Applying and comparing empirical and full Bayesian models in study of evaluating relative risk of suicide among counties of Ilam province

Behzad Mahaki, Yadollah Mehrabi1, Amir Kavousi2, Youkhabeh Mohammadian3, Mehdi Kargar4

Department of Biostatistics, School of Health, Isfahan University of Medical Sciences, Isfahan, 1Department of Epidemiology, School of Public Health, 2Department of Sciences, School of Health, Safety and Environment, Shahid Beheshti University of Medical Sciences, 3Department of Clinical Psychology, School of Behavioral Sciences and Mental Health, Iran University of Medical Sciences, Tehran, 4Ward 2, Ebne-Sina Hospital, Shiraz University of Medical Sciences, Shiraz, Iran

ABSTRACT

Introduction: Disease mapping includes a set of statistical techniques that provides maps based on estimates of diseases rates. Bayesian ones are the most important models in this field. They consider prior information on changes in the disease rates in overall map and spatial pattern of the disease. These include a broad range of models with their own formulation, characteristics, strengths, and weaknesses. In the present study, we explain and compare three important and widely-used Bayesian models in the study of evaluating relative risk of suicide in Ilam province. Materials and Methods: In this applied-ecological research, suicide incidence in Ilam province in 2008 and 2009 was analyzed by use of Gamma-Poisson, Log-normal, and BYM Bayesian models. Models were fitted to data using WinBUGS software. Results: Fitting the three models showed that Darenshahr and Shirvan-Chrdavol had the highest and the lowest relative risk of suicide, respectively (relative risks based on Gamma-Poisson, Log-normal, and BYM models were 2.243, 2.275, and 2.279 for Darenshahr and 0.321, 0.321, and 0.319 for Shirvan-Chrdavol, respectively). Conclusion: Despite some differences in estimates, the ranks of relative risks in counties in all three models are the same. The counties based on the relative risks of suicide from the most to the least are: Darenshahr, Ilam, Dehloran, Eyvan, Abdanan, Mehran, Malekshahi, and Shirvan-Chrdavol.

Key words: BYM, disease mapping, empirical bayes, Ilam, relative risk, suicide

INTRODUCTION

Disease mapping contains a set of statistical methods, which leads to generate accurate maps based on estimations of incidence, prevalence, and mortality of diseases.[1,2] The oldest example of disease mapping is the address of cholera victims based on distance from water resources, which was presented in 1854 by John Snow.[3] Due to importance of geographic distribution of disease rates in determination of risk factors, there is raised concern about the disease mapping and risk assessment in recent years.[3] The most important aims of disease mapping contains describing spatial variation...
in prevalence of diseases in order to formulate hypothesis about causes of diseases, specification of high risk areas which needed increased notice and interference, generating an accurate map of disease risk in area for the purpose of better resource allocation, and risk estimation as well as producing disease atlase.\cite{1,2,4}

There are different models and methods to develop maps of diseases including simple statistic illustration, informal methods, basic models, Bayesian models, multilevel models, etc.

Bayesian approach to disease mapping consists of prior information about variation in disease rates, in addition to observed events in each area. It also could consider spatial pattern of disease in which closer geographic areas have more similar disease rates.\cite{5,9}

Bayesian models in disease mapping comprise variously wide range with every one of them has its own formulation, characteristics, advantages, and disadvantages. This reveals the necessity of evaluating and comparing these models. Several studies were conducted about comparison of these models. However, except a few and limited studies, disease mapping models were not compared, and their general condition was not investigated. Clayton and Kaldor made comparisons about a limited set of models (Empirical Bayesian and full Bayesian models). They concluded that in spite of some differences among relative risk estimations in all methods, the order of relative risk is similar in the whole mapping area. Other efforts have been carried out to find the sensitivity analysis of full Bayesian models that are confined to a limited range of Bayesian models.\cite{10-17}

Regarding the importance of Empirical and full Bayesian models, this paper describes and compares three Bayesian models in study of relative risk estimation of suicide in counties of Ilam province during 18 months from March of 2007 till October of 2008.

