the observed data was assessed using the concordance correlation coefficient (ρ_C) (Lin 1989; Lin 2000).

Results

The observed and unsmoothed malaria incidence rate over a 13 year period (Figure 7.3) was highest around the North-eastern border of South Africa, the areas adjoining southern tip of

Mozambique reporting the highest incidence. The border with Swaziland experiences very low incidence, and intermediate incidence is recorded along the borders with Zimbabwe and Botswana. The area of the Kruger National Park has a very low resident population.

Region A, roughly the "Type C" municipality of Umkhanyakude, contains three districts that report 90% of all KwaZulu-Natal cases (65% from Ingwavuma, 20% from Ubombo 5%, from Hlabisa, listed from North to South). Region B, roughly Ehlanzeni, reports 95% of all Mpumalanga cases (62% cases in Nkomazi, 19% in Barberton, 8% in Nelspruit, 6% in Witrivier, and 3% in Pelgrimsrust, listed East to West). Analysis region C corresponds to two municipalities and reports 42% of cases in Limpopo. The Bohlabela and Mopani municipalities had to be combined because the border bisects several districts. Region D, approximately Vhembe, reports 54% of Limpopo cases. Regions E (roughly Capricorn) and F (Waterberg) are low-incidence regions, reporting only < 1% and 3.8% of Limpopo cases respectively.

Drug resistance increased in each province, first to CQ, then to SP, but was finally reduced to near-zero after change to ACT (Figure 7.6). While Mpumalanga and Limpopo were still using CQ, SP resistance in KwaZulu-Natal had already increased around 40%. Mpumalanga and Limpopo switched from SP mono-therapy to ACT before SP resistance had reached 10%. Because of this Mpumalanga decided to use SP as the partner drug when initially changing to ACT. Details of the predicted curves are shown in the caption of Figure 7.6.

Figure 7.4 clearly shows that the observed annual malaria case incidence (calculating from July to June) is much higher in KwaZulu-Natal than in the other two provinces. Though Mpumalanga and Limpopo report comparable absolute numbers of cases (96 102 and 74 730 respectively), the four regions in Limpopo are much more populous, resulting in very low

overall observed incidence rates. Incidence in the five analysis regions is displayed on different scales in Figure 7.5, together with the summer rainfall for each season. Some peaks in observed incidence coincided with peaks in rainfall, but this pattern was not consistent.

Table 7.3 shows the results of the spatio-temporal model, first fitted on 18 years of data, then fitted on 24 years, over six analysis regions. Observed and predicted log incidence is shown for each analysis region in Figure 7.7, using the predictions from Model A, fitted on 18 years of data and forecast for six. Though the fit of the model, when calculated over all regions, was high ($\rho_C = 0.86$, Table 7.4), the confidence intervals of the predictions diverged substantially in the validation period (Figure 7.7). The accuracy of the predictions differed between regions (Table 7.4). In region B incidence could be predicted most accurately from the available co-variates, while in region E there was no significant correlation between the observations and predictions. Drug resistance was only a significant predictor in the model fitted on all data.

Discussion

Over the analysis period a dramatic upward trend in malaria incidence was observed in South Africa (Figure 7.4). This trend was most pronounced in KwaZulu-Natal and least pronounced in Limpopo. Climatic variables together with trends in drug resistance accounted for much of the spatial and temporal variation across all analysis regions (Table 7.4). Rainfall provides important breeding sites for the dominant local mosquito vectors (*Anopheles arabiensis*), particularly in summer when warm temperatures allow rapid larval development. Years with particularly high rainfall sometimes, but not always, coincide with a peak in malaria incidence (Figure 7.5). Temperature determines extrinsic parasite development not just in summer, but also after the main mosquito breeding season; warmer winters presumably boost the reservoir of parasites that start the new transmission season in early spring (Craig *et al* 2004b).

Much of the region-specific temporal trends however remained unexplained. The predictions from Model A (Figure 7.7) appear reasonable on the log scale, but when converted to actual incidence, the observed values were over- or under-predicted by up to two orders of magnitude. Since the variance in each individual region is much lower than the overall variance, the good overall model fit hides important region-specific discrepancies. Since malaria control is implemented at the provincial level, the lack of accuracy in the regional predictions limits the usefulness of the results for decision-making.

Furthermore, the model, fitted on the beginning of the series, certainly did not perform very well predicting ahead in time. Predictions during the validation period were both inaccurate and uncertain. A model fitted on all data not surprisingly predicted the data post 2000 much better, but then of course the forecasting benefit is lost.

The weak fit in the validation period (2000 to 2005) can be explained partially by the tremendous changes that were taking place in malaria control at this time. The epidemic of 1999/2000 in KwaZulu-Natal was probably caused by a combination of a failed drug (Figure 7.6), a failed insecticide and re-invasion of the highly anthropophilic, previously eliminated, and now insecticide resistant *Anopheles funestus* (Hargreaves *et al* 2000). The epidemic was followed by substantial reductions in incidence, a combined result of drug policy changes (Barnes *et al* 2005), return to DDT for IRS (Maharaj *et al* 2005) and the launch of the malaria control programme of the Lubombo Spatial Development Initiative (LSDI) in neighbouring Mozambique (Sharp *et al* 2007). Some of these factors could not be included numerically in the analysis.

The drug-resistance values used in the regression model were estimated from another model, based on very few observations (Table 7.1). The uncertainty in these estimates was not

allowed for in the regression model. Not allowing for the error in the independent variables generally leads to an over-estimate in the standard errors of the model coefficients; had the errors been allowed for, the confidence intervals might have been smaller, and the predictions even more divergent.

Though the suspected source of infection (based on the investigations by the malaria agents) was not recorded in a third of cases, it seems likely that migration has had a significant impact on malaria in all the three provinces, particularly Mpumalanga (Table 7.5). Though migration has varied over time, a major determinant of the geographical pattern seems to have been the ongoing introduction of infections from Mozambique, which presumably replenishes the local parasite reservoir despite intensive control in South Africa. Not only are the neighbouring areas of southern Mozambique highly endemic, but until recently there was no effective control in this country. Conversely, intensive malaria control in Zimbabwe, Botswana and Swaziland, coupled with lower general malaria risk, presumably contributed to lower reported incidence along the other borders. The number of cases from other endemic countries has been very low throughout the study period.

Drug resistance appears not to have been a major determinant of either the spatial or temporal patterns in incidence rates (Table 7.3). Major changes in resistance levels occurred late in the study period, when other factors were also changing rapidly. CQ resistance emerged about eight years later in Mpumalanga than in KwaZulu-Natal (Figure 7.6), only really starting to spread after the drug had already failed and been replaced by SP in KwaZulu-Natal. Nevertheless, the spread of CQ resistance, once started, appears to have continued at similar rates in both provinces. This raises the question of whether CQ resistant strains were present in Mpumalanga early on, but remained at low frequencies for longer, or whether resistant strains only reached Mpumalanga eight years later. The latter explanation is not entirely

unlikely given negligible direct influx of infected people from KwaZulu-Natal. Diffusion of resistant strains from KwaZulu-Natal across the international borders also seems unlikely due to low levels of transmission in Swaziland and low drug pressure in Mozambique.

Baseline surveys of SP resistance were carried out when the drug was introduced in KwaZulu-Natal and a few months after introduction in Mpumalanga. In Mpumalanga the baseline resistance to SP was twice that of KwaZulu-Natal, though low, suggesting that resistant genes were already present and were selected for immediately when SP was introduced. By 2001 all three provinces were using effective drugs, and reductions in malaria incidence were seen in each province (Figure 7.4), most notably in KwaZulu-Natal. By this time KwaZulu-Natal in particular was also benefiting from both direct and cross-border effects of the LSDI malaria control programme (Sharp *et al* 2007), which has extended IRS to neighbouring areas of Mozambique and intensified detection and treatment of infections.

Mpumalanga (region B) saw an excess of observed cases beyond what could be explained by the co-variates (a positive spatial random effect). This is likely due to the frequency of imported malaria in this province, where half or more of the infections were probably not contracted locally (Table 7.5). Limpopo on the other hand, reported lower incidence in all regions than predicted based on the co-variates. Regions E and F included large and fairly densely populated areas where malaria risk is extremely low (Figures 7.3 and 7.4), so that the prediction, which assumes uniform conditions, over-estimated the overall incidence. In regions C and D the weather stations that provided the climate data were located in the highest malaria risk areas of these regions respectively: respectively Phalaborwa in the hot and humid ‘low-veld', and Messina on the Limpopo river, thus the deficit of cases actually observed in regions C and D.

Strongly positive temporal random effects were estimated in the seasons of 1988/89 and 2000/01 (and not in the remaining years). Both these years followed after a change in drug policy in Region A (Table 7.1), where incidence had reached epidemic proportions with the rise of drug resistance (Figures 7.4 and 7.6). This excess of cases suggests that it may take a year after introducing an effective drug, until incidence returns to ‘normal', or what can be explained by other risk factors.

Conclusion

Malaria in South Africa is mainly found along the northern eastern borders, with the highest incidence reported along the border with Mozambique where there has been no control until recently. Weather station data was able to explain much of the spatial and temporal variation in the incidence data. Region-specific temporal patterns on the other hand largely appear to reflect variations in the effectiveness of control, which in turn reflect emergence of drug and insecticide resistance and policy changes. Predicted incidence, on a regional or provincial level, diverged quite substantially from the reported incidence, perhaps more so than would be useful for planning purposes. In this context - a situation highly modified by malaria control - it seems clear that malaria forecasting systems driven by climate data are not very useful. Even when they incorporate factors that are proxy for the effectiveness of malaria control, predictions were not very accurate. Since some important non-climatic determinants of malaria transmission are difficult if not impossible to quantify, accurate predictions of incidence remain elusive.

Acknowledgements

Sincere thanks go to Colleen Fraser for assistance in the malaria data aggregation, and to the South African Medical Research Council for financial support. We also wish to thank the South African malaria control programmes for their tireless efforts to protect our people against this disease, and for the meticulous ongoing surveillance without which this study would not have been possible.

Table 7.1 Introduction dates and number of available resistance surveys for chloroquine (CQ), sulphadoxine-pyremthamine (SP) and artemisinine combination therapy (ACT) in the three malarious provinces of South Africa.

		KwaZulu-Natal	Mpumalanga	Limpopo
CQ	Introduced in:	1980 †	1980 †	1980 †
	Number of surveys	11	7	1
SP	Introduced in:	February 1988	October 1997	June 1999 ‡
	Number of surveys	3	2	1
ACT §	Introduced in:	January 2001	January 2003	October 2004

† Though CQ was introduced in the 1940's, 1980 was used as introduction date for the curve fitting.

‡ SP was introduced gradually between January 1998 and December 2000; June 1999 represents the midpoint.

§ Mpumalanga first changed from SP to an artemether / SP combination, then to artemether / lumefantrine (Coartem) in 2005. KwaZulu-Natal and Limpopo changed from SP directly to Coartem.

Table 7.2 Weather stations selected to represent the six analysis regions (A to F), in the three malarious provinces of South Africa: KwaZulu-Natal (KZN), Mpumalanga (MP) and Limpopo (LP). The South African Weather Services station numbers in square brackets, distance from the main station is shown in km, followed by the dates of data. The currently operating stations are underlined.

Region	Main station (rain & temperature)	Rainfall filler stations	Temperature filler stations
A, KZN	Makatini [0411323 2]	Pongola [0410410 1] 40km 05/1994, 01/2004, 12/2004	Pongola [0410410 1] 40km 05/1994, 01/2004, 12/2004
B, MP	Nelspruit 0555750 9 07/1993-12/2004	Nelspruit-Agr [0555837 5] 7km 01/1986-02/1991 Nelspruit-Friedenheim [0555866 5] 10km, 03/1991-01/1993 Mayfern [0556088 4] 13km 02/1993-06/1993, other minor gaps	Mthethomusha 30km Pretoriuskop [0556460 8] 51km Vaalhoek [0594494 0] 87km 01/1986-06/1993
C, LP	Phalaborwa-Mun [0681266A8] 2km, 01/1981-04/1982 Phalaborwa [0681266B2] 05/1982- 06/1988 Phalaborwa [0681266C7] 08/1988-06/1990 Phalaborwa [0681236 1] 09/1990-11/1991 Phalaborwa [0681266 3] 08/1993-12/2004	Shimuwini [0681493 6] 30km 08,12/1991, 10/1992-07/1993, other minor gaps	Talamati [0596063 7] 80km Satara [0639474 9] 81km Shingwedzi [0725756A8] 97km 12/1991-07/1993, other minor gaps
D, LP	Messina-Macuville [0809706 X] 01/1981-12/2004	Messina-Pol [0810081 3] 18km 01/1992-01/1993	Waterpoort-Brenhilde [0764880 2] 60km, 01/1992-01/1993
E, LP	Pietersburg-Wk [0677802 0] 01/1980-11/1985 Pietersburg-Wk [0677802a5] 12/1985-04/1992 Pietersburg Wo [0677802BX] 05/1992-12/2004		Mara [0722099 1] 80km 04/1992
F, LP	Sentrum Ysterpan [0630616 8] 01/1984-12/2004	Groenvlei-Skl [0630556 9] 3km 05/1988, 03&09/1991	Ellisras [0674311 6] 70km 05/1988, 03&09/1991

† The quoted IRR is the estimated effect on case incidence of an increase in resistance (or seroprevalence) of 1%.

Table 7.3 Incidence Rate Ratios, with 95% credible intervals, estimated from two spatio-temporal models, for each of the three coefficients included in the model. Credible intervals that do not overlap with unity, correspond to statistical significance and are marked with (*).

Variable	Incidence Rate Ratio	95% credible interval
Model A: fitted on 18 years, predicted for last 6 years		
summer rainfall monthly mean (mm)	1.01	1.00 - 1.02*
temperature annual mean (°C)	3.39	2.58 - 4.34*
Drug resistance (%)†	1.02	0.996 - 1.05
Model B: fitted on all 24 years		
summer rainfall monthly mean (mm)	1.01	1.00 - 1.02*
temperature annual mean (°C)	2.70	2.26 - 3.30*
Drug resistance (%)†	1.03	1.01 - 1.04*

† The quoted IRR is the estimated effect on case incidence of an increase in drug resistance of 1%.

Table 7.4 Concordance correlation coefficients (ρ_C) between observed and predicted log incidence for two models; the number of observations and the 95% confidence intervals are shown in parentheses; $p < 0.005$ (**), $p < 0.05$ (*).

