

https://dspace.org/

Who dies where, when and why? Modelling determinants and space-time risk of infant, child and adult mortality in rural South Africa, 1992-2008

Sartorius, Benn

2012-05-18
http://hdl.handle.net/10539/11469

Downloaded from a DSpace repository. https://dspace.org/

Who dies where, when and why?

Modelling determinants and space-time risk of infant, child and adult mortality in rural South Africa, 1992-2008

Benn Sartorius

A thesis submitted to the Faculty of Health Sciences, University of the Witwatersrand, Johannesburg, in fulfilment of the requirements for the degree of Doctor of Philosophy (by publication)

Johannesburg, 2011

Candidate declaration

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

---------- ----------

Benn Sartorius Date

Dedication

To my family who know me.

Abstract

Rationale: There is a lack of reliable data in developing settings to inform policy and practice to make best use of direct scarce resources. Health and socio-demographic surveillance systems have the potential to address this information gap. While levels and trends of mortality in rural South Africa have previously been documented, the complex spatial-temporal patterns and risk factors for correlated mortality data have not been fully examined. This will contribute to policy, interventions and programmes.

Aims: To apply various advanced spatial methodologies and Bayesian modelling to longitudinal health and socio-demographic surveillance data in order to: better understand the dynamics of mortality in space-time; identify and correctly quantify risk factors for mortality using intrinsically correlated longitudinal data; and relate disparities in risk factor distributions with spatial mortality risk. Using these findings to: elucidate the public health implications and better inform programme planning and resource allocation.

Methods: The Agincourt health and socio-demographic surveillance system, located in a rural sub-district in northeast South Africa close to the Mozambique border, is based on the continuous demographic monitoring of an entire geographically-defined population. Vital events − deaths, births, in- and out-migrations − are updated annually. The site covers an area in excess of 400km^2 and contains 25 villages, ~14,700 households and ~85,000 individuals. Various simple to more advanced spatial techniques were used to identify significant mortality “hotspots” in space and time. Multivariate Bayesian models were used to assess the effects of the most

significant covariates on age-specific mortality and develop predictive models to enable mapping of the mortality outcome. Spatial correlation was modelled using village-specific random effects which were considered as latent observations of a spatial Gaussian process. Correlation between any pairs of village locations was considered as an exponential function of their distance and modelled by the covariance matrix of the process. Temporal correlation was introduced by yearly random effects modelled via an autoregressive process of various order. The Bayesian framework was used to specify the models and Markov chain Monte Carlo simulation was applied to estimate the parameters.

Key findings: Significant increases in mortality in most age groups were observed, especially from the late 1990's onwards, largely due to the increase in HIV/AIDS and tuberculosis. Results suggest strong geographical and temporal clustering of age-specific mortality in distinct foci, showing that mortality can vary within a small geographical area such as the Agincourt sub-district. The study confirmed several known risk factors and identified additional novel predictors of mortality. Significant differences in the risk factor profiles of the identified mortality "hotspots" included: higher Mozambican concentration, higher proportion of mother deaths, higher household mortality burden and higher mortality of household head, higher infectious disease mortality specifically HIV/TB and diarrhoea, higher proportion of temporary labour migrants with longer duration spent away, lower education and higher poverty.

Implications: The impact of HIV/AIDS on mortality dynamics within this rural setting is striking. Findings indicate the need for interventions to be targeted at higher risk households and villages with respect to both direct and indirect effects of the HIV

epidemic. These high risk clusters also displayed significant differences in risk factor profiles. Risk maps can be used by decision makers for the design and implementation of interventions to alleviate the mortality burden. Interventions that target the mother-infant pair and increase access to a variety of services for more vulnerable “high mortality” households are urgently needed. Important interventions include prevention of mother to child transmission (PMTCT), antiretroviral therapy (ART) rollout, water and sanitation, and screening for and control of non-communicable disease risk factors. Increased distance to nearest health facility emerged as a significant risk factor among adults and highlights the importance of geographical access to ART rollout. The strong spatial clustering of diarrhoeal and malnutrition mortality in children represents a breakdown or absence of basic services, such as provision of water and sanitation, that needs to addressed. Recommendations from this study have implications for other rural settings within South Africa and potentially beyond.

Keywords: age-specific mortality, trends, spatial-temporal risk, determinants, demographic surveillance system, rural, South Africa

Table of contents

	Page No.
Title page	1
Candidate declaration	2
Dedication	3
Abstract	4-6
Table of contents	7-9
List of original papers	10
List of tables	11
List of figures	12-13
List of abbreviations	14-15
1.0 Background	16-31
1.1 Justification of the study	16-23
1.2 Spatial-temporal modelling in the public health sector	23-25
1.3 Mortality estimates and trends: sub-Saharan Africa, South Africa and the Agincourt sub-district	25-29
Summary	30
Overall contribution	31
2.0 Literature review	32-56
2.1 Methods for and applications of spatial analysis for disease and mortality	32-47
2.2 Ethics of mapping and confidentiality	47-48
2.3 Health and Socio-demographic Surveillance System (HDSS), verbal autopsy and geographic information systems (GIS) – strengths and limitations	48-52
2.4 Risk factor for age-specific mortality	52-56
2.5 Thesis framework	57

3.0 Aims and objectives 58-59

4.0 Thesis themes 60

5.0 Data and methods 61-78

5.1 Study area 61-62

5.2 Data collected in the Agincourt Health and Socio-demographic Surveillance System (HDSS) 62-63

5.3 Village household mapping and Geographic Information Systems (GIS) 63-64

5.4 Data quality control 64-65

5.5 Data cleaning and management 65-68

5.6.1 Study population 68

5.6.2 Study period 69

5.7 Data analysis 69-78

5.7.1 Descriptive analysis: demographics, mortality rates and cause-specific fractions (CSF) 69

5.7.2 Temporal trends: descriptive and analytical 69-70

5.7.3 Risk factor analysis using Bayesian geostatistical modelling 70-74

5.7.4 Spatial analyses and risk maps 74-77

5.7.5. Significant differences in determinant profile in high risk clusters 77

5.7.6 Software 78

6.0 Results 79-111

6.1 Descriptive Statistics 79-83

6.2 Temporal trends in age- and cause-specific mortality 83-93

6.3 Major risk factors for age-specific mortality 94-98

6.4 Spatial distribution of age-specific mortality, 1992-2007 98-107

6.5 Space and time distribution of all-cause age-specific mortality 107-108

6.6 Potential proximate reasons for the observed high risk clusters 108-111

7.0 Discussion 112-146

7.1 All-cause mortality 112-113

7.2 Leading causes of death 113-114

7.3 Temporal, spatial and spatial-temporal distribution of mortality 114-124

7.4 Risk factors for mortality in Agincourt Health and Socio-demographic Surveillance System (HDSS) 125-141

7.5. Distribution of underlying risk factors and implications for policy 142-146

7.6 Limitations 146-148

8.0 Conclusions 149-153

Acknowledgements 153-154

References 155-213

Appendix 1 (model formulations and WinBUGS code) 214-224

Appendix 2 (papers) 225-

List of original Papers

The thesis is based on the following papers:

I. Sartorius B, Kahn K, Vounatsou P, Collinson MA and Tollman SM (2010). Space and time clustering of mortality in rural South Africa (Agincourt HDSS), 1992-2007. Global Health Action Supplement 1. DOI: 10.3402/gha.v3i0.5225.

II. Sartorius B, Kahn K, Vounatsou P, Collinson MA and Tollman SM (2010). Young and vulnerable: spatial-temporal trends and risk factors for infant mortality in rural South Africa (Agincourt), 1992-2007. BMC Public Health, 10: 645 DOI: 10.1186/1471-2458-10-546.

III. Sartorius B, Kahn K, Vounatsou P, Collinson MA and Tollman SM (2011). Survived infancy but still vulnerable: Spatial-temporal trends and risk factors for child (1-4) mortality in rural South Africa (Agincourt), 1992-2007. Geospatial Health, 5(2): 285-295.

IV. Sartorius B, Kahn K, Vounatsou P, Collinson MA and Tollman SM (2011). Dying in their prime: spatial-temporal trends and risk factors for adult mortality in the Agincourt rural sub-district, South Africa, 1992-2008. In manuscript.

List of tables

Table 1: List of global and local spatial methods

Table 2: Geocoding of households that dissolved prior to introduction of the Agincourt GIS system, 1992-2002

Table 3: Demographic profile of study sample by age group

Table 4: Top five causes-of-death by age group

Table 5: Trend coefficients and significance for infants and child mortality rates

Table 6: Trend coefficients and significance for leading causes of death among infants and children combined, 1992-2007

Table 7: Mortality trend coefficients, significance and rate differences by 5 year age group among younger adults (15-49 years) and period

Table 8: Trend coefficients and significance for leading causes of death among young adults (15-49), 1992-2007

Table 9: Trend coefficients and significance for all-cause and leading causes of death among older adults (50-64)

Table 10: All-cause bivariate (Stata) and multivariate risk factor analyses for age-specific mortality using Bayesian geostatistical modelling (WinBUGS)

Table 11: Mortality rates (95percent Poisson confidence intervals [a]) per 1,000 person-years by age-group and village

Table 12: Significant clusters of all-cause mortality by age-group using the space-time scan analysis scanning for high mortality rates only

Table 13: Strengths and limitations (in theory and as applied to longitudinal Agincourt HDSS data) of different and increasingly complex spatial analysis methods

Table 13: Comparison of the advantages and disadvantages of various longitudinal survival modelling approaches

List of figures

Figure 1: Maps showing the regional location of the Agincourt Health and socio-Demographic Surveillance Site

Figure 2: Log hazard of mortality by yearly age

Figure 3: Predicted log hazard of death among adults aged 15-49 (linear line represents expected trend in developed setting)

Figure 4: Infant and child mortality rates by year, 1992-2008

Figure 5: Perinatal and neonatal mortality rates by year, 1992-2008

Figure 6: Mortality rate of mothers dying in infants' first year per 1000 infant person-years, 1992-2007

Figure 7: Selected cause-specific mortality rates among infants and children by year, 1993-2007

Figure 8: Adult mortality rates by year, Agincourt sub district, 1992-2008

Figure 9: Selected annual cause-specific mortality rates among younger adults (15-49 years), 1993-2007

Figure 10: Annual older adult (50-64 years) mortality rates, 1992-2008

Figure 11: selected cause-specific mortality rates among older adults (50-64) by year, Agincourt sub district, 1992-2007

Figure 12: Villages with significantly high or low mortality rates based on 95percent Poisson confidence limits, significant Moran's I local spatial autocorrelation and Kulldorff spatial scan clusters of all-cause mortality in infants and children

Figure 13: Villages with significant mortality rates based on 95percent Poisson confidence limits, significant Moran spatial autocorrelation and Kulldorff spatial scan clusters of all-cause mortality in adults (15-49 years)

Figure 14: Villages with significant mortality rates based on 95percent Poisson confidence limit comparison to overall rate

Figure 15: Villages with significant excess mortality based on individual age-group results using Poisson confidence intervals (section 6.4.1)

Figure 16: Maps of all-cause mortality risk by age group within the Agincourt sub district based on baseline models without covariates

Figure 17: Geographic risk of adult mortality adjusting for distance to nearest health facility at prediction point using a univariate Bayesian spatial kriging model that includes distance to health facility as the risk factor (2km band increments with blue being the lowest risk, >6km or high risk areas are represented in red)

List of abbreviations

AIDS	Acquired immune deficiency syndrome
ARI	Acute respiratory infection
ARMA	Autoregressive moving average
ART	Anti-retroviral therapy
CAR	Conditional Autoregressive
CDC	Centers for Disease Control
CD4	Cluster of differentiation 4
CI	Confidence interval
COD	Cause-of-death
DIC	Deviance information criterion
GARMA	Generalised autoregressive moving average
GIS	Geographic Information System
HAART	Highly active anti-retroviral therapy
HDSS	Health and Socio-Demographic Surveillance Systems
HIV	Human Immunodeficiency Virus
HR	Hazard ratio
ICD	International Classification of Diseases
INDEPTH	International Network for the Demographic Evaluation of Populations and Their Health in Developing Countries
Km	Kilometres
MCA	Multiple correspondence analysis
MCMC	Markov Chain Monte Carlo
MDG(s)	Millenium Development Goals

MTCT	Mother to child transmission
PCA	Principal Component's analysis
PMTCT	Prevention of mother to child transmission
PY	Person-years
OR	Odds ratio
ORT	Oral rehydration therapy
RR	Relative risk
SAR	Simultaneous Autoregressive
SES	Socio-economic status
STI	Sexually transmitted infection
SQL	Structured Query Language
SSA	sub-Saharan Africa
TB	Tuberculosis
VA	Verbal autopsy
WHO	World Health Organization

1.0 Background

This thesis deals with the spatial-temporal distributions of mortality in a rural South African setting, identifying determinants that may be driving mortality and finally relating the identified spatial hotspots with their determinant profile to suggest tailored interventions. Given the intrinsic geostatistical and longitudinal nature of the data, sophisticated spatial-temporal and Bayesian methodologies are needed to accurately quantify risk and are used in this thesis. Previous work has not utilised these methodologies which has limitations discussed in detail in subsequent sections.

1.1 Justification of the study

The following section deals with the justification for this study which commences with a discussion of the value of health and socio-demographic surveillance systems (HDSS) in a developing country context where data are often not available. The justification then highlights how little proper spatial analysis and risk factor modelling have been undertaken for HDSS how the growing number of HDSS have contributed to filling the gap in national data, as well as informing policy interventions.

1.1.1 Lack of data and value of Health and socio-Demographic Surveillance Systems (HDSS) mortality data

Countries that monitor mortality and its causes are among those that have made substantial progress in health. Reliable statistics on mortality, its causes and trends are in high demand for assessing the global and regional health situation. Reliable

mortality data are a prerequisite for planning health interventions (Ruzicka and Lopez, 1990), yet such data are often not available or reliable in developing countries, including those in sub-Saharan Africa. (Lopez, 1990; Mathers et al., 2005). In particular, information on Cause-of-death in Sub Saharan Africa (SSA) is either unreliable or nonexistent (Lopez, 1990; Kaufman et al., 1997; Hammer et al., 2006). In this regard, the data collected by infrequent population censuses in this region are not suitable for health policy or for research directed at understanding rapidly evolving health transitions.

Infant and child mortality are linked to maternal health which, in turn, is linked to socio-economic status (SES), health service access and the quality of service provision. The Millenium Development Goal (MDG) 4 focuses on reducing the under-5 mortality by two-thirds between 1990 and 2015. Reliable and timely estimates of childhood mortality are critically important in SSA in order to develop public health policy. In particular, disease-specific mortality rates are useful for detecting the protective effect of a specific intervention (Greenwood, 1999; Cooper et al., 1998). However, in many rural African settings, children do not die in medical facilities and thus denominator data are not recorded in a manner suitable for research purposes (Cooper et al., 1998; Greenwood, 1999). Hospital-based data of child mortality and disease-specific fatality rates, moreover, are often incorrectly extrapolated to interpret child mortality in surrounding communities (Snow et al., 1994). An outcome of these deficiencies, therefore, is that inappropriate policy is developed to direct scarce resources to lower risk communities, as well as an inability to evaluate the effectiveness of intervention programs.

The level of adult mortality is also an important indicator or proxy for the overall assessment of mortality in a population, as well as for evaluating variations in access to health care services and for planning health interventions (Rosero-Bixby, 1991; Timaeus, 2001). Despite the important implications of adult mortality estimates, this information has either been neglected or is biased in many African countries because of unreliable data or inappropriate models (Hill, 1999; United Nations, 2002). Because of these limitations, adult mortality estimates for public health purposes have been compromised by an assumption that the survival disadvantage after childhood is small (Murray et al., 1992). In addition to the socio-economic burden of adult mortality, its impact on the young and elderly survival is also a major issue. The large estimated number of orphans in South Africa (Government of South Africa, 2007; Budlender et al, 2008; UNGASS, 2010) further magnifies the need to assess the impact of adult mortality that precipitates both emotional, as well as financial hardships. The impact of HIV, in particular, has significantly reduced income per capita in 80 percent of all affected households (Collins et al., 2007). Furthermore, the impacts of gender and nationality are important considerations that need to be integrated with morbidity and mortality data in order to understand rapidly evolving adult health transitions (Kahn et al., 2006).

In the absence of accurate and comprehensive registries of vital events for the majority of the region's inhabitants, longitudinal studies of defined population-based cohorts represent the only realistic strategy to fill this void in basic public health information. Health and socio-Demographic Surveillance Systems (HDSS), which monitor the entire population of a defined geographical area, provide longitudinal health data and serve as a basis for health research. HDSS are being increasingly used

to prospectively monitor demographic events, as well as provide accurate age-specific mortality rates. In this regard, a number of countries, including South Africa, have established HDSS to collect and analyze health data in selected sites with the view to filling in the gaps of national data, setting health priorities and for directing policy based on sound longitudinal evidence (INDEPTH Network[1], 2002; Sankoh and Binka, 2004). The number of HDSS has increased over time and at present INDEPTH comprises 42 HDSS in 19 countries (www.indepth-network.org).

Cause-of-death data are a critical input to formulating good public health policy. In most parts of Africa, a verbal autopsy (VA) is the only means to determine the probable cause-of-death and classify cause-specific mortality data (Hammer et al., 2006). A VA is also a valuable tool for assessing longitudinal mortality trends and serves as a platform to test and evaluate health interventions (INDEPTH Network, 2002). It provides a better understanding of chronic diseases, the conditions for which little is known. Non-communicable diseases account for a significant proportion of adult deaths in sub-Saharan Africa (Tollman et al., 2008), yet the empirical bases for public health policies and interventions are essentially absent. Despite the HIV epidemic, non-communicable disease among adults is also rapidly increasing in many developing countries due to ageing and health transitions (Kaufman et al., 1997; Tollman et al., 2008,).

[1] "The International Network for the Demographic Evaluation of Populations and Their Health in Developing Countries (INDEPTH) is a global network of members who conduct longitudinal health and demographic evaluation of populations in low- and middle-income countries (LMICs). INDEPTH aims to strengthen global capacity for Health and Demographic Surveillance Systems (HDSS), and to mount multi-site research to guide health priorities and policies in LMICs, based on up-to-date scientific evidence." http://www.indepth-network.org

1.1.2 Methodological value of spatial-temporal and Bayesian risk factor analysis

Recent advances in data availability and analytic methods have created new opportunities to improve the analysis of disease on a local, national or regional basis (Walter, 2000). The widespread use of geographic information systems (GIS) that are factored into statistical packages, have further encouraged spatial data analysis.With the development of Markov Chain Monte Carlo (MCMC) methods and software such as WinBUGS (a generic Bayesian software platform that has built in tools to account for location data), spatial-temporal Bayesian approaches are being applied to the analysis of many social and health problems in addition to disease mapping and modelling (Best et al., 2005). The reason for this is that it provides a platform for incorporating prior knowledge assumptions (based on objective observable data), adjusting for spatial -temporal correlation and uncertainty (unstructured and structured heterogeneity) in the modelling process, and to model both the observed data and random variables.

The availability of geo-referenced health data, advances in statistical methodology, and developments in geographic information systems (GIS), are increasingly reflected in epidemiological research (Elliot et al., 2000). In particular, the use of spatial-temporal analysis has increased in recent years (Elliot et al., 2000), especially for malaria risk and transmission (Kleinschmidt, 2001; Gemperli, 2003a; Gemperli, 2003b; Gething et al., 2006). Despite the growing applications of spatial methodology

in malaria research, few studies have analysed spatial variation of all-cause and cause-specific mortality, with little or no work on HDSS longitudinal data.

Classical statistical methods for risk factor analysis assume that the outcomes are independent. However longitudinal HDSS data have inherent spatial and temporal characteristics which violate this assumption. Objects in close proximity are often more alike. Consequently, one must include the effects of spatial proximity when performing statistical inference on such processes. Including these spatial effects is important for the efficient estimation of parameters, prediction, and the design of sampling networks (Wikle et al., 2002). Thus, common exposures (measured or unmeasured) may influence mortality similarly in households of the same geographical area, introducing spatial correlation in mortality outcomes. Repeated data are also expected to be correlated in time. Standard statistical methods, however, assume independence of outcome measures, such as mortality, thus ignoring correlation bias for two reasons. Firstly, the standard error of the covariates is underestimated, thereby overestimating the significance of the risk factors. Secondly, estimates of the mortality outcomes are incorrect at the locations where data are lacking. Geostatistical models relax the assumption of independence and assume that spatial correlation is a function of distance between locations. They are highly parameterized models and their full estimation has only become possible in the last decade by formulating them within a Bayesian framework that estimates the parameters using Markov chain Monte Carlo (MCMC) simulation (Diggle et al., 1998). The literature indicates a paucity of spatial analysis, or Bayesian geostatistical risk factor modelling, has been employed to analyse HDSS longitudinal data.

1.1.3 Policy value of spatial-temporal and risk factor analyses

The identification of geographical clusters of high risk mortality is an important policy issue that has received limited attention, especially the ability to rapidly identify individuals, households and villages at elevated risk. This study contributes to other literature that investigates mortality and its risk factors that are important from a public health perspective (Gobalet and Thomas 1996). The study also provides guidance regarding the distribution of health services, and other spatially-targeted interventions for disease control, mortality reduction and resource allocation in rural sub-Saharan Africa.

The mortality gap between wealthier and poorer children is still unacceptably high, or even increasing, in many countries worldwide (Wagstaff, 2000). The targeting of health interventions to poorer individuals and high risk communities is, therefore, important for achieving equity. Successful approaches include subsidized health care and health inputs, improved geographic access to health interventions in poor communities, and social marketing (Victora et al., 2003). This study contributes to the targeting of mortality in Africa by incorporating methodology that explains the spatial-temporal dynamics of adult mortality and its associated risk factors. In particular, Benzler and Sauerborn (1998) recommend that when general population-wide intervention programmes are too expensive to implement, it is necessary to limit such efforts to high risk units where particular adverse health effects are most likely to occur.

Despite the high cost of HIV/AIDS prevention for children (Schwartländer et al 2001; Stover et al., 2002; Stringer et al 2004), the lifetime treatments costs of HIV infected infants are far higher (Giraudon et al., 1999; Adomakoh et al., 2002; Sansom et al., 2006). Effective preventative measures, such as prevention of mother to child transmission (PMTCT), therefore, need to be prioritized and targeted to reduce transmission rates to below 2 percent (European Collaborative Study, 2005; Naver et al., 2006; Newell et al., 2007) and improve child survival. Spatial analysis and mapping is an effective method of depicting the spread or retreat of disease and mortality over space and time (Clarke et al., 1996).

1.2 Spatial temporal modelling in the public health sector

Bayesian geostatistical modelling is both a powerful and statistically robust tool for identifying high prevalence areas in a heterogeneous and imperfectly known environment (Clements et al., 2006). Analysis of spatially indexed data is common in biomedical and epidemiological research due to the effect of geographical location on health-related outcomes. An increasing body of literature on spatial analysis of health outcomes in developing countries, moreover, has been motivated by the availability of geo-referenced data, and by the recent advances in methods and software that can implement such complex models (Walter 2000; Brezger et al., 2005; Rezaeian et al., 2007). Previous studies applying Bayesian and other spatial analyses have demonstrated their policy and decision-making value (Best et al., 2005; Gething et al., 2006). Examples include malaria prevention and control (Kleinschmidt, 2001; Diggle et al., 2002; Gemperli et al. 2003a; 2003b; 2006; Mabaso et al., 2006; Kazembe et al., 2007a; Noor et al., 2009; Yeshiwondim et al., 2009; Gosoniu et al., 2006; 2008; 2010;

Zacarias et al., 2010), control of other neglected tropical or parasitic diseases such as schistosomiasis (bilharzia), onchocerciasis and trachoma (Gyapong et al., 2002; Zhang et al., 2008; Raso et al., 2005a and b; Clements et al., 2006; Clements et al., 2008a; 2008b;Taylor et al., 2009; Baker et al., 2010; Brooker et al., 2010; Clements et al., 2010a; 2010b) and for targeting of mortality interventions (Balk et al., 2003; Adebayo et al., 2004; Balk et al., 2004; Gemperli et al., 2004; Hsu et al., 2004; Kandala 2006; Kazembe et al., 2007a; Kazembe and Namangale 2007b; Becher 2010; Sartorius et al., 2010a; 2010b; 2011).

Social, policy and technological changes influence the types of risk factors to which populations are exposed, typically shifting the major causes of disease and death. It is estimated that non-communicable diseases will cause over 70 percent of all deaths worldwide by 2020, compared with an estimated 15percent of deaths from communicable diseases (Gwatkin et al., 1999). The mapping of geographical variations in the risk of communicable and non-communicable diseases aimed at advancing etiological hypothesis and targeting of interventions, will increase the importance of future geographically driven epidemiological studies. In this regard, GIS data incorporated in spatial or geostatistical techniques, will become increasing important to target interventions. Spatial analysis and GIS can reveal spatial variations and distribution patterns of disease and mortality (all-cause and cause-specific) much more effectively than the tabular form with no spatial structure or proximity built into it (World Health Organization (WHO), 2009). Its efficiency in visualizing the distribution of health-related problems also allows policy makers to target resources more efficiently. Disease maps highlight low- and high risk areas and environmental factors (physical and/or socio-cultural) that contribute to disease causation and

mortality. Furthermore, GIS overlay techniques and relating the distribution of identified risk factors with observed mortality can assist policy makers when tailoring interventions to a given location (Jerrett et al., 2010).

1.3 Mortality estimates and trends: sub-Saharan Africa, South Africa and the Agincourt sub-district

Large reductions in child mortality occurred in low and middle-income countries towards the end of the last century, however, more than 10 million children still die every year in these countries (Ahmad et al, WHO). In particular, infant and childhood deaths remain high the developing world, particularly sub-Saharan Africa (SSA). In 1990, there was a 20-fold difference in infant deaths between sub-Saharan Africa and industrialized countries (180 versus 9 deaths per 1000 live births). In 2000, this difference had increased to 29-fold with mortality rates of 175 and 6 per 1000 children respectively (UNICEF 2001). Furthermore, approximately 420 000 children in 2007 became infected with HIV, mostly through mother-to-child transmission (MTCT) in resource poor settings, particularly SSA (Dabis et al., 2000; Newell, 2001; UNICEF et al., 2008). There is, however, significant heterogeneity within SSA with respect to differential age patterns and trends in childhood mortality rates which have increased since the 1990s (Ahmad OB et al, 2000; WHO, 2002). This trend has mainly been attributed to the effects of the HIV/AIDS epidemic and to the spread of chloroquine-resistant malaria (Müller and Garenne, 1999; Adetunji, 2000; Trapé, 2001). Other, more easily preventable, causes-of-death among children younger than five years include diarrhoea, pneumonia, measles and malnutrition. (Black et al., 2003; Tulloch, 1999).

Increased HIV/AIDS infection, as well as the re-emergence of infectious diseases, especially in sub-Saharan Africa, has been accompanied by a resurgence of drug-resistant strains of malaria and tuberculosis closely associated with HIV/AIDS (Cohen 2000; Morens et al., 2004). Furthermore, some infectious diseases are linked to several types of cancers. Hepatitis B and C, for instance, are risk factors for liver cancer, and the human papilloma virus is a precursor for cervical cancer (Ikeda et al., 1998; Niederau et al., 1998; Walboomers et al., 1999). In addition, countries in SSA are increasingly subject to diseases consequent on changing behavioural patterns including smoking and alcohol consumption. Lifestyle-related diseases, as well as diseases linked to ageing are, therefore, also associated with adult mortality in both developed and developing countries (WHO 2002; 2003).

The emergence of HIV/AIDS has precipitated a substantial decline in life expectancy in many African countries, especially in adults aged between 15 and 60 years. The severity of the HIV/AIDS impact is magnified because this disease targets working adults, who are responsible for the welfare of the young and elderly (Murray et al., 1992) who comprise more than 50 percent of the population of SSA (United Nations 2006). Although estimating these mortality levels has been compromised by a lack of reliable data, adult mortality is higher in SSA than in the rest of the world (Mathers et al. 2006). Using models that take HIV/AIDS into account, Murray et al. (2003) estimate the probability of dying between the ages of 15 and 60 in SSA in 2000 had risen by approximately 0.15 when compared to 1990 (Lopez et al., 2006). In 2001, moreover, adult mortality from communicable and reproductive diseases appears to have increased in SSA with HIV/AIDS accounting for 40percent and 45percent of

adult male and female mortality respectively (Mathers et al. 2006). However, this international picture conceals regional variations, since HIV/AIDS and many tropical diseases, such as malaria, do not affect all African countries to the same extent (Grant and De Cock, 1998; Kalipeni 2000).

In South Africa, HIV has produced a substantial decline in life expectancy in both urban rural settings (Bradshaw et al, 2000; Kahn et al., 2007a). The HIV effect on child mortality may be more noticeable in South Africa due to the lower underlying non-HIV mortality than in other parts of Africa. Increases in mortality in rural South Africa were most prominent in children (0–4 years) and young adults (20–49 years) where increases of two- and fivefold respectively were reported in the 1990's and early 2000's. Kahn et al. (2007a), for instance, showed a prominent increase in child mortality due to HIV over this period from 39/1000 to 77/1000. Similarly, Garrib et al. (2006) found high levels of infant mortality in 2006, 67.5 per 1000 person-years, and estimated HIV/AIDS as the single largest cause-of-death in the under-5 age-group at 41percent. Gender differences in mortality patterns were particularly evident with greater increases in female mortality in most adult age groups (Kahn et al., 2007a). Adult mortality in South Africa also increased during the period 1992 to 2005 due to infectious and parasitic diseases largely driven by HIV/TB (Tollman et al., 2008). Furthermore, age-specific mortality from non-communicable disease also increased significantly in adults who were 30 years and older; the change in younger age-groups was not significant. The prominent increase in all-cause mortality, therefore, has been driven by the large increase in infectious and parasitic disease (HIV) and a modest increase in non-communicable disease (Tollman et al., 2008). However, little spatial

analyses of these patterns have been undertaken, thus, justifying a starting point this study.

Injuries and violent death are also an important cause of adult mortality in Africa, particularly adult males. Violent deaths are mostly related to a specific local or regional context and often neglected given its lower relative health priority status in SSA and especially South Africa (Meel, 2004). Furthermore, the difference between rural and urban areas in terms of living conditions and health care, suggests that adult mortality in rural areas may differ from urban areas (Clifford and Brannon, 1985; Ali et al., 2007). In this context, HDSS, though not representative at national level, contribute to a better understanding of population levels and trends, particularly with regards to adult mortality, while also providing informative data on cause-of-death (Kaufman et al. 1997; Garenne and Cantrelle, 1997; Ngom et al., 2001; Pison 2005).

Studies in several developing countries have shown high rates of premature mortality in adults (WHO 2003; Lopez et al., 2006). The risk of a 15-year-old person dying before reaching 60 years of age is 25percent for men and 22percent for women in developing countries, more than double than that in the developed world, where the respective figures are 12percent and 5percent (Murray et al., 1992). A contrast between low-mortality developing countries such as China (with more than one-sixth of the world's population) and high mortality countries in Africa (with one-tenth of the global population) illustrates the extreme diversity in health conditions among developing countries. Less than 10percent of deaths in China occur below 5 years of age compared with 40percent in Africa. Conversely, 48percent of deaths in China occur beyond age 70, compared with only 10percent in Africa (Lee, 2003). In addition

to the neglected problem of adult death which requires attention in its own right, the health of adults is essential for the wellbeing of the young and the elderly. Thus, the need to understand the trend and causes of adult mortality to develop national and international policies cannot be underestimated (Kitange, 1996).

Mortality-related analyses in Agincourt, and other HDSS, have largely documented the changing trends and patterns of all-cause and cause-specific mortality by age and sex. These analyses have also quantified certain risk factors such as comparisons between Mozambican immigrants and local South Africans, and variations by gender and socio-economic status using conventional methodologies (Tollman et al., 1999a; Garenne et al., 2000a; Hargreaves et al., 2004; Sankoh et al., 2006; Zwang et al., 2007; Kahn et al., 2007a; Cook et al., 2008; Tollman et al., 2008). However, understanding risk factors and their distribution can play a central role in predicting and preventing premature mortality by guiding policy and targeting interventions. An accurate quantification of previously documented and new risk factors is, therefore, possible using a Bayesian approach.

Summary

In the background and justification, I have shown:

- the dearth of mortality data in developing settings where accurate information regarding spatial-temporal distribution and associated predictors is most needed to inform policy and practice, and to appropriately direct scarce resources
- the, as yet, untapped potential of health and socio-demographic surveillance systems to contribute data on spatial-temporal mortality patterns and related risk or predictive factors
- previous studies that have documented levels and trends for mortality in sub-Saharan Africa, South Africa and the Agincourt sub-district
- that there are sophisticated methods that can be applied to elucidate complex spatial-temporal patterns of mortality, with examples of how this has been applied within the public health sector
- that most previous analyses have not used methods for correlated data when quantifying risk factors, and that these can be correctly quantified using Bayesian geostatistical and temporal models
- that there has been policy value in the past when these or similar techniques have been applied, especially in resource poor settings.

Overall contribution

This thesis will contribute through the testing, refinement and application of various classical and Bayesian spatial-temporal analysis and statistical modelling of risk factors for a very large geostatistical longitudinal data such as an HDSS. The novel application of methodologies within the public health sector will contribute to a better understanding of factors related to mortality and how to better quantify them for correlated geostatistical and longitudinal data. This study contributes to the development of public health interventions by targeting clusters of adverse health outcomes that appear to aggregate geographically over time. This study will also contribute to the tracking and targeting of other emerging (or re-emerging) communicable diseases that are compromising achievements made in developing countries (Sen and Bonita, 2000). In particular, space-time modelling and mapping can be an effective tool for public health authorities and epidemiologists in showing and monitoring diffusion patterns of communicable diseases and in searching for infectious agents. There is, thus, the need to identify disparities in the distribution of mortality and their related risk factors in space and time in order to guide policy interventions and programmes. The methods developed, assessed and used in this thesis, therefore, contribute to our understanding of risk factor modelling of large correlated longitudinal data.

2.0 Literature review

The literature review extends from the history and development of spatial analysis techniques, to their applications in various settings such as health and socio-demographic surveillance systems (HDSS) and risk factors for age-specific mortality. Using the relevant literature, I finally present a broad overall conceptual framework for this thesis.

2.1 Methods and applications of spatial analysis for disease and mortality

2.1.1 Early disease mapping

The earliest disease maps were constructed more than two hundred years ago (Pickle, 2002; Elliot, 2004) when John Snow developed cholera maps to describe the epidemic in London (Snow, 1855). Other early examples include a spot map of yellow fever in 1798 and an unpublished disease map of the world produced in 1792 (Barrett, 2000). Maps of disease rates in different countries started to emerge in the 1800's that characterized the spread and possible causes of outbreaks of infectious diseases such as yellow fever and cholera (Walter, 2000). Most of the first disease maps identified residence location of cases by either a dot or a small bar (McLeod, 2000). Patterns were identified by visual examination and the case locations were compared to those of suspected or known risk factors. A well known early example is that of cholera cases plotted in relation to water pumps in London (Snow, 1855). With time these methods "grew in complexity, sophistication, and utility" (Elliot, 2004).

2.1.2 Development of spatial analysis

Early evidence of manual and then computerized spatial information capturing and analysis were demonstrated in the disciplines of cartography, surveying and geography (Coppock and Rhind, 1991) This evidence, however, was not formalized until the late twentieth century due to a lack of appropriate databases and software (Bailey and Gatrell, 1995; Pickle 2002). Since then spatial analytic techniques have been developed in many disciplines including biology, epidemiology, sociology, demography, statistics, geoinformatics, computer science, mathematics, and scientific modelling (Bailey and Gatrell, 1995). Due to recent advances, the geographic information system (GIS) has become a common software feature that is extensively applied in the modern analytic toolbox and now provides public health researchers and policy makers with an excellent platform to explore the spatial nature of data (Bailey and Gatrell, 1995; Ricketts 2003). Computer science has also been enhanced by the application of algorithms (Mitchell, 1998) that have been complimented by advances in processing capacity. Spatial analysis in the biological sciences includes studies of species distributions (Segurado and Araujo, 2004; Guisan and Thuiller, 2005; Elith et al., 2006), and animal movement (Boveta and Benhamou, 1988; Wua et al., 2000; Patterson et al., 2009; Tang and Bennett, 2010). Other spatial analysis applications include ecological studies of vegetation (Dale 1999), remote sensing imagery in vegetation mapping (Xie 2008), and ecological studies of spatial population dynamics (Tilman and Kareiva, 1997). The potential of remote sensing data for use in the field of epidemiology and the control of tropical diseases was first noted by Cline (1970) and has since been extensively applied in animal and human epidemiological research (Stein et al., 1999; Brooker et al., 2006; Rinaldi et al., 2006;

Hay et al., 2006; Danson et al., 2008). Further developments of spatial analysis include spatial econometrics (Anselin et al., 2004; Arbia and Baltagi, 2009; Anselin, 2010), disease mapping and spread in epidemiology (Basáñez et al., 2004; Pfeiffer et al., 2008), and mortality and health care delivery (Walsh et al.,1997; Klauss et al., 2005). Finally, spatial analysis has been greatly enhanced by new mathematical, modelling and statistical techniques that have contributed to the development of advanced spatial statistics (Whittle, 1954; Vecchia, 1988; Gelfand et al., 1990; Diggle, 1998; Elliot et al., 2000; Walter, 2000; Rue et al., 2002; Wikle et al., 2002; Higdon et al., 2003; Gemperli, 2003b; Banerjee et al., 2004).

Spatial analysis in the early part of the last century was hampered by a lack of appropriate statistical methods, lack of data and software (Pickle, 2002; Bell, 2002). Identification of cases for early maps, dating back to the late 18th and 19th centuries, was made by individual physicians, and rates could not be calculated in the absence of an area-wide population count. When health outcome data became available on a national level in the 1930'and 1940's, statistical methods for their analysis soon followed. The next decade saw the development of key statistical methods to evaluate spatial autocorrelation or clustering (Moran, 1948; Geary, 1954) This formative work was then extended to the detection of disease clustering (Mantel, 1967) and space-time interactions (Knox, 1964). Modern spatial analysis now focuses on computer-based techniques because of the large amount of data, the power of modern statistical and geographic information science (GIS) software, and the complexity of the computational modelling. Kulldorff and Nagarwalla (1995), for example, developed a new method for the detection of spatial and space-time clusters of disease. "The proposed test can detect clusters of any size, located anywhere in the study region. It

is not restricted to clusters that conform to predefined administrative or political borders. The test can be used for spatially aggregated data, as well as when exact geographic co-ordinates are known for each individual" (Kulldorff and Nagarwalla, 1995). A free software to implement this methodology called SaTScan is available at http://www.satscan.org/.

Other statistical methods for spatial analysis also developed in parallel to those in epidemiology. Geostatistical methods such as kriging arose from the need to interpolate and predict in the geologic sciences, for example to produce a surface rendition of soil content (Krasilnikov et al., 2008) or to predict where oil drilling would be successful (Hohm, 1988). These methods were initially for lattice point data (regularly-spaced samples), however, extensions allowed application to irregularly-spaced data (Nielsen, 1994). Prediction models were also developed for small area estimation from national survey data (see Ghosh and Rao 1994; Rao 2003). The goal of small area estimation is to predict responses in non-sampled areas, similar to geostatistics, but the method includes explanatory covariates in the regression model that ignores spatial correlation in the data.

2.1.3 Current methods

Before I review each of the methods used in my thesis in more detail, I first make a theoretical distinction between the two broad spatial statistic measures and further list the methods under their relevant sections in Table 1. These include

a) Global spatial measures or statistics

-A single average value which applies to the entire data set (study area)

-Same pattern or process occurs over the entire geographic area

b) Local spatial measures or statistics

-A unique value calculated for each location (or observational unit)

-Different patterns or processes may occur in different parts of the region

-Geographical approaches to estimating local spatial segregation and local measures of segregation phenomena were suggested in the 1980s (Getis and Ord, 1992) and have developed in parallel with geostatistical approaches (Cressie, 1993).

Table 1: List of global and local spatial methods

Global methods	Local methods
Joint Count Statistic (Goodchild, 1986) Kendall τ (Kendall, 1970)	Local Indicators of Spatial Association (LISA) (Anselin, 1995): Local Moran's I_i Local Geary's *C* Local $G_i(d)$: developed by Ord and Getis (1995) Local $G_i^*(d)$:developed by Ord and Getis (1995) Generalised forms: GLISA (Bao and Henry 1996)
Weighted K-Function (Getis, 1984)	Local K-Function (Getis, 1984)
Global Moran's *I* (Moran, 1948): well known test for autocorrelation	Kulldorff's spatial and space-time scan statistics
Global Geary's *c* (Geary, 1954): well known test for spatial autocorrelation	Bayesian areal (lattice) models --Conditional Autoregressive (CAR) (Besag 1974) --Simultaneous Autoregressive (SAR) (Cressie, 1993)
General G(d) statistic (Getis and Ord ,1992)	Bayesian point pattern (geostatistical) models Bayesian areal and geostatistical models --General Spatio-Temporal Model (Cressie and Huang, 1999; Stein 2005) --Spatio-Temporal Dynamic Models (Wikle 1998, 2003)

The individual methods reviewed below start with simpler techniques and increase in complexity and sophistication. Local methods are used (Moran's I, Kulldorff spatial clustering and Bayesian geostatistical kriging) for all analyses in the publications. This is because we were less interested in whether or not clustering/autocorrelation was occurring globally but rather where the specific all-cause and cause-specific hotspots were within the site.

2.1.3.1 Simple exploratory: areal aggregation, rates

Generally disease or mortality proportions or rates are aggregated by different levels of government administrative unit boundaries or demarcations (e.g. provincial, district and sub-district). The most basic calculation is the crude mortality rate. However, a comparison of crude rates between different villages would essentially be meaningless because more deaths would be expected in villages with an older population structure. An age-adjusted rate scale, therefore, is more meaningful for proper comparison. The actual value of any standardized rate is only meaningful in comparison to other rates that have also been standardized in the same way (Rothman et al., 2008). The two methods most often used to adjust epidemiologic rates are the direct and indirect methods (Fleiss 1981). The advantage of the indirect method is that it may be used for sparsely populated areas which would have age-specific rates too unreliable for the direct method of standardization. However, direct standardization retains the rank order and the proportional differences of the age-specific rates between places. Further limitations of these exploratory approaches include loss of spatial information (particularly important for local public health officials concerned with identifying local "hotspots[2]"), influence of confounders (age), small numbers and hence the reliability of estimates (Pickle 2002; Ghosh et al., 2004; Ghosh et al., 2006; Lawson, 2009).

[2] Location or area with excess mortality

2.1.3.2 Spatial smoothing or filtering

Spatial smoothing or filtering is a non-parametric exploratory spatial analysis technique and has proven useful in disease or mortality mapping (Kafadar, 1996; Talbot et al., 2000; Pickle 2002; Best et al., 2005). The main purpose of these two-dimensional smoothing algorithms (linear or non-linear) is to remove background random noise so that the underlying spatial pattern for a given set of data can be seen. Spatial filtering produces spatial density estimates based on health events that have been observed at individual locations and are a valuable tool for exploring the spatial distribution of cases or deaths in relation to persons at risk. Thus, smoothing methods make use of information from neighbouring points or areas to improve the estimated value for each point. In public health policy, this can be particularly valuable in justifying potential target locations for focusing scarce intervention resources.

The main problem with smoothing methods is how to define spatial neighbours. With temporal smoothing this is simpler as there is a clear ordering of points either side of the given time point. However in a spatial context this becomes more complex as one is now dealing with two dimensions. The choice of spatial neighbours can impact the analysis, especially when areas vary greatly in size and shape (Kafadar, 1996; Talbot et al., 2000; Pickle 2002; Best et al., 2005). Furthermore, none account for possible explanatory variables i.e. they do not include the capacity to adjust for potential confounding variables. Most do not permit inverse variance weights, making them inappropriate for rate and count data except perhaps as a crude first look at the patterns or for areas such as census tracts where population sizes are roughly equal

(Pickle 2002). Thus, their application is limited to producing descriptive maps that must be prepared and interpreted with care.

2.1.3.3 Hypothesis testing of counts: Poisson confidence intervals, spatial autocorrelation and spatial clustering

To simply identify demarcated areas (e.g. government administrative units) or villages (in the case of an HDSS) in which the death rate is significantly above average either overall or by year, an exact 95percent confidence interval (CI) for each (area) rate can be calculated based on the Poisson distribution of the observed number of deaths in a given year or period (Esteve et al., 1994; Sankoh et al., 2001). An area or village rate can be considered significantly above average if the overall rate of the respective year or period is below the lower value of the confidence interval for that area or village rate. This procedure has been commonly used in descriptive epidemiology (Pickle et al., 1987).

The use of spatial-temporal analysis has increasingly been applied in epidemiological research in recent years (Elliot et al., 2000). Advances in data availability and analytic methods have also created new opportunities for investigators to improve on the traditional reporting of disease at national or regional scale by studying variations in disease occurrence rates at a local (small-area) scale (Walter, 2000). Spatial analytical techniques and models, moreover, are often used in epidemiology to identify spatial anomalies or hotspots in disease or mortality regions. These analytical approaches can be used to not only identify the location of such hotspots, but also describe their spatial patterns.

When studying spatially-related objects or characteristics, the first step generally involves describing the regional characteristics that differentiate areas (e.g. villages or households) from each another, and then moves onto the analysis of spatial interrelationships (Douven and Scholten, 1995; Rezaeian et al., 2007). Common spatial techniques used in health research include disease mapping, clustering techniques, the identification of risk factors through map comparisons and regression analysis (Rezaeian et al., 2007). Spatial clustering techniques are important for statistical consideration, and form the initial steps in the development of models for predicting disease risk areas. Disease risk hotspots located close to one another tend to share similar disease or mortality risk factors, because they share similar environments and are also often connected by the spread of communicable diseases via vectors or host dispersal (Waller and Gotway, 2004).

Spatial autocorrelation analysis includes a class of methods which measure spatial dependence, the association between a value at a particular location and values for nearby or adjacent areas. It is useful for finding disease clusters based on area data (Goodchild, 1986; Anselin,1995; Cromley and McLafferty, 2002;). Geary's C (Geary, 1954) and Moran's I (Moran, 1948) are two commonly used methods for areal patterns or spatial autocorrelation. By comparing adjacent area values, they assess the level of large-scale clustering. They have been frequently applied to examine areal clusters, including for example for cancer (Moore and Carpenter, 1999; Mohebbi et al., 2008), other diseases (Fang et al., 2006; Lombardo and Buckeridge , 2007; Chaikaew et al., 2009;), mortality (Uthman et al., 2009; Tottrup et al., 2009; Borden and Cutter, 2008) and other applications such as clustering of traffic accidents, crime and poverty

(Chainey and Ratcliffe, 2005; Holt, 2007; Tsai et al., 2009; Anselin et al., 2008; Erdogan, 2009).

Among the most important exploratory methods for epidemiology and public health are those which identify significant clusters in space and/or time (Ripley, 1977; Diggle, 1983; Alexander and Boyle, 1996; Hjalmars et al., 1994; 1996; 1999, Kulldorff, 1995; 1997; Waller and Gotway, 2004). Spatial, temporal, and space-time scan statistics are now commonly used to detect and evaluate statistically significant spatial clusters in multiple disciplines. The K function introduced by Ripley (1977) is one such method for general clustering in point pattern and is discussed in detail by Diggle (1983) and Waller and Gotway (2004). Another technique which takes this one step further is the kernel intensity function developed by Kelsall and Diggle (1995) which can be used to test for clustering as well as the presence and location of local clusters. Further discussion and applications of this method can be found in (Waller and Gotway, 2004; Wheeler, 2007). A more commonly used clustering technique employs the spatial or space-time scan statistic developed by Kulldorff et al. (1998). This application is embedded in a free and easy to use software called SaTScan which is widely used in an increasing number of applications including epidemiology (Hjalmars et al., 1996, Britton, 1997; Sankoh et al., 2001, , Huang et al., 2007, Wheeler, 2007; Takahashi et al., 2008; Tanser, 2009; Ruiz-Moreno et al., 2010; Pascual, 2010; Becher, 2010) and other research fields and minimizes the problem of multiple statistical tests. SaTScan is also useful for determining which clusters merit further investigation and which clusters are likely to occur by chance alone.

The exploratory analysis of spatial data aims to describe spatial patterns using inferential statistics (for example, to determine whether the occurrence of mortality is random or not), and for the development of hypotheses. However, these techniques do not answer the question as to what may influence the spatial patterns and spatial modelling is better suited to predict mortality rates (e.g. at unsampled locations) (Diggle et al., 1998). These Bayesian geostatistical approaches have the advantage over traditional spatial prediction or interpolation methods through a robust and comprehensive handling of spatial structure (incorporating spatial dependency) and the uncertainty associated with predicted patterns.

2.1.3.4 Spatial modelling (Bayesian kriging)

The identification of clusters described in the section above does not provide any causal explanation for the patterns detected and hence spatial modelling is required. Recent advances in data availability and analytic methods have also created new opportunities for investigators to improve on the traditional reporting of disease at national or regional scale by studying variations in disease occurrence rates at a local (small-area) scale (Walter, 2000). With the advent of health data captured at a fine geographical resolution (as is the case with HDSS data), small area disease mapping studies have become an established technique in epidemiology (Best et al., 2005).

A study by Sankoh et al (2002) demonstrated that the mapping of mortality rates using Bayesian smoothing techniques is a useful graphical supplement to these analytical methods for disease cluster investigations. In this regard, these techniques address the issue of heterogeneity in the population at risk and it is therefore

recommended for use in the explorative mapping of mortality. This and future research efforts in this area will, thus, use Bayesian kriging to produce smooth maps of mortality risk (Gelfand et al., 1999). As mentioned earlier, underlying risk factors (both quantified and un-quantified) drive the spatial and temporal risk clustering observed in this study. Common exposures may influence mortality similarly in households of the same geographical area, introducing spatial correlation in mortality outcomes. Repeated data are also expected to be correlated in time. Standard statistical methods assume independence of outcome measures (e.g. mortality data) and ignoring correlation introduces bias in the analysis. Recent developments indicate Bayesian techniques to be the appropriate methodology for taking account of this spatial and temporal dependence (Wikle and Royle, 2002; Diggle et al., 1998). These risk factor studies in the Agincourt sub-district will employ Bayesian geostatistical models to correctly quantify risk factors for mortality by age group. These spatial models allow one to estimate with accuracy the amount of spatial and non-spatial variation in the data, assess factors associated with spatial correlation and test hypotheses under the presence of spatial correlation.

More complete studies comparing these various methods are needed to provide specific recommendations for general and specific tests of clustering. However, it is clear that clustering tests should not be used for health data unless they account for varying population sizes across areas (Goovaerts and Jacquez, 2004). In addition, some adjustment for multiple comparisons should be made whenever necessary, such as when different sized moving windows are tried. Whenever possible, it seems that use of a Monte Carlo method to compute the significance level of the test is to be preferred over asymptotic results based on questionable assumptions. These tests

provide a useful preliminary evaluation of clusters in a given area and for generating hypotheses; however, this should be followed by further careful investigation to confirm the existence and importance of the identified patterns (Pickle, 2002).

With the development of Markov Chain Monte Carlo (MCMC) methods and software such as WinBUGS, Bayesian spatial modelling approaches are being applied to the analysis of many social and health problems in addition to disease mapping and modelling. The Bayesian approach takes into account not only the raw data but also any prior knowledge that supplements the given data. The prior information may include all possible evidence and results from previous studies and data. These assumptions can be updated when new evidence or information is discovered. Bayesian approaches also assume that neighbouring areas are more likely to be similar than remote areas (Rytkönen, 2004).

Despite the inherent strengths of Bayesian methods, certain criticisms exist. I discuss some of the major issues now one by one. One major concern is the choice of subjective priors (beliefs) which influence the models posterior estimates. The choice of word “subjective” however is not entirely accurate and should rather be referred to as informative. The choice of informative priors which are study specific is often a major issue raised with Bayesian modeling as they are often not deemed transferrable between studies. An informative prior can be problematic if not based on previous “objective data”. In words of Gelman: Prior distributions can be informative while still being constructed from objective data”. Furthermore, given the paucity of this modeling approach using longitudinal demographic surveillance data we have erred on the side of caution and have further elected to use non-informative priors in this

research which further removes any subjective model manipulation by the researcher. A further criticism is the computation complexity of Bayesian methods. This however also applies to conventional classical methods and "are the price we pay for analyzing large and complex data sets" (Gelman, 2008). Furthermore, some argue that the model formulation, fitting and checking in a Bayesian framework is automatic and not necessarily rigorous. Bayesian methods are often scrutinized in the second step (inference), however Gelman argues that the process as a whole is definitely not automatic and the real challenge comes in terms of constructing a model that adequately and objectively depicts reality and rigorously assessing that the fit of the model is correct. As you will see in the methods part of this thesis I have also spent much time ensuring the fit the Bayesian geostatistical models was adequate based on prediction and Bayesian credibility intervals.

The use of GIS based Bayesian spatial modelling has particularly added value in the field of epidemiology. Furthermore, the applications of Bayesian methods to disease mapping, risk assessment and prediction within spatial epidemiological research, are numerous (Besag and Newell, 1991; Bernardinelli et al., 1995; Biggeri et al., 1995). Bayesian modelling techniques can be successfully used in descriptive mapping analyses to produce, for example, maps of posterior means (smoothed incidence values) and maps of posterior probabilities (Best et al.,, 2001; Congdon, 1997; 2003). However, until recently, there were not many software packages suitable for Bayesian analyses, and the building of complex spatial models, therefore, requires specialist programming work (Bland and Altman, 1998). Full Bayesian estimation employing Monte Carlo techniques can be used to predict multi-dimensional disease patterns and to provide more realistic significance levels of statistical tests. Recent major

improvements in computational speed have also facilitated the merging of epidemiology, geostatistics and survey sampling, to provide powerful new methods for the spatial analysis of disease patterns (Ghosh et al., 1998).

2.2 Ethics of mapping and confidentiality

The presentation of sensitive information in maps must protect the subject's confidentiality. One common approach to protect privacy called "computational disclosure control," includes both aggregation of data values in the dataset before analysis, and cell suppression in a table after analysis (Sweeney, 1997; Rudolph et al., 2006). However, even aggregated data may need to be suppressed. One example of aggregation and suppression are the restrictions applied to data retrieved from the Compressed Mortality File (CMF) on Centres for Disease Control (CDC) WONDER on the Web. Counts and rates are suppressed when the single-year count is less than or equal to five for counties with a total population that is less than 100,000. However, when the data is aggregated over three or more years, there is no suppression of small counts even when the count is less than five (http://wonder.cdc.gov).

The increasing use of linked social-spatial and health-spatial data has raised significant concerns regarding the ability to protect the confidentiality of research participants and the potential stigmatisation that may arise if sensitive information were released. Rural areas present an amplified problem in that settlements are fewer, more dispersed and distinct than in urban areas, and higher levels of buffering are thus required to ensure confidentiality and limit disclosure risk (Leah et al., 2005). Presenting information cartographically is a very useful tool for quickly ascertaining

complex spatial patterns visually, yet disclosure risks are associated with this form of presentation (Leah et al., 2005). Increased layers such as borders and roads when displayed on a map add to the security threat.

One approach to protect sensitive individuals or locations is to add a random error to the longitude and latitude before display on a dot map. This jittering of the location is documented for users of the map and the jittering must be sufficient to ensure confidentiality. More traditional is the presentation of spatial statistics in the form of a chloropleth or isopleth maps [3] so that individual locations are never mapped for presentation (Frumkin, 2010).

Confidentiality risks exist when health data are disaggregated at a fine scale (Curtis et al., 2010). Various studies describe geographical masks that protect the confidentiality of health records, when appropriately used, while permitting many important geographically-based analyses (Armstrong et al., 1999; Curtis et al., 2010). They explore transformation-masking methods, aggregation-masking methods, nearest-neighbour masks, and the replacement of geographic identifiers with contextual information of specific interest to the data user.

2.3 Health and socio-Demographic Surveillance Systems (HDSS), verbal autopsy and geographic information systems (GIS) – strengths and limitations

[3] A choropleth map is a thematic map in which areas are shaded or patterned in proportion to the measurement of the statistical variable being displayed on the map, such as population density or per-capita income. In other words it uses geographic units of analysis to aggregate individual case data as smoothed incidence for that area or unit rather than plotting point data.

Many countries have set up sites to collect demographic and health data on selected populations in a longitudinal follow-up (INDEPTH Network, 2002; Sankoh and Binka, 2004). Kahn (2006 and the National Research Council (2002) suggest that the advantages of a HDSS include:

- measurement of and changes in demographic events such as fertility, mortality and cause-specific disease burden
- identification of health transitions
- large sample sizes that enable observation of rare events
- longitudinal follow-up of all members of a community and therefore provide and platform along with the ability to evaluate the impact of interventions
- embedding experimental study designs in a relatively homogeneous community
- determine the sequence of events and therefore assessing causality
- promotion of community participation and strengthening local capacity
- benefit from an established research infrastructure through attracting other scientists, addition of variables to existing protocols at relatively low cost and promotes effective capacity-building with respect to research skills and scientific leadership in developing settings
- increased scientific and policy value the longer the follow-up period continues.

Limitations of HDSS include:

- the fact that one cannot easily generalise findings to a broader area or make inferences due to the restricted geographic focus (due to economic reasons) in many instances (Kahn, 2006; National Research Council, 2002)

- HDSS covering only a small proportion of the total population (usually 10,000–100,000 individuals in most cases) and are not a random sample of the country's population. Therefore, it is not easily possible to derive wider or nationally representative estimates for epidemiological or demographic parameters from these (Hammer et al., 2006)
- sub-group differences being difficult to elucidate, given the relatively homogeneous community and resulting limited community variation;
- HDSS tending to be resource-intensive (funding requirements, design and planning, participation of study subjects, and time)
- mobile migrant sub-populations being difficult to track and follow up
- the representativity of individuals also being compromised the longer they are observed or studied (so called "Hawthorne effect[4]").

A verbal autopsy (VA) is conducted on every death each census and the INDEPTH Network and WHO have developed standardised VA tool across sites (INDEPTH, 2003; WHO, 2007). A VA is an approach used to obtain a probable underlying cause of death by interviewing lay respondents on the signs and symptoms experienced by the deceased before their death. In many HDSS the VA instrument has been improved over time with the addition of questions relating to HIV/AIDS, refining questions for cardiovascular related symptoms as well as the extension of various sections, for example, maternal deaths, lifestyle practices an occupation (Kahn et al., 2007b). The VA tool has also been extensively validated in terms of accuracy and reliability. In Agincourt Health and Socio-demographic Surveillance System (HDSS) this has been done twice thus far (Kahn et al., 2000; Gerritson et al., in manuscript 2011) with the

[4] The "Hawthorne effect" is a form of reactivity whereby subjects modify their behaviour in response to the fact that they are being studied.

latter looking at deaths specifically between 2001-2005 to access sensitivity and specificity with regards to HIV/AIDS in particular.

In both instances the VA has performed well against the gold standard of a hospital doctor diagnosis. Another study in Agincourt HDSS (Fottrell et al., 2010) has also assessed the VA using a standardised computer based probabilistic Bayesian model, namely InterVA[5] (Byass et al., 2003; 2006). As mentioned above general expert physicians have been used to interpret the VA and to assign individual causes-of-death. However, this process is time consuming and not always repeatable. Computer based approaches such as InterVA have the potential to address these issues.

The limitations to VA use include the need for medically trained personnel to assign a cause-of-death, the limited list of causes that can be assigned, and poor sensitivity or specificity for certain causes-of-death (Garenne and Fauveau, 2006). For the latter this is primarily problematic with diseases that have less specific symptoms (HIV in children, malaria in adults, and cancers). Lastly the coding of VA causes has not been systematic or standardised which could be improved by the inclusion of rules for cause-of-death hierarchy (underlying, immediate and contributing causes of death) (Garenne and Fauveau, 2006).

The development and introduction of full Geographic Information System (GIS) databases in most HDSS has strengthened the potential and application of spatial analysis, as well as field work management (Kahn et al., 2007b). The potential of GIS

[5] InterVA is a suite of computer models to facilitate interpreting VA's.

for HDSS analysis has also been effectively demonstrated in other settings such as the African centre (for example: Tanser et al., 2009; Cooke et al., 2010)

2.4 Risk factors for age-specific mortality

The analysis of risk factors for child and adult mortality in developing countries is fundamental for the design of interventions, for monitoring their performance and for achieving the Millenium Development Goals (MDG).

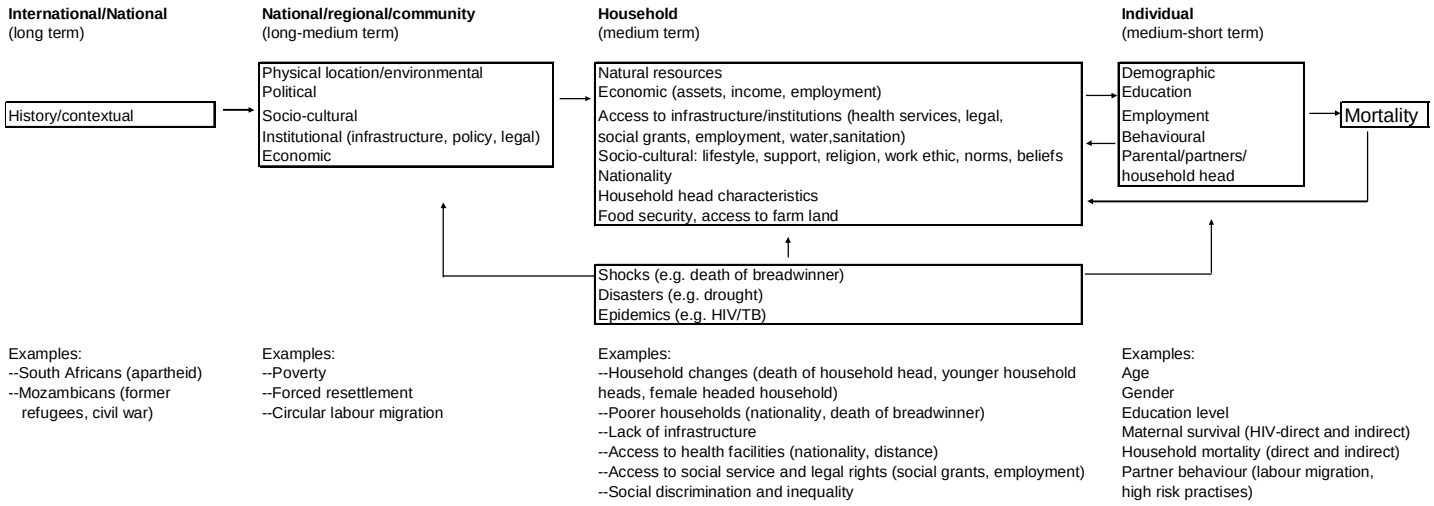
A risk factor analysis in medical geography should be informed by a conceptual that incorporates factors that are most important to the spatial distribution of mortality (and disease) in a given context. According to Oppong et al. (2009) "The disease ecology framework provides an explanation for this uneven geographic distribution of diseases and can be extended to a population level as well (Huynen M et al., 2005). In its most basic form, the disease-ecology framework argues that any disease may be attributable to three sets of factors–genetics, environment, and behavior." The risk factor analysis in this thesis is based on a combined disease–political ecological framework which involves biological, social, economic, behavioural and environmental factors (Mayer, 2000) as well as political and socio-cultural influence (Mayer, 1996) and which I feel suitably encompasses the most important factors (determinants) is this setting. We further construct the framework to take into account the various levels at which determinants operate on mortality. We have included the political component due to the specific historical context in the early to mid-1980s. Many Mozambicans fleeing the civil war settled in the eastern parts of the study site originally in "refugee settlements" that lacked infrastructure. Many of these

Mozambican immigrants elected to stay in South Africa - nearly a third of the current population is of Mozambican origin – and have gradually become more assimilated over time. Nevertheless, they remain a vulnerable sub-group of the population. Significant differentials in health outcomes exist within the Agincourt study population with former Mozambican refugees representing a particularly disadvantaged group (Dolan et al, 1997; Hargreaves et al, 2004; Kahn, 2006). According to Collinson et al. (2006): "In 1993, group refugee status was granted to Mozambicans who had fled the conflict, yet access to water, sanitation, labour markets and legal rights has remained persistently poor for most (Dolan et al, 1995). Nevertheless, uptake to voluntary repatriation programmes has been low (Hargreaves et al, 2004)." The neighbourhood health effect is thus in play in this stetting as Mozambicans settled in generally distinct villages i.e. 'effect' that attributes of local residential environments, or neighbourhoods have on a variety of health and social outcomes (for example: Kawachi and Berkman (eds) 2003; Diez-Roux 2001; Diez-Roux and Mair 2010).

HIV/AIDS and temporary labour migration are other important factors in this conceptual framework. HIV/AIDS has reduced life expectancy in many African countries in recent times, with South Africa and this site severely affected (Bradshaw et al, 2000; Kahn et al., 2007a). The association between temporary migration (often labour related) and HIV infection (and thus mortality) has been shown by several authors in South Africa (Jochelson et al, 1991) (Lurie, 2000.

Based on the context for this specific setting and existing theoretical frameworks, I have attempted to develop an appropriate conceptual framework (see schema below) that illustrates the interrelated nature of key factors in this context.

