ISSN NO. 2320-5407

Journal homepage: http://www.journalijar.com

INTERNATIONAL JOURNAL OF ADVANCED RESEARCH

RESEARCH ARTICLE

Relative Risk Estimation of HIV/AIDS among Different Major Cities of Pakistan

Syed Rizwan Ul Haq Tirmizi[1*], Nasiruddin Khan[1] and Azhar Mahmood[2]

1. Department of Mathematics, University of Karachi, Karachi-75270, Pakistan.

2. National University of Sciences and Technology, Islamabad, Pakistan..

Manuscript Info

Manuscript History:

Received: 22 February 2015
Final Accepted: 25 March 2015
Published Online: April 2015

Key words:

HIV/AIDS, multilevel models, Pakistan.

Corresponding Author

Syed Rizwan Ul Haq Tirmizi

Abstract

The Pakistan HIV/AIDS pandemic is made up of multiple epidemics spatially located in different parts of the country with most of them having the potential of being sustained into the future given information on some risk factors. It is hoped that the findings of this research will be a ready tool in the hands of policy makers in the formulation of policy and design of programs to combat the epidemic in the country. Uncertainties surrounding assumptions of infection intensity can be minimized using Bayesian methods.

INTRODUCTION

Pakistan is situated in the South –Asian region also known as the Indian subcontinent. It was declared an Islamic republic after gaining independence from the British Colonial Rule in 1947. Its geographical position situates Afghanistan to its west, china and central Asia to the north, India to the east and Arabian Sea on its south. Out of the countries bordering it, only Iran is among the low prevalence group while India and Afghanistan are both high risk high prevalence countries. As per the last official census of Pakistan conducted in 1998, Pakistan had a total population of 132 million with an average annual growth rate of 2.69% (USAID-2004). This puts the estimated population figure in 2014 close to the mark of 170 million. The major portion of the population is young with 43% of it under age of 15 and 53% lying in the bracket of 15-64 years. Official figures for literacy are slightly more than 50% but realistic figures by different news reports and analysts put it around 30% with relatively higher literacy in males than females. More than 97% confess the Muslim faith with Christians, Hindus, Sikh, and parses making the rest of the 3 percent.

The idea behind this section is that areas close to one another in geographical space share the same environmental, socio-economic, cultural and demographic factors that influence disease rates and are more likely to share similar relative risks. Ignoring this dependence, where it exists, will result in the standard errors of the ecological regression coefficients being too small if the dependence is positive or too large if the dependence is negative. Thus, we need to reflect this prior knowledge in the model by incorporating a spatial component into the multilevel models.

Methodology

In fact a gamma prior distribution is mathematically convenient for the relative risk but with few drawbacks, it is not allowing to show spatial correlation between the risks of close areas. While A log-normal model which is very flexible for the relative risk.

A log-normal model using statistical tool WINBUGS:

$$y_{i\sim poisson\ (e_i\theta_i)}$$
$$\log\theta_i = \alpha + v_i$$
$$v_i \sim N(0, \tau_v^2)$$

Now considers mortality rates in eight cities of Pakistan listed above. The data is shown below:

S/No	Major Cities/Zones	No of HIV/AIDS cases	No Of deaths (approx)
1	Karachi	2388	716
2.	Hyderabad	2268	680
3.	Larkana	4240	1272
4.	Lahore	1405	421
5.	Faisalabad	1152	645
6.	Sargodha	1829	548
7.	Peshawar	1072	321
8.	D.G.Khan	1436	431

Source: HASP IBBS round III

Table 1. Number of HIV/AIDS cases in the major cities of Pakistan.