Region	Derivation period	Validation period	Entire period
Model A: fitted on 18 years, predicted for 6 years			
All regions	0.90** (n=81; 0.86, 0.94)	0.76** (n=36; 0.64, 0.89)	0.86** (n=117; 0.81, 0.9)
A	0.68** (n=18; 0.44, 0.92)	0.17 (n=6; -0.19, 0.52)	0.5** (n=24; 0.2, 0.79)
B	0.78** (n=12; 0.62, 0.94)	0.22 (n=6; -0.24, 0.67)	0.62** (n=18; 0.4, 0.84)
C	0.58** (n=13; 0.27, 0.89)	-0.53 (n=6; -1.29, 0.23)	0.54** (n=19; 0.26, 0.83)
D	0.49* (n=13; 0.13, 0.85)	-0.22 (n=6; -0.54, 0.09)	0.46** (n=19; 0.16, 0.76)
E	0.11 (n=12; -0.35, 0.57)	-0.28 (n=6; -0.79, 0.23)	0.11 (n=18; -0.28, 0.5)
F	0.56** (n=13; 0.24, 0.87)	-0.08 (n=6; -0.32, 0.15)	0.39* (n=19; 0.09, 0.69)
Model B: fitted on all 24 years			
All regions	0.89** (n=81; 0.85, 0.93)	0.97** (n=36; 0.95, 0.99)	0.92** (n=117; 0.89, 0.94)
A	0.64** (n=18; 0.4, 0.89)	0.88** (n=6; 0.7, 1.05)	0.74** (n=24; 0.56, 0.92)
B	0.77** (n=12; 0.6, 0.95)	0.86** (n=6; 0.69, 1.02)	0.81** (n=18; 0.68, 0.94)
C	0.45** (n=13; 0.15, 0.75)	0.82** (n=6; 0.52, 1.12)	0.53** (n=19; 0.26, 0.8)
D	0.46* (n=13; 0.1, 0.81)	-0.07 (n=6; -0.86, 0.71)	0.54** (n=19; 0.25, 0.82)
E	0.1 (n=12; -0.32, 0.52)	-0.1 (n=6; -0.84, 0.63)	0.1 (n=18; -0.26, 0.45)
F	0.59** (n=13; 0.3, 0.88)	0.27 (n=6; -0.35, 0.88)	0.57** (n=19; 0.32, 0.82)

Table 7.5 Suspected source of infections of malaria cases reported in the three malarious provinces of South Africa, as a percentage of the total number of cases with source reported, by province and by decade.

	Local	**Mozambique**	**Botswana, Zimbabwe, Swaziland**	**Other countries**
KwaZulu-Natal				
1980's	87.15	12.64	0.17	0.03
1990's	89.94	9.37	0.32	0.36
2000's	87.93	10.55	0.56	0.96
Mpumalanga				
1980's	48.51	51.4	0.07	0.02
1990's	56.84	42.9	0.22	0.04
2000's	30.35	68.79	0.61	0.24
Limpopo				
1980's	65.85	32.43	1.21	0.51
1990's	85.85	10.88	2.27	1
2000's	93.16	4.17	2.22	0.46

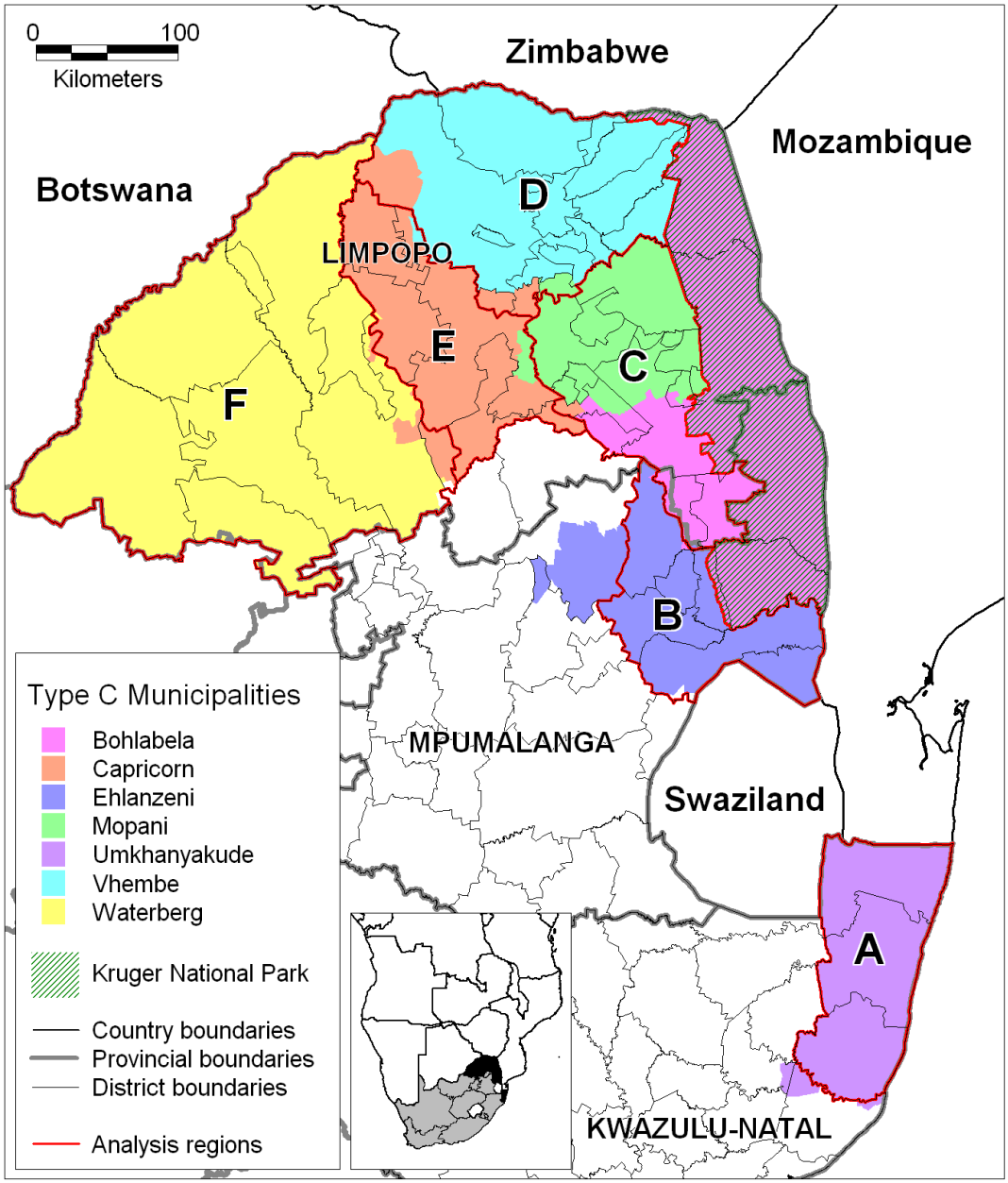


Figure 7.1 Study area showing political boundaries and analysis regions (A to F), in north-eastern South Africa.

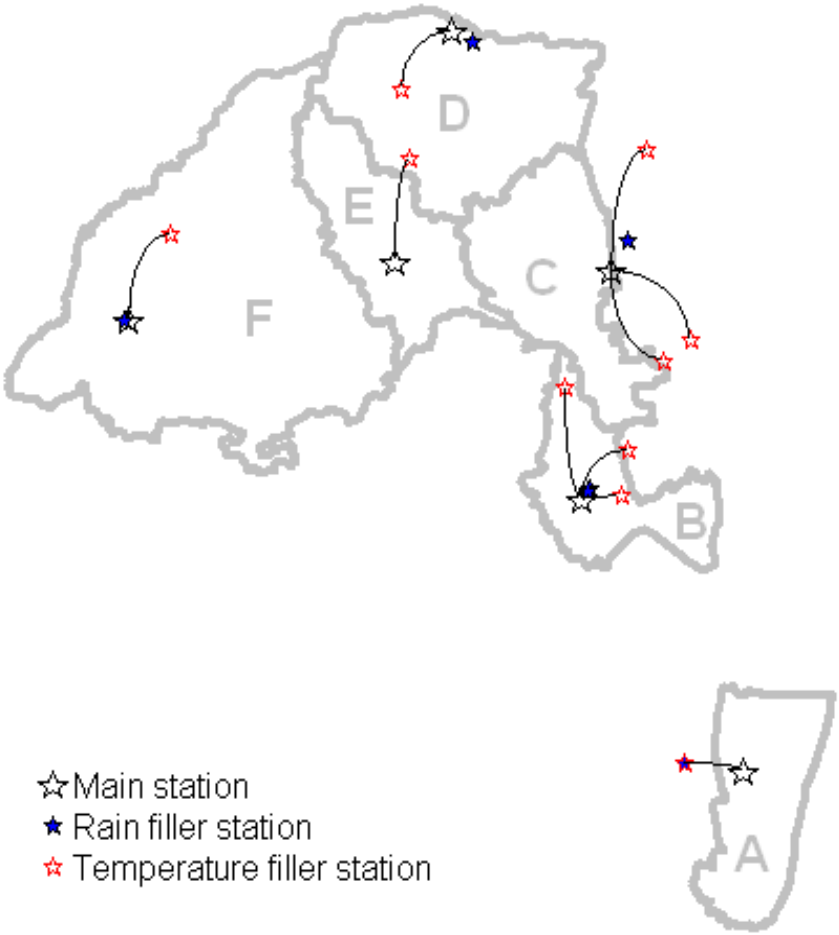


Figure 7.2 Selected weather stations and filler stations within malarious regions (A to F) used in the analysis, in north-eastern South Africa.

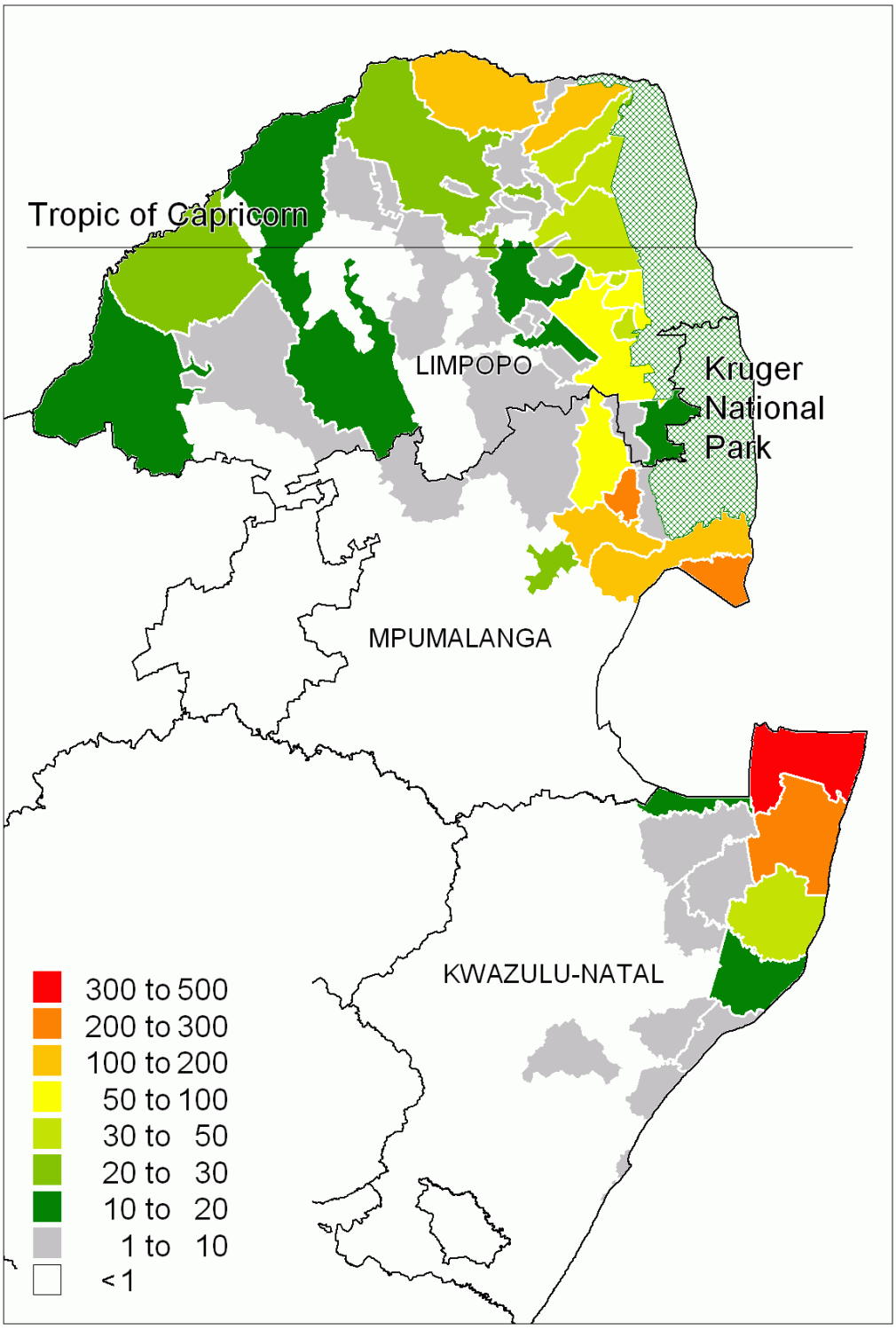


Figure 7.3 Total observed, unsmoothed, malaria cases in South Africa, of all ages, for the time period 1992 - 2004, per 1000 people, by magisterial district.

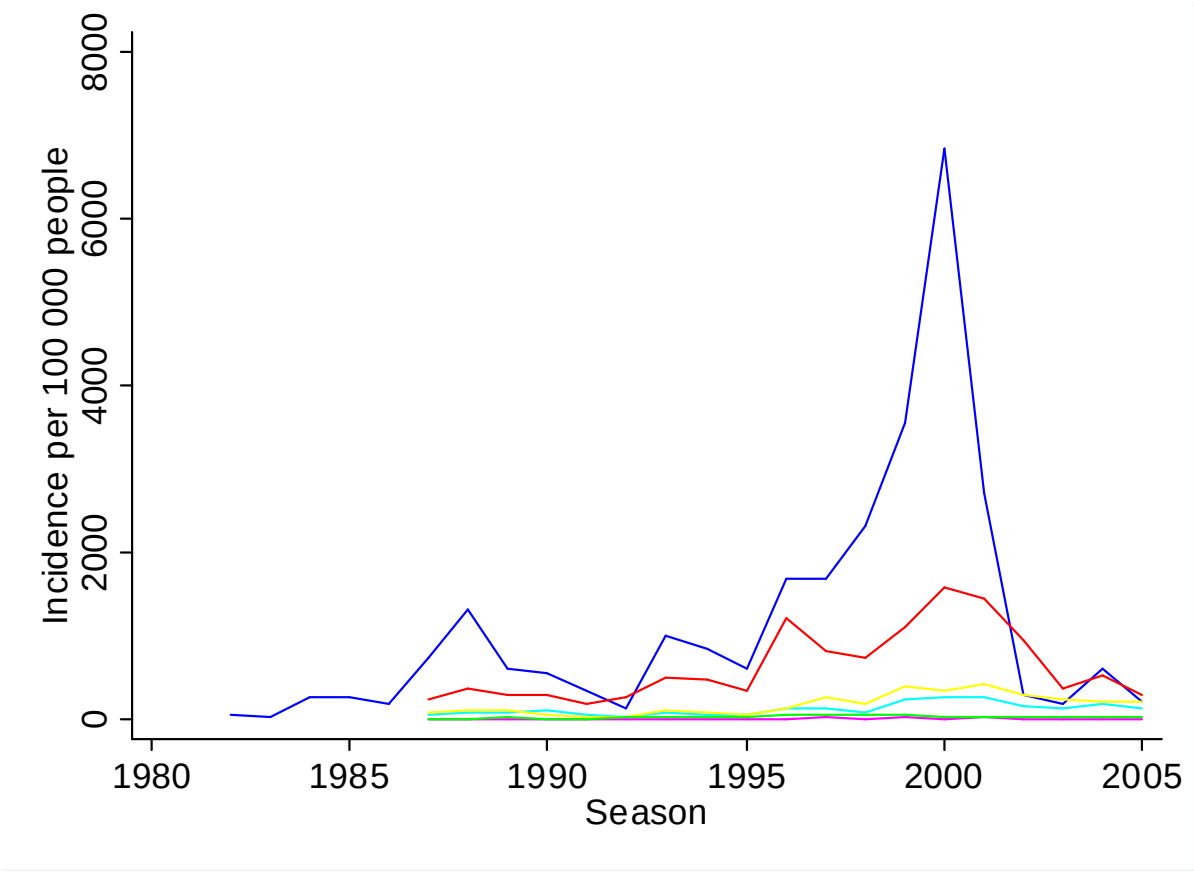


Figure 7.4 Reported total population malaria incidence in South Africa, by season (July to June), per 100 000 people, in analysis regions A (dark blue), B (red), C (light blue), D (yellow), E (pink) and F (green).