Conceptual framework for determinants of mortality - an adapted disease-ecology-political model (Mayer, 1996; 2000; Diez-Roux and Mair 2010)



At the individual level, an example biological determinant of infant survival is gender. Typically male infants have a higher risk of mortality in infancy than female infants. Explanations for this gender difference are dominated by biological factors (Waldron, 1998) and form part of core analytical frameworks for assessing infant and child survival (Mosley and Chen, 1984). Maternal death and resulting orphan hood are well established predictors for poor infant and child health and survival (Miller et al., 2007; Watts et al., 2007) both due to the direct and indirect impacts of HIV. The impact of breastfeeding practices on infant survival is complex especially in settings of high HIV prevalence (Coovadia et al., 2007; WHO, 2010). The impact of proximate factors (household), socioeconomic and environmental factors acting directly or indirectly through them, on mortality has been studied for several decades. Studies have

identified individual- and household-level factors as key determinants of infant and child survival (Becher et al., 2004, Hammer et al., 2006). Child mortality in developing countries is mainly associated with measurable socioeconomic conditions such as poor living conditions (Manda, 1999). Poor children are more likely to be exposed to health risks such as unhygienic or unsafe environments, ingestion of unsafe water and lack of access to sanitation (WHO, 2002; Ezzati et al., 2002; Black et al., 2003)), and have less resistance to disease because of malnutrition. These inequities are further compounded by reduced access to health care. Other proximate determinants of child mortality include maternal factors (age, education, health-related behaviours such as pace of childbearing, child spacing, birth intervals, death), nutritional, environmental contamination, injury, and health care access and quality (Black et al., 2003).

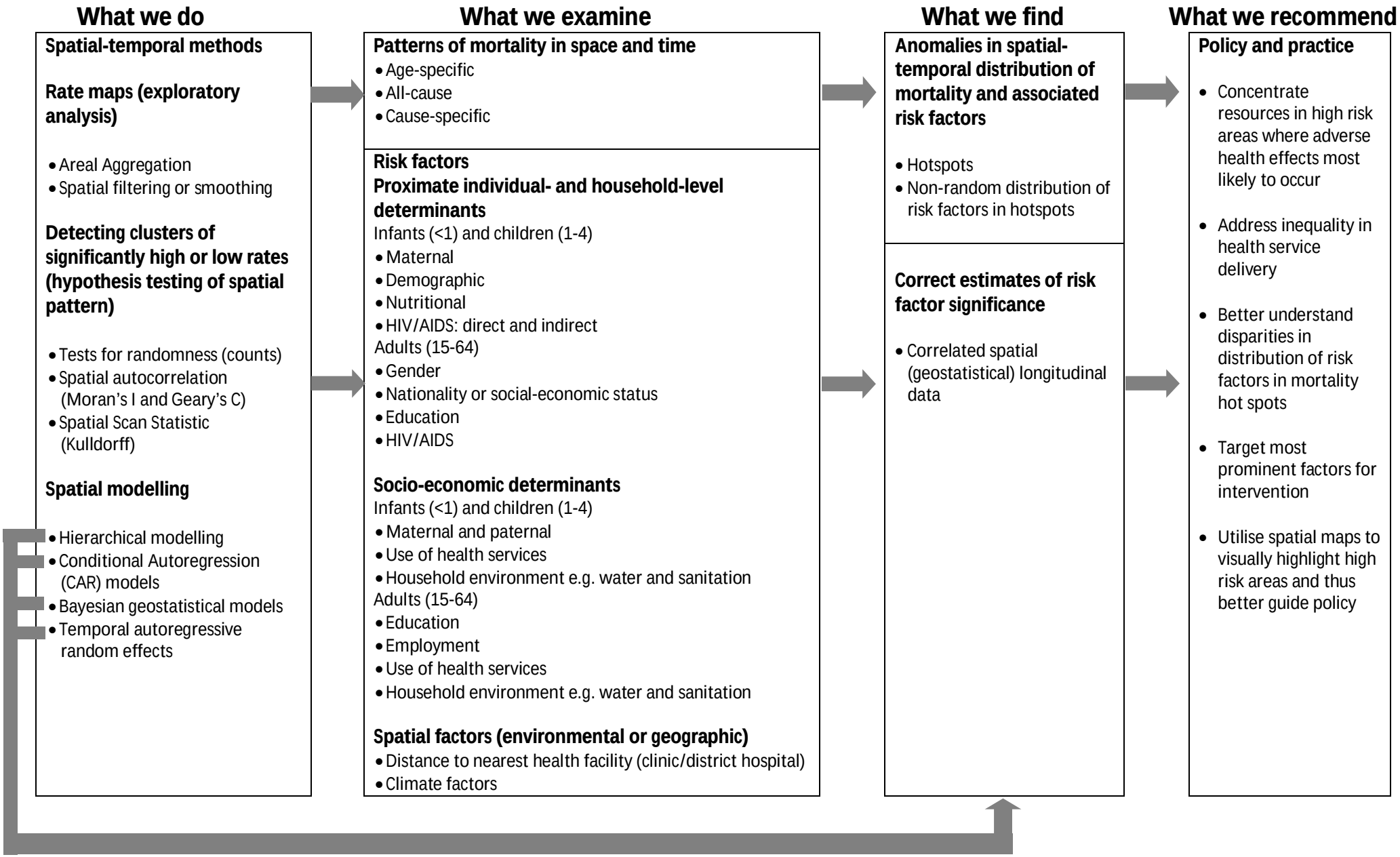
The level of adult mortality is an important indicator for the overall assessment of the mortality pattern in a population, for evaluating access to quality health care services and planning health care interventions (Rosero-Bixby, 1991; Timæus, 2001). The principal focus of global public health efforts over the past several decades has been on improving child health and survival, thus few studies have assessed risk factors for adult mortality in Africa where data are often lacking (Gakidou et al., 2004). Large review studies have indicated the following to be prominent risk factors for adult mortality: high blood pressure and cholesterol, and key behavioural activities: tobacco and alcohol consumption and unsafe sex (WHO, 2002). As many adults circular migrant for labour in this site and engage in high risk activities (Lurie, 2000; Collinson, 2006), behavioural aspects thus form a key component in this framework.

A recent example using data from the Mlomp HDSS in Senegal highlighted certain gender and behavioural differences in adult mortality risk (Duthe and Pison, 2008). Another from the Butajira HDSS in Ethiopia found higher adult mortality associated with low literacy in a household, poor economic status and lack of women's decision making (Fantahun et al., 2008).

Many studies indicate that environmental or geographic factors play an important role in mortality (Balk et al., 2003), including population density (Root, 1997), climate (Patz et al., 2000), disease environment (Root, 1999), and urban residence (Woods, 2003). However, few studies have incorporated potential environmental factors that are explicitly spatial i.e. derived from geographic databases. Spatial variables include simple constructs, such as distances from households or communities (e.g., to nearest clinic or water source) and environmental characteristics that have their own geographic boundaries (e.g., types of farming or land cover).

A number of publications have analysed household-level HDSS data to identify determinants of all-cause (, Becher et al., 2004, Hargreaves et al., 2004; Adazu et al., 2005) and cause-specific mortality like malaria and HIV/TB (Tollman et al., 1999a, Minh et al., 2003; Hammer et al., 2006) and spatial-temporal trends (Delaunay et al., 2001; Kynast-Wolf et al., 2002, Baiden et al., 2006). Furthermore, very few studies have assessed risk factors for adult mortality using HDSS data. None of these studies have employed geostatistical modelling. Instead models for independent data have been used, which is problematic for reasons discussed earlier.

2.5 Thesis framework



3.0 Aims and objectives

3.1 Broad aims

To apply various advanced spatial methodologies and Bayesian modelling to longitudinal health and socio-demographic surveillance data in order to: better understand the dynamics of mortality in space-time; identify and correctly quantify risk factors for mortality using intrinsically correlated longitudinal data; and relate disparities in risk factor distributions with spatial mortality risk.

To elucidate the public health implications that can better inform programme planning and resource allocation.

3.2 Specific objectives

For the rural Agincourt sub-district population from 1992-2008, to:

- elucidate temporal trends in age-specific mortality
- identify significant clusters of age-specific mortality in space-time by applying advanced spatial analysis techniques
- identify and correctly estimate significant risk factors for age-specific mortality using Bayesian inference to analyse large spatially and temporally correlated longitudinal data
- develop smoothed maps of predicted mortality risk for spatially relevant predictors using Bayesian kriging

- assess significant differences between spatial hotpots in the distribution of underlying risk factors for age-specific mortality identified through risk factor modelling.

4.0 Thesis themes

Themes	Integrating narrative	Paper I (all age-groups)	Paper II (infants <1 year)	Paper III (children 1-4 years)	Paper IV (adults 15-64 years)
Spatial-temporal analysis of mortality	Age and cause-specific temporal mortality trends. Comparison of spatial techniques to identify mortality hotspots.	Identify specific villages and clusters of villages with significantly different all-cause or age-specific mortality in space-time.	Bayesian kriging used to produce smooth maps of all-cause and cause-specific infant mortality risk in space alone. Distinct foci identified.	Bayesian kriging used to produce smooth maps of all-cause and cause-specific child mortality risk in space and time. Distinct and emerging foci identified over time.	Bayesian kriging used to produce smooth maps of all-cause and cause-specific adult mortality risk. Bayesian kriging risk map constructed based on distance to nearest health facility; distance found to be a risk factor.
Methods to: **(i) identify and correctly estimate risk factors for correlated mortality data** **(ii) identify differences in risk factor profile between spatial mortality hotpots**	Comparison of risk factors, based on Bayesian models for specific age groups (proximate, socio-economic, spatial factors). Significant differences in risk factor distributions by age group detected between high and lower risk village hotspots or clusters.		Application and testing of Poisson and negative binomial Bayesian random effect risk factor models. Significant classical and novel risk factors emerged.	Application and testing of discrete time (monthly) event history Bayesian random effect risk factor models. Significant classical and novel risk factors emerged.	Application and testing of non-parametric and parametric survival time risk factor models incorporating Bayesian random effects. Significant classical and novel risk factors emerged. Significant differences in risk factor distributions detected between high and lower risk village hotspots or clusters.
Applications for health policy and practice	Excess mortality identified in particular locations, and non-random distribution of underlying risk factors, are policy relevant findings Further research required to better guide policy.	Distinct villages and groups of villages with excess or emerging mortality identified. Require up-scaling of general health and development interventions.	Risk maps can guide policy to vulnerable villages. Strengthen existing health and development interventions and target mother-infant dyad (prevention of vertical transmission or PMTCT[6]; maternal survival). Inequalities in infectious disease mortality risk by nationality needs attention.	Efforts to prevent vertical transmission of HIV (PMTCT) should target high risk villages as should programmes to increase survival of mothers and fathers. Inequalities in infectious disease mortality risk by nationality is an important issue.	Complex interaction of factors driving adult communicable disease mortality: higher HIV and household mortality, longer distance from nearest health facility, lower socio-economic status and education level, higher labour migration. Access to ART for underserved villages is a priority.

[6] Prevention of Mother To Child Transmission (PMTCT) of HIV

5.0 Data and methods

The following section describes the study area and population, the data and its processing and analyses performed in this thesis.

5.1 Study area

The Agincourt Health and Socio-demographic Surveillance System (HDSS) is located in a sub-district in northeast South Africa (Figure 1). It is a poor rural setting and comprises a mix of former Mozambican refugees, migrant workers and a more stable permanent local population. The site was demarcated in 1992 when a baseline census was conducted (Tollman et al., 1999b; Collinson et al., 2002). The key feature of a HDSS is the continuous demographic monitoring of an entire geographically-defined population - in the case of Agincourt HDSS vital events are updated annually. It has since 1993 thus provided a unique longitudinal record of trends in mortality, fertility, in- and out-migration and household composition (Garenne et al., 2000b; Collinson et al., 2006; Kahn et al., 2007a; Garenne et al., 2007;). The site now covers an area in excess of 400km^2 and contains 25 villages, ~14 700 households and ~85 000 individuals.

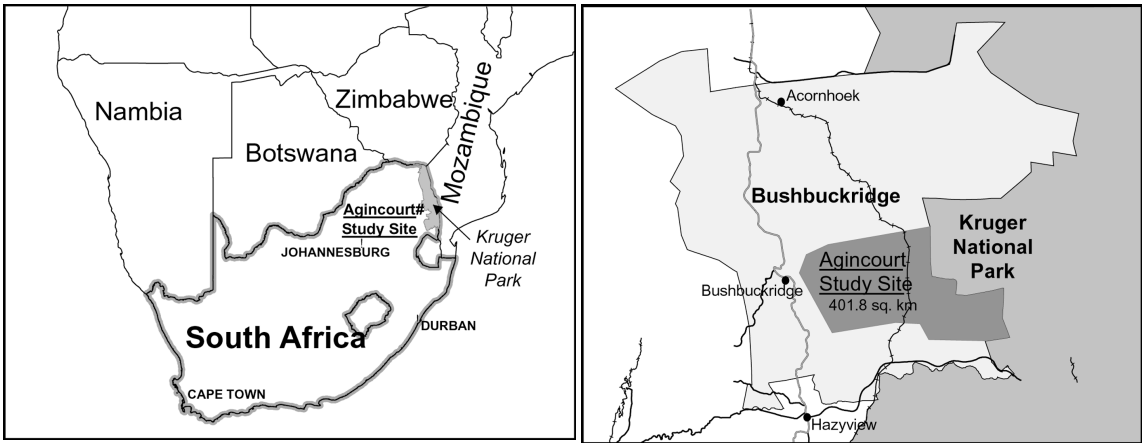


Figure 4: Maps showing the regional location of the Agincourt Health and socio-Demographic Surveillance Site [7]

5.2 Data collected in the Agincourt Health and Socio-demographic Surveillance System (HDSS)

Routine census updates

The baseline census was conducted in 1992 at which time each household was visited and information captured for every resident. The main tool of health and socio-demographic surveillance in Agincourt is a thorough annual update of:

- household memberships
- individual status variables such as relation to household head, nationality, marital-, residence- and education status
- vital events such as pregnancy outcome, death, in- and out-migration and a full maternity history captured for women aged 15-54 years of age. Enquiry into key variables important to each vital event is undertaken.

[7] Kahn K et al. Research into health, population and social transitions in rural South Africa: Data and methods of the Agincourt Health and Demographic Surveillance System. Scandinavian Journal of Public Health, 2007; 35(Suppl 69): 8–20.

Since the baseline, 14 census update rounds have been conducted, the last 8 at strictly annual intervals (1999 to 2009). A verbal autopsy (VA) is conducted on every death to determine its probable cause (Kahn et al., 1999) and has been show to perform well in terms of sensitivity and specificity based on a previous validation study (Kahn et al, 2000).

Additional data collected in Agincourt

Special census modules were introduced in 2000 and provide explanatory variables on topics relevant to understanding and monitoring transitions and thus can be repeated to assess changes over time. Examples include the labour participation module (conducted in 2000, 2004 and 2008), temporary migration module (2002 and 2007), child social grant uptake module (2002, 2005 and 2008), health care utilization (2003, 2006), food security (2004 and 2008) and marital status (2005 and now regularly updated during every census round). A measure of household socio-economic status is gained through an asset survey conducted every 2 years (2001, 2003, 2005, 2007 and 2009).

5.3 Village household mapping and Geographic Information Systems (GIS)

In 1992 hand drawn maps of each village were used; these were updated regularly to incorporate new dwellings and other infrastructural changes. In 1997 village maps were digitised and geo-referenced through an innovative approach that made it possible to construct queries at the household level (le Sueur, 1997). In 2003, a GIS with geo-referencing of every household was introduced. The coordinate data, aerial photographs and HDSS household identifiers were developed into a GIS database.

Maps are updated and edited annually to take account of spatial changes, including the introduction of new dwelling.

5.4 Data quality control

The Agincourt data management system, with a data model as per the standards of the INDEPTH population reference data model, was upgraded in 2002 from Microsoft Access to Microsoft SQL Server. In 2001 the operational database was converted to SQL server, which enabled a higher standard of database technology including data protection and improved database querying and extraction. Overall, these developments ensure a more robust and stable environment for both data volume and complexity.

A custom-made data entry programme, that resembles that data forms for ease of entry, links to the background SQL Server database. Data is captured via simultaneous data entry with multiple networked computers and the system has numerous built-in validation checks. Basic analyses are produced such as village fact sheets, community-feedback information, sampling frames, and denominator information. Further data-cleaning and demographic analyses are conducted to ensure reliable population data.

Data are stored in related tables: the main or so-called ‘Individuals’ table stores key information on all individuals; while the ‘Residence’ table provides information on individual residence episodes, and a’ Membership’ table records information on entry and exit from a particular household should it dissolve. Tables also exist for each vital event category. A range of status observation and special module tables record other

point or cross sectional observations that describes the status of individuals or households at a particular point in time.

To ensure data quality, supervised household visits and random duplicate visits are conducted. Random duplicate visits are conducted by the supervisor for 2percent of the population. From these data, quality can be assessed, and error rates can be computed. Furthermore, form-checking occurs in a structured system at four levels of the field organization. The checks become more detailed as the form progresses through the system. An error is returned to the fieldworker for correction, and, where necessary, a revisit is done.

5.5 Data cleaning and management

5.5.1 Geocoding of households that dissolved from 1992 to 2002

The major limitation is the fact that in the past prior to the development of the full GIS database (in Agincourt this occurred in 2003), many households that dissolved were not recorded in this database but were on old hand drawn paper based maps. However, as will be seen in the methods, this has been reduced to some extent for Agincourt prior to 2003.

Prior to the introduction of the full GIS in 2003, many of the households located on the older paper-based maps that had dissolved prior to 2003 did not have geo-coordinates. In order to bring these household locations into my analytic datasets, queries were run on the database to extract the list of these pre-2003 dissolved

households that had missing coordinates. Old village paper-based maps from 1995-2002 were used to locate these dissolved households and ascertain their position relative to existing households with assigned locations and geo-coordinates. Table 1 below describes the results of this exercise.

Table 2: Geocoding of households that dissolved prior to introduction of the Agincourt GIS system, 1992-2002

Type clean	Total # households	percent
Location digitised	1103	46.82
Linked to existing location	670	28.44
Not located on paper based map	536	22.75
Multiple dissolve	47	1.99
Total	2356	100.00

An SQL script was written to update locations of households that dissolved before 2003 and that were linked to a current existing household identification number.

More than 75percent of coordinates were identified for households that dissolved prior to introduction of the full GIS system in 2003 and that were not previously geocoded. Given the need for spatial coordinates of households in the Bayesian geostatistical analyses (described later), this increased the sample size for the purposes of my PhD from ~87percent to 97percent.

5.5.2 Village boundary extraction and centroid calculation

Using ArcGIS (software housing Agincourt GIS data) we constructed polygons of all village boundaries and extracted vertices. The centroid (point: Cx, Cy) of a non-self-intersecting closed polygon (Burke, 1988), defined by n vertices (x0,y0), (x1,y1), ..., (xn−1,yn−1), was calculated as follows:

$$c_x = \frac{1}{6A} \sum_{i=0}^{N-1} (x_i + x_{i+1})(x_i y_{i+1} - x_{i+1} y_i)$$

$$c_y = \frac{1}{6A} \sum_{i=0}^{N-1} (y_i + y_{i+1})(x_i y_{i+1} - x_{i+1} y_i)$$

5.5.3 Cleaning of VA database

This involved running SQL scripts to identify out-of-domain values and various logical errors (e.g. temporal inconsistencies) in the verbal autopsy (VA) database. The original paper VA forms where then pulled and checked. Cleaning SQL scripts were developed and run to correct values in the main VA database. At the same time, constraints for future data entry and logical checks or triggers (SQL query that cross checks field values during or after entry) were built into the database as well as the VA access front-end to limit future data entry errors. Disparities between values in VA and main Agincourt HDSS database were checked and cleaned.

5.5.4 Checking and hierarchical cleaning of assigned ICD10 codes

With the assistance of Professor Kathleen Kahn, I performed a clean for all ICD-10 cause-of-death (COD) codes. This involved both checking the text description of COD versus the assigned ICD-10 codes, reassessment of selected VA questionnaires as well as a hierarchical restructuring of main, immediate and contributing causes of death where necessary. This exercise was undertaken to improve quality of COD data.

5.5.5 Data extraction using Microsoft SQL

Four different analytical datasets were extracted for the four papers included in this PhD. This was done using Microsoft SQL Server 2005 which houses the Agincourt databases. Since the method of analysis (see statistical models below) was different for the four papers, differently structured datasets has to be constructed and extracted. The mortality clustering dataset comprised age-specific mortality by year and village with precise estimates of person time contributed as denominators (paper I). The infant dataset consisted of person time (in days) contributed by an infant at a given location along with mortality outcome in the first year of life (paper II). An event history (discrete monthly time segments) dataset structure was used for children (1-4 years) until censoring or death (paper III). A continuous survival time approach was used for adult mortality which track time segments at given locations along with mortality outcomes or censoring (paper IV). Relevant COD codes were linked using common individual identifiers in the main Agincourt HDSS and verbal autopsy databases. Time varying covariate datasets were constructed and merged onto core datasets (infant, child and adult) in a time sensitive manner.

5.6.1 Study population

The following age groups were used in thee analyses – namely infants (<1 year), children (1-4 years), young to middle-age adults (15-49 years), and older adults (50-64 years).

5.6.2 Study period

Papers I-III cover the period 1992-2007; Paper IV covers 1992-2008.

5.7 Data analysis

5.7.1 Descriptive analysis: demographics, mortality rates and cause-specific fractions (CSF)

Data on population size, structure, and deaths were extracted from the Agincourt HDSS using Microsoft SQL Server 2005. Data cleaning were done in Stata 10.0. Precise person-years (PY) at risk by age, gender, year, and village were used as the denominator. Observation dates were used for the calculation of person-time as they are the most reliable. Cause-of-death classifications were based on ICD-10 main or underlying cause-of-death code based on VA assessment. Top five causes of death (plus fractions) were listed for each of the specific age groups. Note for infants, death in the perinatal period refers to the first 7 days of life, while neonatal refers to the first 28 days of life.

5.7.2 Temporal trends: descriptive and analytical

Age-specific all-cause and cause-specific mortality trends, using relevant person time denominators, were assessed and depicted graphically. For the age-specific trend analyses significant all-cause and cause-specific mortality trends were assessed univariately (yearly trend term) using the analytical model specific for that age group.

Significance was assessed at the 5percent level. Significant village-specific annual mortality rate trends were analyzed by using a simple Poisson regression model containing person time exposure, a constant and yearly temporal trend term (dos Santos, 1999).

5.7.3 Risk factor analysis using Bayesian geostatistical modelling

5.7.3.1 Outcome variables

Infants – defined as observed mortality within the first 365.25 days of life.
Children - defined as observed mortality between one and 4.999 years of age.
Young to middle-age adults - defined as observed mortality between 15 and 49.999 years of age
Older adults - defined as observed mortality between 50 and 64.999 years of age

5.7.3.2 Explanatory variables

Demographics (gender, nationality), time period, season, maternal factors (former refugee status, age at pregnancy, death of mother during their off-springs infancy or childhood, education) and fertility factors (parity, birth intervals, sibling death), household factors (size mortality experience, household head demographics, socio-economic status based on household assets, food security), health seeking (distance to nearest health facility, antenatal clinic attendance), migration patterns, and household elevation (climatic proxy) were included as covariates.

Socio-economic status (SES) index construction for the risk factor models: socio-economic data collected in the Agincourt HDSS is based on information on living conditions and assets, building materials of main dwelling, water and energy supply, ownership of modern appliances and livestock, and means of transport available (Kahn, 2008). SES quintiles were constructed using a Multiple Correspondence Analysis (MCA). The weights used for the MCA index were those from the first dimension. I used an MCA based index after a detailed analysis of the properties of three measures (other two using absolute asset count and Principal Component Analysis (PCA)), because it is better suited to categorical data (Howe et al., 2008).

5.7.3.3 Dataset structure and modelling approaches

Different dataset structures and thus the corresponding modelling approach were used for the various age groups. For infants, a negative binomial model with an offset of time in days contributed in the first year was used. For children (1-4), a monthly discrete time logistic or event history approach was used that tracked any changes of selected covariates in the given intervals. A monthly time interval was used as it was a better approximation of the risk than using a yearly interval. For the adult models, a non-parametric and parametric survival modelling approaches were adopted that split episodes of time for any relevant changes in selected covariates e.g. change of location or household. For all approaches, standard regression models were fitted to identify the most significant risk factors univariately. All covariates significant at 10percent significance level were entered in the Bayesian geostatistical models. For the child and adult models an unstructured individual level random effect was incorporated using a normal distribution centred on zero and a non-informative

inverse gamma prior for the variance. Temporal correlation was taken into account via autoregressive temporal random effects (see description below). A spatial correlation parameter was incorporated in village-level random effect and modelled by assuming that the random effects are distributed according to a Multivariate Normal distribution with a variance-covariance matrix related to the variogram of the spatial process (Diggle et al, 2002). Markov chain Monte Carlo simulation was used to estimate the model parameters (Gelfand and Smith, 1990). Full details of the various modelling specification along with the WinBUGS codes to implement each are provided in Appendix 2.

5.7.3.4 Model assessment and validation

Deviance Information Criterion (DIC) (Spiegelhalter et al., 2002) was used as the first step in comparison of model fit and the one giving the lowest DIC was chosen. Models were then also validated by fitting the models for a subset of the data and predicting outcomes for the remainder. Credibility intervals were constructed and the model providing the best predictions (along with low DIC) were used as the final model. In Bayesian statistics, a credible interval is a posterior probability interval which is used for interval estimation in contrast to point estimation (confidence intervals). In other words, the credibility interval refers to the distribution of parameter values while a confidence interval pertains to estimates of a single value. All the final multivariate models were also fitted using the conventional form in Stata and full set of diagnostics run to ensure that none of the assumptions were violated.

5.7.3.5 Temporal random effects

To model the temporal random effects various approaches, namely standard autoregressive moving average (ARMA) with priors for the AR(1) and AR(2) processes as defined by Schotman (1994) and Zeller (1996) respectively, as well as a generalised autoregressive moving average (GARMA) approach (Benjamin et al., 2003) were tested and the one providing the best fit for a given model and age group (based on Deviance Information Criterion (DIC) were used in the subsequent analyses. The lower the DIC the better the fit of the model. To further validate particular models we excluded data from last observation year and used the remainder to predicate the outcomes for the excluded year. The predictive accuracy of the developed model was assessed in terms of the percentage of infants with correctly predicted outcome within Bayesian credible intervals of selected probability coverage using R software.

5.7.3.6 Spatial random effect

As mentioned above we assumed that the spatial random effect (w_j) had a multivariate normal distribution, $w_j \sim MVN(\underline{0},\Sigma)$, with variance-covariance matrix Σ expressed as a parametric function of distance between pairs of the village centroids points. We also assume an isotropic stationary spatial process, where $\Sigma mn = \sigma_w^2 \exp(-\varphi d_{mn})$, d_{mn} is the Euclidean distance between villages *m* and *n*, σ_w^2 is the geographical variability known as the sill, φ is a smoothing parameter that controls the rate of correlation decay with increasing distance and measures the range of geographical dependency. We specified φ as a uniform distribution between φ_{min} and φ_{max} (Gelfand et al.,

2003). The range is defined as the minimum distance at which spatial correlation between locations is below 5percent. This distance can be calculated as 3/*u* meters. A non-informative gamma prior was adopted for σ_w^2 with a mean and variance of 0.01.

5.7.4 Spatial analyses and risk maps

5.7.4.1 Exploratory: village rates

We calculated the mortality rates by village and year by dividing the observed number of deaths by the total person-years contributed in village i (i =1,..., 21) at year j (j = 1992,..., 2008). To identify villages in which the mortality rate was significantly above average in time, we constructed exact 95percent confidence intervals (CI) for each rate using the Poisson distribution of the observed number of events i.e. deaths (Este`ve et al., 1994). Village mortality was considered significantly above average for a given year if the overall rate for the given year was below the lower value (α=0.025) of the mortality rate CI for that village (Pickle et al, 1987).

5.7.4.2 Hypothesis testing: Spatial autocorrelation

To further describe spatial mortality patterns as well as age-specific mortality patterns, two common spatial autocorrelation coefficients were calculated: Moran's I, a product-moment coefficient; and Geary's c, a standardized squared-distance measure (Cliff and Ord, 1973; Sokal and Oden, 1978a; 1978b; Cliff and Ord, 1981). Moran's I and Geary's c often lead to similar conclusions, and were reasonably compatible in the present study. I therefore only report the results for Moran's I.

5.7.4.3 Hypothesis testing: Kulldorff clustering technique

In this study, the Kulldorff spatial scan statistic (Kulldorff, 1997) was used to identify space-only clusters of high mortality by age-group in the Agincourt HDSS overall for the entire aggregated period of1992-2008. A circular window is imposed on a map by the statistic and the centre of the circle moves across the study region. This window is centered on each of the possible grid points (village centroids) positioned throughout the study region; the radius of the circle changes continuously between zero and a specified upper limit and is thus flexible both in location and size. The maximum radius size of this window is based upon population at risk. The default maximum-size value, as recommended in the SaTScan user guide (SaTScan), is set to 50% of the total population at risk. In other words, with this setting a reported cluster can contain at most 50% of the total population at risk (Chen et al., 2008). Each of these circles can thus contain a different set and number of neighbouring villages, and each of the circles is a potential cluster of age-specific deaths in the Agincourt study area. However limitations of choice of the scaling parameters do still represent a problem with this approach and are reviewed in great detail in Chen et al (2008). A village is captured in the cluster if it lies within the circle. The spatial scan statistic calculates the likelihood of observing the number of deaths inside and outside each circle, and the one with the maximum likelihood is defined as the most likely cluster i.e. least likely to have occurred by chance (tests the null hypothesis that the risk of dying is the same in all villages in the study area). The spatial scan statistic was also extended into a space-time scan statistic (Hjalmars et al., 1996). The window imposed by the statistic on the study area is cylindrical with a circular geographical base and height

corresponding to time. The centre is again one of several possible village centroids located throughout the Agincourt study area and the height reflects the time interval. The cylindrical window is then moved in space and time. This was applied to the Agincourt HDSS data for the period 1992-2008 (time aggregation of 1 year) to identify high space-time clusters only. The following age groups were used: <5 years, 5-14, 15-49, 50-64, and 65+. Person-time by age group, gender, and village was used as the denominator. To ensure sufficient statistical power, the number of Monte Carlo replications was set to 19,999. The p-value of the statistic is obtained through Monte Carlo hypothesis testing. SaTScan gives the most likely cluster with a corresponding p-value (significance was set at the 5percent level). If other clusters not overlapping with the most likely cluster are identified (secondary, tertiary, etc.), these are also given with their corresponding p-values. Maps showing all significant non-overlapping clusters were constructed in MapInfo Professional 9.5. Larger circles do not represent greater risk clusters but rather contain a larger number of neighbouring villages i.e. extend over larger geographical area.

5.7.4.4 Spatial modelling: Bayesian kriging using a spatially structured covariate namely distance to nearest health facility

Simulation based Bayesian kriging (Gelfand et al., 1999) at regular grid prediction points within the site was used to produce smoothed maps of all-cause and cause-specific adult mortality risk within the entire HDSS site. All-cause and cause-specific baseline models were used that included no covariates except a constant and spatial random effect. Spatial risk of adult mortality adjusting for straight-line distance to nearest health facility at the prediction point (< or >=6km) used a univariate Bayesian

spatial kriging model that included this distance as a covariate. Model estimates were exponentiated to relevant measures of association.

5.7.4.5 Modifications of maps to preserve confidentiality

In this study we removed all geographically identifying features (administrative and village boundaries, village centroids, roads) from the mortality maps that were developed. For all other significant mortality clusters (e.g. cause-specific), tables describing relative cluster location within the site were used to further protect those villages with high HIV burden. This is especially pertinent in a small geographic area such as the Agincourt sub-district where there is high prevalence and stigmatisation of conditions such as HIV/AIDS.

5.7.5. Significant differences in determinant profile in high risk clusters

We classified households and villages contained with significant age specific mortality clusters based on Kulldorff's Spatial Scan Statistic as 1's compared to the remainder of the site 0's. Based on this binary variable I then compared the risk factor profile (based on significant risk factors from the Bayesian multivariate models) of high risk areas to the remainder in an attempt to identify significant differences in the spatial distribution of the determinants.

5.7.6 Software

Data extraction and management was done using Microsoft SQL Server 2005 and STATA 10.0 SE. The analysis was carried out in Microsoft Excel, STATA 10.0 SE, WinBUGS and R. Covariance was assessed using STATA 10.0 SE (using local Moran's I option in the spat* routines). Bayesian models were run in WinBUGS and posterior estimates exported to R for model fit and validation. The risk predictions (kriging) of the fitted spatial models were mapped in MapInfo Professional 9.5.

6.0 Results

In this section, I provide some descriptive statistics for the study population; examine the trend coefficients for the leading causes-of-death of four groups: infants, children, young adults and older adults; examine the major risk factors for each of the four groups; present a spatial distribution of age-specific mortality and finally relate the distribution of risk factors to the identified mortality hotspots.

6.1 Descriptive Statistics

This section commences with a description of the demographic and mortality profile of the study sample including the log hazard of mortality, the leading causes-of-death and trends in age-specific mortality.

6.1.1 Demographic and mortality profile

The demographic and mortality profile of the study samples are provided in Table 3. Overall 9,035 deaths occurred during 1992-2008, based on 1,110,166 person-year time contributed, at an overall crude mortality rate of 8.1 per 1,000 person-years. The highest mortality rates occurred among infants followed by the older adult (50-64) age group (29 and 19 per 1,000 person-years respectively). The mortality rate among children and younger adults (15-49) was similar at 5.7 and 6.9 per 1,000 person-years respectively. Among infants 216 deaths occurred during the perinatal[8] period and 251

[8] Perinatal period: last period of gestation up to first 7 days of life.

in the neonatal[9] period i.e. majority occurred in the perinatal or early neonatal phase. The overall perinatal and neonatal mortality rates were 7.6 and 8.8 per 1,000 person-years respectively. Among adults (15-64) mortality rates showed a steady increase by 5-year grouping with a non-linear excess in 30-34 and 35-39 age groups due to HIV/AIDS.

Table 3: Demographic profile of study sample by age group

Characteristic	Infants	Children (1-4)	Adults (15-49)	Adults (50-64)	Overall [b]
Denominator (person-years)	28,470	116,729	549,030	70,864	1,110,166
Female (percent)	16,030 (50.4)	20,838 (50.3)	99,994 (56.6)	4,062 (59.5)	576,680 (51.9)
South African (percent)	20,382 (64.1)	25,848 (62.3)	66,926 (66.9)	4,292 (62.9)	117,448 (64.1)
Deaths (percent of overall deaths)	826 (9.1)	669 (7.4)	3,798 (42)	1,337 (14.8)	9,035 (100)
Mean age at death (std. dev.)	126.2 days (112.7)	2.1 years (0.9)	34.6 years (8.5)	57.3 years (4.5)	41.8 years (26.8)
Median age at death (IQR)	99 days (16-217)	1.9 years (1.4-2.6)	34.8 years (28.4-41.5)	57.0 years (53.5-61.4)	40.7 years (24.2-63.8)
Mortality rate [a]	29.0	5.7	6.9	18.9	8.1

a: per 1,000 person-years
b: includes 5-14 and 65+ age groups

Hazard of death (Figure 2) increases in the first few years of life then rapidly decreases and reaches its lowest hazard round the age of 8 years. The hazard of mortality starts rising once more round puberty (13 years) and steadily increases thereafter.

[9] Neonatal period: within the first 28 days of life.



Figure 1: Log hazard of mortality by yearly age

While we expect to see a linear increase in the log hazard with increasing age through adulthood (Figure 3), a non-linear pattern is observed to indicating an increased hazard among younger adults due to the HIV epidemic (Figure 2).

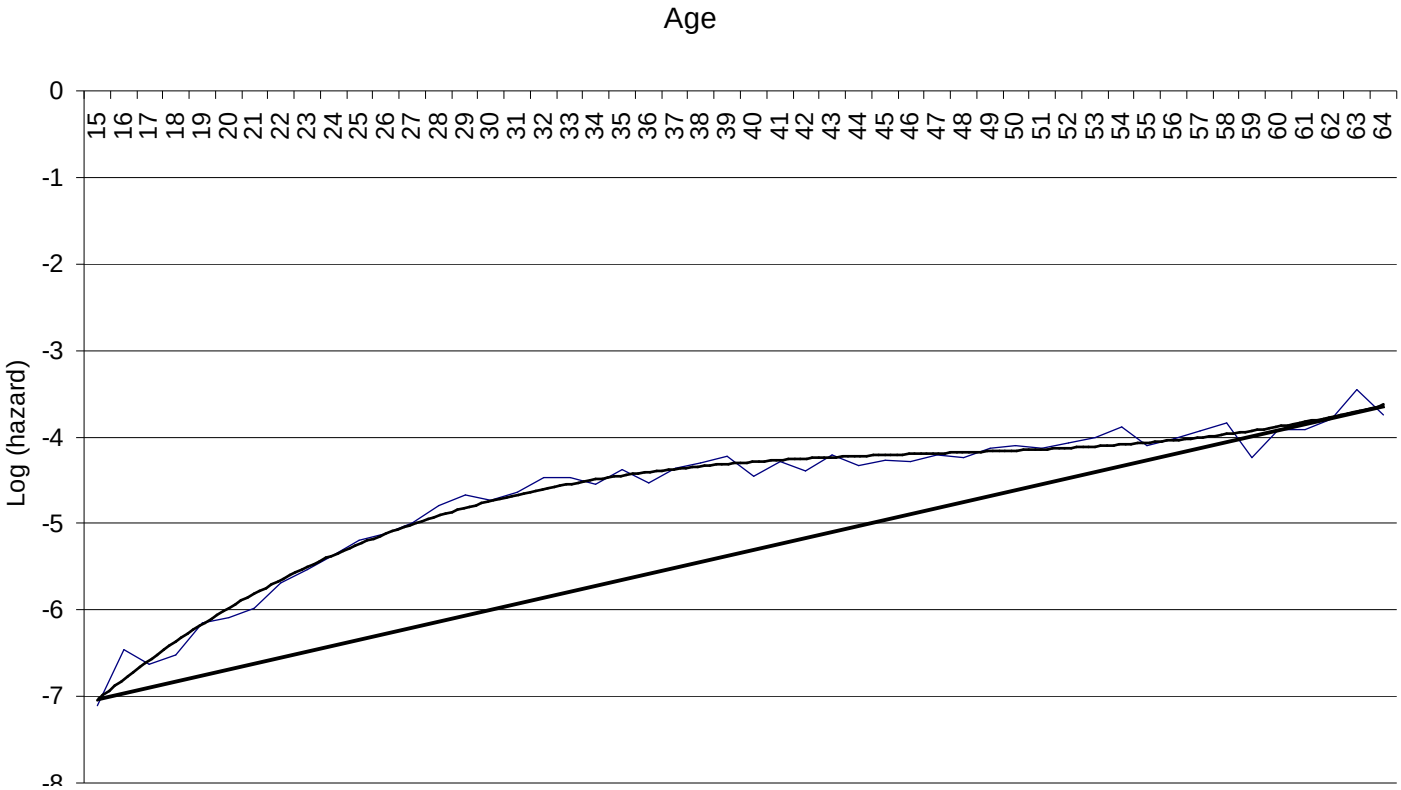


Figure 2: Predicted log hazard of death among adults aged 15-49 (linear line represents expected trend in developed setting)

6.1.2 Leading causes of death by age group, 1992-2007

The leading Cause-of-death in all age groups (Table 4) was HIV/TB. Among children the second most prominent Cause-of-death was diarrhoea or malnutrition. Among younger adults (15-49) external cases of death, namely assault and transport accidents, featured as the second and third top causes-of-death, with lifestyle-related diseases following. In the older adult age group (50-64) following HIV/TB, chronic non-communicable diseases featured prominently.

Table 4: Top five causes-of-death by age group

Rank	Infants (<1) No. (percent)	Children (1-4) No. (percent)	Adults (15-49) No. (percent)	Adults (50-64) No. (percent)
1	HIV/TB 132 (16.6)	HIV/TB 192 (29.5)	HIV/TB 1,545 (43)	HIV/TB 317 (24.7)
2	Diarrhoea or malnutrition 104 (13.1)	Diarrhoea or malnutrition 164 (25.2)	Assault 164 (4.6)	Vascular 159 (12.4)
3	ARI or Pneumonia 103 (13)	External 39 (6)	Transport accident 134 (3.7)	Neoplasms 88 (6.9)
4	Perinatal condition [a] 93 (11.7)	ARI or Pneumonia 27 (4.2)	Vascular 109 (3)	Digestive 54 (4.2)
5	Congenital 25 (3.1)	Congenital 13 (2)	Neoplasms 106 (2.9)	Suicide 35 (2.7)

a: based on ICD-10 main cause-of-death only i.e. P00-P96 (not inclusive of date of birth and date of death timing)

6.2 Temporal trends in age- and cause- specific mortality

Infants and children

A significant increase in the infant and child mortality rates was observed over the study period (Table 5, Figure 4). Infant mortality rates increased from 10.7 to 56.0 per 1000 person-years between 1992 and 2008 while child mortality rates increased from 4.5 to 7.2 per 1000 person-years over the same period.

Table 5: Trend coefficients and significance for infants and child mortality rates

Category	Trend coefficient (β)	p-value
Infants [a]	0.09	<0.001
Perinatal [a]	0.05	0.003
Perinatal [b]	0.07	<0.001
Neonatal [a]	0.06	0.001
Neonatal [b]	0.09	<0.001
Children [a]	0.05	0.001

a: for period 1992-2008 (see Figures 4,5)
b: for period 1996-2008 (see Figure 5)

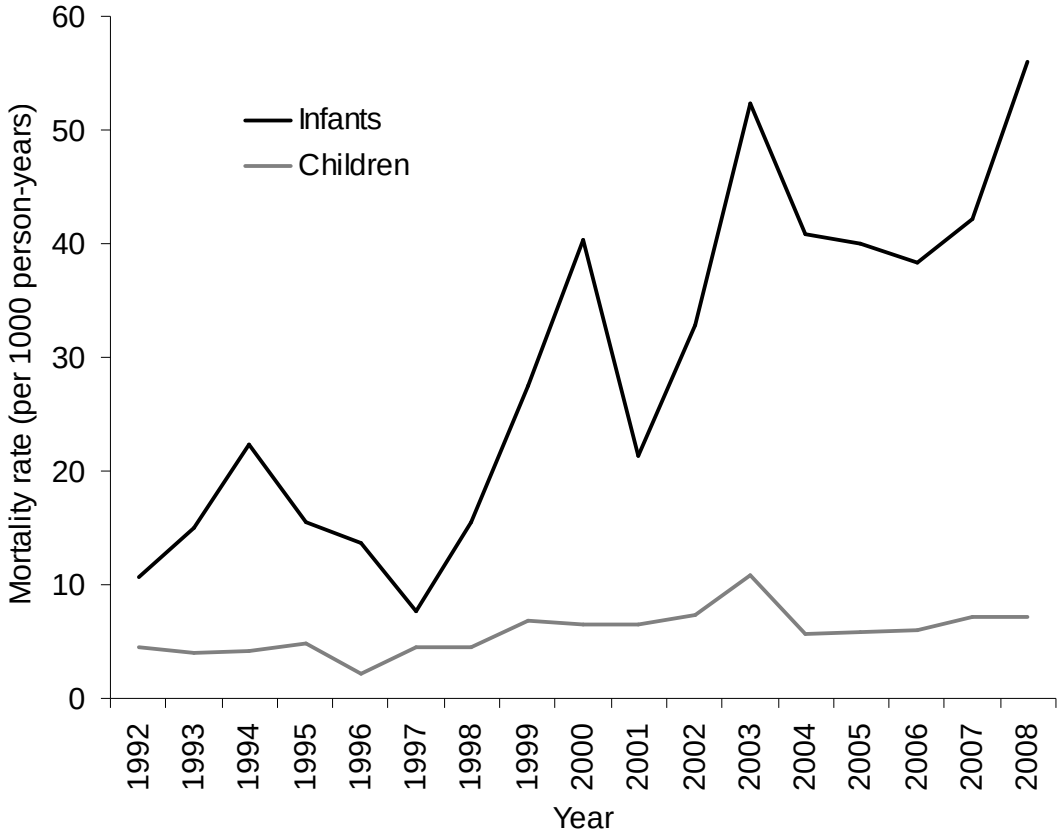


Figure 4: Infant and child mortality rates by year, 1992-2008

The majority of neonatal deaths occurred during the early neonatal period or first 7 days of life. During 1992-1996 a steady decrease in the perinatal and neonatal mortality rate was observed from 10.7 to 2.3 and 10.7 to 2.8 per 1,000 person-years respectively (Figure 5). However from 1997 onwards, concurrent with the rise in infant HIV mortality, a significant increasing trend became apparent with the neonatal and perinatal mortality rates reaching 21.7 and 13.2 per 1,000 person-years respectively.

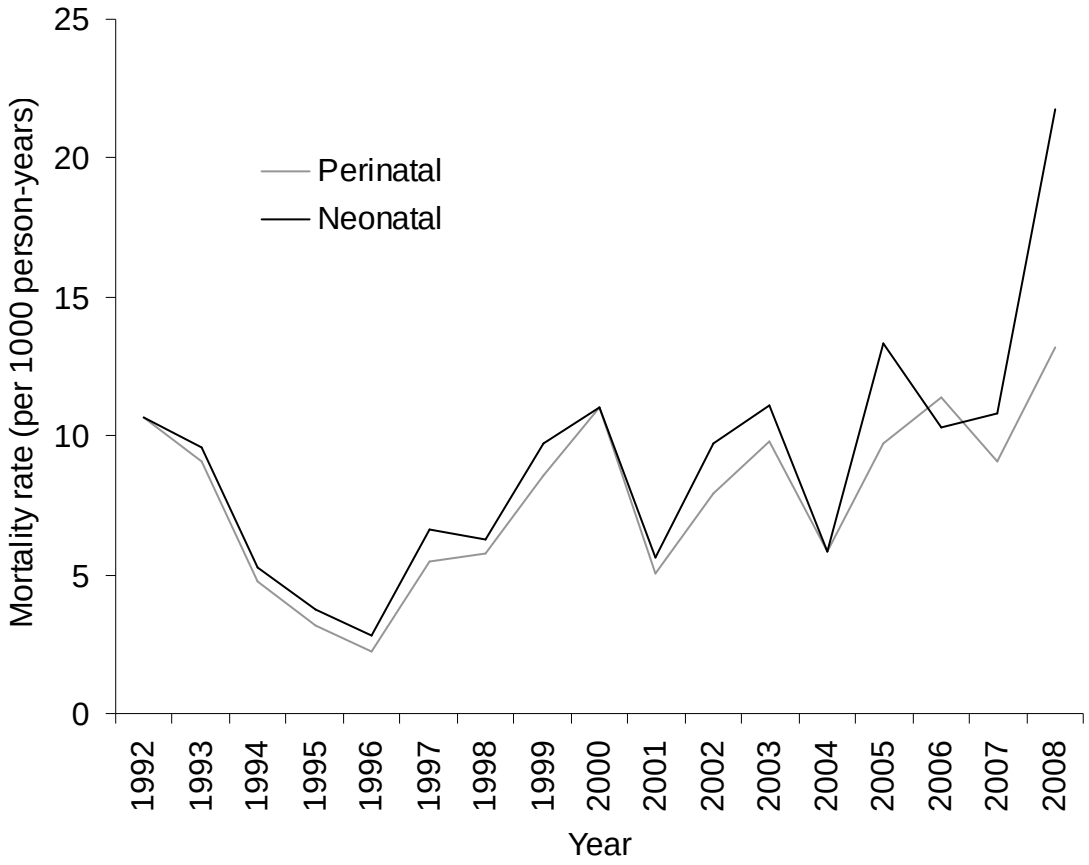


Figure 5: Perinatal and neonatal mortality rates by year, 1992-2008

A significant ($p<0.001$) and striking increase in the mortality rate of mothers dying in the infants' first year was observed (Figure 6), again from 1998 onwards.

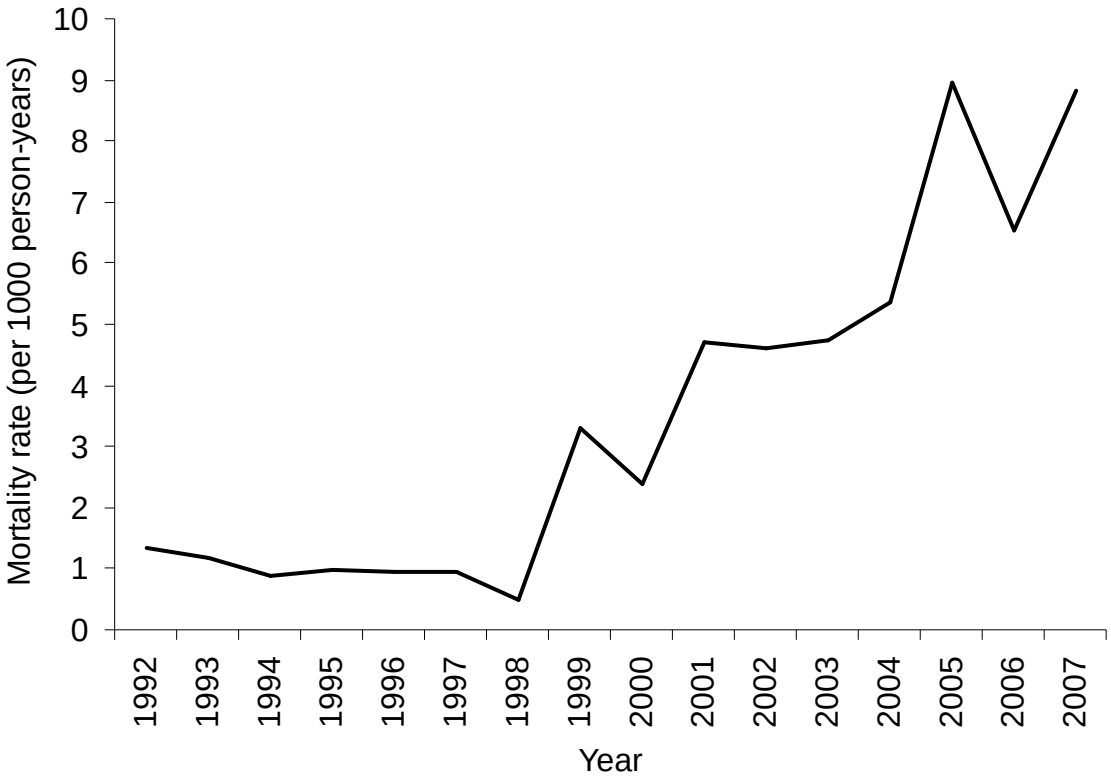


Figure 6: Mortality rate of mothers dying in infants' first year per 1000 infant person-years, 1992-2007

Trend coefficients demonstrate the leading causes-of-death, (Table 6) and show a significant increase in deaths due to HIV/TB and acute respiratory infection (ARI) or pneumonia among infants and children (Figure 7). This trend was more pronounced for HIV/TB. A significant (at 10percent level) increasing trend was also observed for deaths due to unknown causes which is concurrent with the rise in HIV mortality (Figure 7, Table 6).

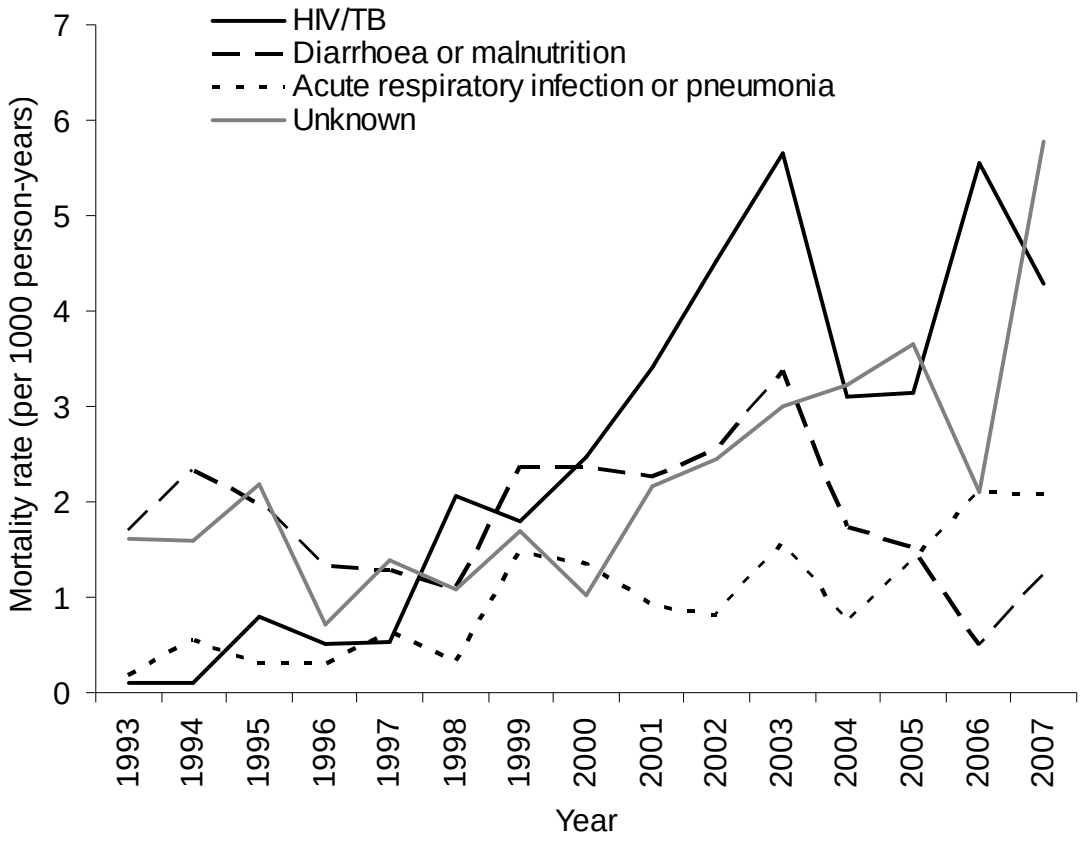


Figure 7: Selected cause-specific mortality rates among infants and children by year, 1993-2007

A significant decrease in deaths due to diarrhoea or malnutrition (Table 6) was observed over the study period. This finding was no longer significant, however, if data for 1992 removed from the trend assessment (β=-0.01, p=0.614). A majority of neonatal deaths occur during the perinatal period or first 7 days of life. A marginally significant increasing trend was observed for congenital deaths between 1992 and 2007. No identifiable trend was observed for external infant and child deaths (Table 6).

Table 6: Trend coefficients and significance for leading causes of death among infants and children combined, 1992-2007

Cause	Trend coefficient (β)	p-value
HIV/TB	0.22	<0.001
Unknown	0.10	0.001
Diarrhoea or malnutrition	-0.05	0.055
ARI or pneumonia	0.12	<0.001
Congenital	0.07	0.073
External	0.05	0.319

Younger adults (15-49 years)

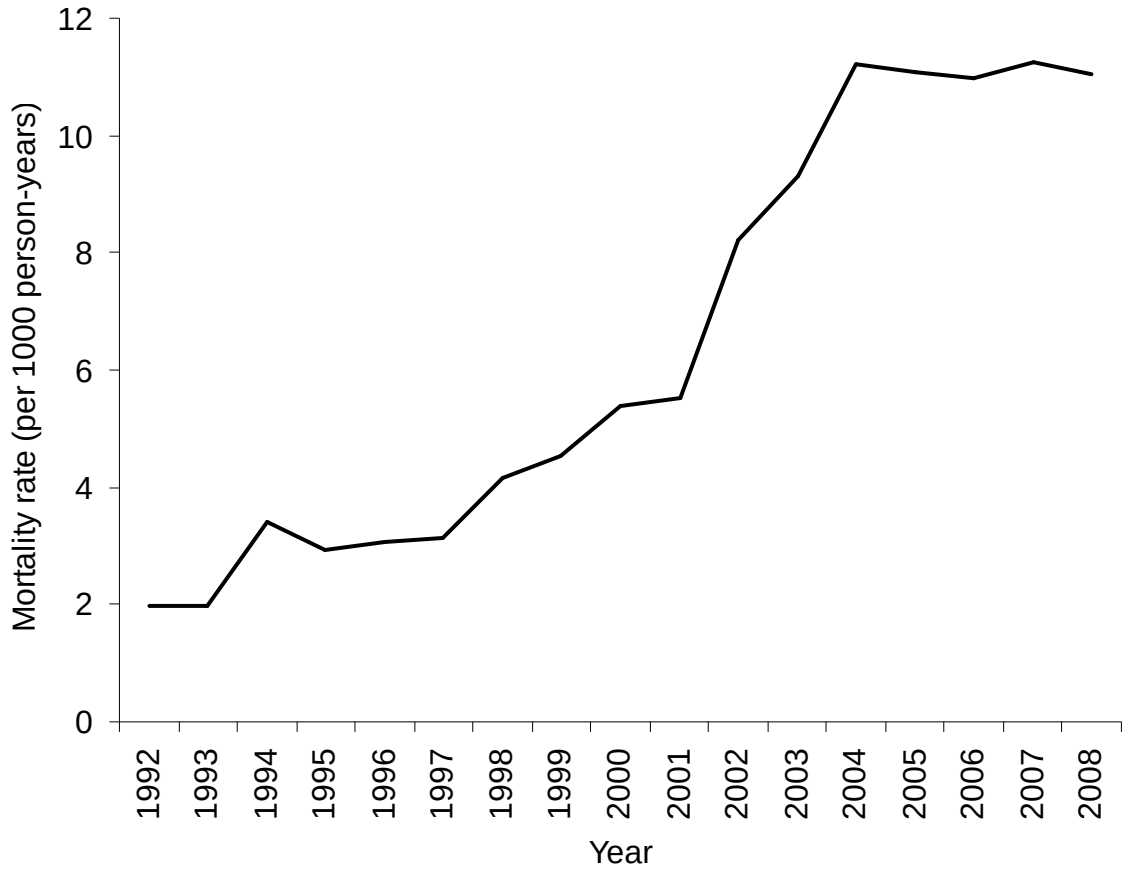


Figure 8: Adult mortality rates by year, Agincourt sub district, 1992-2008

A significant increasing trend for younger adult mortality overall was observed, especially from 1992 to 2004 (Figure 8). Thereafter a levelling out of the mortality rate occurred at approximately 11 deaths per 1000 person-years which was relatively unchanged at 2008. A similar pattern was observed by 5-year age-group. The largest increasing trend coefficients were observed in the 25-29, 30-34 and 35-39 age groups (Table 7). Only the 15-19 age group did not show a significant increasing trend at the 5percent level. The largest absolute rate changes by 5 year age group when comparing the period 1992-1996 to 2004-2008 occurred in the 30-34 and 35-39 age groups (Table 7).

Table 7: Mortality trend coefficients, significance and rate differences by 5 year age group among younger adults (15-49 years) and period

Age group	Trend coefficient (β)	p-value	Rate (1992-1996)	Rate (2004-2008)	Rate change
15-19	0.03	0.073	1.31	1.61	+0.31
20-24	0.10	<0.001	1.33	4.81	+3.48
25-29	0.14	<0.001	2.50	11.77	+9.27
30-34	0.15	<0.001	3.03	18.41	+15.38
35-39	0.13	<0.001	4.85	21.02	+16.17
40-44	0.11	<0.001	5.59	20.41	+14.82
45-49	0.10	<0.001	6.69	21.83	+15.14

A significant and large increasing trend was observed for younger adult deaths due to HIV/TB (leading cause-of-death in this age group) while a significant decreasing trend was observed for deaths due to assault (Table 8, Figure 9).

Table 8: Trend coefficients and significance for leading causes of death among young adults (15-49), 1992-2007

Cause	Trend coefficient (β)	p-value
HIV/TB	0.24	<0.001
Unknown	0.11	<0.001
Assault	-0.04	0.030
Transport accidents	-0.004	0.879
Vascular	-0.02	0.266
Neoplasms	-0.02	0.404

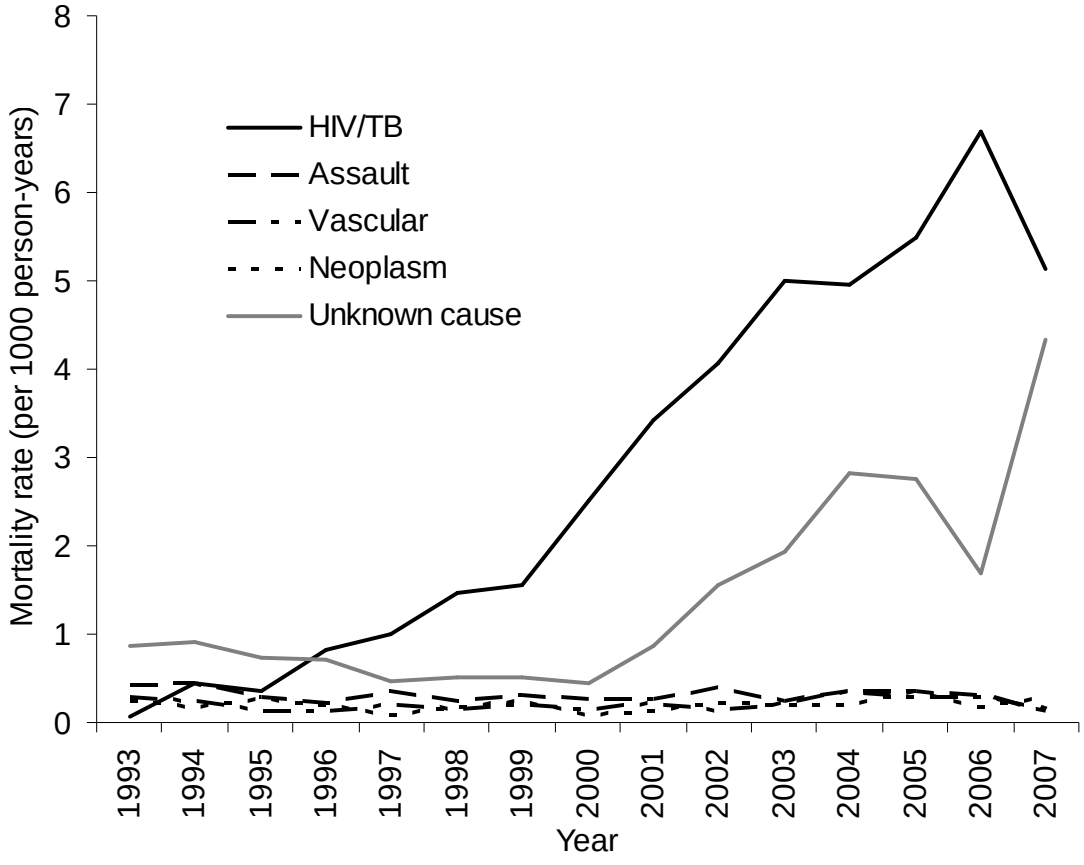


Figure 9: Selected annual cause-specific mortality rates among younger adults (15-49 years), 1993-2007

Concurrent to the significant rise in HIV/TB mortality we also observed a significant increase in deaths due to unknown causes. No significant trend was detected for deaths due to transport accidents, vascular diseases or cancer (neoplasm) over the period 1992-2007 (Table 8).

Older adults (50-64 years)

A significant increase in mortality, (Figure 10), was observed among older adults (50-64 years) over the study period. The rate was relatively stable around 10-15 deaths per 1000 person-years until 2000; from 2000 to 2005 a pronounced rise occurred to a level of approximately 28 deaths per 1000 person-years; it then appeared to stabilize until 2008. This change occurred for all 5 year subgroups of this category although the change was more pronounced for 50-54 and 50-59 year age groups (Figure 10).

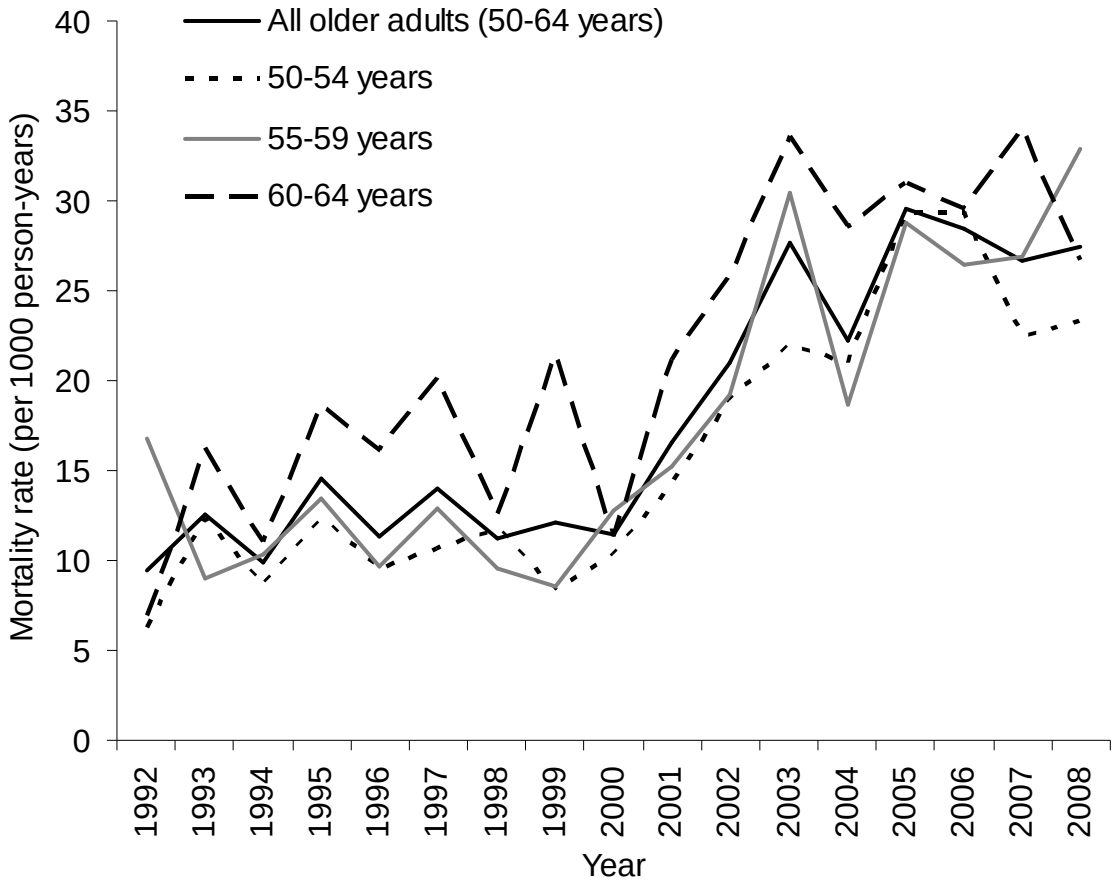


Figure 10: Annual older adult (50-64 years) mortality rates, 1992-2008

The most significant increasing trend was for HIV/TB related deaths (Table 9). A concurrent significant increase in mortality was also observed for unknown causes. A significant increasing trend was identified for deaths due to neoplasms between 1994 and 2007 (Figure 11).