Major Cities/Zones	**node**	**Mean**	**Sd**	**MC error**	**2.5%**	**median**	**97.5%**	**sample**
Karachi	theta[1]	0.3008	0.01124	1.19E-4	0.2791	0.3008	0.3233	8000
Hyderabad	theta[2]	0.3005	0.01133	1.2E-4	0.2791	0.3002	0.3231	8000
Larkana	theta[3]	0.3006	0.008359	9.151E-5	0.2845	0.3005	0.3174	8000
Lahore	theta[4]	0.301	0.01423	1.747E-4	0.274	0.3006	0.3293	8000
Faisalabad	theta[5]	0.5488	0.02226	3.031E-4	0.5065	0.5483	0.5929	8000
Sargodha	theta[6]	0.3005	0.01251	1.316E-4	0.2764	0.3003	0.3257	8000
Peshawar	theta[7]	0.3012	0.01616	1.874E-4	0.2704	0.3009	0.334	8000
D.G.Khan	theta[8]	0.3015	0.0142	1.632E-4	0.2742	0.3013	0.33	8000

Table 2. Relative risk estimation among the cities of Pakistan

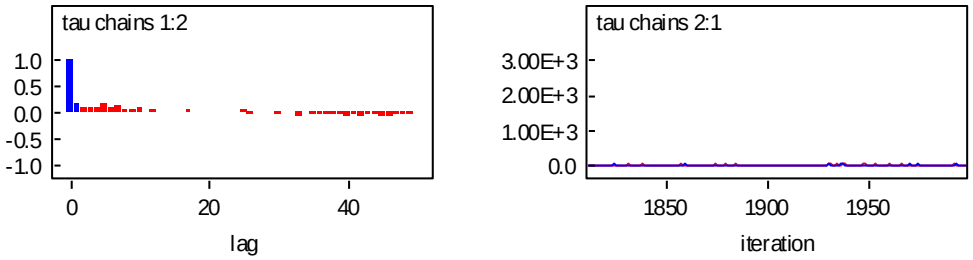

Figure 1. Autocorrelation and Gelman and Rubin convergence diagnostic plot for τ in the log normal model.

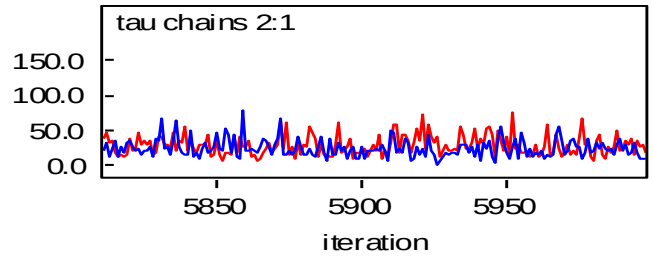

Figure 2. Gelman and Rubin convergence diagnostic for parameter τ in the log normal model after 4000 more

iteration

For convergence, it is required more than 2000 iterations. The Gelman and Rubin diagnostic plot for the parameter τ_v, which is shown in figure 2 displays autocorrelation for the parameter τ_v.

