Kuwait J. Sci. **46 (4)** pp. 33-46, 2019

Socio-economic factors of misconception about HIV/AIDS among ever-married women in Punjab: A comparison of non-spatial and spatial hierarchical Bayesian Poisson model

Majida Rashid[1,*], Dr. Sohail Chand[2]

[1]*PhD Scholar, University of the Punjab.*

[2]*Associate Professor, University of the Punjab.*

[*]*Corresponding Author: mrs.majidajawad@gmail.com*

Abstract

Combating HIV/AIDS is the third goal of Sustainable Development Goals and has become an increasing health concern in Pakistan. On 25 April 2019, the nearby organization in Larkana locale was cautioned by media reports of a flood in human immunodeficiency infection (HIV) cases among youngsters in Ratodero Taluka, Larkana region, Sindh region, Pakistan. From 25 April through 28 June 2019, a sum of 30,192 individuals have been screened for HIV, of which 876 were discovered positive. Eighty-two percent (719/876) of these were underneath the age of 15 years. The only way to combat HIV transmission is to provide accurate knowledge about how the disease is spread to the general public and especially to women of child bearing age (15-49 years). Prevention of HIV and misconceptions about its transmission are related. Therefore, the present study aimed to identify the spatial distribution of three misconceptions of HIV transmission (transmission by mosquito bite, supernatural means and sharing food with an HIV positive person). This study also provides the core socio-economic factors needed to stop misconceptions about HIV/AIDS transmission, which will help in reducing the spread of the epidemic in Pakistan. Spatial and non-spatial Bayesian Hierarchical models were applied. The results show that the Conditional Autoregressive Bayesian Hierarchical Models (Spatial Model) are a more appropriate regression model in the presence of spatial dependence.

Keywords: Bayesian hierarchical model and conditional autoregressive Bayesian hierarchical model; Geary's c; HIV/AIDS; misconception; Moran I; Spatial autocorrelation.

1. Introduction

The inclusion of the spatial effect in traditional statistical modeling is an important field, especially as it is related to Geographical Information System (GIS), which provide the foundation for spatially-indexed outcomes and covariates within epidemiological and environmental models. The spatial models have been applied on a range of different subject areas, such as disease mapping (Best *et al*., 2005), water quality (Zeilhofer *et al*., 2006), air pollution monitoring (Lee & Shaddick, 2010), and ecology (Hoef *et al*., 2006). These approaches provide more accurate models in the presence of spatial dependence by incorporating a spatial autocorrelation structure. The transmission of infectious disease is closely linked to the concept of spatial and spatial-temporal proximity, as transmission is more likely to occur if the at-risk individuals are close in a spatial and temporal sense (Pfeiffer *et al*., 2008)

Important research on Bayesian approaches to modelling disease count at the small-area level include, but are not limited to, Manton *et al.* (1981), Tsutakawa (1988), Besag *et al.* (1991), Marshall (1991), Clayton & Bernardinelli (1992), Breslow & Clayton (1993), Lawson (1994), and Ghosh *et al.* (1998). Central to this method is the inclusion of random effect terms to account for unobserved, spatial features within the data. The appeal of this approach is that it does not produce a single (global) spatial autocorrelation coefficient as is the case with the Conditional Autoregressive Model and Simultaneously Autoregressive Model approaches. Rather, spatially correlated random-effect terms are estimated for each area units, thereby allowing the analyst to identify aggregations of spatial units in which the incidence of disease is not explained by the model.

Multilevel Bayesian models have been used to investigate the spatial distribution of testicular and prostate cancer in Britain (Toledano *et al.*, 2001; Jarup *et al.*, 2002), breast cancer in Greece (Vlachonikolis *et al.*, 2002), insulin-dependent diabetes mellitus in Austria (Schober *et al.*, 2001), stroke and cardiovascular disease in Great Britain (Maheswaran *et al.*, 2002), multiple sclerosis in Italy (Pugliatti *et al.*, 2002), low birth weights in Papua New Guinea (Mueller *et al.*, 2002), malaria in South Africa (Kleinschmidt *et al.*, 2002), and BSE in Great Britain (Stevenson *et al.*, 2005).

In this paper, we compared two Bayesian approaches, i.e. Bayesian Hierarchical Poisson Models (BHPM) and the Spatial Conditional Bayesian Hierarchical Poisson Models (SCBHPM) to modelling misconception indicators of HIV/AIDS in Punjab, Pakistan for the year 2010-2011. Combatting the spread of HIV is one of the world's most serious healthcare challenges. According to the Global Burden of Disease Study Report (GBD) from 2016, the number of HIV infections in Pakistan grew at an average of 17.6 percent a year from 8,360 to 45,990 during the period from 2005 to 2015. This makes it the highest increase in the world (GBD, 2016). The latest report of WHO showed that the estimate (2017) was 150,000 people living with HIV (PLHIV) while in 2018, 21,000 new PLHIV cases were recorded.

When individuals have knowledge about how HIV is spread, HIV/AIDS cases decline and epidemics are avoided (Mondal *et al*., 2016). Correct knowledge about HIV transmission increases safer sexual behavior and is considered an important step toward behavioral change. At the same time, misconceptions can prevent individuals from safer sexual behavior and keep them from taking appropriate action against HIV acquisition and transmission. Misconceptions regarding HIV/AIDS are still very prevalent in the general public. A lot of research has been carried out in developing and developed countries that investigates the relationship between sociodemographic risk factors and correct knowledge about HIV transmission (Letamo 2007), (Mizanur *et al*., 2009), (Mondal *et al*., 2016), (Ochako *et al*., 2011), (Rauf *et al*., 2010).

Other studies have examined the determinants of the transmission of HIV/AIDS virus. Letamo (2007) applied the Logistic Regression Model (LRM) on 9,272 ever married women in Bangladesh. Letamo identified the most prominent determinants of misconception in Bangladesh, which included limited access to mass media, poor economic conditions, living in less developed areas (rural areas), women with poorly educated husbands, a woman's education level, and the age of ever-married women these finding also consistence with (Mondal *et al* (2016). The multivariate logistic regression model (MLRM) was used by Mizanur (2009) to investigate adolescents' knowledge about sexually transmitted diseases (STDs) including HIV/AIDS. The study found that 54.8% of the adolescents had never heard about AIDS. He also revealed that an adolescent's age, years of schooling and knowledge of STDs appeared to be important predictors of an awareness about AIDS (Mizanur *et al*., 2009).

Although these studies successfully examined the factors that influence the knowledge of HIV transmission at individual- and country-level variables, a direct comparison between individual studies is often unrealistic. This is because these studies were performed in different national contexts. In addition, they may not include similar measures or adjust for the same variables (Hailu 2016). Another aspect which is of great interest and ignored in these studies was the clustering and the hierarchical structure of data in the population, both of which could not be captured by using simple LRM.