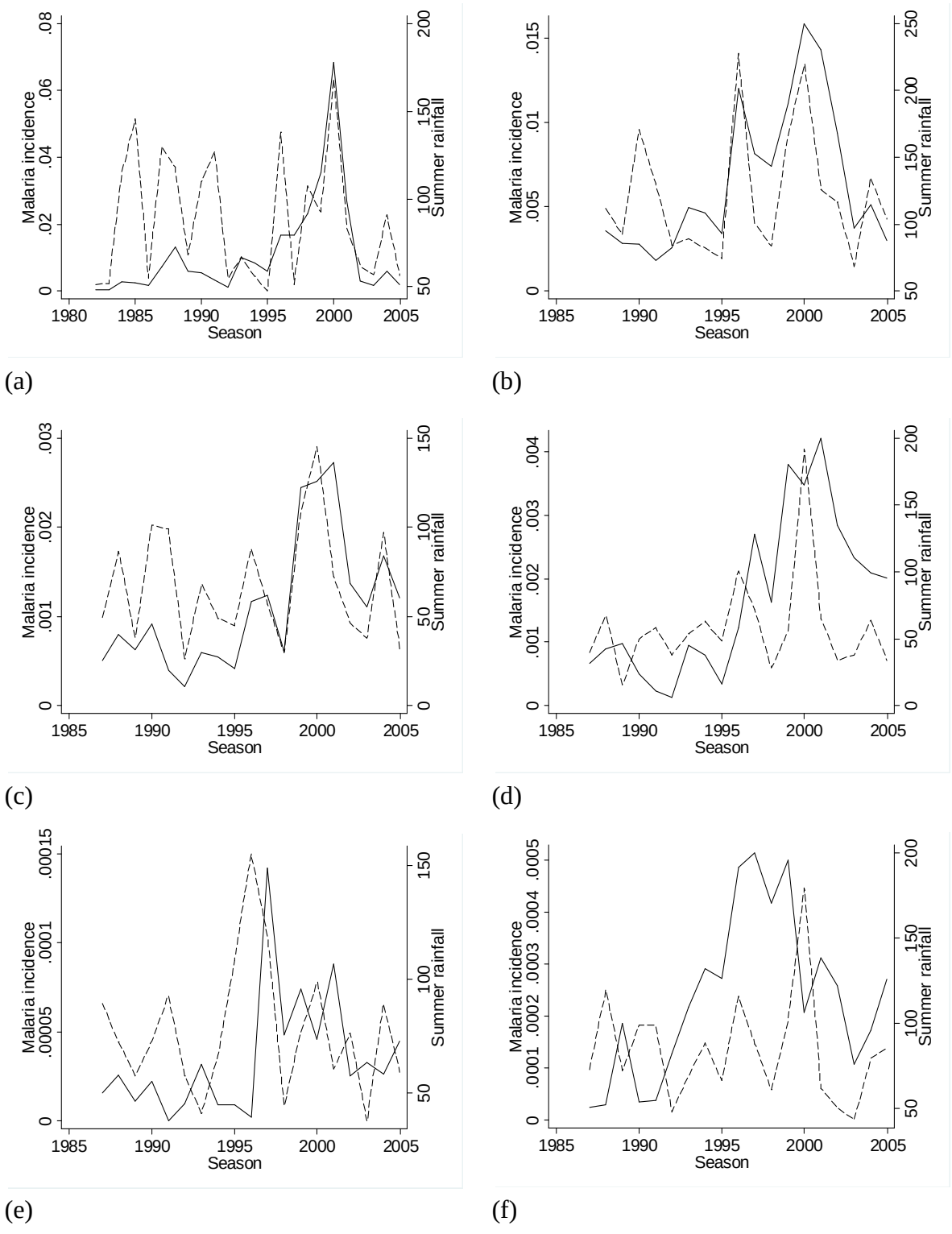


Figure 7.5 Observed total population malaria incidence (solid line) and summer rainfall (dashed line), in South Africa , by season (July to June), for analysis regions A (a); B (b); C (c); D (d); E (e) and F (f).

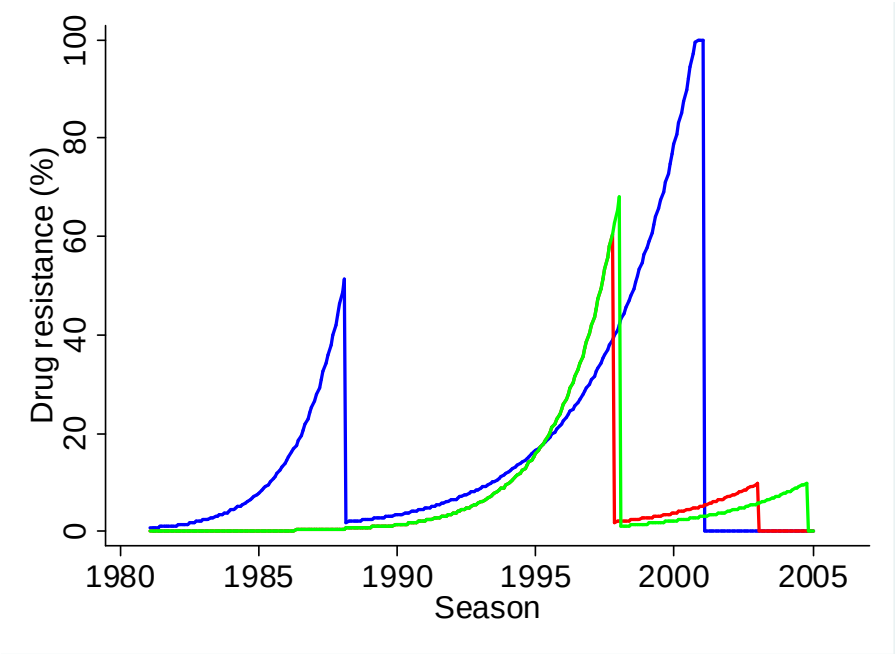


Figure 7.6 Modelled drug resistance curves, to chloroquine (CQ), then sulphadoxine-pyramithemine (SP), then artemisinin combination therapy (ACT), for three provinces in South Africa: KwaZulu-Natal (blue), Mpumalanga (red) and Limpopo (green). The curves were as follows: resistance to CQ in KwaZulu-Natal, = $1.2(1.5^{t})$, in Mpumalanga and Limpopo = $0.01(1.6^{t})$; resistance to SP in all three provinces = $1.9(1.4^{t})$; *t* is the time the drug was introduced (*t* for CQ was taken as 1980).

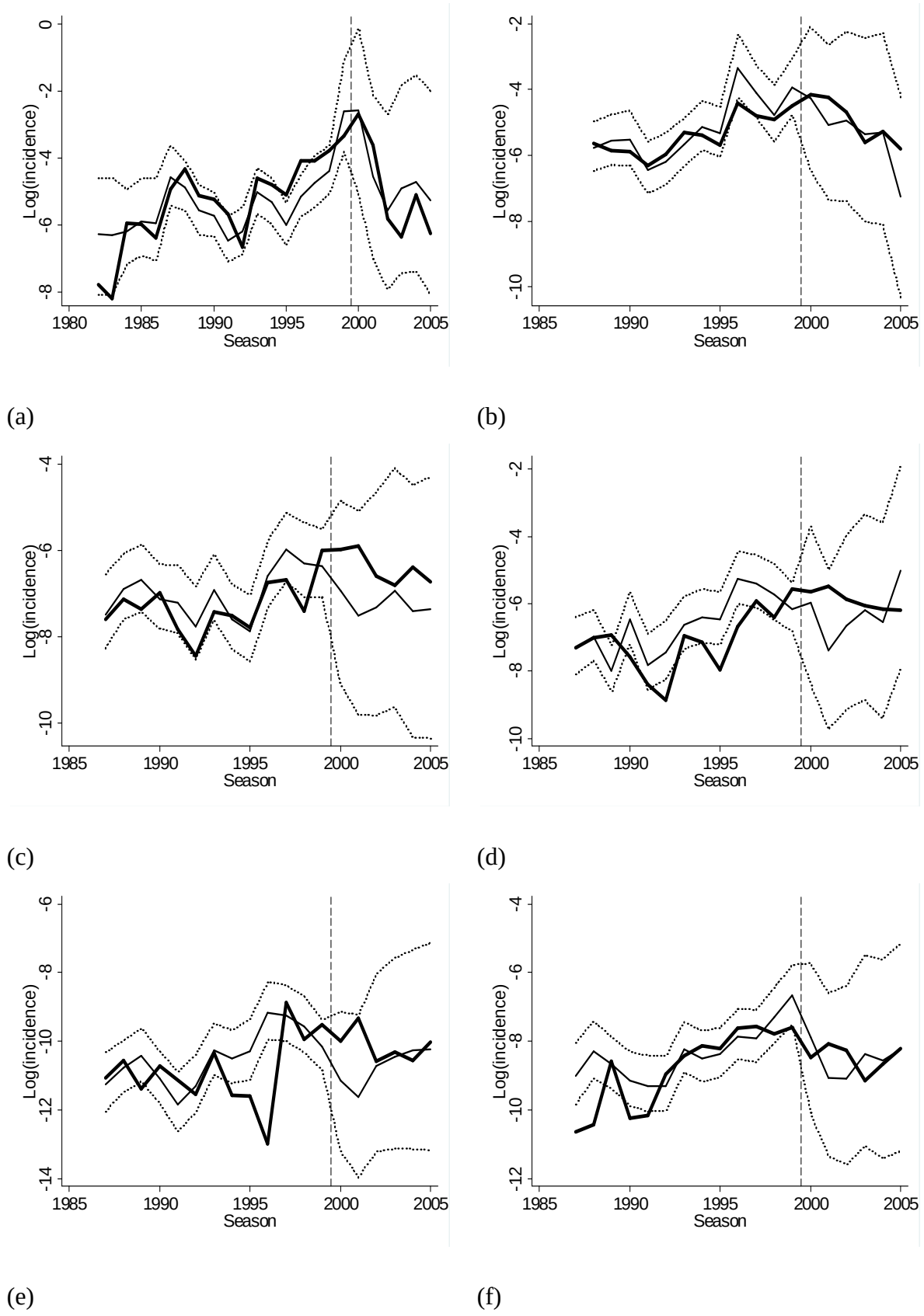


Figure 7.7 Observed (heavy solid line) and predicted (light solid line) total population incidence in six analysis regions in South Africa, on the log scale, with the upper and lower 95% credible intervals (dotted lines): region A (a); B (b); C (c); D (d); E (e) and F (f). The model was fitted on 18 years and validated on the last six years, the division indicated by the vertical dashed line.

Chapter 8

Discussion and Conclusions

Determinants of malaria

The three-way interaction between the disease agent of malaria (*Plasmodium* spp), its mosquito vector (*Anopheles* spp) and the human host, is affected by a range of genetic, behavioural, environmental and anthropogenic factors (Chapter 1). Each of these determinants varies in time and space, with different magnitudes and frequencies, so that the prediction of malaria transmission rates, in both dimensions, becomes a complex undertaking.

In sub-Saharan Africa and Southern Africa as a whole, the distribution of malaria can be estimated more or less successfully based only on mean climatic conditions (Chapter 2). This highlights the importance of temperature and rainfall in limiting transmission. Temperature has a powerful, non-linear effect on exothermic development rates of both *Plasmodium* (Macdonald 1957; Detinova 1962) and the mosquito larvae (Jepson *et al* 1947; Bayoh & Lindsay 2003; Bayoh & Lindsay 2004), and on the mosquito (Muir 1988). Rainfall affects the availability of breeding sites (Gillies & de Meillon 1968), while both temperature and rainfall affect humidity, which is another important determinant of mosquito survival (Muir 1988).

Climatic factors not only determine the distribution of *endemic* malaria, but also likely distribution and frequency of *epidemic* malaria in Africa (Cox *et al* 1999) and the nature of intra-annual seasonality (Mabaso *et al* 2007). Epidemics are particularly likely in marginal areas marked by extreme intra- and inter-annual climatic variation (Najera 1974), where

unusually suitable conditions can trigger epidemics in mostly non-immune populations (Thomson & Connor 2001b). Epidemic malaria is important in Botswana and South Africa, both of which lie in the marginal / epidemic zone of Africa, which is marked by strong seasonality.

While many studies have so far addressed the spatial distribution of malaria risk, this thesis specifically aimed to include the question of spatial variation across time. Temporal effects can also lead to potentially devastating malaria epidemics. The ability to predict epidemics is something that control programme managers in these regions sorely wish for. It has been suggested that malaria epidemics caused by meteorological factors can be predicted from climatic indicators and a process model for a malaria epidemic early warning system have been proposed (Myers *et al* 2000). Temporal effects can also lead to inter-annual variation that would be considered to be within the "normal" range, but which still has serious implications for the management and control of malaria.

Investigating trends in malaria over time requires sufficient retrospective evidence not only on malaria but also on any factors that may cause, or contribute to, the temporal trends, be they biotic or abiotic. Rogers *et al*. (2002), discussing satellite-derived environmental indices, comment that "until we can dissect quantitatively the roles played by extrinsic and intrinsic factors [..], we cannot use these new tools to forecast outbreaks". Indeed there are several problems with over-simplified approaches which try to establish statistical association between long term health data and climate variability and / or climate change, without looking for possible alternative explanations for observed trends (Chapters 3 to 7).

Comprehensive monthly malaria incidence data from South Africa - 36 years in KwaZulu-Natal province, and 18 years in Mpumalanga and Limpopo province, provided a unique

opportunity to examine the association of climatic and non-climatic factors with short- and long-term trends in malaria incidence (Chapters 5 to 7). A similar opportunity was provided by the regular prevalence surveys carried out by the Botswana Ministry of Health between 1961 and 1997, and collated within the MARA project (Chapters 3 and 4).

