Bayesian spatial analysis and disease mapping of leishmaniasis outbreak in Afghanistan, 2003-2009

Oyelola A. Adegboye
Department of Science and Mathematics,
The American University of Afghanistan, Kabul
aadegboye@auaf.edu.af

Abstract

A number of methods from epidemiology, geo-statistics and small area modeling are available for the analysis of disease incidence rates. A simple model assumed a Poisson log-linear relationship between disease rates and other covariates while using random effects to model extra variation in Poisson model. Bayesian modeling has the advantage of allowing the exact analysis of random effect and coefficient models. This study estimates of the incidence of leishmaniasis at the provincial level. Leishmaniasis is the third most common vector-borne disease and it is contracted through bites from sand flies. There are about 250,000 estimated new cases of incidence of cutaneous leishmaniasis in Afghanistan and 67,000 cases in Kabul, thus making the city having largest incidence worldwide. Spatial hierarchical Bayesian models were used to model the over-dispersion of the relative risk of leishmaniasis and take into account the risk dependence between close areas. Bayesian Hierarchical Poisson models were constructed to quantify the influence of hypothesized risk factors on provincial level relative risk of the disease. To take into account the lack of independence of the risk between areas, random components were used. Also geographical constraints operating under the outbreak of leishmaniasis in Afghanistan was explored. The final model include altitude and two random components (intercept and slope) using a conditional autoregressive (CAR) prior and has a deviance information criterion (DIC) of 247.761. Spatial scan statistics confirmed cluster in North-Eastern and South-Eastern region of Afghanistan with a p-value of <0.0001. The results confirmed geographical heterogeneity of leishmaniasis and thus will allow governmental and non-governmental organization to make good health policies and choose areas to implement control measures against Leishmaniasis in a more efficient way.

Key words: Spatial hierarchical Bayesian analysis, Leishmaniasis, Spatial distribution, Relative risk, disease mapping

Background

Leishmaniasis is the third most common vector-borne disease and very important protozoan infection. It is contracted through bites from sand flies and can result in chronic and non healing sores. This mostly occurs on any exposed skin and can be disfiguring and painful. The burden of the disease is overwhelming; the psychological effect can be disturbing. In some society, women infected with this disease are stigmatized and deemed unsuitable for marriage and motherhood (Reithinger, et al., 2005). The World Health Organization (WHO) in 2000 reported that there are estimated 1.5 million annual cases of leishmaniasis worldwide and Afghanistan, Algeria, Saudi Arabia, Brazil, Iran, Iraq, Peru and Syria accounted for over 90% of the cases (Michael, et al., 2008).

There are about 250,000 estimated new cases of incidence of cutaneous leishmaniasis in Afghanistan and 67,000 cases in Kabul, thus making the city having largest incidence worldwide (Reithinger, et al., 2003). Humanitarian relief efforts began in 2002 after the fall Taliban in Afghanistan and more than US $2 billion having been spent on health sector, nevertheless Afghanistan health indicators has witnessed very little improvement (Toby, et al., 2006; Reithinger and Paul, 2007).

A number of methods from epidemiology, geostatistics and small area modeling are available for the analysis of disease incidence rates. A simple model will assume a Poisson log-linear relationship between disease rates and other covariates while using random effects to model extra variation in Poisson model. Bayesian modeling has the advantage of allowing the exact analysis of random effect and coefficient models. The impact of environmental influences on leishmaniasis cannot be ruled out and human activities on the environment play a significant role in the dispersion and vectors thereby changing the geographical distribution of the disease.

The major interest of this study is modelling the transmission dynamics of leishmaniasis, quantification and predicting the disease incidence rate across provinces. Estimates of the incidence of leishmaniasis at the provincial level were provided and also, explored what geographical constraints operate under the outbreak of leishmaniasis in Afghanistan. Spatial hierarchical Bayesian models were used to model the over-dispersion of the relative risk of leishmaniasis and take into account the risk dependence between close areas. Bayesian Poisson models were constructed to quantify the influence of hypothesized risk factors on district/provincial level relative risk of the disease. To take into account the lack of independence of the risk between areas, a random component will be used.