There are multiple reasons for revising these studies as such revisions are necessary to acquire more and more precise data – which, in turn, plays a pivotal role in saving lives especially in cases involving women's health. To the best of our knowledge, there are very few or no studies that have concentrated on misconceptions about HIV transmission in Pakistan using nationally or provincially representative data, particularly Multiple Indicator Cluster Survey (MICS) 2010-11 data. Therefore, the present study aimed to identify the factors among different socio-economic and demographic factors affecting misconception indicators regarding HIV transmission among ever-married women in Punjab, Pakistan. Hopefully the findings of this empirical study will contribute to the development of an enhanced ever-married women's health framework in Pakistan, including recommendations for the development of educational interventions to decrease misconceptions and enhance empowerment with respect to HIV prevention strategies.

2. Material and methods

A secondary data set utilized in our analysis was taken by the MICS 2010 and which was conducted by the Punjab Bureau of Statistics by the government of Punjab, Pakistan in collaboration with the United Nations Development Program (UNDP) and the United Nations Children's Fund (UNICEF). In total, 102,545 households from 36 districts and 150 tehsils (sub-districts) were interviewed. For the study, 85,502 ever-married women

(EM) were interviewed about their comprehensive knowledge regarding the transmission of HIV/AIDS. For the statistical methodology, a unit of analysis is misconception and three misconception-related variables were selected from the MICS 2010 data to evaluate the misconception about the transmission of HIV/AIDS. Eighteen socio-economic covariates were employed to determine possible relationships with the misconception indicators. The descriptive analysis of response variables and covariates used in this study is presented in Table 1.

Table 1. Descriptive analysis of response variable and covariates

Description	Abbreviation	Min.	Max.	Mean	Std. Deviation
Response Variables (Misconception Indicators)					
Number of ever-married women who think HIV cannot be transmitted by mosquito bite	Mosquito	3.00	532.00	83.93	79.66
Number of ever-married women who think HIV cannot be by supernatural means	SN	3.00	636.00	103.11	94.58
Number of ever-married women who think HIV cannot be transmitted by sharing food with an HIV positive person	SF	2.00	563.00	73.13	75.26
Covariates					
Demographic Variable					
Average family size	size	5.10	7.70	6.26	0.44
Average age of EM women	age_mean	31.29	35.06	33.41	0.75
Adolescent birth rate	adb	1.00	100.00	36.14	14.99
Women's Education (%)					
Literate women	literacy	7.70	73.40	43.63	15.34
EM women who attended primary school	primary	3.90	33.03	18.69	5.48
EM women who attended middle school	middle	2.11	17.35	9.13	3.65
EM women who attended high school	matric	1.56	24.47	11.13	4.83
EM women who attended secondary school level school	A. Matric	0.89	30.00	8.93	4.47
Economic Wellbeing (%)					
EM women living in urban areas	U. Area	0.00	100.00	38.49	13.07
Households owner with assets (house, agricultural land or animals)	ownership	63.80	99.80	93.42	5.49
Households receiving government benefits	safetynet	0.10	62.30	6.72	9.17
Households with TVs	TV	23.66	91.57	63.32	15.26
Health Care (%)					
Hepatitis patients	Hp	0.20	4.30	1.14	0.69
TB patients	TB	0.10	2.30	0.45	0.33
Women receiving postnatal care	postnatal	7.00	65.00	36.40	12.89
Types of Contraception (%)					
EM women using female sterilization	FS	3.50	20.40	10.61	3.61
EM women with intrauterine devices	IUD	0.50	12.10	3.36	1.78
EM women's husbands using condoms	condom	0.40	18.70	7.98	4.22

A Choropleth map is a commonly used technique for representing aggregated data. Researchers use it in order to visualize aggregated data. This study first used the Choropleth map to observe a spatial variation among the distribution of three types of indicators for misconceptions about HIV transmission. After that, global measures of spatial autocorrelation (Moran's I and Geary's c) were applied to further examine the spatial effects. Moran's I coefficient of autocorrelation is like Pearson's correlation coefficient as it quantifies the similarity of an outcome variable among areas that are defined as spatially related (Moran, 1950). Moran's, I statistics (Moran, 1950) is as follows:

$$I = \frac{n\sum_i\sum_j W_{ij}(Z_i-\bar{Z})(Z_j-\bar{Z})}{(\sum_i\sum_j W_{ij})\sum_k(Z_k-\bar{Z})^2}, \quad (1)$$

where Z_i is the disease rate of an area, W_{ij} is a measure of the closeness of area i and j, and n is the number of regions in the study. Thus, spatial weights can be seen as a list of weights that are indexed by a list of neighbors and where the weight of the link between i and j is the k^{th} element of the i^{th} weights list component. k tells us which of the i^{th} neighbor list component value is equal to j. If j is not present in the i^{th} neighbor list component, j is not a neighbor of i. Consequently, some weights W_{ij} in the $\boldsymbol{W}$ weight matrix representation will set to zero, where j is not a neighbor of i.

Once the study area's neighbor list sets were established, we assigned spatial weights to each relationship. We used binary weights to construct the weight matrix for the analysis. The following rules were applied:

$w_{ij} = 1$ if i and j is shared a common boarder

$w_{ij} = 0$ otherwise

A binary weight matrix was used in which $w_{ij} = 1$ if i and j shared a common boarder and zero otherwise.

Another weighted estimate for spatial autocorrelation known as Geary's c, or Geary's contiguity (Geary, 1954) was also utilized. This contiguity is quite different from Moran's I. Moran's I considers the similarity between neighboring regions, whereas Geary's c considers the similarity between pairs of regions (Pfeiffer, *et al*., 2008). Geary's c is:

$$c = \frac{(n-1)\sum_i^n\sum_j^n w_{ij}\left(y_i - y_j\right)^2}{2(\sum_i^n(y_i-\bar{y}))^2(\sum_i^n\sum_i^n w_{ij})} \quad (2)$$

The range of Moran's I is +1 to -1, wherein a positive value indicates the positive relationship, a negative value shows a negative relationship. and a zero value indicates no relationship. Similarly, Geary's c considers similarity between pairs of regions with range zero to two. Zero indicates a perfect positive spatial autocorrelation, while two indicates a perfect negative spatial autocorrelation for any pairs of regions.

Bayesian Hierarchical Poisson Models (BHPM) and Spatial Conditional Bayesian Hierarchical Poisson Models (SCBHPM) were applied to identify the socio-economic determinants of three misconception factors. Selection of the best model were based on the Deviance Information Criteria (DIC). R 3.2.2 version software was used to complete the analysis.