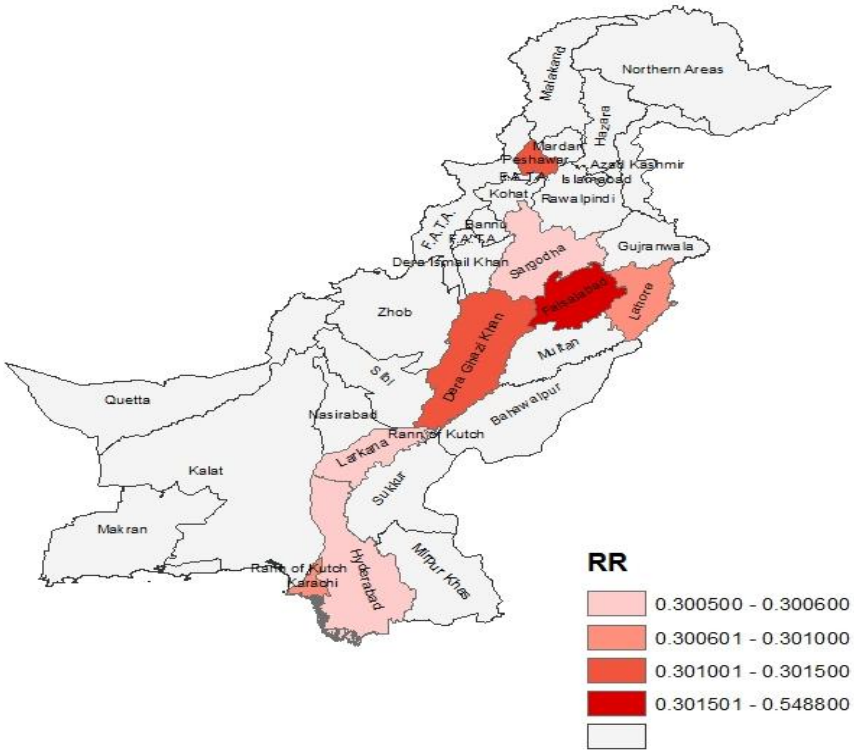

Figure 3. Map illustration log normal model for relative risk

Monitoring Convergence

To establish convergence when fitting, we used three different criteria; the history trace plots, Gelman-Rubin diagnostics (Gilks WR-1996) and the Monte Carlo error as a percentage of the posterior standard deviation. To achieve this, we ran two parallel chains using different starting values with the aim of obtaining an equilibrium distribution of the Markov chain (Gilks WR-1996). From this point of equilibrium, the joint distribution of the sample values is expected to converge to joint posterior distribution. Further iteration from this stationary point produces dependent sample assumed to have come from the posterior distribution. The period from the first iteration till convergence to the posterior distribution is called the *burn-in period*. This burn-in period is usually discarded and further iterations done in order to obtain samples from the joint posterior distribution for posterior inference. Monitoring the convergence of every parameter in a multi-parameter model is not practical; therefore we need to make a decision on the relevant parameters to monitor.

Using the trace or time series plots to monitor convergence, the patterns produced by the parallel chains were observed until they overlap and remain so as the number of iterations increases. The stabilization of this overlap indicates convergence.

Autocorrelation and Gelman and Rubin convergence diagnostic plot for τ in the log normal model is shown in Figure 1. Two parallel chains (the red and the blue lines) were run simultaneously for 2000 iterations from different starting points. For the fixed part, the beta parameters differ significantly in the convergence behaviour. While β[6] reached convergence at an early iterative stage, β [5] is yet to reach convergence even after 4,000 iterations. This problem of convergence of the parameters could be overcome by centering the parameters (Spiegelhalter D,J-2004). Convergence in the random part of the model was easier to achieve than that of the fixed part. As can be seen from the plots, zone variances converged at the early stage of the iteration and remained stable to the end.

We also monitored the convergence of the iterative sampling using the Gelman-Rubin convergence test. The time series plot of the components of the test is shown in Figure 2 . Figure 2 shows that most of the parameters reached convergence at the 5,000 iterations. However, 10,000 iterations were used as the burn-in period for this model. After

convergence, we ran further iterations in order to improve the inference on the posterior estimates.

Normal-Normal model

The spatial variation of HIV prevalence rates in Pakistan may be more distinct if the multilevel structure of the data is incorporated. Also the large variation at the site level may be indicative of the spatial structuring of HIV prevalence in the country.

We shall seek to break down the influences on the distribution of HIV infection into eight major cities as per availability of data from the reliable source. The geopolitical boundaries imposed on the areas of Pakistan are artificial, individuals in areas close to each other tend to share common socio-cultural, religious and behavioral factors that influence the spread of HIV. Therefore spatial smoothing of the HIV relative risk distribution might remove any variation imposed on the data as a result of the geopolitical groupings. Employing the techniques of multilevel modeling also makes it possible to account for the interclass correlation effects between the neighborhood groupings.

A Bayesian analysis of the relative risk using Normal- Normal model is as follows:

S/No	Major Cities / Zones	Log rr estimates	Standard Statistical error
1	Karachi	0.3008	0.01124
2.	Hyderabad	0.3005	0.01133
3.	Larkana	0.3006	0.008359
4.	Lahore	0.3001	0.01423
5.	Faisalabad	0.5488	0.02226
6.	Sargodha	0.3005	0.01251
7.	Peshawar	0.3012	0.01616
8.	D.G.Khan	0.3015	0.0142