Leishmaniasis incidence in Afghanistan

Data on cases of leishmanaisis incidence reported to Health Management Information System (HMIS) and Ministry of Public Health (MoPH) in Afghanistan for the period of 2003 to 2009 were collected and aggregated at province level. Data from HMIS and MoPH included a total of 148,564 new cases leishmaniasis observed annually in Afghanistan over the period 2003-2009. Also collected were population sizes for this period from Central Statistics Organization (CSO) of Afghanistan and the latitude and longitude of the central district was supplied by Afghanistan Information Management Services (AIMS). There are 34 provinces in Afghanistan; cases of leishamaniasis were not available in some provinces (figure 1). The incidence of leishmaniasis disease in Afghanistan has been on the rise especially in Kabul (figure 2). While the number cases of the disease reported across the provinces has not been consistent, Kabul province has recorded a steady increase since 2003 and accounted for about 30% of the total cases (figure 2).

Model Based analysis

Exploratory data analysis on incident leishmaniasis in Afghanistan showed geographical disparity in the occurrence of the disease (figure 1). The standard incidence rate (SIR) was calculated for each province of Afghanistan and were mapped in figure 3. The map showed areas with high and low risk, with dark regions indicating high risk of leishmaniasis whereas the light regions indicate low risk.

Standard incidence rates suffers a set back because of small area count, there are extreme rates where the populations are smallest and geographically close areas tend to have similar disease rates (Clayton and Bernardinelli, 1992). Leishamaniasis disease is a non-contagious vector borne disease and the observed count of cases at provincial level is assumed to occur independently and follows a Poisson distribution.

In order to overcome the drawbacks of SIR, a model based on Spatial Hierarchical Bayesian (SHB) analysis which will combine the specific provincial rate and the influence of neighborhood was formulated. Furthermore, the altitude of the province capital was included as covariate in the SHB analysis. Breslow and Clayton, 1993 proposed an approach of using random effect and random coefficients in generalized linear mixed models (GLMMs) to model extra variation in the Poisson model.

Spatial Hierarchical Bayesian Modeling

When the observed data are scarce, a maximum likelihood (ML) approach to the above model may lead to unstable and largely uninformative ML estimates of the area-specific linear trends due to Poisson sampling variation (Bernardinelli et al., 1995). Bayesian modeling has the advantage of allowing the exact analysis of random effect and coefficient models (Lawson and Zhou, 2005). Several authors have prescribed the use of Spatial Hierarchical Bayesian models for different disease epidemics, (Allepux et al., 2007: Lawson and Zhou, 2005: Stevenson et. al., 2005 and Durr et al., 2005). SHB Poison models were used to model the over-dispersion of the parameter of interest (SIR) and take into account the risk dependence between close areas. Bayesian Poisson models were constructed to quantify the influence of hypothesized risk factors on province level relative risk of leishmaniasis disease. To take into account the lack of independence of the risk between provinces, a random component will be used.

All models were implemented in WINBUGS Software using Gibbs sampling (Lunn, et al., 2000), this allow the iterative exploration of the posterior surface and leads to a set of parameter values rather than a single value which is typical of ML methods (Lawson and Zhou, 2005).

Prior distribution

When modeling using Bayesian framework, one needs to specify a *prior* distribution for the observed data. Several assumption for prior distributions were explored in this study, namely flat distribution thus providing a non-informative prior, gamma distribution, conditional autoregressive (CAR) distribution. The prior in the spatial model was similar to that proposed by Durr et al., 2005, with alpha0 assumed to follow a flat distribution, and unstructured variability parameter was assumed to follow a normal distribution with mean 0 and a precision variable while the structured variability term was allowed to depend on the neighbors, this is sometimes called convolution Gaussian distribution or intrinsic Gaussian CAR (Besag et al., 1991).

Model selection and assessment

Different classes of Bayesian Poison hierarchical models of increasing complexity were formulated; the models follow closely approaches by Lawson and Zhou, 2005 and Stevenson et al., 2005. This model includes a random component with and without spatial structure. Several adjustments were made to the model to give rise to what was called spatially smoothed and non-spatially smoothed model and explored with varying prior distributions for the random effect.