Table 9: Trend coefficients and significance for all-cause and leading causes of death among older adults (50-64)

Cause	Trend coefficient (β)	p-value
All-cause [a]	0.07	<0.001
HIV/TB [b]	0.19	<0.001
Unknown [b]	0.06	<0.001
Vascular [b]	0.01	0.785
Neoplasm [b]	0.07	0.007
Digestive [b]	-0.04	0.185
Suicide [b]	0.02	0.859

a: for the period 1992-2008
b: for the period 1992-2007

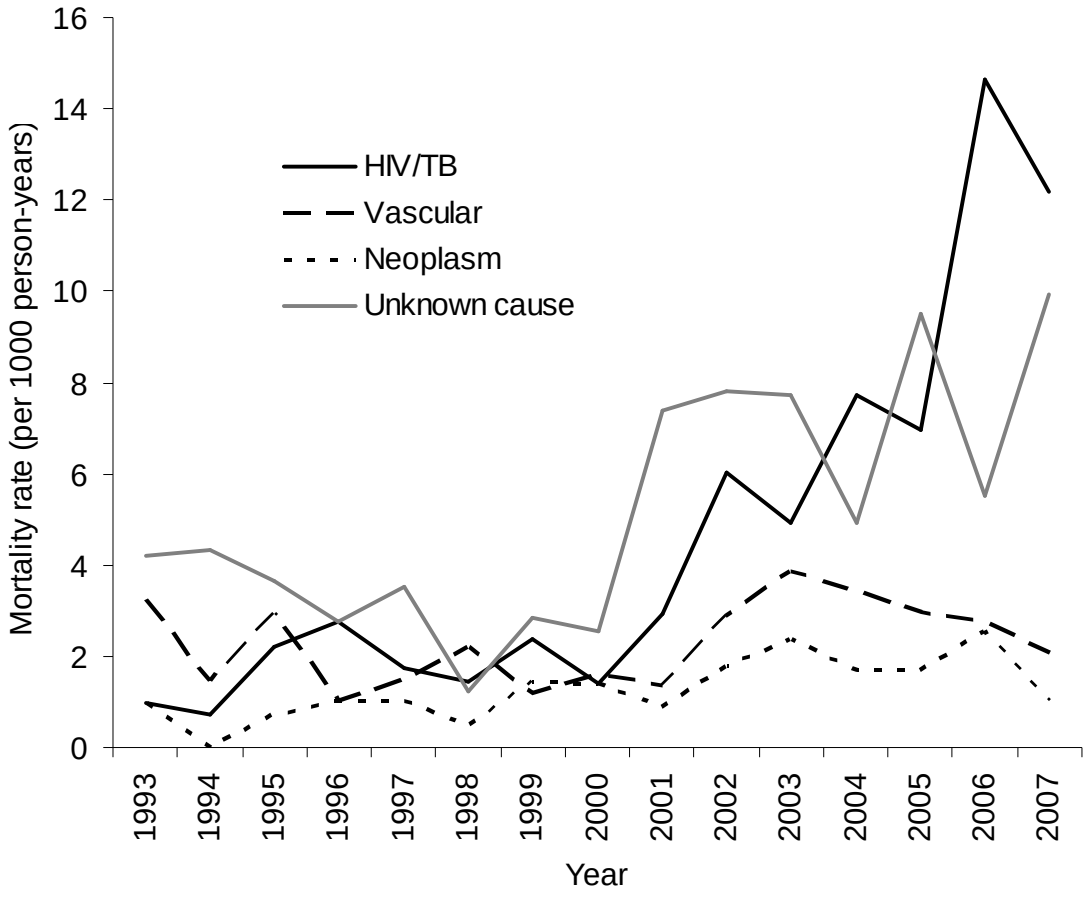


Figure 11: selected cause-specific mortality rates among older adults (50-64) by year, Agincourt sub district, 1992-2007

6.3 Major risk factors for age- specific mortality

This section first discusses the major risk factors influencing age-specific mortality with respect to infants, children and adults. A multivariate model is then developed to illustrate specific relationships.

Infants

Later year of birth, mother dying in the infant's first year (especially due to HIV/TB), higher number of cumulative household deaths, previous birth being stillborn and previous birth interval less than 1 year emerged as highly significant risk factors for all-cause infant mortality (Table 10), especially mother death in the infant's first year and cumulative household deaths. Death of previous child was also a significant risk factor at the 10percent level. Male gender and increasing parity were no longer significant risk factors for infant mortality following multivariate adjustment. No significant association was observed between infant mortality and household socio-economic status, increasing distance to nearest health facility and climate (using elevation as a proxy which corresponds to the rainfall gradient in the sub-district).

Children

Mother death between the child's first and fifth birthdays, particularly due to HIV/TB, was the most prominent risk factor from the multivariate analysis (Table 10), followed by father death due HIV/TB, four or more children aged less than 5 years living within the household, Mozambican origin of the mother, and winter season.

Increasing age of child remained highly protective. As with infants, no significant association was observed between mortality risk and increased distance to nearest health facility. In contrast to infants, however, a significant and increasing trend of protective association was observed with increasing household socio-economic status based on the univariate findings.

Adults

The most prominent risks for 15-49 year mortality following multivariate adjustment were later time period, male gender, being a migrant, increasing number of other household deaths, household head death, and distance to nearest health facility (>6km) (Table 10). Increasing wealth of household, household head being male and older than 40 years were significant and prominent protective factors. No significant difference was observed between Mozambicans and South Africans. Villages with a mortality proportion of HIV/TB above the median value remained at a significantly higher risk.

Table 10: All-cause bivariate (Stata) and multivariate risk factor analyses for age-specific mortality using Bayesian geostatistical modelling (WinBUGS)

Factors	Infants (<1) IRR[a] (95 percent CI)	Children (1-4) OR[a] (95 percent CI)	Young adults (15-49) HR[a] (95 percent CI)	Older adults (50-64) HR[a] (95 percent CI)
Temporal				
• Year	**1.06 (1.03,1.08)[b]**	**1.05 (1.04,1.07)[b]**	**1.15 (1.14,1.16)[b]**	**1.14 (1.12,1.16)[b]**
Proximate individual- and household-level determinants				
• Increasing age in years	n/a	**0.50 (0.46,0.55)[b]**	**1.07 (1.06,1.07)[b]**	**1.04 (1.02-1.06)[b]**
• Winter season	n/a	**1.33 (1.12,1.57)[b]**	---	---
• Male gender	1.03 (0.87,1.21)[b]	1.07 (0.91-1.26)[d]	**1.47 (1.38,1.56)[b]**	**3.03 (2.64-3.48)[b]**
• Mozambican (maternal for infants and children)	1.12(0.70,1.79)[d]	1.12 (0.90,1.38)[b]	0.94 (0.87,1.02)[b]	**0.59 (0.50-0.70)[b]**
• Maternal death (in first year for infants or 1-4 for children)				
Not due to HIV/TB	**6.01 (3.18,11.37)[b]**	**5.45 (2.99,8.75)[b]**	n/a	n/a
Due to HIV/TB	**30.78 (12.13,78.11)[b]**	**15.11 (8.39,24.54)[b]**	n/a	n/a
• Migrant: ≥ 6months outside site per year (maternal status for infants and children)	0.77 (0.591,1.25)[b]	1.15 (0.88,1.50)[d]	**1.15 (1.08,1.23)b**	**1.23 (1.04-1.44)b**
• Paternal death (prior birth to within first year for infants or 1-4 for children)				
Not due to HIV/TB	1.25 (0.03,58.97)[d,e]	1.51 (0.75,2.54)[b]	n/a	n/a
Due to HIV/TB	[e]	**2.19 (0.95,4.06)[b]**	n/a	n/a
• Tertiary education (maternal for infants and children)	0.14 (0.06,0.37)[c]	0.49 (0.23-1.04)[c]	---	---
• Secondary or higher for adult models	n/a	n/a	0.98 (0.91,1.05)[d]	**0.56 (0.40-0.77)[b]**
• Cumulative other household deaths	**8.24 (6.41, 10.59)[b]**	**1.28 (1.07,1.52)[b]**	---	---
• Number of other household deaths	**59.25 (46.87,74.91)[f]**	**9.16 (8.01,10.46)[f]**	**1.04 (1.02,1.07)[b]**	[d]
• Household head:				
Male	0.86 (0.63,1.17)[d]	**0.78 (0.60-1.02)[b]**	**0.53 (0.50,0.57)[b]**	**0.40 (0.35-0.45)[b]**
Death	1.77 (0.03,113.22)[d,e]	1.41 (0.95,2.09)[c]	**3.66 (3.36,3.98)[b]**	**6.74 (5.70-7.92)[b]**
>=40 years of age	1.02 (0.85,1.22)[b]	1.11 (0.90,1.37)[d]	**0.62 (0.58,0.66)[b]**	**0.15 (0.13-0.17)[b]**
Mozambican	1.12 (0.97,1.31)[d]	**1.71 (1.40,2.09)[b]**	1.04 (0.98,1.11)[d]	0.76 (0.66,0.86)[c]

• Parity	0.91 (0.80,1.05)[d]	0.99 (0.90-1.08)[d]	n/a	n/a
• Death of previous sibling	1.17 (0.84,1.65)[b]	**1.63 (1.08-2.45)[b]**	n/a	n/a
• Breastfed	**0.21 (0.17,0.26)[b]**	0.88 (0.58,1.33)[d]	n/a	n/a
• Increasing birth weight	0.42 (0.36,0.50)[c]	0.77 (0.59,1.00)[c]	n/a	n/a
• Other child born less than one year prior	**4.61 (1.34,15.81)[b]**	[c,e]	n/a	n/a
• Four or more children in the household	**1.33 (1.02,1.72)[f]**	**1.44 (1.13,1.80)[b]**	n/a	n/a
Socio-economic determinants				
• Household SES (MCA) quintile for infant and child models; tertile for adult models				
Most poor	1	1	1	1
More poor	0.81 (0.42,1.57)[d]	0.74 (0.51-1.07)[d]	---	---
Poor	0.91 (0.47,1.78)[d]	0.63 (0.43-0.92)[c]	**0.76 (0.69,0.84)[b]**	0.78 (0.58-1.05)[b]
Less poor	0.91 (0.47,1.76)[d]	0.60 (0.41-0.87)[c]	**---**	---
Least poor	0.98 (0.50,1.91)[d]	0.43 (0.28-0.66)[c]	**0.65 (0.58,0.71)[b]**	0.78 (0.58-1.03)[b]
Unknown	[f]	[f]	**1.48 (1.26,1.71)[b]**	1.01 (0.75-1.35)[b]
Environmental or geographic factors				
• Distance to nearest health facility (>6km)	1.18 (0.63,2.21)[d]	0.46 (0.17,1.24)[d]	**5.34 (3.11-9.98)[b]**	1.20 (0.17,8.63)[b,e]
• Climatic proxy (elevation in metres)	0.998 (0.994,1.002)[d]	0.999(0.997,1.000)[d]	0.998(0.997,0.998)[c]	0.998(0.997,0.999)[c]

a: measure of association based on analytical data structure and model type for given age-group; incidence rate ratio (IRR), odds ratio (OR); hazard ratio (HR)
b: significant in bivariate analysis, estimate following multivariate adjustment
c: significant (at 10percent level) in bivariate analysis but substantial missing data so not included in final multivariate model
d: not significant at 10 percent level in bivariate analysis
e: very small numbers
f: not run in Bayesian multivariate framework - either due to co linearity or discovered post publication

Most prominent risks for 50-64 year mortality following multivariate adjustment were later time period, male gender, being a migrant and death of household head (Table 10). Increasing wealth of household, household head being male and older than 40 years were again prominent protective factors. Mozambicans appeared to have significantly lower risk in this age group when compared to South Africans. In contrast to the findings for younger adults, following multivariate adjustment in the 50-64 year model, distance to nearest health facility (>6km) was no longer a significant risk factor.

6.4 Spatial distribution of age-specific mortality, 1992-2007

This section shows the spatial distribution of age-specific mortality. This is first illustrated by assuming a Poisson distribution before using the Kulldorff spatial scan statistic and Moran's I spatial autocorrelation statistic. The spatial distribution of mortality is then developed using Bayesian kriging before showing clusters of space and time distribution of age-specific mortality

6.4.1 Rates and test for randomness of counts assuming a Poisson distribution for all-cause mortality (excludes 5-14 and 65+ years)

By comparing individual village mortality rates to the overall area rate using 95percent CI's based on a Poisson distribution, we observed that three villages in the south-east (lower east grid in Table 11) and one in upper central had significantly higher mortality rates. Two villages with significantly lower mortality rates were also found, both in the upper central region. Similarly, when comparing age-specific

mortality rates we observed significant excess and deficit counts with particular villages emerging as high risk across all age groups (one village in the upper central region of the site and three villages in the lower east). Graphical depictions of this are provided in figures 12 to 15, along with significant spatial differences based on the Kulldorff spatial clustering scan statistic and Moran's I local spatial autocorrelation statistic.

Table 11: Mortality rates (95percent Poisson confidence intervals [a]) per 1,000 person-years by age-group and village

Village	Location	Combined (0-4,15-49) mortality rate (95percent CI)	Infant (<1) mortality rate (95percent CI)	Child (1-4) mortality rate (95percent CI)	Adult (15-49) mortality rate (95percent CI)	Adult (50-64) mortality rate (95percent CI)
16	Upper west	8.17 (7.34-9.07)	19.41 (13.19-27.55)	4.1 (2.63-6.1)	7.08 (6.18-8.08)	17.48 (13.86-21.75)
3	Lower west	7.90 (7.18-8.66)	21.04 (15.46-27.97)	4.51 (3.19-6.19)	**6.16 (5.42-6.97)**	20.23 (16.68-24.3)
1	Upper central	**7.70 (7.07-8.36)**	**17.70 (13.22-23.21)**	4.29 (3.16-5.68)	6.69 (6.01-7.43)	16.25 (13.33-19.62)
2	Upper central	**7.32 (6.43-8.3)**	23.22 (15.66-33.14)	5.97 (4.03-8.52)	**5.77 (4.84-6.82)**	**14.38 (10.57-19.12)**
6	Upper central	9.00 (8.07-10.00)	27.55 (19.94-37.11)	5.18 (3.52-7.36)	6.86 (5.91-7.91)	**24.33 (19.35-30.2)**
13	Upper central	8.35 (7.42-9.37)	25.3 (17.81-34.87)	5.43 (3.69-7.7)	6.56 (5.57-7.66)	18.24 (13.98-23.38)
14	Upper central	9.12 (7.83-10.56)	32.63 (21.5-47.48)	3.72 (1.86-6.65)	7.21 (5.87-8.76)	21 (14.78-28.94)
21	Upper central	**13.05 (10.79-15.66)**	29.51 (17.49-46.63)	**13.19 (7.94-20.6)**	**10.79 (8.45-13.56)**	31.08 (13.42-61.25)
4	Lower central	8.39 (7.46-9.40)	27.64 (19.36-38.26)	4.43 (2.78-6.71)	6.04 (5.13-7.08)	23.08 (18.33-28.69)
9	Lower central	8.00 (7.22-8.85)	18.62 (12.82-26.15)	3.89 (2.56-5.66)	6.99 (6.14-7.93)	17.58 (13.94-21.88)
10	Lower central	8.72 (7.87-9.63)	21.91 (16.04-29.22)	4.01 (2.74-5.66)	**8.01 (7.04-9.08)**	19.17 (14.91-24.25)
18	Lower central	8.22 (6.58-10.13)	22.28 (11.12-39.87)	5.46 (2.72-9.76)	6.64 (4.86-8.86)	18.54 (11.48-28.34)
7	Upper east	9.14 (7.97-10.43)	29.18 (19.54-41.91)	4.61 (2.74-7.29)	7.91 (6.63-9.36)	18.62 (13.3-25.36)
11	Upper east	8.55 (7.86-9.29)	**33.29 (26.66-41.06)**	4.79 (3.55-6.34)	6.8 (6.07-7.58)	18.84 (15.46-22.75)
5	Lower east	8.34 (7.30-9.48)	20.71 (13.27-30.82)	4.28 (2.62-6.61)	6.95 (5.83-8.21)	22.45 (16.71-29.51)
8	Lower east	**9.82 (8.98-10.71)**	30.01 (23.01-38.47)	6.15 (4.58-8.08)	**7.95 (7.06-8.92)**	21.55 (17.52-26.24)
12	Lower east	**10.42 (9.05-11.95)**	24.33 (15.59-36.2)	**9.43 (6.57-13.12)**	**8.61 (7.12-10.32)**	20.75 (14.19-29.29)
15	Lower east	**10.15 (9.08-11.31)**	25.52 (18.15-34.89)	5.88 (4.07-8.21)	**8.55 (7.37-9.87)**	21.25 (16.26-27.3)
17	Lower east	8.63 (7.47-9.90)	24.64 (16.37-35.61)	**9.07 (6.51-12.3)**	**5.69 (4.56-7.01)**	18.28 (13.06-24.9)
19	Lower east	8.50 (6.97-10.26)	28.73 (17.03-45.41)	7.47 (4.5-11.67)	7.1 (5.39-9.18)	**11.06 (6.19-18.25)**
20	Lower east	9.41 (7.67-11.42)	26.64 (14.91-43.95)	**9.77 (6.12-14.79)**	6.54 (4.79-8.72)	18.7 (11.08-29.55)
Overall	---	8.62 (8.41-8.84)	24.68 (22.93-26.53)	5.36 (4.94-5.8)	7.03 (6.8-7.26)	19.17 (18.12-20.26)

a: Statistically significant rates (high or low) are bolded

6.4.2 Kulldorff spatial scan statistic and Moran's I spatial autocorrelation statistic

Towards the south-east corner of the site a statistically significant (at 5percent level) cluster of higher mortality comprising 5 villages was observed for the period 1992-2007 (observed deaths=1831, expected deaths=1706, RR=1.09, p=0.025).

We used the Kulldorff Spatial Scan statistic to identify clusters of villages with higher age-specific mortality. For this we employed an elliptical (instead of circular) scanning window which allows the angle of the ellipse to be varied by the algorithm. This method yielded the following:

Infants: only one significantly higher infant mortality cluster of five villages emerged, which encompassed all villages running closest to the Kruger National Park boundary on the eastern side of the sub-district (observed deaths= 223, expected deaths= 176, RR= 1.38, p=0.034).

Children (1-4 years): two significant clusters of higher child mortality were found using the Kulldorff spatial scan statistic (Figure 12). The villages in red are those with significantly higher than expected counts of child mortality based on Poisson confidence intervals shown in Table 11: one cluster of five villages in the south-east corner (observed deaths= 151, expected deaths=101, RR=1.66, p<0.001) and a single village in the upper central region (observed deaths= 19, expected deaths=8, RR=2.51, p=0.043). The Moran I spatial autocorrelations statistic identified only one significant cluster of high-low (discordant or negative spatial autocorrelation) mortality in the

village in the upper central region of the sub-district. This means the one village in the upper central with excess mortality is surrounded by villages with low mortality.

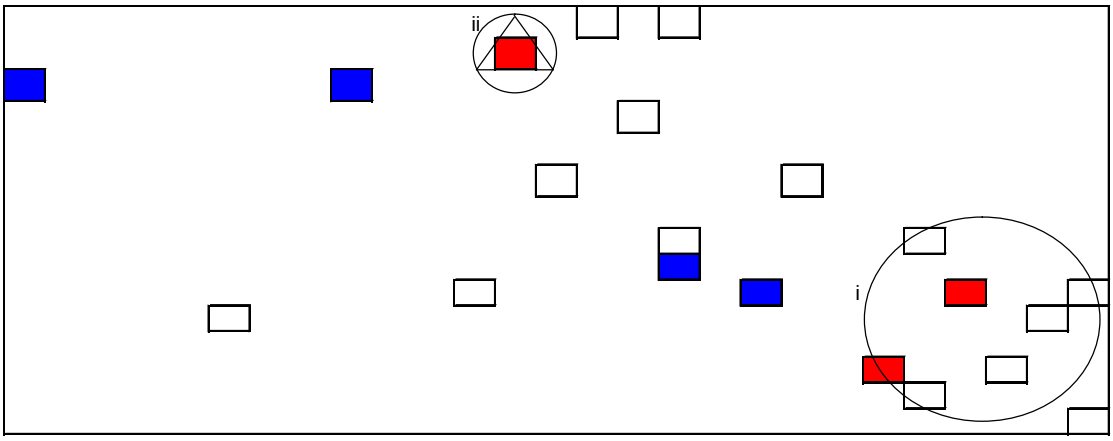


Figure 12: Villages with significantly high or low mortality rates based on 95percent Poisson confidence limits, significant Moran's I local spatial autocorrelation and Kulldorff spatial scan clusters of all-cause mortality in infants and children

Triangle: Moran's I spatial-autocorrelation High-low ($p<0.001$)

Circle i: primary high cluster; Kulldorff Scan statistic ($p<0.001$)

Circle ii: secondary high cluster; Kulldorff Scan statistic ($p=0.011$)

Adults (15-49 years): two significant clusters of higher mortality were observed in similar locations to that observed for children (1-4 years) above based on the Kulldorff spatial scan statistic (Figure 13). One village in the central region of the site had significant excess mortality (Table 11) based on Poisson confidence intervals when compared to the site overall. Though detected by Kulldorff spatial scan statistic it was not statistically significant using this approach. Two significant clusters of mortality were identified using the Kulldorff spatial scan statistic: one cluster consisting of five villages in the south-east corner (observed deaths= 790, expected deaths=702, RR=1.16, p=0.016) and a cluster consisting of one village in the upper

central region (observed deaths= 73, expected deaths=48, RR=1.55, p=0.025). As with child mortality rates, the Moran I spatial autocorrelation statistic only detected on significant cluster of high-low mortality in the village in the upper central region.

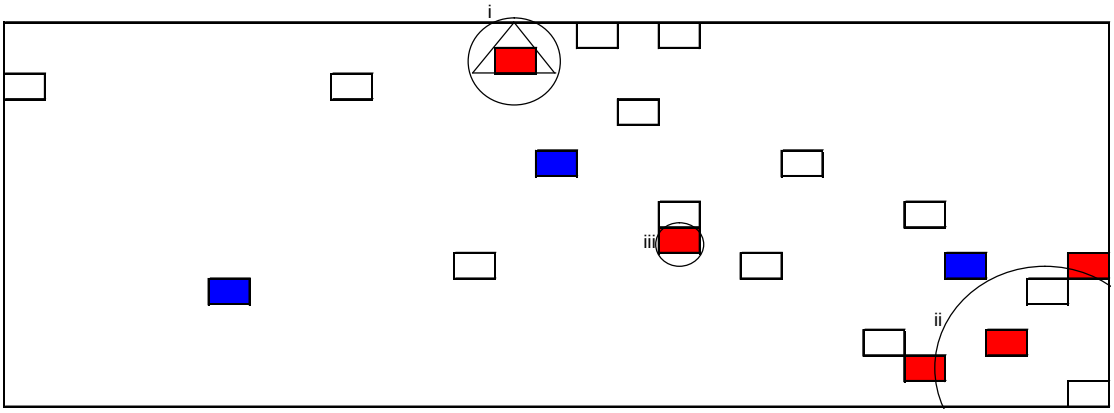


Figure 13: Villages with significant mortality rates based on 95percent Poisson confidence limits, significant Moran spatial autocorrelation and Kulldorff spatial scan clusters of all-cause mortality in adults (15-49 years)

Triangle: Moran's I spatial-autocorrelation High-low ($p<0.001$)

Circle i: primary high cluster; Kulldorff Scan statistic ($p=0.016$)

Circle ii: secondary high cluster; Kulldorff Scan statistic ($p=0.025$)

Circle iii: tertiary high cluster; Kulldorff Scan statistic ($p=0.586$)

Adults (50-64 years): no significant clusters were detected using the Kulldorff scan or Moran I spatial autocorrelation statistics. One village in the upper central region did display significant excess mortality with the lower limit of that villages Poisson 95percent confidence interval being greater than the overall mortality rate for all villages combined. Similarly two villages were significantly below the expected overall mortality rate (shown in blue in Figure 14).

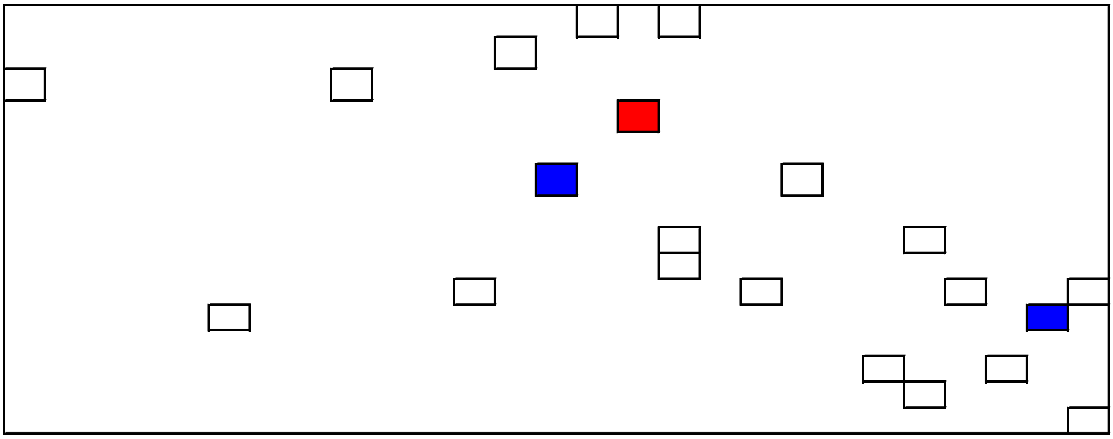


Figure 14: Villages with significant mortality rates based on 95percent Poisson confidence limit comparison to overall rate

A composite of all significant excess or deficit mortality counts for the various age-groups can be seen in Figure 15. There appears to be a “corridor” of mortality starting in the upper central region of the site and runs through the villages in the south-east part of the site. These villages are the ones closest to the Kruger National Park boundary between South Africa and Mozambique. Possible reasons for this observed pattern will be considered in the discussion section.

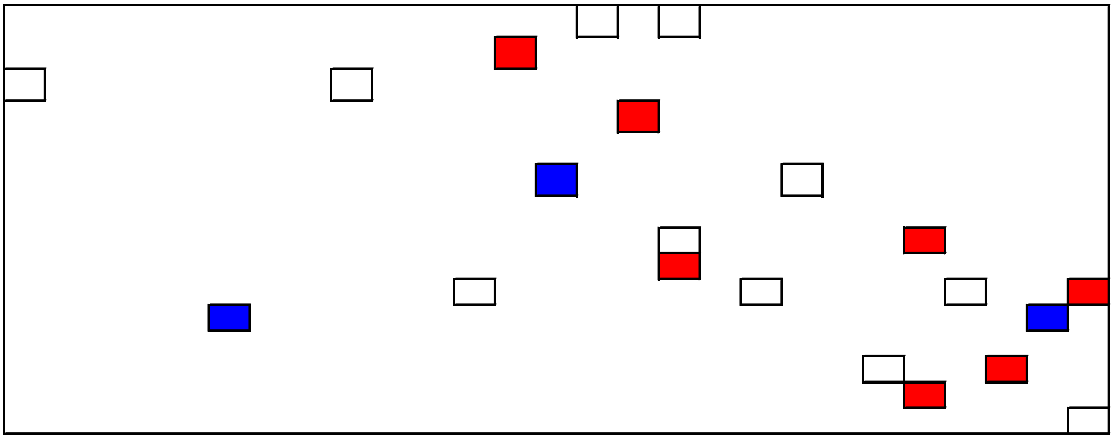


Figure 15: Villages with significant excess mortality based on individual age-group results using Poisson confidence intervals (section 6.4.1)

6.4.3 Bayesian kriging

The risk estimates based on age-specific Bayesian kriging prediction models showed a similar distribution of higher risk as shown by the cluster and village rate findings (Figure 16), especially with regards to infants and children. The similarity in findings for adults when comparing village rates and 95percent CI's, Kulldorff high risk clusters and Bayesian kriging yielded a slightly different pattern and indicated some heterogeneity between the techniques. For example two villages with significantly higher mortality rates among adults aged 15-49 years (based on comparing the village specific lower confidence limit for its mortality rate to the overall aggregated rate) appeared as lower risk predictions using the Kriging approach.

Five distinct foci of higher mortality in the 15-49 age-group were observed using Bayesian kriging (Figure 16). Three are in the central to upper central region of the site and two in the south east. These correlate to areas with higher risk of infectious disease mortality in this age group, largely HIV/TB. A very similar pattern was seen in the 50-64 year age-group when compared to 15-49 years though with one minor difference in that one village in the south east was no longer at higher risk and one additional village in the upper central region emerged as high risk. Similarly, this distribution is largely driven by HIV/TB mortality. Higher non-communicable disease mortality risk was observed in one particular village in the upper central region of the site. Based on the prediction model that adjusted for straight-line distance to health facility from prediction points using a univariate Bayesian kriging model that included this risk factor we can see that two villages, one in the upper and the other in the lower south east region, appear to have a higher mortality risk as a function of

increased distance to the nearest local clinic in the Agincourt sub-district (Figure 17). We also observe that there are parts of other villages that appear to be at a large distance from the nearest health facility.

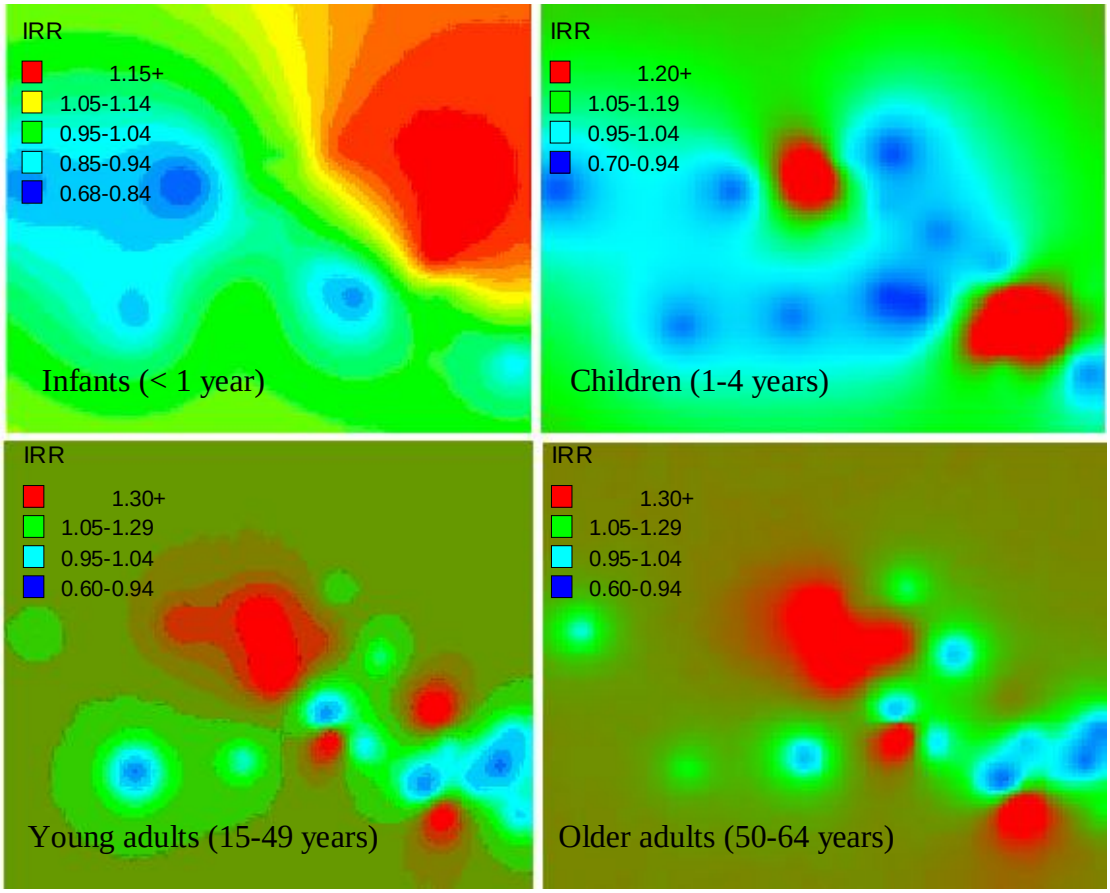


Figure 16: Maps of all-cause mortality risk by age group within the Agincourt sub district based on baseline models without covariates

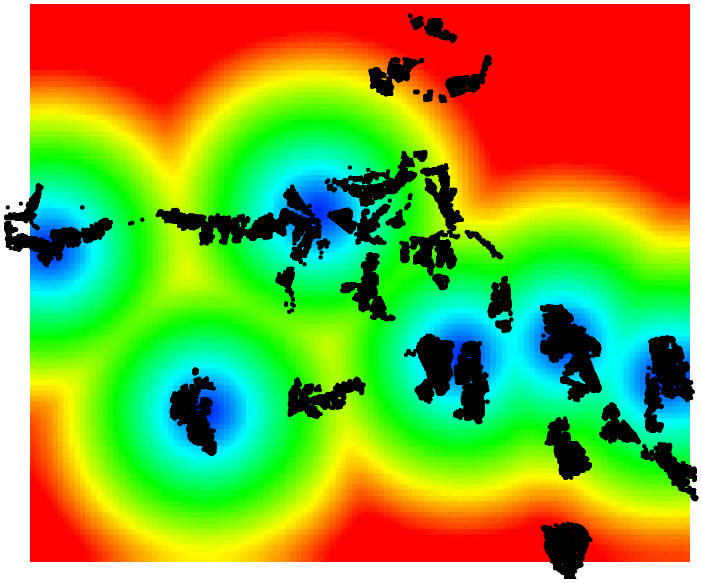


Figure 17: Geographic risk of adult mortality adjusting for distance to nearest health facility at prediction point using a univariate Bayesian spatial kriging model that includes distance to health facility as the risk factor (2km band increments with blue being the lowest risk, >6km or high risk areas are represented in red)

6.5 Space and time distribution of all-cause age-specific mortality

Significant spatial-temporal clusters of age-group specific all-cause mortality can are presented in Table 12. A significant space time cluster of higher all-cause mortality was observed among infants in four villages in the south east corner of the site during 2001-2007 (105 observed cases, 66 expected, RR=1.67, p=0.012). A significant space time cluster of higher all-cause mortality was observed among children (1-4 years) in four villages in the south east corner of site during 1997-2004 (63 observed cases, 34 expected, RR=2.04, p=0.005). During 2001-2007, a significant high cluster of adult mortality (15-49 years) was observed in the south east comprising five villages (548 observed cases, 323 expected, RR=1.82, p<0.001) as well as another cluster of five villages in the lower central/east region during the same period (483 observed cases,

318 expected, RR=1.60, p<0.001). Significant clusters of higher older adult mortality (50-64 years) were observed in similar areas during similar periods (Table 12).

Table 12: Significant clusters of all-cause mortality by age-group using the space-time scan analysis scanning for high mortality rates only

Age-group	Years	Number of Villages	Location within site	Observed cases	Expected cases	Relative risk (RR)	p-value[a]
<1	2001-2007	11	Central/South east	286	195	1.67	<0.001*
1-4	1997-2004	4	South east	63	34	2.04	0.005*
15-49	2001-2007	11	South east	1293	822	1.90	<0.001*
15-49	2001-2007	9	West/Central	1139	809	1.60	<0.001*
50-64	2002-2007	12	Upper-mid Central/South east	368	248	1.69	<0.001*

a: significant at 5percent level (*);

6.6 Potential proximate reasons for the observed high risk clusters

Reasons for the observed high risk mortality spatial clusters in infant, children and adult age-groups are listed below. The significance comparison is based on the comparison of the distribution of various predictors in aggregated high risk versus low risk village clusters as identified using the Kulldorff spatial scan statistic.

Infants

The high infant mortality cluster, comprising five villages along the Kruger National Park Boundary, when compared to the lower mortality cluster had a significantly:

- Higher number of cumulative household deaths;
- Higher number of deaths of previous children;
- Higher incidence of household heads dying;
- More Mozambican household heads;
- More infant deaths due to HIV/TB and diarrhoea or malnutrition;
- Higher number of mothers dying of HIV/TB at the 10percent level;
- Lower duration of breastfeeding.

There was no significant difference in father deaths prior to birth or in infants first year, however numbers were small and it appears there might be a higher incidence in the high risk cluster (IRR=1.41,p=0.126). Interestingly households in the high risk cluster were significantly closer to health facilities on average than in the lower cluster with a mean of 1.75 versus 2.55 kilometres (km's) respectively.

Children

The high child mortality cluster had a significantly:

- Higher number of infant deaths due to HIV (OR=1.75; p=0.020);
- Higher number of infant deaths due to diarrhoea or malnutrition (OR=2.54; p<0.001);

- Higher proportion of Mozambican mothers and household heads (OR=7.07; p<0.001 and OR=7.32; p<0.001 respectively) ;
- Higher proportion previous sibling deaths (OR=1.47; p<0.001);
- Higher mortality rate in all other age groups combined (p<0.001);
- Younger household heads (OR=0.78; p<0.001);
- Higher proportion of households greater than 6km to nearest health facility (OR=1.44; p<0.001);
- Higher proportion of male headed households (OR=1.44; p<0.001);
- Higher mother deaths (while child aged <5) due to undetermined causes, probably related to increase of HIV (OR=2.17; p=0.077). This is expected due to limitations of the VA tool determining HIV-related cause-of-death and often non-specificity of symptoms leading to undetermined cause assessment by the physicians;
- Lower mean maternal education years (5.39 versus 7.75 years, p<0.001);
- Lower proportion of migrants (OR=0.87; p<0.001).

No significant differences between these clusters were identified with regards to mother or father death due to any cause or due to HIV/TB while their children were aged <5 years. There was no significant difference with regards to proportions of preceding births being less than 1 year prior. There was almost a marginally significant higher number of migrant months in the high risk cluster of villages when compared to the remainder (OR=1.004; p=0.106).

Adults

The high younger adult (15-49 years) mortality cluster had a significantly:

- Lower proportion of individuals with secondary or higher level education (OR=0.54; p<0.001). Average education years per individual in high cluster areas was 7.31 years versus 9.27 in lower risk clusters;
- Higher odds of a household belonging to lower socio-economic status category (OR=0.66; p<0.001);
- Higher odds of having a Mozambican household head (OR= 2.60; p<0.001);
- Higher likelihood of having a male household head (OR= 1.06; p=0.001);
- Lower likelihood of having adults involved in government, professional or private sector skilled labour (OR=0.80; p<0.001);
- Lower likelihood of having a household head 40 years or older (OR=0.61; p<0.001);
- Higher number of migrant months (OR=1.005; p=0.039) - especially high number of migrant months in the one village in the upper central area which consistently emerged as mortality hotspot (average of 4.1 migrant months per year per individual in this village compared to 3.3 in all other villages combined; p<0.001). This village in particular also had a significantly lower mean age.
- More likely to have a higher household proportion of deaths attributed to HIV/TB (p<0.001).

There was no significant difference in terms of male headed households or proportions of household head deaths.

7.0 Discussion

The following section discusses the findings of this thesis in relation to the literature, makes some recommendations and proposes future related research questions that need to be addressed. This section is ordered as per the results in section 6 but certain key findings in one or more of the published papers of this thesis (see Appendix) may also be discussed. Finally, a number of conclusions are reached.

7.1 All-cause mortality

Key findings

- High mortality in Agincourt, though slightly lower than national average
- Highest mortality rates in infants and elderly
- Excess mortality in younger to middle age adult age groups due to HIV/AIDS

The results indicate that overall 9,035 deaths occurred in the Agincourt sub-district during the study period 1992-2008 (overall crude mortality rate of 8.1 per 1,000 person-years). This is lower than the average crude mortality rate of 13.3 per 1,000 population described for the country as a whole during a similar period 1994-2008 (Health Systems Trust, 2010). The highest mortality rates occurred among infants followed by the older adult age group age group (50-64), with 29 and 19 deaths per 1000 person-years respectively. The mortality rate among children (1-4) and younger adults (15-49) was similar at 5.7 and 6.9 per 1,000 person-years respectively. A high burden of perinatal mortality was observed with a significant increasing trend from

1996 onwards. It does, however, appear that the early neonatal mortality rate in this sub-district (approximately 8 per 1,000 person-years for the aggregated study period) is lower than the regional estimate given by WHO for 2004 of approximately 34 per 1000 live births (WHO, 2006). Among adults (15-64) log mortality hazard showed a strong non-linear excess between ages 15-55 due to HIV/AIDS (Figures 1 and 2). Generally one expects to observe a linear increase in the log mortality hazard in the adult age range.

7.2 Leading causes of death

Key findings

- Most prominent cause of death HIV/TB
- Emerging non-communicable disease mortality in older age groups
- External causes of death (violent or accident related) prominent among younger adults

Earlier work in South Africa and Agincourt has shown the profound impact of HIV/TB on mortality across most age groups (Bradshaw et al., 2000; Kahn et al., 2007a, Tollman et al., 2008). The results show that infectious diseases are the most prominent cause-of-death with the leading cause-of-death in all age groups being HIV/TB. Approximately half of all child (<5) and younger adult (15-49) deaths could be attributed to infectious causes. Among children the second most prominent cause-of-death was diarrhoea or malnutrition. Mortality from non-communicable disease has also increased significantly in adults 30 years and older in the rural Agincourt sub-district (Tollman et al., 2008). Violent or transport related mortality also featured

prominently in younger adults. In the older age groups however, lifestyle or non-communicable related mortality (e.g. vascular disease, neoplasms) featured prominently following HIV/TB. Thus, despite mortality in this sub-district being dominated by communicable diseases, non-communicable diseases are evident and emerging (Tollman et al., 2008).

7.3 Temporal, spatial and spatial-temporal distribution of mortality

Key findings

- Significant temporal mortality increases in most age groups due to HIV/TB particularly from the late 1990's
- Levelling out of mortality around 2005 possibly linked to ART access outside the sub-district
- Strong space-time differences in age-specific mortality detected even within a small area such as the Agincourt sub-district
- Distinct clustering of higher mortality risk in the eastern and upper central area
- Strong association with former Mozambican refugee location
- Emerging higher risk cluster in upper central - newer village with a young and highly mobile migrant population
- Strengths and weaknesses of different methodologies
- Recommendations for future space-time work

7.3.1 Summary of findings

The results show significant temporal increases in mortality in most age groups in Agincourt, especially from the late 1990's onwards, largely due to the increase in HIV/AIDS. This is similar to findings from other parts of South Africa during similar periods (Dorrington et al., 2001; Hosegood et al., 2004).

There was a significant worsening of infant mortality over time as well as child mortality, particularly between 1996 and 2003. The increase was observed from 1998 onwards, and can be largely attributed to the HIV epidemic and its impact on mortality (Newell, 2001; Tollman et al., 2008), both directly (vertical transmission of HIV) and indirectly (death of a caregiver, resultant loss of wealth). HIV/AIDS deaths due to diarrhoea also remained prominent. Interventions to reduce child mortality must target infectious causes, specifically HIV and diarrhoea. A substantial and significant increase in adult mortality was observed, as has been documented in other parts of the country, over the same period (Dorrington et al., 2001, Hosegood et al., 2004). The increase began around 1999 and seemed to plateau around 2006. Given the younger adult ages at which these additional deaths are occurring (Figures 2 and 3 with log hazard of mortality by yearly age) and the change in the cause-of-death profile, this rise can largely be attributed to HIV/AIDS. This is similar to findings from the Africa Centre in northern Kwazulu Natal where an increase in adult mortality in the late 1990's, with the largest cause-of-death being HIV/TB, was documented (Hosegood et al., 2004). The levelling out of adult mortality is possibly linked to the ART rollout which began in South Africa round 2004, though this plateau in adult mortality in Agincourt preceded ART rollout in this area. This could

be a data effect due to small numbers or as a result of individuals accessing ART outside the site pre Agincourt rollout. The results, supported by other findings, indicate that Government should rapidly augment its current plan to prevent and treat HIV/AIDS. ART rollout started in the Agincourt sub-district in 2007 and, as a result of this delay, many unnecessary deaths have probably occurred. Future studies in this area can better assess the post impact of ART rollout on mortality, as well as specific villages or areas were equity of access may be an issue.

There is renewed interest in the spatial clustering of infectious disease and mortality, especially in poor areas with limited resources. Spatial variations of age-specific mortality remain substantial in developing countries, yet little proper spatial analysis of longitudinal HDSS data undertaken. Results showed strong geographical and temporal differences in mortality even within a small area such as the Agincourt sub-district (~400km^2). Several statistically significant clusters of higher all-cause and cause-specific mortality rates were identified among the 21 villages both in space and space-time. This distribution is being driven by a complex web of interacting factors that have likely increased communicable disease mortality (HIV) and non-communicable disease mortality (in the older age-group) in specific risk areas. A strong geographical pattern with regards to higher infectious disease mortality risk (particularly HIV/TB and diarrhoea) and former Mozambican settlements lying to the east of the site was generally observed. According to the spatial cluster analyses the south-east and upper central regions of the site were consistently identified as high risk areas for most age groups thus indicating a definite non-random element to the mortality distribution in this rural sub-district. From the space-time cluster analysis we observed that most of the significant mortality clusters appeared during the later

period (1999-2007) with none in the earlier period (1992-1998) (Table 12). Significant increases in mortality rates particularly in <5, 15-49, and 50-64-year age groups were observed (Tollman et al., 2008). Hence this newly described temporal increase in spatial clustering is also linked to the increase in mortality over the time period. According to the space-time cluster analysis, the south east and upper central regions of the site appear to have the highest mortality risk. The results inform policies to address health inequalities in the Agincourt sub district and improve access to certain health services. Specific efforts to prevent vertical transmission of HIV should target specific villages, as well as interventions to promote mother and father in childhood. A significant space cluster of older adult (65+) mortality (as well as a space-time cluster of higher mortality among 50-64 year olds during 2002-2006) was observed toward the west of the site in the later period. More affluent South African villages are concentrated towards the west of the site. A recent study by Tollman et al. (2008) found a significant increase in the mortality rate due to non-communicable diseases in adults 30+ when comparing the baseline period of 1992-1994 to 2002-2005 (RR=1.22, p=0.026) . Thus, an emergence of lifestyle related disease and mortality has been documented and the spatial analysis point to specific areas within the site where this may be most prominent.

The highest infant mortality risk was in villages bordering the eastern part of the site. With regard to the geographical distribution of infectious infant deaths (particularly HIV/TB) there was a distinct spatial pattern of mortality with an increasing gradient towards the east of the site where communities appear to be at increased risk and where suitable interventions need to be directed accordingly. Diarrhoea and malnutrition-related mortality was clustered in the south east of the site suggesting

greater problems with clean water and sanitation. Based on an environmental survey conducted some years before , Mozambican refugee settlements in the southeast were consistently worse off with respect to access to water, sanitation, waste disposal; in addition they had fewer schools, poor quality of housing and were particularly isolated from public transport (Dolan et al., 1997). Similar findings have been documented by Collinson more recently (2010). It would appear therefore that the concentrated burden of mortality has not been alleviated and that service provision to more marginalised and poorer communities remains inadequate. A study assessing child mortality disparities between ethnic groups in eleven countries of sub-Saharan Africa found close links to economic inequity and differential use of health services (Brockerhoff and Hewett, 2000). Service provision in this sub-district needs to be fully assessed and strengthened by local government as disparities are still evident.

Spatial autocorrelation and clustering analysis indicated two distinct pockets of higher child mortality burden towards the southeast and upper central parts of the site. This was consistent with the cause-specific mortality risk distribution we observed for HIV/TB, diarrhoea/malnutrition and ARI/pneumonia. Distinct spatial foci of increased adult mortality risk were observed in almost identical foci to that of child mortality. According to the space-time analysis, recent trends suggest that interventions need to first target villages in the southeast (high risk throughout the period) and upper central (emerging risk from 2000 onwards) areas of the site where the highest mortality burden lies. This evolving space-time risk distribution is likely being driven by the evolving HIV epidemic, a temporal phenomenon previously documented in this area (Tollman et al., 2008). One village in the upper central region appears to be at consistently higher risk across all age groups. This village has a

significantly younger and more highly mobile population, potentially engaging in higher risk behaviour with more time spent away as described by (Collinson, 2010).

Addressing the concentrated spatial-temporal burden of mortality requires a multifaceted approach that includes the provision of clean water and sanitation, ensuring maternal survival, promoting maternal, infant and child nutrition, and strengthened primary care services for mothers, infants and children. A comprehensive approach to the prevention of mother-to-child transmission (PMTCT) in these infant and child mortality hotspots is needed. This would include a combination of antiretroviral therapy (ART) from early pregnancy, elective caesarean section and highly active anti-retroviral therapy (HAART) for mother or infants during breastfeeding (Patwari, 1999; Coovadia et al., 2007; WHO, 2010), can significantly reduce transmission rates in this sub-district (European Collaborative Study, 2005; Naver et al., 2006; Newell et al., 2007). The persistent high burden of diarrhoea and malnutrition related mortality in the south east of the site is of concern. Routine testing and improved water supply to villages within this region is required. Rehydration fluid and dietary management are key aspects in the treatment of acute diarrhoea. The capabilities and resources of health facilities, specifically those situated near the south east corner of the sub district, to effectively manage children presenting with diarrhoea and/or malnutrition needs to be assessed and improved (Sartorius el al., 2011).

Some of the spatial distribution of mortality risk described above can be partly explained by differences in nationality within the site. Mpumalanga Province in northeast_South Africa was an important destination for refugees fleeing the civil war

in Mozambique from 1983 onwards. A formal peace agreement was signed in 1992, yet despite voluntary repatriation programmes, by 2000 it was estimated that more than 200 000 former Mozambican refugees were still inhabitants in the province (Johnston 2000). A previous study by Hargreaves (Hargreaves et al., 2004) demonstrated higher mortality rates among children from former Mozambican refugee households when compared to South Africans in the Agincourt sub-district. They concluded that lack of legal status and poorer socio-economic status (SES) of many former Mozambican refugees partly explains this disparity. They remain a vulnerable group and policy amendments are needed to address any inequity and differential access to various services. Other underlying factors driving this and other disparities with regards to mortality distribution need to be assessed in more detail, as well as the distribution of other proximate, socio-economic or spatial driven predictive factors.

This study has also demonstrated the usefulness of Bayesian geostatistical survival or event history models in assessing risk factors and producing smooth maps of infant, child and adult mortality risk in a health and socio-demographic surveillance system. These generated risk maps can be used by decision makers for the design and implementation of an ART rollout. Given that a larger distance from the nearest health facility was a risk in adults, the usefulness of Bayesian kriging can also be highlighted by the generation of risk prediction maps at locations throughout the site and that distinct household and village foci of higher risk were identified. This will be discussed in more detail in the risk factor section.

7.3.2 Methodological comparison and future work

The various strengths and limitations of the different spatial analysis techniques used in this thesis (both in theory and as identified in practice with application now to Agincourt HDSS data) are listed in Table 13. This study has demonstrated the usefulness of these different spatial analysis techniques, namely Kulldorff's spatial-temporal scan statistic for clusters (Kulldorff and Nagarwalla 1995; Kulldorff et al., 1997) and Moran's I (Moran, 1948) for autocorrelation, in highlighting high risk areas within a rural sub-district as a guide for targeting health interventions. Overall the Kulldorff Scan statistic appeared to perform best when identifying significant clusters of age-specific mortality in space and space time, as it can identify aggregates of higher risk villages that singularly may not have had a significantly high mortality rate. There is a future need to identify the most efficient, accurate and simple spatial analysis technique(s) to identify significant space-time anomalies of a given outcome (mortality, divorce etc) for routine HDSS use. There are several spatial analysis techniques (from simple Poisson confidence intervals; to intermediate spatial autocorrelation and clustering; and advanced geostatistical modelling), all of which have certain intrinsic advantages and disadvantages (Table 13). For example the main limitation of the circular Kulldorff's spatial scan statistic is the detection of irregularly shaped clusters. The proposed way forward would be the use of a simulated longitudinal dataset with predefined anomalies embedded and testing the various spatial techniques in terms of efficiency, sensitivity and specificity to detect these predefined anomalies. Subsequent the best technique would be used on empirical. The overall objective of this would be to identify which is the most suitable for

longitudinal HDSS data and then make recommendations for routine use within the INDEPTH network.

Future studies that will employ the Kulldorff Spatial Scan statistic will also use the ellipsoid scanning window which will remove some of the limitations of the circular variant described above as the angle of an ellipsoid can also be changed and can highlight clustering along a linear boundary such as the border of the site with the Kruger National Park. We will also employ the novel and more sophisticated approaches suggested by (Duczmal, 2004; Duczmal et al., 2007) to further improve the detection of irregularly shaped clusters (Table 13) and compare performance with the Kulldorff Spatial Scan statistic. A recently modified and improved variant of the Moran's I spatial autocorrelation statistic (Jackson et al., 2010) has proven to be more powerful for identifying local and global clusters and will also be tested in the near future.

Table 13: Strengths and limitations of different and increasingly complex spatial analysis methods (in theory and as applied to Agincourt longitudinal HDSS data)

	Village-specific mortality rates (95 percent Poisson CI's)	**Spatial autocorrelation: standard Moran's I**	**Spatial and space-time clustering: Kulldorff Scan Statistic**	**Bayesian modelling or kriging**
Strengths	Significant high or low mortality can be observed in single and multiple aggregated villages by age group and period	Identify location and spatial scale of aggregations of unusual values such as clusters of high and low mortality (hot and cold spots) (Boots, 2002) Example of this is the identification of a single incongruent village in upper central region of site with higher mortality risk	Significant mortality clustering of one or multiple combined locations can be observed in space and space-time Lower penalty associated with the elliptic spatial scan statistic (Kulldorff, 2006) Novel approaches (Duczmal, 2004; Duczmal et al., 2007) such as the annealing strategy and genetic algorithm for dealing with irregularly shaped clusters. Perform faster, have less variance and more flexible than elliptic scan (Duczmal et al., 2007)	Produce smooth maps of mortality risks (can incorporate irregular nature of mortality distribution) Prediction of mortality at unsampled locations based on prediction of spatially relevant covariates (e.g. distance to nearest health facility, climatic or environmental factors) High potential to model changes in tropical areas for diseases such as malaria and forecast predicted change in malaria distribution within HDSS given climatic or environmental changes
Limitations	Provides no spatial structure or relative position Minor excesses in single villages that would be significant when clustered with nearer villages (with similar traits) is not taken into account and would be missed Small numbers make it difficult to identify significant differences	Lack of sensitivity (minor excesses significant over a group or cluster of nearer villages not taken into account and would be missed) Does not account for population heterogeneity or density (Jackson et al., 2010)	Lack of specificity – given that clusters are as circular windows: villages with lower mortality surrounded by higher mortality villages are included in the cluster but will have different characteristics; when searching for clustering along a geographic boundary (e.g. river, geographical boundary such as the Kruger National Park) then a circular window may not be appropriate (Sankoh et al., 2001) Inability or penalisation of circular statistic to detect irregularly shaped mortality patterns (Duczmal et al., 2006)	Given large standard errors assigned to prediction estimates, difficult to assess significant differences or excess risk between locations With HDSS we assume that all households are mapped, hence limited potential for Bayesian kriging or prediction at unsampled locations for spatially relevant factors Limited potential in an area (such as Agincourt) where tropical or environmentally driven disease not prominent

			Results from this method are sensitive to the selection of scaling parameters (Waller and Gotway, 2004; Fukuda et al, 2005; Chen et al., 2008).	

7.4 Risk factors for mortality in Agincourt HDSS

Key findings

- Known multilevel risk factors confirmed and novel predictors identified
- Confirms the adapted disease-ecology theoretical framework I developed for this setting in the literature review
- Mother survival and high household mortality burden are key determinants of infant and child survival outcomes
- Direct and indirect or proximate effects of HIV/AIDS are evident
- Higher risk among children (1-4 years) associated with Mozambican nationality, socio-economic status and death due to diarrhoea and malnutrition
- Complex interplay of factors related to migrants appears to be driving adult mortality due to HIV/TB
- Distance to nearest health facility is a significant risk factor in adults but not among infants and children

7.4.1 Confirmation of known factors and identification of novel predictors

Certain multilevel determinants are now discussed in detail in relation to age-specific mortality. Key individual and household level determinants have been confirmed and certain novel determinants have emerged. Certain complex multilevel dynamics are driving mortality (overall and age-specific) in this sub-district. I know discuss some of the key multilevel aspects of mortality in relation to the age groups as well as other key determinants which can be related back to the framework in section 2.4.

Infants, children and mothers

A mother's death in the infant's first year or childhood (1-4) was a major risk factor in this study, as was higher numbers of cumulative household deaths. Other reports indicate that infants who survive the death of the mother, only have a 10 percent chance or less of living past the age of one year (Koblinsky et al., 1994; Ronsmans et al., 2010). Death of the child's mother (and father to a smaller extent) between their first and fifth birthdays (specifically due to HIV/TB) was also a prominent risk factor. The results confirmed the importance of other known risk factors such as age of mother, birth spacing, season, village and ethnic group (Mosley et al., 1984; Hobcraft et al, 1985; Binka et al, 1995; Ronsmans, 1996; Kuate Defo, 1997; Manda, 1999; Hargreaves at al., 2004; Gemperli et al., 2004; Becher et al., 2004; Hammer et al., 2006,). The association between socio-economic status, maternal education and mortality has been previously described (Farah et al,1982; Moser et al., 2005; Schellenberg et al., 2008; Chowdhury et al., 2010). Higher education may result in better health awareness and utilization of health facilities (Jain, 1988; Kuate-Defo and Diallo, 2002; Becher et al., 2004), as well as translate into higher income and the ability to purchase goods and services that improve child's health (Schultz, 1979). Education, moreover, may influence longer birth intervals and possibly higher maternal ages (Cleland and van Ginneken, 1989; Becher et al., 2004; Schellenberg et al., 2008). A significant protective association between higher household SES and reduced infant mortality risk was, however, not observed in this study. This has been shown elsewhere and may be explained by the fact that unlike endogenous maternal and demographic factors that substantially influence an infant's risk of death, the effects of SES factors on mortality increase as the child gets older due to exogenous

factors over which parents have more control (Manda, 1999). The higher risk of child mortality in the winter season is most likely due to the increase in respiratory illness, as well as environmental or household pollution due to the burning of fuel (e.g. coal, wood, paraffin) for indoor heating and cooking. Results from a previous study in South Africa suggest that exposure to cooking and heating smoke from polluting fuels is significantly associated with under 5 mortality (Wichmann and Voyi, 2006).

Direct vertical transmission of HIV or the loss of the primary care giver(s) leading to to impaired child care and indirectly through reduced household income are the most likely explanations for this finding. About half of children infected with HIV through vertical transmission develop AIDS symptoms and die within two years if they are not on ART (UNAIDS, 2002). Published work from this thesis (see Figure 4 in Sartorius et al., 2010b; and Figure 3 in Sartorius et al., 2011) clearly illustrates this impact on infant and childhood mortality risk in this study population. Prevention of vertical transmission of HIV and survival of mothers during infancy and childhood in high prevalence villages needs to be urgently addressed, including expanded antenatal testing (improve elective access to testing and ARV treatment to mothers and fathers in these settlements), prevention of mother-to-child transmission, and improved access to antiretroviral therapy. There is also a need to assess and improve the capacity of district hospitals for emergency obstetric and newborn care. Persisting risk factors, including inadequate provision of clean water and sanitation, are yet to be fully addressed. A comprehensive approach to prevention of mother to child transmission PMTCT, including a combination of ART from early pregnancy, elective caesarean section and avoidance of breastfeeding, can significantly reduce

transmission rates in this sub-district (European Collaborative Study, 2005; Naver et al., 2006; Newell et al., 2007). However, in resource-poor settings such as the Agincourt sub-district, the risks of such surgical procedures need to be taken into consideration before recommending caesarean sections as a feasible strategy for PMTCT. The avoidance of breastfeeding must also be balanced against the risks associated with replacement feeding such as cost and lack of access to clean water (Thior et al., 2006; Coovadia and Bland, 2007) which we have shown to be a problem in the south east region of the site. Breastfeeding had a protective effect on all-cause, as well as diarrhoea and malnutrition-related infant mortality. Breastfeeding protects infants through decreased exposure to contaminated water and food, optimal nutrition, and improved resistance to infection, however, there is risk of HIV transmission through breast milk. In South Africa, the Department of Health policy with regards to breastfeeding practices by HIV positive mothers has evolved in response to emerging research (Coovadia and Bland, 2007). Current recommendations are to breastfeed exclusively during the first six months with administration of anti-retrovirals to HIV positive mothers (WHO, 2010), especially those mothers with low CD4 counts. Mothers or infants receiving highly active anti-retroviral therapy (HAART) prophylaxis should continue prophylaxis for one week after breastfeeding has ceased (WHO, 2010). Infant mortality due to diarrhoea, malnutrition and their interaction is a complex problem in poor, high HIV prevalence African settings. Addressing this issue requires a multifaceted approach that includes provision of clean water and sanitation, promotion of infant nutrition, and strengthened primary care services for mothers and infants to reduce the risk of HIV transmission through breast milk (Patwari, 1999).

The mothers' physical presence or absence had a significant impact on infant mortality. Conversely, the mother being a temporary migrant (largely work-related) proved to be significantly protective and mothers who spent an increasing number of months resident in the site appeared to increase infant mortality risk. Brockerhoff (1994) proposes how maternal rural-urban migration may affect children as a result of three types of living arrangements, namely: children may remain in the village as foster-children in the care of their fathers or other relatives; children may accompany or follow their mothers to towns or cities; and children born after migrant mothers settle in an urban area may remain there through the first few years of life. In this study, infants born to mothers who have permanently relocated to urban areas would not necessarily be captured by the HDSS. However, this information would be captured if they later migrated back into the site or if a household respondent in which the mother previously resided was asked details about the absent individual(s) and any children born to her. Bledsoe and Anastasia (1992), when reviewing evidence from West Africa, suggest that while fostered children may be disadvantaged compared to biological children (in terms of access to health care and nutrition), they may still be better off than if they had accompanied their migrant mothers. By staying home, these children avoid exposure to infectious diseases during a vulnerable period of their life, have continued access to economic resources of a non-migrant father, and benefit from remittances received from the migrant mother (Bledsoe and Anastasia, 1992) – as well as better health care, nutrition and enhanced maternal health knowledge (Hildebrand and McKenzie, 2004). In our study, migrant mothers had significantly higher education and came from households with significantly higher SES which may explain the protective effect of mothers' migration. According to Collinson et al. (2001) there has been an increasing trend in the number of temporary female labour

migrants since 1997 in the Agincourt sub-district, a poor area with limited employment opportunities with resulting pressures to migrate and remit wages back to the rural household.

Death of household members, other than the mother or father, also appeared to be a significant risk in this study. Previous research indicates that HIV has had great (if not the greatest) impact at the household level in terms of dissolution and reduced economic status (Hosegood et al., 2004; Yamano and Jayne, 2004). Thus, the death of these members places additional strain (for example financial burden of payment for medical services and funerals) on the household which negatively impacts on the child's health outcome. Few studies have quantified the excess risk of infant or child mortality associated with other household deaths. High mortality burden households are vulnerable and require both financial and social support to reduce the indirect impact on their children (Hosegood et al., 2007). The household is forced to take on the costs of the funeral, loss of income and labour and, in some cases, the leadership skills of a lost household head. A previous study by Hosegood (2007) has shown that the loss of the household head can have a pivotal effect on other household members. In most instances, our research has indicated that female headed households result in greater mortality risk in the various age groups examined. The socio-economic impacts of AIDS on households include both direct and indirect costs following the premature death of an adult. The direct costs are medical prior to death, and the funeral cost following death (Booysen, 2002; Hosegood et al., 2007) while the indirect cost relates to income lost from a working age individual, along with the impact of this on food security, child schooling and loss of assets (Booysen, 2002).