Suicide is a kind of death committed mainly by one’s own intention.\cite{18} It can be considered as a continuum based on its severity. It can begin from safe and preventable behavior and end up to severe and unpreventable ones.\cite{19} According to the World Health Organization report, the worldwide number of suicide attempts leading to death was one million people annually, and the number of suicide commitment was 20-30 times more than this.\cite{20,21} It is also estimated that around 1,530,000 million people will die due to suicide, and 10-20 times more than this figure will commit suicide by the year 2020.\cite{22} It means that, averagely, one death due to suicide per 20 seconds and one suicide commitment per one to two second will happen by the year 2020. The average worldwide suicide rate is estimated 14.5 per 100,000 people. Lithuania and Russia have the first ranks of suicide in the world. In these countries 46 and 41 suicides per 100,000 (one hundred thousand) people have been committed suicide in these countries respectively.\cite{22}

Suicide as a social anomaly has a high prevalence in Iran too.\cite{24} According to reports from Welfare Organization of Iran, the world rank of suicide for Iran is 58. International statistics indicates that suicide rate in Iran was 9 per 100,000 people, which contributed 1% of total number of deaths.\cite{22} Recently, among all provinces of Iran, Ilam province has shown one of the highest suicide rates.\cite{25} This province located in the west of Iran contains eight counties, and its population has been 545,787 people in year 2006. There are some differences among its counties according to weather, cultural, and social conditions. In addition, some of the counties near the border of Iraq have more post-war problems of 8-year Iraq-Iran war for a long time. Therefore, because of importance of suicide as a complex psychological phenomenon, influenced by individual and environmental factors, investigation and comparison of suicide rate in all counties of Ilam province seems necessary and worthwhile to study.\cite{26-28}

Concerning the importance of suicide and its high rate in Ilam province and existence of only some descriptive reports about suicide incidence, it is essential to investigate geographical variation of suicide incidence rates among counties of Ilam province using advanced statistical models.

**MATERIALS AND METHODS**

In this applied ecological study, suicide data of Ilam province collected by Ilam University of Medical Science during 18 months from March of 2007 till October of 2008 were analyzed. Then, relative risk estimation of suicide incidence was calculated. Relative risk of a disease is calculated by use of proportion of having a disease in a group with exposure to a risk factor to a group without this risk factor.\cite{28} In this study, we consider relative risk as proportion of the cases committed suicide to expected numbers.\cite{1} Data analysis was carried out using Gamma-Poisson, log-normal, and BYM Bayesian models in WinBUGS software. Estimates from each model were compared with one another as well as with standardized mortality ratio as a classic non-Bayesian model. The results of Bayesian models were compared by use of Deviance information criterion (DIC).

First of all, regarding the importance of these methods, specific properties, and some differences among them, it is essential to explain briefly theoretical principles, advantages, disadvantages, and characterization of each model.

**Diseases mapping models**

**Standardized mortality ratio using maximum likelihood estimation**

The first step to evaluate a disease incidence within an area is to estimate the expected rate of incidence in that area. Then, we should compare observed cases with expected one.\cite{1,2} The proportion of observed cases to expected one is called Standardized Mortality Ratio (SMR). In studies related to diseases incidence, we use the proportion of observed to expected disease incidence, which is called Standardized Incidence Ratio (SIR). This ratio is an estimation of relative risk within each area.\cite{1,2}
In order to utilize this method, the map should be divided into \( n \) non-overlapped adjacent areas (\( i = 1, 2, \ldots, n \)). \( O_i \) and \( E_i \) are the number of observed and expected events in area \( i \), respectively. During the study, \( E_i \) was assumed constant and known and calculated by multiplying the overall crude incidence rate by the area population or its estimation.

Relative risk of disease in area \( i \) was shown as \( \theta_i \). In disease mapping, it is assumed that the number of events between areas are independent and follow Poisson distribution with mean \( E_i \theta_i \). Therefore, the likelihood for \( O_i \) is calculated by:

\[
L(\theta) = \prod_i \frac{\exp(-E_i \theta_i)}{O_i!}^{O_i} \prod_i \theta_i^{O_i} \exp\left(-\sum_i E_i \theta_i\right)
\]

Also, logarithm of likelihood function is:

\[
l(\theta) = \ln L(\theta) = \sum_i \ln O_i - \sum_i E_i \theta_i
\]

So, the estimation of maximum likelihood is:

\[
\hat{\theta} = \text{SMR}_i = \frac{O_i}{E_i}
\]

Although use of SMR and SIR is so common, there is some substantial shortcoming related to these estimations.\(^{[2]}\) To overcome these difficulties, statistical smoothing models could be used.\(^{[13]}\) Hierarchical Bayesian models play an important role in modeling the complexity of data structure in spatial epidemiology and can overcome SMRs shortcomings.\(^{[10]}\) In addition to observed numbers, Bayesian approach in disease mapping contains prior information about geographical variation of disease rates in whole map.\(^{[9]}\) It can be also considered disease spatial correlation (the tendency of nearer areas to the same rates of disease).