Table 3. Relative Risks among the eight largest cities

The model incorporating the spatial effects is given as

$$\log(RR_i) \sim N(\beta_i)$$
$$\beta_i \sim N(\theta, \tau^2)$$

RR_i= relative risk for the city i

σ_i=Standard estimated error for RR_i

β_i = relative risk for the city i

θ = Pooled RR

τ^2= city level heterogeneity for the true RR

$\theta \sim$Normal $(0.1x10^6)$

$1/_{\tau^2} \sim Gamma(0.001,0.001)$

We specified non informative priors for the mean and variance i.e. normal prior with large variance for θ and the Gamma distribution for $1/_{\tau^2}$.But a Gamma distribution is positive and have very big standard deviation with the above parameters. We do not required to mention prior for β_i's because we already put a Normal distribution assumption on them.

S/no	Major Cities / Zones	Node	Mean	Sd	MC error	2.5%	Median	97.5%	samples
1	Karachi	beta[1]	0.3013	0.01122	2.696E-4	0.2801	0.3011	0.3236	2000
2.	Hyderabad	beta[2]	0.3012	0.01125	2.134E-4	0.279	0.3011	0.3234	2000
3.	Larkana	beta[3]	0.3012	0.00824	1.771E-4	0.285	0.3013	0.3167	2000
4.	Lahore	beta[4]	0.3022	0.01357	2.934E-4	0.2754	0.3022	0.3283	2000
5.	Faisalabad	beta[5]	0.5344	0.02279	5.19E-4	0.4889	0.5342	0.5802	2000
6.	Sargodha	beta[6]	0.3011	0.01226	3.12E-4	0.2767	0.3011	0.3246	2000
7.	Peshawar	beta[7]	0.3024	0.01589	3.588E-4	0.2713	0.3025	0.3326	2000
8.	D.G.Khan	beta[8]	0.3022	0.01391	3.021E-4	0.2756	0.3022	0.3298	2000
		Tau	0.0968	0.03829	0.001181	0.0533	0.08927	0.1772	2000
		theta	0.2337	0.2337	0.003771	0.2023	0.6377	1.123	2000

Table 4. Bayesian estimate of the true city-specific RR .

The mean of beta[i] in the above table represent the Bayes estimate of the true city-specific RR .What happens if we amend the prior distribution on θ to be more informative?
Let
$\text{theta} \sim \text{dnorm}(2,10)$
Notice how the Bayesian estimate is pulled closer to 2, the prior mean.

node	Mean	sd	MC error	2.5%	median	97.5%	start	sample
theta	0.3836	0.1648	0.01477	0.2816	0.3583	0.5489	1	2000

Table 5. baysian estimate closer to 2.
For fixed – effect
$r_i \sim Binomial(p_i, n_i)$
Where r_i= No of deaths in city i and n_i= No of HIV/AIDS cases. Here we assume that true deaths probabilities (P_i) are independent (i.e fixed effects) for each city i. It is assumed to be equivalent to the standard non informative prior distribution for the $P_i's$ namely:
Priors:

$$P_i \sim Beta(1,1)$$

It is to be noted that Beta distribution is continuous distribution and takes values between 0 and 1.
Random-effect:

$$\mathrm{r_i} \sim \mathrm{Binomial}(\mathrm{P_i}, \mathrm{n_i})$$

$$\log\left[\frac{P_i}{1-P_i}\right] \sim normal(\mu_i, \tau^2)$$

In this model, we assume that the logit of each mortality's rate is dependent to each other.
Standards priors:

$$\mu \sim Nomal(0.1 \times 10^6)$$
$$1/\tau^2 \sim \mathrm{Gamma}(0.001, 0.001)$$

Results:

For fixed effect analysis

S/No	Major Cities/ Zones	**node**	**mean**	**Sd**	**MC error**	**2.5%**	**median**	**97.5%**	**sample**
1	Karachi	p[1]	0.2999	0.009496	1.44E-4	0.2815	0.2997	0.3192	4000
2.	Hyderabad	p[2]	0.3002	0.009364	1.411E-4	0.2822	0.2999	0.3186	4000