Former Mozambican refugees and child mortality

Much of the spatial concentration of infant and child mortality in villages bordering the east of the site can be partly explained by the migration patterns of former Mozambican refugees who now constitute about a third of the Agincourt HDSS population (Kahn, 2006). These people entered South Africa via the Kruger National Park, a game conservation area situated between the eastern border of the site and Southern Mozambique. The self-settled Mozambican settlements are concentrated to the east of the site with predominantly South African villages to the west. Despite equity- orientated policies in South Africa, for example, free primary health care for children and pregnant women, it is generally children at the lowest risk of mortality who are able to access services and are easiest to reach (Twine et al., 2007). Although primary health care is free, transport costs are high; this is a barrier as former Mozambican's are generally further away from health services and have less access to paid work. We found a significant risk for all-cause child mortality associated with having parents of Mozambican origin. Mozambican villages are more vulnerable and isolated with weaker infrastructure; Mozambican households experience poorer access to housing, water and sanitation, social grants, education and health services; and social discrimination (Dolan et al, 1997; Kahn, 2006; Kahn et al., 2007a). These children are the least likely to receive interventions and those who receive one intervention are most likely to receive subsequent ones. Hargreaves et al (2004) suggest the higher mortality rates among children of former Mozambican refugees could be due to the fact that Mozambican households generally are generally poorer (three times more likely to be in the poorest socio-economic quintile) with fewer resources

than South African households. Until recently the lack of legal status meant Mozambican children could not access the child support grant which puts them at further disadvantage relative to South African children. Policy needs to be amended to address differential access to services and inequalities. However, our study indicates that these mortality patterns are not driven by nationality alone, a finding supported by Hargreaves et al. (2004) who found no difference in mortality rates between South African and former Mozambican infants between 1992 and 2000, despite significant differences in the 1-4 year age group. Other factors include Mozambican origin of mother for certain infectious causes, maternal death in first year of infant's life, lower maternal education, poor quality of and limited access to neonatal care, poor antenatal clinic attendance, and increased vulnerability of households with a high mortality burden. These factors should be better elucidated and quantified in order to contribute meaningfully to policy and programmes.

We examined health service access with respect to primary health care generally and antenatal care specifically. Distance to the nearest primary health care facility was not a risk factor in this study with regards to infant and child mortality. Antenatal clinic attendance and the number of antennal clinic visits were significantly protective, with no difference between South Africans and former Mozambican refugees. These findings suggest that factors other than geographic access may be the key to understanding the risks associated with health care utilisation. These could include the quality of care, level of available care (primary versus secondary), cost and social barriers. In South Africa, primary health care for children under the age of six is free, as is antenatal care. However, financial costs associated with transport and opportunity costs associated with lengthy waiting time (Bigdeli and Annear, 2009) are

some of the barriers described in this setting (Twine et al., 2007; Goudge et al., 2009). Twine el al. (2007) showed that the poorest households were less likely to apply for social support grants than those in higher socioeconomic strata due to barriers such as distance from government offices, lack of official documentation and education of caregiver and household head.

The strong clustering of diarrhoeal or malnutrition-related (food security) mortality risk in former Mozambican refugee settlements suggests worse water and sanitation infrastructure facilities which need the attention of local government. Previous research has shown a strong link between poor nutrition and infection (Katona and Katona-Apte, 2008). Oral rehydration therapy (ORT) and dietary management are key aspects in the treatment of acute diarrhoea, particularly those episodes which persist. The capabilities and resources of health facilities to effectively manage children with diarrhoea and/or malnutrition needs to be strengthened as do referral systems where necessary.

Adults

Adult survival is influenced by several factors examined in this study. Low SES and few employment opportunities locally have led to adults migrating externally for work. Migration patterns have been shown to influence HIV risk (Lurie, 2000). A study in Agincourt found that over 90 percent of men perceived little or no personal risk of HIV infection (Collinson et al., 2006). Males were much more likely to be labour migrants than females. The level of reported risk behaviour among migrants depended on the frequency of return and those long-distance migrants who return once or twice

a year report more partners than those who worked in nearby destinations (Collinson et al., 2006). They also found, however, that resident employed men also reported more partners and high levels of male labour migration, coupled with a low frequency of long-distance migrants returning home and low personal HIV risk perception, are potentially contributing to an explosive spread of HIV in this and other rural settings (Collinson et al., 2006). In this regard, they suggest that strategies (though very difficult to implement in reality) to enable more frequent contact between migrant men and their rural families are urgently needed as are prevention and awareness raising activities (Collinson et al., 2006).

Mozambican nationality was not a risk factor in the 15-49 age-group following multivariate adjustment, likely due to the adjustment for lower education and SES in the model which is more evident among Mozambicans and settlements further away from health facilities. The risk for adult mortality was higher in poorer villages suggesting that the benefits of improved health care are not evenly distributed throughout the study area. Another study in South Africa on the impact of adult mortality on household dissolution and migration in rural Kwazulu Natal, suggests that poorer households and households affected by adult deaths were more vulnerable (Hosegood et al., 2004).

Mozambicans in the older age group had a significantly lower risk of mortality when compared to South Africans. This suggests that, given the generally lower SES of Mozambican households, lifestyle-related mortality among South African households may be driving risk for mortality due to non-communicable diseases. A study in this

area showed that non-communicable disease remains evident despite the pronounced impact of HIV/TB (Tollman et al., 2008).

Overall and age-specific mortality risk was higher in households headed by women and where the head was less than 40 years of age. In our study, adult mortality was significantly higher for males than that for females and the spatial pattern of male and female mortality risk was identical, in contrast to a study in China which found different spatial patterns of mortality risk by gender (Ali et al., 2007). A cycle of increased risk of male death (migrancy and lifestyle related disease in 50-64 year olds), leading to more female-headed households is likely. Given that a male household head conferred a survival advantage to adults and that male migrants are at increased risk, this could potentially compound adult mortality over time as household head dynamics change. This needs to be assessed in more detail, along with the impact of adult mortality on orphans and elderly mortality. Female-headed households require support since they appear to be more vulnerable and at higher risk for adult mortality.

7.4.2 Distance to nearest health facility: risk for age-specific mortality

Several studies relate geographic access to the use of health facilities. Members of distant communities use facilities less than those living nearer, but this does not necessarily translate into increased mortality risk (Stock, 1983; Becher et al, 2008). A recent study in Kenya found that, despite significant spatial variations in child mortality, these were not correlated with distance to health facilities (Moïsi et al., 2010). They concluded that geographic access to curative services did not influence

population-level mortality given the density of health facilities in Kenya. They also suggest that when distance access targets are met, further improvements in child survival can only be achieved through renewed investigation of the social, behavioural and quality-of-care factors that obstruct access to health services. Similarly, in rural South Africa, there is an urgent need to evaluate and assure a high level of health service quality; assess and strengthen referral patterns for emergency obstetric, infant and child health care; and identify other barriers to accessing these and other government services. In this study no significant change in risk was found with increasing distance of household from nearest primary health care clinic or district hospitals. The same holds true for infant mortality (Sartorius et al., 2010b). This suggests that quality of health services may be influencing child mortality more than geographic access. Evaluation of primary health care services with attention to quality improvements is needed.

Adult mortality hotspots appear to be affected by differential health care access in this rural setting. Larger distance from the nearest health facility had a significantly higher risk associated with adult mortality. This has been shown in a study on adult mortality in China (Ali et al., 2007). Specific foci of higher mortality risk based on Bayesian kriging which included distance from nearest health facility highlighted specific villages at increased adult mortality risk. Given that increased distance to nearest health facility was a significant risk, this has major implications for ART rollout and inequalities with regards to access. The impact of the ART rollout which started in this area in 2007 and proximity to the more distant district health facilities also needs to be accessed in more detail in future studies.

7.4.3 Methodological comparison and future work

A summary of the advantages and disadvantages of the different methods used are discussed in Table 13. Future work that builds on current work includes the following:

There is a need to identify the most efficient, accurate and simple way to correctly assess risk factors for a given outcome (mortality, divorce etc) using correlated longitudinal HDSS data. Firstly, this relates to the best temporal or longitudinal data structure to use e.g. various discrete time lengths or continuous time with time varying breaks. Secondly, to assess which is the best routine modelling approach that should be used e.g. clustered, generalized linear, generalized linear latent and mixed or random effects modelling. Longitudinal data (such as HDSS) are correlated both in space (closer households are more likely to be similar) and time (repeated measurement on individuals or households). If this correlation is not taken into account, the standard errors of the risk factors can be underestimated thereby over estimating significance. Various methods exist, from simple to very complex, that allow one to adjust for this correlation when performing the analysis. However some of the more complicated techniques (such as Bayesian geostatistical and temporal autoregressive modelling) can take very long time to run. The proposed next steps would be to use basic datasets, both simulated and actual, to identify the best temporal structure and modelling approach for HDSS longitudinal data which best estimates the magnitude and significance of the risk factors in terms of accuracy and computational time for example. The major objective would be to make a recommendation regarding which approach and data structure could best be used routinely, by the INDEPTH network for example, and in computationally limited settings.

More advanced analysis of distance to the nearest health facility as an age-specific mortality risk in Agincourt HDSS was developed using a proper road network analysis. Previous studies in Agincourt and other HDSS that have assessed the impact on distance to nearest health facility as a risk factor for mortality have generally limited the calculation to a straight-line Euclidean distance from the individual's household. In many instance no association has been found and this may be purely due to how this measurement was defined i.e. straight-line. The straight-line distance suffers from limitations that have been described in the literature (Noor et al., 2006). For example this measurement assumes that individuals travel in a straight line and use the nearest health facility, which is often not the case as the patients' actual use characteristics may differ substantially (Gething et al. 2004; Guargliardo et al. 2004)". More complex, as well as accurate measurements (for example a transport network model which accounts for road type, topography and barriers (Noor et al., 2006) exist to estimate the shortest distance to quantify this predictive factor. Future analyses of the physical access to the nearest health facility or external district hospital on mortality in Agincourt and other HDSS, should utilise these more advanced and accurate approaches.

There is also a need to better relate the spatial-temporal distribution of the identified risk factors with the observed space-time anomalies or hot and cold spots. Developing attributable fractions of specific risk factors to individual or aggregated high risk clusters and how the distribution of this varies within and between villages and/or the other spatial units of aggregation is needed. This could also allow one to more efficiently quantify significant differences in the distribution of predictors or risk

factors between 'hot-hot' and 'hot-cold' spots, thereby strengthening policy applications.

Table 13: Comparison of the advantages and disadvantages of various longitudinal survival modelling approaches

Category	**Unrepeated measures [a]: Standard binary logistic (or Poison or Negative Binomial) model (with time offset)**	**Repeated measures [b]: Discrete time (monthly, quarterly or yearly) event history model**	**Repeated measures [b]: Continuous time survival or hazard model**
Advantages	Computationally simple Suitable for age groups with short time span or unrepeated measures (e.g. infants) or cross sectional data	Straightforward handling of time changing covariates Can allow for non-proportional hazards / easily allow for unstructured and structured estimation of the hazard function at each discrete time point (Muthen and Masyn, 2005) More transparent than continuous time methods (Steele, 2005) Essentially logistic regression models which are familiar to most disciplines and more easily interpreted Often most useful and natural in many settings (social and behavioural science) where time is most likely to be measured discretely e.g. school years (Muthen and Masyn, 2005 More capable of handling or analysing ties than continuous time modelling approaches (Box-Steffensmeier and Jones, 1997) Approximation of a continuous time process improves as the discrete intervals become smaller (Yamaguchi, 1991)	Frequently used in many settings (Hougaard, 2000) Truly dynamic analysis – risk of experiencing an event at a certain time point Can deal with censored observations Covariates may change value during the observation period

Disadvantages	Not suitable for multiple or repeated measures per individual (long time spans) Can't incorporate time varying covariates	Loss of precision of event time in longer discrete time periods used (averaging of risk over the period) Creating person-period dataset: --complex data manipulation --large datasets especially as discrete time period shortens (e.g. monthly) --long computation times for complex random effects models	If censoring mechanism related to process or outcome under study then bias introduced (non-random censoring) Complex to split continuous time episodes to reflect all covariate changes Assumption that only one event can occur at any given time is problematic when durations are measured in reasonably broad time intervals i.e. can't easily handle co-occurrence of events (Yamaguchi, 1991). Estimation procedures for continuous-time models thus need to be adapted if there are tied event times (Hosmer and Lemeshow, 1999)

a: overall probability of an event
b: longitudinal progression of the probability that event will occur

7.5 Distribution of underlying risk factors and implications for policy

Key findings

- Proportion of Mozambicans, number of household deaths and levels of poverty were significantly elevated in high risk child and younger adult mortality clusters
- Concentration of a younger migrant population in upper central region driving higher child and adult HIV/TB mortality
- Targeting the RDP village in the upper central region of the site (emerged as at high risk for mortality and HIV especially in the later period). This is a highly mobile sub-population with a high number of circular migrants. The blanket intervention could be health education messages that focus on high risk sexual behaviour when away increasing risk for HIV infection, the risk this poses to their partners when they return, and the consequent risk of vertical transmission to future children arising from this partnership. This village should also be targeted to ensure proper antenatal support to address risk of vertical transmission and PMTCT. The loss of 'bread winners' also further impoverishes these household which subsequently has indirect impacts on their children.
- Blanket up-scaling for entire sub-district: ensuring equitable and high quality antenatal care, PMTCT to prevent vertical transmission, and go identify HIV positive mothers for ART treatment.
- Given the high concentration of diarrhoeal and malnutrition-related child mortality in a cluster of four villages in close proximity in the south-east corner, there is urgent need to assess and scale-up water and sanitation facilities in these communities

This study should be regarded as a first step in prioritizing areas for follow-up public health efforts and evaluating their impact: targeting of vertical prevention of HIV/TB and antiretroviral rollout in significant child and adult mortality clusters; spatial

assessment of antiretroviral therapy (ART) rollout that started in this area in 2007 along with any villages or areas not accessing equitably; and assessment and provision of adequate water and sanitation in the child mortality clusters particularly in the south-east where diarrhoeal mortality appears high. Mother-to-child HIV transmission prevention efforts in areas with high child mortality needs to be emphasized, along with other interventions.

As discussed earlier there is generally a higher concentration of Mozambicans to the east and south east regions of the sub-district which proximately explains part of the observed or predicted space and space-time mortality risk. It also appears that the burden of communicable disease mortality (specifically HIV/TB and diarrhoea) is highest in these areas, thus leading to the all-cause space and space-time findings. Settlements comprising former Mozambican refugees, as well as selected South African villages appear to have an increased risk of and infectious-related mortality. Suitable interventions such as ART and the assessment and provision of adequate water and sanitation need to be directed to these villages to address existing inequalities. A higher cluster of old adult mortality (50-64) was observed in the west of the site and older Mozambicans had a lower risk of mortality when compared to their South African counterparts. Thus, the South Africans' higher standard of living may be contributing to a relatively higher spatial risk of non-communicable disease. This will be investigated in future.

Addressing health inequities in populations is a major challenge (Feachem, 2000), and research that documents and quantifies inequities is needed to inform policies to close health gaps in the developing world. Evidence on reducing inequities within countries

is growing; successful approaches include those that improve geographic access to health interventions in poor communities, subsidize health care and health inputs for the poor, and empower poorer communities (Gwatkin et al., 2004; Marmot 2005). The results of our study indicate the need for interventions in villages to the east of the site, many of which have a large proportion of former refugees, to reduce the higher burden of infant deaths due to infectious and parasitic causes. HAART for HIV began in 2007 in this district; hence its impact cannot be captured during the time frame of this study. This research does, however, provide useful insight into spatial-temporal mortality patterns before HAART rollout and will allow post-rollout assessment of its impact on infant mortality. Such evaluation has the potential to identify areas needing improved access to treatment, specifically prevention of mother-to-child transmission and anti-retroviral therapy.

Of concern is the high number of neonatal deaths (particularly in the perinatal period), their gradual increase over the study period, and the highest risk area being in close proximity to a health facility. This suggests problems of service quality rather than geographic access, and highlights the need to assess and improve the capacity of sub-district health facilities for antenatal, emergency obstetric and newborn care; improve coverage of deliveries by skilled birth attendants; and advise mothers on appropriate care-seeking for sick babies. Part of the perinatal mortality burden observed may relate to maternal HIV since the same village experienced the highest risk for neonatal and infant mortality. A meta-analysis (Brocklehurst and French, 1998) found an association between maternal HIV infection and adverse perinatal outcomes, including low birth weight and pre-term delivery.

When comparing the high risk age-specific village clusters combined as identified by the Kulldorff Spatial Scan statistics versus the remainder of the villages, significant differences emerged with regards to the distribution of underlying risk or protective factors.

A complex interaction of factors appears to be driving adult mortality in space-time. Villages with higher HIV, large distance from nearest health facility, low SES, low education, high household mortality burden, and high migrancy rates (increased risk behaviour) appear to be driving communicable disease mortality particularly HIV. Mozambican nationality was protective in the older age-group (50-64) indicating that the generally lower SES of Mozambicans may be protective relative leading to more affluent South Africans in terms of lifestyle related non-communicable disease mortality. The risk maps can guide decision makers regarding an ART rollout, specifically to six higher risk villages. Given that increased distance to nearest health facility was a significant risk, this has major implications for ART rollout and perpetuating inequalities with regards to access.

One village in the upper central region of the site repeatedly emerged as a significant hotspot of mortality. This village had the highest number of migrants and migrant months compared to other villages. It is also a "Reconstruction and Development Programme (RDP)"[10] village with a highly mobile population within the sub-district. Temporary or labour migrants are more vulnerable to HIV than more settled populations and this has been shown in other African and southern African countries

[10] The Reconstruction and Development Programme (RDP) is a South African socio-economic policy framework implemented by the African National Congress (ANC) government. One component involves replacing all shacks and informal settlements in South Africa by providing low cost housing for the poor.

(Brockerhoff and Biddlecom, 1999; Lurie at al., 2003). A complex interaction of factors that surround this mobility predisposes the migrant (and thus his/her partner or partners) to greater risk of acquiring HIV. This has been shown to be due to poverty, high sexually transmitted infection (STI) rates, the presence or absence of circumcision, high partner change rates and increasing the number of higher risk partners (commercial sex workers) (Williams et al., 2002; White, 2003). According to Lurie (2003) migrant men are much more likely to have multiple sexual partners and engage in high-risk behaviour.

The findings indicate the need for more detailed research regarding the underlying risk factors (at individual, household or community level) that may drive the observed spatial-temporal all-cause and cause-specific mortality patterns. For example when planning interventions for both communicable and non-communicable mortality, one should disentangle the intricate set of factors that differentiate settled South Africans, former Mozambican refugees and labour migrants.

7.6 Limitations

A limitation of the study is the potential to miss infant deaths, particularly neonatal deaths, which would underestimate the overall infant mortality burden. Infants that are born and then die during the 12 months between HDSS census update rounds may not be reported, particularly if the mother migrated out of the household; similarly, death among in-migrant infants who die before they are enumerated in the annual household census may be missed. However, infant death ascertainment has improved in the

study site (Kahn, 2006), and the proportion of infants who were in-migrants decreased significantly over time, reducing the bias towards the end of the study period.

Determination of cause-of-death through verbal autopsy is more problematic for diseases that have less specific symptoms such as HIV/AIDS (Garenne and Fauveau, 2006). The prevalence of HIV infection in a population and the resulting rate of HIV-associated co-morbidity and death due to malnutrition in children, for example, may affect the performance (such as specificity) of the tool. Thus, it is likely that the HIV burden is underestimated due to the misclassification of deaths as AIDS-related conditions such as malnutrition or diarrhoea, or their being placed in the "unknown cause" category. The significant increase in the number of infant deaths attributed to unknown causes since the late 1990s (Figure 2) is concurrent with the rise in HIV-related mortality in the area. Levels of stigma associated with HIV are high in South Africa, particularly prior to the introduction of HAART. The ability to make a diagnosis on VA depends, in large part, on the quality of information provided by the respondent. This may have been compromised in some cases in an effort to disguise HIV as a likely cause-of-death, partly explaining the increase in unknown causes.

Physician-coded verbal autopsies have known limitations (Murray et al., 2007) and misclassification could have occurred in our data, especially with regards to underestimating non-specific HIV/AIDS-related mortality. One limitation of this study is the HIV/TB related deaths misclassified by VA as unknown (R99), as demonstrated by the significant increasing trends in this classification over time (Figures 7,9,11), which would underestimate the true burden and also suggest that a physician-coded verbal autopsy relies heavily on household recall of medical records

and related information, limiting its applicability in low-resource settings (Murray et al., 2007). However, a previous validation study of the VA in Agincourt HDSS has shown that it performs well in this high HIV prevalence setting (Kahn et al., 2000). These found that for HIV/TB combined the sensitivity, specificity and PPV were all high (78, 80 and 85 percent respectively). Other studies have also confirmed that VA data can be used to reasonably estimate the distribution of AIDS- and non-AIDS-related deaths even in a rural population with relatively low levels of education (Doctor and Weinreb, 2003). One area that needs to be addressed is how the tool could be modified to take into account the influence of HIV/AIDS when estimating the sensitivity for conditions such as diarrhoea, ALRI/pneumonia and malnutrition that are unrelated to HIV/AIDS.

8.0 Conclusions

Demographic surveillance systems provide a viable method for the collection of reliable data on vital events in rural sub-Saharan Africa. They add value given the paucity or lack of reliable routine mortality data or statistics in Sub-Saharan Africa, which often renders important health or mortality related issues, along with their determinants, invisible to policymakers and donors. Such estimates are critical for the design and implementation of effective public health programmes in rural sub-Saharan Africa including South Africa. The value of HDSS data have been demonstrated and the results of the thesis have provided valuable information for service planning and prioritization in this and other rural areas of South Africa. The high burden of HIV in this population (leading cause-of-death), as is the case with most of South Africa, is clearly evident. By estimating the true spatial and temporal distribution of the age-specific mortality burden in rural northeast South Africa, this study has shown variation across a relatively small geographical area. Various spatial analysis techniques (ranging from very simple to highly advanced) were employed and their relative strengths and weaknesses identified with recommendation for future research to progress forward. The thesis employed Bayesian geostatistical models in order to identify risk factors, correctly estimated the standard errors (significance) of these risk factors and produced smoothed maps of age-specific mortality risk from spatially correlated longitudinal mortality data in a health and socio-demographic surveillance system. Findings indicate the need for interventions targeted at villages with excess age-specific mortality risk due to both a direct and indirect impact of HIV. A few essential interventions include improving prevention of mother-to-child transmission programmes, and antiretroviral therapy for HIV positive mothers to

ensure their survival during their infants' and child's critical first year(s) of life, targeting and improving access to various services for vulnerable households (for example households with high mortality burden, suffered shock of key adult(s) deaths or loss of pivotal household head), adult education (with regards to unsafe sex) and improved testing of highly mobile or migrant individuals. From our study, it is clearly inadequate to consider maternal health separately from infant and neonatal health. This is consistent with other studies which showed that maternal health directly affects infants' health (Newell et al., 2004). Policy should thus have greater emphasis on interventions targeting the mother-infant pair. We also conclude that the non-random clustering of infant mortality due to diarrhoea and malnutrition in the south-east part of the site represents a breakdown in basic services (or, indeed, their absence); there is hence need to assess and improve water and sanitation in these villages. The high level of perinatal mortality, in some instances in close proximity to health facilities, is of concern, indicating a need to strengthen the capacity of sub-district facilities for emergency obstetric and newborn care. Recommendations from this study will have applications to other rural settings within South Africa and potentially beyond.

Based on the space-time analysis, the southeast and upper central regions of the site appear to have the highest mortality risk. These maps are particularly helpful in identifying high mortality areas to guide efficient allocation of limited resources in child survival and other programs. Risk factor results can contribute to policies to address health inequalities and improve access to health services. Targeted efforts to prevent the vertical transmission of HIV in specific villages needs to be undertaken, as well as efforts to promote the survival of mothers and fathers, both which emerged

as prominent risk factors for child mortality. The distribution of adult mortality is driven by a complex web of interacting factors that have likely increased communicable disease mortality (HIV/TB) and non-communicable disease mortality (in the older adult age-group) in specific risk areas.

The impact of HIV on mortality dynamics within this rural setting is striking. The findings of this study indicate that particular villages, households and individuals, are at higher mortality risk with significant differences in their risk factor profiles. Risk maps can be used by decision makers for the design and implementation of interventions to alleviate this burden and reduce disparities. Interventions that target the mother-infant pair and increase access to various services for more vulnerable "high mortality" households are needed. Important interventions include PMTCT, ART rollout, water and sanitation, and screening for and control of non-communicable disease risk factors. Increased distance to nearest health facility, a significant risk factor among adults, highlights the importance of geographical access to ART rollout. The strong concentration of diarrhoeal and malnutrition mortality in children represents a breakdown or absence of basic services, such as provision of water and sanitation, that needs to addressed. Recommendations from this study have implications for other rural settings within South Africa and potentially beyond.

Our study has demonstrated the considerable potential of spatial statistical methods for analyzing HDSS event data; this is a contribution since few studies have used geostatistical modelling on HDSS event data (e.g. mortality). This study has also demonstrated that space-time methods can be used to identify anomalies of a given outcome (in this case mortality) in time which further helps guide policy and intervention both as to where the problem is and when it occurs. Within the resources in the study setting we have also demonstrated that simple to advanced methods can be applied that detect significant hotspots of a given outcome in space and time. These methods can and have been applied to non-disease outcomes in this setting such as poverty (Sartorius et al., 2011) and household dissolution (Sartorius et al., 2011 in press). These methods can also be easily extended to other areas (INDEPTH centres, national census data) and to other social problems (marriage dissolution). Future post-doctoral work will compare various space-time methods (including more recent advances) on actual and simulated data in order to recommend routine use both in INDEPTH and in the broader public health arena.

The analytical determinant models used in this thesis have definite applications to other studies within this site (Sartorius et al., 2011), in the wider INDEPTH network as well as other longitudinal datasets. Future post-doctoral work will focus on identifying and making recommendations with regards to which would be the best for routine use and given computational/resource limitations in certain African settings.

Finally a better integration of the two methodologies above, which I touched on briefly in this thesis, will be a key component that should prove invaluable for policy

makers. That is “knowing the where, when and why in specific locations (e.g. households and villages)” that will allow for highly structured and targeted interventions as well as policy programs.

Acknowledgements

I would like to thank my supervisors, Professor Kathleen Kahn and Dr. Penelope Vounatsou, for providing expert support in the different components of my PhD. In a personal capacity, many thanks Kathy for your guidance, mentoring and support throughout this process, I learnt a lot and have emerged on the other side a better researcher and writer I hope. It was a pleasure to have you as a supervisor. I would also thank my dad, Professor Kurt Sartorius, though not his field of expertise (though I suspect he now knows more about mortality modelling than he might have wished to), for his common sense, help and guidance during my PhD.

This work was supported by a PhD fellowship from the South African Centre for Epidemiological Modeling and Analysis (SACEMA), a National Research Foundation (NRF) Centre of Excellence, as well as research fellowships from the MRC/Wits Rural Public Health and Health Transitions Research Unit (Agincourt) through the Wellcome Trust (grant no. 069683/Z/02/Z), UK, and the Swiss-South Africa Joint Research Programme (project no. JRP IZLSZ3_122926). Additional funding was provided by a Wits Faculty of Health Sciences Medical Research Endowment Fund (MREF) (grant no. SARB000). The Agincourt HDSS and GIS was funded by the Wellcome Trust, UK (grant no. 058893/Z/99/A, 069683/Z/02/Z, 069683/Z/08/Z), the University of the Witwatersrand, the South African Medical Research Council, and

the Andrew W. Mellon Foundation, USA. P. Vounatsou was supported by the Swiss National Science Foundation (project no. 325200_118379).

References

Adazu K, Lindblade KA, Rosen DH, Odhiambo F, Ofware P, Kwach J, Van Eijk AM, Decock KM, Amornkul P, Karanja D, Vulule JM, Slutsker L. (2005). Health and demographic surveillance in rural western Kenya: a platform for evaluating interventions to reduce morbidity and mortality from infectious diseases. Am J Trop Med Hyg.; 73(6):1151-8.

Adebayo SB, Fahrmeir L, Klasen S (2004) Analyzing infant mortality with geoadditive categorical regression models: a case study for Nigeria. Econ Human Biol 2: 229–244.

Adetunji J (2000). Trends in under-5 mortality rates and the HIV/AIDS epidemic. Bulletin of the World Health Organisation 78, 1200-1206.

Adjuik M, Smith T, Clark S, Todd J, Garrib A, Kinfu Y, Kahn K, Mola M, Ashraf A, Masanja H, Kubaje A, Sacarlal J, Alam N, Marra A, Gbangou A, Mwageni E, Binka F (2006). Cause specific mortality rates in developing countries. Bulletin of the World Health Organization; 84:181-188.

Adomakoh S, Kumar A, St. John A (2002). Reducing mother to child transmission of HIV-1 in Barbados. Cost-effectiveness of the PACTG 076 protocol in a middle income, low prevalence setting. International Conference on AIDS. Int Conf AIDS. Jul 7-12

Ahmad OB, Lopez AD, Inoue M (2000). The decline in child mortality: a reappraisal. Bulletin of the World Health Organisation 78, 1175–91.

Alexander F, Boyle P (1996). Methods for investigating localized clustering of disease, IARC Scientific Publications No 135, Lyon.

Ali M, Jin Y, Kim DR, De ZB, Park JK, Ochiai RL, Dong B, Clemens JD, Acosta CJ (2007). Spatial risk for gender-specific adult mortality in an area of southern China. Int. J. Health Geogr.;6:31.

Anselin L (1995) Local indicators of spatial association--LISA. Geographical Analysis, 27:93-115.

Anselin L, Florax R, Rey S (2004). Advances in spatial econometrics: methodology, tools and applications. Springer-Verlag Berlin and Heidelberg GmbH and Co. KG.

Anselin L, Griffiths E, Tita G (2008). Crime mapping and hot spot analysis. In: Wortley R, Mazerolle L (eds) Environmental criminology and crime analysis. Willan Publishing, Cullompton, Devon, pp 97–116

Anselin L (2010). Thirty years of spatial econometrics. Papers in Regional Science 89: 3-25. DOI: 10.1111/j.1435-5957.2010.00279.x

Antunes J, Waldman E (2002). Trends and spatial distribution of deaths of children aged 12–60 months in Sa˜ o Paulo, Brazil, 1980–98. Bulletin of the World Health Organization 2002; 80:391-398.

Arbia G, Baltagi B (2009). Spatial Econometrics. Series: Studies in Empirical Economics. Physica-Verlag, Heidelberg.

Armstrong M, Rushton G and Zimmerman D (1999). Geographically masking health data to preserve confidentiality. Statistics in Medicine; 18: 497–525.

Baiden F, Hodgson A, Adjuick M, Adongo P, Ayaga B and Binka F (2006). Trend and causes of neonatal mortality in the Kassena-Nankana district of northern Ghana, 1995-2002. Tropical Medicine and International Health; 11: 532-539.

Bailey TC, Gatrell AC (1995). Interactive spatial data analysis. Harlow: Longman, 1995.

Baker MC, E Mathieu, F M Fleming, M Deming, J D King, A Garba, J B Koroma, M Bockarie, A Kabore, D P Sankara, D H Molyneux (2010). Mapping, monitoring, and surveillance of neglected tropical diseases: towards a policy framework. Lancet 2010; 375(9710):231-8.

Balk D, Pullum T, Storeygard A, Neuman M (2004). A spatial analysis of childhood mortality in West Africa. Popul Space Place 10: 175–216.

Balk B, Pullum T, Storeygard A, Greenwell F, and Neuman M (2003). Spatial Analysis of Childhood Mortality in West Africa. Calverton, Maryland, USA: ORC Macro and Center for International Earth Science Information Network (CIESIN), Columbia University.

Banerjee S, Carlin B and Gelfand A (2004). Hierarchical Modelling and Analysis for Spatial Data. Chapman and Hall/CRC.

Bao, Shuming and Mark S. Henry (1996). Heterogeneity issues in local measurements of spatial association.Geographical Systems, 3: 1-13.

Barrett FA (2000). Finke's 1792 map of human diseases: the first world disease map? Social Science and Medicine 50(7-8):915-21.

Basáñez MG, Marshall C, Carabin H, Gyorkos T (2004). Bayesian statistics for parasitologists. Trends Parasitol 20, 85-91.

Becher H, Müller O, Jahn A, Gbangou A, Kynast-Wolf G, Kouyaté B (2004). Risk factors of infant and child mortality in rural Burkina Faso. Bull World Health Organ.; 82(4):265-73.

Becher H, Kynast-Wolf G, Sie A, Ndugwa R, Ramroth H, Kouyate B, Müller O (2008). Patterns of malaria: cause-specific and all-cause mortality in a malaria-endemic area of west Africa. Am J Trop Med Hyg 78, 106-113.

Becher H (2010). Guest Editorial: Analyses of mortality clustering at member HDSSs within the INDEPTH Network – an important public health issue. Glob Health Action; 3: 10.3402/gha.v3i0.5476.

Belgrano A (1998). Spatial Ecology: The Role of Space in Population Dynamics and Interspecific Interactions. - Review - book reviews. Ecology.

Bell BS (2002). Spatial analysis of disease--applications. Cancer Treat Res. 2002;113:151-82.

Benjamin M, Rigby R, Stasinopoulos D (2003). Generalized autoregressive moving average models. J Am Stat Association; 98 [461]: 214-223.

Benzler J, Sauerborn R (1998). Rapid risk household screening by neonatal arm circumference. Results from a cohort study in Burkina Faso. Tropical Medicine and International Health; 3: 62-74.

Bernardinelli L, Clayton D, Pascutto C, Montomoli C, Ghislandi M, Songini M (1995). Bayesian analysis of space-time variation in disease risk. Stat Med; 14: 2433-2443.

Besag J (1974). Spatial interactions and the statistical analysis of lattice systems (with discussion). In Journal of the Royal Statistical Society, Series B, volume = 36, pp. 192-236, 1974.

Besag J, Newell J (1991). The Detection of Clusters in Rare Diseases. Journal of the Royal Statistical Society Series a-Statistics in Society; 154: 143-155.

Best N, Richardson S, Thomson A (2005). A comparison of Bayesian spatial models for disease mapping. Statistical Methods in Medical Research 14, 35-39.

Bigdeli M; Annear P (2009). Barriers to access and the purchasing function of health equity funds: lessons from Cambodia. Bull World Health Organ;87:560–564

Biggeri A, Marchi M (1995). Case-Control Designs for the Detection of Spatial Clusters of Diseases. Environmetrics; 6: 385-393.

Binka FN, Maude GH, Gyapong M, Ross DA, Smith PG (1995). Risk factors for child mortality in northern Ghana: a case-control study. International Journal of Epidemiology 24, 127-35.

Black R, Morris S, Bryce J (2003). Where and why are 10 million children dying every year? Lancet; 361: 2226-34.

Bland JM, Altman DG (1998). Bayesians and frequentists. BMJ; 317: 1151-1160.

Bledsoe C, Anastasia B (1992). Child fosterage and child mortality in sub-Saharan Africa: some preliminary questions and answers. In Mortality and Society in Sub-

Saharan Africa. Edited by: Etienne van de Walle, Gilles Pison and Mpembele Sala-Diakanda. Oxford: Clarendon Press.

Boots B (2002). Local measures of spatial association, Ecoscience 9(2): 168–176.

Borden K and Susan L Cutter (2008). Spatial patterns of natural hazards mortality in the United States. International Journal of Health Geographics; 7:64doi:10.1186/1476-072X-7-64.

Booysen, F. (2002). The socio-economic impact of HIV/AIDS on households in South Africa: Pilot study in Welkom and Qwaqwa, Free State Province. Working paper, University of the Free State.

Box-Steffensmeier J and Jones B (1997). Time is of the Essence: Event History Models in Political Science. American Journal of Political Science 41(4): 1414-1461.

Boveta P and Benhamou S (1988). Spatial analysis of animals' movements using a correlated random walk model. Journal of Theoretical Biology.; 131(4): 419-433. doi:10.1016/S0022-5193(88)80038-9

Bradshaw D, Timaeus I, Dorrington R, Nannan N (2000). Reversal in adult mortality trends in South Africa. International Conference on AIDS. Int Conf AIDS. ; 13: abstract no. MoPeD2508.

Brezger A, Kneib T, Lang S (2005) BayesX: Software for Bayesian Inference based on Markov Chain Monte Carlo simulation techniques. J Statist Software 14: 11.

Britton T (1997). Tests to detect clustering of infected individuals within families. Biometrics; 53(1): 98-109.

Brockerhoff M (1994). The impact of rural-urban migration on child survival. Health Transition Review; 4: 127–149.

Brockerhoff M and Biddlecom A (1999). Migration, sexual behaviour and the risk of HIV in Kenya. International Migration Review 33, 833-856.

Brockerhoff M, Hewett P (2000). Inequality of child mortality in sub-Saharan Africa. Bull World Health Organ; 78:30–41

Brocklehurst P, French R (1998). The association between maternal HIV infection and perinatal outcome: a systematic review of the literature and meta-analysis. Br J Obstet Gynaecol.; 105[8]: 836-48.

Brooker S, Clements ACA (2006), Bundy DAP. Global epidemiology, ecology and control of soil-transmitted helminth infections. Advances in Parasitology;62:221–261.

Brooker S, Hotez PJ, Bundy DAP (2010). The Global Atlas of Helminth Infection: Mapping the Way Forward in Neglected Tropical Disease Control. PLoS Negl Trop Dis 4(7): e779. doi:10.1371/journal.pntd.0000779

Budlender D, Proudlock P and Jamieson L (2008). Developing social policy for children in the context of HIV/AIDS: A South African case study. Children's Institute Case Study Number 3. Cape Town: Children's Institute, University of Cape Town, and Community Agency for Social Enquiry.

Burke P (1988). Calculating the area and centroid of a polygon. URL available at - http://paulbourke.net/geometry/polyarea/

Byass P, Fottrell E, Huong DL, Berhane Y, Corrah T, Kahn K, Muhe L, Van DD (2006). Refining a probabilistic model for interpreting verbal autopsy data. Scandinavian Journal of Public Health; 34(1):26-31.

Byass P, Huong DL, Minh HV (2003). A probabilistic approach to interpreting verbal autopsies: methodology and preliminary validation in Vietnam. Scandinavian Journal of Public Health; 31(Suppl.62):32-37.

Chaikaew N , Tripathi N and Souris M (2009). Exploring spatial patterns and hotspots of diarrhea in Chiang Mai, Thailand. International Journal of Health Geographics; 8:36doi:10.1186/1476-072X-8-36

Chainey S, Ratcliffe JH (2005). GIS and crime mapping. Wiley, London

Chen J, Roth R, Naito A, Lengerich E, MacEachren AM (2008). Geovisual analytics to enhance spatial scan statistic interpretation: an analysis of U.S. cervical cancer mortality. International Journal of Health Geographics, 7:57

Clarke KC, McLafferty SL, Tempalski BJ (1996). On epidemiology and geographic information systems: a review and discussion of future directions. Emerg Infect Dis; 2: 85-92.

Cleland J, van Ginneken JK (1989). Maternal education and child survival in developing countries: the search for pathways of influence, Social Science and Medicine 27, 1357–1368.

Clements AC, Lwambo NJ, Blair L, Nyandindi U, Kaatano G, Kinung'hi S, Webster JP, Fenwick A, Brooker S (2006). Bayesian spatial analysis and disease mapping: tools to enhance planning and implementation of a schistosomiasis control programme in Tanzania. Trop Med Int Health.;11(4):490-503.

Clements A, Brooker S, Nyandindi U, Fenwick A, and Blaird L (2008a). Bayesian spatial analysis of a national urinary schistosomiasis questionnaire to assist geographic targeting of schistosomiasis control in Tanzania, East Africa. Int J Parasitol.; 38(3-4): 401–415.

Clements AC, Barnett AG, Nyandindi U, Lwambo NJ, Kihamia CM, Blair L (2008b). Age and gender effects in self-reported urinary schistosomiasis in Tanzania. Trop Med Int Health.;13(5):713-21.

Clements ACA, Kur LW, Gatpan G, Ngondi JM, Emerson PM, et al. (2010a). Targeting Trachoma Control through Risk Mapping: The Example of Southern Sudan. PLoS Negl Trop Dis 4(8): e799. doi:10.1371/journal.pntd.0000799

Clements A, Deville MA, Ndayishimiye O, Brooker S, and Fenwick A (2010b). Spatial co-distribution of neglected tropical diseases in the East African Great Lakes region: revisiting the justification for integrated control. Trop Med Int Health. ; 15(2): 198–207.

Cliff A. and Ord, J.K (1973). Spatial Autocorrelation. Pion, London.
Cliff, A. D. and Ord, J. K (1981). Spatial Processes: Models and Applications. Pion, London.

Clifford WB, Brannon YS (1985). Rural–urban differences in mortality. Rural Sociology.; 50:210–224.

Cline B (1970). New eyes for epidemiologists: aerial photography and other remote sensing techniques. American Journal of Epidemiology.;92:85–89.

Cohen M (2000). Changing patterns of infectious disease. Nature; 406: 726-767.

Collins DL, Leibbrandt M (2007). The financial impact of HIV/AIDS on poor households in South Africa. AIDS.;21 Suppl 7:S75-81.

Collinson MA, Tollman SM, Garenne M, Kahn K (2001). Temporary female migration and labour force participation in rural South Africa. A working paper of the Agincourt Health and Population Unit, University of the Witwatersrand; A working paper of Princeton University's, Center for Migration and Development.

Collinson MA, Mokoena O, Mgiba N, Kahn K, Tollman SM, Garenne M, Shackleton S, Malomane E (2002). Agincourt Demographic Surveillance System (Agincourt DSS). In: Sankoh O, Kahn K, Mwageni E, Ngom P, Nyarko P (eds). Population and health in developing countries - Volume 1: Population, health and survival at INDEPTH sites. Ottawa: IDRC:197-206.

Collinson MA, Tollman SM, Kahn K, et al (2005). Highly prevalent circular migration: households, mobility and economic status in rural South Africa. In: Tienda M, Findlay SE, Tollman SM, et al., eds. African migration in comparative perspective. Johannesburg: Wits University Press.

Collinson, M., Lurie, M., Kahn, K., Johnson, A., and Tollman, S. (2006). Health Consequences of migration: Evidence from South Africa's Rural North-East (Agincourt). In: Tienda, M., Findley, S., Tollman, S., and Preston-Whyte, E. (eds.). Africa on the move. African Migration and Urbanisation in Comparative Perspective. Johannesburg: Wits University Press: 308-328.

Collinson MA (2010). Striving against adversity: the dynamics of migration, health and poverty in rural South Africa. Glob Health Action. 2010; 3: 10.3402/gha.v3i0.5080.

Congdon P (1997). Bayesian models for spatial incidence: a case study of suicide using the BUGS program. Health and Place
3(4): 229-247.

Congdon P (2003). Applied bayesian modelling. Chapter 7 - Models for spatial outcomes and geographical association. John Wiley and Sons.

Cook I, Alberts M, Burger S, Byass P (2008). All-cause mortality trends in Dikgale rural South Africa, 1996-2003. Scand J Public Health.;36(7):753-60. Epub 2008 Jul 22.

Cooke G, Tanser F, Bärnighausen T, Newell ML (2010). Population uptake of antiretroviral treatment through primary care in rural South Africa. BMC Public Health; 10:585.

Cooper RS, Osotimehin B, Kaufman JS, Forrester T (1998). Disease burden in sub-Saharan Africa: what should we conclude in the absence of data? Lancet 351: 208–210.

Coovadia H, Bland R (2007). Preserving breastfeeding practice through the HIV pandemic. Tropical Medicine and International Health 12, 1–18.

Coovadia H, Rollins N, Bland M, Little K, Coutsoudis A, Bennish M, Newell ML (2007). Mother-to-child transmission of HIV-1 infection during exclusive

breastfeeding in the first 6 months of life: an intervention cohort study. The Lancet; 369: 110

Coppock J, Rhind D (1991). The history of GIS. In Maguire, D.]., M. F. Goodchild, and D. W. Rhind, eds. GeograPhical information systems, principles and applications, vol. I. Harlow, Essex, U.K.: Longman.

Cressie, NAC (1993). Statistics for Spatial Data, Wiley & Sons, NY.

Cressie N and Huang HC. Classes of nonseparable spatio-temporal stationary covariance functions. In Journal of American Statistical Association, volume 94, pp. 1330-1340 1999.

Cromley, EK, McLafferty, SL (2002). GIS and Public Health. The Guilford Press, New York, London: pp.130-157.

Curtis A, Mills J, Agustin L and Cockburn M (2010). Confidentiality risks in fine scale aggregations of health data. Computers, Environment and Urban Systems; 35(1): 57-64.

le Seur D, Ngxongo S, Sharp B, et al (1997). Towards a spatial rural information system. Durban: Health Systems Trust and Medical Research Council.

Dabis F, Leroy V, Castetbon K, Spira R, Newell M, Salamon R (2000). Preventing mother-to-child transmission of HIV-1 in Africa in the year 2000. AIDS; 14: 1017–1026.

Dale, M (1999). Spatial pattern analysis in plant ecology. Cambridge: Cambridge University Press.

Danson FM, Armitage RP, Marston CG (2008). Spatial and temporal modelling for parasite transmission studies and risk assessment. Parasite.;15(3):463-8.

Delaunay V, Etard J, Preziosi M, Marra A and Simondon F (2001). Decline of infant and child mortality rates in rural Senegal over a 37-year period (1963-1999). International Journal of Epidemiology; 30: 1286-1293.

Diez-Roux A (2001). Investigating neighbourhood and area effects on health. Am J Public Health 91: 1783-1789.

Diez-Roux A, Mair C (2010). Neighborhoods and health. Ann N Y Acad Sci 1186: 125-145.

Diggle P (1983). Statistical analysis of spatial point patterns. London: Academic press.

Diggle P, Moyeed R and Tawn J (1998). Model-based Geostatistics (with discussion). Applied Statistics, 48: 299-350

Diggle P, Moyeed R, Rowlingston B, Thomson M (2002). Childhood Malaria in the Gambia: a Case-study in Model-based Geostatistics. Applied Statistics; 51: 493-506

Doctor HV, Weinreb AA (2003). Estimation of AIDS adult mortality by verbal autopsy in rural Malawi. AIDS. 2003 Nov 21;17(17):2509-13.

Dolan C, Tollman S, Nkuna V, Gear JSS (1997). The links between legal status and environmental health: A case study of Mozambican refugees and their hosts in the Mpumalanga (Eastern Transvaal) Lowveld, South Africa. Int J Health Hum Rights 1997; 2:62–84.

dos Santos Silva I (1999). Cancer Epidemiology: Principles and Methods. International Agency for Research on Cancer (IARC), Lyon, France.

Dorrington R, Bourne D, Bradshaw D, Laubscher R, Timæus IM (2001). The Impact of HIV/AIDS on Adult Mortality in South Africa. Technical Report. South Africa: Burden of Disease Unit, Medical Research Council. http://www.mrc.ac.za/bod/.

Douven W, Scholten HJ (1995). Spatial analysis in health research. In The Added Value of Geographical Information Systems in Public and Environmental Health. Edited by: de Lepper MJC, Scholten HJ, Stern RM. Boston, MA: Kluwer Academic Press; pp: 117-134.

Duthé G, Pison G (2008). Adult Mortality in a Rural Area of Senegal: Non-Communicable Diseases Have a Large Impact in Mlomp. Demographic Research, 19(37): 1419-1448.

Duczmal L, Assunção R (2004). A simulated annealing strategy for the detection of arbitrarily shaped spatial clusters. Comput. Statist. Data Anal.
45, 269–286.

Duczmal L., Kulldorff M., Huang L (2006). Evaluation of spatial scan statistics for irregularly shaped clusters. J. Comput. Graph. Statist. 15 (2), 1–15.

Duczmal L, Cançadob A, Takahashic R, Bessegatoa L (2007). A genetic algorithm for irregularly shaped spatial scan statistics. Computational Statistics and Data Analysis 52: 43 – 52.

Elith, J, Graham C, Anderson R, Dudik M, Ferrier S, Guisan A, et al. (2006). Novel methods improve prediction of species' distributions from occurrence data. Ecography 29:129-151.

Elliot P, Wakefield J, Best N, Briggs D (2000). Eds: Spatial Epidemiology-Methods and Applications Oxford University Press London

Elliott P and Wartenberg D (2004). Spatial Epidemiology: Current Approaches and Future Challenges. Environ Health Perspect. 112(9): 998–1006.

Erdogan S (2009). Explorative spatial analysis of traffic accident statistics and road mortality among the provinces of Turkey. Journal of Safety Research; 40(5):341-351 doi:10.1016/j.jsr.2009.07.006

Estève J, Benhamou E, Raymond L (1994). Statistical Methods in Cancer Research, Vol. IV, Descriptive Epidemiology. International Agency for Research on Cancer (IARC), Lyon, France.

European Collaborative Study (2005). Mother-to-child transmission of HIV infection in the era of highly active antiretroviral therapy. CID 40, 458–465.

Ezzati M, Lopez A, Rogers A, Vander Hoorn S, Murray C (2002). Comparative Risk Assessment Collaborating Group. Selected major risk factors and global and regional burden of disease. Lancet; 360: 1347-60.

Ezzati M, Lopez A, Rodgers A, Murray C (2004). Comparative quantification of health risks: global and regional burden of disease attributable to selected major risk factors. Geneva: World Health Organization.

Fang L, Yan L, Liang S, Vlas SJD, Feng D, Han X, Zhao W, Xu B, Bian L, Yang H, Gong P, Richardus JH, Cao W (2006). Spatial analysis of hemorrhagic fever with renal syndrome in China. BMC Infectious Diseases; 6:77.

Fantahun M, Berhane Y, Högberg U, Wall S, Byass P (2008). Young adult and middle age mortality in Butajira demographic surveillance site, Ethiopia: lifestyle, gender and household economy. BMC Public Health; 8: 268.

Farah AA, Preston SH (1982). Child mortality differentials in Sudan, Population and Development Review 8, 365–383.

Feachem R. (2000) Poverty and inequity: a proper focus for the new century. Bull World Health Organ.; 78: 1–2.

Fleiss J. L. (1981). Statistical Methods for Rates and Proportions. Second ed. New York: John Wiley and Sons.

Fottrell E, Kahn K, Ng N, Sartorius B, Huong DL, Van Minh H, Fantahun M, Byass P (2010). Mortality measurement in transition: proof of principle for standardised multi-country comparisons. Trop Med Int Health.; 15(10):1256-65.

Frumkin H (2010). Environmental Health: From Global to Local, 2nd Edition. ISBN: 978-0-470-40487-4. Jossey-Bass.

Fukuda Y, Umezaki M, Nakamura K, Takano T (2005). Variations in societal characteristics of spatial disease clusters: examples of colon, lung and breast cancer in Japan. International Journal of Health Geographics, 4:16.

Garenne M., Cantrelle P (1997). Three Decades of Research on Population and Health – The Orstom Experience in Rural Senegal – 1962-1991. In: Das Gupta M., Aaby P., Pison G., Garenne M. (eds).Prospective Community Studies in Developing Countries. Oxford Clarendon Press: 233-252.

Garenne M, Kahn K, Tollman S, Gear J (2000a). Causes of death in a rural area of South Africa: an international perspective. J Trop Pediatr; 46(3):183-90.

Garenne M, Tollman S, Kahn K (2000b). Premarital fertility in rural South Africa: A challenge to existing population policy. Studies in Family Planning; 31(1): 47-54.

Garenne M, Fauveau V (2006). Potential and limits of verbal autopsies. Bulletin of the World Health Organization., 84: 161-256.

Garenne M, Tollman SM, Kahn K, Collinson MA (2007). Fertility trends and net reproduction in Agincourt, rural South Africa, 1992-2004. Scand J Public Health Suppl.; 69: 68–76

Garrib A, Jaffar S, Knight S, Bradshaw D, Bennish ML (2006). Rates and causes of child mortality in an area of high HIV prevalence in rural South Africa. Trop Med Int Health, 11(12):1841-8.

Geary R (1954). The contiguity ratio and statistical mapping. The Incorporated Statistician 5:115-45.

Gelfand A, Smith A (1990). Sampling-based approaches to calculating marginal densities. Journal of the American Statistical Association 1990; 85: 398-409

Gelfand, A. E., Sahu, S. K., and Carlin, B. P. (1995). Efficient parameterization for normal linear mixed models. *Biometrika*, 82, 479-88.

Gelfand, A. E., Sahu, S. K., and Carlin, B. P. (1996). Efficient parameterizations for generalised linear models (with discussion). In Bayesian Statistics 5, (ed. J. M. Bernardo, J. O. Berger, A. P. Dawid, and A. F. M. Smith), pp. 165-80. Clarendon Press, Oxford, UK

Gelfand A, Ravishanker N, Ecker M (1999). Modelling and inference for point-referenced binary spatial data, In: Generalized Linear Models: A Bayesian Perspective, Dey D., Ghosh S., Mallick B., eds. Marcel Dekker Inc. pp373–386.

Gelfand A, Vounatsou P (2003). Proper multivariate conditional autoregressive models for spatial data analysis. Biostat. 4, 11-25.

Gelman A, Carlin JB, Stern HS, Rubin DB (2003). Bayesian Data Analysis, second edition. London: Chapman and Hall.

Gelman A (2008). Response to discussion of “Objections to Bayesian Statistics”. Bayesian Analysis 3: 1-11.

Gemperli A, Vounatsou P (2003a). Fitting generalised linear mixed models for point-referenced spatial data. J Modern Appl Statist Method; 2: 497–511.

Gemperli A (2003b). Development of spatial statistical methods for modelling point-referenced spatial data in malaria epidemiology. In PhD thesis University of Basel: Faculty of Philosophy and Natural Sciences.

Gemperli A, Vounatsou P, Kleinschmidt I, Bagayoko M, Lengeler C, Smith T. (2004). Spatial patterns of infant mortality in Mali; the effect of malaria endemicity. Am J Epidemiol.; 159: 64–72.

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

Gething PW, Noor AM, Gikandi PW, Ogara EA, Hay SI, Nixon MS, Snow RW, Atkinson PM (2006). Improving imperfect data from health management information systems in Africa using space-time geostatistics. PLos Med, 3: e271.

Getis, A (1984) Interaction Modeling Using Second-order Analysis. Environment and Planning A 16: 173-183.

Getis, A and Ord, JK (1992). The Analysis of Spatial Association by Use of Distance Statistics, Geographical Analysis, 24: 189-206.

Ghosh M, Rao J (1994). Small-area estimation - an appraisal. Statistical Science 9(1):55-76.

Ghosh M, Natarajan K, Stroud T, Carlin B (1998). Generalized linear models for small-area estimation. Journal of the American Statistical Association 93(441):273-82.

Ghosh S , Mukhopadhyay P and Lu J (2006). Bayesian analysis of zero-inflated regression models. Journal of Statistical Planning and Inference, 136(4), 1360-1375.

Giraudon I, Leroy V, Msellati P, Elenga N, Ramon R, Welffens-Ekra C, Dabis F (1999). The costs of treating HIV-infected children in Abidjan, Ivory Coast, 1996-1997. Article in French Sante, 9(5):277-81.

Gobalet J, Thomas R (1996). Demographic data and geographic information systems for decision making: The case of public health. Population Research Policy Review; 15: 537-548.

Goodchild MF (1986). Spatial Autocorrelation, Concept and Techniques in Modern Geography. Norwich, UK:Geo Books.

Goovaerts P and Jacquez G (2004). Accounting for regional background and population size in the detection of spatial clusters and outliers using geostatistical filtering and spatial neutral models: the case of lung cancer in Long Island, New York. International Journal of Health Geographics 2004, 3:14 doi:10.1186/1476-072X-3-14

Gosoniu L, Vounatsou P, Sogoba N, Smith T (2006). Bayesian modelling of geostatistical malaria risk data. Geospatial Health 1, 126-139.

Gosoniu L, Vounatsou P, Tami A, Nathan R, Grundmann H, Lengeler C (2008). Spatial effects of mosquito bednets on child mortality. BMC Public Health. 8, 356.

Gosoniu L, Veta AM, Vounatsou P (2010) Bayesian Geostatistical Modeling of Malaria Indicator Survey Data in Angola. PLoS ONE 5(3): e9322. doi:10.1371/journal.pone.0009322

Goudge J, Gilson L, Russell S, Gumede T, Mills A (2009). The household costs of health care in rural South Africa with free public primary care and hospital exemptions for the poor. Trop Med Int Health; 14(4):458-67.

Government of South Africa (2007). HIV and AIDS and STI Strategic Plan for South Africa, 2007-2011. URL available at: http://www.info.gov.za/otherdocs/2007/aidsplan2007/index.html.

Grant A, De Cock K (1998). The growing challenge of HIV/AIDS in developing countries. British Medical Bulletin; 54: 369-381.

Greenwood B (1999). Malaria mortality and morbidity in Africa, 1999. Bull World Health Organ 77: 617–618.

Guisan A, Thuiller W (2005). Predicting species distribution: offering more than simple habitat models. Ecol. Lett.;8: 993–1009. doi:10.1111/j.1461-0248.2005.00792.

Gwatkin D, Bhuiya A, Victora C (2004). Making health systems more equitable. Lancet; 364: 1273-1280.

Gwatkin DR, Guillot M, Heuveline P (1999).The burden of disease among the global poor. Lancet; 354: 586-589.

Gyapong JO, Kyelem D, Kleinschmidt I, Agbo K, Ahouandogbo F, Gaba J, Owusu-Banahene G, Sanou S, Sodahlon YK, Biswas G, Kale OO, Molyneux DH, Roungou JB, Thomson MC, Remme J (2002). The use of spatial analysis in mapping the distribution of bancroftian filariasis in four West African countries.Ann Trop Med Parasitol.; 96(7):695-705.

Hammer G, Somé F, Müller O, Kynast-Wolf G, Kouyaté B, Becher H (2006). Pattern of cause-specific childhood mortality in a malaria endemic area of Burkina Faso. Malaria Journal; 5: 47.

Hammer GP, Kouyaté B, Ramroth H, Becher H (2006). Risk factors for childhood mortality in sub-Saharan Africa. A comparison of data from a Demographic and Health Survey and from a Demographic Surveillance System. Acta Trop.;98(3):212-8.

Hargreaves J, Collinson M, Kahn K, Clark S, Tollman S (2004). Childhood mortality among former Mozambican refugees and their hosts in rural South Africa. International Journal of Epidemiology; 33: 1271-1278.

Hay SI, Tatem AJ, Graham AJ, Goetz SJ, Rogers DJ (2006). Global environmental data for mapping infectious disease distribution. Advances in Parasitology; 62:37–77.

Higdon, D., Lee, H.K.H. and Holloman, C (2003). Markov chain Monte Carlo-based approaches for inference in computationally intensive inverse problems, Bayesian Statistics 7. Proceedings of the Seventh Valencia International Meeting: 181-197.

Hildebrandt N, McKenzie D (2004). The effects of migration on child health in Mexico. Mimeo. Stanford University.

Hill KH (1999). The measurement of adult mortality: an assessment of data availability, data quality and estimation methods. In: Chamie, J. and Cliquet, R. (eds.). Health and mortality issues of global concern: Proceedings of the Symposium on Health and Mortality, Brussels, 19-22 November 1997. Brussels: CBGS, Population and Family Study Centre, Flemish Scientific Institute: 72-83.

Hjalmars U, Kulldorff M and Gustafsson G (1996). Childhood leukemia in Sweden: using GIS and a spatial scan statistic for cluster detection. Stat Med; 15: 707-715

Hjalmars U., Kulldorff M., Gustafsson G (1994). Risk of acute childhood leukaemia in Sweden after the Chernobyl reactor accident, BMJ; 309(6948):154-7

Hjalmars U., Kulldorff M., Wahlqvist Y., Lannering B (1999). Increased incidence rates but no space-time clustering of childhood astrocytoma in Sweden, 1973-1992: a population-based study of pediatric brain tumors, Cancer; 85(9):2077-90

Hobcraft JN, McDonald JW, Rutstein SO (1985). Demographic determinants of infant and early child mortality: A comparative analysis. Population Studies 39(3), 363–85.

Hohm M (1988). Geostatistics and petroleum geology M E Hohm in Geostatistics and petroleum geology. Kluwer Academic Publishers, The Netherlands.

Holt JB (2007). The topography of poverty in the United States: a spatial analysis using county-level data from the Community Health Status Indicators project. Prev Chronic Dis;4(4). http://www.cdc.gov/pcd/issues/2007/ oct/07_0091.htm.

Hosegood V., McGrath N., Herbst K., Timæus I.M (2004). The impact of adult mortality on household dissolution and migration in rural South Africa. AIDS. 2004;18(11):1585–1590.

Hosegood V, Vanneste AM, Timaeus IM (2004). Levels and causes of adult mortality in rural South Africa: the impact of AIDS. AIDS; 18:663–671.

Hosegood V, Preston-Whyte E, Busza J, Moitse S, Timaeus I (2007). Revealing the full extent of households' experiences of HIV and AIDS in rural South Africa. Soc Sci Med.; 65(6): 1249–1259.

Hosmer D and Lemeshow S (1999). Applied Survival Analysis: Regression Modeling of Time to Event Data. New York: Wiley.

Hougaard, P. (2000). Analysis of multivariate survival data. Springer: New York.

Howe L, Hargreaves J and Huttly S (2008). Issues in the construction of wealth indices for the measurement of socio-economic position in low-income countries. Emerging Themes in Epidemiology, 5:3.

Hsu C, Jacobson H and Mas F (2004). Evaluating the disparity of female breast cancer mortality among racial groups - a spatiotemporal analysis. International Journal of Health Geographics; 3:4.

Huang L, Kulldorff M, Gregorio D (2007). A spatial scan statistic for survival data. Biometrics; 63(1):109-18.

Huynen M, Martens P, Hilderink H (2005). The health impacts of globalization: a conceptual framework. Globalization and Health, 1:14.

Ikeda K, Saitoh S, Suzuki Y, Kobayashi M, Tsubota A, Koida I, Arase Y, Fukuda M, Chayama K, Murashima N, Kumada H (1998). Disease progression and hepatocellular carcinogenesis in patients with chronic viral hepatitis: a prospective observation of 2215 patients. J Hepatol; 28: 930-8.

INDEPTH Network (2002). Population and Health in Developing Countries. Volume One: Population, Health and Survival at INDEPTH Sites. IDRC, Ottawa, Canada. 356pp: ISBN: 0-88936-948-8

INDEPTH Network (2003). INDEPTH Standardized Verbal Autopsy questionnaire. Revised August 2003 (http://www.indepth-network.org/core_documents/indepthtools.htm).

Jackson M, Huang L, Xie Q, and Tiwari R (2010). A modified version of Moran's I. Int J Health Geogr. 9: 33.

Jain A (1988). Determinants of regional variation in infant mortality in rural India. In: Infant Mortality in India: Differentials and Determinants, Jain A., Visaria L., eds. Sage Publications, pp127–167.

Jerrett M, Gale S, Kontgis C (2010). Spatial Modelling in Environmental and Public Health Research. Int. J. Environ. Res. Public Health; 7, 1302-1329; doi:10.3390/ijerph7041302

Jochelson K, Mothobeli M, Leger J (1991). Human immunodeficiency virus and migrant labor in South Africa. Int. J. Hlth. Serv. 21, 157-173.

Johnston N (2000). The Regularisation of Former Mozambican Refugees in South Africa: Experiences and Lessons. Acornhoek, Refugee Research Programme.

Kafadar K (1996). Smoothing geographical data, particularly of disease. Statistics in Medicine; 15: 285-96.

Kahn K, Tollman SM, Garenne M, Gear JSS (1999). Who dies from what? Determining cause of death in South Africa's rural north-east. Tropical Medicine and International Health; 4(6): 433-441.

Kahn K, Tollman SM, Garenne M, Gear JS (2000). Validation and application of verbal autopsies in a rural area of South Africa. Tropical Medicine and International Health; 5(11), 824-831.

Kahn K (2006). Dying to make a fresh start: mortality and health transition in a new South Africa [dissertation]. Umeå University, Faculty of Medicine, Public Health and Clinical Medicine, 2006.

Kahn K, Tollman S, Thorogood M, Connor M, Garenne M, Collinson M, Hundt G (2006). Chapter 5: Older Adults and the Health Transition in Agincourt, Rural South Africa: New Understanding, Growing Complexity. Aging in Sub-Saharan Africa: Recommendation for Furthering Research. National Research Council (US) Committee on Population; Cohen B, Menken J, editors. Washington (DC): National Academies Press (US).

Kahn K, Garenne M, Collinson M, Tollman S (2007a). Mortality trends in a new South Africa: Hard to make a fresh start. Scand J Public Health Suppl, 69:26-34.

Kahn K, Tollman SM, Collinson MA, Clark SJ, Twine R, Clark BD, Shabangu M, Gómez-Olivé FX, Mokoena O, Garenne ML (2007b). Research into health, population and social transitions in rural South Africa: Data and methods of the

Agincourt Health and Demographic Surveillance System. Scand J Public Health Suppl; 69: 8–20.

Kahn K (2008). Dying to make a fresh start: Mortality and health transition in a new South Africa. Umeå University Medical Dissertations New Series No 1056 ISSN 0346-6612 ISBN 91-7264-173-8.

Kalipeni E (2000). Health and disease in Southern Africa: a comparative and vulnerability perspective. Social Science and Medicine; 50: 965-983.

Kandala N (2006). Bayesian geoadditive modelling of childhood morbidity in Malawi. Appl Stochast Mod Business Industr 22: 139–154.

Katona P and Katona-Apte J (2008). The Interaction between Nutrition and Infection. Clin Infect Dis.; 46 (10): 1582-1588.

Kazembe L, Kleinschmidt I, Holtz T, and Sharp B (2006). Spatial analysis and mapping of malaria risk in Malawi using point-referenced prevalence of infection data. Int J Health Geogr.; 5: 41.

Kaufman J, Asuzu M, Rotimi C, Johnson O, Owoaje E, and Cooper R (1997). The absence of adult mortality data for sub-Saharan Africa: a practical solution. Bull World Health Organ.; 75(5): 389–395. PMCID: PMC2487019.

Kawachi I and Berkman LF (eds) (2003). Neighbourhoods and Health. New York: Oxford University Press Inc, USA.

Kazembe L, Appleton C, Kleinschmidt I (2007a). Spatial analysis of the relationship between early childhood mortality and malaria endemicity in Malawi. Geospatial Health, 2[1]: 41-50.

Kazembe LN, Namangale JJ (2007b) A Bayesian multinomial model to analyse spatial patterns of childhood co-morbidity in Malawi. Europ J Epidemiol 22: 545–556.

