

Space and time clustering of mortality in rural South Africa (Agincourt HDSS), 1992–2007

Benn Sartorius^{1,4*}, Kathleen Kahn^{1,2,4}, Penelope Vounatsou³, Mark A. Collinson^{1,2,4} and Stephen M. Tollman^{1,2,4}

¹MRC/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; ²Centre for Global Health Research, Epidemiology and Global Health, University of Umeå, Umeå, Sweden; ³Swiss Tropical Institute, Basel, Switzerland; ⁴INDEPTH 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

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 km² 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

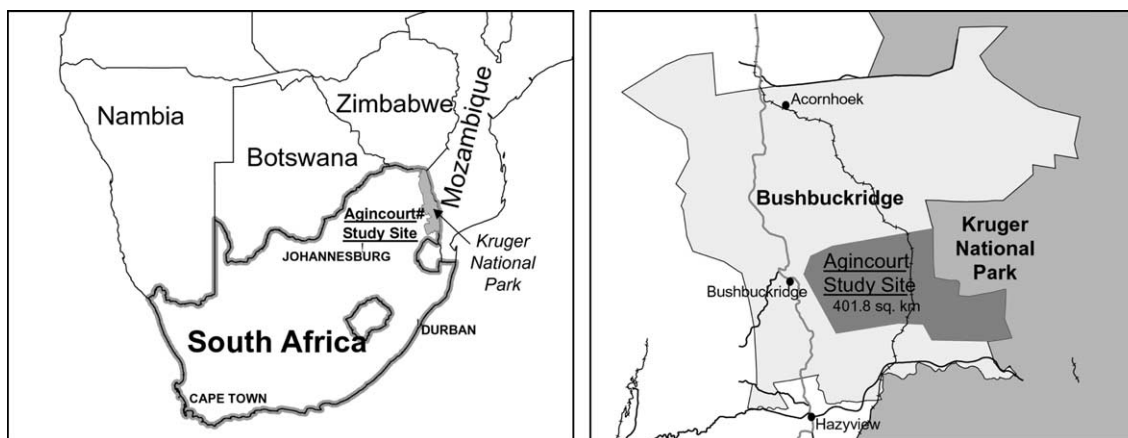


Fig. 1. Maps showing the regional location of the Agincourt health and socio-demographic surveillance site (Source: Kahn K et al. Research into health, population and social transitions in rural South Africa: data and methods of the Agincourt health and socio-demographic surveillance system. *Scand J 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 ($i=1, \dots, 21$) at year j ($j=1992, \dots, 2007$). 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). Kulldorff 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

Table 1. Crude mortality rates (per 1,000 person-years) by age group and gender, Agincourt sub-district, 1992–2007

Age group	Females				Males				Overall			
	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

Village	Total deaths	Person years	Death rate per 1,000 person years					Rate change (95%CI) α
			Overall (95% CI)	1992–1995	1996–1999	2000–2003	2004–2007	
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