**Gamma-poisson model**

One of the earliest examples of Bayesian disease mapping is Gamma-Poisson model. Its resulted estimation is included in Empirical Bayesian estimations.\(^{[10]}\) This model assumes that the number of deaths in areas follow a Poisson distribution with mean \( E_i \theta_i \). The prior distribution for relative risks is also considered Gamma \((a, b)\). To obtain posterior distribution, Gamma prior distribution for relative risks is combined with Poisson likelihood. So, relative risk have posterior distribution Gamma \((a + O_i, b + E_i)\) with mean \( E_i \theta_i | O_i, a, b = \frac{a + O_i}{b + E_i} = w_i \text{SMR}_i + (1 - w_i) \frac{a}{b} \) which \( w_i = \frac{E_i}{b + E_i} \).

Thus, posterior mean for area \( i \) is weighted average of SMR for area \( i \) and overall relative risk that weights are conversely related to SMR variance. If the observed and expected counts are high, the estimator tends to SMR, but when they are low, it tends to prior mean.\(^{[5,8,31]}\)

The advantages of this method are: 1. This model estimates full posterior, which allows hypothesis testing and calculates confidence intervals. 2. Even if the real distribution of relative risks is Gamma, this model can estimate mean and variance using maximum likelihood method, which is more valuable than moment method.\(^{[5,8,31]}\) The main disadvantage of this method is its inability to consider spatial correlation.

**Log-normal model**

Although Gamma prior distribution for relative risk is proper from mathematical point of view, it could be limiting because of difficulty with entering covariates. It is also not possible to consider spatial correlation in closer areas.\(^{[10]}\) Log-normal model is more flexible than Gamma-Poisson model for relative risks of disease:

\[
y_i \sim \text{Poisson}(E_i \theta_i)
\]

\[
\log(\theta_i) = \alpha + u_i + v_i
\]

In this model, \( v_i \) is included in order to consider spatial correlation and similarity of adjacent areas.

**BYM model**

This model was introduced by Clayton and Kaldor and extended by Besag, York, and Mollie,\(^{[8,32]}\) and its formulation is described as:

\[
y_i \sim \text{Poisson}(E_i \theta_i)
\]

\[
\log(\theta_i) = \alpha + u_i + v_i
\]

In this model, relative risk parameter is splitted into three components:

- \( \alpha \): Trend component that is overall level of relative risk.
- \( U_j \): (spatial overdispersion) spatial correlated heterogeneity: It is logical that the close areas have similar relative risks. To take this similarities into account, the random variable \( U_j \) which is uncorrelated with other \( \{U_j\} \) is included in the model. For this component, spatial correlation structure is used where estimates for relative risk in each area are dependent on adjacent areas. The conditional autoregressive model proposed by Besag et al. is:

\[
[u_i | u_j, i \neq j, \tau^2_v] \sim N(\bar{u}_i, \tau^2_v)
\]

That, \( \bar{u}_i = \sum_j u_i w_{ij} \) and \( \tau^2_v = \sum_j w_{ij} \).

If \( i \) and \( j \) were adjacent, \( w_{ij} = 1 \), otherwise \( w_{ij} = 0 \).

- \( \psi \): (Non-spatial overdispersion (spatial uncorrelated heterogeneity)): By the formulation of the model for spatially correlated heterogeneities, the variance is dependent on the number of neighbors and independence couldn’t be defined well. In order to justify this problem, another component \( (\psi) \) is introduced that is an uncorrelated overdispersion parameter. The prior distribution of this parameter is: \( \psi \sim N(0, \tau^2_v) \)
Both $\tau^2_u$ and $\tau^2_v$ parameters control variability of $u$ and $v$. To analysis of full Bayesian, the prior distributions should be determined for these parameters. Bernardinelli has suggested Gamma prior distributions for these parameters.\textsuperscript{[12,13]}

### RESULTS

Information about Ilam counties population, suicide numbers based on 2006 national census conducted by Statistical Center of Iran and registered suicide cases in each county during 18-month study period from March 2007 till October 2008 is shown in Table 1.

The expected number of suicide cases in each county is calculated by multiplying the provincial crude incidence rate by the county population or its estimation (third column multiplied by fifth column in Table 1). The results are presented in Table 2.