Kelsall JE, Diggle PJ (1995): Non-parametric estimation of spatial variation in relative risk. Statistics in Medicine; 14:2335-2342.

Kendall MG (1970). Rank Correlation Methods, 4th edition, London, Griffen.

Kenneth J. Rothman, Sander Greenland, Timothy L. Lash (2008). Modern epidemiology 3rd edition. Lippincott, Williams and Wilkins.

Kitange HM, Machibya H, Black J, Mtasiwa DM, Masuki G, Whiting D, Unwin N, Moshiro C, Klima PM, Lewanga M, Alberti KG, McLarty DG (1996). Outlook for survivors of Childhood in Sub-Saharan Africa: adult mortality in Tanzania, Adult Morbidity and Mortality Project. BMJ, 312:216-20.

Klauss G, Staub L, Widmer M, Busato A (2005). Hospital service areas -- a new tool for health care planning in Switzerland. BMC Health Serv Res. 9;5:33.

Kleinschmidt I (2001). Spatial statistical analysis, modelling and mapping of malaria in Africa. In PhD thesis University of Basel: Faculty of Philosophy and Natural Sciences.

Knox G (1964). Detection of space-time interactions. Appl Stat 13:25-9.

Koblinsky MA, Tinker A, Daly P (1994). Programming for safe motherhood: a guide to action, Health Policy and Planning; 9:252-66.

Krasilnikov, P, Carré, F and Montanarella, L. (2008). Soil geography and geostatistics: Concepts and Applications. Office for Official Publications of the European Communities, Luxembourg.

Kuate Defo B (1997). Effects of infant feeding practices and birth spacing on infant and child survival: a reassessment from retrospective and prospective data. Journal of Biosocial Sciences 29, 303-26.

Kuate-Defo B, Diallo K (2002). Geography of child mortality clustering within African families. Health and Place.; 8:93-117.

Kulldorff M, Nagarwalla N (1995). Spatial disease clusters: detection and inference. Stat Med. 1995 Apr 30;14(8):799-810.

Kulldorff, M. (1997). A spatial scan statistic. Comm. Statist. Theory Methods 26 (6), 1481–1496.

Kulldorff M (1998). Statistical methods for spatial epidemiology: Tests for randomness, pp 49-62 in Gatrell, A.C. and Löytönen, M. GIS and Health, Taylor and Francis, London.

Kulldorff M, Rand K, Gherman G, Williams G, DeFrancesco D (1998). SatSCan – Software for the Spatial and Space-time Scan Statistics, Version 2.1. Bethesda, Madison: National Cancer Institute.

Kulldorff M, Huang L, Pickle L, Duczmal L (2006). An elliptic spatial scan statistic. Statist. Med. 25, 3929–3943.

Kynast-Wolf G, Sankoh O, Ghangoun A, Kouyatè B, Becher H (2002). Mortality patterns, 1993-1998, in a rural area of Burkina Faso, West Africa, based on the Nouna demographic surveillance system. Tropical Medicine and International Health; 7: 349

Lawson A (2009). Bayesian Disease Mapping. Hierarchical Modeling in Spatial Epidemiology. CRC Press.

le Sueur D, Ngxongo S, Sharp B, Martin C, Fraser C, Teuschner M, Tollman S, Green C, Tsoka1 J, Solarsh G and Mmzava1 A (1997). Towards a spatial rural information system. Durban: Health Systems Trust and Medical Research Council. URL available at: http://www.hst.org.za/publications/towards-spatial-rural-information-system

VanWey L, Rindfuss R, Gutmann M, Entwisle B, and Balk D (2005). Confidentiality and spatially explicit data: Concerns and challenges. Available from URL:http://www.pnas.org_cgi_doi_10.1073_pnas.0507804102.

Lee JW (2003). Child survival: a global health challenge. Lancet; 362:262.

Lombardo JS, Buckeridge DL (eds) (2007). Disease surveillance — a public health informatics approach. New Jersey: John Wiley and Sons. 458 pp. ISBN 978-0-470-06812-0. Stat Methods Med Res; 18(2): 224-225 doi: 10.1177/09622802090180

Lopez AD (1990). Causes of death: an assessment of global patterns of mortality around 1985. World Health Stat Q 1990;43:91-104.

Lopez A, Mathers C, Ezzati M, Jamison D, Murray C (2006). Global and regional burden of disease and risk factors, 2001: systematic analysis of population health data. The Lancet; 367: 1747-1757.

Lurie M (2000). Migration and AIDS in Southern Africa: A Review. South African Journal of Science, 96(6):343-347.

Lurie M, Williams B, Zuma K, Mkaya-Mwamburi D, Garnett G, Sweat M, Gittelsohn J, and Abdool Karim S (2003). Who infects whom? HIV-1 concordance and discordance among migrant and non-migrant couples in South Africa. AIDS. Vol.17:2245–2252.

Mabaso M, Vounatsou P, Midzi S, Da Silva J and Smith T (2006). Spatio-temporal analysis of the role of climate in inter-annual variation of malaria incidence in Zimbabwe. Int J Health Geogr. 2006; 5: 20.

Mantel N (1967). The detection of disease clustering and a generalized regression approach. Cancer Research 27:209-20.

Manda S (1999). Birth intervals, breastfeeding and determinants of childhood mortality in Malawi. Social Science and Medicine 48, 301-312.

Mathers CD, Fat DM, Inoue M, Rao C, Lopez AD (2005). Counting the dead and what they died from: an assessment of the global status of cause of death data. Bull World Health Organ.; 83(3):171-7. Epub 2005 Mar 16.

Mayer J (1996). The political ecology of disease as one new focus for medical geography. Progress in Human
Geography 20, 441-456.

Mayer J (2000). Geography, ecology and emerging infectious diseases. Social Science & Medicine 50; 937-952.

McLeod KS (2000). Our sense of Snow: the myth of John Snow in medical geography. Social Science and Medicine 50(7-8):923-35.

Meade M (1977), Medical geography as human ecology: the dimension of population movement. The Geographical Review 67, no.4, pp.379- 393.

Meade M and Earickson RJ (2000), Medical geography. Second edition. New York/London: The Guilford Press.

Meel BL (2004). Incidence and Patterns of Violent and/or Traumatic Deaths between 1993 and 1999 in the Transkei Region of South Africa. The Journal of Trauma: Injury, Infection, and Critical Care; 57(1): 125-129.

Miller C, Gruskin S, Subramanian S, Heymann J (2007). Emerging health disparities in Botswana: examining the situation of orphans during the AIDS epidemic. Social Science and Medicine; 64, 2476-2486.

Minh H, Byass P and Wall S (2003). Mortality from cardiovascular disease in Bavi District, Vietnam. Scandinavian Journal of Public Health; 31: 26-31

Mitchell J S B (1998). Geometric shortest paths and network optimization. Handbook of Computational Geometry, Elsevier Science, Sack J-R, Urrutia J, eds. (2000) Chapter 15, 633-701. URL available at http://www.ams.stonybrook.edu/~js.

Mohebbi M , Mahmoodi M , Wolfe R, Nourijelyani K , Mohammad K, Zeraati H and Fotouhi A (2008). Geographical spread of gastrointestinal tract cancer incidence in the Caspian Sea region of Iran: Spatial analysis of cancer registry data . BMC Cancer.; 8: 137.

Moïsi J, Gatakaa H, Noor A, Williams T, Bauni E, Tsofa B, Levine O, Scott JA (2010). Geographic access to care is not a determinant of child mortality in a rural Kenyan setting with high health facility density. BMC Public Health; 10:142

Moore DA, Carpenter TE (1999). Spatial Analysis methods and Geographic Information Systems: Use in Health Research and Epidemiology. Epidemiologic Reviews. 21(2): 143-161.

Moran P (1948). The interpretation of statistical maps. Journal of the Royal Statistical Society, Series B 10:243-51.

Morens D, Folkers G, Fauci A (2004). The challenge of emerging and re-emerging infectious diseases. Nature; 430: 242-249.

Moser KA, Leon DA, Gwatkin DR (2005). How does progress towards the child mortality millennium development goal affect inequalities between the poorest and least poor? Analysis of Demographic and Health Survey data. BMJ;331:1180–2.

Mosley WH, Chen LC (1984). An analytic framework for the study of child survival in developing countries. Population and Development Review 10, 25- 45.

Moïsi J, Gatakaa H, Noor A, Williams T, Bauni E, Tsofa B, Levine O, Scott J A (2010). Geographic access to care is not a determinant of child mortality in a rural Kenyan setting with high health facility density. BMC Public Health; 10:142

Mulholland E, Smith L, Carneiro I, Becher H, Lehmannd D (2008). Equity and child-survival strategies. Bulletin of the World Health Organization; 86 (5)

Murray C, Yang G, and Qiao X (1992). Adult Mortality: Levels, Patterns, and Causes. In The Health of Adults in the Developing World, ed. R. G. A. Feachem, T. Kjellstrom, C. J. L. Murray, O. Mead, and M. A. Phillips. New York: Oxford University Press.

Murray C, Ferguson B, Lopez A, Guillot M, Salomon J and Ahmad O (2003). Modified logit life table system: principles, empirical validation and application. Population Studies, 57(2):1-18.

Murray C, Lopez A, Feehan D, Peter S, Yang G (2007). Validation of the Symptom Pattern Method for Analyzing Verbal Autopsy Data. PLoS Med 4(11): e327. doi:10.1371/journal.pmed.0040327

Muthén B and Masyn K (2005). Discrete-time survival mixture analysis. Journal of Educational and Behavioral Statistics, 30, 27-28.

Müller O, Garenne M (1999). Childhood mortality in sub-Saharan Africa. The Lancet 353, 673.

National Research Council (2002). Leveraging longitudinal data in developing countries: Report of a workshop.Workshop on Leveraging Longitudinal Data in

Developing Countries Committee, Committee on Population.Durrant VL, Menken J, eds. Division of Behavioral and Social Sciences and Education. Washington, DC: National Academy Press.

Navér L, Lindgren S, Belfrage E, Gyllensten K, Lidman K, Gisslén M, Ehrnst A, Arneborn M, Bohlin AB (2006). Children born to HIV-1-infected women in Sweden in 1982-2003: trends in epidemiology and vertical transmission. J Acquir Immune Defic Syndr.; 42(4):484-9.

Newell M (2001). Prevention of mother-to-child transmission of HIV: challenges for the current decade. Bulletin of World Health Organization;, 79: 1138–1144.

Newell ML, Brahmbhatt H, Ghys PD (2004). Child mortality and HIV infection in Africa: a review. AIDS; 18 2:S27–S34.

Newell ML, Huang S, Fiore S, Thorne C, Mandelbrot L, Sullivan JL, Maupin R, Delke I, Watts DH, Gelber RD, Cunningham CK; PACTG 316 Study Team (2007). Characteristics and management of HIV-1-infected pregnant women enrolled in a randomised trial: differences between Europe and the USA. BMC Infectious Diseases; 7: 60.

Ngom P, Binka FN, Phillips JF, Pence B, Macleod B (2001). Demographic surveillance and health equity in sub-Saharan Africa. Health Policy Plan,16(4):337-44.

Niederau C, Lange S, Heintges T, Erhardt A, Buschkamp M, Hürter D, Nawrocki M, Kruska L, Hensel F, Petry W, Häussinger D (1998). Prognosis of chronic hepatitis C: results of a large, prospective cohort study. Hepatology; 28: 1687-95.

Nielsen A (1994). Analysis of Regularly and Irregularly Sampled Spatial, Multivariate, and Multi-temporal Data. Ph.D. Thesis. IMM/Technical University of Denmark.

Noor A, Gething P , Alegana V , Patil A , Hay S , Muchiri E, Juma E and Snow R (2009). The risks of malaria infection in Kenya in 2009. BMC Infectious Diseases; 9:180doi:10.1186/1471-2334-9-180

Oppong JR and Harold A (2009). Disease, Ecology, and Environment, in A Companion to Health and Medical Geography (eds T. Brown, S. McLafferty and G. Moon), Wiley-Blackwell, Oxford, UK. doi: 10.1002/9781444314762.ch5

Pascual D, Pla F, Sánchez JS (2010). Cluster validation using information stability measures. Pattern Recognition Letters 31 (2010) 454–461.

Patwari AK (1999). Diarrhoea and malnutrition interaction. Indian J Pediatr.; 66(1 Suppl): S124-34.

Patz, J., D. Engelberg, and Last J (2000). The effects of changing weather on public health. Annual Review of Public Health 21(1): 271–307.

Penchansky R, Thomas J (1981). The concept of access – definition and relationship to consumer satisfaction. Med care 19(2), 127-140.

Pfeiffer D, Stevenson M, Robinson T, Stevens K, Rogers D (2008). Spatial analysis in epidemiology. Oxford University Press.

Pickle LW, Mason TJ, Howard N, Hoover RR, Fraumeni JF (1987). Atlas of US Cancer Mortality among Whites: 1950–1980. DHHS Publication No (NIH) 87–2900.

Pickle LW (2002). Spatial analysis of disease. In: Beam C, ed. Biostatistical Applications in Cancer Research. Boston: Klewer Academic Publishers.

Pison G (2005). Population observatories as sources of information on mortality in developing countries. Demographic Research, 13: 301-334.

Ord, J.K and Getis, A., (1995) Local Spatial Autocorrelation Statistics: Distribution Issues and an Application, *Geographical Analysis*, 27(4): 286-306

Ranta J (2001). On probabilistic models for surveillance and prediction of disease incidence with latent processes: case studies on meningococcal outbreaks, childhood diabetes and poliomyelitis, in Faculty of Science, University of Helsinki, Helsinki., Finland.

Roa J (2003). Small area estimation. John Wiley and Sons.

Raso G, Matthys B, N'Goran EK, Tanner M, Vounatsou P, Utzinger J (2005a). Spatial risk prediction and mapping of Schistosoma mansoni infections among schoolchildren living in western Côte d'Ivoire. Parasitology 131, 97-108.

Raso G, Utzinger J, Silué KD, Ouattara M, Yapi A, Toty A, Matthys B, Vounatsou P, Tanner M, N'Goran EK (2005b). Disparities in parasitic infections, perceived ill health and access to health care among poorer and less poor schoolchildren of rural Côte d'Ivoire. Trop Med Int Health; 10(1):42-57.

Raso G, Vounatsou P, Gosoniu L, Tanner M, N'Goran EK, Utzinger J (2006). Risk factors and spatial patterns of hookworm infection among schoolchildren in a rural area of western Côte d'Ivoire, International Journal for Parasitology, 36(2): 201-210.

Rezaeian M, Dunn G, St Leger S, Appleby L (2007). Geographical epidemiology, spatial analysis and geographical information systems: a multidisciplinary glossary. J Epidemiol Community Health;61:98-102 doi:10.1136/jech.2005.043117

Ricketts T (2003). Geographic information systems and public health. Annual Review of Public Health; 24: 1-6.

Rinaldi L, Musella V, Biggeri A, Cingoli G (2006). New insights into the application of geographical information systems and remote sensing in veterinary parasitology. Geospatial Health;1:33–47.

Ripley B (1977). Modelling Spatial Patterns (with discussion), Journal of the Royal Statistical Society, B, 39, 172–212.

Roberts GO and Sahu SK (1997). Updating schemes, correlation structures, blocking and parameterization of the Gibbs sampler. Journal of the Royal Statistical Society, Series B, 59, 291-317.

Ronsmans C (1996). Birth spacing and child survival in rural Senegal. International Journal of Epidemiology 25, 989-97.

Ronsmans C, Chowdhury ME, Dasgupta SK, Ahmed A, Koblinsky M (2010). Effect of parent's death on child survival in rural Bangladesh: a cohort study, Lancet; 375: 2024–2031

Root G (1997). Population density and spatial differentials in child mortality in Zimbabwe.Social Science and Medicine 44(3): 413–421.

Root G (1999). Disease environments and sub-national patterns of under-five mortality in sub-Saharan Africa. International Journal of Population Geography 5: 117–132.

Rosero-Bixby L (1991). Socioeconomic development, health interventions and mortality decline in Costa Rica. Scand J Soc Med Suppl; 46:33-42.

Rothman K, Greenland S, Lash T (2008). Modern epidemiology. Section I: basic concepts. Lippincott Williams and Wilkins.

Rudolph B, Shah G, Love D (2006). Small Numbers, Disclosure Risk, Security, and Reliability Issues inWeb-based Data Query Systems. J Public Health Management Practice; 12(2), 176–183

Rue H and Tjelmeland H (2002). Fitting Gaussian Markov random fields to Gaussian fields. Scandinavian Journal of Statistics 29: 31-49.

Ruiz-Moreno D, Pascual M, Emch M, Yunus M (2010). Spatial clustering in the spatio-temporal dynamics of endemic cholera. BMC Infectious Diseases; 10:51doi:10.1186/1471-2334-10-51

Ruzicka LT, Lopez AD (1990). The use of cause-of-death statistics for health situation assessment: national and international experiences. World Health Stat Q.; 43(4):249-58.

Rytkönen M (2004). Not all maps are equal: GIS and spatial analysis in epidemiology. International Journal of Circumpolar Health; 63(1): 9-24.

Schotman P (1994). Priors for the AR (1) model: parameterization issues and time series considerations. Econometric Theory. 10, 579-595.

Sokal RR and Oden NL (1978a). Spatial autocorrelation in biology. 1. Methodology. Biol J Linn Soc 10:199–228.

Sokal RR and Oden NL (1978b). Spatial autocorrelation in biology. 2. Some biological implications and four applications of evolutionary and ecological interest. Biol J Linn Soc 10:229–249.

Sankoh O, Yé Y, Sauerborn R, Müller O and Becher H (2001). Clustering of childhood mortality in rural Burkina Faso. International Journal of Epidemiology; 30: 485-492.

Sankoh O, Berke O, Simboro S and Becher H (2002). Bayesian and GIS Mapping of Childhood Mortality in Rural Burkina Faso. Control of Tropical Infectious Diseases, Uni-Heidelberg Discussion Paper, 03/2002.

Sankoh O, Binka, F (2004). INDEPTH network: generating empirical population and health data in resource-constrained countries in the developing world. In: Becher, H., Kouyat´e, B. (Eds.), Health Research in Developing Countries. A Collaboration Between Burkina Faso and Germany. Springer-Verlag, Heidelberg, pp. 21–32.

Sankoh OA, Ngom P, Clark SJ, de Savigny D, Binka F (2006). Chapter 7: Levels and Patterns of Mortality at INDEPTH Demographic Surveillance Systems. In: Jamison DT, Feachem RG, Makgoba MW, Bos ER, Baingana FK, Hofman KJ, Rogo KO,

editors. Disease and Mortality in Sub-Saharan Africa. 2nd edition. Washington (DC): World Bank.

Sansom SL, Anderson JE, Farnham PG, Dominguez K, Soorapanth S, Clark J, Sukalac T, Earp MJ, Bohannon B, Fowler MG; PSD Consortium (2006). Updated estimates of healthcare utilization and costs among perinatally HIV-infected children. J Acquir Immune Defic Syndr.; 1;41(4):521-6.

Sartorius B, Kahn K, Vounatsou P, Collinson MA and Tollman SM (2010a). Space and time clustering of mortality in rural South Africa (Agincourt HDSS), 1992-2007. Global Health Action Supplement 1. DOI: 10.3402/gha.v3i0.5225.

Sartorius B, Kahn K, Vounatsou P, Collinson MA and Tollman SM (2010b). Young and vulnerable: spatial-temporal trends and risk factors for infant mortality in rural South Africa (Agincourt), 1992-2007. BMC Public Health; 10: 645 DOI: 10.1186/1471-2

Sartorius B, Kahn K, Vounatsou P, Collinson MA and Tollman SM (2011). Survived infancy but still vulnerable: Spatial-temporal trends and risk factors for child (1-4) mortality in rural South Africa (Agincourt), 1992-2007. Geospatial Health 2011, 5(2): 285-295.

Sartorius K, Sartorius B, Tollman S, Schatz E, Kirsten J, Collinson M (2011). Rural Poverty Dynamics and Refugee Communities in South Africa: A Spatial–Temporal Model. Population, space and place. DOI: 10.1002/psp.697

SaTScan TM User Guide for version 7.0. http://www.satscan.org/techdoc.html

Schellenberg JA, Mrisho M, Manzi F, Shirima K, Mbuya K, Mushi A, Ketende SC, Alonso A, Mshinda H, Tanner M, Schellenberg D (2008). Health and survival of young children in southern Tanzania. BMC Public Health, 8:194.

Schotman P (1994). Priors for the AR [1] model: parameterization issues and time series considerations. Econometric Theory.; 10:579–595. doi: 10.1017/S0266466600008677.

Schultz TP (1979). Interpretation of relations among mortality, economics of the household and the health environment. In: Proceedings of the Meeting on Socio-Economic Determinants and Consequences of Mortality Mexico City, June 19–25 Geneva: World Health Organization.

Schwartländer B, Stover J, Walker N, Bollinger L, Gutierrez JP, McGreevey W, Opuni M, Forsyth S, Kumaranayake L, Watts C, Bertozzi S (2001). Resource needs for HIV/AIDS. Science; 292: 2434–36.

Segurado P, Araújo MB (2000). An evaluation of methods for modelling species distributions. J. Biogeogr. 2004;31:1555–1568. doi:10.1111/j.1365-2699.2004.01076.
Sen K, Bonita R. Global health status: two steps forward, one step back. Lancet; 356: 577-582.

Skrondal A, Rabe-Hesketh S (2004). Generalized latent variable modelling. Boca Raton: Chapman and Hall/CRC.

Snow J. (1855). On the Mode of Communication of Cholera. 2nd ed. London: John Churchill

Snow RW, Schellenberg JR, Forster D, Mung'ala VO, Marsh K (1994). Factors influencing admission to hospital during terminal childhood illnesses in Kenya. Int J Epidemiol 23: 1013–1019.

Spiegelhalter D, Best N, Carlin B, van der Linde A (2002). Bayesian measures of model complexity and fit. J R Stat Soc 64(B), 583-639.

Steele F (2005).Event history analysis. A National Centre for Research Methods Briefing Paper. University of Bristol.

Stein A, van der Meer F, Gorte B (1999). Spatial statistics for remote sensing, Kluwer Academic Publishers, The Netherlands.

Stein ML (2005). Space-time covariance functions. Journal of American Statistical Association, 100: 320-321.

Stock R (1983). Distance and the utilization of health facilities in rural Nigeria. Soc Sci Med 17, 563-70.

Stover J, Walker N, Garnett GP, Salomon JA, Stanecki KA, Ghys PD, Grassly NC, Anderson RM, Schwartländer B (2002). Can we reverse the HIV/AIDS pandemic with an expanded response? Lancet; 360: 73–77.

Stringer E, Sinkala M, Ng'uni C, McFarlane YA, Chipimo M, Myzece E, Chi BH, Stringer JS (2004). In an urban African setting, the cost-effectiveness of PMTCT services improves as services expand. International Conference on AIDS [15th : 2004: Bangkok, Thailand]. Int Conf AIDS. 2004 Jul 11-16; 15: abstract no. TuPeC4933.

Sweeney L (1997). Weaving technology and policy together to maintain confidentiality. Journal of Law, Medicine & Ethics;25:98-110.

Takahashi K, Kulldorff M, Tango T, Yih K (2008). A flexibly shaped space-time scan statistic for disease outbreak detection and monitoring. International Journal of Health Geographics; 11;7:14.

Talbot T, Kulldorff M, Forand S, Haley V (2000). Evaluation of spatial filters to create smoothed maps of health data. Statistics in Medicine; 19: 2399-2408.

Tang W, David A. Bennett. Agent-based Modeling of Animal Movement: A Review. (2010). Geography Compass 4 (7): 682-700. DOI: 10.1111/j.1749-8198.2010.00337

Tanser F and Wilkinson D (1999). Spatial implications of the tuberculosis DOTS strategy in rural South Africa: a novel application of geographical information system and global positioning system technologies. Tropical Medicine and Internationa

Tanser F, Bärnighausen T, Cooke G, and Newell ML (2009). Localized spatial clustering of HIV infections in a widely disseminated rural South African epidemic. Int. J. Epidemiol. 38 (4): 1008-1016. doi: 10.109

Taylor MJ, Awadzi K, Basáñez MG, Biritwum N, Boakye D, Boatin B, Bockarie M, Churcher TS, Debrah A, Edwards G, Hoerauf A, Mand S, Matthews G, Osei-Atweneboana M, Prichard RK, Wanji S, Adjei O (2009). Onchocerciasis Control: Vision for the Future from a Ghanian perspective. Parasit Vectors.; 2(1): 7.

Thior I, Lockman S, Smeaton LM, Shapiro RL, Wester C, Heymann SJ, Gilbert PB, Stevens L, Peter T, Kim S, van Widenfelt E, Moffat C, Ndase P, Arimi P, Kebaabetswe P, Mazonde P, Makhema J, McIntosh K, Novitsky V, Lee TH, Marlink R, Lagakos S, Essex M; Mashi Study Team (2006). Breastfeeding plus infant zidovudine prophylaxis for 6 months vs formula feeding plus infant zidovudine for 1 month to reduce mother-to-child HIV transmission in Botswana: a randomized trial: the Mashi Study. JAMA. 16; 296(7):794-805.

Tilman, David, and Peter Kareiva, editors. (1997). Monographs in Population Biology, 30. Princeton University Press. New Jersey.

Timæus I (2001). Commentary: Mortality decline in Senegal. Int J Epidemiol; 30:1294-1295.

Patterson T, Basson M, Bravington M, Gunn J (2009). Classifying movement behaviour in relation to environmental conditions using hidden Markov models. Journal of Animal Ecology 78:6, 1113-1123

Tollman, S.M., Herbst K, Garenne M (1995). The Agincourt Demographic and Health Study: Phase 1. Johannesburg: Health Systems Development Unit, Department of Community Health, University of the Witwatersrand.

Tollman SM, Kahn K, Garenne M, Gear JS (1999a). Reversal in mortality trends: evidence from the Agincourt field site, South Africa, 1992-1995. AIDS. 18; 13(9):1091-7.

Tollman SM, Herbst K, Garenne M, Gear JS, Kahn K (1999b). The Agincourt demographic and health study--site description, baseline findings and implications. S Afr Med J. 89(8), 858-64.

Tollman SM, Kahn K, Sartorius B, Collinson MA, Clark SJ, Garenne ML (2008). Implications of mortality transition for primary health care in rural South Africa: a population-based surveillance study. Lancet; 372(9642):893-901.

Tottrup C, Tersbol B, Lindeboom W and Meyrowitsch D (2009). Putting child mortality on a map: towards an understanding of inequity in health. Trop Med Int Health.; 14(6):653-62.

Trapé JF (2001). The public health impact of chloroquine resistance in Africa. American Journal of Tropical Medicine and Hygiene 64 (supplement), 12-17.

Tsai PJ, Lin ML, Chu CM, and Perng CH (2009). Spatial autocorrelation analysis of health care hotspots in Taiwan in 2006. BMC Public Health; 9: 464.

Tulloch J (1999). Integrated approach to child health in developing countries. The Lancet 354 (supplement 2), 16-20.

Twine R, Collinson MA, Polzer TJ, Kahn K (2007). Evaluating access to a child-oriented poverty alleviation intervention in rural South Africa. Scand J Public Health; 69: 118-27.

Uthman OA, Yahaya I, Ashfaq K, Uthman MB (2009). A trend analysis and sub-regional distribution in number of people living with HIV and dying with TB in Africa, 1991 to 2006. Int J Health Geogr. 24;8:65.

UNAIDS (2002). Paediatric HIV infection and AIDS: UNAIDS point of view. (UNAIDS best practice collection. Point of view.) Geneva: UNAIDS.

UNGASS (2010). South Africa UNGASS Country Progress Report

UNICEF, UNAIDS and WHO. Children and AIDS; Second Stocktaking Report. New York: UNICEF.

UNICEF (2001). Progress since the world summit for children: a statistical review. New York: UNICEF. http://www.unicef.org/pubsgen/ wethechildren-stats/sgreport_adapted_stats_eng.pdf.

United Nations (2002). Methods for estimating adult mortality. New York: Population Division, Department of Economic and Social Affairs of United Nations.

United Nations (2006). World population prospects: The 2005 revision. New York: Population Division, Department of Economic and Social Affairs of United Nations. (http://esa.un.org/unpp/).

Vecchia AV (1988). Estimation and identification for continuous spatial process. Journal of the Royal Statistical Society: Series B 50, 297-312.

Victora C, Vaughan JP, Barros FC, Silva AC, Tomasi E (2000). Explaining trends in inequities: evidence from Brazilian child health studies. The Lancet 356, 1093-1098.

Victora C, Wagstaff A, Schellenberg JA, Gwatkin D, Claeson M, Habicht JP (2003). Applying an equity lens to child health and mortality: more of the same is not enough. Lancet;: 362: 233–41.

Walboomers JM, Jacobs MV, Manos MM, Bosch FX, Kummer JA, Shah KV, Snijders PJ, Peto J, Meijer CJ, Muñoz N (1999). Human papillomavirus is a necessary cause of invasive cervical cancer worldwide. J Pathol 189(1): 12–19.

Wagstaff A (2000). Socioeconomic inequalities in child mortality: comparisons across nine developing countries. Bulletin of the World Health Organisation 78, 19–29.

Waldron I (1998). Too Young to Die: Genes or Gender? Sex differences in infant and early childhood mortality: Major causes of death and possible biological causes. New York: United Nations.

Walker N, Schwartländer B, Bryce J (2002). Meeting international goals in child survival and HIV/AIDS. Lancet; 360: 284–89.

Waller LA, Gotway CA (2004). Applied Spatial Statistics for Public Health Data. Hoboken, NJ: John Wiley and Sons.

Walsh SJ, Page PH, Gesler WM (1997). Normative models and healthcare planning: network-based simulations within a geographic information system environment. Health Serv Res.; 32(2):243-60.

Walter S (2000). Disease mapping: a historical perspective. Spatial Epidemiology - Methods and Applications (Elliott P, Wakefield J, Best N, Briggs DJ, Eds). Oxford: Oxford University Press, 223–252

Watts H, Gregson S, Saito S, Lopman B, Beasley M and Monasch R (2007), Poorer health and nutritional outcomes in orphans and vulnerable young children not explained by greater exposure to extreme poverty in Zimbabwe. Tropical Medicine and International Health, 12: 584–593. doi: 10.1111/j.1365-3156.2007.01832.

Wheeler D (2007). A comparison of spatial clustering and cluster detection techniques for childhood leukemia incidence in Ohio, 1996 – 2003. International Journal of Health Geographics, 6:13doi:10.1186/1476-072X-6-13

White R (2003). What can we make of an association between human immunodeficiency virus prevalence and population mobility? International Journal of Epidemiology. 32:753–754

Whittle P (1954). On Stationary Processes in the Plane. Biometrika 41: 434 -449.

Wichmann J and Voyi K, 2006. Influence of Cooking and Heating Fuel Use on 1–59 Month Old Mortality in South Africa. Matern Child Health J 10, 553–561.

WHO, (2002). The world health report 2002: reducing risks, promoting healthy life. Geneva: World Health Organization.

WHO (2003). The world health report 2003 – shaping the future. Chapter 1: Global Health: today's challenges. Geneva: World Health Organization. http://www.who.int/whr/2003/chapter1/en/index.html.

WHO (2007). WHO Statistical Information System (WHOSIS). Verbal Autopsy Standards: Ascertaining and attributing causes of death. Geneva: World Health Organization.

WHO, (2007). Reaching the Poor: Challenges for Child Health in the Western Pacific Region. Chapter 7: Barriers to access to child health care. Geneva: World Health Organization

WHO (2009). Role of GIS in public health mapping. http://www.who.int/health_mapping/en/ Geneva: World Health Organization

WHO (2010). Guidelines on HIV and infant feeding. Principles and recommendations for infant feeding in the context of HIV and a summary of evidence, Geneva: World Health Organization. http://www.who.int/child_adolescent_health/documents/9789241599535/en/index.html, accessed

Wikle CK, Berliner LM and Cressie N (1998). Hierarchical Bayesian space-time models. In J. Envr. Ecol. Stat., 5: 117-154.

Wikle C and Royle J (2002). Spatial statistical modelling in biology. Encyclopaedia of Life Support Systems, EOLSS Publishers Co. Ltd

Wikle CK (2003). Hierarchical Bayesian models for predicting the spread of ecological processes. In Ecology, 84: 1382-1394.

Williams B, Gouws E, Lurie M, Crush J (2002). Spaces Of Vulnerability: Migration and HIV/AIDS In South Africa. Migration Policy Series No. 24 South African Migration Project.

Woods R (2003). Urban-rural mortality diff erentials: An unresolved debate. Population and Development Review 29(1): 29–46.

Wua H, Bai-Lian Li, Timothy A. Springerc and William H. Neill (2000). Modelling animal movement as a persistent random walk in two dimensions: expected magnitude of net displacement. Ecological Modelling 132(1-2): 115-124.

Xie Y (2008). Remote sensing imagery in vegetation mapping: a review. J Plant Ecol. 1(1): 9-23.

Yamano T and Jayne T (2004). Measuring the impact of working-age adult mortality on small-scale farm households in Kenya. World Development.; 32(1):91–119.

Yamaguchi K (1991). Event History Analysis. Newbury Park: Sage.

Yates R (2010). Women and children first: an appropriate first step towards universal coverage. Bull World Health Organ; 88(6): 474–475.

Yeshiwondim A, Sucharita Gopal S, Hailemariam A, Dengela D and Patel H (2009). Spatial analysis of malaria incidence at the village level in areas with unstable transmission in Ethiopia. International Journal of Health Geographics; 8:5 doi:10.1186/1476-072X-8-5.

Zacarias O and Andersson m (2010). Mapping malaria incidence distribution that accounts for environmental factors in Maputo Province – Mozambique. Malaria Journal, 9:79doi:10.1186/1475-2875-9-79

Zeller A (1996). An introduction to Bayesian inference and econometrics. John Wiley and Sons Ltd/Inc: New York.

Zhang Z, Carpenter TE, Chen Y, Clark AB, Lynn HS, Peng W, Zhou Y, Zhao G, Jiang Q (2008). Identifying high-risk regions for schistosomiasis in Guichi, China: a spatial analysis. Acta Trop.; 107(3):217-23.

Zwang J, Garenne M, Kahn K, Collinson M, Tollman SM (2007). Trends in mortality from pulmonary tuberculosis and HIV/AIDS co-infection in rural South Africa (Agincourt). Trans R Soc Trop Med Hyg.; 101(9):893-8.

Appendix 1 (model formulations and WinBUGS code)

The various analytical modelling approach formulations as well as WinBUGS code are provided below. All parameter specifications follow on after all model formulations.

--Model specifications and parameters

Negative binomial spatial-temporal modelling approach (Sartorius et al., 2010)

Let Y_{it} and p_{it} be the status and probability of mortality of individual i in year of birth t. We assume that Y_{it} arises from a negative binomial distribution, that is Y_{it} ~NegBin[p_{it}, r], where p_{it} *is* the probability that individual *i* at location *si* is dead and *r* is the parameter that quantifies the amount of extra Poisson variation. We modelled the probability of death [p_{it}] as follows: logit (p_{it}) = $\beta_0 + \beta X_{it} + \varphi_{it} + \alpha_t$

Discrete time logistic spatial-temporal modelling approach (Sartorius et al., 2011)

Let Y_{it} and p_{it} be the mortality status of individual i and time t. We assume that arises from a Bernoulli distribution, $Y_{it} \sim Be(p_{it})$. We model covariates X_{it}, village-specific random effect φ_{it}, individual level random effect μ_i and temporal random effect α_t as follows: logit (p_{it}) = $\beta_0 + \beta X_{it} + \mu_i + \varphi_{it} + \alpha_t$

Parametric spatial modelling approach

Note that a non-parametric Cox modelling approached was also tested, though violated the proportional hazards assumption. Various parametric distributions were also tested based on best fit and predictive power. We assumed a parametric Weibull distribution for the survivor function, where t_{ikj} is the failure time of an individual i (for censored observations the survival distribution is a truncated Weibull with an upper bound corresponding to the censoring time) for residence episode k at location j with covariate vector X_{ik} and β is a vector of unknown regression coefficients and including a village level spatial random effect w_j in the exponent of the hazard model as follows

$t_{ikj} \sim \text{Weibull}(\rho,\mu_{ikj})$ $\quad i = 1,...,N;$

with a baseline hazard function of the form

$l_0(t_{ikj}) = \rho t_{ikj}^{r-1}$

and means for the various models as follows

$\log(\mu_{ikj}) = \beta_0 + \beta X_{ik} + \varphi_{it}$

Where applicable in the various model formulations:

- β_0 is the constant (model dependent) when all covariates are zero (i.e. the constant)
- X_{it} denotes the covariates
- β is the vector of regression coefficients
- φ_{it} the village-specific random effect
- α_t the temporal random effect.
- μ_i the individual level random effect

We assume that φ_{it} has a multivariate normal distribution, $\varphi_{it} \sim MVN(\underline{0},\Sigma)$, with variance-covariance matrix Σ. We also assume an isotropic stationary spatial process, where $\Sigma kl = \sigma_w^2 \exp(-\varphi d_{kl})$, d_{kl} is the Euclidean distance between villages *k* and *l*, σ_w^2 is the geographical variability known as the sill, φ is a smoothing parameter that controls the rate of correlation decay with increasing distance and measures the range of geographical dependency. Both a noninformative gamma prior was adopted for the exponential correlation function, phi [φ], which is the smoothing parameter that controls the rate of correlation decay, as well as uniform prior with a distribution limit between φ_{min} and φ_{max} based on calculation which factors in the minimum and maximum estimated distance between locations i.e. meters between village centroids in this case (Gelfand and Vounatsou, 2003).

The individual level random effect, μ_i, was modelled as a Normal distribution, $\mu_i \sim N(\underline{0}, \sigma_u^2)$. We also used hierarchical centering to improve convergence of selected models. Hierarchical centering were each stochastic variable is considered to arise from a stochastic mean. Gelfand et al. (1995,1996) argue that this procedure often improves convergence, and further evidence is provided by Roberts and Sahu (1997).

Temporal random effects were also used at monthly or yearly intervals to account for temporal correlation. A first order year level autoregressive temporal random effect (α_t) was modeled as simple AR(1) model where the random effects α_t may be written

as $\alpha_t \mid \alpha_{t-1} \sim N(\rho\alpha_{t-1},\sigma^2_\alpha)$,where σ^2_α is the temporal dispersion parameter and ρ as

temporal autocorrelation with $|\rho| \leq 1$ (Schotman, 1994). It simple terms it assumes that deaths at month or year t are influenced by deaths at month t-1.

A second order year level autoregressive temporal random effect (α_t), for t=1 to 16 years, was modelled as a normal distribution with mean α_{mean} [t=3,..,16] = ρ0 + ρ[1]*α[t-1] + ρ[2]*α[t-2] and a noninformative gamma distribution for the variance parameter. The first two autoregressive terms were specified as α_{mean} [1] <- ρ0 + l[1] and α_{mean} [2] <- ρ0 + ρ[1]*α[1] + l[2]. Noninformative normal prior distributions were adopted for the ρ and l coefficients (Zeller, 1996).

We assumed:

- non-informative Normal distributions for the $\beta_{0\ and}$ β parameters ~ N(0, 1 or 0,1).
- inverse gamma priors for all σ^2 of the various random effects

The range (φ) is defined as the minimum distance at which spatial correlation between locations is below 5% i.e. spatial correlation is significant within this distance (example: Raso et al., 2006). This distance can be calculated as 3/φ meters based on exponential correlation decay.

--WinBUGS code for the three model specifications

Note covariates are represented by: "b[1]*X1[i]+ … +b[8]*X8[i]"

Negative binomial spatial-temporal model

model {

```
for (i in 1:N){
        infant_deaths[i] ~ dnegbin(p[i],r)
        p[i] <- r/(r+mu[i])
        log(mu[i]) <-  log(offset[i]) + b0 + alpha[year[i]] + w[village[i]] +
        b[1]*X1[i]+....+b[7]*X7[i]
                }
# AR(2) process
for (j in 1:2){
        l[j] ~ dnorm(0,0.01)
        rho[j] ~ dnorm(0,1)
        }
rho.0 ~ dnorm(0,0.001)
tau.e ~ dgamma(1,1)
sigma2.e <- 1./tau.e
alphamean[1] <- rho.0 + l[1]
alphamean[2] <- rho.0 + rho[1]*alpha[1] + l[2]
for (t in 3:16){
        alphamean[t] <- rho.0 + rho[1]*alpha[t-1] + rho[2]*alpha[t-2]
        }
for (k in 1:16){
        alpha[k]~ dnorm(alphamean[k],tau.e)
        }
# Spatial process
w[1:21] ~spatial.exp(m[], x[], y[], tau.w, phi,1)
for (j in 1:21){
```

```
        m[j] <- 0
        }
tau.w ~ dgamma(1,1)
sigma2.w <- 1/tau.w
phi ~ dunif(0.0001,0.0017)
range <- 3/phi
#  Dispersion parameter
r ~ dgamma(1,0.1)
#  Constant
b0 ~ dnorm(0,0.1)
# Covariates
for (j in 1:7) {
        b[j] ~ dnorm(0,0.1)
        RR[j] <- exp(b[j])
                }
# Spatial kriging
for (j in 1:M){
        w.pred[j] ~ spatial.unipred(0,x.pred[j],y.pred[j], w[])
        p.pred[j]<-exp(w.pred[j])
                }
}
```

Discrete time (monthly) logistic models

--Spatial-temporal: AR(1)

Model

```
#  Likelihood
```

```
{
for (i in 1:N){
        u[i]~dnorm(0,tau.u)
        for (t in 1:n.obs[i]){
        child_died[i,t] ~ dbern(p[i,t])
        logit(p[i,t]) <- b0 +
        b[1]*X1[i]+ … +b[8]*X8[i]+u[i]+w[village[i,t]]+alpha[year[i,t]]
        }
}
# Individual random effect
tau.u ~ dgamma(0.01, 0.01)
sigma2.u <- 1/tau.u
# Spatial process
w[1:21] ~ spatial.exp(m[], x[], y[], tau.w, phi,1)
for (j in 1:21){
        m[j] <- 0
        }
tau.w ~ dgamma(0.01, 0.01)
sigma2.w <- 1/tau.w
phi ~ dunif(0.0001,0.0017)
range <- 3/phi
#  AR(1) process
for (k in 2:16){
        alphamean[k-1]<-rho*alphayear[k-1]
        alpha[k] ~ dnorm(alphamean[k-1],tau.e)
```

```
        }
alpha[1] ~ dnorm(0,tau.a)
tau.a <- (pow(tau.e,2)/(1-pow(rho,2)))
tau.e ~ dgamma(0.01,0.01)
sigma2.a <-1./tau.a
sigma2.e <-1./tau.e
rho~dunif(-5,5)
#  Constant
b0 ~ dnorm(0,0.1)
#  Covariates
for (j in 1:8) {
        b[j] ~ dnorm(0,0.1)
        OR[j] <- exp(b[j])
        }
# Spatial kriging
for (j in 1:M){
        w.pred[j] ~ spatial.unipred(0,x.pred[j],y.pred[j], w[])
        p.pred[j]<-exp(w.pred[j])
        }

}
--Spatial-temporal: AR(2)
Model
{
for (i in 1:N){
```

```
        for (t in 1:n.obs[i]){
                child_died[i,t] ~ dbern(p[i,t])
                logit(p[i,t]) <- v[subject[i,t]] +
                b[1]*X1[i]+ … +b[8]*X8[i]+u[i]+w[village[i,t]]++ w[village[i,t]] +
                alpha[year[i,t]]
                }
        }
# Hierarchical centering of individual random intercepts
for (j in 1:N){
        v[j]~dnorm(b0,tau.v)
        }
tau.v ~ dgamma(1,1)
sigma2.v<-1./tau.v
#  Spatial process
w[1:21] ~ spatial.exp(m[], x[], y[], tau.w, phi,1)
for (j in 1:21){
        m[j] <- 0
        }
tau.w ~ dgamma(1,1)
sigma2.w <- 1/tau.w
phi ~ dunif(0.0001,0.0017)
range <- 3/phi
#  AR(2) process
for (j in 1:2){
        l[j] ~ dnorm(0,1)
```

```
        rho[j] ~ dnorm(0,1)
        }
rho.0 ~ dnorm(0,1)
tau.e ~ dgamma(1,1)
sigma2.e <- 1./tau.e
alphamean[1] <- rho.0 + l[1]
alphamean[2] <- rho.0 + rho[1]*alpha[1] + l[2]
for (t in 3:16){
        alphamean[t] <- rho.0 + rho[1]*alpha[t-1] + rho[2]*alpha[t-2]
        }
for (k in 1:16){
        alpha[k]~ dnorm(alphamean[k],tau.e)
        }
#  Constant
b0 ~ dnorm(0,1)
#  Covariates
for (j in 1:8) {
        b[j] ~ dnorm(0,1)
        OR[j] <- exp(b[j])
        }
}
```

Parametric spatial model

```
model
{
        for(i in 1:N) {
```

```
#  Individual random effect
            u[i]~dnorm(0,tau.u)
             for(k in 1:n.obs[i]){
                    t[i,k] ~ dweib(rho, mu[i,k])I(tcen[i,k],)
                     log(mu[i,k])<-b0+ b[1]*X1[i]+ … +b[8]*X8[i]+u[i]+w[village[i,k]]
                     }
           }
# Spatial process
  w[1:25] ~ spatial.exp(m[], x[], y[], tau.w, phi,1)
  for (j in 1:25){
             m[j] <- 0
             }
# Priors
tau.w ~ dgamma(1,1)
sigma2.w <- 1/tau.w
phi ~ dunif(0.0001,0.0017)
range <- 3/phi
tau.u~dgamma(1,1)
sigma2.u<-1./tau.u
b0~dnorm(0,1)
rho~dgamma(1,1)
# Spatial kriging
for (j in 1:M){
          w.pred[j] ~ spatial.unipred(0,x.pred[j],y.pred[j], w[])
          p.pred[j]<-exp(w.pred[j])
```

}

}

Appendix 2 (papers)

I. Sartorius B, Kahn K, Vounatsou P, Collinson MA and Tollman SM (2010). Space and time clustering of mortality in rural South Africa (Agincourt HDSS), 1992-2007. Global Health Action Supplement 1. DOI: 10.3402/gha.v3i0.5225.

II. Sartorius B, Kahn K, Vounatsou P, Collinson MA and Tollman SM (2010). Young and vulnerable: spatial-temporal trends and risk factors for infant mortality in rural South Africa (Agincourt), 1992-2007. BMC Public Health, 10: 645 DOI: 10.1186/1471-2458-10-546.

III. Sartorius B, Kahn K, Vounatsou P, Collinson MA and Tollman SM (2011). Survived infancy but still vulnerable: Spatial-temporal trends and risk factors for child (1-4) mortality in rural South Africa (Agincourt), 1992-2007. Geospatial Health, 5(2): 285-295.

IV. Sartorius B, Kahn K, Vounatsou P, Collinson MA and Tollman SM (2011). Dying in their prime: spatial-temporal trends and risk factors for adult mortality in the Agincourt rural sub-district, South Africa, 1992-2008. In manuscript.

INDEPTH Mortality Clustering Supplement

Space and time clustering of mortality in rural South Africa (Agincourt HDSS), 1992–2007

Benn Sartorius[1,4]*, Kathleen Kahn[1,2,4], Penelope Vounatsou[3], Mark A. Collinson[1,2,4] and Stephen M. Tollman[1,2,4]

[1]MRC/Wits Rural Public Health and Health Transitions Research Unit (Agincourt), School of Public Health, Faculty of Health Sciences, University of the Witwatersrand, Johannesburg, South Africa; [2]Centre for Global Health Research, Epidemiology and Global Health, University of Umeå, Umeå, Sweden; [3]Swiss Tropical Institute, Basel, Switzerland; [4]INDEPTH Network, Accra, Ghana

***Background*:** Detailed information regarding the spatial and/or spatial–temporal distribution of mortality is required for the efficient implementation and targeting of public health interventions.
***Objectives*:** Identify high risk clusters of mortality within the Agincourt subdistrict for targeting of public health interventions, and highlight areas for further research.
***Design*:** Mortality data were extracted from the Agincourt health and socio-demographic surveillance system (HDSS) for the period 1992–2007. Mortality rates by age group and time were calculated assuming a Poisson distribution and using precise person-time contribution estimates. A spatial scan statistic (Kulldorff) was used to test for clusters of age group specific all-cause and cause-specific mortality both in space and time.
***Results*:** Many statistically significant clusters of higher all-cause and cause-specific mortality were identified both in space and time. Specific areas were consistently identified as high risk areas; namely, the east/south-east and upper east central regions. This corresponds to areas with higher mortality due to communicable causes (especially HIV/TB and diarrheal disease) and indicates a non-random element to the distribution of potential underlying causative factors e.g. settlements comprising former Mozambican refugees in east/south-east of the site, corresponding higher poverty areas, South African villages with higher HIV prevalence, etc. Clusters of older adult mortality were also observed indicating potential non-random distribution of non-communicable disease mortality.
***Conclusion*:** This study has highlighted distinct clusters of all-cause and cause-specific mortality within the Agincourt subdistrict. It is a first step in prioritizing areas for further, more detailed research as well as for future public health follow-on efforts such as targeting of vertical prevention of HIV/TB and antiretroviral rollout in significant child and adult mortality clusters; and assessment and provision of adequate water and sanitation in the child mortality clusters particularly in the south-east where diarrheal mortality appears high. Underlying causative factors need to be identified and accurately quantified. Other questions for more detailed research are discussed.

Keywords: *all-cause mortality*; *demographic surveillance*; *clustering*; *spatial–temporal*

Received: 19 April 2010; Revised: 16 May 2010; Accepted: 18 May 2010; Published: 30 August 2010

Reliable statistics on mortality, its causes and trends are in high demand for assessing the global and regional health situation and developing appropriate interventions. Countries that monitor mortality and its causes are among those that have made substantial progress in health. In the absence of routine mortality statistics (especially sub-Saharan Africa), health and socio-demographic surveillance site (HDSS) data provide a valuable source for estimating all-cause adult mortality and mortality trends. An additional benefit of HDSS implementing the verbal autopsy (VA) is that they are often the only data in many countries to monitor cause-specific mortality of a population on a longitudinal basis (1).

In 1992, the Agincourt subdistrict of Bushbuckridge was demarcated by Wits University as a site for health and socio-demographic surveillance (HDSS) and a baseline census conducted that same year (2–4). Life

Global Health Action 2010. © 2010 Benn Sartorius et al. This is an Open Access article distributed under the terms of the Creative Commons Attribution-Noncommercial 3.0 Unported License (http://creativecommons.org/licenses/by-nc/3.0/), permitting all non-commercial use, distribution, and reproduction in any medium, provided the original work is properly cited. Citation: Global Health Action Supplement 1, 2010. DOI: 10.3402/gha.v3i0.5225

expectancy among both males and females in Agincourt has significantly and steadily decreased (12 years in females and 14 years in males). The increases in mortality were most prominent in children (0–4) and young adults (20–49) where increases of two- and fivefold, respectively, have been observed when comparing mortality rates from 1992–1993 to 2002–2003. Gender differences in mortality patterns are also evident with more marked increases in females in most adult age groups (5). According to a study by Tollman et al. (6), comparing periods from 1992–1994 to 2002–2005, the increase in infectious and/or parasitic (I&P) disease mortality was significant in all age and sex groups except children aged 5–14 years (increase in HIV and tuberculosis mortality was significant however) and the elderly (65+). With respect to increased I&P disease mortality, the change was driven by HIV/TB. Age-specific mortality from non-communicable disease increased significantly in adults who were 30 years and older; the change in younger age groups was not significant. Thus the prominent increase in all-cause mortality is being driven by the large increase in I&P disease (HIV) and a modest increase in non-communicable disease (6). However, few true spatial analyses have been undertaken and thus offer an area for more detailed research within this site.

Benzler and Sauerborn (7) suggest that when population-wide intervention programs are too expensive to implement, it is necessary to limit such efforts to high risk units where certain adverse health effects are the most likely to occur. Therefore, investigating the distribution of adverse health outcomes in a population (whether random or not) should be an important objective before starting a program for primary or secondary prevention of communicable disease. It is necessary to determine whether there are clusters where adverse health outcomes seem to aggregate. If this is the case, there is a need to identify them by means of simplified scores and to develop specific health strategies targeted at these clusters (8).

Use of spatial–temporal analysis has increasingly been applied in epidemiological research in recent years (9). Advances in data availability and analytic methods have created new opportunities for investigators to improve on the traditional reporting of disease on a national or regional scale by studying variations in disease occurrence rates at a local (small-area) scale (10). Among the most important exploratory methods for epidemiology and public health are those that identify significant clusters in space and/or time (11–15). Spatial, temporal, and space–time scan statistics are now commonly used to detect and evaluate statistically significant, spatial clusters. These methods can be analyzed by using the space–time scan statistic (SaTScan™) software (16), which is used widely in an increasing number of applications including epidemiology (8, 13, 17–19) and other research fields and minimizes the problem of multiple statistical tests. SaTScan™ is useful for determining those cluster alarms that merit further investigation and those clusters that are likely to occur by chance. Despite growing applications of spatial methodology, fewer studies have analyzed spatial variation of all-cause and cause-specific mortality, with little or no work on DSS longitudinal data. Analysis in Agincourt HDSS has thus not utilized a proper spatial (and spatial–temporal) analysis of mortality or other outcomes, especially since this and other HDSS sites keep track of the coordinates of all households and update these regularly.

This study will aim to identify clusters of all-cause and cause-specific mortality within the Agincourt subdistrict, which will be important for local and national health departments to minimize morbidity and mortality through timely and spatially directed implementation of prevention and control measures in a resource limited rural area. It will also thus direct future research efforts in terms of identifying the underlying reasons (risk factors) for the observed clustering both in space and time.

Study population and methods

Description of Agincourt health and socio-demographic surveillance site (HDSS)

The Agincourt health and socio-demographic surveillance site (HDSS), established in 1992, contains a blend of former Mozambican refugees, migrant workers and a more stable permanent population (2). The site is situated in the northeast of South Africa (Fig. 1), covers an area in excess of 400 km^2 and consists of 21 villages with approximately 11,700 households and a population of 70,000 people at the end of 2007. A full geographic information system (GIS) exists for village boundaries (20) and households within the site and is updated annually. The study population comprised all individuals within the site during the period 1992–2007.

Verbal autopsy (VA) and cause-specific categories

A VA is conducted on every death to determine its probable cause (21). The Agincourt VA tool was first validated in the mid-1990s (22) and again in 2006 with particular reference to HIV/AIDS and tuberculosis (manuscript in preparation). Cause-specific fraction analysis of main or underlying causes of death was limited to 1992–2006 as assessments of VAs for 2007 have not yet been completed.

Mortality rate trends over time

Data on population size, structure, and deaths were extracted from the Agincourt HDSS using Microsoft SQL Server 2005. Data cleaning were done in Stata 10.0. Precise person-years (PY) at risk by age, gender, year, and village were used as the denominator. Observation dates

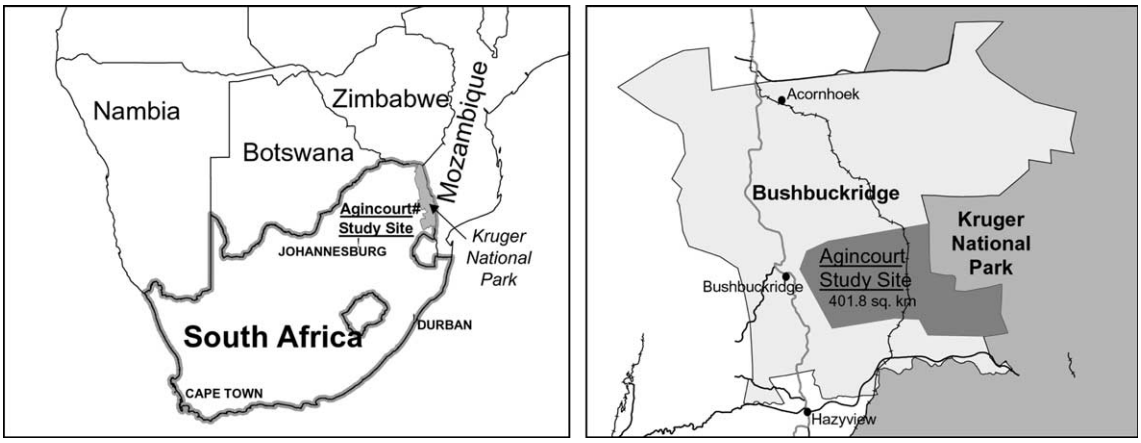


Fig. 1. Maps showing the regional location of the Agincourt health and socio-demographic surveillance site (Source: Kahn K et al. Research into health, population and social transitions in rural South Africa: data and methods of the Agincourt health and socio-demographic surveillance system. Scand J Public Health 2007; 35: 8–20).

were used for the calculation of person–time as they are the most reliable. We calculated the mortality rates by village and year by dividing the observed number of deaths by the total person-years contributed in village i ($i=1,\ldots, 21$) at year j ($j=1992,\ldots, 2007$). To identify villages in which the mortality rate was significantly above average in time, we constructed exact 95% confidence intervals (CI) for each rate using the Poisson distribution of the observed number of events i.e. deaths (23). Village mortality was considered significantly above average for a given year if the overall rate for the given year was below the lower value ($\alpha=0.025$) of the mortality rate CI for that village (24). Temporal trends in rates were analyzed in Stata by using a simple Poisson regression model containing person–time exposure, a constant and temporal (annual) trend term (25).

Mortality clustering technique (Kulldorff Scan statistic and SaTScan™)

In this study, the Kulldorff spatial scan statistic (26) was used to identify space-only clusters of high mortality only by age-group in the Agincourt HDSS overall for the entire aggregated period (1992–2007). A circular window is imposed on a map by the statistic and the center of the circle moves across the study region. This window is centered on each of the possible grid points (village centroids) positioned throughout the study region; the radius of the circle changes continuously between zero and a specified upper limit and is thus flexible both in location and size. Each of these circles can contain a different set and number of neighboring villages, and each of the circles is a potential cluster of age-specific deaths in the Agincourt study area. A village is captured in the cluster if it lies within the circle. The spatial scan statistic calculates the likelihood of observing the number of deaths inside and outside each circle, and the one with the maximum likelihood is defined as the most likely cluster i.e. least likely to have occurred by chance (tests the null hypothesis that the risk of dying is the same in all villages in the study area). Kuldorff et al. (13) also extended the spatial scan statistic into a space–time scan statistic. The window imposed by the statistic on the study area is cylindrical with a circular geographical base and height corresponding to time. The center is again one of several possible village centroids located throughout the Agincourt study area and the height reflects the time interval. The cylindrical window is then moved in space and time. This was also applied to the Agincourt HDSS data for the period 1992–2007 (time aggregation of 1 year) to identify high space–time clusters only. The following age groups were used: <5 years, 5–14, 15–49, 50–64, and 65+. Person–time by age group, gender, and village was used as the denominator. To ensure sufficient statistical power, the number of Monte Carlo replications was set to 19,999. The p-value of the statistic is obtained through Monte Carlo hypothesis testing. SaTScan™ gives the most likely cluster with a corresponding p-value (significant was set at the 5% level in this study). If other clusters not overlapping with the most likely cluster are identified (secondary, tertiary, etc.), these are also given with their corresponding p-values. Maps showing all significant non-overlapping clusters were constructed in MapInfo Professional 9.5. Larger circles do not represent greater risk clusters but rather contain a larger number of neighboring villages i.e. extend over larger geographical area. Village centroids were not displayed to preserve confidentiality in a small geographic area.

Results

Description of mortality in study sample

During 1992–2007 the highest mortality rates were observed among children, 50–64 and 65+ (9, 19, and 46 per 1,000 person-years, respectively) (Table 1). Similar

Citation: Global Health Action Supplement 1, 2010. DOI: 10.3402/gha.v3i0.5225

Table 1. Crude mortality rates (per 1,000 person-years) by age group and gender, Agincourt sub-district, 1992–2007

	Females				Males				Overall			
Age group	Deaths	Person years	Death rate	95% CI for death rate	Deaths	Person years	Death rate	95% CI for death rate	Deaths	Person years	Death rate	95% CI for death rate
<5	654	71,435.9	9.2	8.47–9.88	686	71,132.48	9.6	8.94–10.39	1,343	142,602.3	9.4	8.92–9.94
5–14	119	140,287.9	0.8	0.7–1.02	131	140,139.7	0.9	0.78–1.11	250	280,438.6	0.9	0.78–1.01
15–49	1,663	261,594.7	6.4	6.06–6.67	1,883	242,948.5	7.8	7.4–8.11	3,546	504,543.9	7.0	6.8–7.26
50–64	511	36,928.2	13.8	12.66–15.09	742	28,445.64	26.1	24.24–28.03	1,253	65,373.88	19.2	18.12–20.26
65+	1,057	28,272.9	37.4	35.17–39.71	967	15,954.2	60.6	56.85–64.55	2,024	44,228.54	45.8	43.79–47.8
Total	4,004	538,519.6	7.4	7.21–7.67	4,409	498,620.5	8.8	8.59–9.11	8,416	1,037,187.2	8.1	7.94–8.29

mortality rates were observed by age group and gender except in 50–64 and 65+ years where males had much higher rates.

Temporal trends

A significant increase in the mortality rate over time was observed from 4.7 deaths per 1,000 person years (95% CI: 4.16–5.32) in 1992 to 12.5 deaths per 1,000 person years (95% CI: 11.35–13.82) in 2007. A significant increase in the mortality rate in all villages over time was observed. Overall there were significantly higher mortality rates in one village in the upper central part of the site and two in the south-east (Table 2). Two villages (both in the south-east part of the site) showed significantly higher mortality rates during specific periods, one in 2000–2003 and the other in 2004–2007. Several villages showed excessive increases in mortality when comparing the rate in the first to last period, with all but two (one in the west and the other in the upper central region) situated toward the eastern part of the site (Table 2).

There were significant increasing trends in mortality for <5, 15–49, and 50–64-year age groups during the period 1992–2007 (Fig. 2). Mortality in the 5–14-year age groups remained constant and at a low level. Significant increases in mortality in the age group 50–64 for both genders occurred but are more pronounced among males. The elderly (65+) had the highest mortality rates, higher (and slightly increasing) in males than females (constant).

Cause-specific fractions (1992–2006)

Among children (<5) and adults (15–49), I&P-related diseases remains the highest causes of death (560 or 48% of 1,165 and 1,306 or 45% of 2,883, respectively). This is largely due to HIV/TB mortality, which accounted for 23% (273) and 40% (1,141) of child and adult mortality, respectively. Diarrhea and acute respiratory illness (ARI) feature as prominent causes of death among children (145 deaths or 12% and 94 or 8%, respectively). HIV/TB featured as a prominent cause of death in the 50–64-year age group (244 or 23% of 1,077).

Vascular disease (all circulatory system disease) and cancer (neoplasm) feature as the most prominent non-communicable causes of death, particularly in older age groups where they accounted for 14% (151) and 7% (76) in those 50–64 years and 22% (401) and 11% (193) in those 65+. Malnutrition is a prominent cause of death among children (8% or 92 deaths). Vehicle accidents followed by assault are the two leading external causes of death (4% or 300 deaths and 2% or 170 deaths overall). Suicide was highest among children aged 5–14 years and adults 15–49 (3% or 6 deaths and 2% or 63 deaths, respectively).

Table 2. Crude mortality rates overall and by village in the Agincourt sub-district, 1992–2007

			Death rate per 1,000 person years					
Village	Total deaths	Person years	Overall (95% CI)	1992–1995	1996–1999	2000–2003	2004–2007	Rate change (95%CI) α
1	828	106,445	7.8 (7.3–8.3)	5.4	4.9	9.4	11.7	6.4 (5.9, 6.8)
2	344	47,642	7.2 (6.5–8.0)	4.5	5.6	8.8	10.0	5.5 (4.9, 6.2)
3	640	82,066	7.8 (7.2–8.4)	4.8	6.9	8.4	11.1	6.3 (5.8, 6.8)
4	421	50,498	8.3 (7.6–9.2)	5.8	5.8	10	11.5	5.7 (5.1, 6.2)
5	315	40,215	7.8 (7.0–8.8)	4.7	6.9	9.1	10.5	5.8 (5.1, 6.5)
6	456	55,304	8.3 (7.5–9.0)	4.1	6.2	9.8	13.0	**8.9 (8, 9.8.0)**
7	282	35,393	8.0 (7.1–9.0)	3.8	5.6	9.6	12.9	**9.1 (8.0, 10.2)**
8	675	74,735	**9.0 (8.4–9.7)**	5.9	6.2	10.1	**13.9**	**8.0 (7.4, 8.6)**
9	550	69,795	7.9 (7.2–8.6)	5.0	5.6	8.5	12.5	**7.4 (6.8, 8.0)**
10	519	66,170	7.8 (7.2–8.6)	5.6	5.9	8.4	11.1	5.5 (5.1, 5.8)
11	738	94,577	7.8 (7.3–8.4)	5.3	5.6	8.3	11.6	6.3 (5.9, 6.6)
12	253	29,467	8.6 (7.6–9.7)	5.5	5.8	10	12.4	6.9 (6.2, 7.5)
13	409	50,503	8.1 (7.3–8.9)	5.1	6.3	10.8	10.2	5.2 (4.6, 5.7)
14	242	27,863	8.7 (7.6–9.9)	4.3	5.9	10.1	14.6	**10.3 (9.0, 11.7)**
15	432	46,907	**9.2 (8.4–10.1)**	6.1	7.2	10.5	12.9	6.8 (6.2, 7.4)
16	527	61,783	8.5 (7.8–9.3)	5.4	6.1	10.1	13.0	**7.7 (7.0, 8.4)**
17	263	34,933	7.5 (6.7–8.5)	5.4	4.5	9.5	12.4	7.0 (5.8, 8.3)
18	115	15,765	7.3 (6.0–8.8)	4.7	6.2	7.6	11.4	6.7 (5.3, 8.3)
19	156	19,049	8.2 (7.0–9.6)	6.3	6.1	8.6	13.8	7.4 (5.7, 9.6)
20	132	16,136	8.2 (6.8–9.7)	5.9	6.1	**12.7**	8.5	2.6 (1.8, 3.6)
21	122	11,990	**10.2 (8.5–12.2)**	–	2.7	8.9	12.5	**9.8 (9.4, 6.1)**
Overall	8,419	1,037,238	8.12 (7.9–8.3)	5.2 (4.9, 5.5)	5.9 (5.6,6.2)	9.4 (9.0,9.7)	12.0 (11.5,12.4)	6.8 (6.6, 6.9)

Note: Bold numbers indicate mortality rates significantly above the average for a given period ($p < 0.05$). α: Rate difference first period (1992–1995) versus last period (2004–2007) except for village 21, which compares first available period (1996–1999) to last period.

