Published: November 2011

Joint Mapping Modelling for Multiple Health Problems in South Africa

Samuel Manda - Senior specialist statistician, Biostatistics Unit, South African Medical Research Council

Analysing multiple disease outcomes

Disease mapping models are used in spatial epidemiological studies to investigate the causes and distributions of diseases. An ecologic investigation is usually adopted to assess disease risk in relation to risk exposure factors measured in a small area such as an electoral ward. When the spatial data have been collected over many time periods, temporal trends are included. The models often incorporate expected disease incidence based on age and gender stratification and account for dependence of disease incidence on contextual area factors such as deprivation and population density. For small areas or rare diseases such as cancer, the usual estimates of disease risk may be unreliable and unstable. This has necessitated identification of models that stabilise risk estimates in small geographic areas by using hierarchical random effects models (1).

In a standard univariate spatial model, the underlying area-specific disease risk is split into two parts; one that has a global mean relating to all the areas *(heterogeneity or unstructured area effect*) and the other that has a mean depending on neighbouring areas' risks and has a variance which is inversely proportional to the number of neighbours (*local or structured area effect*) (2). Recently, there has been considerable methodological and application research on modelling and analysing multiple disease outcomes using joint mapping models. These extensions enable analysts to make an assessment on similarities as well as differences between risk factors among diseases purportedly sharing common risk profiles (3-4). These models, by combining data from different diseases, improve precision and efficiency of estimates, especially for rare diseases.

Most of the studies looking at mapping of health problems in the Sub-Saharan African region have concentrated on using univariate spatial models (5-7). Thus, there is a need to use and apply joint mapping models to measure co-morbidities of common illnesses in the region. A few studies measuring co-morbidity of diseases by using multivariate spatial modelling within the region have recently appeared (8-9). This article aims to show the utility of joint mapping models in estimating co-morbidities in two important health problems in South Africa.

The Shared-Spatial Component Model: Theory

An analyst can use a multivariate normal model to assess covariances and correlations within and between diseases underlying spatial risks (10). However, we use a shared component model which fits common and disease-specific unobserved and unmeasured spatial risks (3). The relative risk of each disease depends on a latent spatial component shared by all the diseases under study and the respective disease-specific latent component. For instance, if we have prevalence data on two diseases, then the risk for each disease is modelled as:

$$\log\left(\frac{P_{i1}}{1-P_{i1}}\right) = \alpha_{i0} + X_i\beta_1 + k\theta_i + \omega_{i1}$$
$$\log\left(\frac{P_{i2}}{1-P_{i2}}\right) = \alpha_{i0} + X_i\beta_2 + {}^{\theta_i}/_{k} + \omega_{i2}$$

where P_{ij} is the prevalence rate of disease j in area i; θ_i is the shared component, common to both diseases; ω_{i1} and $\omega_{i2}$are the disease-specific spatial risk components, respectively. The unknown parameter $k > 0$ is included to allow for a differential gradient for the shared component for the two diseases. The ratio of the scaling parameters k to $1/k$ compares the weight of disease one relative to disease two associated with the shared component. Usually the spatial components are modelled using both local dependence structures as well as unstructured heterogeneous effects to capture possible extra-variation in the data not captured by all the terms in the model affecting the risks.

Examples of Multiple Health Outcomes in South Africa

We illustrate the application of a shared-component model to two multiple health outcomes in South Africa. The first health problem concerns HIV and Syphilis prevalence data among pregnant women attending public antenatal clinics (ANC) in South Africa between 2007 and 2009 (11). The second concerns four vascular diseases: high blood pressure; heart attack or angina; stroke and high blood cholesterol available from the 2003 South African Demographic and Health Survey (12). The four chronic diseases are becoming a growing public health problem associated with lifestyle (smoking, alcohol, lack of physical activity) and dietary patterns. We measure distribution of disease at the health-district level. A district is the basis unit through which the delivery of Primary Health Care is undertaken in South Africa.

The number of ANC attendees that were sampled per district ranged from 51 to 2627. The prevalence of HIV and Syphilis ranged from 0 to 46.4% and from 0 to 12.6%, respectively. This shows great variation in both HIV and Syphilis prevalence

between the districts. However, some of the estimates, especially for Syphilis, were based on small sample sizes. Observed geographical trends indicated that HIV prevalence was highest in the North-Eastern parts of the country and lowest in the South-Western parts. Lower HIV prevalence rates were found in the least populated and most rural areas compared to the metropolitan areas. Syphilis prevalence shows contrasting patterns to those observed for HIV prevalence, with high rates in the South-Western parts of the country.