In South Africa, cross-border people movements, agricultural activities and changes in the case reporting system, available information on HIV prevalence, as well as control-related factors namely the emergence of insecticide and anti-malarial drug resistance, and subsequent control policy changes, were reviewed and their potential effect on malaria transmission examined (Chapters 6 and 7). In Botswana the analysis was limited to climatic and environmental factors (Chapters 3 and 4). In both countries the calendar year would have split the malaria transmission season in two, therefore the data were aggregated into periods starting in July and ending in June.

Climatic factors

An introductory remark is appropriate at this point, concerning the spatial-temporal analysis of malaria data against climate, namely the availability of climatic data with corresponding temporal and spatial extent and frequency. There are generally three main alternatives: monthly temperature and rainfall coverages, that have been interpolated from weather station data, using different techniques and sometimes incorporating other important factors such as elevation (Chapters 2 to 4); actual weather station data (Chapters 5 and 7); or remotely sensed data which offer surrogates of climate variables (Chapter 3). Each data source has weaknesses, related to availability (in terms of accessibility and timeliness), completeness, temporal and spatial extent and resolution, interpretation and accuracy. Unfortunately there is no optimal solution.

A previous analysis of the *spatial* heterogeneity of malaria incidence in northern KwaZulu-Natal, South Africa (Kleinschmidt *et al* 2001b) showed that case incidence in 1994/95 was significantly related to average winter rainfall, average winter maximum temperature and also inversely related to the distance of the nearest mapped water body. A subsequent *temporal* analysis of malaria case data (Kleinschmidt *et al* 2001c) failed to detect significant relationships of malaria incidence between 1993 and 1997, and in small geographical areas, with remotely sensed climate indicators. Further analysis was recommended.

We have now analysed total annual malaria cases, as well as inter-seasonal variability (calculated as the between year rate ratio), against a range of climatic indicators obtained from three weather stations in the highest malaria incidence area of KwaZulu-Natal (Chapter 5). No evidence was found of association between case *totals* and climate. However, the inter-annual *variation* in case numbers could be explained significantly by several climatic variables. The two most significant ones were mean maximum daily temperatures from January to October preceding the current season, or otherwise mean seasonal temperature, and total rainfall during the current summer months (December to March).

In an examination of the entire malarious area of South Africa (Chapter 7), summer rainfall and mean annual temperature were significant predictors of the observed temporal and spatial variability in malaria incidence. In South Africa, despite intense malaria control operations, climate still appears to be a major driving force of inter-annual variation in malaria incidence, though it did not seem to be linked to overall transmission levels.

When examining malaria prevalence, rather than incidence, in Botswana, climatic factors - notably again summer rainfall and mean annual temperature - were significant predictors of the *spatial* variation (Chapter 3), but not of the *temporal* variation (Chapter 4) in malaria

prevalence. The coverage of malaria control operations in 1961/62 (the period focussed on in Chapter 3) was still rather incomplete, while after 1974 (the period covered in Chapter 4) malaria control was intense and coverage more or less complete. It is not clear whether climate failed to explain prevalence over time because prevalence is less sensitive to temporal changes in climate (than incidence for example), or whether it is the intensive control that broke down the relationship between climate and prevalence. Presumably this would only be clarified through a time-space analysis of the complete prevalence data set, possibly even including the three surveys available in the 1940's, before any control operations were going on in Botswana.

Temporal coincidence of peaks and troughs alone cannot offer evidence of causal links (Chapters 3 and 5). Specifically, in this case, the link between a meteorological measurement, such as rainfall, and malaria incidence is not direct. Rather the two are connected through a web of biological processes (Chapter 1), which are non-linear and ill-defined in practice. Having said this, summer rainfall and mean temperature have emerged repeatedly, in this thesis (Chapters 3, 5 and 7), as well as in other studies, as important predictors of malaria. These two variables in particular are highly plausible in terms of the aetiology of the disease (Chapter 5).

Climate definitely has a limiting effect on malaria transmission, particularly average conditions. In this thesis climate repeatedly emerged as an important and consistent explanatory variable of the spatial distribution of malaria, in various contexts and given different types of data. Temporal trends in malaria transmission however were not so easily linked to climatic variations, even though extreme events, such as wide-spread epidemics following high rainfall and flooding, are clearly precipitated by extreme weather. It seems

therefore that non-climatic effects exert a powerful effect on malaria transmission, particularly, it appears, on the temporal distribution.

Non-climatic factors

While climate probably acts as the main ultimate limiting factor in the spatial and temporal distribution of malaria, non-climatic factors can alter or override the effects of climate at different levels and scales. Environmental details, hydrological factors such as the presence of surface water where rainfall is low, agricultural practice, deforestation and other human activities, have the power to affect distribution or rate of malaria transmission beyond what can be explained by a knowledge of general climate.

The important role of the presence and effectiveness of malaria control, both South Africa and Botswana, has already been mentioned. Control fundamentally modifies the basic transmission pattern as it emerged in response to the underlying environmental, climatic and human conditions, presumably over millennia. Conversely, in KwaZulu-Natal, the failure of control due to drug and insecticide resistance, led to a temporary return of the highly effective vector *An. funestus* and a partial return to the hyper-endemic transmission state that characterized the northern part of this province before 1940. Subsequent policy changes brought the worsening situation back under control. In Botswana malaria risk today is also much lower than it was in the 1940's, thanks to an effective national control programme (Mabaso *et al* 2004). The South African example proves again that malaria risk can quickly regain, or at least approach, historical levels if control fails. There are other examples of this around the world, where malaria re-invaded areas where it had been eradicated previously, following the interruption or suspension of control activities (Sharma 1996; Chadee *et al* 1999; Povoa *et al* 2003).

The 30-year malaria case data series in KwaZulu-Natal referred to above, as well as the shorter series in the other two provinces, revealed strong medium and long term trends, which could not be explained by climate. In KwaZulu-Natal simple, single-variable linear regression analysis showed significant association between total annual cases and the measured level of drug resistance, as well as relative measures of HIV infection (Chapter 6). The same was true for the whole country, while accounting for temporal and spatial correlation (Chapter 7). Agricultural factors and insecticide resistance appear to have affected the level of malaria transmission at certain periods and to some degree, but numerical analysis of these was not possible due to paucity of data (Chapter 6).

Migration also affects malaria incidence in KwaZulu-Natal and Mpumalanga, both of which border on Mozambique (Chapters 6 and 7). Being a border area was found to be a significant risk factor in a previous small-area analysis of incidence data in KwaZulu-Natal (Kleinschmidt *et al* 2002). Cross-border movement of infected people, from uncontrolled to controlled areas leads to the continued introduction of parasites, and the supplementation of the local parasite reservoir.

Migration also needs to be considered with respect to the transportation of drug resistant strains. A recent study (Roper *et al* 2004) inferred the global spread of drug resistance by comparing the genetics of different parasite populations and concluded that drug resistance was probably introduced to Africa from South East Asia. That chloroquine failed about eight years earlier in KwaZulu-Natal than in Mpumalanga and Limpopo (Chapter 7) suggests that chloroquine resistant strains first arrived in this province. Mpumalanga and Limpopo appear to have been spared for almost a decade by a combination of fortunate factors, including low transmission in neighbouring Swaziland, low drug pressure in neighbouring Mozambique, minimal influx of malaria from KwaZulu-Natal, tight border control with other malaria

endemic neighbouring countries, and a political climate that was uninviting to immigrants from African countries where chloroquine was already spreading by the mid-1980's. Chloroquine resistance was reported in Mozambique as early as 1975 (Pillay & Bhoola 1975) and definitely by the mid-1980s (Schapira & Schwalbach 1988), so given the high rate of imported malaria from this country (Chapters 6 and 7), it is possible that despite negligible drug pressure in Mozambique, resistant strains eventually crossed the border into Mpumalanga.

It was suggested (Chapter 6) that one may expect stronger correlation between malaria variability and climate variability where the natural state has not been greatly altered by malaria control. This was found to be the case in Botswana: climatic variation failed to account for the spatial and temporal variation in prevalence in Botswana at a time when malaria control was intense (Chapter 4), even though climate factors were able to explain much of the spatial variation in Botswana before the comprehensive coverage of control (Chapter 3).

In 2004 a Delphi workshop in South Africa set out to decide to what extent the policy change to an artemisinin-based combination antimalarial was responsible for the marked decline in malaria morbidity and mortality in KwaZulu-Natal following the epidemic of 1999/2000 (Muheki *et al* 2004). The need for such a workshop illustrates just how complex the interplay of various climatic and non-climatic risk factors is in determining malaria morbidity and mortality. Even where relevant data are available, the question can not be answered with finality. Since some important non-climatic determinants of malaria transmission are difficult if not impossible to quantify, we may remain at last with unexplained trends that still require “expert opinion” to identify and hopefully, to address.

Statistical and spatial methods

A progressive development of spatial and statistical methods employed for mapping malaria risk was seen in a series of publications, which focussed initially on Kenya (Snow *et al* 1998) and Mali (Kleinschmidt *et al* 2000). An early spatial statistical analysis of prevalence data from West Africa (Kleinschmidt *et al* 2001a) was followed by a more sophisticated analysis which included central Africa (Gemperli *et al* 2006a). Analysis of the East African (Omumbo *et al* 2005) data moved into a different direction, drawing on remote sensing rather than geo-statistical techniques, particularly discriminant analysis and Fourier-transformed remotely sensed data. In all of these studies climatic, environmental and hydrological factors were considered in an attempt to explain observed spatial variations in the prevalence data, and to predict to unobserved locations.

While developing a geo-statistical risk model from point-referenced malaria prevalence in Botswana (Chapter 3), a formerly neglected issue was addressed, namely the variable selection process. A staged process of variable selection and model formulation was demonstrated, which proved to be a practical, though not necessarily the optimal solution. Instead of using once-off automated step-wise variable selection, it was repeated on multiple bootstrap samples drawn from the data. This method made it possible to identify the most consistent and stable explanatory variables. Selection frequency provided an objective rationale for choosing one variable above another, and to choose between similar and strongly correlated indicators.

Recent developments in Bayesian spatial modelling software packages such as geoR (Christensen & Ribeiro 2002) or Winbugs (Anon. 2004b), have made the analysis of point-referenced spatial data more widely accessible, and have so opened the door to detailed and appropriate analysis of the MARA prevalence data. In 2001 this still presented an obstacle

(Kleinschmidt 2001). Bayesian methods are so useful when analysing malaria data against environmental variables, because the complexity of the malaria transmission system outlined in Chapter 1 involves so very many uncertainties, both in the nature and degree of observed or suspected associations, and Bayesian methods model this uncertainty in a formal way.

Uncertainties in the model parameters are treated as probability distributions. The shapes of the various ("prior") distributions are initially specified only as belonging to a particular family, based on the type of data or parameter involved. As the observed data are taken into consideration, the curves are then free to take on any shape ("posterior" distributions), within generous limits ("precision"). The uncertainty in any predictions and modelled risk maps is also reflected in the form of probability distributions, which reduce the risk of wrong or misleading conclusions being drawn.

Markov chain Monte Carlo algorithms allow these uncertainties to be estimated, by drawing random samples of any unknown parameter, within a wide possible range, and estimating the probability of any particular combination of these randomly generated values, given the observed data. By repeating this many times over, probability distributions are generated for each unknown parameter, which provides an overall more accurate picture than frequentist statistical methods. Gibbs sampling, used by the Winbugs software, by always sampling from the probability distributions that are being established as the sampling continues, ensure that the most probable values are being sampled most often, resulting in the model over time converging upon the most probable combination of parameter values.

Spatial tiers of modelling malaria

Modelling of malaria involves different approaches, targeting different spatial scales (Chapters 1 and 2). The first, continental, tier defined the broad distribution of disease based

on climatic conditions in an average year (Chapter 2). Other studies at this scale have followed. Tanser *et al*. (2003) for example developed a different kind of distribution model, that defined the likely absence or presence of malaria transmission on a monthly basis, also based on long-term mean climate profiles. The number of months during which transmission was possible then supplied a measure of the possible duration and timing of the average malaria transmission season, and in some areas a bi-annual seasonality pattern. This model later provided the transmission season required to infer malaria transmission intensity in West Africa, where the Garki transmission model was used to convert age-prevalence curves into estimates of entomological inoculation rates (Gemperli *et al* 2006a).

The second, sub-continental level, the distribution at the periphery is refined using annual data sets for higher temporal resolution, taking into account differences between major malaria ecological zones (MARA/ARMA). Data on epidemic malaria in the African highlands was collated and analysed within the Highlands Malaria Project (Cox *et al* 1999), a sister-project of MARA. A map of epidemic risk was produced.

The third tier of modelling considers malaria distribution at the regional or national level. This tier of modelling has been the target of many different investigations, mostly under the umbrella of the MARA project, or at least, involving data collected by the MARA project. Malaria risk models have been produced for Mali (Kleinschmidt *et al* 2000; Gemperli *et al* 2006b), West (and central) Africa (Kleinschmidt *et al* 2001a; Gemperli *et al* 2006a), Kenya (Snow *et al* 1998), East Africa (Omumbo *et al* 2005), and now for Botswana (Chapter 3).

The fourth tier of spatial modelling, at a scale of 30km^2 and below, has been addressed by various investigations seeking to identify localized risk factors, such as the location of breeding sites and their impact on various malaria risk indices (Trape *et al* 1993; Smith *et al*

1995; Thompson *et al* 1997). The usefulness of highly detailed mapping of malaria cases, such as at the house-hold level, as illustrated in KwaZulu-Natal (Hay *et al* 2000b), is debatable. In some cases it may allow highly targeted control activities, such as treatment or drainage of breeding sites, but in general this level of investigation is arguably mostly academic in nature.

Further application of malaria models

Maps produced of the malaria distribution model (Chapter 2) have evoked much interest and have been used widely, mainly for visual representation of malaria risk in Africa, by global institutions such as the World Health Organization (Anon. 2003c; Gordon *et al* 2004) or in the United States President's Malaria Initiative (Anon. 2007e) for example. The model has even been used, rightly or wrongly, for guiding prescription malaria prophylaxis for travellers, and posters of this model in particular have been distributed to pharmacies in many countries by GlaxoSmithKline.

In contrast to hand-drawn historical expert opinion maps, geographical models of malaria distribution that have been produced through an explicit numerical process have major advantages. For one it is possible to combine them with other geographical data coverages, such as a population distribution model. This has been done to estimate continental populations at risk of malaria, as well as disease burden and mortality (Snow *et al* 1999a; Snow *et al* 2003). The Africa Malaria Report (Anon. 2003c), among others, has drawn on these estimates for the countries' situation analyses. The population at risk estimated by the distribution model have also been used to make estimates for upscaling of control, for example estimating the number of insecticide treated nets required to cover to reach country targets (Miller *et al* 2007). The widespread use of the malaria distribution model illustrates

above all the desperate need for -, the potential usefulness of -, and despite decades of research, the dearth of - reliable large-scale malaria risk maps and estimates of malaria risk.