The main reason for using a Bayesian approach is that it allows for better representation and takes in a fuller account of the uncertainties related to models and parameter values. In contrast, most decision analyses based on maximum likelihood (or least squares) estimation involve fixing the values of parameters. These parameters may, in actuality, have an important bearing on the final outcome of the analysis and for which there is considerable uncertainty. One of the major benefits of the Bayesian approach is the ability to incorporate prior information. While other approaches use "prior" information by specifying levels or ranges of individual parameters for use in sensitivity analysis, the Bayesian approach forces the analyst to look at historical data sets or to canvass expert knowledge to determine what is known about the biological parameters and processes.

In Bayesian analysis, data are conceptually considered fixed with some distribution of parameters to be estimated (Martin, 2003). For this analysis, the misconception indicators are count data. Therefore, BHPM was used to model the misconception factors and can be expressed as follows:

$$MHIV_k|\lambda_k = Poisson\ (\lambda_k) \quad (3)$$

$$\log(\lambda_k) = \boldsymbol{X_k\beta}, \quad (4)$$

where MHIV is the response variables and are assumed to be statistically independent within each cluster or hierarchy, and k = 1,2 … 135 tehsils. $\boldsymbol{\beta}$ represents the regression parameters and $\boldsymbol{X_k}$ is the vector of covariates.

The likelihood function (LF) for each of the tehsil of Punjab and corresponding covariates is

$$p(\mathrm{MHIV_k}|\boldsymbol{X_k},\boldsymbol{\beta} = Poisson\ (\lambda_k), \quad (5)$$

where $p(.|.)$ denotes a conditional probability mass function. The Poisson density is evaluated at the specific values of $MHIV_k$ with a corresponding mean parameter λ_k. The nineteen (19) parameters, $\beta_1, \beta_2, \beta_3, \dots .. \beta_{19}$ correspond to an intercept and eighteen covariates (see Table 1), respectively. The choice of prior distributions within Bayesian models is an important concept. Often prior distributions are chosen to provide only limited information about the parameters, and noninformative or flat/vague priors are chosen. On the other hand, informative prior distributions are useful to ensure particular effects. In the case of the regression model, a Poisson where predictors are to be included within the model, it is commonplace to assume an independent zero-mean Gaussian distribution with a variance $\tau^2_{\beta_k}$ for a regression parameter β_k, that is, $\beta_k \sim \mathrm{N}(0, \tau^2_{\beta_k})$. Variance parameters for random effects τ^2 are often assigned apparently weakly informative conjugate inverse-gamma priors. That is, $\tau^2 \sim$ Inverse Gamma (a, b), where common specifications are a=1.0, b = 0.0001 or $\mathrm{a} = 2.0, \mathrm{b} = 1.0$. This can be written as:

$$\pi(\beta_1), \pi(\beta_2), \pi(\beta_3), \dots .., \pi(\beta_k) = normal\ (0, \sigma^2 = \tau^2_{\beta_k}) \quad (6)$$

Using the Bayes' theorem, the likelihood function and prior distribution determine the posterior distribution of $\beta_1, \beta_2, \beta_3, \dots .. \beta_{19}$, where prior is parametrized by the hyper-parameter. The Markov chain Monte Carlo (MCMC) obtains samples from the desired posterior distribution.

Recalling equation 3, where only one level is considered, namely the tehsil k, for the level 2 hierarchical model for BHPM (Draper, 1996) can be expressed as:

$$\mathrm{MHIV_{kl}}\ |\lambda_{kl} \sim Poisson\ (\lambda_{kl}) \quad (7)$$

$$\log(\lambda_{kl}) = \boldsymbol{X_{kl}\beta} + \gamma_{kl}\ . \quad (8)$$

Here, $\gamma_{kl} \sim normal\ (0, \sigma^2)$, k represents the tehsil ($i.e. k = 1, 2, \dots 135$), and l denotes the clustering effects (district 1, 2, 3, … ,36) the likelihood function is the same as in equation 5. However,, there is an additional parameter due to the clustering effect which is the random effect γ_{kl}. After incorporating the prior, the variance parameter becomes:

$$\pi(\beta_1), \pi(\beta_2), \pi(\beta_3), \dots .., \pi(\beta_k) = normal\ (0, \sigma^2 = \tau^2_{\beta_{kl}}) \quad (9)$$

The Bayesian approach is used to estimate the SCBHPM by specifying the values for the distribution of hyper-prior. This allows the variability of the hyperprior parameters among the tehsils. Since tehsils are close to each other and share common socio-economic aspects, a similar misconception count may be expected in the neighboring tehsils. Conditional Intrinsic Gaussian autoregressive (CIGA) models are the most widely used techniques to explain this prior knowledge (Best, *et al*., 1999).

Recalling equations 3 and 4, SCBHPM models are calculated as:

$$MHIV_k|\lambda_k\ = f(mhiv_k|\lambda_k) \quad (10)$$

$$ln(\lambda_k)\ = X_k^T\beta + O_k + \psi_k. \quad (11)$$

Here, the response variables $MHIV_k$ come from an exponential family of distributions $f(mhiv_k|\lambda_k, \nu^2)$, and CARBayes is from the Poisson family, because response variables are countable, with a natural log link function. ψ_k is a spatially structured component that includes a set of random effects that come from conditional autoregressive models ($\phi = \phi_1, \phi_2, \dots, \phi_k$). This is a special case of the Gaussian Markov Random Field (GMRF) and is expressed as $\phi \sim N(\boldsymbol{0}, \tau^2\boldsymbol{Q}(\boldsymbol{W}, \rho)^{-1})$, where $\boldsymbol{Q}(\boldsymbol{W}, \rho)$ is a precision matrix that is an intrinsic model. The binary specification of $\boldsymbol{W}$ was used and 1 was assigned if the areal units shared a common boundary, otherwise the answer is 0. The CARBayes priors (Leroux, *et al*., 2000) are:

$$\psi_k\ = \ \phi_k \quad (12)$$

$$\phi_k|\boldsymbol{\phi_{-k}}, \boldsymbol{W}, \tau^2,$$

$$\rho \sim N\left(\frac{\rho\sum_1^K w_{ki}\phi_i}{\rho\sum_1^K w_{ki} + 1 - \rho}, \frac{\tau^2}{\rho\sum_1^K w_{ki} + 1 - \rho}\right)$$

$$\tau^2 \sim Inverse - Gamma(a, b),$$

$$\rho \sim Uniform(0, 1),$$

Adding a district-level (level 2) effect in equation 11, the SCBHPM model at level 2 is given as:

$$MHIV_{kl}|\lambda_{kl} = f(mhiv_{kl}|\lambda_{kl}). \tag{13}$$

The link function for level 2 is given as

$$ln(\lambda_{kl}) = \boldsymbol{X}_{kl}^{T}\boldsymbol{\beta} + O_{kl} + \psi_{kl}$$
$$\boldsymbol{\beta} \sim N(\boldsymbol{\lambda}_{\beta}, \boldsymbol{\Sigma}_{\beta}.$$