For the four vascular diseases, the number of sampled adults in a health district ranged from 10 to 900. High blood pressure and heart rate were more prevalent than stroke and high blood cholesterol. There were 15 districts out of the 52 that had a prevalence of 0% for stroke, and 13 out of 52 had a prevalence of 0% for high blood cholesterol. High prevalence of high blood pressure and heart attacks are in the districts in the South-Western parts of the country and the lowest in the North-Eastern parts. For stroke and high blood cholesterol, prevalence rates appear to be relatively evenly distributed across the country.

The shared-spatial component model: practice

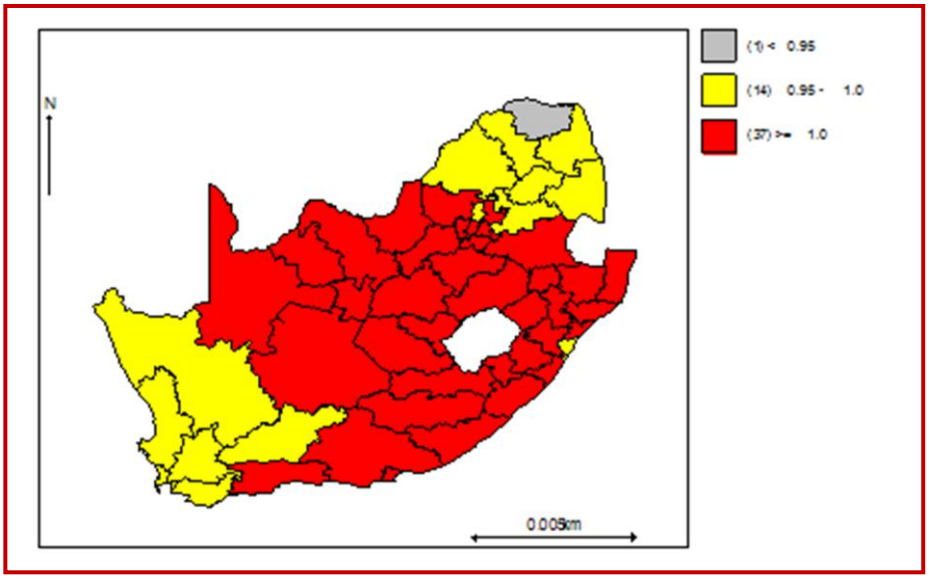
The observed prevalence maps (not shown) showed a large amount of variation especially for the rare diseases, which makes it difficult to discern any geographical trends in prevalence rates. Thus, it was necessary to use spatial models to smooth out the instability in the risk estimates. For both sets of multiple health outcomes, a shared component model (1) was fitted. In the case of HIV and Syphilis three spatial risk effects were used: a shared component that can be interpreted as representing risky sexual behaviours (e.g. multiple and concurrent sexual partners and unprotected sex) and two disease-specific components representing unmeasured risk factors associated with the individual diseases. In addition the effects of two contextual district level factors, social and material deprivation and population density, were measured. For the four vascular diseases, a shared component that can be interpreted as representing nutritional and lifestyle factors and four disease-specific risk components were used. In addition, individual subject risk factors were controlled for in the vascular diseases models. The district-level effects that remained after controlling for the effects of the included observed and measured risk factors were assessed on the disease-odds scale where values above 1 indicate increased prevalence in the corresponding district.

Table 1 shows the effects of deprivation and population density on the prevalence of HIV and Syphilis. For each disease, the effects shown are the odds of disease prevalence for different levels of the factor compared to the odds of the prevalence in the reference category (lowest), whereby a ratio value higher or lower than 1 indicates that the odds of disease prevalence increases or decreases compared to the reference category. The values in brackets are ranges of the ratio of the odds, indicating a significant difference between the odds if the range does not include 1. Thus, high values of deprivation and population density are associated with increasing HIV prevalence. On the other hand, Syphilis prevalence is inversely associated with both deprivation and population density. This shows that HIV and Syphilis have contrasting dependence on the two contextual factors.

Table 1: Estimates from a joint model for antenatal HIV and Syphilis Prevalence, South Africa, 2007-2009

Covariate effects		**HIV Prevalence**	**Syphilis Prevalence**
Yearly Prevalence (%) (median and 95% CI)	2007	13.76 (8.97 – 19.83)	7.76 (4.57 – 13.18)
	2008	13.75 (8.93 – 19.80)	5.40 (3.16 – 9.31)
	2009	13.79 (8.96 – 19.82)	5.34 (3.10 – 9.21)
Deprivation (OR and 95% CI)	I (Lowest)	1.00 (-,-)	1.00 (-,-)
	II	1.54 (1.08 – 2.34)	0.79 (0.43 – 1.34)
	III	1.98 (1.27 – 3.36)	0.52 (0.27 – 0.93)
	IV	1.59 (1.11 – 2.47)	0.32 (0.17 – 0.60)
	V (Highest)	1.80 (1.10 – 3.12)	0.28 (0.15 – 0.56)
Population Density (OR and 95% CI)	I (Lowest)	1.00 (-,-)	1.00 (-,-)
	II	1.59 (1.04 - 2.17)	0.67 (0.40 - 1.12)
	III	1.48 (0.99 – 2.19)	0.49 (0.27 – 0.92)
	IV	1.53 (1.01 – 2.10)	0.41 (0.23 – 0.72)
	V (Highest)	1.97 (1.25 – 2.92)	0.45 (0.24 - 0.80)