For all model three chains were ran to help assess convergence and was visualized with time series plots and Gelman-Rubin Statistics. Deviance information criterion (DIC) was used to compare all models and model with the smallest DIC is said to be the "best fit".

Areas of high risk

Several tests are available for spatial randomness that adjusts for an uneven background population. Such test statistics are used to test whether or not the geographical distribution of disease is random. Previous studies have shown that the spatial scan statistic has good power in detecting hot spot clusters (Kulldurf, 1997). SaTScan is a software program written to implement the scan statistic; it can be used to find clusters in space and/or time (Kulldurf, 1997).

Results

Cases of leishmaniasis were collected from 34 provinces in Afghanistan over a period starting from 2003 to 2009. A total of 148,564 new cases were reported to HMIS and MoPH during this period with Kabul recording the highest number (30%), followed by Kandahar (13%) and Balkh (10%) of all new cases. As indicated by the graph right of figure 4, there is a weak association between leishmaniasis cases and altitude of the provincial Centrum. North Eastern region of Afghanistan recorded the highest incidence of the disease, while South Western region recorded no cases of the disease (figure 1). The expected number of cases was estimated for each province and mapped as displayed in figure 5.

The standard incidence rate provides an assessment of excess risk in a province that would be expected. The map of the standard incidence rate also indicated geographical disparities in the risk of the disease; the North Eastern region has high risk of the disease more than the rest of the country. These crude rates (SIR) must be interpreted carefully and maybe misleading because they are influenced by the population size of the regions and neighboring provinces.

As mentioned before, SIR has drawbacks because it assumed provinces as independent. However, in a spatial point of view, this is not the case and we are more interested in the more global spatially distribution of the number of cases of leishmaniasis. A spatial hierarchical Bayesian analysis with random components to take into account the lack of independence of the risk between provinces was formulated.

Several models with different complexities were explored that include random component that is non-spatially structured heterogeneity and spatial structured heterogeneity. The model follow that of Besag et al., 1991: the unstructured variability parameter (ui) was assumed to follow a normal distribution, and the structured variability term followed the multivariate normal conditional autoregressive (CAR) distribution. The non-spatially smoothing adjusted the relative risk estimates for province with low numbers towards the overall mean, while including spatially structured heterogeneity term was to condition the smoothing on neighboring provinces. The results of the models are discussed below:

$$\log(\mu_i) = \log(E_i) + \alpha_0 + (\beta_1 + b_{1i}) * \text{Alt} + u_i + v_i \qquad 5$$

All models were run in WINBUGS via Gibbs sampling and the posterior distributions were estimates. Three chains were run and of 15,000 iterations and the convergence were checked by trace plot and visualized by Gelman and Rubin plot. Model selection was done by looking at the Deviance Information Criterion (DIC) to assess the goodness-of-fit and model complexity (Spiegelhalter et al., 2002). Table 1 summarizes the models with their level of complexity and value of DIC (the smaller the better). The results for the final models with the smallest DIC were selected for presentation in table 2.The models with smaller DIC values are those with altitude as covariate and with random components (Table 1). The final model selected as the best include altitude as covariate and two random components that is, random intercept (unstructured and structured) and random slope. The chosen model has a DIC value of 247.761 and allow for over-dispersion and spatial correlation through the use of conditional autoregressive prior. The unstructured heterogeneity term (U_i) followed a normal distribution with a mean of 0 and variance σ^2, while the structured heterogeneity term (V_i), estimated using a normal distribution with a provincial dependent mean and weighted variance (weighted by adjacent provinces). In this model SIR are smoothed locally towards the mean risk in the set of neighbouring areas (Lorenzo-Luaces et al., 2009).

Three chains were run of 5000 iterations as burn-in phase and 15000 iterations were run to estimate the variables. Convergence was visually assessed by Gelman and Rubin plot and convergence was satisfied. This model has a DIC value of 247.761, table 2 showed the summary statistics for the precision terms and posterior summaries. The relative risk (RR) from the final model is displayed in Figure 6, higher RR was observed in North-Eastern, central and South-Eastern.