Here, X_{kl}^{T}, O_{kl} denotes a vector of eighteen covariates and an offset for tehsil k within district l. ψ_{kl} represent the spatial variation, which is common to all tehsil within each district. Lee (2017) defined the spatial and tehsil level variations as:

$$\psi_{lk} = \phi_{lk} + \xi_{\lambda(l,k)}$$
$$\phi_{l}|\boldsymbol{\phi}_{-l} \sim N\left(\frac{\rho \sum_{k=1}^{l} w_{lk}\phi_{k}}{\rho \sum_{k=1}^{l} w_{lk} + 1 - \rho}, \frac{\tau^2}{\rho \sum_{k=1}^{l} w_{lk} + 1 - \rho}\right),$$
$$\zeta_{r} \sim N(0, \sigma^2) \quad for\ all\ r,$$
$$\tau^2, \sigma^2 \sim Inverse - Gamma\ (a, b),$$
$$\rho \sim Uniform\ (0,1)$$

In this case, the study region Punjab (P) is divided into l distinct areal units or districts $\{P_l = P_1, P_2, P_3, \dots, P_{36}\}$, and the data are available on k tehsil within the area (district l). Thus, for the district unit P_l, there are k different response variables being modelled by the incorporation of both tehsil (individual) level and spatial variations. $\zeta_{\mu(l,k)}$ is a random effect for a tehsil level variation and which provides an independent and identically distributed zero-mean Gaussian Prior with a constant variance σ^2 (Lee, 2017).

3. Results

Punjab, Pakistan is divided into nine divisions, thirty-six districts and one hundred fifty tehsils (tehsils were adjusted to 135 to compare with the shape file). A total of 85,502 EM women were interviewed, of which only 26.77% of EM women have heard about HIV/AIDS.

Figure 1 shows the statistics from 2011 and 2014 on the knowledge of HIV/AIDS transmission. According to the data, 83.1% of EM women think that HIV/AIDS can be transmitted by mosquito bite. Similarly, for 2011, 79.3% and 85.3% of EM women responded that HIV/AIDS can be transmitted by supernatural means and by sharing food with an HIV/AIDS positive person respectively. This data point has shown a downward trend since 2014. Figure 2, gives the information about the knowledge of mother-to-child transmission of HIV. The statistics shows that during 2014, 29.5% of EM women believe that HIV/AIDS can be transmitted from mother-to-child during pregnancy. Similarly, 27.7% and 36.3% of EM women believed that HIV can be transmitted to mother-to-child during delivering baby and by breastfeeding respectively. The knowledge of mother-to-child HIV transmission was quite improved in 2014 as compare to 2011.

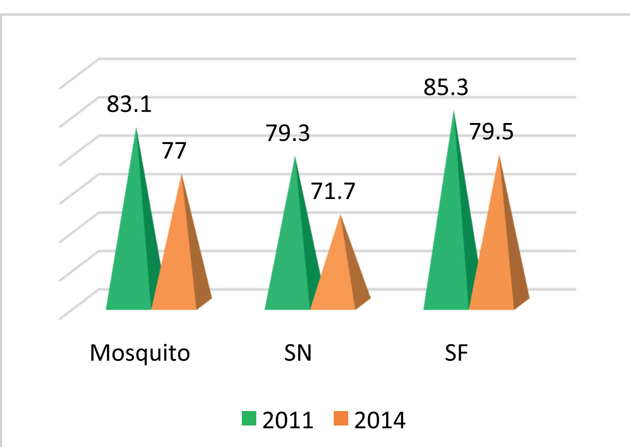

Fig. 1. Knowledge of HIV transmission

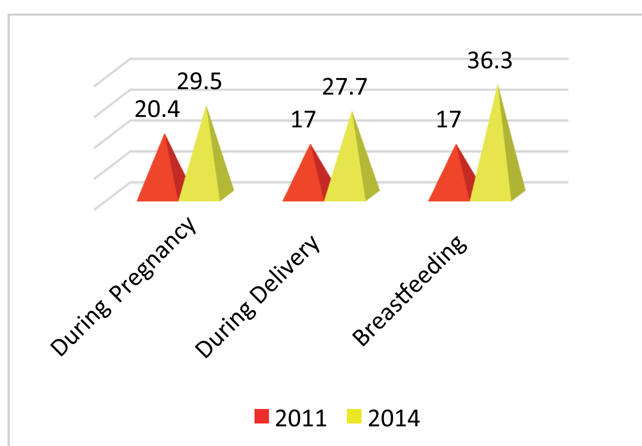

Fig. 2. Knowledge of Mother-to-Child HIV Transmission

Figure 3 shows that the percentage of misconception indicators was similar in tehsils located in the same district. Similarly, the distribution of misconception indicators in the districts showed an analogous trend with a shared communal division.

Figure 4 shows the Choropleth maps. These illustrate that the tehsils, within a shared boundary had similar estimates. The highest misconception was observed in the South-West zone. Lower estimates of misconception were observed in different locations in the northeastern and northwestern parts of the province.

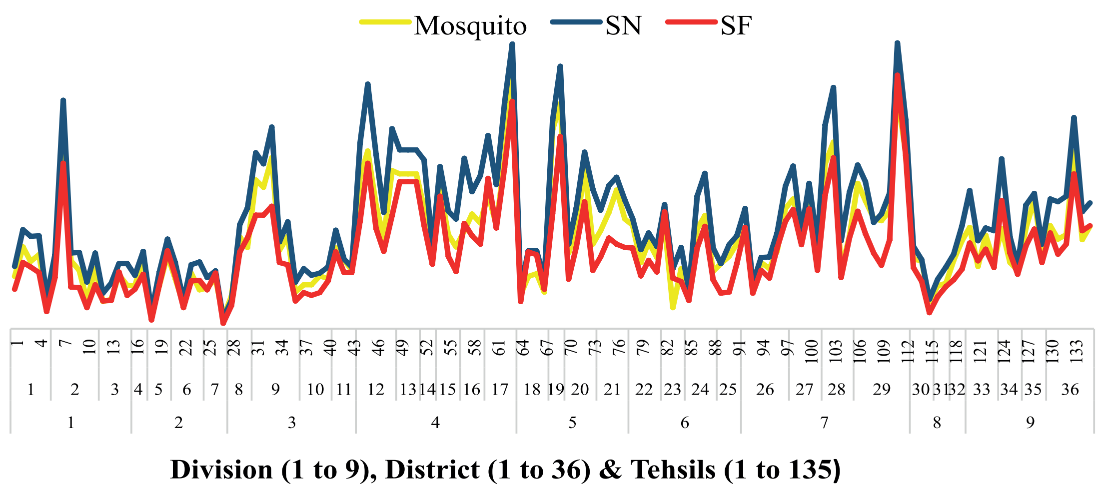

Fig. 3. Distribution of three types of misconception indicators regarding HIV transmission (Mosquito, SN & SF)