For estimating relative risk of suicide, at first, estimation based on standardized incidence ratio as classical non-Bayesian model was calculated. According to this estimation, the relative risks in counties from highest to lowest ones are Dare-shahr (2.311), Ilam (1.158), Dehloran (1.074), Eyvan (0.839), Abdanan (0.677), Mehran (0.471), Shirvan-Cherdavel (0.302), and Malekshahi (0.293).

Then, in order to apply and compare Bayesian models in estimating relative risk of suicide in counties of Ilam province, dates were fitted to Gamma-Poisson (GP), Log-Normal (LN), and full Bayesian (BYM) models by use of WinBUGS software. Estimation of relative risk for suicide in each county was calculated and shown in Tables 3-5 along with standard deviation and confidence interval. Using these results, in addition to compare counties relative risks, we can compare three mentioned models. According to estimations made by fitting these models, Dare-shahr County has shown the highest relative risk. Relative risk estimations of this county using Gamma-Poisson, Log-Normal, and BYM model were 2.243, 2.275, and 2.279, respectively. It indicates the higher risk than expected and the average rate in the province. Shirvan-Cherdavel County showed the lowest relative risk of suicide.

Estimation of relative risk for this county based on three mentioned models were 0.321, 0.321, and 0.319 respectively that indicates the lower relative risk than expected and the average rate in this province. Regarding the confidence intervals resulted from the three models, Dare-shahr and Ilam Counties shown to be significantly higher risk rather than average province (relative risk more than 1) and Shirvan-Cherdavel, Malekshahi, and Mehran were found to be lower risk rather than the average amount in the province (relative risk lower than 1).

Moreover, the three models were compared by use of DIC criterion [Table 6]. This criterion is obtained by summing two other criteria, which represents goodness of fit and model complexity respectively. Based on the DIC, there was no certain difference between these models.

### CONCLUSION

Although the estimations obtained from Gamma-Poisson, Log-Normal, and BYM model represented some differences of Standardized Incidence Ratios according to either amounts and counties incidence ranks, they showed approximately similar estimations for suicide relative risk. These results are in accordance with Clayton and Kaldor’s findings in comparing Bayesian models. They showed that in spite of some differences among relative risk estimations, the ranks for each area are same in all three models.\textsuperscript{[81]} In addition, confidence intervals resulted from fitting this three models showed slight difference. Standard deviations of these estimations were approximately the same. In all three models, the highest standard deviation belonged to Dare-shahr County, which has the most number of suicides, and the lowest one was seen in Shirvan-Cherdavel County with the least ones.

One justification for similarities between results of the models could be related to the low number of study areas and registered incidence cases. If the counts in lower levels such as districts or villages are available, their comparisons will be more valuable than recent results.

According to the estimations resulted from fitting the models, the ranks of Ilam Province counties from highest...
of counties like districts or villages. Additionally, it’s better to add time term to the study in order to investigate temporal variation along with spatial variation.

REFERENCES

1. Lawson AB, Browne WJ, Carmen L, Rodeir. V. Disease Mapping with WinBUGS and MLwiN. New York: John Wiley and Sons; 2003.
2. Lawson AB, Biggeri AB, Boehning D, Lesaffre E, Viel JF, Clark A, et al. Disease mapping models: An empirical evaluation. Stat Med 2000;19:2217-41.
3. Snow J. On the Mode of Communication of Cholera. 2nd ed. London: Churchill Livingstone; 1854.
4. Pickle L, Mungiole M, Jones G, White A. Exploring spatial patterns of mortality: The new atlas of United States mortality. Stat Med 1999;18:3211-30.
5. Carlin BP, Louis TA. Bayes and Empirical Bayes methods for data analysis. London: Chapman and Hall; 1996.
6. Manton K, Woodbury M, Stallard E. A variance components approach to categorical data models with heterogeneous mortality rates in North Carolina counties. Biometrics 1981;37:259-69.
7. Tsutakawa R, Mixed model for analyzing geographic variability in mortality rates. J Am Stat Assoc 1988;83:37-42.
8. Clayton D, Kaldor J. Empirical Bayes estimates of age-standardized relative risks for use in disease mapping. Biometrics 1987;43:671-91.
9. Bernardo JM, Smith AF. Bayesian Theory. New York: John Wiley and Sons; 1994.
10. Mahaki B, Mehrabi Y, Kavousi A, Akbari ME, Waldhoer T, Schmid VJ, et al. Multivariate disease mapping of seven prevalent cancers in Iran using a shared component model. Asian Pac J Cancer Prev 2011;12:2353-8.
11. Clayton D, Kaldor J. Empirical Bayes estimates of age-standardized relative risks for use in disease mapping. Biometrics 1987;43:671-81.
12. Marshall R. Mapping disease and mortality rates using empirical Bayes estimators. Applied Stat 1991;40:283-94.
13. Tsutakawa RK, Shoop GL, Marienfeld C. Empirical Bayes estimation of cancer mortality rates. Stat Med 1985;4:201-12.
14. Mollie A, Richardson S. Empirical Bayes estimates of cancer mortality rates using spatial models. Stat Med 1985;4:201-12.
15. Asmorian NS, Ruzitalab A, Amir K, Masoud S, Mahaki B.
Area-to-Area Poisson Kriging analysis of mapping of county-level esophageal cancer incidence rates in Iran. Asian Pac J Cancer Prev 2013;14:11-3.