A further benefit of numerically defined distribution models is that they can be reproduced or modified by other authors, or that they can be applied to different spatial climate data sets. As such they provide a baseline against which climate change scenarios could be evaluated. Thomas *et al*.(2004) for example re-calculated the distribution model (Chapter 2) using a different climate data base (provided by the Intergovernmental Panel on Climate Change) as well as climate data predicted by the second generation Hadley Centre coupled global climate model, specifically the medium-high scenario, to examine the potential effect of climate change on malaria distribution in Africa. What emerged was a mixed picture, confirming the complexity and spatial heterogeneity of the transmission system in Africa. However, the authors concluded rather surprisingly that in the next few decades malaria distribution in Africa was more likely to shrink than to expand.

Tanser *et al*. (2003) also recalculated their continental seasonality model for three different Hadley Centre climate change scenarios, and concluded that, though the extent of malaria may change but little, transmission seasons may increase in duration and that person-months of exposure may well increase, largely in areas of existing transmission. The publication of these findings further fuelled the debate between researchers who make dire predictions of the fall-out of climate change and global warming on malaria and other vector-borne diseases, and those who urge caution and reason and a more rigorous, evidence-based approach to this issue.

The distribution model (Chapter 2) has also been applied in South Africa, to predict potential changes in malaria distribution due to climate change, as part of a vulnerability and adaptation

assessment in the South African county study on climate change (Craig & Sharp 2000). Both the potential future distribution and the increase in the number of people at risk was derived. The results from this study have been used subsequently to estimate the potential unmitigated economic impact of climate change (Turpie *et al* 2002; van Rensburg & Blignaut 2002). The potential future malaria distribution and populations at risk received wide press coverage, particularly following a cabinet meeting where the results were presented.

What neither parliament nor the public were reminded of, was that in South Africa malaria has been near eradicated, and that even without climate change, malaria could spread again to its previous level and extent, if control were to cease and existing health system were to collapse. Malaria mortality estimates by magistrates in KwaZulu-Natal from November 1931 to June 1932 totalled 22 132 (population at risk = < 1m) (Le Sueur *et al* 1993). That is >20 000 *deaths* of malaria, as opposed to <2000 *infections* (and around 15 deaths) in 2004, and around 20 000 *infections* during the most serious epidemic during the past 35 years (Chapter 5). Some fear that wildly pessimistic and highly "sellable" predictions all too often eclipse the voice of evidence and reason, provoking occasional articles such as "Hot topic or hot air?" (Hay *et al* 2002b), "Global warming and malaria: a call for accuracy" (Reiter *et al* 2004), or "Climate of fear: global-warming alarmists intimidate dissenting scientists into silence" (Lindzen 2006).

It is unfortunate that in the context of malaria the climate change debate attracts such a disproportionate amount of political attention, obscuring issues that are actually quite straight-forward. All too often the climate change debate unduly postpones the problem to some future date, diverting valuable expertise and funds, rather than dealing with the situation now, with ways and means at our disposal today. Experience has shown that monumental improvements are possible, even while global temperatures are rising. Experience has also

shown that, with or without climate change, malaria can spread and re-invade areas where was previously eradicated. The problem of malaria in Africa is enormous and has been for a very long time; climate change can hardly make matters worse.

Conclusion

As suggested by the title, this thesis contributes towards understanding how malaria transmission risk is distributed in space and over time, and what factors might explain the observed heterogeneities. This question has been addressed at different spatial scales, from the continental down to the sub-national, with focus on two countries in Southern Africa where seasonal and inter-annual variation plays an important role. After examining the large-scale distribution using an inductive approach that built on the theoretical link between climate and malaria transmission, actual malariometric data - both regional incidence and point-referenced prevalence data - were analysed. The long-term effect of mean climatic conditions was considered, as well as effects of inter- and intra-annual variations. The implications for control planning has been highlighted.

Whether inter-annual malaria incidence in general, and malaria epidemics specifically, can be predicted in advance, with sufficient accuracy and time to help plan health care and control, remains to be seen. The findings of this thesis certainly emphasise that in addition to shorter-term variation, which seems to be driven by climate in many cases, malaria transmission is largely determined by non-climatic factors. This appears to be particularly true where the natural malaria endemicity has been modified by control interventions. As the drive to control malaria in Africa continues and intensifies, the need for long-term surveillance of not merely malaria transmission, but also of the coverage and effectiveness of control interventions, will grow.

References

ANON. 1938. *Malaria areas in the Union of South Africa.* Department of Public Health, Union of South Africa, Pretoria.

ANON. 1993. *Implementation of the global Malaria Control Strategy* (Report). WHO Technical Report Series 839, World Health Organization, Geneva.

ANON. 1994. *Vitamin A Deficiency in Kenya. A report of the National Macronutrients survey* (Report). UNICEF, Kenya Country Office, Government of Kenya, Nairobi.

ANON. 1995. *Africa Data Sampler,* Edition 1 (CD-ROM). World Resources Institute, Washington, DC.

ANON. 1996. *Report of a workshop on prevention and control of malaria epidemics in southern Africa, Windhoek, Namibia, 19-23 August 1996* (Report). STP/MAL2, World Health Organization, Geneva.

ANON. 1998a. *Botswana malaria cases: district data 1982 - 1994* (Data file). Botswana Ministry of Health, Gaborone.

ANON. 1998b. *GTOPO30 global digital elevation model* (Internet). Center for Earth Resources Observation and Science, United States Geological Survey, Sioux Falls, South Dakota (http://edc.usgs.gov/products/elevation/gtopo30/gtopo30.html).

ANON. 1998c. *Malaria case records by district, 1983-1997* (Data file). Department of Health South Africa / Medical Research Council.

ANON. 1999a. *Population census, 1996* (CD-ROM). Statistics South Africa, Pretoria (http://www.statssa.gov.za).

ANON. 1999b. *SPSS for Windows,* Version 10. SPSS Inc., Chicago (http://www.spss.com/).

ANON. 2001a. *Pathfinder Advanced Very High Resolution Radiometer (AVHRR) data* (Internet). National Oceanic and Atmospheric Administration (NOAA), Goddard Distributed Active Archive Center (http://daac.gsfc.nasa.gov/data/dataset/).

ANON. 2001b. *Stata Statistical Software,* Version 7. Stata Corporation, College Station, Texas.

ANON. 2003a. *Census 2001* (CD-ROM). Statistics South Africa (http://www.statssa.gov.za).

ANON. 2003b. *Malaria epidemics: forecasting, prevention, early detection and control. From policy to practice* (Report). WHO/HTM/MAL/2004.1098, World Health Organization, Geneva.

ANON. 2003c. *The Africa Malaria Report 2003* (Report). World Health Organization, Geneva (http://www.rbm.who.int/amd2003/amr2003/amr_toc.htm).

ANON. 2004a. *The World Health Report 2004* (Report). World Health Organization, Geneva (http://www.edscuola.com/archivio/handicap/whr_2000.htm).

ANON. 2004b. *WinBUGS,* Version 1.4.1. Medical Research Council (UK) and Imperial College of Science, Technology and Medicine, London (http://www.mrc-bsu.cam.ac.uk/bugs).

ANON. 2004c. *Epidemiological Fact Sheets: on HIV/AIDS and sexually transmitted infections, South Africa* (Report). UNAIDS/WHO Working Group on Global HIV/AIDS and STI Surveillance (http://data.unaids.org/Publications/Fact-Sheets01/southafrica_EN.pdf).

ANON. 2007a. *CPC Merged Analysis of Precipitation (CMAP)* (Internet). Climate Prediction Center, National Centers for Environmental Prediction, NOAA/ National Weather

Service, Camp Springs. (http://www.cpc.ncep.noaa.gov/products/global_precip/html/wpage.cmap.shtml).

ANON. 2007b. *Global Land 1-KM AVHRR Project* (Internet). Center for Earth Resources Observation and Science, United States Geologic Survey, Sioux Falls (http://edcsns17.cr.usgs.gov/1KM/1kmhomepage.html).

ANON. 2007c. *Historical climate data* (Internet). South African Weather Service, Pretoria (http://www.weathersa.co.za).

ANON. 2007d. *HIV prevalence (%) (antenatal): percentage of women surveyed testing positive for HIV* (Internet). Health Systems Trust (http://www.hst.org.za/healthstats/13/data).

ANON. 2007e. *The United States President's Malaria Initiative* (Internet). (http://www.fightingmalaria.gov).

ADLER R.F., Huffman J.G., Chang A., Ferraro R., Xie P., Janowiak J., Rudolf B., Schneider U., Curtis S., Bolvin D., Gruber A., Susskind J. and Arkin P. 2003. The Version 2 Global Precipitation Climatology Project (GPCP) Monthly Precipitation Analysis (1979-Present). *Journal of Hydrometeorology* ***4***: 1147-1167.

AKAIKE H. 1973. Information theory and an extension of the maximum likelihood principle. In: Petrov, B. N. and F. Csaki (eds.) . *Second International Symposium on Information Theory,* Akademiai Kiado, Budapest, pp 267-281.

ANDERSON J.R., Hardy E.E., Roach J.T. and Witmer R.E. 1976. *A land use and land cover classification system for use with remote sensor data.* Unpublished document.

ANDERSON R.M. and May R.M. 1991. Indirectly transmitted microparasites. In: *Infectious Diseases of Humans: Dynamics and Control*, Oxford University Press, Oxford, pp 374-429.

APPLETON C.C., Sharp B.L. and Le Sueur D. 1995. Wetlands and water-related parasitic diseases of man in southern Africa. In: Cowan, G. (ed.) . *Wetlands in South Africa*, Department of Environmental Affairs and Tourism, Pretoria.

AUSTIN P.C. and Tu J.V. 2004. Bootstrap methods for developing predictive models. *The American Statistician* ***58***: 131-137.

BABYAK M.A. 2004. What you see may not be what you get: a brief, nontechnical introduction to overfitting in regression-type models. *Psychosomatic Medicine* ***66***: 411-421.

BARNES K.I., Durrheim D.N., Little F., Jackson A., Mehta U., Allen E., Dlamini S.S., Tsoka J., Bredenkamp B., Mthembu D.J., White N.J. and Sharp B.L. 2005. Effect of artemether-lumefantrine policy and improved vector control on malaria burden in KwaZulu-Natal, South Africa. *Public Library of Science Medicine* ***2***: e330.

BAYOH M.N. and Lindsay S.W. 2003. Effect of temperature on the development of the aquatic stages of *Anopheles gambiae sensu stricto* (Diptera: Culicidae). *Bulletin of Entomological Research* ***93***: 375-381.

BAYOH M.N. and Lindsay S.W. 2004. Temperature-related duration of aquatic stages of the Afrotropical malaria vector mosquito *Anopheles gambiae* in the laboratory. *Medical and Veterinary Entomology* ***18***: 174-179.

BLOLAND P.B., Wirima J.J., Steketee R.W., Chilima B., Hightower A. and Breman J.G. 1995. Maternal HIV infection and infant mortality in Malawi: evidence for increased mortality due to placental malaria infection. *AIDS* ***9***: 721-726.

BOUMA M.J. 2003. Methodological problems and amendments to demonstrate effects of temperature on the epidemiology of malaria. A new perspective on the highland epidemics in Madagascar, 1972-89. *Transactions of the Royal Society of Tropical Medicine and Hygiene* ***97***: 133-139.

BOUMA M.J., Sondorp H.E. and Van der Kaay H.J. 1994. Climate change and periodic epidemic malaria. *The Lancet* ***343***: 1440.

BOX G.E.P. and Pierce D.A. 1970. Distribution of residual autocorrelations in autoregressive-integrated moving average time series models. *Journal of the American Statistical Association* ***65***: 1509-1526.

BOYD M.F. 1949. *Malariology: A Comprehensive Survey of All Aspects of This Group of Diseases From a Global Standpoint*, Saunders, Philadelphia.

BREDENKAMP B.L., Sharp B.L., Mthembu S.D., Durrheim D.N. and Barnes K.I. 2001. Failure of sulphadoxine-pyrimethamine in treating *Plasmodium falciparum* malaria in KwaZulu-Natal. *South African Medical Journal* ***91***: 970-972.

BRYCE J., Roungou J.B., Nguyen-Dinh P., Naimoli J.F. and Breman J.G. 1994. Evaluation of national malaria control programmes in Africa. *Bulletin of the World Health Organization* ***72***: 371-381.

CHADEE D.D., Beier J.C. and Doon R. 1999. Re-emergence of *Plasmodium malariae* in Trinidad, West Indies. *Annals of Tropical Medicine and Parasitology* ***93***: 467-475.

CHARLWOOD J.D., Kihonda J., Sama S., Billingsley P.F., Hadji H., Verhave J.P., Lyimo E.O., Luttikhuizen P.C. and Smith T. 1995. The rise and fall of *Anopheles arabiensis* (Diptera: Culicidae) in a Tanzanian village. *Bulletin of Entomological Research* ***85***: 37-44.

CHAYABEJARA S., Sobti S.K., Payne D., and Braga F. 1975. *Malaria situation in Botswana* (Report). AFR/MAL/144, World Health Organization, Regional Office for Africa.

CHRISTENSEN O.F. and Ribeiro P.J.J. 2002. geoRglm - a package for generalised linear spatial models. *R News **2***: 26-28.

CHRISTIE M. 1958. A method for the study of larval populations of *Anopheles gambiae* and other pool-breeding mosquitoes. *Annals of Tropical Medicine and Parasitology **48***: 271.

COHEN C., Karstaedt A., Frean J., Thomas J., Govender N., Prentice E., Dini L., Galpin J. and Crewe-Brown H. 2005. Increased prevalence of severe malaria in HIV-infected adults in South Africa. *Clinical Infectious Diseases **41***: 1631-1637.

COLUZZI M. 1999. The clay feet of the malaria giant and its African roots: hypotheses and inferences about origin, spread and control of *Plasmodium falciparum*. *Parassitologia **41***: 277-283.

CONCATO J., Feinstein A.R. and Holford T.R. 1993. The risk of determining risk with multivariable models. *Annals of Internal Medicine **118***: 201-210.

COX J., Craig M.H., Le Sueur D., and Sharp B.L. 1999. *Mapping Malaria Risk in the Highlands of Africa. MARA/HIMAL Technical Report* (Report). London School of Hygiene and Tropical Medicine, London.

CRABB C. 2002. Blaming malaria rise on climate change is simplistic. *Bulletin of the World Health Organization **80***: 334-335.

CRAIG M.H., Kleinschmidt I., Le Sueur D. and Sharp B.L. 2004a. Exploring 30 years of malaria case data in KwaZulu-Natal, South Africa, Part II: the impact of non-climatic factors. *Tropical Medicine and International Health **9***: 1258-1266.

CRAIG M.H., Kleinschmidt I., Nawn J.B., Le Sueur D. and Sharp B.L. 2004b. Exploring 30 years of malaria case data in KwaZulu-Natal, South Africa, Part I: the impact of climatic factors. *Tropical Medicine and International Health* ***9***: 1247-1257.

CRAIG M.H. and Sharp B.L. 2000. *Health Section. Part 1: Malaria* (Report). South African Country Study on Climate Change Vulnerability & Adaptation Assessment. Department of Environmental Affairs and Tourism, Pretoria.