Spatial only analysis

All-cause

Toward the south-east corner of the site, a statistically significant (at 5% level) cluster of higher mortality comprising five villages was observed for the period 1992–2007 (observed deaths = 1,831, expected deaths = 1706, RR = 1.09, $p = 0.025$).

All-cause by age group

With the exception of 65+ mortality, all significant clusters of higher mortality were in the south-east corner of the site. There were clusters of higher child (<5) and adult mortality (15–49 years) in one village in the upper central region (Table 3). No significant clusters were identified for <1, 5–14, and 50–64-year age groups.

Spatial–temporal analysis

All-cause

There were three statistically significant space–time clusters of higher all-cause mortality. The most likely cluster was situated in the south-east corner and comprised six villages for the period 2002–2007 (observed deaths = 1,155, expected deaths = 789, RR = 1.54, $p < 0.001$) using the space–time scan statistic. A secondary cluster of 7 villages was situated in the upper central to east region of the site during the period 2001–2007 (observed deaths = 1,237, expected deaths = 898, RR = 1.44, $p < 0.001$); while a tertiary cluster of three villages was situated in the central/west region during 2002–2007 (observed deaths = 1,038, expected deaths = 742, RR = 1.46, $p < 0.001$).

All-cause by age group

Spatial–temporal clustering of age-specific all-cause mortality can be seen in Table 4. Significant space–time clusters of higher all-cause mortality were observed among children in six villages in the upper central region (mostly likely) of the site during 1999–2006 (233 observed cases, 148 expected, RR = 1.70, $p < 0.001$); and in five villages in the south-east (secondary cluster) during the same period (227 observed cases, 150 expected, RR = 1.62, $p = < 0.001$). During 2001–2007, three significant clusters of high adult mortality (15–49) were observed. Most likely cluster during 2001–2007 was in the south-east corner of the site comprising seven villages

Citation: Global Health Action Supplement 1, 2010. DOI: 10.3402/gha.v3i0.5225

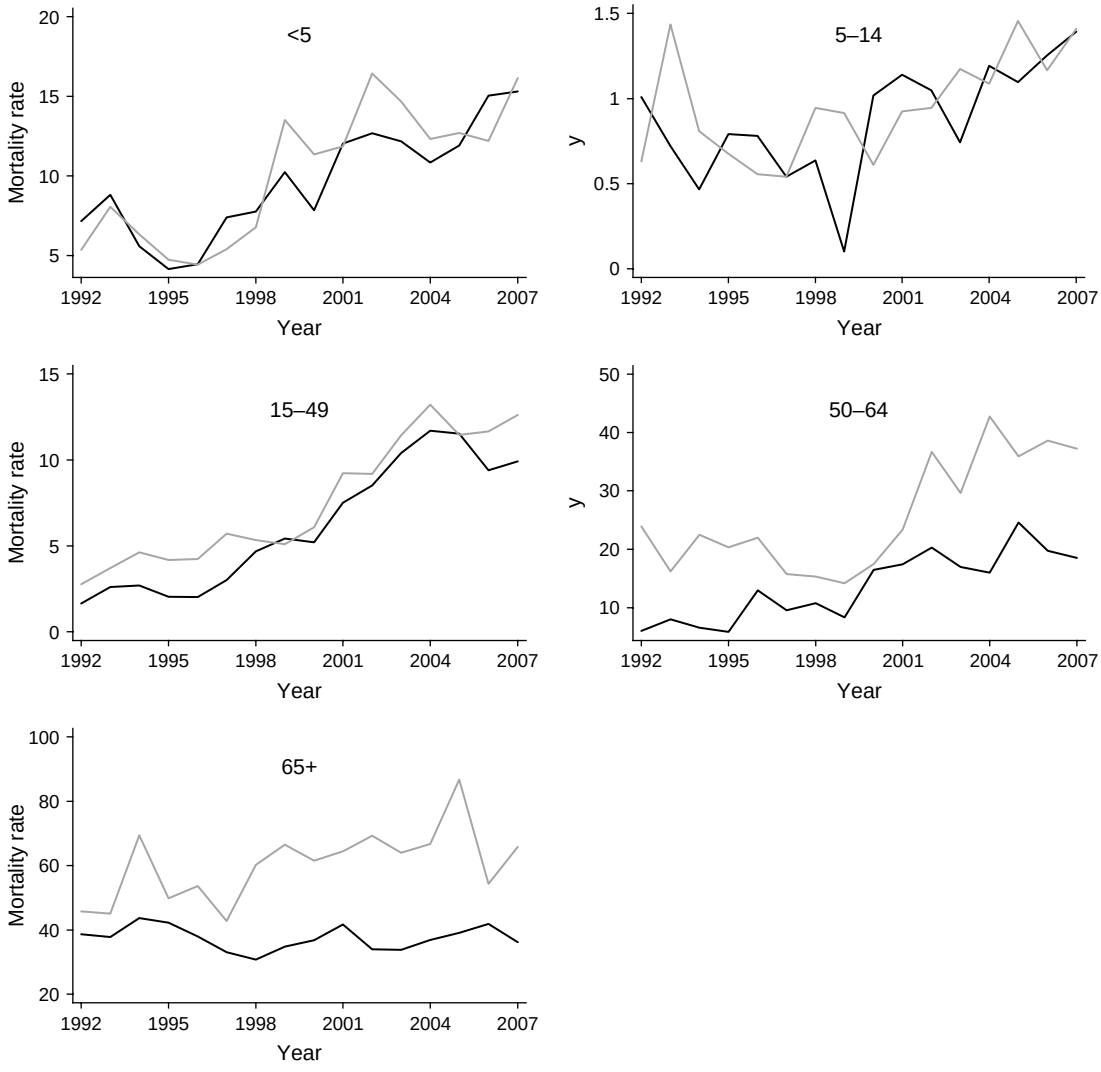


Fig. 2. Crude mortality rates (per 1,000 person years) by gender (female = black, male = light gray), age group and year, Agincourt sub-district, 1992–2007.

(638 observed cases, 385 expected, RR = 1.80, $p < 0.001$); a secondary cluster of seven villages in the upper central/east region during the same period (602 observed cases, 402 expected, RR = 1.60, $p < 0.001$); and a tertiary cluster of three villages in the west/central region during 2003–2007 (426 observed cases, 278 expected, RR = 1.60, $p < 0.001$). Significant clusters of higher older adult mortality (50–64) were observed in similar areas during similar periods (Table 4). No significant space–time clusters of all-cause mortality were identified for the 5–14 and 65+ age groups. Graphical depictions of clusters by age group can be seen in Fig. 3.

Discussion

Demographic surveillance systems provide a viable method for the collection of reliable data on vital events in rural sub-Saharan Africa, especially in the absence of accurate routine mortality statistics. Increasingly, there is renewed interest in the spatial clustering of infectious disease and mortality, especially in poor areas with limited resources. Little proper spatial analysis of longitudinal HDSS data has been done thus far. This study has demonstrated the usefulness of Kulldorff's scan statistic in highlighting high risk areas within the Agincourt sub-district for future targeting of health interventions, as well as focusing more detailed research regarding the underlying risk factors (at individual, household or community level) that may be driving these spatial–temporal all-cause and cause-specific mortality patterns. This study should be regarded as a first step in prioritizing areas for follow-up public health efforts and evaluating their impact (e.g. ARV rollout started in this area in 2007).

Increasing trends in mortality were observed in most age groups (<5, 15–49, and 50–64 years) during the period 1992–2007, largely due to the HIV epidemic. As

Table 3. Clusters of all-cause mortality by age group, using the purely spatial analysis scanning for high mortality rates, Agincourt sub-district, 1992–2007

Age group	Type	Number of villages	Location within site	Observed cases	Expected cases	Relative risk (RR)	*p*-Value
<5	Most likely	7	South east corner	526	439	1.32	<0.001
<5	Secondary	6	Upper central	35	18	2.02	0.011
15–49	Most likely	5	South east corner	790	703	1.16	0.013
15–49	Secondary	1	Upper central	73	48	1.55	0.025
65+	Most likely	9	Central/West	1,039	961	1.17	0.024

can be seen in the cause-specific fractions, all I&P mortality (mainly HIV) is the leading cause of death in this population. Thus, mother-to-child HIV transmission prevention in clusters with high child mortality needs to be undertaken, along with other interventions.

Several statistically significant clusters of higher all-cause and cause-specific mortality rates were identified among the 21 villages within the Agincourt sub-district both in space and space–time. The south-east and upper central regions of the site were consistently identified as high risk clusters, i.e. a non-random distribution. Former Mozambican refugees (about a third of the Agincourt population) entered South Africa via the Kruger National Park situated along the eastern border of the site and settled in this area. Kahn indicates that they are a vulnerable subgroup, poorer in more isolated villages with less infrastructure and generally further away from health facilities, with poor access to water and sanitation as well as labor markets (27). It also appears that the burden of communicable disease mortality (specifically HIV/TB and diarrhea) is highest in these areas (upper central and south-east for HIV/TB and south-east for diarrhea), this all leading to the all-cause space and space–time findings. Thus, settlements comprising former Mozambican refugees as well as selected South African villages appear to have increased risk of I&P-related mortality. Suitable interventions such as ARV treatment and assessment and provision of adequate water and sanitation need to be directed to these villages to overcome existing inequalities. More detailed research to elucidate the exact risk factors and the relative contribution of each needs to be undertaken. The confounding effect of settlement specific socio-economic status (SES) in space and time also needs to be adjusted for in future studies.

From the space–time analysis we observed that most of the significant mortality clusters appeared during the later period (1999–2007) with none in the earlier period (1992–1998) (Table 4). Significant increases in mortality rates particularly in <5, 15–49, and 50–64-year age groups were observed (6). Hence this temporal increase in spatial clustering – a newly described phenomenon – is also linked to the increase in mortality over the time period.

A significant space cluster of older adult (65+) mortality (as well as a space–time cluster of higher mortality among 50–64 year olds during 2002–2006) was observed toward the west of the site in the later period. The study by Tollman et al. (6) also found a significant increase in the mortality rate from non-communicable diseases in adults 30+ from 1992–1994 to 2002–2005 (RR = 1.22, $p = 0.026$). As noted, most of the self-settled Mozambican settlements are to the east of the site with more of the South African settlements to the west. According to Hargreaves et al. (28), Mozambican households generally have a lower standard of living than South

Table 4. Clusters of all-cause mortality by age group using space–time scan analysis scanning for high mortality rates, Agincourt sub-district, 1992–2007

Age group	Type	Years	Number of villages	Location within site	Observed cases	Expected cases	Relative risk (RR)	*p*-Value
<5	Most likely	1999–2006	6	Upper central/east	233	148	1.70	<0.001
<5	Secondary	1999–2006	5	South east	227	150	1.62	<0.001
15–49	Most likely	2001–2007	7	South east corner	638	385	1.80	<0.001
15–49	Secondary	2001–2007	7	Upper central/east	602	402	1.60	<0.001
15–49	Tertiary	2003–2007	3	West/central	426	278	1.60	<0.001
50–64	Most likely	2002–2007	4	Upper east	151	94	1.69	<0.001
50–64	Secondary	2003–2006	1	South east	54	25	2.22	<0.001
50–64	Tertiary	2002–2006	3	Central/West	154	109	1.47	0.026

Citation: Global Health Action Supplement 1, 2010. DOI: 10.3402/gha.v3i0.5225



Fig. 3. Significant clusters of all-cause mortality by age group in Agincourt sub-district using space–time analysis (Note: Figure shows administrative boundary of lower Bushbuckridge municipality; 1 = most likely cluster, 2 = secondary cluster, 3 = tertiary cluster).

African households and were three times more likely to fall in the poorest quintile than South African households. Thus the South Africans' higher standard of living may be contributing to a relatively higher spatial risk of non-communicable disease. This needs to be investigated in more detailed future studies.

The increasing use of linked social-spatial and health-spatial data raises significant concerns regarding the confidentiality of research participants and the stigmatization that may arise if sensitive information were released. This is especially true in a small geographic area such as the Agincourt HDSS. Rural areas present an additional problem in that settlements are fewer, more dispersed and thus more distinct than in urban areas. Hence higher levels of buffering are required to ensure confidentiality and limit disclosure risk (29). Presenting information cartographically is a useful tool for ascertaining complex spatial patterns visually, yet disclosure risks are associated with this form of presentation (29). Increased layers (e.g. borders, roads, etc.) displayable on a map add to the security threat. In this study we removed all geographically identifying features (administrative and village boundaries, roads) from the subset of all-cause mortality maps that were developed. For other significant mortality clusters, tables describing their relative location within the site were rather used to further protect those villages with high HIV burden.

Exploratory analysis of spatial data aims to describe spatial patterns using inferential statistics (occurrence of mortality for example is random or not), and to develop of hypotheses. However, it does not answer the question as to what may be influencing the spatial patterns, while spatial modeling (incorporating spatial dependency) is better suited to predict mortality rates (e.g. at unsampled locations). A study by Sankoh et al. (30) demonstrated that mapping of mortality rates using Bayesian smoothing techniques is a useful graphical supplement to spatial analytical methods as it addresses the issue of heterogeneity in the population at risk. Future research will thus use Bayesian kriging, as suggested by Gelfand et al. (31), to produce smooth maps of mortality risk. As mentioned earlier, underlying risk factors (both quantified and unquantified) drive the spatial (and temporal) risk clustering observed in this study. Common exposures may influence mortality similarly in households of the same geographical area, introducing spatial correlation in mortality outcomes. Longitudinal data are also expected to be correlated in time. Standard statistical methods assume independence of outcome measures (e.g. mortality events) and overlook correlation biases. Recent developments recommend Bayesian techniques as the appropriate methodology for taking account of this spatial and temporal dependence. Future risk factor studies in the Agincourt subdistrict will employ Bayesian geostatistical models to correctly quantify risk factors for mortality by age group.

This study underscores the need for an exploratory approach to assess geographic and temporal patterns (both historical and emerging) in all-cause mortality within a relatively small geographic area such as the Agincourt sub-district. It highlights villages requiring more targeted health interventions, raising detailed questions regarding cause-specific and spatial–temporal changes as well as the risk factors that may drive the observed all-cause mortality patterns.

Acknowledgements

This work was supported by a grant from the INDEPTH Network as well as by a PhD fellowship from the South African Centre for Epidemiological Modelling and Analysis (SACEMA). Additional funding was provided by the MRC/Wits Rural Public Health and Health Transitions Research Unit (Agincourt) through The Wellcome Trust, UK [Grant No. 069683/Z/08/Z] and the Swiss South African Joint Research Programme. SaTScan™ is a trademark of Martin Kulldorff. The SaTScan™ software was developed under the joint auspices of (a) Martin Kulldorff, (b) the National

Cancer Institute, and (c) Farzad Mostashari of the New York City Department of Health and Mental Hygiene.

Conflict of interest and funding

The authors have not received any funding or benefits from industry to conduct this study.

References

1. Hammer G, Somé F, Müller O, Kynast-Wolf G, Kouyaté B, Becher H. Pattern of cause-specific childhood mortality in a malaria endemic area of Burkina Faso. Malar J 2006; 5: 47.
2. Tollman SM, Herbst K, Garenne M, Gear JSS, Kahn K. The Agincourt demographic and health study: site description, baseline findings and implications. S Afr Med J 1999; 89: 858–64.
3. Tollman SM. The Agincourt field site – evolution and current status. S Afr Med J 1999; 89: 853–8.
4. Collinson M, Mokoena O, Mgiba N, Kahn K, Tollman S, Garenne M, et al. Agincourt demographic surveillance system. In: Sankoh O, Kahn K, Mwageni E, Ngom P, Nyarko P, eds. Population and health in developing countries. Population, health, and survival at INDEPTH sites. Vol. 1. Ottawa: International Development Research Centre; 2002, pp. 197–206.
5. Kahn K, Garenne M, Collinson M, Tollman S. Mortality trends in a new South Africa: hard to make a fresh start. Scand J Public Health 2007; 35: 26–34.
6. Tollman S, Kahn K, Sartorius B, Collinson M, Clark S, Garenne M. Implications of mortality transition for primary health care in rural South Africa: a population-based surveillance study. Lancet 2008; 372: 893–901.
7. Benzler J, Sauerborn R. Rapid risk household screening by neonatal arm circumference. Results from a cohort study in Burkina Faso. Trop Med Int Health 1998; 3: 62–74.
8. Sankoh O, Yé Y, Sauerborn R, Müller O, Becher H. Clustering of childhood mortality in rural Burkina Faso. Int J Epidemiol 2001; 30: 485–92.
9. Elliot P, Wakefield J, Best N, Briggs D, editors. Spatial epidemiology-methods and applications. London: Oxford University Press; 2000.
10. Walter S. Disease mapping: a historical perspective. In: Elliott P, Wakefield J, Best N, Briggs DJ, eds. Spatial epidemiology – methods and applications. Oxford: Oxford University Press; 2000 pp. 223–52.
11. Alexander F, Boyle P. Methods for investigating localized clustering of disease. IARC Scientific Publications No 135, Lyon, France, 1996.
12. Hjalmars U, Kulldorff M, Gustafsson G. Risk of acute childhood leukaemia in Sweden after the Chernobyl reactor accident. BMJ 1994; 309: 154–7.
13. Hjalmars U, Kulldorff M, Gustafsson G. Childhood leukemia in Sweden: using GIS and a spatial scan statistic for cluster detection. Stat Med 1996; 15: 707–15.
14. Hjalmars U, Kulldorff M, Wahlqvist Y, Lannering B. Increased incidence rates but no space–time clustering of childhood astrocytoma in Sweden, 1973–1992: a population-based study of pediatric brain tumors. Cancer 1999; 85: 2077–90.
15. Kulldorff M. Statistical methods for spatial epidemiology: tests for randomness. In: Gatrell AC, Löytönen M, eds. GIS and health. London: Taylor & Francis; 1998 pp. 49–62.
16. Kulldorff M, Rand K, Gherman G, Williams G, DeFrancesco D. SaTScan – software for the spatial and space–time scan statistics, Version 2.1. Bethesda, Madison: National Cancer Institute; 1998.
17. Britton T. Tests to detect clustering of infected individuals within families. Biometrics 1997; 53: 98–109.
18. Huang L, Kulldorff M, Gregorio D. A spatial scan statistic for survival data. Biometrics 2007; 63: 109–18.
19. Takahashi K, Kulldorff M, Tango T, Yih K. A flexibly shaped space–time scan statistic for disease outbreak detection and monitoring. Int J Health Geogr 2008; 7: 14–27.
20. Tollman, SM, Herbst K, Garenne M. The Agincourt demographic and health study: phase 1. Johannesburg: Health Systems Development Unit, Department of Community Health, University of the Witwatersrand, 1995.
21. Kahn K, Tollman SM, Garenne M, Gear JSS. Who dies from what? Determining cause of death in South Africa's rural northeast. Trop Med Int Health 1999; 4: 433–41.
22. Kahn K, Tollman SM, Garenne M, Gear JSS. Validation and application of verbal autopsies in a rural area of South Africa. Trop Med Int Health 2000; 5: 824–31.
23. Estève J, Benhamou E, Raymond L. Statistical methods in cancer research, Vol. IV, descriptive epidemiology. Lyon, France: International Agency for Research on Cancer (IARC); 1994.
24. Pickle LW, Mason TJ, Howard N, Hoover RR, Fraumeni JF. Atlas of US cancer mortality among whites: 1950–1980. DHHS Publication No (NIH) 87–2900, 1987.
25. dos Santos Silva, I. Cancer epidemiology: principles and methods. International Agency for Research on Cancer (IARC), Lyon, France, 1999.
26. Kulldorff M. A spatial scan statistic. Commun Stat Theory Methods 1997; 26: 1481–96.
27. Kahn K. Dying to make a fresh start: mortality and health transition in a new South Africa [dissertation]. Umeå University, Faculty of Medicine, Public Health and Clinical Medicine, 2006.
28. Hargreaves J, Collinson M, Kahn K, Clark S, Tollman S. Childhood mortality among former Mozambican refugees and their hosts in rural South Africa. Int J Epidemiol 2004; 33: 1271–78.
29. VanWey LK, Rindfuss RR, Gutmann MP, Entwisle B, Balk DL. Confidentiality and spatially explicit data: concerns and challenges 2005; Available from: http://www.pnas.org_cgi_doi_10.1073_pnas.0507804102 [cited March 2010].
30. Sankoh O, Berke O, Simboro S, Becher H. Bayesian and GIS mapping of childhood mortality in rural Burkina Faso. Control of tropical infectious diseases, Uni-Heidelberg Discussion Paper, March 2002.
31. Gelfand A, Ravishanker N, Ecker M. Modeling and inference for point-referenced binary spatial data. In: Dey D, Ghosh S, Mallick B, eds. Generalized linear models: a Bayesian perspective. New York: Marcel Dekker Inc; 1999 pp. 373–86.

***Benn Sartorius**
School of Public Health
University of the Witwatersrand
7 York Road
Parktown, Gauteng, Johannesburg, South Africa
Tel: +27 72 326 0595
Fax: +27 11 717 2084
Email: bennsartorius@hotmail.com

Citation: Global Health Action Supplement 1, 2010. DOI: 10.3402/gha.v3i0.5225

BMC Public Health

RESEARCH ARTICLE Open Access

Young and vulnerable: Spatial-temporal trends and risk factors for infant mortality in rural South Africa (Agincourt), 1992-2007

Benn KD Sartorius[1*], Kathleen Kahn[1,2,3], Penelope Vounatsou[4], Mark A Collinson[1,2,3], Stephen M Tollman[1,2,3]

Abstract

Background: Infant mortality is an important indicator of population health in a country. It is associated with several health determinants, such as maternal health, access to high-quality health care, socioeconomic conditions, and public health policy and practices.

Methods: A spatial-temporal analysis was performed to assess changes in infant mortality patterns between 1992-2007 and to identify factors associated with infant mortality risk in the Agincourt sub-district, rural northeast South Africa. Period, sex, refugee status, maternal and fertility-related factors, household mortality experience, distance to nearest primary health care facility, and socio-economic status were examined as possible risk factors. All-cause and cause-specific mortality maps were developed to identify high risk areas within the study site. The analysis was carried out by fitting Bayesian hierarchical geostatistical negative binomial autoregressive models using Markov chain Monte Carlo simulation. Simulation-based Bayesian kriging was used to produce maps of all-cause and cause-specific mortality risk.

Results: Infant mortality increased significantly over the study period, largely due to the impact of the HIV epidemic. There was a high burden of neonatal mortality (especially perinatal) with several hot spots observed in close proximity to health facilities. Significant risk factors for all-cause infant mortality were mother's death in first year (most commonly due to HIV), death of previous sibling and increasing number of household deaths. Being born to a Mozambican mother posed a significant risk for infectious and parasitic deaths, particularly acute diarrhoea and malnutrition.

Conclusions: This study demonstrates the use of Bayesian geostatistical models in assessing risk factors and producing smooth maps of infant mortality risk in a health and socio-demographic surveillance system. Results showed marked geographical differences in mortality risk across a relatively small area. Prevention of vertical transmission of HIV and survival of mothers during the infants' first year in high prevalence villages needs to be urgently addressed, including expanded antenatal testing, prevention of mother-to-child transmission, and improved access to antiretroviral therapy. There is also need to assess and improve the capacity of district hospitals for emergency obstetric and newborn care. Persisting risk factors, including inadequate provision of clean water and sanitation, are yet to be fully addressed.

* Correspondence: benn.sartorius@wits.ac.za
[1]MRC/Wits Rural Public Health and Health Transitions Research Unit (Agincourt), School of Public Health, Faculty of Health Sciences, University of the Witwatersrand, Johannesburg, South Africa
Full list of author information is available at the end of the article

BioMed Central

© 2010 Sartorius et al; licensee BioMed Central Ltd. This is an Open Access article distributed under the terms of the Creative Commons Attribution License (http://creativecommons.org/licenses/by/2.0), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

Background

Infant mortality is an important health indicator of a population given its strong link to socio-economic status (SES), health service access and quality, and maternal health.

In the absence of vital events registration, health and socio-demographic surveillance (HDSS) data provide a valuable source for estimating mortality rates, trends and risk factors. HDSS sites implementing the verbal autopsy (VA) to determine probable cause of death are often the only means in most developing and many middle-income countries to observe cause-specific mortality of a population on a longitudinal basis and are a valuable tool for assessing trends in burden of disease [1,2].

Diarrhoea, pneumonia, malnutrition and malaria are the leading causes of death among infants in low income countries [3,4]. Birth asphyxia and neonatal sepsis are responsible for most neonatal deaths [3]. These diseases, that can be largely prevented or effectively treated at relatively low cost, cause almost 95% of preventable infant and child deaths [1]. HIV/AIDS has emerged as a major cause of death among infants in recent years, though in few countries outside of Africa [5].

In 1990, there was a 20-fold difference in the rate of infant deaths between sub-Saharan African and industrialized countries (180 versus 9 deaths per 1000 live births). In 2000, this difference had increased to 29-fold with mortality rates of 175 and 6 per 1000 children respectively [6]. This is because many sub-Saharan African countries have seen reversals in child mortality trends in recent years due to HIV/AIDS. In 2007, approximately 420 000 children became infected with HIV [7], mostly through mother-to-child transmission (MTCT) [8,9] in resource poor settings particularly sub-Saharan Africa. Kahn et al showed a doubling of child mortality due to HIV in a rural South African population (Agincourt sub-district) between 1992 and 2003 from 39/1000 person-years to 77/1000 [10]. Garrib et al in 2006 found very high levels of infant mortality in another rural area of South Africa, 67.5 per 1000 person-years, with HIV/AIDS estimated as the single largest cause of death in the under-5 age-group (41% of deaths) [11]. Thus interventions to reduce infant and child mortality are urgently required. A study in Zambia estimated that the cost per averted infection was approximately US$890 [12]. According to a study in Barbados the lifetime cost of treating an HIV infected child is US$ 8,665 [13]. This is much lower than estimates from the US where the cost for perinatally infected infants was USD 113,476 for 9 years of survival, US$ 151,849 for 15 years, and US$ 228,155 for 25 years [14]. According to a study in the Ivory Coast, the mean cost of treatment was € (euros) 254 per child-year for infected children, €108 more than the mean cost of treatment for HIV-negative children born to HIV-positive mothers (a 74% increase in treatment costs) [15]. Thus despite the costs associated with HIV/AIDS prevention among young children [16,17], lifetime treatments costs of HIV infected infants are far higher; hence preventive measures need to be prioritized and targeted to those at high risk in poor, resource limited settings.

Effective interventions such as prevention of mother to child transmission (PMTCT) are available. A comprehensive approach to PMTCT can reduce transmission rates to below 2% [18-20]. Yet health care access and inequity remain widespread problems in economically disadvantaged areas [21]. Mpumalanga Province in northeast South Africa was an important destination for refugees fleeing the civil war in Mozambique from 1983 onwards. A formal peace agreement was signed in 1992, yet despite voluntary repatriation programmes, by 2000 it was estimated that more than 200,000 former Mozambican refugees were still inhabitants in the province [22]. A study by Hargreaves et al [23] demonstrated higher mortality rates among children from former Mozambican refugee households when compared to those from South African-headed households in the Agincourt sub-district. They concluded that lack of legal status and poorer SES of Mozambican refugees partly explains this disparity.

Inequalities in health outcomes or access to services and benefits can occur across space and time. In some situations this can reflect a compositional effect with variations merely reflecting the different groups that inhabit different locations [24]. However, certain inequalities in child health outcomes are avoidable and unjust. These may reflect underlying inequities in the distribution of wealth, resources and social privilege in a given society, rather than an individual's choice or behaviour. To be fair, society must strive to achieve equal opportunities for all children regardless of parental status (education, SES) and geographical location. High-quality services for children that bridge the social divide are an important means of achieving equity goals. If South Africa is to achieve the Millennium Development Goals by 2015, including MDG 4 to reduce child mortality, then there is need to scale-up coverage rapidly with access to high quality health care and social support, particularly in the most poor and marginalised communities [25]. When population-wide intervention programmes are too costly to implement, it becomes necessary to target such efforts to high risk areas where adverse health events are the most likely to occur [26]. To address inequity in child survival, service planners need to understand the underlying socio-demographic

profile and other factors contributing to high risk. Spatial-temporal mapping of high risk communities identifies those with greatest need, rather than those that are easiest to reach [27]. This provides evidence on where to target interventions for greatest impact [28] and generates hypotheses on the determinants of increased risk.

With the development of Markov Chain Monte Carlo (MCMC) methods and Bayesian software such as WinBUGS, geostatistical spatial-temporal modelling has increasingly been applied in epidemiological research [29], especially with regards to malaria risk and transmission. Gemperli et al (2004) carried out a Bayesian spatial analysis of infant mortality in Bali that confirmed well-known risk factors and found a spatial pattern of infant mortality that showed a clear relationship with established foci of malaria transmission [30-32]. Despite growing applications of spatial methodology in malaria research, fewer studies have analysed spatial variations in population dynamics including all-cause and cause-specific mortality, with little or no work on longitudinal data collected in relatively small geographic areas covered by health and socio-demographic surveillance.

Many individual and household level factors have been identified as key determinants of infant and child mortality. Since objects in close proximity are often more alike, common exposures (measured or unmeasured) may influence mortality similarly in households of the same geographical area, introducing spatial correlation in mortality outcomes. Repeated measurements on individuals and households are also expected to be correlated in time. Standard statistical methods assume independence of outcome measures, for example mortality data. Ignoring this correlation introduces bias in the risk analysis as the standard error of the risk factors is underestimated, thereby overestimating significance. Bayesian geostatistical models relax the assumption of independence by, for example, incorporating random effects to measure spatial correlation as a function of distance between locations.

The aim of this study was to assess changes in infant mortality patterns in rural northeast South Africa over time, determine mortality risk factors and produce cause-specific mortality maps to identify high risk areas. These insights can provide guidance on the best allocation of limited resources to reduce infant mortality in this and similar areas of the country.

Methods

Study area and population

The Agincourt health and socio-demographic surveillance system (HDSS), established in 1992, extends over an area of about 400 km^2 and consists of 21 villages with approximately 11,700 households and a population of 70,000 people at the end of 2007 (Figure 1). A full geographic information system (GIS) covers all households within the site and is updated annually. For these analyses the study population consisted of all infants who were either born or migrated into the site between 1992 and 2007 and who either survived or died in their first year of life.

Outcome measures

A verbal autopsy (VA) was conducted on every death to determine its probable cause [34]. Interviews administered by trained lay fieldworkers were assessed independently by two physicians to determine probably cause-of-death. Where consensus could not be reached, a third independent medical assessment was made. The VA was first validated in the mid-1990s [35] and again in 2006 with particular reference to HIV/AIDS related mortality. International Classification of Diseases (ICD-10) was used to classify main or underlying, immediate and contributory causes of death. For this study, cause-specific analysis was limited to main causes from 1992-2006 as VA's had not yet been assessed for 2007.

Explanatory variables

Covariates included: infant demographic variables (gender, nationality); 5-year time periods; maternal factors (former refugee status, age at pregnancy, death in first year of child's life, education); fertility factors (parity, birth intervals, sibling death); household mortality experience, socio-economic status (SES) and food security; distance to health facility; antenatal clinic attendance; and household elevation (climatic proxy). Every two years since 2001, an asset survey was conducted in all households within the HDSS [36]. Information on living conditions and assets, building materials of main dwelling, water and energy supply, ownership of modern appliances and livestock, and means of transport available were recoded (one being higher SES and zero lower status), summed to give an overall score for a household, and then used to construct wealth quintiles for SES ranked by increasing score from most to least poor.

Statistical analysis

The negative binomial is an alternative for the commonly used Poisson distribution, often regarded as the default distribution for integer count data. The Poisson assumes that expected mean equals its variance. The negative binomial differs from the Poisson distribution in that it allows for the variance to exceed the mean. Since the negative binomial distribution has one more parameter than the Poisson distribution, the second parameter is used to adjust the variance independently of the mean. Our data displayed evidence of being highly overdispersed and thus the negative binomial model was chosen.

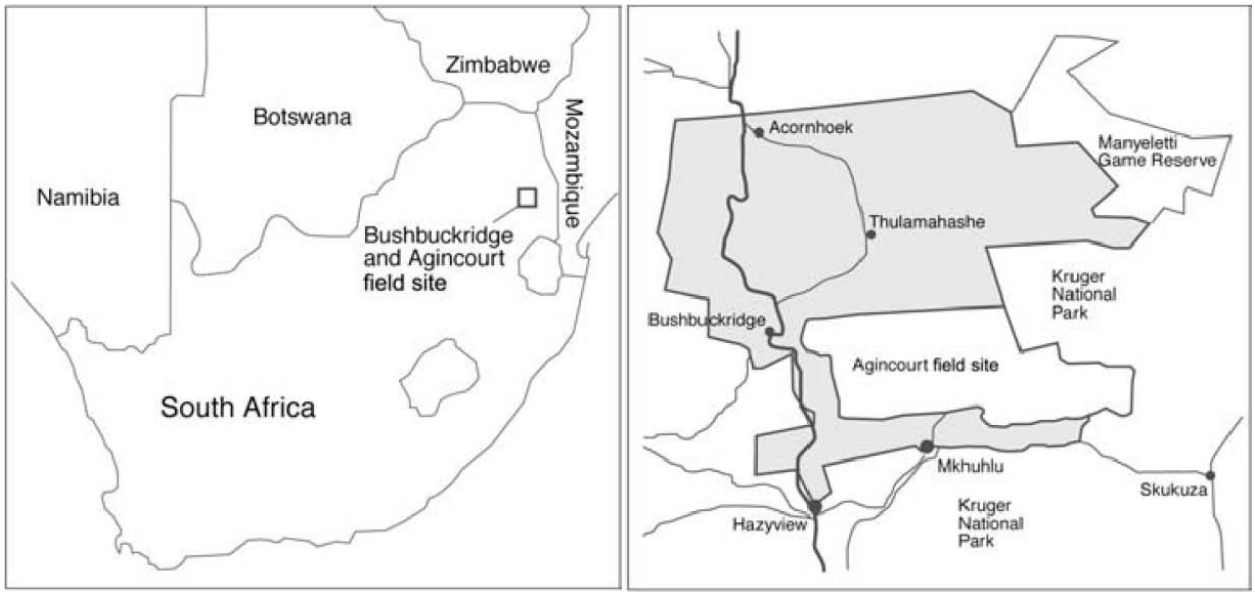


Figure 1 Location of the Agincourt HDSS site [33], South Africa.

A preliminary negative binomial regression analysis was carried out to assess the relationship between infant mortality and each covariate. Covariates significant at the 10% level (without substantial missing values) were then incorporated into the multivariate model.

The multivariate Bayesian negative binomial model was fitted in WinBUGS to examine the association between the significant covariates and all-cause infant mortality. Observation dates were used to calculate the person-days contributed by each infant (offset). Spatial random effects were used at a village level to take into account spatial correlation. Temporal random effects were also used at yearly intervals to account for temporal correlation. Village specific random effects were modelled via a multivariate Gaussian process (multivariate Gaussian distribution with covariance matrix expressed as a parametric function of distance between pairs of village centroid points) [37]. Standard Bayesian autoregressive (AR) approaches, with priors for the AR (1) and AR(2) processes defined by Schotman [38] and Zeller [39] respectively, as well as a Poisson generalized autoregressive moving average (GARMA) approach [40], were tested to model the temporal random effects. Various order models for the AR and MA terms were assessed and the one that best fitted the data was used. MCMC simulation was employed to estimate the model parameters [41]. Further details of the statistical modelling approach are given in the appendix.

Model assessment and validation

Deviance Information Criterion (DIC) [42] was used as the first step in comparison of model fit and the one giving the lowest DIC was chosen. Models were then also validated by fitting the models for 1992-2006 and predicting outcomes for all infants in 2007. Credibility intervals were constructed and the model providing the best predictions (along with low DIC) were used as the final model. The negative binomial models, particularly the AR(1) and AR(2) to model the temporal random effect, provided the lowest DIC (8618.07 and 8617.34 respectively) by some margin when compared to other approaches such as GARMA. In Bayesian statistics, a credible interval is a posterior probability interval which is used for interval estimation, in contrast to point estimation (confidence intervals). In other words, the credibility interval refers to the distribution of parameter values while a confidence interval pertains to estimates of a single value. In this study the negative binomial AR (2) predicted the outcome much better than the AR(1) model based on these Bayesian credibility intervals. Thus the AR(2) process was used in the final multivariate model.

Risk maps

A baseline model was used that included no covariates but a constant and site-specific (village centroid) random effect. All identifying features (village centroids, geographic boundaries) were removed, and the prediction area expanded irregularly (~740 km^2) to double the normal size, in order to ensure confidentiality and avoid stigmatizing of villages. The HIV/TB map is not shown for this reason. Simulation-based Bayesian kriging [43] at prediction points (regular grid) within the site was used to produce maps of mortality risk for the whole

HDSS site. Model estimates were exponentiated to represent incidence rate ratios (IRR).

Software

Data extraction and management was done using Microsoft SQL Server 2005. The analysis was carried out in STATA 10.0, WinBUGS and R. The predictions of the fitted spatial models were mapped in MapInfo Professional 9.5.

Results

Demographic profile of study sample

Between 1992 and 2007 31,804 infants were either born or migrated into the Agincourt HDSS. Of these, 26,000 (81.8%) were born within the site and half (50.4%) were female. Just under two-thirds were South African citizens (20,375; 64.2%) and a little over one-third were born to Mozambicans (11,356; 35.8%). There were 737 infant deaths (2.3%) giving an overall mortality rate of 24.7 per 1,000 person years; of these, 175 deaths were within the perinatal period and 202 within the neonatal period.

Cause of death (1992-2006)

The top causes of death among infants, as assessed by verbal autopsy, were HIV/TB (n = 116), acute diarrhoea or malnutrition (n = 91), acute respiratory infection (ARI) or pneumonia (n = 82) and septicemia (n = 20). In total 300 infant deaths were attributed to infectious and/or parasitic causes. During 1992-2006, 230 infants (33.6%) had an unknown cause of death.

Temporal trends by cause

There was a significant increasing trend in the infant mortality rate over the study period (IRR = 1.09, 95%CI: 1.05-1.12, $p < 0.001$) (Figure 2). A significant increasing trend (at 10% level) was also observed for all-cause neonatal (first 28 days of life) mortality rate (IRR = 1.04, 95%CI: 1.00-1.08, $p = 0.068$), particularly from 1996 onwards. Mean time to death among neonates was 4.49 days (SD 6.05) indicating that most occur in the perinatal period (first 7 days of life). Between 1992 and 2006 the infant mortality rate due to HIV/TB significantly increased from 0 to 10.95 deaths per 1000 person years (IRR = 1.27, 95%CI: 1.17-1.38, $p < 0.001$), the increase commencing from about 1998. No significant changes were observed for infant deaths due to acute diarrhoea or malnutrition and ARI or pneumonia. A significant increasing trend was observed with infant deaths attributed to unknown (R99) causes (IRR = 1.27, 95%CI: 1.19-1.36, $p < 0.001$), particularly from 1998 onwards. A significant (IRR = 1.14, $p < 0.001$) and striking increase in the mortality rate of mothers dying in the infants' first year was also observed (Figure 3), again from 1998 onwards.

Univariate analysis

Later time period, higher number of cumulative household deaths, death of previous sibling and mother dying in the first year of infant's life were large and highly significant risk factors for infant mortality (Table 1). Male gender and increasing birth parity were also found to be significant risk factors. Breast feeding had a protective

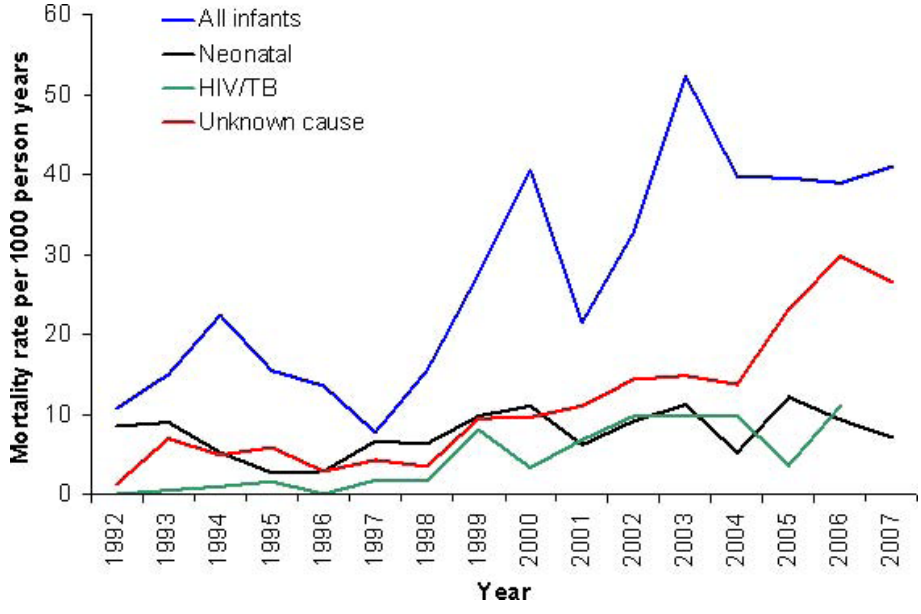


Figure 2 All-cause neonatal and infant mortality rates per 1,000 person years, Agincourt sub-district 1992-2007.

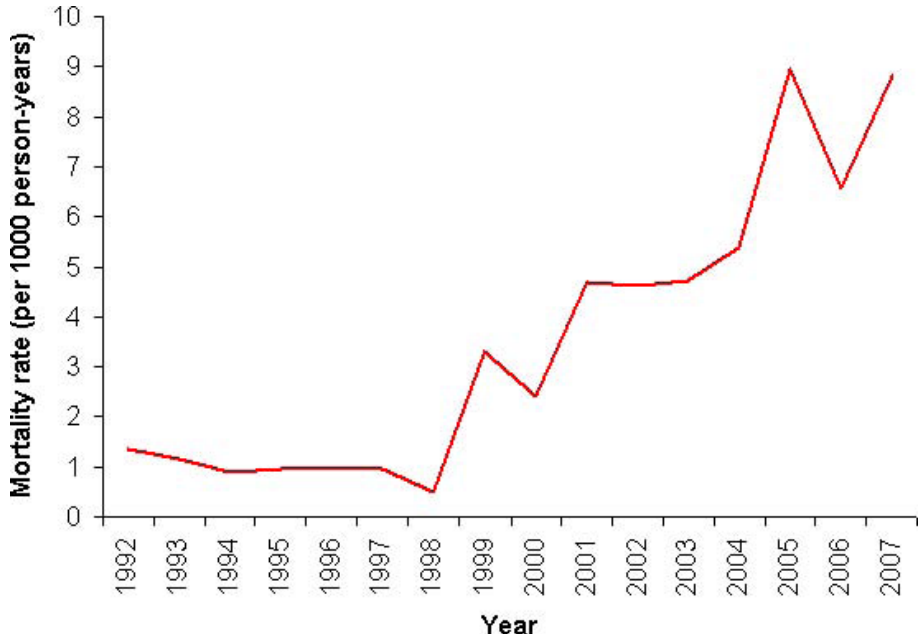


Figure 3 Mortality rate of mothers dying in infants' first year per 1000 infant person-years, Agincourt sub-district, 1992-2007.

influence on all-cause as well as diarrhoea and malnutrition-related infant mortality (Table 1). Increasing infant weight at birth also had a significantly protective effect. High (post-secondary) level of maternal education, mother attending antenatal clinic and increasing number of antenatal clinic visits were found to be significantly protective. Mozambican origin of mother was not found to be a risk factor for all-cause infant mortality. However, mother having arrived from Mozambique post 1992 was found to be a significant risk factor for death due to infectious and/or parasitic causes. Increasing distance to nearest health facility was not a significant risk factor and no differential health care access was observed by nationality (South African versus Mozambican). Mother being a migrant was found to be significantly protective and, conversely, increasing number of months spent resident in the site per year by the mother was found to be a risk. Further, migrant mothers were found to be significantly more educated than mothers permanently resident in the study site ($p < 0.001$) and came from households with a significantly higher SES ($p = 0.0025$). No significant difference was found in antenatal clinic attendance between permanent and migrant mothers.

There was a strong non-linear reduction in the probability of death as the infant progresses to the end of their first year (Figure 4). This risk over time was much higher for those infants whose mother died in their first year, and remained elevated for the remainder of the first year.

A significant increase in the number of years of maternal education as well as antenatal clinic attendance was observed over the study period (both $p < 0.001$). However, significant increases in the number of mothers dying in their infants' first year, number of maternal deaths (< = 42 days after infants date of birth), as well as other household deaths over time was also observed (all $p < 0.001$).

Almost a third (30.2%, n = 91) of mothers who died in the infants' first year died of HIV/TB; this was a significant risk factor for infant mortality (IRR = 164.7, $p < 0.001$). Approximately 44% of the mothers that died in the infant's first year died of unknown causes, many probably unclassified HIV-related deaths.

Household water supply consisting of raw natural water (river, pond or dam) was a risk factor (IRR = 16.50, $p = 0.010$) for deaths due to acute diarrhoea and malnutrition, although numbers were small. Mother being of Mozambican origin also proved to be a significant risk factor for infant death due to diarrhoea or malnutrition (IRR = 1.66, $p = 0.019$).

Multivariate analysis

Later year of birth, mother dying in the infant's first year, higher number of cumulative household deaths and previous birth being stillborn remained highly significant in the all-cause multivariate model (Table 2). Following multivariate adjustment, large IRR values were again observed for mother death in the infant's first year, and cumulative household deaths. Death of

Table 1 All-cause univariate risk factor analysis for infant mortality in the Agincourt sub-district, 1992-2007

Factor	n	IRR	p-value	signif
Temporal				
1 year continuous	31,804	1.23	<0.001	*
5 year period	31,804			
1992-1996	10,744	1.00		
1997-2001	10,624	4.61	<0.001	*
2002-2006	10,436	10.76	<0.001	*
Demographic				
Male gender	31,804	1.86	0.006	*
Mother refugee status	29,068			
South African citizen	18,746	1.00		
Mozambican origin	10,322	1.12	0.638	
Breast feeding and birth weight				
Breast fed	23,890/25,697	0.21	<0.001	*
Breast fed (diarrhoea & malnutrition)	23,890/25,697	0.38	0.001	*
Increasing birth weight (kilograms)	15,235[β]	0.42	<0.001	*
Maternal				
Mother Mozambican in-migrated post 1992 [α]	10,322	4.89	0.004	*
Mother status	31,041			
Mother in same household	27,076	1.00		
Mother not in household	2,488	0.01	<0.001	*
Mother death	1,477	5.79	<0.001	*
Mother residency status				
Permanent (> = 6 months in site)	28,852	1.00		
Migrant	25,200	0.45	0.059	#
Other	2,035	0.36	0.067	#
Increasing number of months resident during the previous 12 months	28,962	1.11	0.001	*
Mother died in child's first year	91/31,804	65.72	<0.001	*
Mother age at pregnancy	27,981	0.99	0.595	
Mother education	16,971			
None or primary	6511	1.00		
Secondary	9561	0.90	0.677	
Tertiary	899	0.13	<0.001	*
Paternal				
Father died in child's first year	59/31,804	0.69	0.834	
Father died before birth	57/31,804	0.98	0.993	
Household head	26,034			
Male gender	16,625	0.86	0.331	

Table 1 All-cause univariate risk factor analysis for infant mortality in the Agincourt sub-district, 1992-2007 *(Continued)*

Mozambican origin	9,972	1.04	0.808	
Age (years)	25,660			
18-29	2,995	1.00		
30-49	13,143	0.93	0.785	
50-64	6,486	1.39	0.243	
65+	3,036	1.43	0.272	
Household morbidity and mortality				
Cumulative number of household deaths in year of birth (continuous)	31,804	14.64	<0.001	*
None	26,444	1.00		
1	4,048	35.51	<0.001	*
2-3	1,232	109.34	<0.001	*
4+	80	131.47	<0.001	*
Number of household admissions in year of birth (continuous)	2,895	1.55	0.093	#
None	2,392	1.00		
1-2	402	1.51	0.464	
3+	101	17.80	0.041	*
Fertility				
Birth parity (continuous)	25,083	1.22	0.170	
1	25,083	1.00		
2-3	16,526	2.32	0.001	*
4+	7,663	0.55	0.313	
Preceding birth interval	8,421	1.05	0.001	*
Post birth interval	7,199	0.51	<0.001	*
Previous birth stillborn	318/27,981	16.71	<0.001	*
Previous sibling died	945/27,981	7.48	0.001	*
Preceding interval sibling death	413	1.23	0.413	
Mother attended antenatal clinic	24,928	0.06	<0.001	*
Number of antenatal clinic visits	17,305	0.84	<0.001	*
Socio-economic status of household				
SES absolute score (quintiles)	9,397			
Most poor	1,584	1.00		
Very poor	1,801	0.81	0.540	
Poor	1,971	0.91	0.780	
Less poor	2,019	0.91	0.780	
Least poor	2,022	0.98	0.940	
Food security status of household				
Predicted food shortage in coming year	3,481			
Same amount of food	1,122	1.00		
More food	377	0.31	0.091	#
Less food	1,982	1.24	0.597	
Distance to nearest health facility				
Minimum distance to health facility (straight-line) from household	25,749			
< 5 km	24,023	1.00		
> = 5 km	1,726	1.18	0.607	
Minimum distance to health facility (network) from village centroid	31,804			

Table 1 All-cause univariate risk factor analysis for infant mortality in the Agincourt sub-district, 1992-2007 *(Continued)*

<5 km	1,220	1.00		
>= 5 km	30,584	1.28	0.684	
Climatic				
Elevation (meters) - rainfall proxy	30,583			
350-399	2,554	1.00		
400-449	14,202	0.43	0.072	#
450-499	4,120	0.25	0.009	*
500-549	7,202	0.17	<0.001	*
550-599	2,505	0.46	0.226	

Total sample size (n = 31,804).

α: infectious and parasitic deaths only.

β: birth weight data only available for this number of infants.

* significant at the 5% level; # significant at the 10% level.

previous child was also a significant risk factor at the 10% level. There was more temporal than spatial correlation in the final spatial-temporal model (0.44 versus 0.09). The spatial-temporal model estimated the range (distance at which spatial correlation ceases) to be 5,224 metres (95%: 1,805-21,420) meters. AR(1) and AR(2) parameters were between -1 and 1 indicating stationarity.

Risk maps

All-cause

Figure 5 shows all-cause and neonatal mortality risk. Note that with increasing distance from locations with observed mortality (i.e. villages), the standard error of the prediction increases. The map of all-cause infant mortality risk reveals the highest risk to be among villages bordering the Kruger National Park to the east of the site, running from the upper central towards the south-east. Distinct foci of high mortality risk can be identified in two villages in particular. Neonatal mortality displayed a similar pattern with 2 distinct foci of higher risk. One of these foci is in close proximity to a health facility. Figure 6 shows a distinct pattern of higher infectious and/or parasitic causes, including HIV/TB, towards the east of the site, again with distinct foci of higher mortality in former refugee settlements. Acute diarrhoea and malnutrition showed a distinct cluster of higher mortality in the south-east. The pattern of ARI or pneumonia infant mortality risk was less distinct, though two foci could be observed to the east of the study area.

Discussion

The results indicate a worsening of infant mortality: year of birth was significantly associated with infant mortality and risk of death increased over the study period. The increase was particularly from 1998 onwards, and can be largely attributed to the HIV epidemic and its impact on mortality in the study area [9,44], both direct (vertical transmission of HIV) and indirect (death of a caregiver). Mother's death in infant's first year was a major risk factor in this study, as was higher numbers of cumulative household deaths. Results confirmed the importance of other known risk factors [23,30]. The protective association between increasing maternal education and infant mortality has been previously described [30,45] and is possibly a result of better health awareness and utilization of health facilities [46], longer birth intervals [47], and higher income which improves infants' health through ability to purchase goods and services [48]. A significant association of higher household SES was however not observed in this study. This has been shown elsewhere and may be explained by that fact that unlike endogenous maternal and demographic factors that substantially influence an infant's risk of death, the effects of SES factors on mortality increase as the child gets older due to exogenous factors which parents have more control over [49].

We examined health service access with respect to primary health care generally and antenatal care specifically. Distance to nearest primary health care facility was not a risk factor in this study. Antenatal clinic attendance and number of ANC visits was significantly protective, with no difference between South Africans and former Mozambican refugees. These finding suggest that factors other than geographic access may be key to understanding the risks associated with health care utilisation. These could include quality of care, level of available care (primary versus secondary), cost and social barriers. In South Africa, primary health care for children under the age of six is free, as is antenatal care. However, financial costs associated with transport and

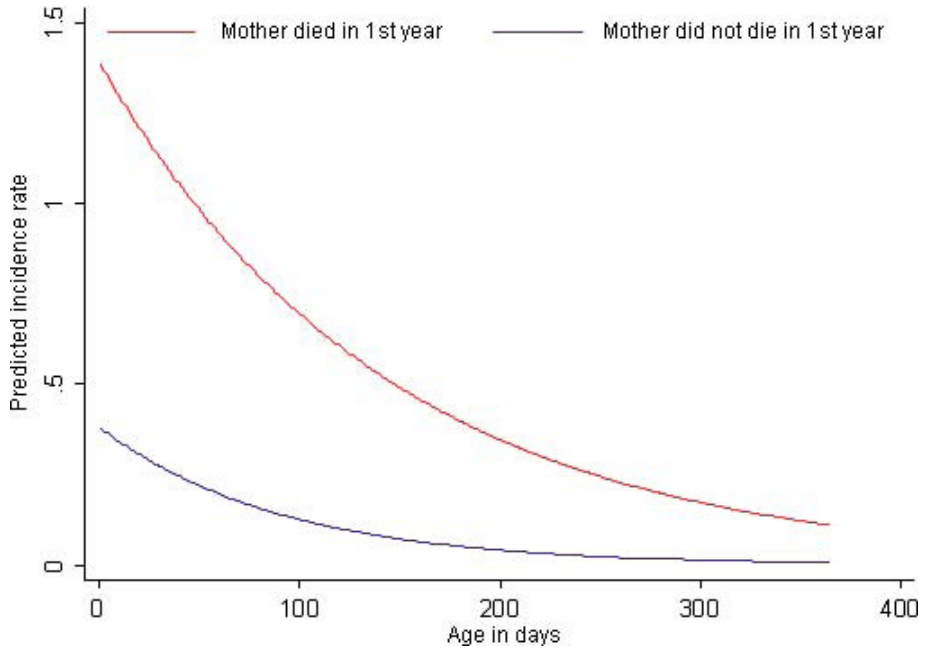


Figure 4 Predicted infant mortality incidence rate by day of life and mother status in first year, Agincourt sub-district.

opportunity costs associated with lengthy waiting time [50] are some of the barriers described in this setting [51,52]. Twine el al showed that the poorest households were less likely to apply for social support grants than those in higher socioeconomic strata due to barriers such as distance from government offices, lack of official documentation and education of caregiver and household head [51].

A recent study in Kenya found that, despite significant spatial variations in child mortality, these were not correlated with distance to health facilities [53]. They concluded that geographic access to curative services did not influence population-level mortality given the density of health facilities in Kenya. They also suggest that when distance access targets are met, further improvements in child survival can only be achieved through renewed investigation of the social, behavioural and quality-of-care factors that may obstruct access to health care services. Similarly in rural South Africa, there is urgent need to evaluate and assure a high level

Table 2 All-cause multivariate risk factor analysis for infant mortality in Agincourt sub-district, 1992-2007, using WinBUGS

	Spatial model		Temporal model		Spatial-temporal model	
Covariate	**IRR [95%CI]**	**signif $^{\alpha}$**	**IRR [95%CI]**	**signif $^{\alpha}$**	**IRR [95%CI]**	**Signif $^{\alpha}$**
Later year of birth	1.09 [1.06,1.13]	*	1.20 [1.03,1.42]	*	1.25 [1.07,1.53]	*
Cumulative household deaths	12.45 [9.41,16.31]	*	12.52 [9.47,16.26]	*	12.59 [9.53,16.49]	*
Male gender	1.09 [0.84,1.4]		1.09 [0.83,1.4]		1.09 [0.84,1.4]	
Mother died in infant's first year	49.82 [7.85,204.7]	*	49.62 [8,189.9]	*	51.11 [8.49,200.8]	*
Pregnancy parity	0.91 [0.80,1.05]		0.94 [0.82,1.08]		0.94 [0.82,1.06]	
Previous child died	2.33 [0.99,4.86]	#	2.07 [0.93,4.07]	#	2.02 [0.89,3.96]	#
Previous birth stillborn	8.13 [2.08,22.99]	*	6.24 [1.55,16.99]	*	6.29 [1.56,17.85]	*
Constant (b0)	-3.24 [-4.60,-1.82]		-0.10 [-2.09,1.65]		-0.63 [-3.08,0.97]	
σ^2 (spatial)	9.15 [4.29,18.33]		—		0.09 [0.03,0.22]	
σ^2 (temporal)	—		0.42 [0.14,1.05]		0.44 [0.14,1.11]	
DIC $^{\beta}$	8680.07		8617.10		8617.34	

α: *significant at the 5% level; #significant at the 10% level.

β: Deviance Information Criterion.

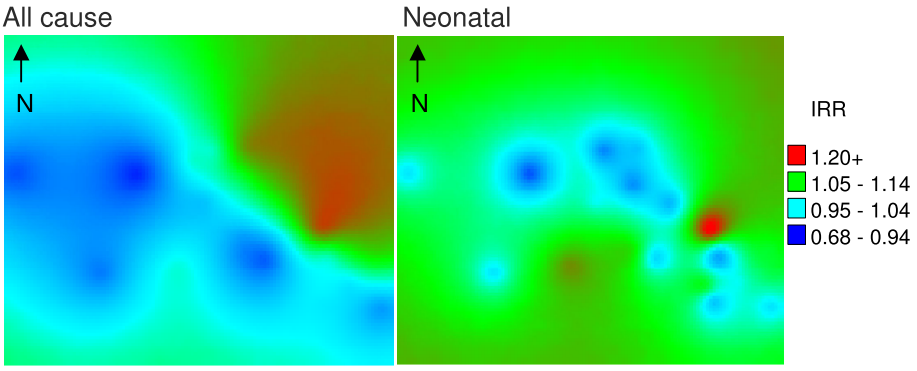


Figure 5 Maps of all-cause and neonatal mortality risk within the Agincourt sub-district 1992-2007, based on baseline models without covariates.

of health service quality, assess and strengthen referral patterns for emergency obstetric, infant and child health care, and identify other barriers to accessing these and other government services.

Mothers' physical presence or absence had a significant impact on infant mortality: mother a temporary migrant (largely work-related) proved significantly protective, while conversely increasing number of months per year spent resident by the mother in the rural site was a risk. Brockerhoff [54] describes how maternal rural-urban migration may affect children through three types of living arrangement: children may remain in the village as foster-children in the care of their fathers or other relatives; children may accompany or follow their mothers to towns or cities; and children born after migrant mothers settle in an urban area may remain there through the first few years of life. (Note that in this study, infants born to mothers in urban areas would not be captured onto the HDSS database unless they later migrated into the rural household). Bledsoe et al, [55] reviewing evidence from West Africa, suggest that while fostered children may be disadvantaged compared to biological children (in terms of access to health care and nutrition), they may still be better off than if they had accompanied their migrant mothers. By staying home, these children avoid exposure to infectious diseases during a vulnerable period of their life, have continued access to economic resources of a non-migrant father, and benefit from remittances received from the migrant mother [55] - as well as better health care, nutrition and enhanced maternal health knowledge [56]. In our study, migrant mothers had significantly higher education and came from households with significantly higher SES which may explain the protective effect of mothers' migration. According to Collinson et al, [57] since 1997 there has been an increasing trend in the number of temporary female labour migrants in the Agincourt sub-district, a poor area with limited employment opportunities with resulting pressures to migrate and remit wages back to the rural household.

The spatial distribution showed marked geographical differences in all-cause mortality risk, indicating variation even within a relatively small area. The highest infant mortality risk was in those villages on the eastern border of the site. Much of this spatial distribution can be explained by the migration patterns of former Mozambican refugees (who constitute about a third of the Agincourt HDSS population) who entered South Africa via the Kruger National Park, a wild game conservation area situated between the eastern border of the site and

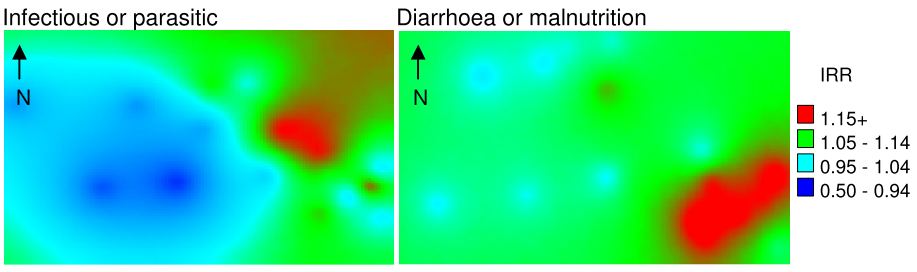


Figure 6 Maps of selected cause-specific infant mortality risk within the Agincourt sub-district, 1992-2006, based on baseline models without covariates.

Southern Mozambique. They remained a vulnerable group, poorer in more isolated villages with less infrastructure and generally further away from health facilities, with poor access to water and sanitation as well as labour markets and legal rights [58]. However, our study indicates that the all-cause infant mortality risk pattern is not being driven by former refugee status alone, a finding supported by Hargreaves et al [23] who found no difference in mortality rates between South African and former Mozambican refugee infants between 1992 and 2000, despite significant differences in the 1-4 year age group. Multiple factors are driving the observed all-cause spatial risk pattern, including Mozambican origin of mother for certain infectious causes, maternal death in first year of infant's life, lower maternal education, poor quality of and limited access to neonatal care, poor antenatal clinic attendance, and increased vulnerability of households with a high mortality burden. These factors should be better elucidated and quantified in order to contribute meaningfully to policy and programmes.

With regard to the geographical distribution of infectious infant deaths (particularly HIV/TB) there was a distinct spatial pattern of mortality with an increasing gradient towards the east of the site where communities appear to have increased risk and suitable interventions need to be directed accordingly. One village in particular had a significantly higher mortality rate (all-cause and HIV) when compared to all other villages. Diarrhoea and malnutrition-related mortality was clustered in the south east of the site suggesting greater problems with clean water and sanitation, services that need to be assessed and addressed by local government. Breastfeeding had a protective effect on all-cause as well as diarrhoea and malnutrition-related infant mortality (Table 1). Breastfeeding protects infants through decreased exposure to contaminated water and food, optimal nutrition, and improved resistance to infection however there is risk of HIV transmission through breast milk. In South Africa, Ministry of Health policy on breastfeeding by HIV positive mothers has evolved in response to emerging research [59]; current recommendations are to breastfeed exclusively during the first 6 months with administration of anti-retrovirals to HIV positive mothers [60], especially those with low CD4 counts. Mothers or infants receiving highly active anti-retroviral therapy (HAART) prophylaxis should continue prophylaxis for one week after breastfeeding has ended [60]. Infant mortality due to diarrhoea, malnutrition and their interaction is a complex problem in poor, HIV prevalent African settings. Addressing this requires a multifaceted approach including provision of clean water and sanitation, promoting infant nutrition, and strengthened primary care services for mothers and infants to reduce the risk of HIV transmission through breast milk [61].

Addressing health inequities in populations is a major challenge [62], and research that documents and quantifies inequities is needed to inform policies to close health gaps in the developing world. Evidence on reducing inequities within countries is growing; successful approaches include those that improve geographic access to health interventions in poor communities, subsidize health care and health inputs for the poor, and empower poorer communities [63]. The results of our study indicate the need for interventions in villages to the east of the site, many of which have a large proportion of former refugees, to reduce the higher burden of infant deaths due to infectious and parasitic causes. HAART for HIV began in 2007 in this district and its impact cannot thus be captured during the time frame of this study. This research does, however, provide useful insight into spatial-temporal mortality patterns before HAART rollout and will allow post-rollout assessment of its impact on infant mortality. Such evaluation has the potential to identify areas needing improved access to treatment, specifically prevention of mother-to-child transmission and anti-retroviral therapy.

Of concern is the high number of neonatal deaths (particularly in the perinatal period), their gradual increase over the study period, and the highest risk area being in close proximity to a health facility. This suggests problems of service quality rather than geographic access, and highlights the need to assess and improve the capacity of sub-district health facilities for antenatal, emergency obstetric and newborn care; improve coverage of deliveries by skilled birth attendants; and advise mothers on appropriate care-seeking for sick babies. Part of the perinatal mortality burden observed may relate to maternal HIV since the same village experienced highest risk for neonatal and infant mortality. A meta-analysis by Brocklehurst et al [64] in 1998 found an association between maternal HIV infection and adverse perinatal outcomes, including low birth weight and pre-term delivery.