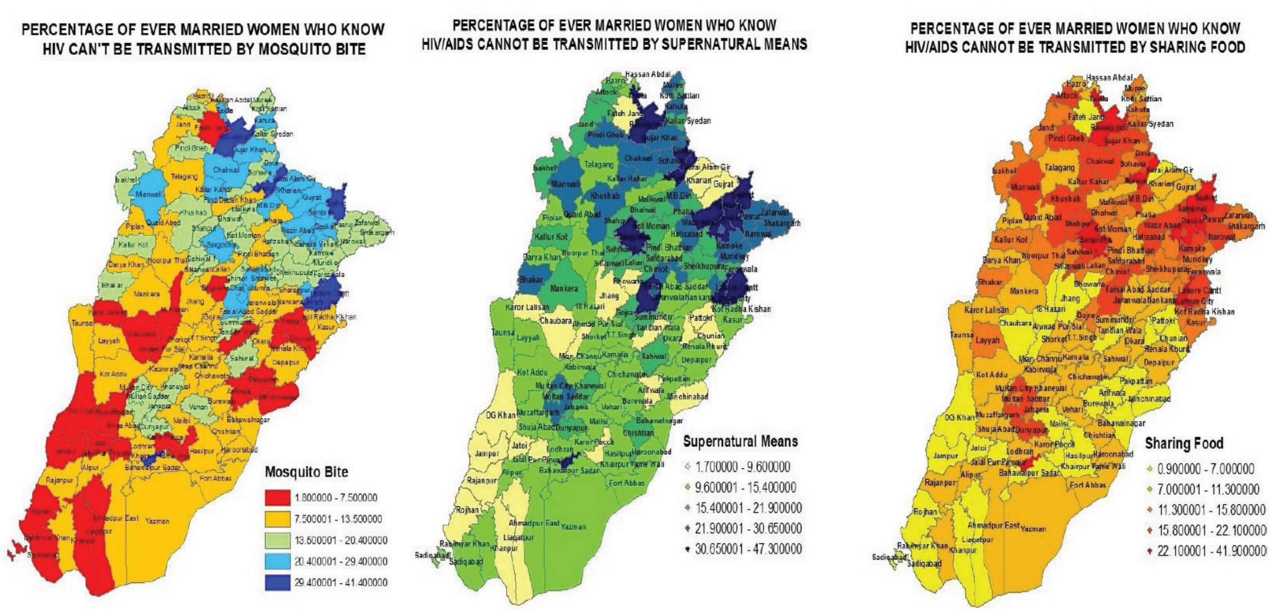

Fig. 4. Maps of misconception indicators about HIV/AIDS

Data generated from Moran's *I* and Geary's c confirmed the presence of spatial clustering. Table 2 shows values generated by Moran's I and Geary's c. The data confirms spatial clustering among the administrative boundaries.

Socio economic determinants of three types of response variable were found by using BHPM. According to Table 3, TB and postnatal are significantly associated with all three misconception factors while U. Area is significant for Mosquito and SN. Similarly, Hp is significantly associated with Mosquito and SF. However, literacy is only significant for SF.

SCBHPMs were used with the binary weight matrix, and weight "1" was assigned to those regions that shared the same boundary and "0" was used for other boundaries. Binary weighting, used with fixed distance, K nearest neighbors, and contiguity spatial relationships, like this study, where all 135 administrative areas (tehsil) are spatially contiguous. Parameter estimates by SCBHPMs analysis (Table 4) shows that more variables are statistically significant as compare to BHPMs. The size, agemean, ownership, TB, adb, and IUD were negatively associated with all three items. Similarly, Matric, safety net, Hp, FS and TV were positively associated with the response variables. Also noticeable is that literacy was only positively significant for SF, while Middle was negatively significant for Mosquito and SF. Furthermore, condom-use was positively associated with SN and SF at a 5% level of significance.

Table 2. Estimates of Spatial Autocorrelation Statistics

Models	MCMC *Moran's I*		MCMC *Geary's c*	
	Statistics	p-value	Statistics	p-value
Mosquito	0.16427	0.00499	0.89872	0.0009
SN	0.17539	0.00099	0.89817	0.08691
SF	0.14686	0.00799	0.94537	0.01998

Table 3. Parameter estimates: Standard error and 95% credibility intervals from Bayesian hierarchical level 2 models

Variables	Estimates (SE) (L-95% CL, U-95% CL)		
	Mosquito	SN	SF
Fixed Part			
(Intercept)	8.62600 (0.12883)	8.27100 (0.13221)	8.07723 (0.13556)
	(1.29915, 16.26183)	(0.43730, 16.77000)	(0.11273, 16.67023)
size	0.17454 (0.00485)	0.22560 (0.00505)	0.18637 (0.00520)
	(-0.16906, 0.46679)	(-0.09286, 0.53740)	(-0.16848, 0.47123)
age_mean	-0.15030 (0.00348)	-0.16400 (0.00355)	-0.19378 (0.00374)
	(-0.34745, 0.05824)	(-0.37020, 0.05799)	(-0.42023, 0.03324)
adb	-0.00266 (0.00011)	-0.00168 (0.00012)	-0.00097 (0.00012)
	(-0.00978, 0.00415)	(-0.00930, 0.00497)	(-0.00831, 0.00686)
literacy	0.01379 (0.00036)	0.01465 (0.00038)	0.02121 (0.00039)
	(-0.00562, 0.03654)	(-0.00928, 0.03708)	(-0.00211, 0.04507)
Primary	-0.02674 (0.00058)	-0.02111 (0.00058)	-0.02115 (0.00058)
	(-0.05877, 0.00757)	(-0.05745, 0.01385)	(-0.05586, 0.01388)
Middle	-0.00744 (0.00118)	-0.01559 (0.00120)	-0.04085 (0.00128)
	(-0.07178, 0.06747)	(-0.09418, 0.05354)	(-0.11632, 0.04084)

