Bayesian approach in Estimating Risk Determinants of Infectious diseases

Thembile Mzolo

University of KwaZulu Natal[1] **& Human Sciences Research Council**[2]
[1]School of Statistics and Actuarial Sciences
[2]Social Aspects of HIV/AIDS and Health

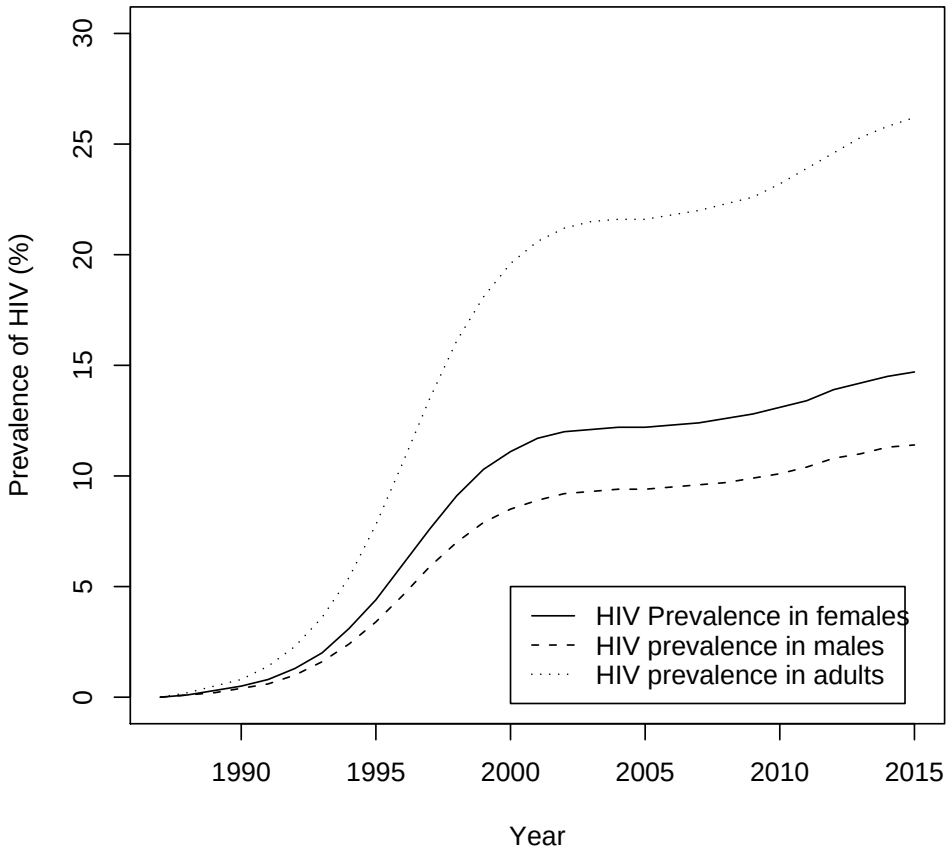
51^{st} Annual SASA Conference, October 2008