3.	Larkana	p[3]	0.3001	0.007067	1.172E-4	0.286	0.3001	0.314	4000
4.	Lahore	p[4]	0.3	0.0124	1.843E-4	0.2757	0.2999	0.3248	4000
5.	Faisalabad	p[5]	0.5603	0.01445	2.548E-4	0.5311	0.5607	0.588	4000
6.	Sargodha	p[6]	0.2998	0.01062	1.504E-4	0.2791	0.2997	0.3207	4000
7.	Peshawar	p[7]	0.2998	0.01421	2.165E-4	0.2727	0.2999	0.3282	4000
8.	D.G.Khan	p[8]	0.3003	0.01219	1.849E-4	0.2769	0.3001	0.3246	4000

Table 6. Fixed effects values

For random effect analysis:

S/no	Major Cities/ Zones	**Node**	**mean**	**Sd**	**MC error**	**2.5%**	**Median**	**97.5%**	**Sample`**
1	Karachi	p[1]	0.3004	0.00954	1.476E-4	0.2819	0.3006	0.3186	4000]
2.	Hyderabad	p[2]	0.3002	0.00963	1.538E-4	0.2817	0.3002	0.3189	4000
3.	Larkana	p[3]	0.3004	0.00692	1.055E-4	0.287	0.3002	0.314	4000
4.	Lahore	p[4]	0.3	0.0125	1.911E-4	0.2773	0.2998	0.3246	4000
5.	Faisalabad	p[5]	0.554	0.01525	2.38E-4	0.5248	0.5543	0.5828	4000
6.	Sargodha	p[6]	0.3001	0.01094	1.85E-4	0.2801	0.3	0.3212	4000
7.	Peshawar	p[7]	0.3	0.01424	2.39E-4	0.2739	0.2996	0.3272	4000
8.	D.G.Khan	p[8]	0.3006	0.0119	1.925E-4	0.278	0.3005	0.3243	4000
		Tau	0.5841	7.51	0.1606	0.2443	0.3961	0.7766	4000
		pop.mean	0.3296	0.03907	5.721E-4	0.2658	0.3276	0.4054	4000

Table 7. Random effect values

Hence using the random effects model, we estimate the mortality mean probability as 0.3296 and standard deviation 0.5841.

Bayesian analysis has the ability to estimate any function of the parameters by examining the corresponding posterior distribution. We can track the mortality rate among the cities.