Matric	0.02533 (0.00090) (-0.02444, 0.07760)	0.03079 (0.00088) (-0.02672, 0.08168)	0.02988 (0.00091) (-0.02109, 0.09091)
A. Matric	0.02210 (0.00085) (-0.03582, 0.07234)	0.01012 (0.00086) (-0.03991, 0.06355)	0.01095 (0.00089) (-0.04053, 0.06489)
U. Area	0.00737 (0.00019) (-0.00460, 0.01828)	0.01129 (0.00019) (-0.00003, 0.02306)	0.01270 (0.00021) (-0.00014, 0.02496)
ownership	-0.02228 (0.00053) (-0.05352, 0.01314)	-0.01650 (0.00058) (-0.05039, 0.02072)	-0.00730 (0.00059) (-0.04545, 0.02561)
safetynet	0.00645 (0.00025) (-0.00742, 0.02147)	0.00523 (0.00024) (-0.01110, 0.01848)	0.00819 (0.00025) (-0.00624, 0.02455)
TV	0.00410 (0.00026) (-0.01223, 0.01943)	0.00279 (0.00027) (-0.01208, 0.02085)	0.00596 (0.00028) (-0.01047, 0.02242)
Hp	0.14930 (0.00269) (-0.02802, 0.31404)	0.13740 (0.00275) (-0.03021, 0.30900)	0.15160 (0.00288) (-0.02487, 0.34347)
TB	-0.41290 (0.00620) (-0.77958, -0.03275)	-0.42030 (0.00627) (-0.79720, -0.01943)	-0.50852 (0.00654) (-0.90365, -0.10128)
postnatal	0.01378 (0.00020) (0.00182, 0.02537)	0.01438 (0.00020) (0.00263, 0.02598)	0.01396 (0.00021) (0.00128, 0.02704)
FS	0.00730 (0.00057) (-0.02733, 0.04222)	0.00384 (0.00055) (-0.02701, 0.04062)	-0.00645 (0.00054) (-0.04150, 0.02381)
IUD	-0.02433 (0.00101) (-0.09304, 0.04394)	-0.00993 (0.00105) (-0.07186, 0.05669)	0.01033 (0.00110) (-0.06008, 0.07744)
Condom	0.00376 (0.00061) (-0.03229, 0.04113)	0.00721 (0.00066) (-0.03229, 0.04678)	-0.00150 (0.00065) (-0.04156, 0.03813)
Random Part			
District variation	0.10690 (0.00175) (0.00784, 0.21310)	0.08031 (0.00172) (0.00001, 0.17940)	0.03577 (0.00177) (0.00001, 0.15120)
Tehsil variation	0.24470 (0.00150) (0.15770, 0.32680)	0.27640 (0.00165) (0.18380, 0.37310)	0.32980 (0.00202) (0.21560, 0.45870)
DIC	**1067.80400**	**1098.54600**	**1045.91900**
5% level of significance		10% level of significance	

Table 4. Parameter estimates: Standard error and 95% credibility Intervals from conditional Bayesian hierarchical level 2 models

Variables		**Estimates (SE) (L-95% CL, U-95% CL)**	
	Mosquito	**SN**	**SF**
Fixed Part			
(Intercept)	14.0717 (8.892e-03) (9.1948, 18.7468)	15.5480 (8.372e-03) (11.1897, 20.3625)	15.6793 (8.101e-03) (11.5842, 20.4434)
size	-0.3921 (4.424e-04) (-0.6249, -0.1307)	-0.3408 (4.398e-04) (-0.6085, -0.1152)	-0.3502 (4.642e-04) (-0.6236, -0.1134)
age_mean	-0.1795 (2.335e-04) (-0.3051, -0.0393)	-0.2337 (2.280e-04) (-0.3624, -0.1111)	-0.2246 (1.998e-04) (-0.3298, -0.1138)

adb	-0.0071 (7.547e-06) (-0.0111, -0.0027)	-0.0047 (8.547e-06) (-0.0094, 0.0003)	-0.0053 (7.944e-06) (-0.0100, -0.0011)
literacy	-0.0026 (2.486e-05) (-0.0151, 0.0116)	-0.0137 (2.450e-05) (-0.0248, 0.0012)	0.0135 (1.880e-05) (0.0031, 0.0236)
Primary	0.0005 (4.744e-05) (-0.0254, 0.0255)	0.0180 (4.826e-05) (-0.0081, 0.0439)	0.0120 (4.457e-05) (-0.0095, 0.0387)
Middle	-0.0556 (7.580e-05) (-0.0962, -0.0115)	-0.0459 (9.035e-05) (-0.0982, 0.0004)	-0.0444 (7.609e-05) (-0.0831, -0.0005)
Matric	0.0746 (6.876e-05) (0.0397, 0.1142)	0.0922 (5.591e-05) (0.0601, 0.1222)	0.0982 (6.590e-05) (0.0618, 0.1362)
A. Matric	0.0116 (7.112e-05) (-0.0315, 0.0479)	0.0216 (7.920e-05) (-0.0222, 0.0636)	0.0090 (7.375e-05) (-0.0332, 0.0504)
U. Area	-0.0011 (1.121e-05) (-0.0073, 0.0050)	0.0016 (1.173e-05) (-0.0048, 0.0080)	0.0014 (1.155e-05) (-0.0052, 0.0078)
ownership	-0.0421 (3.854e-05) (-0.0619, -0.0209)	-0.0449 (3.365e-05) (-0.0617, -0.0255)	-0.0478 (3.964e-05) (-0.0689, -0.0256)
safetynet	0.0230 (2.245e-05) (0.0107, 0.0356)	0.0268 (2.287e-05) (0.0137, 0.0392)	0.0281 (2.267e-05) (0.0149, 0.0398)
TV	0.0267 (2.309e-05) (0.0139, 0.0392)	0.0219 (2.145e-05) (0.0099, 0.0331)	0.0212 (2.286e-05) (0.0094, 0.0353)
Hp	0.2121 (2.271e-04) (0.0881, 0.3389)	0.1293 (2.328e-04) (0.0045, 0.0331)	0.1424 (2.337e-04) (0.0153, 0.2745)
TB	-0.7863 (5.100e-04) (-1.0731, -0.5055)	-0.6687 (5.297e-04) (-0.9779, -0.3936)	-0.6783 (4.843e-04) (-0.9462, -0.4089)
postnatal	0.0037 (1.519e-05) (0.0050, 0.0119)	0.0040 (1.446e-05) (0.0043, 0.0118)	0.0047 (1.471e-05) (0.0034, 0.0130)
FS	0.0419 (5.168e-05) (0.0163, 0.0733)	0.0621 (5.398e-05) (0.0328, 0.0923)	0.0593 (5.364e-05) (0.0310, 0.0908)
IUD	-0.0548 (7.818e-05) (-0.0988, -0.0112)	-0.0661 (6.890e-05) (-0.1037, -0.0272)	-0.0698 (7.495e-05) (-0.1116, -0.0280)
condom	0.0266 (5.044e-05) (-0.0017, 0.0555)	0.0503 (4.873e-05) (0.0252, 0.0790)	0.0491 (5.113e-05) (0.0191, 0.0755)
Random Part			
District variation	0.7901 (0.001028) (0.4121, 1.5308)	1.0077 (0.001366) (0.5109, 2.0013)	0.9848 (0.001295) (0.5071, 1.9158)
Tehsil variation	0.3385 (0.0002451) (0.2329, 0.5028)	0.3830 (0.0002607) (0.2691, 0.5572)	0.3835 (0.0002645) (0.2684, 0.5606)
Spatial variation	0.5588 (0.000731) (0.1572, 0.9282)	0.5306 (0.0007354) (0.1473, 0.9194)	0.5310 (0.0007344) (0.1457, 0.9182)
DIC	**352.7824**	**548.7202**	**832.2264**

SCBHPM provided the estimates after incorporating spatial and hierarchical effects. The results show an intense decrease in DIC at level 2, which is evidence that the model incorporating the spatial and hierarchical effects provides the better estimate.