16. Lawson AB, Biggeri AB, Boehning D, Lesaffre E, Viel JF, Bertollini R. Disease Mapping and Risk Assessment for Public Health.New York: John Wiley and Sons; 1999.

17. Lawson AB, Cressie NA. Spatial statistical methods in environmental epidemiology. In: Sen PK, Rao CR, editors. Vol. 18. Elsevier: Handbook of Statistics: Bioenvironmental and Public Health Statistics; 2000. p. 357-96.

18. Andrivian V. Frequency of suicide attempts and its relation with main risk factors among schizophrenic patients in Golestan hospital. The first Adventure Seminar Abstracts book. Iran: Lorestan University of Medical Science; 2000. p. 2.

19. Esfandiari GH. The risk factors of suicide: A review on previous studies. Journal of Nursing and Midwifery. Vol. 37. Iran: Shahid Beheshti University of Medical Sciences and Health Services; 2001;24-9.

20. World Health Organization. World Report on violence and health. Geneva: WHO; 2002.

21. Farhadi A. A sociological and psychological affecting variables on suicide in Lorestan. The first Adventure Seminar Abstracts book. Iran: Lorestan University of Medical Science; 2000. p. 61.

22. Ahmadi A, Hajiahmadi M. An epidemiological report on successful suicide in Mazandaran province in 1990-91. J Mazandaran Univ Med Sci. 2000;10:8-12.

23. Bertolote JM, Fleischmann A. Global perspective in the epidemiology of suicide. Suicidology 2002;7:2.

24. Ghoreishi SA, Mousavinasab N. Systematic review of researches on suicide and suicide attempt in Iran. Iranian Journal of Psychiatry and Clinical Psychology 2008;14:115-121.

25. Janghorbani M, Sharifirdad G. Completed and attempted suicide in Ilam, Iran (1995-2002): Incidence and associated factors. Arch Iran Med 2005;8:119-26.

26. Mann JJ, Waternaux C, Haas GL, Malone KM. Toward a clinical model of suicidal behavior in psychiatric patients. Am J Psychiatry 1999;156:181-9.

27. Morselli E. Suicide: An essay on comparative statistics. In: Paul K, editor. London original work. IlSuicidio. Italy: Published in Milan; 1879. p. 1881.

28. Flisher AJ, Parry CD, Bradshaw D, Juritz JM, Seasonal variation of suicide in South Africa. Psychiatry Res 1997;66:13-22.

29. Leon Geordies. Epidemiology. Translated by Sabbaghian, Holakousi. Tehran: Gap Publication. 2005.

30. Simonof JS. Smoothing methods in statistics. New York: Springer; 1996.

31. Wolpert R, Ickstadt K, Poisson-gamma random field models for spatial statistics. Biometrika 1998;85:251-67.

32. Besag J, York J, Mollie A. Bayesian image restoration with two applications in spatial statistics. Ann Inst Stat Math 1991;43:1-21.

33. Bernardinelli L, Clayton D, Montomoli C. Bayesian estimates of disease maps: How important are priors? Stat Med 1995;14:2411-31.

34. Jamshidzadeh F, Rafiei H, Yasemi MT, Rahimi AS, Azam K. Trend of suicide and attempts for suicide in Ilam province during 1994-2001. J Soc Welfare 2004;3: 325-348.

Source of Support: Nil, Conflict of Interest: None declared