A limitation of the study is the potential to miss infant deaths, particularly neonatal deaths, which would underestimate the overall infant mortality burden. Infants that are born and then die during the 12 months between HDSS census update rounds may not be reported, particularly if the mother migrated out of the household; similarly, death among in-migrant infants who die before they are enumerated in the annual household census may be missed. However, infant death ascertainment has improved in the study site [36], and the proportion of infants who were in-migrants decreased significantly over time, reducing the bias towards the end of the study period. Determination of cause of death through verbal autopsy is more problematic for diseases that have less specific symptoms such as HIV/

AIDS [65]. The prevalence of HIV infection in a population and the resulting rate of HIV-associated co-morbidity and death due to malnutrition in children, for example, may affect the performance (such as specificity) of the tool. Thus it is likely that the HIV burden is underestimated due to the misclassification of deaths as AIDS-related conditions such as malnutrition or diarrhoea, or there being placed in the "unknown cause" category. The significant increase in number of infant deaths attributed to unknown causes since the late 1990s (Figure 2) is concurrent with the rise in HIV-related mortality in the area. Levels of stigma associated with HIV are high in South Africa, particularly prior to the introduction of HAART. The ability to make a diagnosis on VA depends, in large part, on the quality of information provided by the respondent. This may have been compromised in some cases in an effort to disguise HIV as a likely cause of death, partly explaining the increase in unknown causes.

Conclusion

By estimating the true spatial distribution of the infant mortality burden in rural northeast South Africa, this study has shown variation across a relatively small geographical area. The approach used Bayesian geostatistical models in order to assess risk factors, correctly estimate the standard errors (significance) of these risk factors and produce smooth maps of infant mortality risk from spatially correlated longitudinal mortality data in a health and socio-demographic surveillance system. Findings indicate the need for interventions targeted at villages with excess infant mortality risk due to both a direct and indirect impact of HIV. Essential interventions include improved prevention of mother-to-child transmission programmes, and antiretroviral therapy for HIV positive mothers to ensure their survival during their infants' critical first year(s) of life. From our study, it is clearly inadequate to consider maternal health separately from infant and neonatal health. This is consistent with other studies which showed that maternal health directly affects infants' health [66]. Policy should thus have greater emphasis on interventions targeting the mother-infant pair. We also conclude that the non-random clustering of infant mortality due to diarrhoea and malnutrition in the south-east part of the site represents a breakdown in basic services (or, indeed, their absence); there is hence need to assess and improve water and sanitation in these villages. The high levels of perinatal mortality, in some instances in close proximity to health facilities, is of concern, indicating need to strengthen the capacity of sub-district facilities for emergency obstetric and newborn care. Recommendations from this study will have applications to other similar rural settings within South Africa and potentially beyond.

Appendix: Statistical Model

Let Y_{it} and p_{it} be the status and probability of mortality of an infant i in year of birth t. We assume that Y_{it} arises from a negative binomial distribution, that is Y_{it} ~NegBin[p_{it}, r], where p_{it} *is* the probability that child *i* at location *si* is dead and *r* is the parameter that quantifies the amount of extra Poisson variation. We modelled the probability of death [p_{it}] as follows:

1. logit (p_{it}) = $\beta_0 + \beta X_{it} + \phi_{it}$ (multivariate spatial model)
2. logit (p_{it}) = $\beta_0 + \beta X_{it} + \alpha_t$ (multivariate temporal model)
3. logit (p_{it}) = $\beta_0 + \beta X_{it} + \phi_{it} + \alpha_t$ (multivariate spatial-temporal model)
4. logit (p_{it}) = $\beta_0 + \phi_{it}$ (spatial kriging model) i.e. constant and spatial random effect with no covariates

where β_0 is the incidence rate where all covariates are zero (i.e. the constant), X_{it} denotes the covariates, β is the vector of regression coefficients, ϕ_{it} the village-specific random effect, μ_i the individual level random effect and α_t the temporal random effect. Following a Bayesian model specification, noninformative normal prior distributions were adopted for the regression coefficients β and an informative (based on estimates from Stata) and non-informative gamma prior distribution for the over-dispersion parameter *r* were adopted and tested [lower DIC dictating which was used]. We assume that ϕ_{it} has a multivariate normal distribution, $\phi_{it} \sim$ MVN ($\underline{0}$,Σ), with variance-covariance matrix Σ. We also assume an isotropic stationary spatial process, where $\Sigma kl = \sigma_w^2 \exp(-\phi d_{kl})$, d_{kl} is the Euclidean distance between villages *k* and *l*, σ_w^2 is the geographical variability known as the sill, ϕ is a smoothing parameter that controls the rate of correlation decay with increasing distance and measures the range of geographical dependency. A noninformative gamma prior was adopted for phi [ϕ], which is the smoothing parameter that controls the rate of correlation decay, as well as uniform prior with a distribution between ϕ min and ϕ max [67]. Both approaches were tested and the approach providing the best fit was then used. The range is defined as the minimum distance at which spatial correlation between locations is below 5%. This distance can be calculated as $3/u$ meters. The second order year level autoregressive temporal random effect (α_t), for t = 1 to 16 years, was modelled as a normal distribution with mean α_{mean} [t = 3,..,16] = $\rho 0 + \rho[1]*\alpha[t-1] + \rho[2]*\alpha[t-2]$ and a noninformative gamma distribution for the variance parameter. The first two autoregressive terms were specified as α_{mean} [1] <- $\rho 0 + l[1]$ and α_{mean} [2] <- $\rho 0 + \rho[1]*\alpha[1] + l[2]$. Noninformative normal prior distributions were adopted for the ρ and l coefficients [34].

MCMC simulation was applied to fit the models. We ran a single chain sampler with a burn-in of 5000 iterations. Convergence was assessed by running the simulation until the Monte Carlo error for each parameter of interest was less than 5% of the sample standard deviation. The chains thereafter were sampled every single iteration until a sample size of 10,000 had been attained.

Acknowledgements

This work was supported by a PhD fellowship from the South African Centre for Epidemiological Modeling and Analysis (SACEMA), a National Research Foundation (NRF) Centre of Excellence. Additional funding was provided by a travel grant from the NRF Knowledge Interchange and Collaboration (KIC), Wits Faculty of Health Sciences Medical Research Endowment Fund (MREF) (Grant number: SARB000) and a research fellowship from the MRC/Wits Rural Public Health and Health Transitions Research Unit (Agincourt) through The Wellcome Trust, UK. Benjamin Clark Provided support in preparing the data. The Agincourt health and socio-demographic surveillance system and geographic information system was funded by the Wellcome Trust, UK (Grant numbers: 058893/Z/99/A, 069683/Z/02/Z, 069683/Z/08/Z), the University of the Witwatersrand, the South African Medical Research Council, and the Andrew W. Mellon Foundation, USA.

Author details

[1]MRC/Wits Rural Public Health and Health Transitions Research Unit (Agincourt), School of Public Health, Faculty of Health Sciences, University of the Witwatersrand, Johannesburg, South Africa. [2]Centre for Global Health Research, Epidemiology and Global Health, University of Umeå, Umeå, Sweden. [3]INDEPTH Network, Accra, Ghana. [4]Swiss Tropical and Public Health Institute, Basel, Switzerland.

Authors' contributions

All authors have read and approved the final manuscript. BKD Sartorius: conception and design, data extraction, data analysis, drafted the manuscript. KK: acquisition of funding, conception and design, policy implications, reviewed manuscript. PV: acquisition of funding, conception and design, statistical support, reviewed manuscript. MAC: conception and design, reviewed manuscript. SMT: acquisition of funding, conception and design, reviewed manuscript

Competing interests

The contents of this paper and the data used for it have not been published elsewhere. The paper is also not in press or under review elsewhere, nor has a similar paper been written by anyone else using the same data and methods.

Received: 15 December 2009 Accepted: 26 October 2010
Published: 26 October 2010

References

1. Hammer G, Somé F, Müller O, Kynast-Wolf G, Kouyaté B, Becher H: **Pattern of cause-specific childhood mortality in a malaria endemic area of Burkina Faso.** *Malaria Journal* 2006, **5**:47.
2. Adjuik M, Smith T, Clark S, Todd J, Garrib A, Kinfu Y, Kahn K, Mola M, Ashraf A, Masanja H, Kubaje A, Sacarlal J, Alam N, Marra A, Gbangou A, Mwageni E, Binka F: **Cause specific mortality rates in developing countries.** Bulletin of the World Health Organization; 2006:**84**:181-188.
3. Black R, Morris S, Bryce J: **Where and why are 10 million children dying every year?** *Lancet* 2003, **361**:2226-2234.
4. Ezzati M, Lopez A, Rodgers A, Murray C: **Comparative quantification of health risks: global and regional burden of disease attributable to selected major risk factors.** Geneva: World Health Organization; 2004.
5. Walker N, Schwartländer B, Bryce J: **Meeting international goals in child survival and HIV/AIDS.** *Lancet* 2002, **360**:284-89.
6. UNICEF: **Progress since the world summit for children: a statistical review.** New York: UNICEF; 2001 [http://www.unicef.org/pubsgen/wethechildren-stats/sgreport_adapted_stats_eng.pdf].
7. UNICEF, UNAIDS, WHO: **Children and AIDS; Second Stocktaking Report.** NY: UNICEF; 2008, 11.
8. Dabis F, Leroy V, Castetbon K, Spira R, Newell M, Salamon R: **Preventing mother-to-child transmission of HIV-1 in Africa in the year 2000.** *AIDS* 2000, **14**:1017-1026.
9. Newell M: **Prevention of mother-to-child transmission of HIV: challenges for the current decade.** *Bulletin of World Health Organization* 2001, **79**:1138-1144.
10. Kahn K, Garenne M, Collinson M, Tollman S: **Mortality trends in a new South Africa: Hard to make a fresh start.** *Scand J Public Health Suppl* 2007, **69**:26-34.
11. Garrib A, Jaffar S, Knight S, Bradshaw D, Bennish M: **Rates and causes of child mortality in an area of high HIV prevalence in rural South Africa.** *Trop Med Int Health* 2006, **11(12)**:1841-8.
12. Stringer E, Sinkala M, Ng'uni C, McFarlane YA, Chipimo M, Myzece E, *et al*: **In an urban African setting, the cost-effectiveness of PMTCT services improves as services expand.** *International Conference on AIDS [15th: 2004: Bangkok, Thailand]. Int Conf AIDS* 2004, 11-16, 15: abstract no. TuPeC4933.
13. Adomakoh S, Kumar A, St John A: **Reducing mother to child transmission of HIV-1 in Barbados. Cost-effectiveness of the PACTG 076 protocol in a middle income, low prevalence setting.** *International Conference on AIDS. Int Conf AIDS* 2002, **14, abstract no. LbPp2210.**
14. Sansom SL, Anderson JE, Farnham PG, Dominguez K, Soorapanth S, Clark J, Sukalac T, Earp MJ, Bohannon B, Fowler MG, PSD Consortium: **Updated estimates of healthcare utilization and costs among perinatally HIV-infected children.** *J Acquir Immune Defic Syndr* 2006, **41(4)**:521-6.
15. Giraudon I, Leroy V, Msellati P, Elenga N, Ramon R, Welffens-Ekra C, Dabis F: **The costs of treating HIV-infected children in Abidjan, Ivory Coast, 1996-1997.** *Article in French Sante* 1999, **9(5)**:277-81.
16. Stover J, Walker N, Garnett G, Salomon JA, Stanecki K, Ghys P, *et al*: **Can we reverse the HIV/AIDS pandemic with an expanded response?** *Lancet* 2002, **360**:73-77.
17. Schwartländer B, Stover J, Walker N, Bollinger L, Gutierrez JP, McGreevey W, *et al*: **Resource needs for HIV/AIDS.** *Science* 2001, **292**:2434-36.
18. European Collaborative Study: **Mother-to-child transmission of HIV infection in the era of highly active antiretroviral therapy.** *CID* 2005, **40**:458-465.
19. Navér L, Lindgren S, Belfrage E, Gyllensten K, Lidman K, Gisslén M, *et al*: **Children born to HIV-1-infected women in Sweden in 1982-2003: trends in epidemiology and vertical transmission.** *Journal of Acquired Immune Deficiency Syndrome* 2006, **42**:484-489.
20. Newell M, Huang S, Fiore S, Thorne C, Mandelbrot L, Sullivan J, *et al*: **Characteristics and management of HIV-1-infected pregnant women enrolled in a randomised trial: differences between Europe and the USA.** *BMC Infectious Diseases* 2007, **7**:60.
21. Victora C, Wagstaff A, Schellenberg J, Gwatkin D, Claeson M, Habicht JP: **Applying an equity lens to child health and mortality: more of the same is not enough.** *Lancet* 2003, **362**:233-41.
22. Johnston N: **The Regularisation of Former Mozambican Refugees in South Africa: Experiences and Lessons.** Acornhoek, Refugee Research Programme; 2000.
23. Hargreaves J, Collinson M, Kahn K, Clark S, Tollman SM: **Childhood mortality among former Mozambican refugees and their hosts in rural South Africa.** *Int J Epidemiol* 2004, **33(6)**:1271-8.
24. WHO: **Reaching the Poor: Challenges for Child Health in the Western Pacific Region. Chapter 7: Barriers to access to child health care.** 2007.
25. Yates R: **Women and children first: an appropriate first step towards universal coverage.** *Bull World Health Organ* 2010, **88(6)**:474-475.
26. Benzler J, Sauerborn R: **Rapid risk household screening by neonatal arm circumference. Results from a cohort study in Burkina Faso.** *Tropical Medicine and International Health* 1998, **3**:62-74.
27. Mulholland E, Smith L, Carneiro I, Becher H, Lehmannd D: **Equity and child-survival strategies.** *Bulletin of the World Health Organization* 2008, **86(5)**.
28. Antunes J, Waldman E: **Trends and spatial distribution of deaths of children aged 12-60 months in Sa˜ o Paulo, Brazil, 1980-98.** Bulletin of the World Health Organization; 2002:**80**:391-398.
29. Elliot P, Wakefield J, Best N, Briggs D: **Spatial Epidemiology-Methods and Applications.** Oxford University Press London; 2000.
30. Gemperli A, Vounatsou P, Kleinschmidt I, Bagayoko M, Lengeler C, Smith T: **Spatial patterns of infant mortality in Mali; the effect of malaria endemicity.** *Am J Epidemiol* 2004, **159**:64-72.

31. Kazembe L, Appleton C, Kleinschmidt I: **Spatial analysis of the relationship between early childhood mortality and malaria endemicity in Malawi.** *Geospatial Health* 2007, **2(1)**:41-50.
32. Gosoniu L, Vounatsou P, Tami A, Nathan R, Grundmann H, Lengeler C: **Spatial effects of mosquito bednets on child mortality.** *BMC Public Health* 2008, **8**:356.
33. Kahn K, Tollman SM, Collinson MA, Clark S, Twine R, Clarke B, *et al*: **Research into health, population and social transitions in rural South Africa: Data and methods of the Agincourt Health and Demographic Surveillance System.** *Scandinavian Journal of Public Health* 2007, **35(Suppl 69)**:8-20.
34. Kahn K, Tollman SM, Garenne M, Gear JS: **Who dies from what? Determining cause of death in South Africa's rural north-east.** *Trop Med Int Health* 1999, **4(6)**:433-441.
35. Kahn K, Tollman SM, Garenne M, Gear JS: **Validation and application of verbal autopsies in a rural area of South Africa.** *Trop Med Int Health* 2000, **5(11)**:824-831.
36. Kahn K: **Dying to make a fresh start: mortality and health transition in a new South Africa [dissertation].** Umeå University, Faculty of Medicine, Public Health and Clinical Medicine; 2006.
37. Diggle P, Tawn J, Moyeed R: **Model-based geostatistics.** *Appl Stat* 1998, **47**:299-350.
38. Schotman P: **Priors for the AR [1] model: parameterization issues and time series considerations.** *Econometric Theory* 1994, **10**:579-595.
39. Zeller A: **An introduction to Bayesian inference and econometrics.** John Wiley and Sons Ltd/Inc: New York; 1996.
40. Benjamin M, Rigby R, Stasinopoulos D: **Generalized autoregressive moving average models.** *J Am Stat Association* 2003, **98(461)**:214-223.
41. Gelfand A, Smith A: **Sampling-based approaches to calculating marginal densities.** *J Am Stat Association* 1990, **85**:398-409.
42. Spiegelhalter D, Best N, Carlin B, van der Linde A: **Bayesian measures of model complexity and fit.** *J R Stat Soc* 2002, **64(B)**:583-639.
43. Gelfand A, Ravishanker N, Ecker M: **Modeling and inference for point-referenced binary spatial data.** In *Generalized Linear Models: A Bayesian Perspective.* Edited by: Dey D, Ghosh S, Mallick B. Marcel Dekker Inc; 1999:373-386.
44. Tollman S, Kahn K, Sartorius B, Collinson M, Clark S, Garenne M: **Implications of mortality transition for primary health care in rural South Africa: a population-based surveillance study.** *Lancet* 2008, **372**:893-901.
45. Farah A, Preston S: **Child mortality differentials in Sudan.** *Population and Development Review* 1982, **8**:365-383.
46. Jain A: **Determinants of regional variation in infant mortality in rural India.** In *Infant Mortality in India: Differentials and Determinants.* Edited by: Jain A, Visaria L. Sage Publications; 1988:127-167.
47. Cleland J, van Ginneken J: **Maternal education and child survival in developing countries: the search for pathways of influence.** *Soc Sci Med* 1989, **27**:1357-1368.
48. Schultz T: **Interpretation of relations among mortality, economics of the household and the health environment.** *Proceedings of the Meeting on Socio-Economic Determinants and Consequences of Mortality Mexico City* Geneva: World Health Organization; 1979, June 19-25.
49. Manda S: **Birth intervals, breastfeeding and determinants of childhood mortality in Malawi.** *Soc Sci Med* 1999, **48(3)**:301-312.
50. Bigdeli M, Annear P: **Barriers to access and the purchasing function of health equity funds: lessons from Cambodia.** *Bull World Health Organ* 2009, **87**:560-564.
51. Twine R, Collinson MA, Polzer TJ, Kahn K: **Evaluating access to a child-oriented poverty alleviation intervention in rural South Africa.** *Scand J Public Health* 2007, **69**:118-27.
52. Goudge J, Gilson L, Russell S, Gumede T, Mills A: **The household costs of health care in rural South Africa with free public primary care and hospital exemptions for the poor.** *Trop Med Int Health* 2009, **14(4)**:458-67.
53. Moïsi J, Gatakaa H, Noor A, Williams T, Bauni E, Tsofa B, Levine O, Scott JA: **Geographic access to care is not a determinant of child mortality in a rural Kenyan setting with high health facility density.** *BMC Public Health* 2010, **10**:142.
54. Brockerhoff M: **The impact of rural-urban migration on child survival.** *Health Transition Review* 1994, **4**:127-149.
55. Bledsoe C, Anastasia B: **Child fosterage and child mortality in sub-Saharan Africa: some preliminary questions and answers.** In *Mortality and Society in Sub-Saharan Africa.* Edited by: Etienne van de Walle, Gilles Pison and Mpembele Sala-Diakanda. Oxford: Clarendon Press; 1992:279-302.
56. Hildebrandt N, McKenzie D: **The effects of migration on child health in Mexico.** Mimeo. Stanford University; 2004.
57. Collinson M, Tollman S, Garenne M, Kahn K: **Temporary female migration and labour force participation in rural South Africa.** A working paper of the Agincourt Health and Population Unit, University of the Witwatersrand. A working paper of Princeton University's, Center for Migration and Development, Working paper #01-10a; 2001.
58. Dolan C, Tollman S, Nkuna V, Gear J: **The links between legal status and environmental health: A case study of Mozambican refugees and their hosts in the Mpumalanga (eastern Transvaal) Lowveld, South Africa.** *International Journal of Health and Human Rights* 1997, **2(2)**:62-84.
59. Coovadia H, Rollins N, Bland M, Little K, Coutsoudis A, Bennish M, Newell ML: **Mother-to-child transmission of HIV-1 infection during exclusive breastfeeding in the first 6 months of life: an intervention cohort study.** *The Lancet* 2007, **369**:1107-1116.
60. WHO: **Guidelines on HIV and infant feeding. Principles and recommendations for infant feeding in the context of HIV and a summary of evidence.** WHO; 2010 [http://www.who.int/child_adolescent_health/documents/9789241599535/en/index.html], accessed 4 Sep 2010.
61. Patwari AK: **Diarrhoea and malnutrition interaction.** *Indian J Pediatr* 1999, **66(1 Suppl)**:S124-34.
62. Feachem R: **Poverty and inequity: a proper focus for the new century.** *Bull World Health Organ* 2000, **78**:1-2.
63. Gwatkin D, Bhuiya A, Victora C: **Making health systems more equitable.** *Lancet* 2004, **364**:1273-1280.
64. Brocklehurst P, French R: **The association between maternal HIV infection and perinatal outcome: a systematic review of the literature and meta-analysis.** *Br J Obstet Gynaecol* 1998, **105(8)**:836-48.
65. Garenne M, Fauveau V: **Potential and limits of verbal autopsies.** Bulletin of the World Health Organization; 2006:**84**:161-256.
66. Newell ML, Brahmbhatt H, Ghys PD: **Child mortality and HIV infection in Africa: a review.** *AIDS* 2004, **18(2)**:S27-S34.
67. Gelfand A, Vounatsou P: **Proper multivariate conditional autoregressive models for spatial data analysis.** *Biostat* 2003, **4**:11-25.

Pre-publication history

The pre-publication history for this paper can be accessed here:
http://www.biomedcentral.com/1471-2458/10/645/prepub

doi:10.1186/1471-2458-10-645
Cite this article as: Sartorius *et al.*: Young and vulnerable: Spatial-temporal trends and risk factors for infant mortality in rural South Africa (Agincourt), 1992-2007. *BMC Public Health* 2010 **10**:645.

Submit your next manuscript to BioMed Central and take full advantage of:

- **Convenient online submission**
- **Thorough peer review**
- **No space constraints or color figure charges**
- **Immediate publication on acceptance**
- **Inclusion in PubMed, CAS, Scopus and Google Scholar**
- **Research which is freely available for redistribution**

Submit your manuscript at
www.biomedcentral.com/submit

BioMed Central

Geospatial Health 5(2), 2011, pp. 285-295

Survived infancy but still vulnerable: spatial-temporal trends and risk factors for child mortality in the Agincourt rural sub-district, South Africa, 1992-2007

Benn Sartorius[1], Kathleen Kahn[1,2,3], Mark A. Collinson[1,2,3], Penelope Vounatsou[4,5], Stephen M. Tollman[1,2,3]

[1]MRC/Wits Rural Public Health and Health Transitions Research Unit (Agincourt), School of Public Health, Faculty of Health Sciences, University of the Witwatersrand, Johannesburg, South Africa; [2]Centre for Global Health Research, Epidemiology and Global Health, University of Umeå, Umeå, Sweden; [3]INDEPTH Network, Accra, Ghana; [4]Department of Epidemiology and Public Health, Swiss Tropical and Public Health Institute, Basel, Switzerland; [5]University of Basel, Basel, Switzerland

Abstract. Targeting of health interventions to poor children at highest risk of mortality are promising approaches for enhancing equity. Methods have emerged to accurately quantify excess risk and identify space-time disparities. This provides useful and detailed information for guiding policy. A spatio-temporal analysis was performed to identify risk factors associated with child (1-4 years) mortality in the Agincourt sub-district, South Africa, to assess temporal changes in child mortality patterns within the study site between 1992 and 2007, and to produce all-cause and cause-specific mortality maps to identify high risk areas. Demographic, maternal, paternal and fertility-related factors, household mortality experience, distance to health care facility and socio-economic status were among the examined risk factors. The analysis was carried out by fitting a Bayesian discrete time Bernoulli survival geostatistical model using Markov chain Monte Carlo simulation. Bayesian kriging was used to produce mortality risk maps. Significant temporal increase in child mortality was observed due to the HIV epidemic. A distinct spatial risk pattern was observed with higher risk areas being concentrated in poorer settlements on the eastern part of the study area, largely inhabited by former Mozambican refugees. The major risk factors for childhood mortality, following multivariate adjustment, were mother's death (especially when due to HIV and tuberculosis), greater number of children under 5 years living in the same household and winter season. This study demonstrates the use of Bayesian geostatistical models for accurately quantifying risk factors and producing maps of child mortality risk in a health and demographic surveillance system. According to the space-time analysis, the southeast and upper central regions of the site appear to have the highest mortality risk. The results inform policies to address health inequalities in the Agincourt sub-district and to improve access to health services. Targeted efforts to prevent vertical transmission of HIV in specific settings need to be undertaken as well as ensuring the survival of the mother and father in childhood.

Keywords: Bayesian inference, autoregressive, geostatistical data, child mortality, kriging, survival, Bernoulli or logistic, spatio-temporal model, health and demographic surveillance, mortality, South Africa

Introduction

Large reductions in child mortality occurred in low-income and middle-income countries towards the end of the last century, however more than 10 million children still die every year. Most childhood deaths occur in the developing world, particularly sub-Saharan Africa (SSA), although there is considerable heterogeneity within the region with countries showing differential trends in levels and age patterns of childhood mortality. Childhood mortality rates have considerably declined over the past decades in much of SSA, but since the 1990s mortality rates have started to increase again in parts of the continent (Ahmad et al., 2000; WHO, 2002). This new trend has mainly been attributed to the effects of the HIV/AIDS epidemic and to the spread of chloroquine-resistant malaria (Müller et al., 1999; Adetunji, 2000; Trapé, 2001). Other prominent causes of death among children include diarrhoea, pneumonia, measles and the underlying cause of malnutrition for deaths among children younger than 5 years. Most of these conditions are either preventable or treatable with low-cost interventions (Tulloch, 1999; Black et al., 2003).

Gaps between mortality of wealthier and poorer children within most countries are unacceptably wide

Corresponding author:
Benn Sartorius
MRC/Wits Rural Public Health and Health Transitions Research Unit (Agincourt), School of Public Health,
Faculty of Health Sciences, University of the Witwatersrand, Johannesburg, South Africa
Tel: +27 11 717-2627; Fax: +27 11 717-2084
E-mail: benn.sartorius@wits.ac.za

and in some areas this gap is increasing (Wagstaff, 2000). Targeting of health interventions to poorer or higher risk individuals and/or communities and ensuring universal coverage are promising approaches for promoting equity. Successful approaches include subsidised health care and health inputs, improved geographic access to health interventions in poor communities and social marketing (Victora et al., 2003). Geostatistical models allow accurately quantifying inequities in health, identifying risk factors for mortality within a population and predicting mortality at unsampled locations, and ultimately generating smoothed maps of mortality risk. Despite the increasing use of Bayesian geostatistical models in risk mapping and prediction of parasitic diseases such as malaria (Gemperli et al., 2006; Gosoniu et al., 2008; Hay et al., 2009; Riedel et al., 2010), schistosomiasis and soil-transmitted helminthiasis (Raso et al., 2005; Clements et al., 2008; Schur et al., 2011), little or no work has been done on mortality data from health and demographic surveillance sites (HDSS). Geostatistical model-based predictions of mortality at non-sampled locations can generate smoothed maps that can be used for identifying clusters of high mortality and assessing effectiveness of interventions.

Child mortality in developing countries is mainly associated with measurable socio-economic conditions such as poor living conditions (Manda, 1999). Poor children are more likely to be exposed to health risks, and have less resistance to disease because of malnutrition and other risk factors typical of poorer communities. These inequities are further compounded by reduced access to health care in the form of preventive and curative interventions. Other proximate determinants of child mortality include maternal factors (e.g. age, pace of childbearing and death of mother), nutritional factors, environmental contamination, injury and health care access and quality (Mosley and Chen 1984a, b; Hobcraft et al., 1985; Binka et al., 1995; Ronsmans, 1996; Kuate Defo, 1997). Issues of access which include accessibility (distance to facility), affordability and acceptability of available health services also heavily contribute to child mortality. Geography and ethnicity can both lead to failure to access health care, and therefore inequity in child survival. A previous study in the Agincourt HDSS, a rural South African population, showed a higher level of childhood mortality (particularly in those children aged 1-5 years) among former Mozambique refugees when compared to South African households and confirmed various other established risk factors (Hargreaves et al., 2004). Indeed, Hargreaves and colleagues (2004) suggested that the lack of legal status, "refugee" villages being more isolated with less infrastructure, and lower wealth of many former Mozambican refugees (three times more likely than host South African households to be in the poorest quintile of the sample) may partly explain this disparity (Kahn, 2008). This should be kept in mind when assessing spatio-temporal trends in child mortality in this area. Another factor is the impact of dysfunctional health services, while communities have geographic access to adequate health services, they may fail to derive any benefit from them (Penchansky et al., 1981; Mulholland et al., 2008). There is still a need for more research in the Agincourt sub-district and other poor rural settings of sub-Saharan Africa to clarify the role of factors such as distance to nearest health facility on child morbidity and mortality as well as to identify any poorly functioning health facilities in addition to the traditional risk factors mentioned above. Given the inherent spatial (households and villages) and temporal (repeated measurements on a child each year) correlation in longitudinal data, Bayesian spatio-temporal models provide the most appropriate methodology for risk factor analysis as they take into account both sources of correlation. Standard approaches on the other hand assume independence of outcomes such as mortality and under/over estimate the magnitude and precision of the of effects of risk factors.

The objectives of this study were to assess changes in child mortality patterns within the rural Agincourt subdistrict over a 15-year period, accurately quantify risk factors and develop all-cause and cause-specific mortality maps to identify high risk areas within the Agincourt area. This will provide guidance on how best to allocate limited resources to reduce child mortality and information on effectiveness of current health policy.

Materials and methods

Study area and population

The Agincourt HDSS was demarcated in 1992 and comprises a mix of temporary migrant workers, former Mozambican refugees and a more stable permanent population (Tollman et al., 1999) (Fig. 1). It covers an area of ~400 km^2 and contains 21 villages with approximately 11,700 households and 70,000 people. A full geographical information system (GIS) exists for all households within the site and is updated annually. The study population consisted of all children (aged 1-4 years) who were either resident, born, or in-migrated into the site between 1992 and 2007.

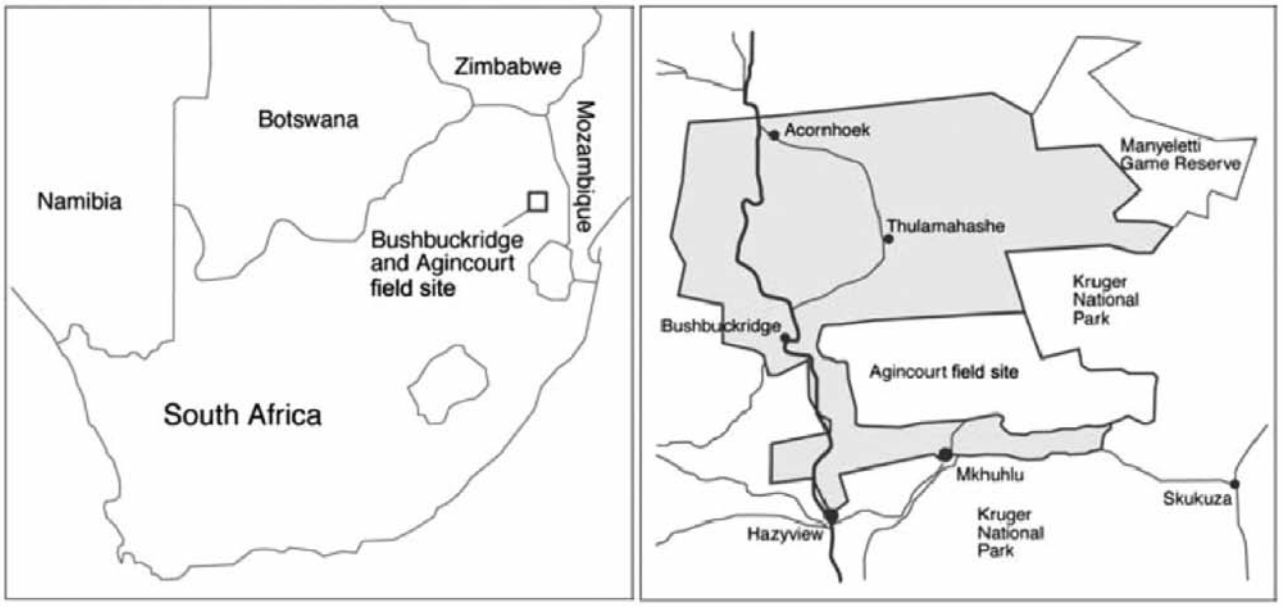


Fig. 1. Location of the Agincourt HDSS site (Kahn et al., 2007), South Africa.

Outcome measures

A verbal autopsy, validated by Kahn et al. (2000) in the mid-1990s and again in 2005, is conducted for each death within the Agincourt HDSS. Lay, trained fieldworkers interviewed the closest caregiver on the signs and symptoms of the terminal illness, as well as lifestyle risk factors and treatment sought. International Classification of Diseases (ICD-10) was used to classify main, underlying and contributing causes of death following independent assessment of the completed questionnaires by two or three physicians. Where consensus was achieved, the diagnosis was accepted as the 'probable cause of death'; otherwise, the death was classified as 'ill-defined'. Cause-specific data (e.g. HIV/AIDS, tuberculosis, acute diarrhoea or malnutrition, acute respiratory infection or pneumonia and accidents) were available from 1992 to 2006.

Explanatory variables

The following factors were included as covariates: demographic (gender, nationality), temporal, maternal (nationality, age at pregnancy, death while child aged 1-4 years and education), fertility (parity, birth intervals and sibling death), household mortality experience, socio-economic status (SES), food security, distance to health facility, household elevation (climate proxy) and health seeking behaviour (antenatal clinic attendance). SES is based on information on living conditions and assets, building materials of main dwelling, water and energy supply, ownership of modern appliances and livestock, and means of transport available (Kahn, 2008). The sum of these scores for certain variables provided each household with an overall absolute score. We also used these factors to construct a weighted score using multiple correspondence analysis (MCA). These scores were then divided into five socio-economic strata ranked by increasing value of the score and corresponding closely to five wealth quintiles: most poor, very poor, poor, less poor and least poor.

Statistical analysis

The time to death was treated as discrete at monthly intervals and Cox proportional hazards models were fitted using dichotomous logistic regression formulations (D'Agostino et al., 1990). Preliminary regression models were applied to assess the relationship between all-cause child mortality with each covariate. Covariates significant at the 10% level (without substantial missing values) were then incorporated into a multivariate geostatistical-temporal model to assess the effects of the most significant covariates on child mortality and to develop a predictive model to enable mapping of the mortality outcome. Spatial correlation was modelled via village-specific random effects, which considered as latent observations a spatial Gaussian process. Correlation between any pairs of village locations were considered as an exponential function of their distance and modelled by the covariance matrix of the process (Diggle et al., 1998). Temporal correlation was introduced by yearly random effects modelled

via an autoregressive process of various order (Schotman, 1994; Zeller, 1996). A Bayesian framework was used to specify the models and Markov chain Monte Carlo (MCMC) simulation was applied to estimate the parameters (Gelfand et al., 1990).

The order of the autoregressive process was assessed using the deviance information criterion (DIC) (Spiegelhalter et al., 2002). Models with the smallest DIC indicate the best fit. Validation was carried out by fitting the models on the subset of data during the period 1992-2006 and calculating the proportion of villages with observed mortality in 2007 within the 95% credible intervals of their corresponding predictive distributions.

Autoregressive models of first and second order had similar DICs and similar predictive ability for 2007. However, the second order process converged faster and was thus used in the final model. Further modeling details are given in the Appendix.

Simulation-based Bayesian kriging (Gelfand et al., 1999) at numerous prediction points within the site was used to produce smoothed maps of all-cause and cause- specific mortality risk within the whole Agincourt HDSS. Predictions were carried out at baseline category of the predictors and during the periods of 1992-1995, 1996-1999, 2000-2003 and 2004-2007. All identifying features (village centroids, boundaries) have been removed from the maps and the prediction area expanded irregularly, in order to ensure confidentiality and avoid stigmatising of villages. The HIV and tuberculosis mortality risk map is also not shown for this reason.

Data extraction and management was done using Microsoft SQL Server 2005. The analysis was carried out in STATA version 10.0 (Stata Corp., 2007), OpenBUGS (Spiegelhalter et al., 1999) and R (R Development Core Team, 2008). The predictions of the fitted spatial models were mapped in Map Info Professional version 9.5 (MapInfo, 2008).

Results

Between 1992 and 2007, there were 46,675 children aged 1-4 years in the Agincourt HDSS. Each child, on average, contributed 23.67 person months of time (standard deviation (SD) 13.31). There were 565 deaths (6.14 per 1,000 person years) with a mean age at death of 2.20 years (SD 0.93 years). The sex ratio was approximately 1:1 (21,733 females or 50.3%) as was the breakdown of children born in the site *versus* those who in-migrated (23,423 or 54.5% for the former). More than one third (15,691 or 36.8%) of the children were born to former Mozambican refugees.

The top causes of death among children in this period (n = 535) were HIV/tuberculosis (n = 136, 25.4%), acute diarrhoea or malnutrition (n = 135, 25.2%), accidents (n = 31, 5.8%) and acute respiratory infection or pneumonia (n = 16, 3.0%). In total 314 (58.7%) deaths were attributed to infectious or parasitic causes. Fifty-seven deaths (10.7%) were classified as unknown (i.e. R95-R99) and 106 deaths (20.3%) had no diagnosis since a verbal autopsy could not be conducted due to lack of a respondent or family refusal. A significant increasing trend in the child mortality rate was observed over the study period ($\beta = 0.05$, $p = 0.001$). This was observed particularly for the 1992-2003 period (Fig. 2).

Bivariate analyses indicate that winter season, having a Mozambican mother, four or more children aged less than 5 years living in the same household, mother or father death during childhood (1-4 years) especially due to HIV/tuberculosis, father death before birth, increasing number of cumulative household deaths and death of previous sibling are significant factors of child mortality (Table 1). Significant protective factors against child mortality were increasing age of child, increasing mothers age, tertiary level education of mother, increased post-birth interval, mother antenatal clinic attendance as well as increased frequency of antenatal clinic attendance and higher SES quintile of household (Table 1). No significant association was observed between distance from household and nearest health facility. For mothers who died while the child was aged 1-4 years (n = 259), the leading cause of death was HIV/tuberculosis (30.5%), while 47.1% where classified as unknown. This was similar for fathers who died before the child was born or up to the child's 5th birthday (n = 543, 28.7% HIV/tuberculosis, 33.5% unknown).



Fig. 2. Child mortality rate per 1,000 person years (with linear trend line) in Agincourt sub-district, 1992-2007.

Table 1. All-cause bivariate and multivariate child mortality risk factor analyses in Agincourt sub-district, 1992-2007.

Factor	Bivariate non-spatial model		Multivariate spatio-temporal model OR (95% BCI)
	n	OR (95% CI)	
Temporal			
By year	1,136,136	1.05 (1.04-1.07)	-
Winter season	1,136,136	1.36 (1.15-1.60)	1.33 (1.12-1.57)
Demographic			
Age (continuous)	1,136,136	0.50 (0.46-0.55)	0.51 (0.46-0.56)
Male (gender)	1,135,970	1.07 (0.91-1.26)	-
Maternal			
Mozambican	13,148/35,845	1.72 (1.39-2.12)	1.15 (0.93-1.40)
Mother migration (increasing number of months spent resident in site in given year of child's life)	665,539	0.93 (0.89-0.96)	-
By age (in years)			
1-1.99	192,999	0.93 (0.88-0.98)	-
2-2.99	190,637	0.90 (0.84-0.96)	-
3-3.99	194,671	1.04 (0.89-1.22)	-
4-4.99	87,232	0.92 (0.80-1.06)	-
Mother death	1,152/1,136,136	5.16 (4.03-6.61)	-
Mother death in 1-4 not due to HIV/tuberculosis	191/1,136,136	7.77 (4.43-13.65)	5.17 (2.84-8.29)
Mother death in 1-4 due to HIV/tuberculosis	66/1,136,136	19.92 (10.40-38.19)	13.52 (7.57-21.68)
Mother's age	1,045,197	0.98 (0.97-0.99)	-
Mother's education	364,244	0.98 (0.96-1.01)	-
None	57,168	1.00	
Primary	96,422	1.01 (0.69-1.47)	-
Secondary	188,891	0.94 (0.67-1.33)	-
Tertiary	21,763	0.49 (0.23-1.04)	-
Paternal			
Father death before birth to 4	580/1,136,136	2.41 (1.53-3.82)	-
Father death before birth to 4 not due to HIV/tuberculosis	430/1,136,136	2.07 (1.06-4.02)	1.44 (0.73-2.41)
Father death before birth to 4 due to HIV/tuberculosis	150/1,136,136	5.06 (2.48-10.41)	1.99 (0.86-3.68)
Household demographics			
Male household head	970,917	0.78 (0.60-1.02)	-
Mozambican household head	967,490	1.61 (1.24-2.09)	-
Household head age (continuous)	959,306	1.00 (0.99-1.01)	-
1-3 children in household aged <5 years	1,008,756	1.00	1.00
≥4 children in household aged <5 years	119,644	1.52 (1.21-1.91)	1.52 (1.18-1.90)
Household size (count of individuals)	1,136,132	1.02 (1.01-1.04)	-
Cumulative number of household deaths	1,136,136		
None	934,491	1.00	
1	155,545	41.32 (30.77-55.49)	-
2-3	43,514	62.95 (45.71-86.71)	-
≥4	2,586	106.90 (59.28-192.77)	-
Fertility			
Pregnancy parity	23,703	0.99 (0.90-1.08)	-
1	14,572	1.00	
2-4	8,560	1.05 (0.86-1.28)	-
≥5	571	0.20 (0.05-0.82)	-
Preceding birth interval	7,078	0.98 (0.95-1.01)	-
Post birth interval	6,925	0.90 (0.82-0.99)	-
Previous birth stillborn	288/23,703	0.68 (0.25-1.84)	-
Previous sibling died	800/23,703	1.63 (1.08-2.45)	-
Mother attended antenatal clinic	20,849/21,215	0.48 (0.28-0.83)	-
Number of antenatal clinic visits	14,119	0.92 (0.88-0.97)	-
Socio-economic status (SES) of household			
SES MCA[δ] score quintiles	358,524		
Most poor	63,390	1.00	
Very poor	69,347	0.74 (0.51-1.07)	-
Poor	73,736	0.63 (0.43-0.92)	-
Less poor	76,396	0.60 (0.41-0.87)	-
Least poor	75,655	0.43 (0.28-0.66)	-
Individual variation (O^2_j)	-	-	0.04 (0.01-0.11)
Spatial variation (O^2_z)	-	-	0.23 (0.10-0.48)
Temporal variation (O^2_t)	-	-	0.29 (0.12-0.68)

α, Infectious or parasitic child deaths only; β, no child deaths; δ, similar findings when using the absolute score index; household head age was not significant; straight distance from household and network distance from village centroid to nearest health facility (by category <5 and ≥5km) were not significant

We found a strong reduction in the probability of death as the child progresses to the age of 5 years (Fig. 3). This probability is much higher for those children whose mother died of HIV/tuberculosis and remains elevated throughout childhood. For those children whose mother died in childhood due to a cause unrelated to HIV/tuberculosis, we observed an elevated risk between 1-3 years of age, thereafter the risk drops close to levels observed for children whose mother did not die during their childhood.

Results of the multivariate Bayesian spatio-temporal model suggest that mother death between the child's first and fifth birthdays, particularly due to HIV/tuberculosis, was the most prominent risk factor from the multivariate analysis (Table 1), followed by father death due to HIV/tuberculosis, four or more children aged less than 5 years living within the household, Mozambican origin of the mother and winter season. Increasing age remained highly protective. The spatio-temporal model estimated the range or distance at which spatial correlation ceased between villages to be approximately 8.23 km (95% credible interval: 1.83 to 27.24 km). The autoregressive term was between -1 and 1 indicating stationarity.

Distinct foci of higher all-cause mortality risk can be observed (Fig. 4) in the central northern and south-eastern parts of the Agincourt study site. The villages in the southeast of the site are comprised mostly of former Mozambican refugees (>90% of village occupants). With increasing distance from the village centroids the standard error of the mortality risk prediction increases.

Child mortality risk due to infectious and/or parasitic causes was higher in the southeast corner of the site towards the Kruger National Park boundary. One additional focus of higher mortality was observed in the central northern part of the site. Distinct foci of reduced child mortality risk due to infectious or parasitic causes were also observed.

HIV/tuberculosis mortality risk showed two distinct foci: one in the upper central part of the site and a grouping in the southeast corner of the site. Acute diarrhoea or malnutrition mortality risk was mostly concentrated in the southeast corner of the site (Fig. 5). A small additional foci was also observed in the upper central part of the site. Mortality risk for acute respiratory infection or pneumonia child deaths was highest in the southeast, upper central and lower southwest of the site (Fig. 5). Accidental mortality risk showed a less distinct pattern however foci of higher mortality risk can be clearly observed across the site (Fig. 5).

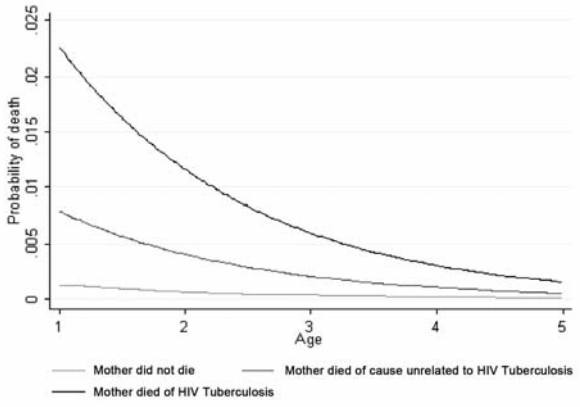


Fig. 3. Probability of death by age during 1-4 years and mother's cause-of-death based on a classical discrete-time logit model which included these two covariates

Fig. 6 shows the distribution of higher risk foci of mortality by period. While the focus of high risk in the southeast corner of the site persists across the period, by 2004 it has spread to surrounding villages. In addition, a new high risk area more upper centrally located emerges in the 2000-2003 period, and intensifies in the next period. Distinct foci of persisting and emerging lower mortality risk are also observed.

Discussion

In this study Bayesian spatio-temporal models were fitted to assess the geographical patterns and trends of all-cause and cause-specific child mortality in Agincourt HDSS. Results confirmed strong geographic differences in mortality risk and the importance of a number of risk factors such as maternal and paternal death (largely due to HIV), poorer SES and high household mortality burden.

A statistically significant increase in child mortality was observed over the study period, particularly between 1996 and 2003 largely due to the HIV epi-

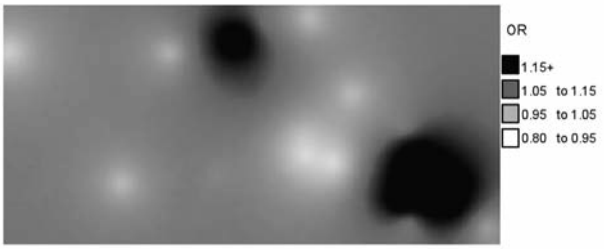


Fig. 4. Smoothed map of all-cause child mortality risk within Agincourt HDSS based on the baseline Bayesian model without covariates.

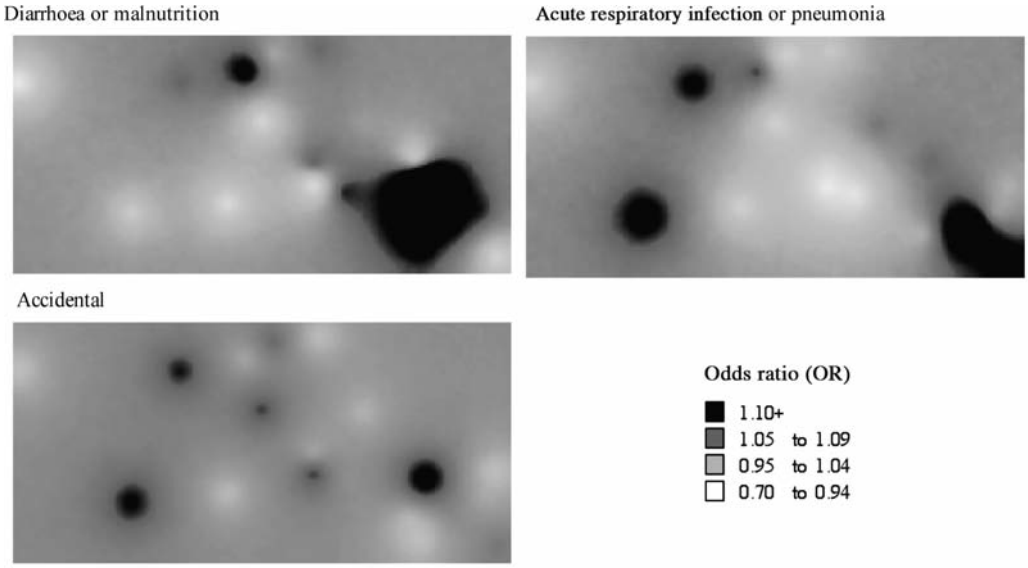


Fig. 5. Smoothed maps of selected cause-specific child mortality risk within Agincourt HDSS based on baseline Bayesian models without covariates.

demic. The leading causes of death were HIV/tuberculosis, followed by diarrhoea/malnutrition. One limitation of this study is the HIV/tuberculosis-related deaths misclassified by the verbal autopsy as unknown, which would underestimate the true burden. Almost half of all child deaths in this area could be attributed to infectious or parasitic causes. Interventions to reduce child mortality targeting infectious causes, specifically HIV and diarrhoea, are therefore urgently needed.

These findings confirm a number of risk factors documented in previous studies (Mosley and Chen, 1984a, b; Hobcraft et al., 1985; Binka et al., 1995; Ronsmans, 1996; Kuate Defo, 1997; Manda, 1999). There was a strong decreasing probability of mortality with increasing age of the child and higher probability of mortality associated with winter season. This is likely due to the increase in respiratory illness during this period, as well as environmental or household pollution due to the burning of fuel (e.g. coal, wood

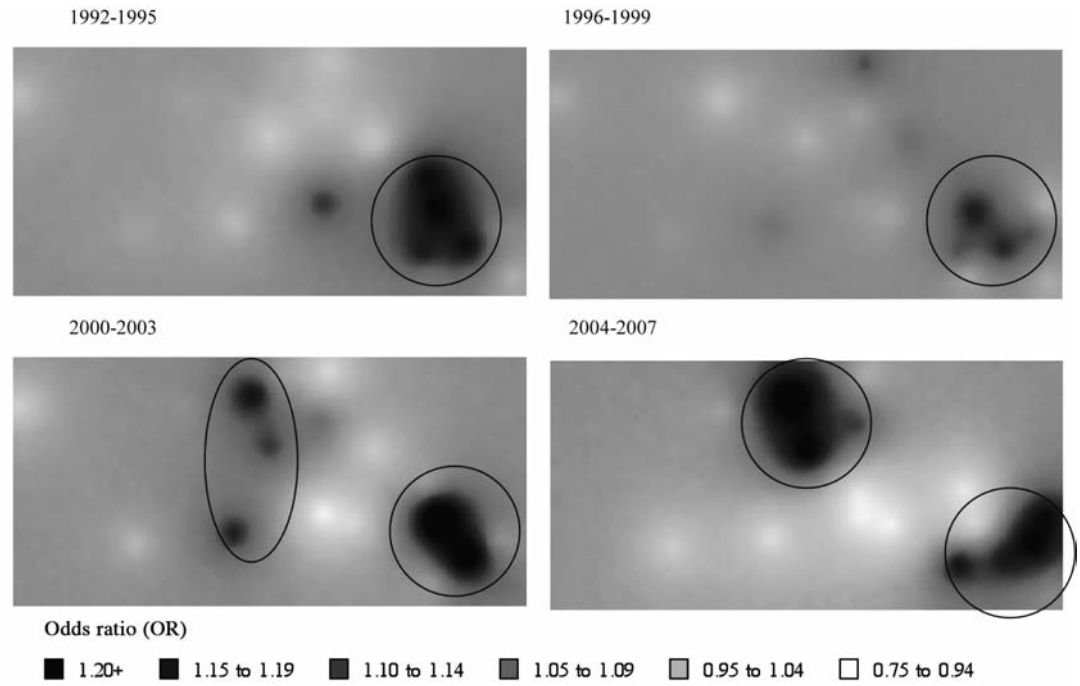


Fig. 6. Smoothed maps of all-cause child mortality risk within Agincourt HDSS by period based on the baseline Bayesian model without covariates.

and paraffin) for indoor heating and cooking. Results from a previous study in South Africa suggest that exposure to cooking and heating smoke from polluting fuels is significantly associated with <5 year mortality (Wichmann et al., 2006).

The spatial analysis indicated two distinct pockets of higher mortality burden towards the southeast and upper central parts of the site. This was consistent with the cause-specific mortality risk distribution we observed for HIV/tuberculosis, diarrhoea/malnutrition and acute respiratory infection/pneumonia. According to the space-time analysis, recent trends indicate the need to target interventions in the southeast corner (high risk throughout the period) and upper central (emerging risk from 2000 onwards) parts of the site which experience the highest mortality burden. This evolving space-time risk distribution is likely being driven by the evolving HIV epidemic.

A strong geographical pattern with regards to higher infectious disease mortality risk (particularly HIV/tuberculosis and diarrhoea) in the former Mozambican settlements lying to the east of the site was observed. We also found a significant risk for all-cause child mortality associated with having parents of Mozambican origin. Many refugees fleeing the civil war in Mozambique from 1983 onwards settled in the northeast of South Africa, including in Mpumalanga province. Despite voluntary repatriation programmes following a formal peace agreement in 1992, it was estimated that by 2000 more than 200,000 former Mozambican refugees were still living in the province (Johnston, 2000). They have remained a poorer and more vulnerable group, living in isolated villages with less or weaker in frastructure, poor access to water and sanitation, generally further away from health facilities and labour markets, with limited legal rights and experience barriers to accessing social grants, education and health services (Dolan et al., 1997, Hargreaves et al., 2004; Kahn, 2008). Hargreaves and colleagues (2004) have demonstrated higher mortality rates among children from Mozambican-headed households when compared to South African-headed households; lack of legal status and poorer SES of many former Mozambican refugees partly explains this disparity. Although primary health care is free, transport costs are high which represents a problem as the Mozambican settlements are generally further away from health services. Lack of legal status means Mozambican children do not have the same access to child support grants, which puts them at a further disadvantage relative to their South African counterparts. Hence, despite equity-orientated health and social policies in the southern part of the study site, it would appear that generally those children who are hardest to reach, both physically and socially, are also those with the highest mortality. Policy amendments are therefore needed to address inequity and any differential access to various services.

Death of the child's mother (and father to a smaller extent) between their first and fifth birthdays (specifically due to HIV/tuberculosis) was the most prominent risk factor in this study. Vertical transmission of HIV and/or loss of the primary care giver (leading to a loss of direct care as well as indirectly through household income lost) are the most likely explanation for this finding. Access to voluntary counselling and testing and anti-retroviral treatment to mothers and fathers in these settlements needs to be increased. About half of children infected with HIV through vertical transmission develop AIDS symptoms and die within 2 years (UNAIDS, 2002). Fig. 3 illustrates this issue. A comprehensive approach to prevention of mother-to-child transmission (PMTCT), including a combination of antiretroviral therapy (ART) from early pregnancy, elective caesarean section and highly active anti-retroviral therapy (HAART) for mother or infants during breastfeeding (Coovadia et al., 2007; WHO, 2010), can significantly reduce transmission rates in this sub-district (European Collaborative Study, 2005; Navér et al., 2006; Newell et al., 2007). In resource-poor settings, the risks and costs of surgical procedures are barriers to considering caesarean sections as a feasible strategy for preventing MTCT. The avoidance of breastfeeding in absence of HAART must also be balanced against the risks associated with replacement feeding such as cost and lack of access to clean water (Thior et al., 2006; Coovadia et al., 2007) which we have shown to be a problem in the south east region of the site.

The strong clustering of diarrhoea/malnutrition-related mortality risk in the southeast corner of the site, again in former Mozambican refugee settlements, points to a deficiency in water and/or sanitation infrastructure, which needs to be further assessed and addressed by local government. We recommend routine testing and improved water supply to reduce these unnecessary deaths. Rehydration fluid (ORT) and dietary management are key aspects in the treatment of acute diarrhoea, particularly those episodes which persist. The capabilities and resources of health facilities, specifically those situated near to the southeast corner of the site, to effectively manage children presenting with diarrhoea and/or malnutrition needs to be improved.

The association between SES, maternal education and mortality has been previously described by Farah and Preston (1982). Higher education may result in better health awareness and utilisation of health facilities (Jain, 1988), higher income and the ability to purchase goods and services that improve children's health (Schultz, 1979), longer birth intervals and, possibly, higher maternal ages (Cleland et al., 1989).

Death of household members other than the mother or father also appeared to be a significant risk in this study. Thus death of these members places additional strain (including financial burden of medical services and funerals, and loss of income) on the household, which negatively impacts on the child's health outcome. High mortality households require both financial and social support to reduce the indirect impact on their children.

There are several studies relating geographical access to use of health facilities. As one would expect, members of communities that are more distant use the facilities less than those that live nearer, but this does not necessarily translate into increased mortality risk (Stock, 1983; Becher et al., 2008). In this study no significant increased risk of child mortality was found with increasing distance of household to the nearest primary health care clinic as well as district hospitals located outside the site. The same holds true for infant mortality (Sartorius et al., 2010). This suggests that quality of health services may be influencing child mortality rather than geographical access. Evaluation of primary health care services with attention to quality improvements is needed.

Conclusion

Our study has demonstrated the considerable potential of spatial statistical methods for analysing longitudinal health and socio-demographic data, and is one of only a few studies to have used geostatistical modelling on HDSS mortality data. Based on the space-time analysis the southeast and upper central regions of the site appear to have the highest child mortality risk at present. These maps are particularly helpful in identifying high mortality areas to guide efficient allocation of limited resources in child survival programmes. The risk factor results can also contribute to policies to address health inequalities in rural South Africa and to improve access to health services. Targeted efforts to prevent vertical transmission of HIV in specific villages need to be introduced, as well as programmes to ensure the survival of the mother and father through children's childhood, as both emerged as prominent risk factors for child mortality.

Acknowledgements

This work was supported by a PhD fellowship from the South African Centre for Epidemiological Modeling and Analysis (SACEMA), a National Research Foundation (NRF) Centre of Excellence, as well as research fellowships from the MRC/Wits Rural Public Health and Health Transitions Research Unit (Agincourt) through the Wellcome Trust (grant no. 069683/Z/02/Z), UK, and the Swiss-South Africa Joint Research Programme (project no. JRP IZLSZ3_122926). Additional funding was provided by a Wits Faculty of Health Sciences Medical Research Endowment Fund (MREF) (grant no. SARB000) and a research fellowship from the MRC/Wits Rural Public Health and Health Transitions Research Unit (Agincourt) through the Wellcome Trust, UK. The Agincourt HDSS and GIS was funded by the Wellcome Trust, UK (grant no. 058893/Z/99/A, 069683/Z/02/Z, 069683/Z/08/Z), the University of the Witwatersrand, the South African Medical Research Council, and the Andrew W. Mellon Foundation, USA. P. Vounatsou was supported by the Swiss National Science Foundation (project no. 325200_118379).

Appendix: Statistical model

Let Y_{ijt} be the mortality status of child i in village j at time interval t. We assume that Y_{ijt} arises from a Bernoulli distribution, $Y_{ijt} \sim \text{Be}(p_{ijt})$ where p_{it} is the probability of death of child i at interval t. We model covariates X_{it}, temporal random effect α_t, village φ_t and child ε_t random effects on the logit of p_{ijt}, that is $\text{logit}(p_{ijt}) = X_{it}^T + \beta + \alpha_t + \varphi_j + \varepsilon_i$, where β is the vector of regression coefficients. We assume that φ_i has a multivariate normal distribution, $\varphi \sim MVN(0, \Sigma)$ with variance-covariance matrix Σ. We also assume an isotropic stationary spatial process, where $\Sigma_{kl} \sim \sigma_\varphi^2 \exp(-\rho d_{kl})$, d_{kl} is the Euclidean distance between villages k and l, σ_φ^2 is the geographical variability known as the sill, ρ is a smoothing parameter that controls the rate of correlation decay with increasing distance and measures the range of geographical dependency (Diggle et al.., 1998). The range is defined as the minimum distance at which spatial correlation between locations is below 5%. This distance can be calculated as $3/\rho$ and is expressed in meters. The year level temporal random effect α_t t = 1, 2, ..., 16, was modelled via a second order stationary autoregressive process that is $\alpha_t \sim N(\gamma_1 \alpha_{t-1} + \gamma_2 \alpha_{t-2}, \sigma_\varphi^2)$, $t > 2$, where $-2 < \gamma_1 < 2$ and $-1 < \gamma_2 < 1 - |\gamma_1|$ (Harvey, 1993) and the child random effects via independent normal distributions $\varepsilon_i \sim N(0, \sigma_z^2)$ with variance σ_z^2. Uniform prior distributions were adopted for γ_1, γ_2, vague Normal distributions for the β, inverse gamma priors for the variance parameters and a uniform prior for ρ. The model was fitted using MCMC in OpenBUGS. A burn in of 5,000 was chosen

and run until the Monte Carlo error was <5% SD for each covariate, thereafter run until sample of 10,000 obtained.

References

Adetunji J, 2000. Trends in under-5 mortality rates and the HIV/AIDS epidemic. Bull World Health Organ 78, 1200-1206.

Ahmad OB, Lopez AD, Inoue M, 2000. The decline in child mortality: a reappraisal. Bull World Health Organ 78, 1175-1191.

Becher H, Kynast-Wolf G, Sie A, Ndugwa R, Ramroth H, Kouyate B, Müller O, 2008. Patterns of malaria: cause-specific and all-cause mortality in a malaria-endemic area of West Africa. Am J Trop Med Hyg 78, 106-113.

Binka FN, Maude GH, Gyapong M, Ross DA, Smith PG, 1995. Risk factors for child mortality in northern Ghana: a case-control study. Int J Epidemiol 24, 127-135.

Black RE, Morris SS, Bryce J, 2003. Where and why are 10 million children dying every year? Lancet 361, 2226-2234.

Cleland J, van Ginneken JK, 1989. Maternal education and child survival in developing countries: the search for pathways of influence. Soc Sci Med 27, 1357-1368.

Clements ACA, Garba A, Sacko M, Touré S, Dembelé R, Landouré A, Bosque-Oliva E, Gabrielli AF, Fenwick A, 2008. Mapping the probability of schistosomiasis and associated uncertainty, West Africa. Emerg Infect Dis 14, 1629-1632.

Coovadia H, Bland R, 2007. Preserving breastfeeding practice through the HIV pandemic. Trop Med Int Health 12, 1-18.

Coovadia H, Rollins N, Bland M, Little K, Coutsoudis A, Bennish M, Newell ML, 2007. Mother-to-child transmission of HIV-1 infection during exclusive breastfeeding in the first 6 months of life: an intervention cohort study. Lancet 369, 1107-1116.

D'Agostino R, Lee M, Belanger A, 1990. Relation of pooled logistic regression to time dependent Cox regression analysis: The Framingham Heart Study. Stat Med 9, 1501-1515.

Diggle P, Tawn J, Moyeed R, 1998. Model-based geostatistics. Appl Stat 47, 299-350.

Dolan C, Tollman S, Nkuna V, Gear J, 1997. The links between legal status and environmental health: a case study of Mozambican refugees and their hosts in the Mpumalanga (Eastern Transvaal) Lowveld, South Africa. Int Health Hum Rights 2, 62-84.

European Collaborative Study, 2005. Mother-to-child transmission of HIV infection in the era of highly active antiretroviral therapy. Clin Infect Dis 40, 458-465.

Farah AA, Preston SH, 1982. Child mortality differentials in Sudan. Popul Dev Rev 8, 365-383.

Gelfand A, Ravishanker N, Ecker M, 1999. Modelling and inference for point-referenced binary spatial data. In: Generalized Linear Models: A Bayesian Perspective, Dey D, Ghosh S, Mallick B, eds. Marcel Dekker Inc. 373-386.

Gelfand A, Smith A, 1990. Sampling-based approaches to calculating marginal densities. J Am Stat Assoc 85, 398-409.

Gemperli A, Vounatsou P, Sogoba N, Smith T, 2006. Malaria mapping using transmission models: application to survey data from Mali. Am J Epidemiol 163, 289-297.

Gosoniu L, Vounatsou P, Tami A, Nathan R, Grundmann H, Lengeler C, 2008. Spatial effects of mosquito bednets on child mortality. BMC Public Health 8, 356.

Hargreaves J, Collinson M, Kahn K, Clark S, Tollman S, 2004. Childhood mortality among former Mozambican refugees and their hosts in rural South Africa. Int J Epidemiol 33, 1271-1278.

Hay S, Guerra CA, Gething PW, Patil AP, Tatem AJ, Noor AM, Kabaria CW, Manh BH, Elyazar IRF, Brooker S, Smith DL, Moyeed RA, Snow RW, 2009. A world malaria map: *Plasmodium falciparum* endemicity in 2007. PLoS Med 6, e1000048.

Hobcraft JN, McDonald JW, Rutstein SO, 1985. Demographic determinants of infant and early child mortality: a comparative analysis. Population Stud 39, 363-385.

Jain A, 1988. Determinants of regional variation in infant mortality in rural India. In: Infant Mortality in India: Differentials and Determinants, Jain A, Visaria L, eds. Sage Publications, 127-167.