The data also show the importance of comprehensive knowledge of HIV transmission among EM women age 15-49 in Punjab, Pakistan. Misconceptions about HIV transmission can be observed everywhere in the world. This study's results concur with other similar transmission studies. A study by (Jittimanee *et al*., 2009) found that many Thai TB patients thought that HIV transmission was possible by mosquito bite and by sharing food with an HIV positive person. In another study, 74.0% of the HIV-infected elderly female patients in a South African hospital thought that they could get the virus through mosquito bites which had previously fed off an HIV infected person (Rauf *et al*., 2010).

The study data show that the average age of ever-married women had an inverse relationship with the misconception indicators, which was similar to results in Rauf *et al*. (2010). Women with higher educational levels were less likely to have misconceptions. (Carey *et al*., 2000). The statistical results also reveal that the number of households which have TVs had a positive relationship with the misconception indicators, which is similar to the findings of Mondal *et al.* (2016).

Two types of contraceptive measures were used in this study in order to find correlation. The study shows a strong relationship with the misconception indicators and types of contraceptive measures. The study strongly showed positive relations among condom used as a contraceptive measure and misconception indicators. However, IUD-use showed a negative association with the response variables.

4. Conclusions

Throughout this paper we have stressed the importance of spatial autocorrelation in traditional statistical modeling. We provided an overview of the techniques that may be used to quantify the effects that explanatory variables have on the spatial distribution of an outcome of interest. This is combined with the hierarchical effect. It was shown that accounting for spatial dependence provides useful benefits in term of enhancing understanding of the factors associated with the distribution of disease/outcome of interest. The analysis shows that the regression coefficients from the models that account for spatial dependence were less precise (compared with those that ignore it). This means that the null hypothesis is less likely to be rejected when it is true (Type I error). The results of the model accounting spatial effect at level two showed that 33.11%, 27.62% & 27.96% of variation were observed due to the spatial effects in Item A, Item B and Item C, respectively.

This study also illustrates the benefits of spatial analysis in determining factors associated with the misconception indicators regarding HIV/AIDS. The statistical results suggested that the predicting misconception indicators about HIV/AIDS on likely covariates—ignoring the spatial autocorrelation by using BHPM—provided fewer statistically significant variables as compared to the model that accounts for the spatial dependences (SCBHPM). This finding supported the argument in Rashid and Chand (2016). These researchers assumed the independence of response variables may result in misleading findings.

ACKNOWLEDGEMENTS

The authors are thankful to the Punjab Bureau of Statistics for providing the data set for this analysis.

References

Besag, J., York, J. & Mollié, A., 1991. Bayesian image restoration with two applications in spatial statistics. Annals of the Institute of Statistical Mathematics, **43**(1), pp. 1-20.

Best, N.G. *et al*., 1999. Bayesian Models for Spatially Correlated Disease and Exposure Data. In: J. M. Bernardo, J. O. Berger, A. P. Dawid & A. F. M. Smith, eds. Bayesian Statistics 6. s.l.:Oxford University Press.

Best, N., Richardson, S. & Thomson, A., 2005. A Comparison of Bayesian Spatial Models for Disease Mapping. Statistical Methods in Medical Research, Volume **14**, 35-59.

Breslow, N.E. & Clayton, D.G., 1993. Approximate inference in generalised linear mixed models.. Journal of the American Statistical Association, **88**(421), 9-25.

Carey, M. *et al*., 2000. Using information, motivational enhancement, and skills training to reduce the risk of HIV infection for low-income urban women: a second randomized clinical trial. Health Psychology, 3-11.

Clayton, D. & Bernardinelli, L., 1992. Bayesian methods for mapping disease risk. In: Geographical and Environmental Epidemiology: Methods for Small Area Studies. Oxford: Oxford University Press.

Draper, D., 1996. Discussion of the Paper by Lee and Nelder. Journal of the Royal Statistical Society, Series B, 662-663.

Finch, W.H., Bolin, J.E. & Kelley, K., 2014. Multilevel Modeling Using R. London: CRC Press.

GBD, 2016. Global Burden of Disease Study 2016, s.l.: Institute for Health Metrics and Evaluation.

Ghosh, M., Natarajan, K., Stroud, T.W.F. & Carlin, B.P., 1998. Generalized linear models for small-area estimation. Journal of the American Statistical Association, **93**(441), 273-282.

Hailu, T.G., 2016. A Multilevel Modeling Analysis of the Determinants and Cross-Regional Variation of the HIV Testing in Ethiopia: Ethiopian DHS 2011. Journal of Biometrics & Biostatistics.

Hoef, J.M.V., Peterson, E. & Theobald, D., 2006. Spatial statistical models that use flow and stream distance. Environmental and Ecological Statistics, December, **13**(4), 449-464.

Jarup, L. *et al*., 2002. Geographical epidemiology of prostate cancer in Great Britain. International Journal of Cancer, 695-699.

Jittimanee, S.X. *et al*., 2009. Social Stigma and Knowledge of Tuberculosis and HIV among Patients with Both Diseases in Thailand. PLoS One.

Kleinschmidt, I., Sharp, B., Mueller, I. & Vounatsou, P., 2002. Rise in malaria incidence rates in South Africa: A small-area spatial analysis of variation in time trends. American Journal of Epidemiology, **155**(3), 257-264.

Lawson, A.B., 1994. On using spatial Gaussian priors to model heterogeneity in environmental epidemiology.. Statistician, Volume **43**, 69-76.

Lee, D., 2017. CARBayes version 5.0: An R Package for Spatial Areal Unit Modelling with Conditional Autoregressive Priors. June.

Lee, D. & Shaddick, G., 2010. Spatial modeling of air pollution in studies of its short-term health effects. Biometrics, December, Volume **4**, 1238-1246.

Leroux, B.G., Lei, X. & Breslow, N., 2000. Estimation of Disease Rates in Small Areas: A new Mixed Model for Spatial Dependence. In: D. B. M. Elizabeth Halloran, ed. Statistical Models in Epidemiology, the Environment and Clinical Trials. New York: Springer-Verlag.