Figure 1: Shared risk component for HIV and Syphlis Prevalence



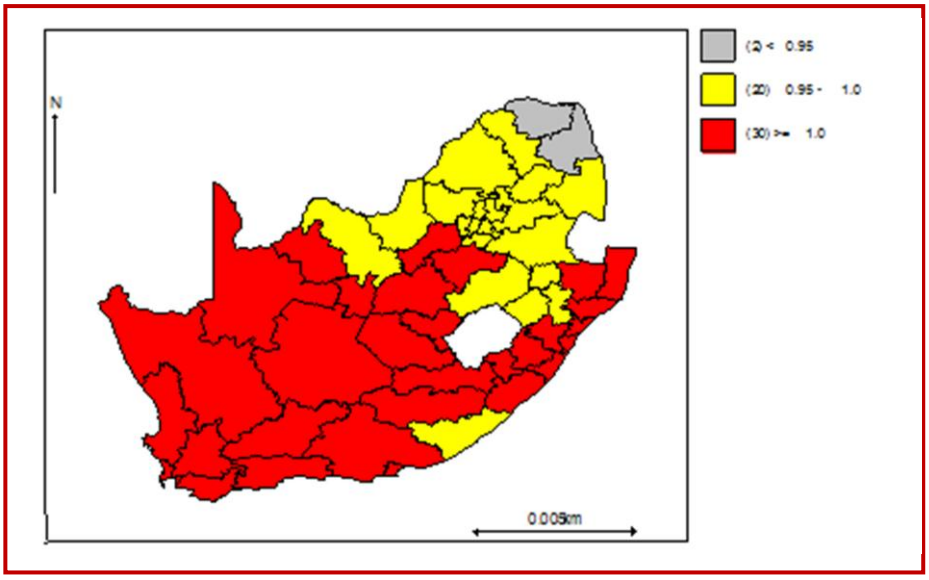
HIV and Syphilis specific-excess risk maps (after measuring the effects of important contextual predictor variables) showed high HIV rates in North-Eastern parts of the country, and high Syphilis in the districts around the South-Western corridor (maps not shown). The shared component (representing risky sexual behaviour) distribution shown in Figure 1 has a larger effect on HIV and Syphilis in the North-Western to South-Eastern corridor of the country.

The results from estimating co-morbidly of the four vascular diseases are presented in Table 2. Following the same interpretation of significance as before, results show that increasing age is positively associated with increased risks of all of the four vascular diseases. Being obese significantly increases the risk for all of the four vascular diseases. Lifestyle factors such as smoking and drinking alcohol appear not to have any adverse effects on the vascular diseases. Other factors (gender, education, population group, urban setting) affected the risks in non-systematic ways (data not shown).

Table 2: Median (95% CI) Odds Ratios estimates from a joint model for four high blood pressure, heart attack, stroke and high blood cholesterol, South Africa, 1998.

	High Blood Pressure	**Heart Attack**	**Stroke**	**High Blood Cholesterol**
Age				
14 - 24 years	1.00	1.00	1.00	1.00
25 - 34 years	2.21 (1.69-2.86)	2.99 (2.03-4.44)	4.84 (1.77-27.95)	2.70 (1.35-6.10)
35 - 44 years	4.45 (3.46-5.76)	3.55(2.44-5.38)	6.13(2.17-33.87)	2.52 (1.22-5.45)
44 – 54 years	10.83 (8.56-13.91)	5.92 (4.11-9.03)	10.33 (3.89-62.76)	5.36 (0.19-11.54)
55 – 64 years	17.35 (13.5-22.4)	9.37 (6.52-14.17)	13.75(4.76-70.09)	4.70(2.36-10.19)
>= 65 years	22.49 (17.58-29.23)	11.04 (7.54-16.54)	14.41 (5.21-80.4)	5.07 (2.48-11.55)
Body Mass Index				
Underweight	0.77 (0.58-1.01)	1.26 (0.90-1.75)	1.03 (0.40-2.26)	0.36 (0.05-1.36)
Normal	1.00	1.00	1.00	1.00
Obese	2.18 (1.93-2.46)	1.21 (1.00-1.46)	1.48 (0.97-2.32)	3.97 (2.58-6.33)
Current smoker (Yes=1, No=0)	0.83 (0.72-0.97)	0.96 (0.74-1.21)	1.31 (0.80-2.10)	0.95 (0.64-1.41)
Current alcohol drinker (Yes=1, No=0)	1.09 (0.96-1.25)	0.94 (0.76-1.16)	0.66 (0.40-1.03)	0.91 (0.61-1.33)