CRAIG M.H., Snow R.W. and Le Sueur D. 1999. A climate-based distribution model of malaria transmission in Africa. *Parasitology Today* ***15***: 105-111.

DE MEILLON B. 1934. Observations on *Anopheles funestus* and *Anopheles gambiae* in the Transvaal. *Publications of the South African Institute for Medical Research* ***6***: 195.

DEACON H.E., Freese J.A. and Sharp B.L. 1994. Drug-resistant *Plasmodium falciparum* malaria in the eastern Transvaal. *South African Medical Journal* ***84***: 394-395.

DEICHMANN U. 1997. *Population Density for Africa in 1990,* Edition 3 (Internet). NCGIA, UCSB, Santa Barbara (http://grid.cr.usgs.gov/clearinghouse/datalist.html).

DETINOVA T.S. 1962. Determination of the epidemiological importance of populations of *Anopheles maculipennis* by their age composition. In: *Age Grouping Methods in Diptera of Medical Importance, With Special Reference to Some Vectors of Malaria,* World Health Organization, Geneva, pp 122-150.

DICKEY D.A. and Fuller W.A. 1979. Distribution of the estimators for autoregressive time series with unit root. *Journal of the American Statistical Association* ***74***: 427-431.

DIETZ K. 1988. Mathematical models for transmission and control of malaria. In: Wernsdorfer, W. H. and I. McGregor (eds.) . *Malaria: Principles and Practice of Malariology,* Churchill Livingstone, New York, pp 1091-1133.

DIGGLE P.J., Tawn J.A. and Moyeed R. 1998. Model-based geostatistics. *Journal of the Royal Statistical Society C* ***47***: 299-350.

DOLO G., Sissoko M.S., Dao A., Traore S.F., Sissoko M., Bouare M., Dicko A., Sogoba N., Dembele, Niar O. and Bagayoko M. 1997. Impact of irrigated rice cultivation on malaria transmission in Niono, Mali. *American Journal of Tropical Medicine and Hygiene Supplement* ***57***: 183-184.

DURRHEIM D.N., Sharp B.L. and Barnes K.I. 2001. Sentinel malaria surveillance - more than a research tool. *South African Medical Journal* ***91***: 968-970.

FREEDMAN M.L. 1953. *Malaria Control* (Report). The Botswana National Archives and Records Services, Gaborone.

FREEMAN T. and Bradley M. 1996. Temperature is predictive of severe malaria years in Zimbabwe. *Transactions of the Royal Society of Tropical Medicine and Hygiene* ***90***: 232.

FREESE J.A., Robinson A., Roper C., Rossouw E.J., Bredenkamp B.L.F. and Gouws E. 2000. *A randomised controlled trial of chloroquine vs sulfadoxine / pyrimethamine for the treatment of falciparum malaria in KwaZulu-Natal.* Unpublished document.

FREESE J.A., Sharp B.L., Bredenkamp B.L., Gouws E., Mthembu D.J., La Grange J.J.P., Kruger P. and Durrheim D.N. 2001. *Operations management of drug effectiveness against Plasmodium falciparum malaria in the endemic areas of South Africa.* Unpublished document.

FREESE J.A., Sharp B.L., Ngxongo S.M. and Markus M.B. 1988. *In vitro* confirmation of chloroquine-resistant *Plasmodium falciparum* malaria in KwaZulu. *South African Medical Journal* ***74***: 576-578.

GARG M.R., Gogtay N.J., Kotwani R.N., Bodhe P.V. and Kshirsagar N.A. 1999. Resurgence of malaria in Mumbai--is escalating chloroquine resistance a cause? *The Journal of the Association of Physicians of India* ***47***: 377-379.

GARNHAM P.C.C. 1988. Malaria parasites of man: life cycles and morphology (excluding ultrastructure). In: Wernsdorfer, W. H. and I. McGregor (eds.) . *Malaria: Principles and Practice of Malariology,* Churchill Livingstone, Edinburgh, pp 61-96.

GARRETT-JONES C. and Grab B. 1964. The assessment of insecticidal impact on the malaria mosquito's vectorial capacity, from the data on the populations of parous females. *Bulletin of the World Health Organization* ***31***: 71-86.

GEMPERLI A., Sogoba N., Fondjo E., Mabaso M., Bagayoko M., Briet O.J., Anderegg D., Liebe J., Smith T. and Vounatsou P. 2006a. Mapping malaria transmission in West and Central Africa. *Tropical Medicine and International Health* ***11***: 1032-1046.

GEMPERLI A. and Vounatsou P. 2003. Fitting generalized linear mixed models for point-referenced spatial data. *Journal of Modern Applied Statistical Methods* ***2***: 497-511.

GEMPERLI A., Vounatsou P., Kleinschmidt I., Bagayoko M., Lengeler C. and Smith T. 2004. Spatial patterns of infant mortality in Mali: the effect of malaria endemicity. *American Journal of Epidemiology* ***159***: 64-72.

GEMPERLI A., Vounatsou P., Sogoba N. and Smith T. 2006b. Malaria mapping using transmission models: application to survey data from Mali. *American Journal of Epidemiology* ***163***: 289-297.

GESLER W. 1986. The uses of spatial analysis in medical geography: a review. *Social Science and Medicine* ***23***: 963-973.

GILL C.A. 1938. *The Seasonal Periodicity of Malaria and the Mechanism of the Epidemic Wave*, Churchill, London.

GILLIES M.T. 1988. Anopheline mosquitos: vector behaviour and bionomics. In: Wernsdorfer, W. H. and I. McGregor (eds.) . *Malaria: Principles and Practice of Malariology*, Churchill Livingstone, Edinburgh, pp 453-485.

GILLIES M.T. and de Meillon B. 1968. *The Anophelinae of Africa South of the Sahara*, The South African Institute for Medical Research, Johannesburg.

GORDON B., Mackay R., and Rehfuess E. 2004. *Inheriting the World: The Atlas of Children's Health and the Environment* (Report). World Health Organization, Geneva.

GOVERE J.M., La Grange J.J.P., Durrheim D.N., Freese J.A., Sharp B.L., Mabuza A., Mngomezulu N. and Bredenkamp B.L. 1999. Sulfadoxine-pyrimethamine effectiveness against *Plasmodium falciparum* malaria in Mpumalanga Province, South Africa. *Transactions of the Royal Society of Tropical Medicine and Hygiene* ***93***: 644.

GREENBERG A.E., Nsa W., Ryder R.W., Medi M., Nzeza M., Kitadi N., Baangi M., Malanda N., Davachi F. and Hassig S.E. 1991. *Plasmodium falciparum* malaria and perinatally acquired human immunodeficiency virus type 1 infection in Kinshasa, Zaire. A prospective, longitudinal cohort study of 587 children. *The New England Journal of Medicine* ***325***: 105-109.

GUERRA C.A., Snow R.W. and Hay S.I. 2006. A global assessment of closed forests, deforestation and malaria risk. *Annals of Tropical Medicine and Parasitology* ***100***: 189-204.

GWADZ and Collins. 1996. Anopheline mosquitoes and the agents they transmit. In: Beaty and Marquardt (eds.) . *The Biology of Disease Vectors*, Colorado University Press.

HADDOW A.J. 1943. Measurements of temperature and light in artificial pools with reference to the larval habitat of *Anopheles (Myzomyia) gambiae* Giles and *A. (M.) funestus* Giles. *Bulletin of Entomological Research* ***34*** : 89.

HANSFORD C.F. 1989. Chloroquine resistance in *Plasmodium falciparum* in KwaZulu, 1983-1988. *South African Medical Journal* ***76***: 546-547.

HARGREAVES K., Koekemoer L.L., Brooke B.D., Hunt R.H., Mthembu J. and Coetzee M. 2000. *Anopheles funestus* resistant to pyrethroid insecticides in South Africa. *Medical and Veterinary Entomology* ***14***: *181- 189.*

HARINASUTA T. and Bunnag D. 1988. The clinical features of malaria. In: Wernsdorfer, W. H. and I. McGregor (eds.) . *Malaria: Principles and Practice of Malariology,* Churchill Livingstone, Edinburgh, pp 709-734.

HARRELL F.E., Jr. 2001. *Regression Modeling Strategies: With Applications to Linear Models, Logistic Regression and Survival Analysis,* Springer, New York.

HAY S.I., Cox J., Rogers D.J., Randolph S.E., Stern D.I., Shanks G.D., Myers M.F. and Snow R.W. 2002a. Climate change and the resurgence of malaria in the East African highlands. *Nature* ***415***: 905-909.

HAY S.I., Guerra C.A., Tatem A.J., Noor A.M. and Snow R.W. 2004. The global distribution and population at risk of malaria: past, present, and future. *The Lancet Infectious Diseases* ***4***: 327-336.

HAY S.I., Myers M.F., Burke D.S., Vaughn D.W., Endy T., Ananda N., Shanks G.D., Snow R.W. and Rogers D.J. 2000a. Etiology of interepidemic periods of mosquito-borne disease. *Proceedings of the National Academy of Sciences of the United States of America* ***97***: 9335-9339.

HAY S.I., Omumbo J.A., Craig M.H. and Snow R.W. 2000b. Earth observation, geographic information systems and *Plasmodium falciparum* malaria in sub-Saharan Africa. *Advances in Parasitology* ***47***: 173-215.

HAY S.I., Rogers D.J., Randolph S.E., Stern D.I., Cox J., Shanks G.D. and Snow R.W. 2002b. Hot topic or hot air? Climate change and malaria resurgence in East African highlands. *Trends in Parasitology* ***18***: 530-534.

HAY S.I., Rogers D.J., Toomer J.F. and Snow R.W. 2000c. Annual *Plasmodium falciparum* entomological inoculation rates (EIR) across Africa: literature survey, Internet access and review. *Transactions of the Royal Society of Tropical Medicine and Hygiene* ***94***: 113-127.

HAY S.I., Snow R.W. and Rogers D.J. 1998. Predicting malaria seasons in Kenya using multitemporal meteorological satellite sensor data. *Transactions of the Royal Society of Tropical Medicine and Hygiene* ***92***: 12-20.

HERBST J.M., Taylor L.A. and Joubert S.M. 1987. Chloroquine resistance in *Plasmodium falciparum* in Natal. *South African Medical Journal* ***72***: 627-629.

HIGHTOWER A.W., Ombok M., Otieno R., Odhiambo R., Oloo A.J., Lal A.A., Nahlen B.L. and Hawley W.A. 1998. A geographic information system applied to a malaria field study in western Kenya. *American Journal of Tropical Medicine and Hygiene* ***58***: 266-272.

HOFFMAN I.F., Jere C.S., Taylor T.E., Munthali P., Dyer J.R., Wirima J.J., Rogerson S.J., Kumwenda N., Eron J.J., Fiscus S.A., Chakraborty H., Taha T.E., Cohen M.S. and Molyneux M.E. 1999. The effect of *Plasmodium falciparum* malaria on HIV-1 RNA blood plasma concentration. *AIDS* ***13***: 487-494.

HUTCHINSON M.F., Nix H.A., McMahan J.P. and Ord K.D. 1995. *Africa - A topographic and climatic database,* Edition 1 (CD-ROM)., Canberra.

JEPSON W.F., Moutia A. and Courtois C. 1947. The malaria problem in Mauritius: the bionomics of Mauritian anophelines. *Bulletin of Entomological Research* ***38***: 177-208.

JUSTICE A.C., Covinsky K.E. and Berlin J.A. 1999. Assessing the generalizability of prognostic information. *Annals of Internal Medicine* ***130***: 515-524.

KALYESUBULA I., Musoke-Mudido P., Marum L., Bagenda D., Aceng E., Ndugwa C. and Olness K. 1997. Effects of malaria infection in human immunodeficiency virus type 1-infected Ugandan children. *The Pediatric Infectious Disease Journal* ***16***: 876-881.

KAMAT V. 2000. Resurgence of malaria in Bombay (Mumbai) in the 1990s: a historical perspective. *Parassitologia* ***42***: 135-148.

KILIAN A.H., Langi P., Talisuna A. and Kabagambe G. 1999. Rainfall pattern, El Nino and malaria in Uganda. *Transactions of the Royal Society of Tropical Medicine and Hygiene* ***93***: 22-23.

KITRON U., Otieno L.H., Hungerford L.L., Odulaja A., Brigham W.U., Okello O.O., Joselyn M., Mohamed-Ahmed M.M. and Cook E. 1996. Spatial analysis of the distribution of tsetse flies in the Lambwe Valley, Kenya, using Landsat TM satellite imagery and GIS. *Journal of Animal Ecology* ***65***: 371-380.

KLEINSCHMIDT I.2001. *Spatial statistical anlaysis, modelling and mapping of malaria in Africa*. Doktor der Philosophie, Philosophisch-Naturwissenschafliche Fakultät, Universität Basel.

KLEINSCHMIDT I., Bagayoko M., Clarke G.P., Craig M. and Le Sueur D. 2000. A spatial statistical approach to malaria mapping. *International Journal of Epidemiology* ***29***: 355-361.

KLEINSCHMIDT I., Omumbo J., Briet O., Van De G.N., Sogoba N., Mensah N.K., Windmeijer P., Moussa M. and Teuscher T. 2001a. An empirical malaria distribution map for West Africa. *Tropical Medicine and International Health* ***6***: 779-786.

KLEINSCHMIDT I., Sharp B., Mueller I. and Vounatsou P. 2002. Rise in malaria incidence rates in South Africa: a small-area spatial analysis of variation in time trends. *American Journal of Epidemiology* ***155***: 257-264.

KLEINSCHMIDT I. and Sharp B.L. 2001. Patterns in age-specific malaria incidence in a population exposed to low levels of malaria transmission intensity. *Tropical Medicine and International Health* ***6***: 986-991.

KLEINSCHMIDT I., Sharp B.L., Clarke G.P.Y., Curtis B. and Fraser C. 2001b. The use of generalised linear mixed models in the spatial analysis of small area malaria incidence rates in KwaZulu-Natal, South Africa. *American Journal of Epidemiology* ***153***: 1213-1221.

KLEINSCHMIDT I., Vounatsou P., Sharp B.L., Curtis B., Hayes S. and Cox J. 2001c. *Space-time models of small area malaria incidence in relation to remote-sensed inter-annual climatic variation.* Unpublished document.

KOVATS R.S. 2000. El Nino and human health. *Bulletin of the World Health Organization* ***78***: 1127-1135.

KOVATS R.S., Bouma M.J., Hajat S., Worrall E. and Haines A. 2003. El Nino and health. *The Lancet* ***362***: 1481-1489.

KRUGER P., Durrheim D.N. and Hansford C.F. 1996. Increasing chloroquine resistance -- the Mpumalanga Lowveld story, 1990-1995. *South African Medical Journal* ***86***: 280-281.

LE SUEUR D.1991. *The ecology, over-wintering and population dynamics of the pre-imaginal stages of the Anopheles gambiae Giles complex (Diptera: Culicidae) in northern Natal, South Africa*. Doctoral thesis, University of Natal, Pietermaritzburg.