Discussion and Conclusion

The spatial scan statistic of Kulldorf was used for cluster detection and test local clusters, this confirm the findings earlier using Bayesian hierarchical analysis. The results from spatial scan statistics, eight provinces (Logar, Pakya, Kabul, Parwan, Khost, Paktika, Kapisa, Wardak) identified as primary cluster and another two (Kandahar, Balkh) has secondary cluster. These provinces are mostly from North-Eastern and South-Eastern and are term regions with high risk of the disease with a statistically significant p-value of <0.0001. The clustered regions were as displayed in Figure 6.

Environment plays an important part in transmission and occurrence of vector borne diseases, with areas in close proximity to each other having similar risk. The use of spatial statistics and geographical information system in the study of geographical heterogeneity in hypothesize of risk lieshmaniasis in Afghanistan is very crucial.

The model in this study suggests excess risk of the disease in the North-Eastern and South-Eastern region of Afghanistan. This model includes non-spatially structured model with unstructured heterogeneity term with a normal prior distribution and gamma hyperpriors for the precision terms and spatially structured term with CAR (allowing for spatial dependencies in the estimation of relative risks) priors plus a random slope. Further epidemiological analysis that includes additional demographic and environmental variables could be explored to obtain a more consistent explanation.

The results confirmed geographical heterogeneity of leishmaniasis and thus will allow governmental and non-governmental organization to make good health policies and choose areas to implement control measures against Leishmaniasis in a more efficient way.

