ISSN: 0975-8585

Research Journal of Pharmaceutical, Biological and Chemical Sciences

Mapping Tuberculosis Incidence Using Mixture Distribution in Khuzestan Province, Iran.

Safoora Gharibzadeh[1], Sarah Gharibzadeh[2], Seyed Mohammad Alavi[3], Rahim Chinipardaz[4], and Mehri Ghafourian[5]*

[1]Ms, PhD Student, Department of Epidemiology And Biostatistics, School of Public Health, Tehran University of Medical Sciences, Tehran, Iran.
[2]MD, Researcher, Ahvaz Jundishapur University of Medical Sciences, Ahvaz, Iran.
[3]Professor, Health Research Institute, Infectious and Tropical Diseases Research Center, Department of Infectious Diseases, Jundishapur University of Medical Sciences, Ahvaz, Iran.
[4]Professor, Statistics Department, Mathematics and Computer Science Department, Shahid Chamran University of Ahvaz, Ahvaz, Iran.
5Professor, Health Research Institute, Infectious and Tropical Diseases Research Center, Immunology Department, Ahvaz Jundishapur University of Medical Sciences, Ahvaz, Iran
.

ABSTRACT

The allocation of resources in the health authority areas is based on a mapping of disease incidence and determination distribution of disease relative risk in small areas. Data on tuberculosis (TB) have been analyzed before, without giving sufficient attention to spatial variations. We aim to find high risk cities of Khuzestan province for TB with expanding a mixture of Poisson distributions for the number of cases and categorize the cities to one of the risk groups of TB. Mixture distributions of Poisson were used to find the number of risk groups, which might have produced by observed counts of TB patients in 23 cities of Khuzestan province. The Number of risk groups and Poisson parameters of each component were estimated with maximum likelihood approach using the computer package C.A.MAN (Computer Assisted Mixture Analysis). Bayesian methods were used to allocate each city to a particular risk group with a programme in R (3.0.0), the statistical software. The results were geographically presented in maps by using ArcGIS (9.3) software. Seven cities showed high risk with traditional method while in the mixture method, it was shown by 5. Our findings demonstrated the usefulness of the mixture models for modeling data with geographical variations.
Keywords: Health resources, Khuzestan, Mapping, Mixture models, TB

Corresponding author

INTRODUCTION

Collected data in biological and medical researches, usually contain variations. Statistical methods identify and separate these variations due to external and unknown factors. The traditional approach assumes a probability density for the variable of interest (e.g., the number of people having the disease), which is used to answer various questions on the (observed) values of the variable and to compute the measure of the variable's variations [1].

Schaltmann and Bohning [2] in a simulation study compared the methods of mixture model with traditional models, such as quantiles and significance method, in order to develop disease mapping. They found that mixture model approach brings higher percentage of correct allocation. When data, such as the number of people having a disease, are collected on small regions- like city wards, census tracts, *etc*.- the variations that are found in these data may be due to spatial proximity between regions or inter-differences in variables of regions. Thus, over-dispersion always presents in the spatially referenced data [2].

Traditional methods apply Standardized Mortality Ratio (SMR) as a tool for epidemiological measure. One of the criticisms of this approach is the categorization based on arbitrary. In addition SMRs are unstable, especially in small areas [3].

Since traditional methods have deficiencies in mapping the incidence of disease, Schaltmann and Bohning [2] suggested Empirical Bayes (EB) methods that have more stable estimates for relative risk. After expanding these methods, Schaltmann and Bohning [2] introduced mixture modeling to reduce random variability of data. Advantages of this method are as follows: yielding a shrinkage estimator by means of posteriors; presenting an estimate of the underlying risk pattern and providing estimation for number of components [2, 4].

It is worth considering that the mixture models are useful for modeling the population heterogeneity common in disease mapping problems because of their robustness [5].

We conducted an investigation to geographically present the map of TB in Khuzestan province using the technique of mixture models. These screening helps the health experts to detect high and low risk cities and identify regions that need extra care.

MATERIAL AND METHODS

Data on patients with TB during 2011-2012 were collected in Khuzestan Deputy of medical care. Diagnostic criteria for TB were at least two sputum smears positive for acid-fast bacilli out of three sputum smears. Basic geographic aggregation unit for mapping in this research was the city, of which there are 23 in Khuzestan province. Total number of people in each city and number of cases reported with Statistical Center of Iran were used for analysis.

We first used traditional methods to categorize Incidence Rate (I.R) into tertiles. Then, Non-parametric maximum likelihood estimation (NPMLE) was used to estimate the parameters of mixture model. After detecting the number of components, using Bayes theorem, each city was allocated to one of the risk levels.