LE SUEUR D., Binka F., Lengeler C., de Savigny D., Snow R.W., Teuscher T. and Touré Y.T. 1997. An atlas of malaria in Africa. *Africa Health* ***19***: 23-24.

LE SUEUR D. and Sharp B.L. 1988. The breeding requirements of three members of the *Anopheles gambiae* Giles complex (Diptera: Culicidae) in the endemic malaria area of Natal South Africa. *Bulletin of Entomological Research* ***78***: 549-560.

LE SUEUR D., Sharp B.L. and Appleton C.C. 1993. Historical perspective of the malaria problem in Natal with emphasis on the period 1928-1932. *South African Journal of Science* ***89***: 232-239.

LE SUEUR D., Sharp B.L., Fraser C. and Ngxongo S.M. 1993. Assessment of the residual efficacy of lambda-cyhalothrin. 1. A laboratory study using *Anopheles arabiensis* and *Cimex lectularius* (Hemiptera: Cimicidae) on treated daub wall substrates from Natal, South Africa. *Journal of the American Mosquito Control Association* ***9***: 408-413.

LEESON H.S. 1931. *Anopheline Mosquitos in Southern Rhodesia*, The London School of Hygiene and Tropical Medicine, London.

LIN L.I.K. 1989. A concordance correlation coefficient to evaluate reproducibility. *Biometrics* ***45***: 255-268.

LIN L.I.K. 2000. A note on the concordance correlation coefficient. *Biometrics* ***56***: 324-325.

LINDBLADE K.A., Walker E.D., Onapa A.W., Katungu J. and Wilson M.L. 1999. Highland malaria in Uganda: prospective analysis of an epidemic associated with El Nino. *Transactions of the Royal Society of Tropical Medicine and Hygiene* ***93***: 480-487.

LINDSAY S.W., Wilkins H.A., Zieler H.A., Daly R.J., Petrarca V. and Byass P. 1991. Ability of *Anopheles gambiae* mosquitoes to transmit malaria during the dry and wet seasons in an area of irrigated rice cultivation in The Gambia. *Journal of Tropical Medicine and Hygiene* ***94***: 313-324.

LINDZEN R. 2006. Editorial: Climate of fear. *The Wall Street Journal* Wednesday, April 12, 2006 12:01 a.m.

LUNN D.J., Thomas A., Best N.G. and Spiegelhalter D.J. 2000. WinBUGS -- a Bayesian modelling framework: concepts, structure, and extensibility. *Statistics and Computing* ***10***: 325-337.

LYSENKO A.Y. and Semashko I.N. 1968. Geography of malaria. In: Lebedew, A. W. (ed.) . *Medical Geography,* Academy of Sciences USSR, Moscow.

MABASO M.L., Craig M., Ross A. and Smith T. 2007. Environmental predictors of the seasonality of malaria transmission in Africa: the challenge. *American Journal of Tropical Medicine and Hygiene* ***76***: 33-38.

MABASO M.L., Kleinschmidt I., Sharp B. and Smith T. 2006a. El Nino Southern Oscillation (ENSO) and annual malaria incidence in Southern Africa. *Transactions of the Royal Society of Tropical Medicine and Hygiene* .

MABASO M.L., Sharp B. and Lengeler C. 2004. Historical review of malarial control in southern African with emphasis on the use of indoor residual house-spraying. *Tropical Medicine and International Health* ***9***: 846-856.

MABASO M.L., Vounatsou P., Midzi S., Da Silva J. and Smith T. 2006b. Spatio-temporal analysis of the role of climate in inter-annual variation of malaria incidence in Zimbabwe. *International Journal of Health Geographics* ***5***: 20.

MABUZA A., Govere J.M., Durrheim D.N., Mngomezulu N., Bredenkamp B.L., Barnes K.I. and Sharp B.L. 2001. Therapeutic efficacy of sulfadoxine-pyrimethamine in uncomplicated *Plasmodium falciparum* malaria three years after introduction in Mpumalanga. *South African Medical Journal* ***91***: 975-978.

MACDONALD G. 1957. *The Epidemiology and Control of Malaria*, Oxford University Press, London.

MACÉ J.M., Boussinesq M., Ngoumou P., Enyegue Oye J., Koéranga A. and Godin C. 1997. Country-wide rapid epidemiological mapping of onchocerciasis (REMO) in Cameroon. *Annals of Tropical Medicine and Parasitology* ***91***: 379-391.

MAHARAJ R.1995. *Effects of temperature and humidity on adults of the Anopheles gambiae complex (Diptera: Culicidae) in South Africa - implications for malaria transmission and control*. Doctoral thesis, University of Natal, Pietermaritzburg.

MAHARAJ R., Mthembu D.J. and Sharp B.L. 2005. Impact of DDT re-introduction on malaria transmission in KwaZulu-Natal. *South African Medical Journal* ***95***: 871-874.

MALONE J.B., Abdel-Rahman M.S., El Bahy M.M., Huh O.K., Shafik M. and Bavia M. 1997. Geographic information systems and the distribution of *Schistosoma mansoni* in the Nile delta. *Parasitology Today* ***13***: 112-119.

MARA/ARMA. 1998. *Towards an atlas of malaria risk in Africa: First technical report of the MARA/ARMA (Mapping Malaria Risk in Africa) collaboration.* (Report). Medical Research Council, Durban.

MARTENS W.J.M.1997. *Health Impacts of Climate Change and Ozone Depletion: An Eco-Epidemiological Modelling Approach*. Doctoral thesis, Maastricht University.

MARTENS W.J.M., Niessen L.W., Rotmans J., Jetten T.H. and McMichael A.J. 1995. Potential impact of global climate change on malaria risk. *Environmental Health Perspectives* ***103***: 458-464.

MBOGO C.M., Mwangangi J.M., Nzovu J., Gu W., Yan G., Gunter J.T., Swalm C., Keating J., Regens J.L., Shililu J.I., Githure J.I. and Beier J.C. 2003. Spatial and temporal heterogeneity of *Anopheles* mosquitoes and *Plasmodium falciparum* transmission along the Kenyan coast. *American Journal of Tropical Medicine and Hygiene* ***68***: 734-742.

MCMICHAEL A.J. and Martens W.J.M. 1995. The health impacts of global climate change: grappling with scenarios, predictive models and multiple uncertainties. *Ecosystem Health* ***1***: 23-33.

MILLER J.M., Korenromp E.L., Nahlen B.L. and Steketee R.W. 2007. Estimating the number of insecticide-treated nets required by African households to reach continent-wide malaria coverage targets. *Journal of the American Medical Association* ***297***: 2241-2250.

MITCHELL T.D., Hulme M. and New M. 2003. *CRU TS 2.0 high-resolution gridded climate data,* Edition 1 (Internet). Climate Research Unit, University of East Anglia, Norwich. (http://www.cru.uea.ac.uk/cru/data/hrg.htm).

MOLINEAUX L. 1988. The epidemiology of human malaria as an explanation of its distribution, including some implications for its control. In: Wernsdorfer, W. H. and I. McGregor (eds.) . *Malaria: Principles and Practice of Malariology,* Churchill Livingstone, Edinburgh, pp 913-998.

MOLINEAUX L. and Gramiccia G. 1980. *The Garki Project: Research on the Epidemiology and Control of Malaria in the Sudan Savanna of West Africa.*, World Health Organization, Geneva.

MOLINEAUX L., Muir D.A., Spencer H.C. and Wernsdorfer W.H. 1988. The epidemiology of malaria and its measurement. In: Wernsdorfer, W. H. and I. McGregor (eds.) . *Malaria: Principles and Practice of Malariology*, Churchill Livingstone, Edinburgh, pp 999-1089.

MOUCHET J., Manguin S., Sircoulon J., Laventure S., Faye O., Onapa A.W., Carnevale P., Julvez J. and Fontenille D. 1998. Evolution of malaria in Africa for the past 40 years: impact of climatic and human factors. *Journal of the American Mosquito Control Association* ***14*** : 121-130.

MUHEKI C., McIntyre D. and Barnes K.I. 2004. Artemisinin-based combination therapy reduces expenditure on malaria treatment in KwaZulu Natal, South Africa. *Tropical Medicine and International Health* ***9***: 959-966.

MUIR D.A. 1988. Anopheline mosquitos: vector reproduction, life-cycle and biotope. In: Wernsdorfer, W. H. and I. McGregor (eds.) . *Malaria: Principles and Practice of Malariology*, Churchill Livingstone, New York, pp 431-451.

MULLER O. and Moser R. 1990. The clinical and parasitological presentation of *Plasmodium falciparum* malaria in Uganda is unaffected by HIV-1 infection. *Transactions of the Royal Society of Tropical Medicine and Hygiene* ***84***: 336-338.

MURRAY C.J. and Lopez A.D. 1997. Mortality by cause for eight regions of the world: Global Burden of Disease Study. *The Lancet* ***349***: 1269-1276.

MYERS M.F., Rogers D.J., Cox J., Flahault A. and Hay S.I. 2000. Forecasting disease risk for increased epidemic preparedness in public health. *Advances in Parasitology* ***47***: 309-330.

NABARRO D.N. and Tayler E.M. 1998. The "roll back malaria" campaign. *Science* ***280***: 2067-2068.

NAJERA J.A. 1974. A critical review of the field application of a mathematical model of malaria eradication. *Bulletin of the World Health Organization* ***50***: 449-457.

NÁJERA J.A., Kouznetsov R.L., and Delacollette C. 1998. *Malaria epidemics: detection and control, forecasting and prevention* (Report). WHO/MAL/98.1084, World Health Organization, Geneva.

NELSON G.S. 1959. *Atlas of Kenya.* Crown Printers, Nairobi.

NEW M. and Hulme M. 1997. *Construction of 3 minute latitude / longitude monthly climate surfaces over Africa for the period 1951-1995* (CD-ROM). Climatic Research Unit, University of East Anglia, Norwich and the Mapping Malaria Risk in Africa Initiative, South African Medical Research Council, Durban. (http://www.mara.org.za./climatecd/info.htm).

OMUMBO J., Ouma J., Rapuoda B., Craig M.H., Le Sueur D. and Snow R.W. 1998. Mapping malaria transmission intensity using geographical information systems (GIS): an example from Kenya. *Annals of Tropical Medicine and Parasitology* ***92***: 7-21.

OMUMBO J.A., Hay S.I., Snow R.W., Tatem A.J. and Rogers D.J. 2005. Modelling malaria risk in East Africa at high-spatial resolution. *Tropical Medicine and International Health* ***10***: 557-566.

ONORI E. and Grab B. 1980. Indicators for the forecasting of malaria epidemics. *Bulletin of the World Health Organization* ***58***: 91-98.

ORKIN F.M. 1999. *Mid-year estimates 1999* (Report). Statistical release P0302, Statistics South Africa, Pretoria.

PATZ J.A., Hulme M., Rosenzweig C., Mitchell T.D., Goldberg R.A., Githeko A.K., Lele S., McMichael A.J. and Le Sueur D. 2002. Climate change: Regional warming and malaria resurgence. *Nature **420***: 627-628.

PATZ J.A. and Kovats R.S. 2002. Hotspots in climate change and human health. *British Medical Journal **325***: 1094-1098.

PILLAY N. and Bhoola R.L. 1975. Probable chloroquine-resistant *Plasmodium falciparum* malaria from Mozambique A case report. *South African Medical Journal **49***: 1443-1444.

POVOA M.M., Conn J.E., Schlichting C.D., Amaral J.C., Segura M.N., Da Silva A.N., dos Santos C.C., Lacerda R.N., de Souza R.T., Galiza D., Santa Rosa E.P. and Wirtz R.A. 2003. Malaria vectors, epidemiology, and the re-emergence of *Anopheles darlingi* in Belem, Para, Brazil. *Journal of Medical Entomology **40***: 379-386.

REITER P. 2000. From Shakespeare to Defoe: malaria in England in the Little Ice Age. *Emerging Infectious Diseases **6***: 1-11.

REITER P., Thomas C.J., Atkinson P.M., Hay S.I., Randolph S.E., Rogers D.J., Shanks G.D., Snow R.W. and Spielman A. 2004. Global warming and malaria: a call for accuracy. *The Lancet Infectious Diseases **4***: 323-324.

ROBERT V., Awono-Ambene H.P., Le Hesran J.Y. and Trape J.F. 2000. Gametocytemia and infectivity to mosquitoes of patients with uncomplicated *Plasmodium falciparum* malaria attacks treated with chloroquine or sulfadoxine plus pyrimethamine. *American Journal of Tropical Medicine and Hygiene **62***: 210-216.

ROGERS D.J., Randolph S.E., Snow R.W. and Hay S.I. 2002. Satellite imagery in the study and forecast of malaria. *Nature* ***415***: 710-715.

ROPER C., Pearce R., Nair S., Sharp B., Nosten F. and Anderson T. 2004. Intercontinental spread of pyrimethamine-resistant malaria. *Science* ***305***: 1124.

ROSS R. 1911. *The Prevention of Malaria*, Murray, London.

SACHS J. and Malaney P. 2002. The economic and social burden of malaria. *Nature* ***415***: 680-685.

SCHAPIRA A. and Schwalbach J.F. 1988. Evaluation of four therapeutic regimens for falciparum malaria in Mozambique, 1986. *Bulletin of the World Health Organization* ***66***: 219-226.

SCHULZE R.E. 1997. *South African atlas of agrohydrology and -climatology* (Report). TT82/96, Water Research Commission, Pretoria.

SHANKS G.D., Biomndo K., Hay S.I. and Snow R.W. 2000. Changing patterns of clinical malaria since 1965 among a tea estate population located in the Kenyan highlands. *Transactions of the Royal Society of Tropical Medicine and Hygiene* ***94***: 253-255.

SHARMA V.P. 1996. Re-emergence of malaria in India. *Indian Journal of Medical Research [A]* ***103***: 26-45.

SHARP B.L., Kleinschmidt I., Streat E., Maharaj R., Barnes K.I., Durrheim D.N., Ridl F.C., Morris N., Seocharan I., Kunene S., la Grange J.J., Mthembu J.D., Maartens F., Martin C.L. and Barreto A. 2007. Seven years of regional malaria control collaboration - Mozambique, South Africa, and Swaziland. *American Journal of Tropical Medicine and Hygiene* ***76***: 42-47.

SHARP B.L. and Le Sueur D. 1996. Malaria in South Africa - the past, the present and selected implications for the future. *South African Medical Journal* ***86***: 83-89.

SHARP B.L., Le Sueur D., Ngxongo S., Bredenkamp B.L., and Wilken G.B. 1993a. *Transmission of malaria and vector control in the Mamfene area, Ubombo district, Natal Province (1986-1992)* (Report). Medical Research Council, Parow.

SHARP B.L., Le Sueur D., Wilken G.B., Bredenkamp B.L., Ngxongo S. and Gouws E. 1993b. Assessment of the residual efficacy of lambda-cyhalothrin. 2. A comparison with DDT for the intradomiciliary control of *Anopheles arabiensis* in South Africa. *Journal of the American Mosquito Control Association* ***9***: 414-420.