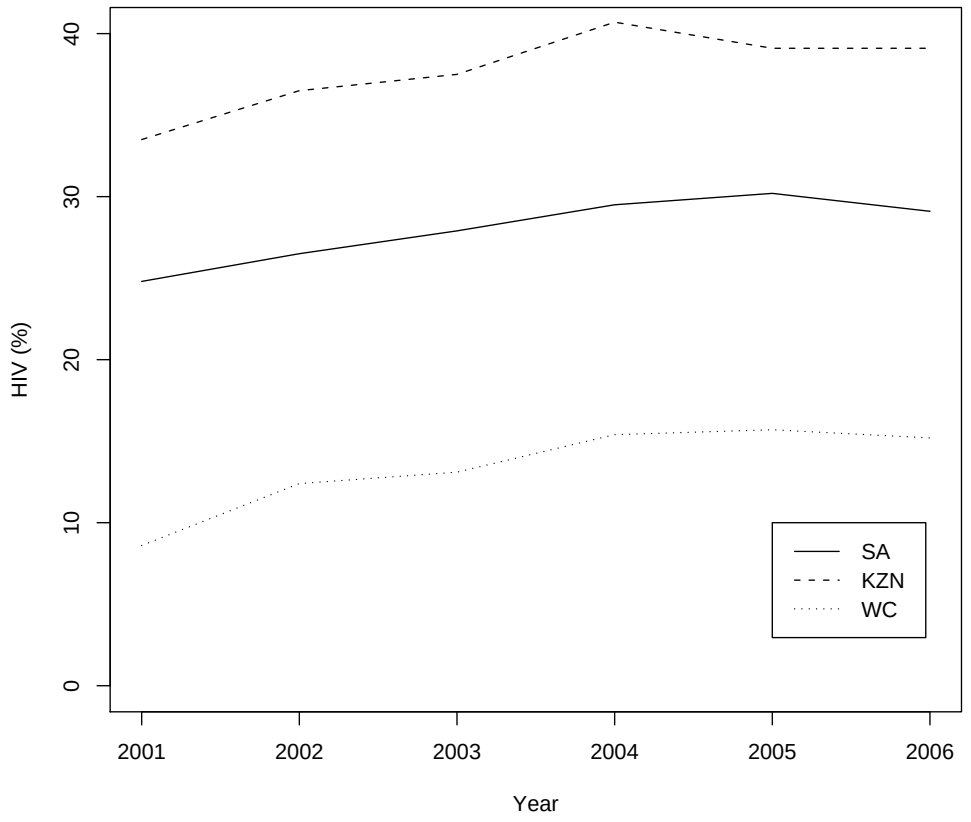
Outline

HIV and TB

- People who are co-infected with both HIV and TB are at increased risk of dying from TB (Bucher 1999; Corbett 2003).
- The burden of TB in countries with high rates of HIV has increased rapidly over the past decade.
- In KZN about 80% of TB patients are also infected with HIV (Gandhi 2006).
- SA has has the highest number of people living with HIV/AIDS in the world (UNAIDS, 2007).
- In 2006 it was reported that KZN has the highest HIV prevalence in SA (DOH, 2006).
- Infection levels vary among provinces and by gender.

HIV status by gender and province





Incidence estimates of TB in South Africa (Weyer 2004)

Province	Incidence	Total cases	% of TB cases	% of HIV cases
Western Cape	1333	58577	2.28	50.4
Northern Cape	822	8033	10.4	52.0
Eastern Cape	1307	102152	1.3	58.8
KwaZulu Natal	1696	173944	0.9	83.4
Limpopo	647	41108	1.6	55.1
Mpumalanga	1052	35977	2.9	77.9
Free State	871	29790	2.9	70.5
Gauteng	1034	85855	1.2	63.6
North West	754	29472	2.6	64.3
SOUTH AFRICA	1084	529320	2.9	66.4

Objectives

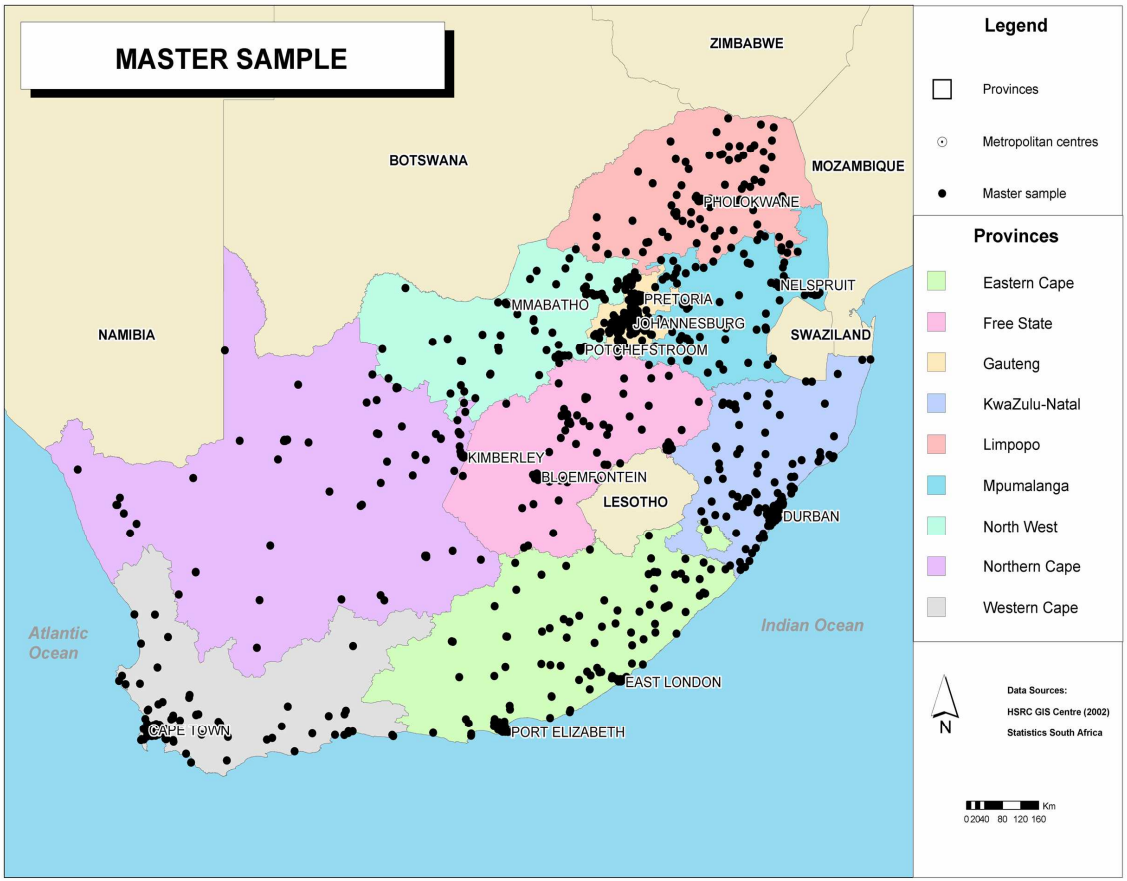
In South Africa, no coherent analysis of the determinants of HIV and TB has been conducted at a national level and thus this study seeks to mend this gab.