At first, traditional methods of disease mapping were used. For *i*th city (i changes between 1 to 23) observed, the number of TB showed with o_i and population with n_i. Dividing observed by population, calculates yearly incidence rate: $I.R = \frac{o_i}{n_i}$. To estimate I.R we use Binomial model for the number of observed TB patients. If we assume fix rate of θ for TB in province, the probability of morbidity for a person equals to θ. Estimation of θ equals to:

$\hat{\theta} = \frac{\sum o_i}{\sum n_i}$ = (total number of observed case of TB in 2012 / total population of Khuzestan in 2012)

With this estimation, the number of cases for TB in *i*th city (*i*=1, 2,…, 23) can be modeled with Binomial model with n_i and θ parameters. The Number of expected cases (e_i) in *i*th city, is calculated by means of binomial distribution $e_i = n_i.\theta$, $i = 1, 2, ..., 23$. The ratio of $\frac{o_i}{e_i}$,) is used frequently in different studies for *i*th area we showed by SMR_i .

When n_i is large and θ small, we can approximate Binomial model with Poisson model. Then number of observed cases of TB for *i*th city have Poisson distribution by means of m_i, this parameter is the function of expected number of cases (e_i) and relative risk (λ_i) of *i*th city. Therefore:

$$o_i \sim Po(\lambda_i e_i)$$

Probability of y in *i*th city equals to:

$$\frac{e^{-\lambda_i e_i}(\lambda_i e_i)^y}{y!}$$

y is non-negative integer [6]. With separate relative risk (λ_i) for each city, number of parameters is as much as the number of cities. λ_i may be assumed to come from a population of parameters with suitable distribution. Recently several models have been presented for describing spatial variation of rates [7-9]. Therefore, we postulate that relative risk of area is random variable that's distribution was mixture of Poisson distribution with parameter of λe_i. λ has discrete distribution probability that total number of components of distribution (k), takes values of $\lambda_1, \lambda_2, ..., \lambda_k$ with probabilities of $p_1, p_2, ..., p_k$. Then:

$$o_i \sim \sum_{j=1}^{k} p_j Poisson(e_i \lambda_j) \ .$$

Above distribution is a Mixture distribution with $Poisson(e_i \lambda_j)$ as a component density and

$$\Pr[\lambda = \lambda_j] = p_j \ , \quad j = 1, 2, ..., k$$

as a mixing distribution. Parameter estimation is obtained with Non-Parametric Maximum Likelihood Estimation (NPMLE) using expectation maximization (EM) algorithm. After that each city was allocated to one of the Risk groups. We calculate posterior probabilities for membership of *i*th city to *j*th group:

$$p_{ij} = \frac{\hat{p}_j \ f(o_i \mid \hat{\lambda}_j)}{\sum_r p_r f(o_i \mid \hat{\lambda}_j)}$$

*i*th city allocates to the group that have maximum value of posterior probability.

To assess how well the model fits the data, firstly the models were run with different number of components, and then according to maximum of Bayesian Information Criterion (BIC) and Akaike Information Criterion (AIC) was selected the best model.

$$BIC = 2L - (2K - 1)\log(n_{obs})$$

$$AIC = 2 * L - \alpha * (2 * K - 1)$$

ISSN: 0975-8585

where L is log-likelihood for each of the models, *K* is the number of components, α=2 and *n* is the number of observed cases for TB [10].

Model fitting is done with C.A.MAN package and calculating posterior probabilities, with the programme written in R (3.0.0). Pattern of disease were plotted on maps with ArcGIS 9.3.

RESULTS

Total population of different cities varies from 24515 (Haftgel) to 1415024 (Ahvaz). Table 1 shows the population and Incidence Rate of TB in cities of Khuzestan in 2012. In traditional method, Tertiles of incidence rate were determined and then SMRs were calculated for 5 cities: Ahvaz, Abadan, Andimeshk, Khoramshahr, Hoveizeh and Shadegan have SMR≥1. Plot 1 represents I.R of TB in Khuzestan (according to Tertiles of I.R). Table 2 represents the results of fitting mixture models with k=2, 3, 4, 5 components. For choosing model with suitable number of components we used AIC and BIC.

Maximum value for AIC occurs in k=2 and BIC in k=3 components, we choose k=3 components as a final model. According to mixture model Lali had low Risk (R.R=0.000) Ahvaz, Abadan, Andimeshk, Khoramshahr and Shadegan have high risk with R.R=1.435. Other cities have Medium risk (R.R=0.601). Value of R.R represents Poisson distribution parameter. According to Table 3 two methods differ in 8 (33.33%) cities. From cities with high risk, Abadan, Khoramshahr and Shadegan are located near the Arvand and Persian Gulf, Ahvaz and Andimeshk were densely populated.

Plot 1- Mapping incidence of TB in cities of Khuzestan according tertiles, 2012.