Figure 2: Shared risk component for high blood pressure, heart attack, stroke and high blood cholesterol prevalence



Vascular disease-specific maps (not shown) of excess risks (after measuring the effects of important subject-level predictor variables) showed that high blood pressure and stroke rates were concentrated highly in the South-Western parts of the country. Heart attack was highly concentrated in the central North-Eastern corridor; and high blood cholesterol had high rates in the top North-Eastern corridor. The distribution of the shared component (representing nutrition and lifestyle) shown in Figure 2 has a larger effect on vascular disease prevalence in the South-Western areas of the country.

Better understanding of co-morbidity

This article has demonstrated the application of recently developed methodology and estimation techniques in spatial epidemiology to model multiple health outcomes. In particular, we used the shared component model to assess common and divergent putative risk factors in the context of the burden of multiple sexually transmitted and multiple vascular diseases in South Africa. The results have shown that HIV and Syphilis have largely divergent risk and spatial factors. The common risk factors are mainly concentrated in the North-Western to South-Eastern corridor. This may suggest that interventions aimed at modifying risky sexual behaviours in this corridor will have greater impact on reducing the burden of HIV and Syphilis. Increasing age and obesity were significantly positively associated with all of the four vascular diseases. The common risk factors were more concentrated in the South-Western parts, implying that lifestyle modifications will make greater impact in reducing the burden of diseases in these areas.

In conclusion, multivariate mapping models provide a better understanding of co-morbidity between health outcomes than using separate univariate models. In particular, the modeller and the analyst of multiple disease outcomes can assess the underlying common and divergent spatial distributions of the diseases to optimally integrate disease management required to address the multiple burden of diseases in South Africa and the Sub-Saharan African region.

Samuel Manda, Senior specialist statistician, Biostatistics Unit, South African Medical Research Council. Areas of interest: the methodological and application biostatistics, covering the EM algorithm estimation, Bayesian statistics, spatial statistics and survival modelling. *samuel.manda@mrc.ac.za*

References:

1. Clayton DG, Kaldor J. Empirical Bayes estimates of age-standardized relative risks for use in disease mapping. Biometrics. 1987: 43(3): 671–681.
2. Besag J, York J, Mollie A. Bayesian image restoration, with two applications in spatial statistics. Ann Inst Statist Math. 1991:43(1): 1-59.
3. Knorr-Held L, Best NG. A shared component model for detecting joint and selective clustering of two diseases. JR Statist Soc A. 2001;164(1): 73-85.
4. Manda SOM, Feltbower RG, Gilthorpe MS. Investigating spatio-temporal similarities in the epidemiology of childhood leukaemia and diabetes. Eur J Epidemiol. 2009: 24(12):743-52.
5. Gemperli A, Vounatsou P, Kleinschmidt I, et al. Spatial patterns of infant mortality in Mali: the effect of Malaria endemicity. American Journal of Epidemiology. 2004;159(1): 64-72.
6. Kandala N-B, Ghilagaber G. A Geo-Additive Bayesian Discrete-Time Survival Model and its Application to Spatial Analysis of Childhood Mortality in Malawi. Quality & Quantity. 2006; 40(6):935–957.
7. Kleinschmidt I, Ramkissoon A, Morris N, et al. Mapping indicators of sexually transmitted infection services in the South Africa public health sector. Trop Med Int Health. 2006; 11(7): 1047–1057.
8. Kazembe LN, Mvula AS, Simoonga C. Joint spatial modelling of common morbidities of childhood fever

and diarrhoea in Malawi. Health & Place. 2009;15(1):165-172.
9. Kandala, N-B, Manda SOM. Assessing Geographic Co-morbidity Associated with Vascular Diseases in South Africa: a joint Bayesian Modeling Approach. In: Ngianga-Bakwin Kandala, Khaled Khatab (eds). Advanced Techniques for Modelling Maternal and Child Health in Africa. Springer (Forthcoming).
10. Manda SOM, Leyland A. An empirical comparison of maximum likelihood and Bayesian estimation methods for multivariate spatial disease model. South African Statistical Journal. 2007; 41: 1-21.
11. National Department of Health. National Antenatal Sentinel HIV and Syphilis Prevalence Survey in South Africa 2008. Pretoria: Department of Health; 2009.
12. National Department of Health, Medical Research Council & Measure DHS. South Africa Demographic and Health Survey 2003. Pretoria: National Department of Health; 2005.