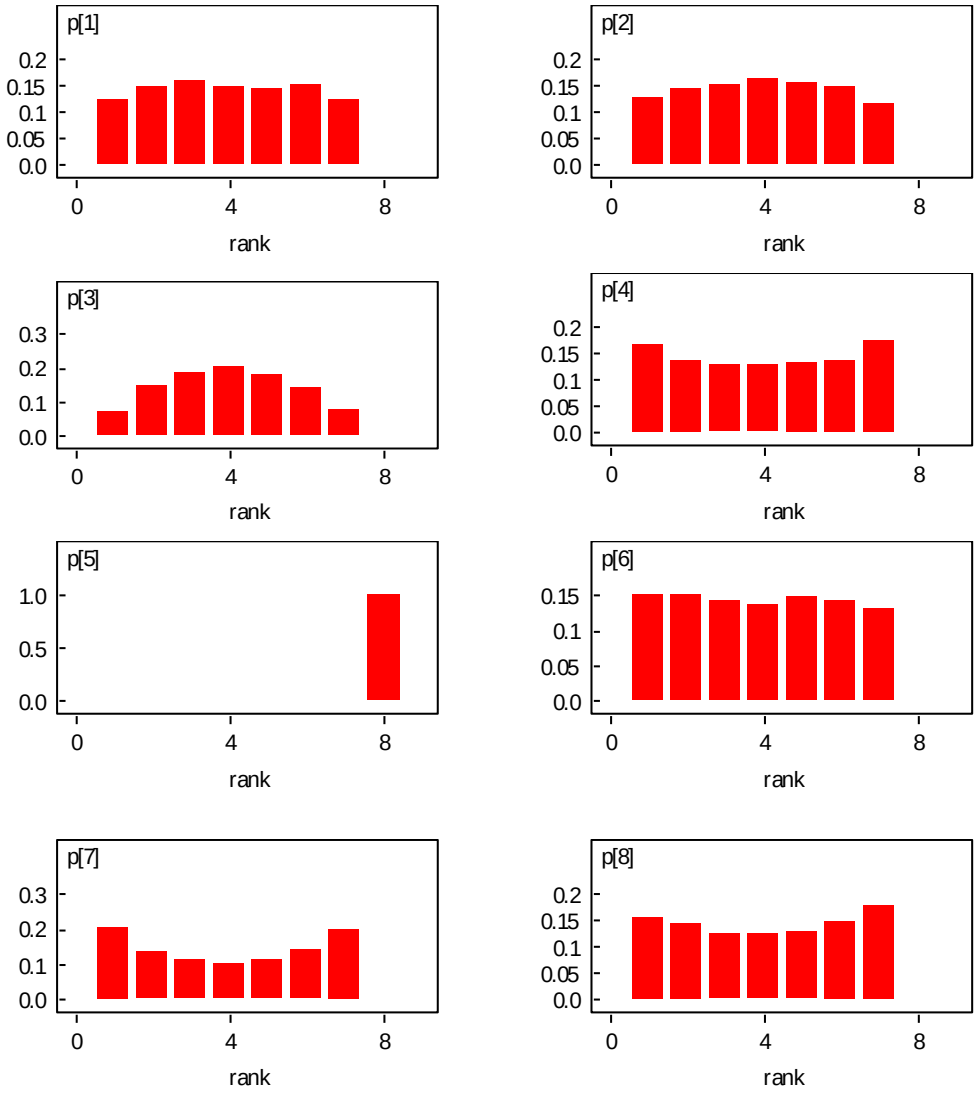

Figure 4. Mortality rate among the cities.

node	mean	sd	MC error	2.5%	Median	97.5%	Sample
Larkana & Karachi	0.493	0.5	0.008406	0.0	0.0	1.0	4000
Karachi & Lahore	0.5172	0.4997	0.008439	0.0	1.0	1.0	4000
Karachi & peshawar	0.5105	0.4999	0.008911	0.0	1.0	1.0	4000
Larkana & Hyderabad	0.5092	0.4999	0.007743	0.0	1.0	1.0	4000

Table 8. Risk intensity estimated data among the cities.

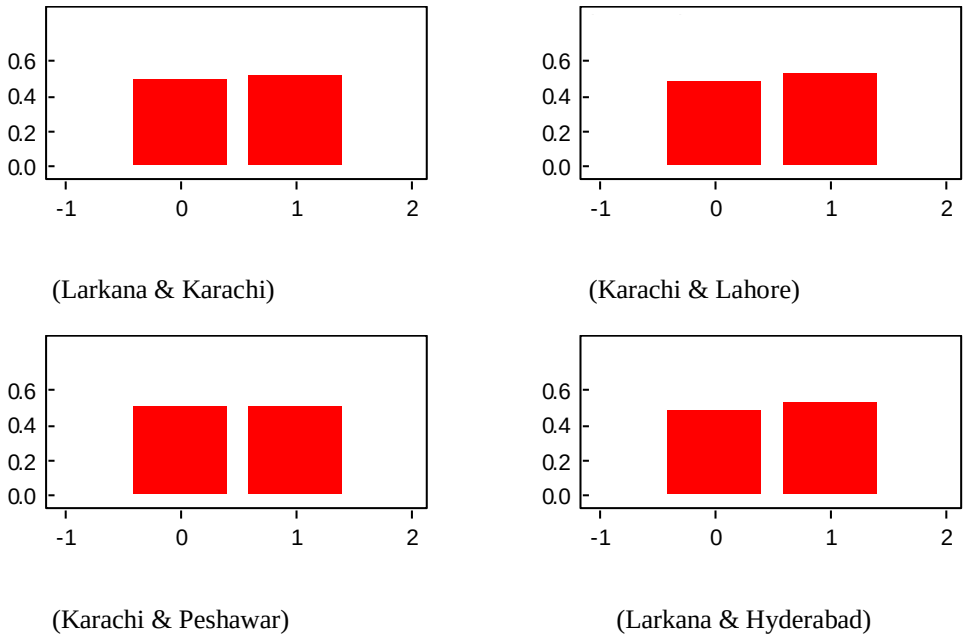

Figure 5 Comparison of HIV/AIDS intensity of risk among the few cities of Pakistan:

Limitations of the research

The availability of data on HIV/AIDS in Pakistan is a serious limitation on this research. The nature of available data, to a great extent, tailored the direction of this research. Data used in the construction models do not cover the period of the inception of the epidemic and the recent years. The data were collected on all HIV/AIDS diagnosed between 2006 - 2009, thus omitting the first reported cases in 1987 and data up to 2005. Also, there was no detailed information about the epidemic in the recent years (2010 - 2013). The aggregation of the data at national level made it impossible to conduct more detailed analysis. The study and modeling of the trend of the epidemic across the various demographic strata of the Pakistan society was hindered by the non-stratification of the data. It is strongly recommended that data be placed on the public domain after striping all patient identities and made more accessible to researchers. Aggregation of data should be avoided as much as possible, at the least; data with complete details should be published by sex and age for each of the provinces of Pakistan.

Conclusion

The aim of this research was to develop epidemic models that could describe and predict the HIV/AIDS epidemic. The spatial effect on the distribution of the risk of HIV infection was studied by fitting spatial multilevel models. Variations in the risk of HIV infection within the eight cities of Pakistan were prominent in the models. Accounting for the spatial effects, as discussed, significantly reduced the random variability in HIV prevalence across the cities. Thus, relative risk estimates obtained using these models are relatively precise and are expected to give an accurate map of the distribution of HIV prevalence in Pakistan.

References

Clayton D and Bernardinelli L. (1992). Bayesian Methods for mapping disease risk. In P. Elliot, J. Cuzick, D. English and R. Stern, editors, Geographical and Environmental Epidemiology: methods for Small-Area Studies.

Gelman A and Rubin DB.(1992). Inference from iterative simulation using multiple sequences. Statistical Science, 7(4):457-511,

Gelman A and Rubin DB. (1992.). A single series from the Gibbs sampler provides *a* false sense of security. In Bayesian Statistics: Bernado JM, Berger JO, Dawid AP and Smith, eds. Oxford University Press.

Goldstein H.(1995). Multilevel Statistical Models. Edward Arnold, London,

Hegarty A and Barry D.(2008). Bayesian disease mapping using product partition models. Statistics in Medicine, 27:3868-3893.

Lawson A, Browne W, and Vidal Rodeiro C L. (2003). Disease Mapping with WinBUGS and MLwiN. West Sussex: John Wiley and Sons Ltd,

Mollie A. (1999). Bayesian and Empirical Bayes Approaches to Disease Mapping. In: Lawson A, Editor. Disease Mapping and Risk Assessment for Public Health. John Willey and Sons Ltd.

National AIDS Control Program.(2010). Ministry of Health Government of Pakistan Islamabad.

Spiegelhalter D J, Thomas A, (2004). Best N, and Lunn D. WinBUGS User Manual, version 2.0. MRC Biostatistics Unit, Cambridge.

UNAIDS (2008). Epidemiology Reference Group. Technical report and recommendations: Improving EPP and spectrum estimation tools for the round of national estimates with specific attention to prevalence facts, uncertainty, etc. Meeting held in London.

UNAIDS Epidemiology Reference Group.(2001). Recommended methodology for the estimation and projection of HIV prevalence and AIDS mortality in the short-term. Meeting held at La Mainaz Hotel, Gex France.

UNAIDS FACT SHEET (2014).

UNAIDS. (2006). Epidmiology Reference Group. Technical report and recommendations: Improving parameter estimation, projection methods, uncertainty estimation and epidemic classification. Meeting held in Prague, Czech Re-public.

WHO. (2003). Methods for estimation/projection of HIV infection and AIDS cases/deaths.HIV/AIDS in Asia and Pacific Region. Annex 4, pages 102- 109.