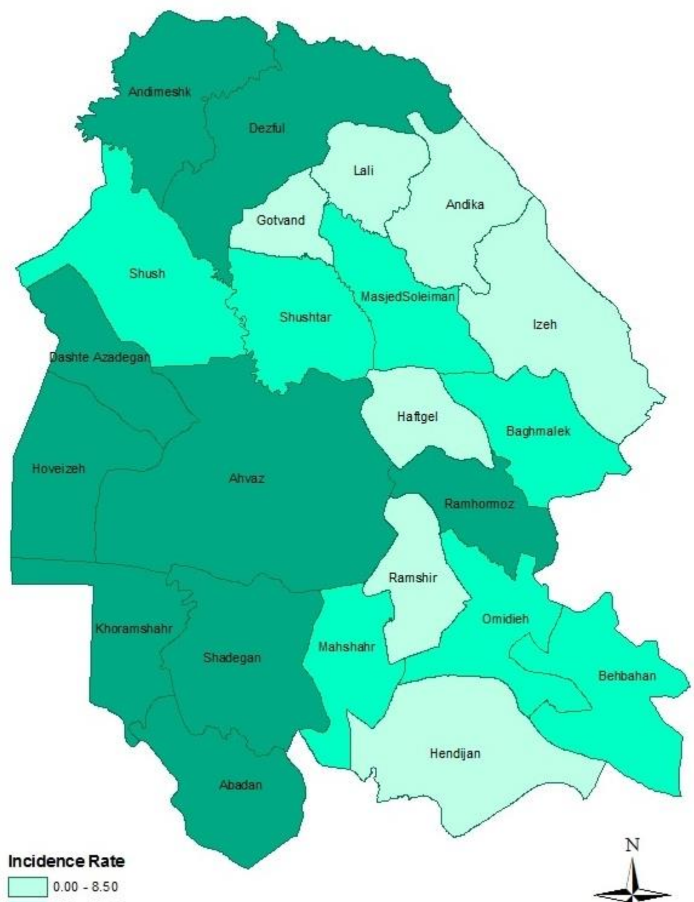


Table 1- Number of cases, population, incidence rate and SMR of TB in 2012

	City	No. TB	Population	I.R	SMR
1	Abadan	103	311250	33.09	1.85
2	Ahvaz	366	1415024	25.87	1.45
3	Andika	1	64420	1.55	.09
4	Andimeshk	41	197257	20.79	1.16
5	Baghmalek	15	111455	13.46	.75
6	Behbahan	28	194018	14.43	.81
7	Dashte Azadegan	16	105609	15.15	.85
8	Dezful	68	473089	14.37	.80
9	Gotvand	5	67041	7.46	.42
10	Haftgel	1	24515	4.08	.23
11	Hendijan	1	38314	2.61	.15
12	Hoveizeh	7	37849	18.49	1.04
13	Izeh	15	208959	7.18	.40
14	Khoramshahr	44	193669	22.72	1.27
15	Lali	0	41795	.00	.00
16	Mahshahr	30	266498	11.26	.63
17	MasjedSoleiman	16	174120	9.19	.51
18	Omidieh	15	104393	14.37	.80
19	Ramhormoz	9	100454	8.96	.50
20	Ramshir	2	48901	4.09	.23
21	Shadegan	29	146215	19.83	1.11
22	Shush	18	211669	8.50	.48
23	Shushtar	18	209182	8.60	.48
	Total	848	4745696	17.86	

Table 2- Parameters of Mixture Poisson with different components and GOF criteria.
Maximum likelihood estimates provided by C. A.MAN
***Selected model**

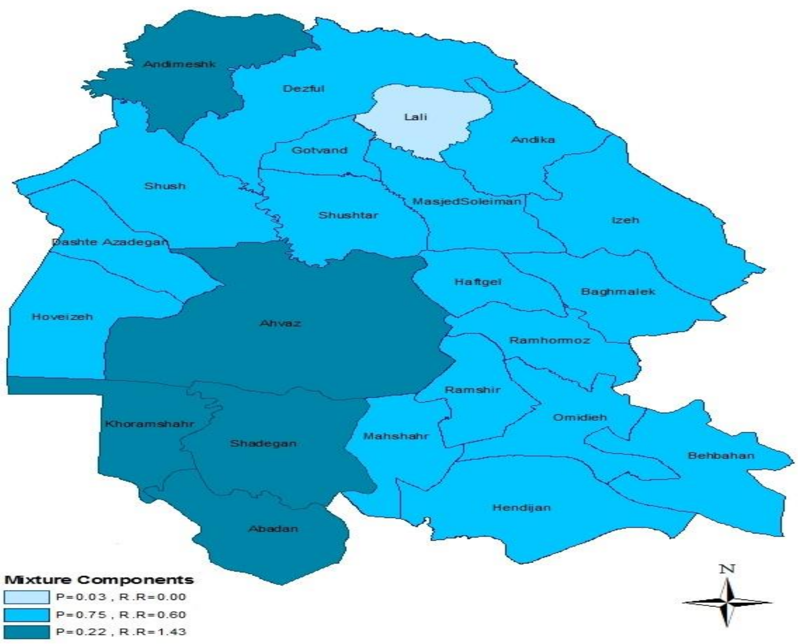
Components	Weight ($\hat{p}_j$)	Parameter ($\hat{\lambda}_j$)	Log-Likelihood	BIC	AIC
K=2	0.0358	0.0000	-90.65	-201.529	-187.30
	0.9642	0.70843			
K=3*	0.0306	0.0000	-88.65	-211.01	-187.3
	0.7461	0.6015			
	0.2233	1.4353			
K=4	0.0306	0.0000	-87.79	-222.78	-189.58
	0.7446	0.6011			
	0.1837	1.3750			
	0.0410	1.8273			
K=5	0.0207	0.0000	-84.97	-230.62	-187.94
	0.4685	0.4536			
	0.3091	0.7727			
	0.1610	1.3864			
	0.0408	1.8272			