Letamo, G., 2007. Misconceptions about HIV prevention and transmission in Botswana. African Journal of AIDS Research, 193-198.

Maheswaran, R., Strachan, D.P., Dodgeon, B. & Best, N.G., 2002. A population-based case-control study for examining early life influences on geographical variation in adult mortality in England and Wales using stomach cancer and stroke as examples. International Journal of Epidemiology, **31**(2), 375-382.

Manton, K.G., Woodbury, M.A. & Stallard, E., 1981. A variance components approach to categorical data models with heterogenous mortality rates in North Carolina counties. Biometric, 259-269.

Marshall, R.J., 1991. Mapping disease and mortality rates using empirical Bayes estimators.. Applied Statistics, **40**(2).

Martin, A.D., 2003. Bayesian Inference for Heterogeneous Event Counts. Sociological Methods & Research, August. 30-63.

Mizanur, M.R., Kabir, M. & Shahidullah, M., 2009. Adolescent Knowledge and Awareness about HIV/AIDS and Factors Affecting them in Bangladesh. Journal of Ayub Medical College Abbottabad, 3-6.

Mondal, N.I., Hoque, N., Chowdhury, R.K. & Hossain, S., 2016. Factors Associated with Misconceptions about HIV Transmission among Ever-Married Women in Bangladesh. Japanese Journal of Infectious Diseases, Volume **68**, 13-19.

Moran, P.A., 1950. Notes on Continuous Stochastic Phenomena. Biometrika, Volume **37**, 17-23.

Mueller, I., Betuela, I. & Hide, R., 2002. Regional patterns of birthweights in Papua New Guinea in relation to diet, environment and socio-economic factors.. Annals of Human Biology, **29**(1), 74-88.

Ochako, R. *et al*., 2011. Trends and Determinants of Comprehensive HIV and AIDS knowledge amoung Urban Young Women in Kenya. AIDS Research and Therapy, 8-11.

Pfeiffer, D. *et al*., 2008. Spatial Analysis in Epidemiology. s.l.:Oxford University Press, UK.

Pugliatti, M., Solinas, G., Sotgiu, S. & Castiglia, P., 2002. Multiple sclerosis distribution in Northern Sardinia – spatial cluster analysis of prevalence. Neurology, **58**(2), 277-82.

Rashid, M. & Chand, S., 2016. Spatial Analysis and Modeling of Infant Mortality Rate Using Conditional Autoregressive Model - A Case Study of Punjab, Pakistan. Pakistan Journal of Medical Research, Volume **55**.

Rauf, W. *et al*., 2010. Knowledge of and misconceptions about the spread and prevention of HIV infection among older urban women attending the Tshwane District Hospital, South Africa. South African Family Practice, 142-148.

Schober, E., Rami-Merhar, B., Waldhör, T. & Karimian-Teherani, D., 2001. Regional distribution of childhood onset diabetes mellitus in Austria – analysis of national registration from 1989 to 1999. Wiener klinische Wochenschrift, 491-495.

Stevenson, M.A. *et al*., 2005. Area-level risks for BSE in British cattle before and after the July 1988 meat and bone meal feed ban.. Preventive Veterinary Medicine, Volume **69**, 129-144.

Tenkorang, E., 2013. Myths and misconceptions about HIV transmission in Ghana: what are the drivers?. Culture, Health and Sexuality, 296-310.

Toledano, M.B. *et al*., 2001. Spatial variation and temporal trends of testicular cancer in Great Britain. British Journal of Cancervolume , Volume **84**, 1482-1487.

Tribune, T.E., 2011. HIV/AIDS: Pakistan has one of Asia's highest HIV prevalence rates. [Online] Available at: http://tribune.com.pk/story/240379/hivaids-pakistan-has-one-of-asias-highest-hiv-prevalence-rates/

Tsutakawa, R.K., 1988. A mixed model for analyzing geographic variability in mortality rates.. Journal of the American Statistical Association, **83**(40).

UNAIDS, 2005. National Aids Programmes: A gude to indicators For Monitoring and Evaluating National Antiretroviral Programmes, s.l.: Joint United Nations Programme on HIV/AIDS (UNAIDS), World Health Organization.

Vlachonikolis, I., Aletra, T. & Georgoulias, V., 2002. Incidence of breast cancer on Crete, 1994-1995. European Journal of Cancer, **38**(4), 574-577.

Weiss, R., 1993. How does HIV cause AIDS?. Science, May. 260.

Zeilhofer, P., Beatriz, E. & Lima, G.A.R., 2006. Spatial Patterns of Water Quality in the Cuiabá River Basin, Central Brazil. Environmental Monitoring and Assessment, December, **123**(1-3), 41-62.

Submitted : 16/01/2019

Revised : 11/04/2019

Accepted : 14/04/2019

العوامل الاجتماعية والاقتصادية التي تؤدي لسوء الفهم حول فيروس نقص المناعة البشرية / الإيدز بين النساء المتزوجات في البنجاب: مقارنة بين نموذج بواسون الهرمي غير المكاني والمكاني

ماجدة راشد، سهيل شاند
جامعة البنجاب، باكستان

الملخص

تُعد مكافحة فيروس نقص المناعة البشرية / الإيدز الهدف السادس من الأهداف التنموية للألفية، وقد أصبحت مصدر اهتمام متزايد بالصحة في باكستان. تُعد باكستان واحدة من إحدى عشرة دولة في منطقة آسيا والمحيط الهادئ والتي يعاني معظم سكانها من فيروس نقص المناعة البشرية. والطريقة الوحيدة لمحاربة هذا الفيروس هي نشر الوعي والمعرفة الدقيقة عن وسائل انتقاله بين الناس وخاصة النساء في سن الإنجاب، أي من 15 إلى 49 سنة. ترتبط الوقاية من الفيروس وسوء الفهم حول وسائل انتقاله ببعضهما البعض. لذلك، تهدف هذه الدراسة إلى تحديد التوزيع المكاني لثلاثة أنواع من عوامل الاعتقاد الخاطئ عن وسائل انتقال فيروس نقص المناعة البشرية (أي انتقاله عن طريق لدغة البعوض والوسائل الخارقة للطبيعة وتبادل المواد الغذائية مع شخص مصاب بالفيروس). توفر هذه الدراسة أيضاً العوامل الاجتماعية والاقتصادية الأساسية لإيقاف الاعتقاد الخاطئ حول وسائل انتقال فيروس نقص المناعة البشرية / الإيدز وستساعد في الحد من انتشار الوباء في باكستان. وقد تم تطبيق نموذج Bayesian الهرمي المكاني وغير المكاني، وأظهرت النتائج أن النماذج الهرمية البيزية ذاتية الانحدار (النموذج المكاني) هي نماذج الانحدار الأكثر ملاءمة في وجود الاعتماد المكاني.