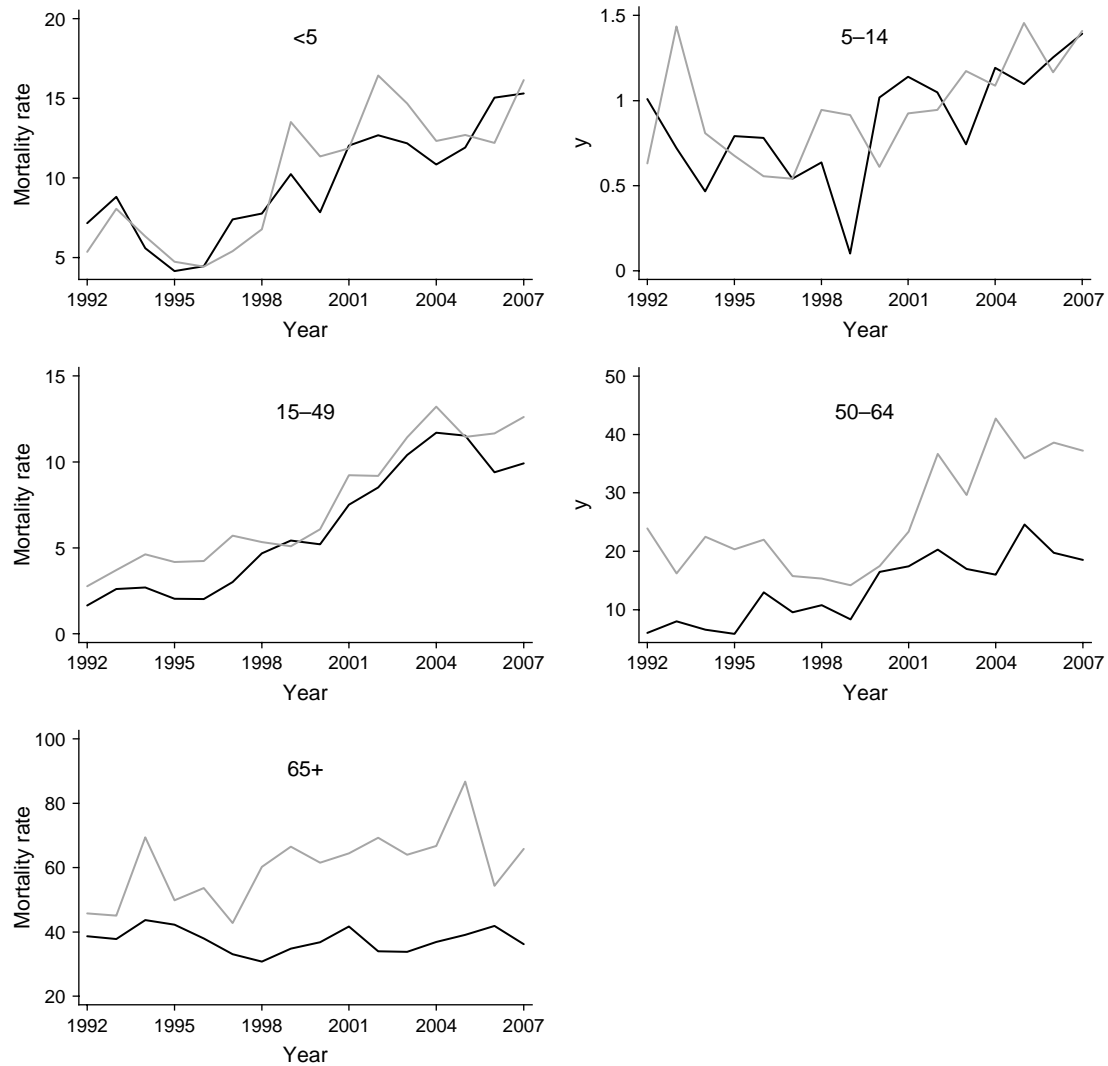


Fig. 2. Crude mortality rates (per 1,000 person years) by gender (female = black, male = light gray), age group and year, Agincourt sub-district, 1992–2007.

(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)	<i>p</i> -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)	<i>p</i> -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

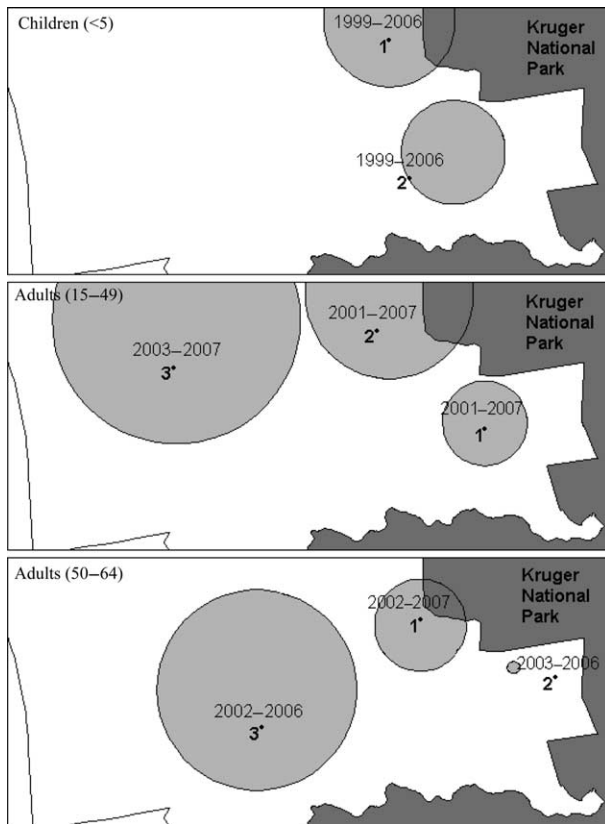


Fig. 3. Significant clusters of all-cause mortality by age group in Agincourt sub-district using space-time analysis (Note: Figure shows administrative boundary of lower Bushbuckridge municipality; 1 = most likely cluster, 2 = secondary cluster, 3 = tertiary cluster).

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 K, 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. Hjalmar U, Kulldorff M, Gustafsson G. Risk of acute childhood leukaemia in Sweden after the Chernobyl reactor accident. *BMJ* 1994; 309: 154–7.
13. Hjalmar 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. Hjalmar 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 north-east. *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