ISSN: 0975-8585

Table 3- comparing Traditional method and mixture model (2012).

Number	City	I.R	Mixture Component
1	Abadan	3	3
2	Ahvaz	3	3
3	Andika	1	2
4	Andimeshk	3	3
5	Baghmalek	2	2
6	Behbahan	2	2
7	Dashte Azadegan	3	2
8	Dezful	2	2
9	Gotvand	1	2
10	Haftgel	1	2
11	Hendijan	1	2
12	Hoveizeh	3	2
13	Izeh	1	2
14	Khoramshahr	3	3
15	Lali	1	1
16	Mahshahr	2	2
17	MasjedSoleiman	2	2
18	Omidieh	2	2
19	Ramhormoz	2	2
20	Ramshir	1	2
21	Shadegan	3	3
22	Shush	2	2
23	Shushtar	2	2

Plot 2- Mapping incidence of TB in cities of Khuzestan according to components of mixture model, 2012.



DISCUSSION

Main results of the current study revealed that alongside a river and densely populated cities have higher risk of TB which is in agreement with results of Chandrasekaran and Arivarignan [1].

This study demonstrated that the mapping of TB could properly be evaluated with mixture models. The importance of mapping for detecting reasons of infectious disease was discussed in different studies [11].

In the present study, using mixture models, distribution of TB in Khuzestan is divided into 3 risk groups. The method used in this study is known as empirical Baysian methods [12]. Rattanasiri et al. [3], Bohninng [13], Chandrasekaran and Arivarignan [1], Gharibzadeh wt al. [6] and Lawson and Clark [11] used this method for mapping of various disease. Different authors use SMR method in their studies but we apply mixture distribution approach to modeling SMR. Strength of this study is solving the problem of choosing the number of components. Previous studies used Likelihood-Ratio Test (LRT) that problems of using this method mentioned by Schaltmann and Bohning [2]. In this study we solve this problem with using two criteria AIC and BIC. In order to get better results, we suggest using several years' information of TB with models such as space-time Mixture Models.

ACKNOWLEDGMENT

This study was supported by Health Research Institute, Infectious and Tropical Diseases Research Center of Ahvaz Jundishapur University of Medical Sciences (Project No 90128).

REFERENCES

[1] Chandrasekaran K , Arivarignan G. Indian J Med Res. 2006; 123: 788-798.
[2] Schlattmann P, Böhning D. Stat Med. 1993; 12: 1943-1950.
[3] Rattanasiri S, Bohning D, Rojanavipart P, Athipanyakom S. Southeast Asian J Trop Med Public Health. 2004; 35: 38-47.
[4] Biggeri A, Marchi M, Lagazio C, Martuzzi M, Böhning D. Stat Med. 2000; 19: 2539-2554.
[5] Militino AF, Ugarte MD, Dean CB. Stat Med 2001: 20: 2035–2049.
[6] Gharibzadeh S, Mahjoub H, Moghim BA, Sadri GH. Hakim. 2010; 12: 11-18.
[7] Clayton D, Kaldor J. Biometrics. 1987; 43: 671-681.
[8] Besag J, York J, Mollié A. Ann Inst Stat Math. 1991; 43: 1-20.
[9] Bernardinelli L, Montomoli C. Stat Med. 1992; 11: 983-1007.
[10] Böhning D, Dietz E, Kuhnert R, Schön D. Stat Methods Appt. 2005; 14: 29-43.
[11] Lawson AB, Clark A. Stat Med. 2002; 21: 359-370.
[12] Maritz JS, Lwin T. Empirical Bayes Methods. 2nd ed, London, Chapman and Hall, 1989.
[13] Bohning D. Environmetrics. 2003; 14: 431-451.