SHARP B.L., Ngxongo S., Botha M.J., Ridl F.C. and Le Sueur D. 1988. An analysis of 10 years of retrospective malaria data from the KwaZulu areas of Natal. *South African Journal of Science* ***84***: 102-106.

SHARP B.L., Streat E., Kleinschmidt I., Le Grange J.J.P., Mthembu D.J., Kunene S., Mabunda S., Maharaj R., Martin C., Maartens F., Booman M., Dlamini Q., Govere J.M., Hattingh I., Durrheim D.N. and Baretto A. 2003. *Regional malaria control.* Unpublished document.

SLEIGH A.C., Liu X.L., Jackson S., Li P. and Shang L.Y. 1998. Resurgence of vivax malaria in Henan Province, China. *Bulletin of the World Health Organization* ***76***: 265-270.

SMITH T., Charlwood J.D., Takken W., Tanner M. and Spiegelhalter D.J. 1995. Mapping the densities of malaria vectors within a single village. *Acta Tropica* ***59***: 1-18.

SMITH T., Killeen G.F., Maire N., Ross A., Molineaux L., Tediosi F., Hutton G., Utzinger J., Dietz K. and Tanner M. 2006. Mathematical modeling of the impact of malaria vaccines on the clinical epidemiology and natural history of *Plasmodium falciparum* malaria: Overview. *American Journal of Tropical Medicine and Hygiene* ***75***: 1-10.

SNOW R.W., Craig M., Deichmann U. and Marsh K. 1999a. Estimating mortality, morbidity and disability due to malaria among Africa's non-pregnant population. *Bulletin of the World Health Organization* ***77***: 624-640.

SNOW R.W., Craig M.H., Deichmann U. and Le Sueur D. 1999b. A continental risk map for malaria mortality among African children. *Parasitology Today* ***15***: 99-104.

SNOW R.W., Craig M.H., Newton C.R.J.C., and Steketee R.W. 2003. *The public health burden of Plasmodium falciparum malaria in Africa: Deriving the numbers.* (Report). Disease Control Priorities Project, Working Paper No. 11, Fogarty International Center, National Institutes of Health, Bethesda, Maryland (http://www.fic.nih.gov/dcpp).

SNOW R.W., Gouws E., Omumbo J.A., Rapuoda B., Craig M.H., Tanser F.C., Le Sueur D. and Ouma J. 1998. Models to predict the intensity of *Plasmodium falciparum* transmission: applications to the burden of disease in Kenya. *Transactions of the Royal Society of Tropical Medicine and Hygiene* ***92***: 601-606.

SNOW R.W. and Marsh K. 1998. The epidemiology of clinical malaria among African children. *Bulletin of Institute Pasteur* ***96***: 15-23.

SNOW R.W., Marsh K. and Le Sueur D. 1996. The need for maps of transmission intensity to guide malaria control in Africa. *Parasitology Today* ***12***: 455-457.

SNOW R.W., Omumbo J.A., Lowe B., Molyneux C.S., Obiero J.O., Palmer A., Weber M.W., Pinder M., Nahlen B., Obonyo C., Newbold C., Gupta S. and Marsh K. 1997. Relation between severe malaria morbidity in children and level of *Plasmodium falciparum* transmission in Africa. *The Lancet* ***349***: 1650-1654.

SOME E.S. 1994. Effects and control of highland malaria epidemic in Uasin Gishu district, Kenya. *East African Medical Journal* ***71***: 2-8.

SPIEGELHALTER D.J., Best N.G., Carlin B.P. and Linde A.V.D. 2002. Bayesian measures of model complexity and fit. *Journal of the Royal Statistical Society B* ***64***: 583-639.

STEKETEE R.W., Wirima J.J., Bloland P.B., Chilima B., Mermin J.H., Chitsulo L. and Breman J.G. 1996. Impairment of a pregnant woman's acquired ability to limit *Plasmodium falciparum* by infection with human immunodeficiency virus type-1. *American Journal of Tropical Medicine and Hygiene Supplement* ***55***: 42-49.

STUCKENBERG B.R. 1969. Effective temperature as an ecological factor in southern Africa. *Zoologica Africana* ***4***: 145-197.

SWELLENGREBEL N.H. 1931. Malaria investigations in some parts of the Transvaal and Zululand. *South African Institute for Medical Research* ***4***: 245-274.

TANSER F.C., Sharp B. and Le Sueur D. 2003. Potential effect of climate change on malaria transmission in Africa. *The Lancet* ***362***: 1792-1798.

THOMAS C.J., Davies G. and Dunn C.E. 2004. Mixed picture for changes in stable malaria distribution with future climate in Africa. *Trends in Parasitology* ***20***: 216-220.

THOMPSON R., Begtrup K., Cuamba N., Dgedge M., Mendis C., Gamage-Mendis A., Enosse S.M., Barreto J., Sinden R.E. and Hogh B. 1997. The Matola malaria project: a temporal and spatial study of malaria transmission and disease in a suburban area of Maputo, Mozambique. *American Journal of Tropical Medicine and Hygiene* ***57***: 550-559.

THOMSON M.C. and Connor S.J. 2001a. *Malaria early warning systems* (Report). A framework for field research in Africa WHO/CDS/RBM/2001.32, Roll Back Malaria / World Health Organization, Geneva.

THOMSON M.C. and Connor S.J. 2001b. The development of Malaria Early Warning Systems for Africa. *Trends in Parasitology* ***17***: 438-445.

THOMSON M.C., Connor S.J., Milligan P. and Flasse S. 1996. The ecology of malaria - as seen from earth-observation satellites. *Annals of Tropical Medicine and Parasitology* ***90***: 243-264.

THOMSON M.C., Connor S.J., O'Neill K. and Meert J.P. 2000. Environmental information for prediction of epidemics. *Parasitology Today* ***16***: 137-138.

THOMSON M.C., Mason S.J., Phindela T. and Connor S.J. 2005. Use of rainfall and sea surface temperature monitoring for malaria early warning in Botswana. *American Journal of Tropical Medicine and Hygiene* ***73***: 214-221.

TRAPE J.F., Lefebvre-Zante E., Legros F., Druilhe P., Rogier C., Bouganali H. and Salem G. 1993. Malaria morbidity among children exposed to low seasonal transmission in Dakar, Senegal and its implications for malaria control in tropical Africa. *American Journal of Tropical Medicine and Hygiene* ***48***: 748-756.

TRAPE J.F. and Rogier C. 1996. Combating malaria morbidity and mortality by reducing transmission. *Parasitology Today* ***12***: 236-240.

TURPIE J., Winkler H., Spalding-Fecher R., and Midgley G. 2002. *Economic Impacts of Climate Change in South Africa: A Preliminary Analysis of Unmitigated Damage Costs* (Report). Energy and Development Research Centre, University of Cape Town.

VAN DER HOEK W., Konradsen F., Amerasinghe P.H., Perera D., Piyaratne M. and Amerasinghe F.P. 2003. Towards a risk map of malaria for Sri Lanka: the importance of house location relative to vector breeding sites. *International Journal of Epidemiology* ***32***: 280-285.

VAN NIEROP W.H., Frean J.A. and Markus M.B. 1996. *In vitro* susceptibilities of field isolates of *Plasmodium falciparum* to chloroquine in Mpumalanga [letter]. *South African Medical Journal* ***86***: 984-985.

VAN RENSBURG J.J.J. and Blignaut J.N. 2002.The economic impact of an increasing health risk due to global warming. In: Department of Economics, University of Pretoria, Pretoria, pp 117-129.

VERHOEFF F.H., Brabin B.J., Hart C.A., Chimsuku L., Kazembe P. and Broadhead R.L. 1999. Increased prevalence of malaria in HIV-infected pregnant women and its implications for malaria control. *Tropical Medicine and International Health* ***4***: 5-12.

VON SEIDLEIN L., Jawara M., Coleman R., Doherty T., Walraven G. and Targett G. 2001. Parasitaemia and gametocytaemia after treatment with chloroquine, pyrimethamine/sulfadoxine, and pyrimethamine/sulfadoxine combined with artesunate in young Gambians with uncomplicated malaria. *Tropical Medicine and International Health* ***6***: 92-98.

WHITE G.B. 1974. *Anopheles gambiae* complex and disease transmission in Africa. *Transactions of the Royal Society of Tropical Medicine and Hygiene* ***68***: 278-301.

WHITWORTH J., Morgan D., Quigley M., Smith A., Mayanja B., Eotu H., Omoding N., Okongo M., Malamba S. and Ojwiya A. 2000. Effect of HIV-1 and increasing immunosuppression on malaria parasitaemia and clinical episodes in adults in rural Uganda: a cohort study. *The Lancet* ***356***: 1051-1056.

WILKINSON D. and Davies G.R. 1997. The increasing burden of tuberculosis in rural South Africa--impact of the HIV epidemic. *South African Medical Journal* ***87***: 447-450.

WILSON D.B. 1956. *Atlas of Tanzania.* Survey Division, Department of Lands and Surveys, Government Printers, Dar es Salaam.

ZADEH L.A. 1965. Fuzzy sets. *Information and Control* ***8***: 338-353.

Curriculum Vitae

Personal details

Name	Marlies Hildegard Craig (nee Tscheuschner)
Date of birth	2 May 1970
Nationality	South African / German

Education

1991	Bachelor of Science with Zoology and Entomology majors, University of Natal, South Africa
1992	Bachelor of Science Honours in Entomology, University of Natal, South Africa
1994	Master of Science in Parasitology, University of Natal, South Africa
2007	Doctor of Philosophy in Epidemiology, University of Basel, Switzerland
1996	Post-graduate Diploma in Adult Education, University of Natal, South Africa

Other courses attended

2006	4 day course on remote sensing and earth observation, with Klaus Hochheim, PhD from Noetix Research Inc, Ottawa, Canada, at the Medical Research Council, Durban, South Africa.
2004	2 ½ day course on spatial statistics, with Prof Noel Cressie & Prof Jay Ver Hoef, organized by the Swiss Statistical Society.

2003 Advanced epidemiology. Annual course offered by the Public Health and Epidemiology department of the Swiss Tropical Institute, Basel.

2003 Introduction to statistics. Annual course offered by the Biostatistics unit of the Swiss Tropical Institute, Basel.

2004 Key Issues in International Health. Annual course offered by the Public Health and Epidemiology department of the Swiss Tropical Institute, Basel.

2002 One week course on infectious diseases modelling, including training on the modelling software *Stella*, School of Health Systems and Public Health, University of Pretoria, South Africa.

1999 One week advanced *Idrisi* and spatial methods training course, by staff of the Idrisi Project, Medical Research Centre, Durban, South Africa.

1996 One week training in GIS and *Idrisi*, at the Idrisi Project, Clark University, Worcester, USA.

Publications

Craig MH, BL Sharp, MLH Mabaso, I Kleinschmidt. 2007. Developing a spatial-statistical model and map of historical malaria prevalence in Botswana using a staged variable selection procedure. *International Journal of Health Geographics* ***6***:44.

Craig MH, I Kleinschmidt, JB Nawn, D Le Sueur and BL Sharp. 2004. Exploring 30 years of malaria case data in KwaZulu-Natal, South Africa: Part I. The impact of climatic factors. *Tropical Medicine and International Health* ***9***:1247–1257.

Craig MH, I Kleinschmidt, D Le Sueur and BL Sharp. 2004. Exploring 30 years of malaria case data in KwaZulu-Natal, South Africa: Part II. The impact of non-climatic factors. *Tropical Medicine and International Health* ***9***:1258–1266.

Craig MH, Bredenkamp BL, Williams CH, Rossouw EJ, Kelly VJ, Kleinschmidt I, Martineau A, Henry GF. 2002. Field and laboratory comparative evaluation of ten rapid malaria diagnostic tests. *Transactions of the Royal Society of Tropical Medicine and Hygiene* ***96****:* 258-265.

Craig MH, RW Snow and D leSueur. 1999. A climate-based distribution model of malaria transmission in sub-Saharan Africa. *Parasitology Today* ***15****:* 105-111.

Craig MH and BL Sharp. 1997. Comparative evaluation of four techniques for the diagnosis of *Plasmodium falciparum* infections. *Transactions of the Royal Society of Tropical Medicine and Hygiene* ***91****:* 279-282.

Mabaso MLH, MH Craig, A Ross and T Smith. 2007. Environmental predictors of the seasonality of malaria transmission in Africa: the challenge. *American Journal of Tropical Medicine and Hygiene* ***76***: 33-38

Mabaso MLH, MH Craig, P Vounatsou and T Smith. 2005. Towards empirical description of malaria seasonality in southern Africa: the example of Zimbabwe. *Tropical Medicine and International Health* ***10***: 909-918.

Moodley I, I Kleinschmidt, BL Sharp, MH Craig and C Appleton. 2003. Temperature-suitability maps for schistosomiasis in South Africa. *Annals of Tropical Medicine and Parasitology* ***97***:617-27.

Snow RW, MH Craig, CR Newton and RW Steketee. 2003. The public health burden of *Plasmodium falciparum* malaria in africa: deriving the numbers. Working Paper No. 11, Disease Control Priorities Project. Bethesda, Maryland: Fogarty International Center, National Institutes of Health.

Sharp SL, MH Craig, B Curtis, A Mnzava and I Kleinschmidt. 2000. Malaria. In Ntuli A., N. Crisp, E. Clarke, P. Barron (eds), *South African Health Review 2000*, chapter 18. Health Systems Trust, Durban, pp 351-364.

Hay SI, JA Omumbo, MH Craig and RW Snow. 2000. Earth observation, geographic information systems and *Plasmodium falciparum* malaria in sub-Saharan Africa. *Special issue of Advances in Parasitology* ***47***: 175-217.

Kleinschmidt I, G P Clarke, M Bagayoko, M H Craig and D le Sueur. 2000. A spatial statistical approach to malaria mapping. *International Journal of Epidemiology* ***29:***355-361.

Coetzee M , MH Craig and D le Sueur. 2000. Distribution of African malaria mosquitoes belonging to the Anopheles gambiae complex. *Parasitology Today* ***16***:74-77.

Snow RW, MH Craig, U Deichmann and D leSueur. 1999. A continental risk map for malaria mortality among African children. *Parasitology Today* ***15:*** 99-104.

Snow RW, MH Craig U Deichmann and K Marsh. 1999. Estimating mortality, morbidity and disability due to malaria among Africa's non-pregnant population. *Bulletin of the World Health Organization* ***77:*** 624-640.

Cox J, MH Craig, D Le Sueur, and BL Sharp. 1999. Mapping Malaria Risk in the Highlands of Africa. MARA/HIMAL Technical Report.

Snow RW, E Gouws, JA Omumbo, B Rapuoda, MH Craig, FC Tanser, D le Sueur and J Ouma. 1998. Models to predict the intensity of Plasmodium falciparum transmission: applications to the burden of disease in Kenya. *Transactions of the Royal Society of Tropical Medicine and Hygiene* ***92:*** 601-606.

Omumbo JA, J Ouma, B Rapuoda, MH Craig, D le Sueur and RW Snow. 1997. Mapping malaria transmission intensity using geographical information systems (GIS); an example from Kenya. *Annals of Tropical Medicine and Parasitology* ***92****:* 7-21.