- Determine the risk determinants of HIV and those of TB.
- Identify risk factors that are common to both epidemics.
- Analyse the data in the study to assess the significance of the various potential risk determinants by means of a series of statistical methods in increasing complexity.
- Incorporate heterogeneity between geographic areas in modelling by means of random effects models and Bayesian hierarchical modelling approaches.

Data description

- Data used is from a household based second-generation surveillance survey of HIV conducted by HSRC in 2005.
- The survey design applied a multi-stage disproportionate, stratified sampling approach based on a master sample of 1000 EAs.
- The sample was stratified by province and locality type of the EAs whereas in urban areas race was used as a third stratification variable.
- The master sample allowed for reporting of results at the level of province, type of locality, age and race group.

Data descrip. *cont*



- Sample weights were introduced.
- Individuals were asked for DBS for HIV testing.
- Data consists of 13 422 households & 10 584 participated in the study.
- 23 275 people aged 2^+ completed interviews & 15 581 were tested for HIV.

Descriptive analysis

Socio-demographic factors

Parameter	Total	HIV+(%)	TB+(%)
Sex of the respondent			
male	6338	10.72	2.55
female	10057	16.67	1.66
Race			
African	9664	17.67	2.43
White	1913	0.52	0.38
Coloured	3013	2.84	1.35
Indian	1772	1.35	0.97
Age			
15 to 24	5708	10.28	1.15
25 to 34	2688	23.96	2.14
35 to 44	2928	18.57	2.70
45 to 54	2375	10.06	2.80
55+	2696	4.03	2.44
Health			
good	12998	13.22	1.03
poor	3086	16.56	6.70
Marital status			
never	7540	16.18	2.00
ever	8571	11.67	2.13
migration			
yes	1563	15.29	2.67
no	14625	13.62	2.00

Parameter	Total	HIV+(%)	TB+(%)
Geotype			
urban formal	9523	11.51	1.83
urban informal	1734	23.72	2.55
rural formal	3710	15.25	2.31
rural informal	1431	12.29	1.93
Education			
primary	4592	14.54	3.58
secondary	10027	14.85	1.58
tertiary	1585	5.53	0.26
Income			
unemployed	5828	19.91	2.74
employed	5168	12.82	1.09
other	5148	7.00	2.25

Bayesian Hierarchical models

- The data is clustered at an EA level.
- By controlling for both fixed and random risk factors we will be able to quantify any excess association between HIV & TB.
- Bayesian methods require prior information to estimate the posterior distribution.
- These methods involve integrating high-dimensional functions.
- Our focus is on the MCMC methods of simulating data.
- The roots of the MCMC methods come from the Metropolis Algorithm (Metropolis & Ulam 1949; Metroplis 1953).
- The Gibbs sampler (Geman & Geman 1984) is a MCMC method that is widely applicable.

Bayesian *cont.*
Gibbs sampling

- The Gibbs sampler is a special case of the Metropolis Algorithm.
- Gibbs sampler is also known as an *alternating conditional sampling* (Gelman 1998).
- In the Gibbs sampler one needs only to consider the univariate conditional distributions.
- These conditional distributions have simple forms and are easier to simulate than complex joint distributions.
- Consider a bivariate random variable (x, y) & suppose we wish to compute one or both marginals, $p(x)$ & $p(y)$.

Bayesian *cont.*
Gibbs sampling

- The sampler starts with some initial value y_0 for y & x_0 for x by generating random variable from the conditional distribution $p(x|y = y_0)$.
- The sampler uses x_0 to generate a new value of y_1. The sampler proceeds as follows:

$$x_i \sim p(x|y = y_{i-1})$$
$$y_i \sim p(y|x = x_i)$$

- This process is repeated k times, generating a Gibbs sampler of length k.