Johnston N, 2000. The Regularisation of Former Mozambican Refugees in South Africa: Experiences and Lessons. Acornhoek, Refugee Research Programme.

Kahn K, 2008. Dying to make a fresh start: mortality and health transition in a new South Africa. Umeå University Medical Dissertations New Series No. 1056, ISSN 0346-6612 ISBN 91-7264-173-8.

Kahn K, Tollman SM, Collinson MA, Clark SJ, Twine R, Clark BD, Shabangu M, Gómez-Olivé FX, Mokoena O, Garenne ML, 2007. Research into health, population and social transitions in rural South Africa: data and methods of the Agincourt health and demographic surveillance system. Scand J Public Health 35 (Suppl 69), 8-20.

Kahn K, Tollman SM, Garenne M, Gear JS, 2000. Validation and application of verbal autopsies in a rural area of South Africa. Trop Med Int Health 5, 824-831.

Kuate Defo B, 1997. Effects of infant feeding practices and birth spacing on infant and child survival: a reassessment from retrospective and prospective data. J Biosoc Sci 29, 303-326.

Manda S, 1999. Birth intervals, breastfeeding and determinants of childhood mortality in Malawi. Soc Sci Med 48, 301-312.

MapInfo, 2008. MapInfo Professional. Version 9.5. Pitney Bowes Software Inc., One Global View, Troy, New York.

Mosley WH, Chen LC, 1984a. An analytical framework for the study of child survival in developing countries. In: Child Survival: Strategies for Research, Mosley WG, Chen LC, eds, New York: Population Council, 25-44.

Mosley WH, Chen LC, 1984b. An analytic framework for the

study of child survival in developing countries. Popul Dev Rev 10, 25-45.

Mulholland E, Smith L, Carneiro I, Becherc H, Lehmannd D, 2008. Equity and child-survival strategies. Bull World Health Organ 86, 399-407.

Müller O, Garenne M, 1999. Childhood mortality in sub-Saharan Africa. Lancet 353, 673.

Navér L, Lindgren S, Belfrage E, Gyllensten K, Lidman K, Gisslén M, Ehrnst A, Arneborn M, Bohlin AB, 2006. Children born to HIV-1-infected women in Sweden in 1982-2003: trends in epidemiology and vertical transmission. J Acquir Immune Defic Syndr 42, 484-489.

Newell ML, Huang S, Fiore S, Thorne C, Mandelbrot L, Sullivan JL, Maupin R, Delke I, Watts DH, Gelber RD, Cunningham CK, PACTG 316 Study Team, 2007. Characteristics and management of HIV-1-infected pregnant women enrolled in a randomised trial: differences between Europe and the USA. BMC Infect Dis 7, 60.

Penchansky R, Thomas J, 1981. The concept of access – definition and relationship to consumer satisfaction. Med Care 19, 127-140.

R Development Core Team, 2008. R: a language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. ISBN 3-900051-07-0, URL http://www.R-project.org.

Raso G, Matthys B, N'Goran EK, Tanner M, Vounatsou P, Utzinger J, 2005. Spatial risk mapping and prediction of *Schistosoma mansoni* infections among schoolchildren living in western Côte d'Ivoire. Parasitology 131, 97-108.

Riedel N, Vounatsou P, Miller J, Gosoniu L Chizema-Kawesha E, Mukonka V, Steketee RW, 2010. Geographical patterns and predictors of malaria risk in Zambia: Bayesian geostatistical modelling of the 2006 Zambia national malaria indicator survey (ZMIS). Malar J 9, 37.

Ronsmans C, 1996. Birth spacing and child survival in rural Senegal. Int J Epidemiol 25, 989-997.

Sartorius BKD, Kahn K, Vounatsou P, Collinson MA, Tollman SM, 2010. Young and vulnerable: spatial-temporal trends and risk factors for infant mortality in rural South Africa (Agincourt), 1992-2007. BMC Public Health 10, 645.

Schotman P, 1994. Priors for the AR (1) model: parameterization issues and time series considerations. Econom Theor 10, 579-595.

Schultz TP, 1979. Interpretation of relations among mortality, economics of the household and the health environment. In: Proceedings of the Meeting on Socio-Economic Determinants and Consequences of Mortality Mexico City, June 19-25 Geneva: World Health Organization.

Schur N, Hürlimann E, Garba A, Traoré MS, Ndir O, Ratard RC, Tchuem Tchuenté LA, Kristensen TK, Utzinger J, Vounatsou P. 2011. Geostatistical model-based estimates of schistosomiasis prevalence among individuals ≤20 years in West Africa. PLoS Negl Trop Dis (in press).

Spiegelhalter D, Best N, Carlin B, van der Linde A, 2002. Bayesian measures of model complexity and fit. J R Stat Soc B 64, 583-639.

Spiegelhalter D, Thomas A, Best N, 1999. WinBUGS Version 1.2 User Manual MRC Biostatistics Unit.

StataCorp, 2007. Stata Statistical Software: Release 10. College Station, TX: StataCorp. LP.

Stock R, 1983. Distance and the utilization of health facilities in rural Nigeria. Soc Sci Med 17, 563-570.

Thior I, Lockman S, Smeaton LM, Shapiro RL, Wester C, Heymann SJ, Gilbert PB, Stevens L, Peter T, Kim S, van Widenfelt E, Moffat C, Ndase P, Arimi P, Kebaabetswe P, Mazonde P, Makhema J, McIntosh K, Novitsky V, Lee TH, Marlink R, Lagakos S, Essex M; Mashi Study Team, 2006. Breastfeeding plus infant zidovudine prophylaxis for 6 months vs formula feeding plus infant zidovudine for 1 month to reduce mother-to-child transmission in Botswana. JAMA 296, 794-805.

Tollman SM, Herbst K, Garenne M, Gear JS, Kahn K, 1999. The Agincourt demographic and health study-site description, baseline findings and implications. S Afr Med J 89, 858-864.

Trapé JF, 2001. The public health impact of chloroquine resistance in Africa. Am J Trop Med Hyg 64 (Supp), 12-17.

Tulloch J, 1999. Integrated approach to child health in developing countries. Lancet 354 (Suppl 2), 16-20.

UNAIDS, 2002. Paediatric HIV infection and AIDS: UNAIDS point of view. (UNAIDS best practice collection. Point of view.) Geneva: UNAIDS.

Victora CG, Wagstaff A, Schellenberg JA, Gwatkin D, Claeson M, Habicht JP, 2003. Applying an equity lens to child health and mortality: more of the same is not enough. Lancet 362, 233-241.

Wagstaff A, 2000. Socioeconomic inequalities in child mortality: comparisons across nine developing countries. Bull World Health Organ 78, 19-29.

WHO, 2002. The world health report 2002: reducing risks, promoting healthy life. Geneva: World Health Organization.

WHO, 2010. Guidelines on HIV and infant feeding. Principles and recommendations for infant feeding in the context of HIV and a summary of evidence. Geneva: World Health Organization. http://www.who.int/child_adolescent_health/documents/9789241599535/en/index.html (accessed 4 September 2010).

Wichmann J, Voyi K, 2006. Influence of cooking and heating fuel use on 1-59 month old mortality in South Africa. Matern Child Health J 10, 553-561.

Zeller A, 1996. An Introduction to Bayesian Inference and Econometrics. John Wiley and Sons Ltd/Inc: New York.

Dying in their prime: spatial-temporal trends and risk factors for adult mortality in the Agincourt rural sub-district, South Africa, 1992-2008

Authors

Benn Sartorius[1]*, Kathleen Kahn[1,2,3], Penelope Vounatsou[4,5], Mark A. Collinson[1,2,3], Stephen M. Tollman[1,2,3]

1 MRC/Wits Rural Public Health and Health Transitions Research Unit (Agincourt), School of Public Health, Faculty of Health Sciences, University of the Witwatersrand, Johannesburg, South Africa; 2 Centre for Global Health Research, Epidemiology and Global Health, University of Umeå, Umeå, Sweden; 3 INDEPTH Network, Accra, Ghana; 4 Department of Epidemiology and Public Health, Swiss Tropical and Public Health Institute, Basel, Switzerland; 5 University of Basel, Basel, Switzerland

*Corresponding author:
Benn Sartorius
School of Public Health,
Faculty of Health Sciences, University of the Witwatersrand,
Johannesburg, South Africa
Tel: +27-11-717-2627; Fax: +27-11-717-2084
E-mail: benn.sartorius@wits.ac.za

Abstract

Background: Adult mortality is an important indicator for the assessment of the overall health and mortality patterns within a population, for evaluating variations in access to health services and planning health care interventions. Despite its important implications adult mortality has been neglected in many developing countries where data are often absent.

Methods: Younger and older adult (15-49 and 50-64 years) data were used from the Agincourt health and socio-demographic surveillance system for the period 1992-2008. Levels, causes and trends of mortality were assessed. A Bayesian hierarchical geostatistical parametric survival model was used to identify factors associated with adult mortality. Simulation-based Bayesian kriging was used to map high or emerging risk areas.

Results: Adult mortality increased significantly with time due to HIV/AIDS. Prominent risk factors were male gender, migrancy, increasing number of household deaths, household head death and distance from household to nearest health facility more than 6km. Protective factors identified were higher household socio-economic status, having a male or older (40 plus years of age) household head, increasing education status in 15-49 age-group and Mozambican nationality in 50-64 age-group. Distinct foci of higher spatial mortality risk were observed in six villages. These were fairly dispersed, except for one cluster of three villages in the upper central region.

Discussion and conclusions: HIV/AIDS has had a major impact on adult mortality in this rural sub-district. There are strong geographical and temporal differences in mortality that vary even across a relatively small area. A complex interaction of factors, for example HIV-poverty-education-migration, appears to be driving communicable adult mortality in space-time and resulting in vulnerable households and villages. Mozambican nationality was protective in the older age-group (50-64) indicating that the generally lower SES of Mozambicans may protect them from lifestyle related non-communicable disease mortality compared to the less poor South Africans. The risk maps generated can be used by decision-makers for planning of ART rollout, specifically ensuring access for the six higher risk villages. Greater distance to nearest health facility was a significant risk, indicating the need to address inequalities related to access.

Keywords: adult mortality; health and demographic surveillance; survival modelling; Bayesian inference; risk kriging; South Africa

Introduction

The level of adult mortality is becoming an important indicator for the assessment of the overall health and mortality status of a population, for evaluating variations in accessibility to quality health care services and planning health care interventions (1-3) Despite the important implications of adult mortality, this phenomenon has been neglected in many developing countries. In addition to the socio-economic burden of younger adult mortality on households and society, it impacts negatively on the young and elderly in a number of direct and indirect ways including funeral and medical expenditures; indirect costs providing care to ill adult children, fostering of grandchildren, and the loss of remittances and income (4-5). Furthermore, burden of disease data is vital to understanding rapidly evolving health transitions. In particular, HIV/AIDS has reduced life expectancy in many African countries in recent times, with South Africa particularly severely affected (6-7). The growing importance of this indicator is also particularly highlighted by the increasing burden of non-communicable diseases among economically productive adults by ageing trends and health transitions (8). In sub-Saharan Africa, despite the HIV epidemic, non-communicable disease among adults is increasing due to population ageing and potentially changes in lifestyle risk factors (9-10). However, knowledge of adult mortality levels to inform local and national policy is hampered by absent or incomplete vital registration systems and delays in death statistics in sub-Saharan Africa (10). Health and socio-demographic surveillance systems HDSS) implementing a verbal autopsy (VA) to determine probable cause of death are therefore often the only means to determine cause-specific mortality on a longitudinal basis (10-11) , and thus a valuable tool for assessing mortality trends. Longitudinal data is particularly useful for monitoring chronic diseases, for which the least is probably known, as they allow one to study the life course and accumulation of risk and also provide accurate time sequences of events and intra-individual change over time which is important for causal inference (12). Given the empirical basis for public health policies and interventions in sub-Saharan Africa is essentially absent, verbal autopsy has great

potential to promote an understanding of adult cause-specific mortality (10), how it is evolving in space and time and what factors may be driving this process.

Knowledge of risk factors driving adult mortality in developing countries is fundamental for planning interventions and monitoring their impact. The use of survival analysis in the health field is extensive; however, few studies have assessed risk factors for adult mortality in Africa where data are often lacking. Spatial and spatial-temporal disease or mortality maps can provide researchers and policy makers with visual images displaying the evolving patterns of adult mortality as well as causal processes and space-time geostatistics can also often help improve missing or imperfect data (13). A number of methods from epidemiology, geostatistics and small area modelling have converged to provide powerful new, though complex and time consuming, models with which are capable of analyzing health data (13-16). In particular, the development of Markov chain Monte Carlo (MCMC) simulation methods and software, such as OpenBUGS, and Bayesian approaches, are being applied to the analysis of social and health problems, disease mapping, modelling or kriging (14). The widespread use of geographic information systems (GIS) that link to statistical packages has also advanced spatial data analysis. While the use of geostatistical spatial modelling in epidemiological research has increased in recent years (16-22), there has been little advanced spatial analysis of adult mortality using longitudinal HDSS data and very few risk factor studies for adult mortality with correct estimation of predictor significance using correlated longitudinal data.

The application of this methodology will contribute to our understanding of factors associated with adult mortality, identifying high risk areas for more precise targeting of public health interventions. Space-time changes in adult mortality can highlight emerging trends and future hot spots. The analysis aims to provide feedback to assess the effectiveness of existing public health policy; and to provide guidance as to the best allocation of limited resources in a poor rural setting where inequalities may exist.

Methods

Study area and population

The Agincourt HDSS, located in northeast South Africa, was established in 1992. There was a baseline census in 1992 that collected data on all individuals and households in the population (23). This has been followed by annual updates of births, deaths, in and out-migrations. It is a poor rural sub-district that includes former Mozambican refugees, temporary migrant workers and a more stable permanent population (24). The site covers an area of about 400km^2 and contains 25 villages, 13,500 households and 84,000 individuals. There is a full geographic information system (GIS), containing locations of all households within the site, which is updated annually. A household is defined as a group of people who reside and eat together, plus the linked temporary migrants who would eat with them on return. Verbal autopsies (VA) were introduced in 1993. A full VA is conducted on every death recorded during the annual census update. This is carried out by a specially trained fieldworker who interviews the closest caregiver of the

deceased (25). VA's are assessed by three medical practitioners. Two doctors independently review the information and assign a probable cause to each death. If the diagnoses agree then it is accepted, however should they differ, then the two clinicians discuss in an effort to reach consensus. If this proves unsuccessful then a third practitioner, blind to earlier findings, assesses the VA. If the third assessment is congruent with one other, it is accepted as the "probable cause of death"; if not, the cause is coded as "ill-defined" (25). Causes of death (main, immediate and/or contributing) are coded to be consistent with the International Classification of Diseases (ICD-10).

The study population comprised all adults (15-49 and 50-64 years) in 21 villages during the period 1992-2008. Data from four new villages added to the site since 2007 were not included in the analysis since death ascertainment was incomplete.

Outcome and explanatory variables

The outcome in this study was defined as time (in years) contributed by an adult during the study period until right censoring (0) or death (1). The time to right censoring was set to either the date of permanent outmigration during the study period or as 31 December 2008 if the individual was present and alive at this endpoint. The explanatory variables included individual level factors (age, gender, nationality, education years, migrant patterns), household level factors (size, household age, deaths, socio-economic status (SES), household head demographics, distance to nearest health facility) and village level factors (size, number of deaths, proportion of deaths attributed to HIV/TB, migration patterns). A temporary migrant is defined as a labor migrant who resides in the household for less than six months of the year but whose return is assumed by the household respondent i.e. retains a significant link to their base household (26). A permanent migrant is defined as a person who enters or leaves a household with a permanent intention of entering or leaving. Household SES s was based on living conditions and assets including building materials of main dwelling, water and energy supply, ownership of modern appliances and livestock, and means of transport. These assets were used to construct an SES index using Multiple Correspondence Analysis (MCA). The MCA index was selected as it makes fewer assumptions about the underlying distributions of the indicator variables and was more suitable given the categorical nature of the HDSS asset status module (27-29).

Risk factor analysis

Selected explanatory variables (e.g. education years by gender or ethnicity) were compared against one another using the standard t-test to better identify any potential significant underlying causal processes as suggested by the risk factor analysis.

The risk factor analysis was split into two age groups, namely 15-49 and 50-64 years to better address the different risk profile associated with each. A preliminary non-parametric Cox survival analysis was initially conducted to assess the relationship between adult mortality and each covariate. Time contributed by adults was split into discrete continuous segments to incorporate any changes to household location.

Covariates significant at the 10% level (without substantial missing values) were then incorporated into the multivariate model. The assumption of proportional hazards was not upheld in the multivariate model and various parametric survival models were tested with the one providing best fit finally adopted.

Given the inherent spatial and temporal correlation of longitudinal HDSS data, problems arise when using standard statistical methods as they assume independence of outcome measures (e.g. mortality). Objects in close proximity are often more alike and common exposures (measured or unmeasured) may influence adult mortality similarly in households of the same geographical area, introducing spatial correlation in mortality outcomes. Including the spatial effect of proximity is important for efficient estimation of parameters and prediction (30). Ignoring this correlation introduces bias in the risk factor analysis as the standard error of the covariates is underestimated, thereby overestimating the significance of the risk factors. Geostatistical models relax the assumption of independence and assume that spatial correlation is a function of distance between locations. They are highly parameterised models and their full estimation has only become possible in the last decade by formulating them within a Bayesian framework (31) and estimating the parameters via Markov chain Monte Carlo (MCMC) simulation. With the development of MCMC methods and software such as OpenBUGS (32), Bayesian approaches are being applied to the analysis of many social and health problems in addition to disease mapping and modelling or kriging (14). Thus Bayesian geostatistical multivariate models are needed to analyse longitudinal data in order to address these problems.

A Bayesian hierarchical geostatistical parametric model, assuming the widely used Weibull distribution for the underlying hazard, was used to examine the multivariate association between the significant covariates and adult mortality. A spatial random effect at the village level was included to take account of spatial correlation and was modelled using a multivariate Gaussian distribution with a covariance matrix expressed as a parametric function of the distance between pairs of village centroids points (31). Furthermore, an unstructured household-level random effect was included to take into account repeated household observations where time episodes were split to incorporate any time varying issues such as change of household physical location. Markov Chain Monte Carlo (MCMC) simulation (33) was employed to estimate the model parameters using OpenBUGS (32). Further details of the statistical modelling approach are given in Appendix 1.

Model assessment

Model comparison in Stata was based on the Akaike Information Criterion (AIC). We selected the model with smallest value of AIC and then graphically examined model fit in Stata using Cox-Snell residual plots. The Deviance Information Criterion (DIC) was used to assess the various multivariate models (34). Both the AIC/DIC are a measure of the relative goodness of fit of a statistical model. Generally the smaller the AIC/DIC, the better the model fit.

Risk maps

Simulation-based Bayesian kriging (35) at numerous grid prediction points within the site was used to produce smoothed maps of all-cause and cause-specific adult mortality risk within the whole HDSS area. All-cause and cause-specific baseline models were used that included no covariates except a constant and spatial random effect. All identifying features (such as village centroids, boundaries) have been removed from the maps to ensure confidentiality and avoid stigmatising of high risk villages. The HIV and tuberculosis mortality risk map is also not shown for the abovementioned reason. Model estimates were exponentiated to hazard ratios (HR) given the time to event (survival) approach. A HR of less than one indicates a lower than baseline risk i.e. fewer than expected events, a value of one indicates baseline risk, while >one indicates higher than baseline risk. A grayscale is applied to the location specific HR prediction with darker areas (black representing maximum risk) reflecting increasingly higher risk and increasingly lighter areas (white representing the minimum risk) indicating lowering risk. A simple map showing the spatial risk of all-cause adult mortality as a function of straight-line distance to nearest health facility using circular buffer zones around health facilities was also constructed.

Software

Data were extracted using Microsoft SQL Server 2005. The analysis was carried out in STATA version 10.0 SE (Stata Corp., 2007) and OpenBUGS (32). The predictions of the fitted age-specific or cause-specific spatial models were constructed in MapInfo Professional version 9.5 (MapInfo, 2008).

.

Results

Descriptive analysis

Between 1992 and the end of 2008 there were 88,509 adults within the 21 villages originally set up as the Agincourt study site. There were 86,883 records as certain individuals changed location within the site on one or more occasions. The mean age at the start of their residence within the site was 26.1 years (Std. Dev. 12.1) with a median age of 22.4 (IQR: 15.1-32.0). The majority were female (49,890 or 57.4%). Of those whose nationality was known, 33.3% or 28 287 were of Mozambican origin with the remainder virtually all being South African. The average number of years of education attained by the adult study sample (15-64 years) was 6.2 (Std. Dev. 4.5) with this mean being significantly higher in males compared to females (6.29 versus 6.06, $p<0.001$). Similarly the mean years of education attained by South Africans was significantly higher than that of their Mozambican counterparts (7.14 versus 4.30, $p<0.001$). Males were significantly more likely to spend on average fewer resident months within the site each year compared to females (mean of 9.76 versus 7.98, $p<0.001$). Similarly, South African adults were more likely to spend on average fewer resident months within site compared to Mozambicans (8.93 versus 9.01, $p<0.001$).

Cause of death (COD)

Of the 4,873 adult deaths (15-64 years) that occurred before 2008, the majority were attributed to communicable causes (44% or 2,134), largely due to HIV/TB (39% or 1,876). In the 15 to 49 and 50 to 64 year age-groups, 43% (1,552) and 25% (324) attributed to HIV/TB respectively. Non-communicable diseases were responsible for 16% (759) of all adult deaths pre-2007. Older Mozambican adults (50-64 years) had a lower proportion of non-communicable disease mortality when compared to South African's, 21% (67) and 33% (317) respectively. Older female adults had a higher proportion of mortality attributed to non-communicable disease than males, 33 % (172) versus 28% (212).

Survival and temporal trend analysis

During the period 1992-2008, there were 5,923 adult deaths, a total of 615,496 person-years and a mortality rate of 962.3 adult deaths per 100,000 person-years. The mean and median time at risk was 6.96 and 5.47 years respectively (range: 0.003-17.0 years). The Kaplan-Meier function estimates a 25% chance of an adult dying after duration of 17.0 years. Mortality significantly increased in both age groups over the study period; however the most pronounced increase occurred from 1999 onwards (Figure 1). Older adult mortality (50-64) was stable overall and by gender (higher in males though) until 1999 (1,161 deaths per 100,000 person-years). Rates increased from 2000 to reach a plateau of 2,977 deaths per 100,000 person-years in 2005 with similar levels observed through to 2008. This change was largely due to the more pronounced increase in mortality rates among males (Figure 1). There was little difference in the mortality rates of younger adult males and females (15-49) and overall a steady significant increase was observed from around 1996 (344 deaths per 100,000 person-years) to reach a maximum in 2007 of 1,152 deaths per 100,000 person-years.

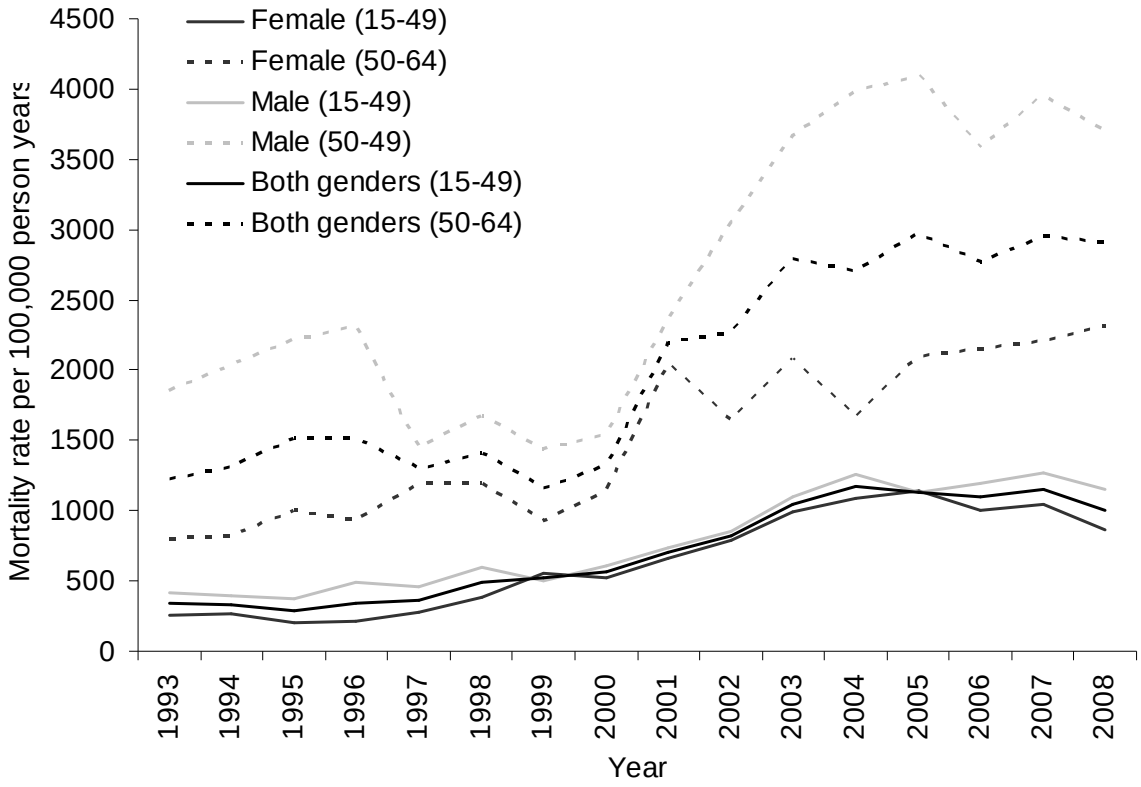


Figure 1: Annual mortality rates per 100,000 person-years by adult age group and gender, Agincourt sub-district, 1993-2008

Univariate analysis

The most prominent and significant risk factors from the univariate analysis were later time period, older age, male gender, being a migrant (i.e. spending six months or more on average outside the Agincourt sub-district each year), increasing number of household deaths, death of household head and villages with a mortality proportion due to HIV/TB above the median value (Table 1). Increasing wealth status of household, male gender of household head and household head age above 40 years were significant protective factors in both younger and older age groups. Increasing education status was protective in the 15-49 year age-group only while Mozambican nationality was protective in the older age-group only (50-64 years). Distance to nearest health facility (>6km) was a significant risk factor in the 15-49 age-group and appeared to increase the mortality risk in the 50-64 year age group (not significant at the 10% level due to small numbers of deaths).

Multivariate analysis

The multivariate analysis for both age-groups using classical and Bayesian approaches yielded similar findings with wider confidence intervals for the Bayesian approach, as would be expected, given the correction for spatial and unstructured correlation (Table 2 and 3).

The most prominent risk factor for 15-49 year mortality following multivariate adjustment were later time period, male gender, being a migrant, increasing number of household deaths, death of household head and distance to nearest health facility (>6km). Increasing wealth of household, household head being male and older than 40 years were significant and prominent protective factors. No significant difference was observed between Mozambicans and South Africans. Villages with a mortality proportion of HIV/TB above the median value remained at a significantly higher risk.

The most prominent risk factors for 50-64 year mortality following multivariate adjustment were later time period, male gender, being a migrant and death of the household head. Increasing wealth of household, household head being male and older than 40 years were again prominent protective factors. Mozambican nationality appeared to have significantly lower risk in this age group compared to South Africans.

Spatial risk maps of adult mortality

The distribution of gender-specific mortality risk based on Bayesian kriging was similar for males and females (Figure 2). Five distinct foci of higher mortality in the 15-49 year age-group can be seen in Figure 3. Three are in the central to upper central region of the site and two in the south east. These correlate to areas with higher risk of infectious disease mortality, (Figure 4), largely HIV/TB, with a strong clustering in three villages in the upper central region. One village in particular has repeatedly emerged as a mortality hotspot, especially in recent years. This village has a significantly younger and highly mobile migrant population ($p<0.001$) compared to all other villages. A similar pattern was seen in the 50-64 age-group except for one village in the south east that was no longer higher risk and one additional village in the upper central region emerged as high risk. This distribution was also largely driven by rising chronic infectious disease (HIV/TB) mortality. Higher non-communicable disease mortality risk in the older adult age-group was observed in three distinct foci in the upper central , central and south east of the site (Figure 4). Based on the circular buffer zones around health facilities (to relate straight-line distance to health facility seen in risk factor analysis) we see that households in five villages in particular appear to have a higher mortality risk as a function of increased distance to the nearest local primary health care clinic (Figure 5). These specific areas are numbered in this figure.

Table 1: Bivariate analysis of risk factors for adult mortality by age groups 15-49 and 50-64 years, Agincourt sub-district, 1992-2008

	15-49			50-64		
Factors	n	HR (95% CI)	p-value	n	HR (95% CI)	p-value
Period α						
1992-1997	55798	Ref		4415	Ref	
1998-2003	24067	2.63 (2.45,2.81)	<0.001	1181	2.00 (1.72,2.33)	<0.001
2004-2008	20129	5.52 (4.93,6.19)	<0.001	1226	3.67 (2.77,4.86)	<0.001
Age α	99994	1.06 (1.06,1.07)	<0.001	6822	1.04 (1.03,1.05)	<0.001
Male gender	46188	1.19 (1.13,1.26)	<0.001	2760	1.83 (1.65,2.05)	<0.001
Mozambican nationality	31219	1.06 (1,1.13)	0.053	2415	0.77 (0.68,0.87)	<0.001
None or primary level maximum education	48531	Ref		6391	Ref	
Secondary or higher level maximum education	50183	0.54 (0.51-0.57)	<0.001	344	0.67(0.50-0.90)	0.007
Average months per years outside site	99233	1.08 (1.08,1.09)	<0001	6779	1.04 (1.03,1.06)	<0.001
Migrant (>=6 months per year on average outside site)	25397	1.74 (1.64,1.85)	<0.001	1254	1.55 (1.35,1.78)	<0.001
Household size	99747	0.97 (0.96,0.97)	<0.001	6797	0.99 (0.98,0.99)	<0.001
Household duration α	99994	0.89 (0.89,0.90)	<0.001	6821	0.92 (0.91,0.93)	<0.001
Number of other household deaths						
None	56492	Ref		4019	Ref	
1-4	42860	1.08 (1.02,1.14)	0.008	2776	1.02 (0.91,1.13)	0.777
5+	642	1.97 (1.56,2.49)	<0.001	27	1.04 (0.52,2.09)	0.912
Household head male	61937	0.63 (0.6,0.67)	<0.001	4054	0.75 (0.67,0.83)	<0.001
Household head age >=40years α	55541	0.82 (0.77,0.86)	<0.001	5390	0.32 (0.28,0.35)	<0.001
Household head Mozambican	32543	1.07 (1,1.13)	0.044	2397	0.76 (0.67,0.87)	<0.001
Household head death (excluding the individual)	3351	4.62 (4.28,4.99)	<0.001	402	3.58 (3.15,4.08)	<0.001
Maximum household MCA SES category						
Most poor	4861	Ref		452	Ref	
Poor	12790	0.55 (0.47,0.63)	<0.001	935	0.81 (0.62,1.06)	0.121
Least poor	32376	0.37 (0.32,0.42)	<0.001	1687	0.71 (0.56,0.91)	0.007
Straight-line distance to nearest health care facility (<5km)	69444	Ref		4471	Ref	
>= 5km	4690	1.07 (0.94,1.21)	0.305	280	1.17 (0.99,1.39)	0.064
Straight-line distance to nearest health care facility (<6km)	69581	Ref		3516	Ref	
>= 6km	4553	7.51 (4.04-13.99)	<0.001	1235	4.97 (0.70-35.34)	0.109
Average migrant months per year per village individual	99994	1.14 (1.03,1.25)	0.008	6822	0.97 (0.8,1.18)	0.769
Proportion of village deaths due HIV/TB (<17%) β	54903	Ref		3789	Ref	
>=17%	45091	1.12 (1.06,1.18)	<0.001	3033	0.99 (0.89,1.11)	0.912

α: at residence censorship start; β: median split

Table 2: Multivariate analysis of risk factors for younger adult mortality (15-49 years), Agincourt sub-district, 1992-2008

	Non-spatial model (Stata)		Non-spatial model (OpenBUGS α)		Spatial model (OpenBUGS α)	
Factors	HR (95% CI)	p-value	HR (95% CI)	signif β	HR (95% CI)	signif β
Period γ						
1992-1997	1.00		1.00		1.00	
1998-2003	2.60 (2.41-2.80)	<0.001	2.47 (2.30,2.65)	*	2.61(2.42,2.83)	*
2004-2008	4.81 (4.24-5.45)	<0.001	3.00 (2.64,3.43)	*	2.90 (2.52,3.32)	*
Age γ	1.07 (1.06-1.07)	<0.001	1.07 (1.06,1.07)	*	1.07 (1.06,1.07)	*
Male gender	1.48 (1.39-1.58)	<0.001	1.46 (1.37,1.56)	*	1.47 (1.38,1.56)	*
Mozambican	0.96 (0.89-1.04)	0.304	0.97 (0.90,1.04)		0.94 (0.87,1.02)	
Migrant δ	1.15 (1.08-1.23)	<0.001	1.18 (1.10,1.26)	*	1.17 (1.09,1.25)	*
Secondary or higher level education	0.90 (0.84-0.97)	0.003	0.95 (0.88,1.03)		0.98 (0.91,1.05)	
Household total	1.00 (0.99-1.00)	0.459	1.00 (1.00,1.01)	#	1.00 (1.00,1.01)	#
Number of other household deaths						
None	1.00		1.00		1.00	
1-4	1.19 (1.11-1.27)	<0.001	1.15 (1.08,1.23)	*	1.17 (1.09,1.25)	*
5+	1.78 (1.40-2.28)	<0.001	1.76 (1.37,2.19)	*	1.81 (1.39,2.30)	*
Household head death	4.38 (4.04-4.75)	<0.001	4.52 (4.16,4.91)	*	4.52 (4.16,4.91)	*
Maximum household SES category						
Most poor	1.00		1.00		1.00	
Poor	0.66 (0.57-0.77)	<0.001	0.80 (0.73,0.88)	*	0.76 (0.69,0.84)	*
Least poor	0.53 (0.46-0.61)	<0.001	0.69 (0.63,0.75)	*	0.65 (0.58,0.71)	*
Unknown	1.97 (1.73-2.25)	0.001	1.45 (1.25,1.67)	*	1.48 (1.26,1.71)	*
Household head male	0.51 (0.48-0.55)	<0.001	0.50 (0.47,0.53)	*	0.50 (0.47,0.53)	*
Household head age >=40years γ	0.63 (0.59-0.67)	<0.001	0.64 (0.60,0.68)	*	0.63 (0.59,0.67)	*
Distance nearest health facility (>= 6km)	5.17 (2.77-9.65)	<0.001	5.34 (3.11-9.98)	*	5.34 (3.11-9.98)	*
Average migrant months per year per village individual	0.97 (0.89-1.06)	0.510	0.96 (0.88-1.07)		0.95 (0.87-1.07)	
Proportion of village deaths due HIV/TB (<17%) β	1.00		1.00		1.00	
>=17%	1.04 (0.98-1.10)	0.221	1.07 (1.02,1.13)	*	1.92 (1.40,2.69)	*
Shape parameter	1.89 (1.84-1.93)	---	1.91 (1.87,1.94)	---	1.93 (1.86,1.99)	---
Individual variation ($\sigma^2\varepsilon$)	---	--	0.09 (0.07,0.12)	---	0.11 (0.08,0.14)	---
Spatial variation ($\sigma^2\varphi$)	---	---	---	---	2.21 (1.09,4.11)	---
AIC(Stata)/DIC(WinBUGS)	28360.7	---	49668.60	---	49576.00	---

α Burn in 5000, then run until MC error <5% of standard deviation for each covariate, thereafter run until sample of 10000 obtained

β * significant at 5% level, # significant at 10% level

γ: censorship start

δ: >=6 months per year on average outside site

ε: median split

Table 3: Multivariate analysis of risk factors for older adult mortality (50-64 years), Agincourt sub-district, 1992-2008

	Non-spatial model (Stata)		Non-spatial model (OpenBUGS α)		Spatial model (OpenBUGS α)	
Factors	HR (95% CI)	p-value	HR (95% CI)	signif β	HR (95% CI)	signif β
Period γ						
1992-1997	1.00		1.00		1.00	
1998-2003	2.59 (2.20-3.05)	<0.001	2.56 (2.15-3.00)	*	2.59 (2.16-3.08)	*
2004-2008	5.48 (4.06-7.40)	<0.001	4.37 (3.10-5.77)	*	4.07 (2.85-5.60)	*
Age γ	1.04 (1.03-1.06)	<0.001	1.04 (1.03-1.06)	*	1.04 (1.02-1.06)	*
Male gender	2.82 (2.49-3.20)	<0.001	2.98 (2.61-3.39)	*	3.03 (2.64-3.48)	*
Mozambican	0.62 (0.54-0.71)	<0.001	0.60 (0.51-0.69)	*	0.59 (0.50-0.70)	*
Migrant δ	1.23 (1.06-1.44)	0.007	1.24 (1.05-1.47)	*	1.23 (1.04-1.44)	*
Secondary or higher education	0.58 (0.43-0.78)	<0.001	0.55 (0.42-0.77)	*	0.56 (0.40-0.77)	*
Household head death	6.09 (5.26-7.05)	<0.001	6.64 (5.65-7.74)	*	6.74 (5.70-7.92)	*
Maximum household MCA SES category						
Most poor	1.00		1.00		1.00	
Poor	0.75 (0.57-0.99)	0.043	0.79 (0.57-1.06)		0.78 (0.58-1.05)	
Least poor	0.73 (0.56-0.95)	0.021	0.77 (0.56-1.05)		0.78 (0.58-1.03)	
Unknown	0.93 (0.71-1.21)	0.571	0.99 (0.71-1.33)		1.01 (0.75-1.35)	
Household head male	0.41 (0.36-0.46)	<0.001	0.40 (0.35-0.45)	*	0.40 (0.35-0.45)	*
Household head age >=40years γ	0.16 (0.15-0.19)	<0.001	0.15 (0.13-0.17)	*	0.15 (0.13-0.17)	*
Household size	1.00 (0.99-1.01)	0.893	1.00 (0.99-1.01)		1.00 (0.99-1.01)	
Shape parameter	1.87 (1.79-1.96)	---	1.89 (1.81-1.98)	---	1.92 (1.83-2.01)	---
Individual variation ($\sigma^2\varepsilon$)	---	---	0.13 (0.08-0.19)	---	0.16 (0.10-0.27)	---
Spatial variation ($\sigma^2\varphi$)	---	---	---	---	0.25 (0.11-0.56)	---
Range (meters)	---	---	---	---	18930 (2350-29610)	---
AIC/DIC	5403	---	10204	---	10193	---

α Burn in 5000, then run until MC error <5% of standard deviation for each covariate, thereafter run until sample of 10000 obtained
β * significant at 5% level, # significant at 10% level
γ: censorship start
δ: >=6 months per year on average outside site

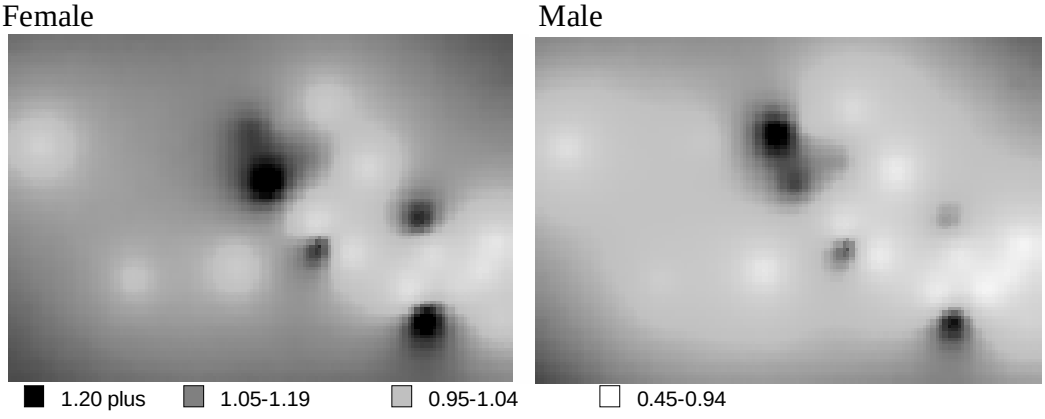


Figure 2: Spatial risk (HR) of gender specific adult mortality (15-64 years) based on prediction estimates from the Bayesian spatial kriging model, Agincourt sub-district, 1992-2008

15-49 50-64

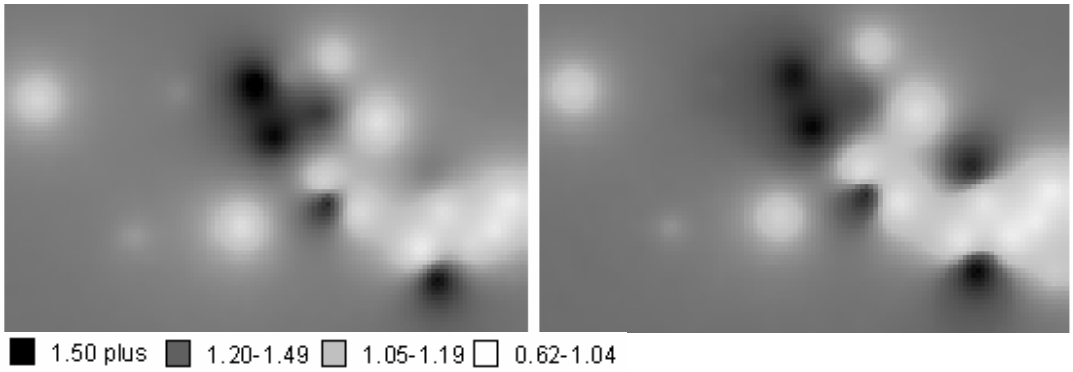


Figure 3: Spatial risk (HR) of age-specific adult mortality (15-49 and 50-64 years) based on prediction estimates from the Bayesian spatial kriging model, Agincourt sub-district, 1992-2008

Communicable (15-49) Non-communicable (15-64)

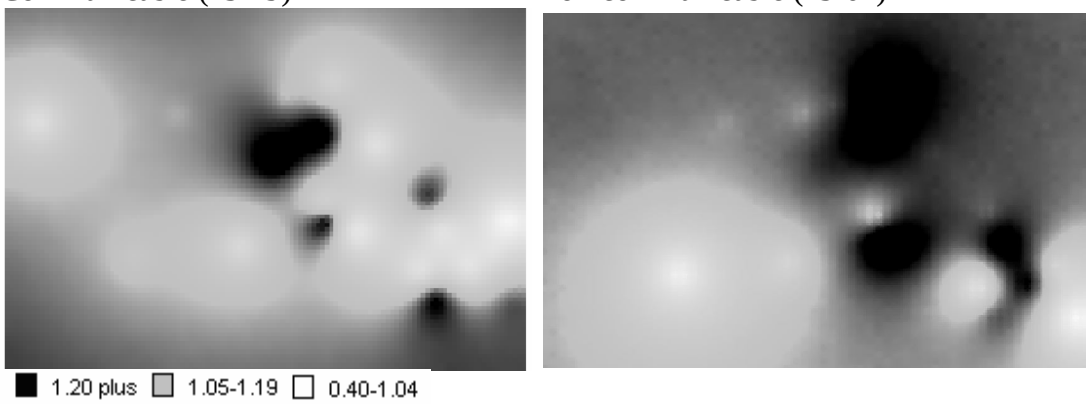


Figure 4: Spatial risk (HR) of broad cause-specific adult mortality by age-group based on prediction estimates from Bayesian spatial kriging models, Agincourt sub-district, 1992-2008

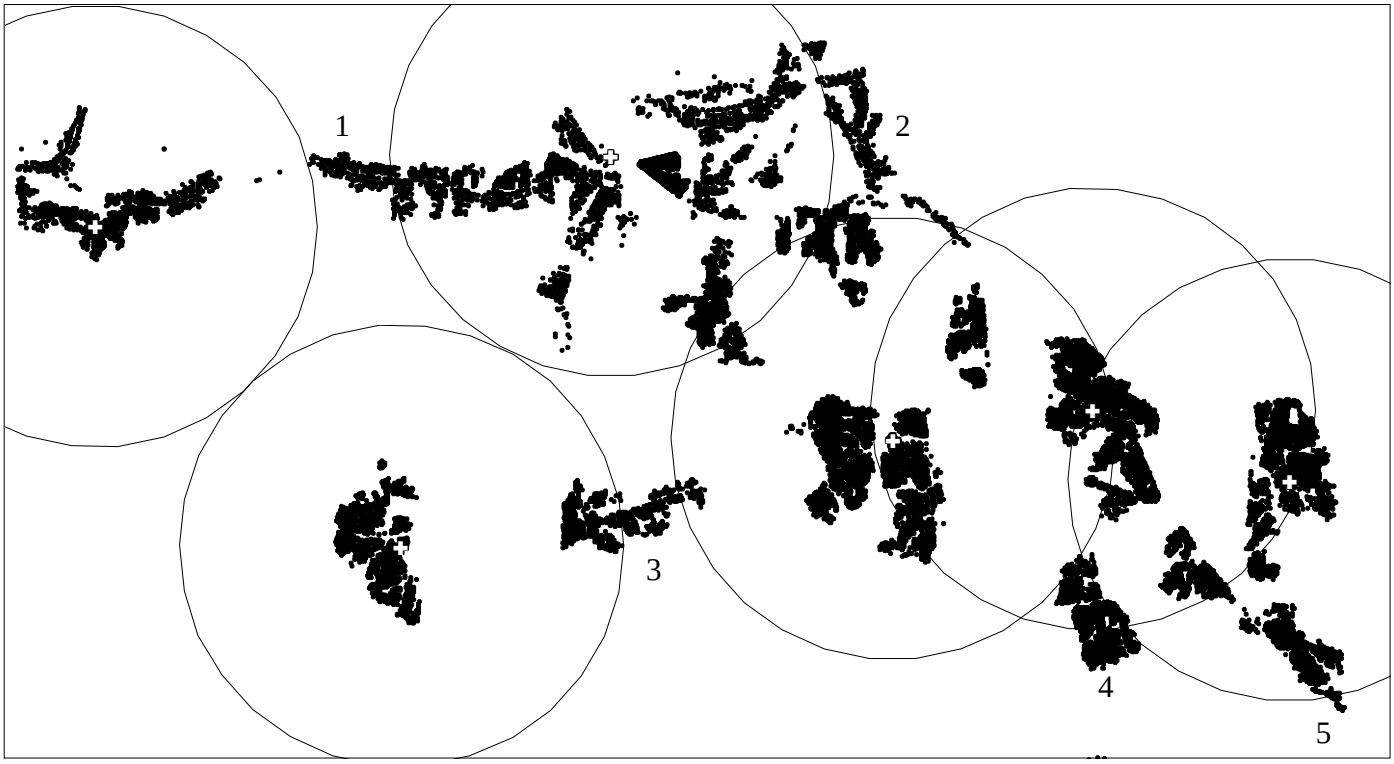


Figure 5: Household (black dots) and health facility (white cross) locations with concentric circles showing the 5km capture zone of these health facilities, Agincourt sub-district. Note: potentially higher risk household areas discussed above are numbered 1-5.

Discussion and conclusions

This study demonstrates the use of Bayesian geostatistical survival models in assessing risk factors and producing smooth maps of adult mortality risk in a health and socio-demographic surveillance system. Results showed strong geographical and temporal differences in adult mortality risk and indicates that these can vary in even a small geographical area. Some key differences in the determinant profile of identified mortality hotspots were also found such as lower age and higher migrancy in one particular village in the upper central region which has emerged as a focal point for HIV/TB in this rural sub-district and impacted surrounding villages.

A substantial and significant increase in adult mortality was observed in the Agincourt sub-district over the study period, in line with other areas of the country and nationally (36-37). Furthermore, the increase appeared to begin suddenly around 1999/2000 and levelled out around 2006. The additional deaths occurred in younger adult ages and were largely attributed to HIV/TB – which was the highest cause of death in both 15-49 and 50-64 year age-groups. This is similar to findings from rural Kwazulu Natal where increases in adult mortality in the late 1990's were determined by verbal autopsy, with the largest cause of death being HIV/TB (37). The levelling out of adult mortality could possibly be linked to ART rollout which began in South Africa round 2004. However the plateau in adult mortality in Agincourt preceded ART rollout which began in this area

only in 2007; before this, individuals would have accessed ART outside the Agincourt sub-district. Future studies will assess the long term impact of ART rollout on mortality.

Distinct spatial foci of increased adult mortality risk were observed in the upper central and south east regions of the Agincourt sub-district - specifically six villages displayed a higher mortality risk based on Bayesian kriging. These risk maps can be used by decision makers for the design and implementation of relevant interventions, in this case ART rollout. The spatial distribution is being driven by a complex web of interacting factors that include increased communicable disease mortality (HIV/TB) and non-communicable disease mortality (in the older age-group) in particular high risk areas.

Villages with higher HIV mortality burden (a proxy for prevalence), greater than 6km's from the nearest health facility, lower socio-economic status (SES), corresponding low levels of education, high household mortality burden, and high labour migration (spending on average more than 6 months per year away). Increased sexual risk behaviour and exposure to HIV appear to be driving communicable disease (HIV/TB) mortality and this has been documented in this area previously (38).

Poor SES and high unemployment have led to external migration for work while low education levels may have led to false perceptions of HIV exposure risk. - a previous study found that over 90 percent of men perceived little or no personal risk of HIV infection (38). This has potentially increased HIV transmission to long term migrants through increased number of partners while away, thus leading to increased transmission to their partners in Agincourt when they return, thus increasing the mortality burden among these households and villages. This in turn impacts these households as they have lost bread winners which further decreasing their SES.

The association between temporary migration and HIV has been shown previously in South Africa (39) with males more likely to be migrants than females. A previous study in Agincourt has shown that the level of reported risk behaviour among migrants varies with frequency of return to their rural home; those long-distance migrants who return once or twice a year report more partners than those who work in nearby destinations (38). However, resident employed men also reported more partners, indicating that increased risk may not be among long term migrants only. Collinson et al concluded that high levels of male labour migration, coupled with a low frequency of return home and low personal HIV risk perception, have led to an explosive spread of HIV in this and other rural settings (38). Strategies that enable more frequent contact between migrant men and their rural families - though very difficult to implement in reality - are urgently needed, as are HIV prevention and awareness raising activities amongst all men in this area (38).

Adult populations in mortality hotspots likely experienced differential access to health care as larger distance from the nearest health facility emerged as a significant risk factor. This has been shown in other studies on adult mortality (40) and specifically with regards to ART access in South Africa (41). This study found that with increasing distance from treatment site adults were significantly less likely to access ART and more likely to die. The finding that increased mortality risk was associated with distance to nearest health

facility has major implications for ART rollout; inequalities in service access need to be addressed. Bayesian kriging predictions based on distance from nearest health facility can highlight specific areas at increased adult mortality risk of adult mortality; this information can be used by policy makers.

Mozambican nationality was no longer a risk factor in the 15-49 year age-group following multivariate adjustment. Previous studies in this area have shown an increased risk of mortality among Mozambican children between the age of 1 and 5 (42-43). Thus is would appears this relative increased risk when compared to South Africans reduces as they enter early adulthood. This is likely due to the adjustment for lower education and SES in the model, more evident among Mozambicans, and to these settlements being further away from health facilities. The risk for adult mortality was higher in the poorest villages (defined by lower SES status from asset status module data) than in wealthier villages, suggesting that the benefits of improved health care are not evenly distributed throughout the study area. Hosegood et al suggest that poorer households trying to cope with adult deaths are more vulnerable to dissolution and migration (44).

Mozambicans in the older age group had a significantly lower risk of mortality compared to South Africans. Given the general lower SES standing of Mozambican households, this suggests that lifestyle-related mortality among more affluent South African households may be driving the observed risk of non-communicable disease-related mortality, that remains evident despite the pronounced impact of HIV/TB (9). Further research into the underlying reasons for lower risk among older adult Mozambicans may inform interventions to reduce risk more broadly

Mortality in all-age groups was higher in households headed by women and where the head was younger than 40 years. In contrast, our study reveals, adult mortality significantly higher for males than females with no difference in the spatial pattern of male and female mortality risk. A cycle of greater male deaths (migrancy and/or lifestyle related disease), leading to more female headed households is likely. Given that male household head conferred a survival advantage to adults and that male migrants are at increased risk, this could potentially compound adult mortality over time as household head dynamics change. This cycle coupled with the direct and indirect impacts of adult mortality on children and the elderly is leading to vulnerable households within this rural sub-district that should be targeted for social support. Female headed households also need to be targeted as they appear to be more vulnerable and at higher risk for child, adult and overall mortality.

Physician-coded verbal autopsies have known limitations, such as relying heavily on household recall of medical records and related information, which affects its applicability in low-resource settings (45). Misclassification could thus have occurred in our data, especially with regards to underestimating non-specific HIV/AIDS related and non-communicable disease. However, a previous validation study of the VA in Agincourt has shown that it performs well in this high HIV prevalence setting (46). This validation found that for HIV/TB (often combined most HDSS analysis of HIV related death due to the difficulty of disentangling deaths due to these two causes) the sensitivity, specificity

and PPV were all high at 78, 80 and 85% respectively. Other studies have also confirmed that VA data can be used to reasonably estimate the distribution of AIDS- and non-AIDS-related deaths even in a rural population with relatively low levels of education (37,47). One area that also needs to be addressed is how the tool could be modified to take into account the influence of HIV/AIDS when estimating the sensitivity for conditions such as diarrhoea, ALRI/pneumonia and malnutrition that are unrelated to HIV/AIDS.

The findings of this study suggest that even in a small geographic area, such as the Agincourt sub-district, we see spatial disparities with regards to mortality risk and that this appear to be related to differences in the underlying risk profile of these individuals, households and villages. The emerging trends are also likely to be impacted by a complex interaction of these factors including potential disparity with regards to health care access and ART. Health programmes need to take account of this when assessing and further planning the comprehensive plan to prevent and treat HIV/AIDS in this and other rural populations.

Acknowledgements

This work was supported by a PhD fellowship from the South African Centre for Epidemiological Modelling and Analysis (SACEMA), a National Research Foundation (NRF) Centre of Excellence, as well as research fellowships from the MRC/Wits Rural Public Health and Health Transitions Research Unit (Agincourt) through the Wellcome Trust (grant no. 069683/Z/02/Z), UK, and the Swiss-South Africa Joint Research Programme (project no. JRP IZLSZ3_122926). Additional funding was provided by a Wits Faculty of Health Sciences Medical Research Endowment Fund (MREF) (grant no. SARB000). The Agincourt HDSS and GIS was funded by the Wellcome Trust, UK (grant no. 058893/Z/99/A, 069683/Z/02/Z, 069683/Z/08/Z), the University of the Witwatersrand, the South African Medical Research Council, and the Andrew W. Mellon Foundation, USA. P. Vounatsou was supported by the Swiss National Science Foundation (project no. 325200_118379).

References

1. Rosero-Bixby L. Socioeconomic development, health interventions and mortality decline in Costa Rica. Scand J Soc Med Suppl 1991;46:33-42.
2. Timaeus I. Commentary: Mortality decline in Senegal. Int J Epidemiol. 2001;30:1294-5.
3. Martens P, Akin S, Maud H, Mohsin R. Is globalization healthy: a statistical indicator analysis of the impacts of globalization on health. Globalization and Health. 2010;6:16.
4. World Bank. Confronting AIDS: Public Priorities in a Global Epidemic. : World Bank Policy Research Report, Washington, D.C; 1997.
5. Beegle K, Filmer D, Stokes A, Tiererova L. Orphanhood and the Living Arrangements of Children in Sub-Saharan Africa. World Bank Policy Research Working Paper 4889. 2009.

6. Bradshaw D, Timaeus I, Dorrington R, Nannan N. Reversal in adult mortality trends in South Africa. International Conference on AIDS2000.
7. Kahn K, Garenne M, Collinson M, Tollman S. Mortality trends in a new South Africa: Hard to make a fresh start. Scand J Public Health Suppl. 2007;69:26-34.
8. WHO. Indicator definitions and metadata. WHO Statistical Information - System - WHOSIS 2008; Geneva. 2008.
9. Tollman S, Kahn K, Sartorius B, Collinson M, Clark S, Garenne M. Implications of mortality transition for primary health care in rural South Africa: a population-based surveillance study. Lancet. 2008;372(893–901).
10. Kaufman J, Asuzu M, Rotimi C, Johnson O, Owoaje E, Cooper R. The absence of adult mortality data for sub-Saharan Africa: a practical solution. Bull World Health Organ. 1997;75(5):389–95.
11. Hammer G, Somé F, Müller O, Kynast-Wolf G, Kouyaté B, Becher H. Pattern of cause-specific childhood mortality in a malaria endemic area of Burkina Faso. Malaria Journal. 2006;5:47.
12. Kuh D, Shlomo Y. A life course approach to chronic disease epidemiology: Oxford University Press; 2004.
13. Gething P, Noor A, Gikandi P, Ogara E, Hay S, Nixon M, et al. Improving imperfect data from health management information systems in Africa using space-time geostatistics. PLoS Med. 2006;3(6):e271.
14. Langford I, Leyland A, Rashbas J, Goldstein H. Multilevel modeling of the geographic distribution of diseases. J R Stat Soc Ser C Appl Stat. 1999;48:253-68.
15. Xia H, Carlin B, Waller L. Hierarchical models for mapping Ohio lung cancer rates. Environmetrics. 1997;8:107-20.
16. Walter S. Disease mapping: a historical perspective. Spatial Epidemiology – Methods and Applications. (Elliott P WJ, Best N, Briggs DJ, Eds), editor: Oxford University Press; 2000.
17. Gosoniu L, Vounatsou P, Sogoba N, Smith T. Bayesian modelling of geostatistical malaria risk data. Geospatial Health. 2006;1:126-39.
18. Kleinschmidt I. Spatial statistical analysis, modelling and mapping of malaria in Africa: University of Basel:; 2001.
19. Gemperli A. Development of spatial statistical methods for modelling point-referenced spatial data in malaria epidemiology: University of Basel; 2003.
20. Gemperli A, Vounatsou P, Kleinschmidt I, et al. Spatial patterns of infant mortality in Mali; the effect of malaria endemicity. Am J Epidemiol. 2004;159:64–72.
21. Kazembe L, Appleton C, Kleinschmidt I. Spatial analysis of the relationship between early childhood mortality and malaria endemicity in Malawi. Geospatial Health. 2007;2(1):41-50.
22. Gosoniu L, Vounatsou P, Tami A, Nathan R, Grundmann H, Lengeler C. Spatial effects of mosquito bednets on child mortality. BMC Public Health. 2008;8:356.
23. Tollman S, Herbst K, Garenne M, Gear J, Kahn K. The Agincourt demographic and health study--site description, baseline findings and implications. S Afr Med J. 1999;89(8):858-64.
24. Tollman S, Herbst K, Garenne M, Gear J, Kahn K. The Agincourt Demographic and Health Study: Site description, baseline findings and implications. South African Medical Journal. 1999;89(8):858-64.

25. Kahn K, Tollman S, Collinson M, Clark S, Twine R, Clarke B, et al. Research into health, population and social transitions in rural South Africa: Data and methods of the Agincourt Health and Demographic Surveillance System. Scand J Public Health. 2007;69(Suppl):8–20.
26. Collinson M, Tollman S, Kahn K, Clark S. Highly prevalent circular migration: Households, mobility and economic status in rural South Africa. Tienda M FS, Tollman S and Preston-Whyte E, editor. Johannesburg: Wits University Press; 2006.
27. Blasius J, Greenacre M. Correspondence analysis and related methods in practice. M. Greenacre JBE, editor. London: Chapman & Hall; 2006.
28. Booysen F, Van Der Berg S, Burger R, Von Maltitz M, Du Rand G. Using an asset index to assess trends in poverty in seven Sub-Saharan African countries. World Development. 2008;36(6):1113-30.
29. Howe L, Hargreaves J, Huttly S. Issues in the construction of wealth indices for the measurement of socio-economic position in low-income countries. Emerging Themes in Epidemiology. 2008;5:3.
30. Wikle C, Royle J. Spatial statistical modelling in biology. . Encyclopaedia of Life Support Systems: EOLSS Publishers Co. Ltd; 2002.
31. Diggle P, Moyeed R, Tawn J. Model-based Geostatistics (with discussion). Applied Statistics. 1998;48:299-350.
32. Spiegelhalter D, Thomas A, Best N. WinBUGS Version 1.2 User Manual MRC Biostatistics Unit. 1999.
33. Gelfand A, Smith A. Sampling-based approaches to calculating marginal densities. J Am Stat Association. 1990;85:398-409.
34. Spiegelhalter D, Best N, Carlin B, van der Linde A. Bayesian measures of model complexity and fit. J R Stat Soc B. 2002;64:583-639.
35. Gelfand A, Ravishanker N, Ecker M. Modeling and inference for point-referenced binary spatial data. Dey D. GS, Mallick B, editor: Marcel Dekker Inc; 1999.
36. Dorrington R, Bourne D, Bradshaw D, Laubscher R, Timæus I. The Impact of HIV/AIDS on Adult Mortality in South Africa2001.
37. Hosegood V, Vanneste A, Timaeus I. Levels and causes of adult mortality in rural South Africa: the impact of AIDS. AIDS. 2004;18(4):663-71.
38. Collinson M, Wolff B, Tollman S, Kahn K. Trends in internal labour migration from the rural Limpopo Province, male risk behaviour, and implications for spread of HIV/AIDS in rural South Africa. J Ethn Migr Stud. 2006;32(4):633-48.
39. Lurie M. Migration and AIDS in Southern Africa: a review. Southern African Journal of Science. 2000;96.
40. Ali M, Jin Y, Kim D, De Z, Park J, Ochiai R, et al. Spatial risk for gender-specific adult mortality in an area of southern China. Int J Health Geogr. 2007;6:31.
41. Ingle S, May M, Uebel K, Timmerman V, Kotze E, Bachmann M, et al. Differences in access and patient outcomes across antiretroviral treatment clinics in the Free State province: A prospective cohort study. S Afr Med J. 2010;100:675-81b.
42. Hargreaves J, Collinson M, Kahn K, Clark S, Tollman S. Childhood mortality among former Mozambican refugees and their hosts in rural South Africa. Int J Epidemiol. 2004;33:1271-8.

43. Sartorius B, Kahn K, Vounatsou P, Collinson M, Tollman S. Survived infancy but still vulnerable: Spatial-temporal trends and risk factors for child (1-4) mortality in rural South Africa (Agincourt), 1992-2007. Geospatial Health. 2011;5(2):285-95.
44. Hosegood V, McGrath N, Herbst K, Timaeus I. The impact of adult mortality on household dissolution and migration in rural South Africa. AIDS. 2004;18(11):1585-90.
45. Murray C, Lopez A, Feehan D, Peter S, Yang G. Validation of the symptom pattern method for analyzing verbal autopsy data. PLoS Med. 2007;4:e327.
46. Kahn K, Tollman S, Garenne M, Gear J. Validation and application of verbal autopsies in a rural area of South Africa. Trop Med Int Health. 2000;5(11):824-31.
47. Doctor H, Weinreb A. Estimation of AIDS adult mortality by verbal autopsy in rural Malawi. AIDS. 2003;17(17):2509-13.
48. Gelfand A, Vounatsou P. Proper multivariate conditional autoregressive models for spatial data analysis. Biostat. 2003;4:11-25.

Appendix 1: Multivariate Statistical Model

We analysed the data assuming a parametric Weibull distribution for the survivor function, where t_{ikj} is the failure time of an adult i (for censored observations the survival distribution is a truncated Weibull with an upper bound corresponding to the censoring time) for residence episode k at location j with covariate vector X_{ik} and β is a vector of unknown regression coefficients and including a village level spatial random effect w_j in the exponent of the hazard model as follows

$$t_{ikj} \sim \text{Weibull}(\rho,\mu_{ikj}) \quad i = 1,...,N;$$

with a baseline hazard function of the form

$$l_0(t_{ikj}) = \rho t_{ikj}^{r-1}$$

and means for the various models as follows

1) multivariate non-spatial model $\log(\mu_{ikj}) = \beta_0 + \beta X_{ik}$
2) mutlivariate spatial model $\log(\mu_{ikj}) = \beta_0 + \beta X_{ik} + w_j$
3) spatial kriging model $\log(\mu_{ikj}) = \beta_0 + w_j$

where β_0 is ~Normal(0,0.1) and w_j has a multivariate normal distribution, $w_j \sim MVN(\underline{0},\Sigma)$, with variance-covariance matrix Σ expressed as a parametric function of distance between pairs of the 25 village centroids points. We also assume an isotropic stationary spatial process, where $\Sigma mn = \sigma_w^2 \exp(-\varphi d_{mn})$, d_{mn} is the Euclidean distance between villages *m* and *n*, σ_w^2 is the geographical variability known as the sill, φ is a smoothing parameter that controls the rate of correlation decay with increasing distance and measures the range of geographical dependency. We specified φ as a uniform distribution between φ_{min} and φ_{max} (48). The range is defined as the minimum distance at which spatial correlation between locations is below 5%. This distance can be calculated as $3/u$

meters. A non-informative gamma prior was adopted for σ_w^2 with a mean and variance of 0.01. The regression coefficients were given non-informative normal priors, namely $\beta \sim \text{Normal}(0,0.1)$.

The shape parameter of the survival distribution (ρ) was given a non-informative gamma distribution with a mean and variance of 1.

MCMC simulation was applied to fit the models. We run a single chain sampler with a burn-in of 5000 iterations. Convergence was assessed by running the simulation until the Monte Carlo error for each parameter of interest was less than 5% of the sample standard deviation. The chains thereafter were sampled every single iteration until a sample size of 10 000 had been attained.