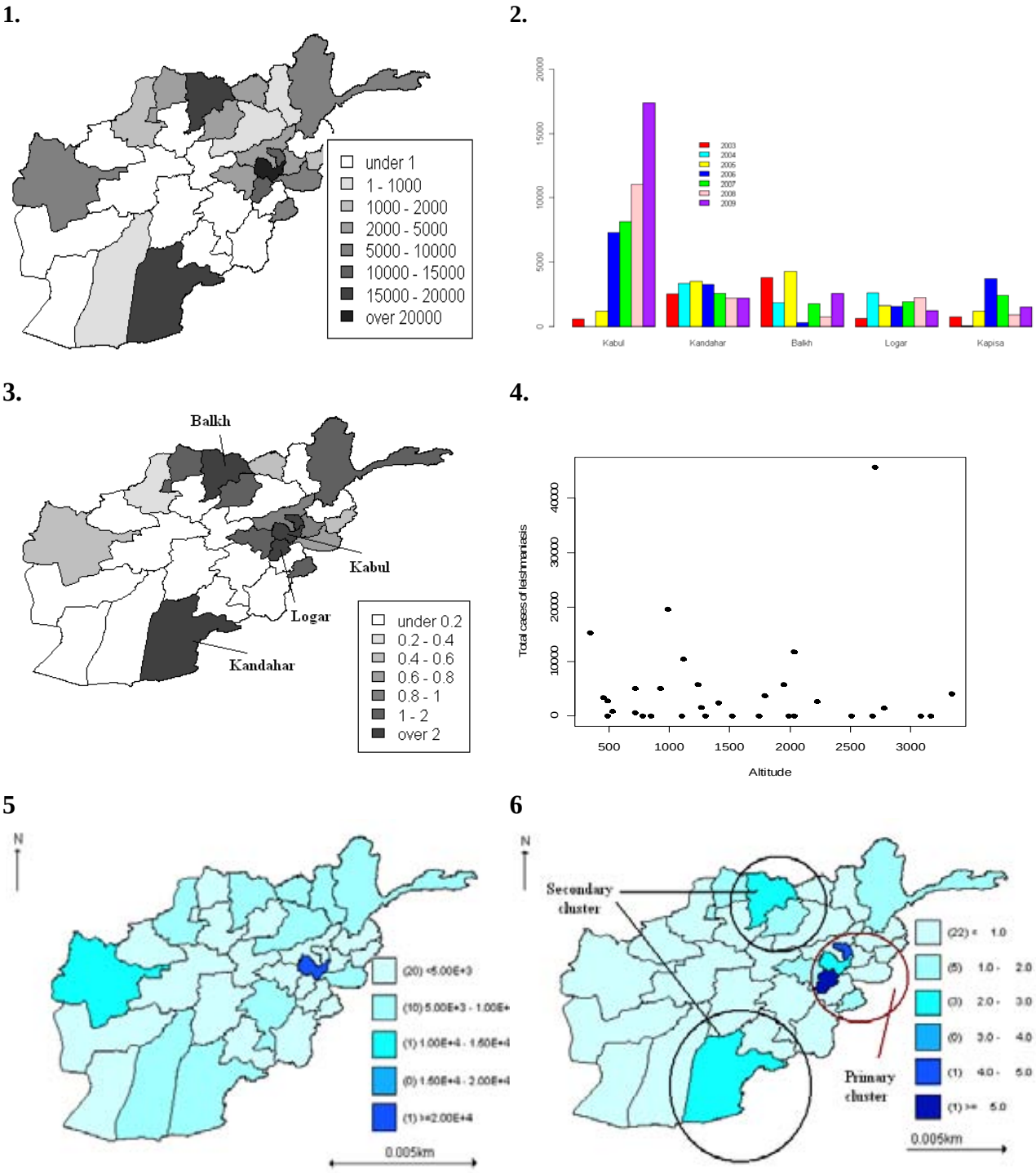

Figure1.Distribution of cases of leishmaniasis incidence in Afghanistan 2003-2009: 2. Trend of reported cases of leishmaniasis disease in five Afghanistan provinces with highest incidence during the period of 2003 to 2009: 3. Standard incidence rate of leishmaniasis in Afghanistan during the period 2003 to 2009: 4. Scatter plots of total cases of leishmaniasis against altitude: 5. Expected incidence of Leishmaniasis in Afghanistan during the period 2003-2009.6: Relative Risk estimated by Hierarchical Bayesian Model for Leishmaniais cases in Afghanistan from 2003 to 2009 with non-structured and spatial random intercept and random slope controlling for Altitude and Population.

Table 1. Summary of results of Hierarchical Bayesian model from WinBUGs with different complexities (Bold face are explored further)

Model	Description	Dbar	Dhat	pD	DIC
Non-spatial random effects:					
Model 1	**Non-structured random intercept**	**245.357**	**201.060**	**44.297**	**289.653**
Model 2	Altitude with non-structured random intercept	254.126	201.025	53.101	307.227
Spatial random effects:					
Model 3	**Altitude with non-structured & spatial random intercept**	**234.162**	**200.808**	**33.353**	**267.515**
Model 4	Spatial random intercept	53654.200	206.538	53447.600	107102.000
Model 5	Non-structured & spatial random intercept	261.268	200.906	60.362	321.631
Model 6	Altitude with spatial Random intercept	71671.000	4662.080	67008.900	138680.000
Model 7	**Altitude with non-structured & spatial (random intercept & slopes)**	**224.612**	**201.463**	**23.149**	**247.761**

Table 2. Posterior summary of results of Hierarchical Bayesian model from WinBUGs : non-spatial regression models; spatial regression model with random intercept only and spatial regression with both random intercept (non-structured & spatial) and slope models

Model	Mean	Standard error	MC error	Credible interval	
				2.5%	97.5%
Intercept	-0.1426	0.1825	0.0125	-0.4883	0.2016
Altitude	480.2533	679.0460	6.3778	-0.0001	0.0002
Variance of random intercept (non-spatial)	19770	7476	107.6000	8349	36120
Variance of random intercept (spatial)	208	489.1000	23.7200	1.7770	1389
Variance of random slope	1176	1629	74.8300	21.4100	5641

REFERENCE

1. Allepuz , A., Lopez-Quilez ,A., Forte, A., Fernandez, G. & Casal, J. (2007). Spatial analysis of bovine spongiform encephalopathy in Galicia, Spain (2000-2005). *Preventive Veterinary Medicine.* 79,174-185.
2. Bernardinelli, L., Clayton, D. & Songini, M. (1995). Bayesian Analysis of Space-Time Variation Disease Risk. *Statistics in Medicine*. 14,2433-2443.
3. Besag, J., York, J, Mollie, A., (1991). Bayesian image restoration with two applications in spatial statistics (with discussion). *Annals of Institute Statistical Mathematics,* 43,1-59
4. Breslow, N and Clayton, D. G. (1993). Approximate inference in generalized linear mixed models. *Journal of American Statistical Association,* 88,9-25.