Bayesian *cont.*
Gibbs sampling

- In order to get the desired total m sample points, one samples the chain:
  - after a sufficient burn-in to remove the effects of the initial sampling values.
  - at a set time points (say, every n samples) following the burn-in.
- The Gibbs sampler converges to a stationary distribution which is a target distribution we are trying to simulate (Tierney 1994).
- Gibbs sampler usually produces chains with smaller autocorrelations than other MCMC samplers (Draper 2000).
- Stationarity formal tests that can be used are Geyer (1992), Geweke (1992), Raftery and Lewis (1992b), etc.

HIV Results

- For HIV model these variables were used: sex, age, race, education, health & condom use at sexual debut.
- For TB model these variables were used: HIV status, sex, education, income and health status.
- Priors for fixed effects were assumed multivariate normal centered at zero.
- Priors for random effects (EA) were assumed to follow a normal distribution.
- The burn-in period of 2000 iterations were allowed for both models.
- An estimated intraclass correlations for HIV and TB are $\rho_{HIV} = 0.169$ and $\rho_{TB} = 0.249$, respectively.

Table: HIV Results

Parameter	Bayesian results Estimate (SE)	GLMMs results Estimate (SE)
Intercept	-6.177 (0.404)	-5.813 (0.404)
Sex of the respondent		
Male	-0.376 (0.072)	-0.346 (0.074)
Female	0	0
Age of respondent		
15 to 24	1.861 (0.160)	1.825 (0.176)
25 to 34	2.438 (0.157)	2.379 (0.168)
35 to 44	1.987 (0.151)	1.943 (0.165)
45 to 54	1.046 (0.164)	1.030 (0.179)
55+	0	0
Race		
African	2.857 (0.360)	2.063 (0.342)
White	0.589* (0.457)	0.312* (0.468)
Coloured	1.080 (0.374)	0.894 (0.363)
Indian	0	0
Education		
None/primary	0.803 (0.164)	0.749 (0.175)
Secondary/matric	0.721 (0.156)	0.715 (0.168)
Tertiary	0	0
Health status		
Good	-0.623 (0.083)	-0.645 (0.086)
Poor	0	0
Condom use at sexual debut		
Yes	-0.358 (0.099)	-0.359 (0.104)
No	0	0

* p-value$>$0.05

Table: TB Results

	Bayesian results	GLMMs results
Parameter	Estimate (SE)	Estimate (SE)
Intercept	-4.098 (0.822)	-3.797 (0.814)
HIV status		
Negative	-1.430 (0.167)	-1.445 (0.172)
Positive	0	0
Sex of the respondent		
Male	0.837 (0.161)	0.863 (0.162)
Female	0	0
Education		
None/primary	2.205 (0.797)	2.0593 (0.789)
Secondary/matric	1.534 (0.796)	1.388* (0.789)
Tertiary	0	0
Income		
Unemployed	0.084* (0.192)	0.188* (0.181)
Employed	-0.789 (0.251)	-0.571 (0.243)
other	0	0
Health status		
Good	-1.598 (0.163)	-1.678 (0.167)
Poor	0	0

* p-value$>$0.05

Discussion

- In SA women are unable to negotiate safe sex with their partners (Worth *et al*. 1990; Stein *et al*. 1990).
- So far it has been reported that HIV is high among women than men based on antenatal clinic attendee studies and other studies (DOH 2007; Shisana *et al*. 2003).
- Individuals in age group 25 to 34 are engaged in highly connected sexual networks.
- Africans are mostly found in informal areas where the socio-economic impact plays a huge role in the spread of HIV.
- Use of condom during sexual debut is an indication of being well informed about risk of HIV.

Discussion *cont*

- TB is one of the leading opportunistic infections in HIV infected individuals.
- Individuals infected with TB are likely to be infected with HIV due to the compromised immune system.
- Male individuals are likely to be exposed to poor working environments than females.
- While those who are educated have enough income to take good care of their health and less likely to contract TB.

Conclusion

- Results from the Bayesian and the GLMMs are quite comparable & this comparability is re-assuring
- Inference drawn from the two modelling approaches provides some degree of confidence in the results.
- The positive correlation observed at an EA level for both HIV (ρ_{HIV}) and TB (ρ_{TB}) indicates that interventions should be aimed at an area level rather than only the individuals.
- Studies that intervene at community level. e.g Mwanza Trial (Grosskurth, 1995) should be encouraged to fight epidemics of diseases such as HIV and TB.

Acknowledgements

- Dr Mwambi, UKZN, supervisor.
- Dr Zuma, HSRC, co-supervisor.
- SACEMA, for funding my MSc studies.
- HSRC, for allowing me to use their rich dataset.
- StatSA and HSRC for sponsoring me to attend this conference.

Thank You