5. Clayton, D. and Bernardinelli, L. (1992) Bayesian methods for mapping disease risk. In: G*eographical and environment epidemiology: methods for small area studies.* eds Elliott, P., Cuzick, J., English, D. and Stern, R. Oxford University Press, Oxford, 205-220.
6. Durr, P. A., Tait, N. & Lawson, A. B. (2005). Bayesian hierarchical modeling to enhance the epidemiological value of abattoir surveys for bovine fasciolosis. *Preventive Veterinary Medicine*. 71,157-172.
7. Lawson, A. B. & Zhou, H. (2005). Spatial statistical modeling of disease outbreaks with particular reference to the UK foot and mouth disease (FMD) epidemic of 2001. *Preventive Veterinary Medicine* 71,141-156
8. Lawson A. B., Browne, W. J., Vidal Rodeiro, C. J. (2003). Disease mapping with WINBUGS and MLwin. Wiley, Chichester.
9. Lorenzo-Luaces Alvarez P, Guerra-Yi ME, Faes C, Galán Alvarez Y, Molenberghs G. (2009). Spatial analysis of breast and cervical cancer incidence in small geographical areas in Cuba, 1999-2003. *European Journal of Cancer Prevention*, 18: 395-403
10. Lunn, D. J., Thomas, A., Best, N. and Spiegelhalter, D. (2000) WinBUGS -- a Bayesian modeling framework: concepts, structure, and extensibility. *Statistics and Computing*, 10:325-337.
11. Kulldorff, M. (1997). A spatial scan statistic. *Communication in Statistics-Theory Methods.* 26, 1481–96.
12. Michael Fauldea, Joachim Schraderb, Gerhard Heyl, Mohammed Arnirih, Achirn Hoerauff, (2008). Zoonotic cutaneous leishmaniasis outbreak in Mazar-e Sharif, northern Afghanistan: An epidemiological evaluation. *International Journal of Medical Microbiology*. 298:543-550
13. Moran, P. A. P. (1950). Notes on continuous stochastic phenomena. *Biometrika*. 37, 17–23.
14. Reithinger R., Mohsen M., Aadil K., Sidiqi M., Erasmus P., Coleman P. G. (2003). Anthroponotic cutaneous leishmaniasis, Kabul, Afghanistan. *Emerging Infectious Diseases*, 9:727-9.
15. Reithinger Richard and Coleman Paul G, (2007).Treating cutaneous leishmaniasis patients in Kabul, Afghanistan: cost-effectiveness of an operational program in a complex emergency setting. *BMC Infectious Diseases*, 7:3
16. Reithinger, R., Aadil, K., Kolaczinski, J., Mohsen, M., Hami, S. (2005). Social impact of leishmaniasis, Afghanistan. *Emerging Infectious Diseases*, 11: 634-6.
17. Spiegelhalter D J, Best N G, Carlin B P and van der Linde A (2002). Bayesian measures of model complexity and fit (with discussion). *Journal of Royal Statistical Society Series B* 64: 583-640
18. Stevenson, M. A., Morris, R. S., Lawson, A. B., Wilesmith, J. W., Ryan, J. B. M., & Jackson, R. (2005). Area-level risk for BSE in British cattle before and after the July 1988 meat and bone meal feed ban. *Preventive Veterinary Medicine*. 69, 129-144
19. Tango, T. (1995). A class of tests for detecting 'general' and 'focused' clustering of rare diseases. *Statistics in Medicine*. 14, 2323–2334
20. Toby Leslie, Sarah Saleheen, Mohammed Sami, Ismail Mayan, Najibullah Mahboob, Kathy Fiekert, Annick Lenglet, Rosalynn Ord and Richard Reithinger (2006). Visceral leishmaniasis in Afghanistan. *CMAJ*, 175:3.