Original Article

Available online at www.bpasjournals.com

Bulletin of Pure and Applied Sciences.
Vol. 38E (Math & Stat.), No.1, 2019. P.41-48
Print version ISSN 0970 6577
Online version ISSN 2320 3226
DOI: 10.5958/2320-3226.2019.00005.5

A GENDER WISE SPATIAL DISTRIBUTION OF MOUTH CANCER USING POISSON-GAMMA MODEL FOR CHENNAI ZONES

P. Sampath[1*], P.R. Jayashree[2], R. Srinivasan[3], R. Swaminathan[4]

Author Affiliation:
[1]Statistical Assistant, Department of Epidemiology and Cancer Registry, Cancer Institute (WIA), Adyar, Chennai, Tamil Nadu 600020, India.
E-mail: kubersambath@yahoo.co.in
[2]Assistant professor, Department of Statistics, Presidency College, Chepauk, Chennai, Tamil Nadu 600005, India.
E-mail: vprjaya@gmail.com
[3]Senior Technical Assistant, National Institute for Research in Tuberculosis, ICMR, Chetpet, Chennai, Tamil Nadu 600031, India.
E-mail: srinivast_r@yahoo.com
[4]Professor and Head, Department of Epidemiology and Cancer Registry, Cancer Institute (WIA), Adyar, Chennai, Tamil Nadu 600020, India.
E-mail: r.swaminathan@cancerinstitutewia.org

***Corresponding Author:**
P. Sampath, Statistical Assistant, Department of Epidemiology and Cancer Registry, Cancer Institute (WIA), Adyar, Chennai, Tamil Nadu 600020, India.

E-mail: kubersambath@yahoo.co.in

Received on 07.09.2018, Revised on 03.11.2018 Accepted on 09.01.2019

Abstract: Cancer is known to be one of the leading causes of mortality in the world. There were about 14.1 million incidences and 8.2 million deaths due to cancer globally. In terms of mouth cancer Age Standardized Rate is 4.0 per 100000 populations worldwide and 7.2 per 100000 populations in India. In Chennai, mouth cancer burden has significantly increased over the past decade irrespective of geographical region. In this paper, the mouth cancer incidence is used to analyze the spatial distribution for high risk and low risk areas of different zones in Chennai by gender for the period of 2004-2013. The aim of this study is to fit a Poisson Gamma model and to explore the Empirical Bayesian and frequentist approach for disease mapping of mouth cancer incidence for Chennai zones by sex. The results of the estimates reveal that the empirical Bayesian estimate is more stable than the conventional frequentist estimates.

Keywords: Disease Mapping, Poisson-Gamma model, Relative risk, Spatial Distribution, Standardized Incidence Ratio/ Standardized Morbidity Ratio.

1. INTRODUCTION

Cancer is known to be one of the non infectious diseases and there were about 14.1 million incidences and also leading causes of mortality globally. The cancer incidences rate is 7% and mouth cancer incidence is 5% in Chennai to the total of its occurrences in India. Mouth Cancer is a manifold stage of disease highly correlated with environmental conditions, lifestyles, biological and social factors. Its incidence pattern is highly influenced by geographic region, cultural and population characteristics and it shows spatial variations. Hence, identification of vulnerable population is more important on the community and to prepare suitable health policies for prevention and control of cancer. The spatial distribution of mouth cancer incidence for different zone levels is still unidentified in Chennai.

The city of Chennai (formerly Madras) is known to exist since 18th century. It lies on the eastern coast of peninsular India at 13° N Latitude and 80° E longitude and is the Gateway to South India. It is the most populated district (46, 46,732 inhabitants, population density of 26,553 inhabitants/km^2 with a predominance of male). The Chennai city, the capital of the State of Tamil Nadu, with 6.44% population proportionate to the State of Tamil Nadu within the metropolitan limits, has an area of about 170 km^2. It is divided into 155 wards comprised of ten zones. According to population census ten year report, the sex ratio has changed from 934 females: 1,000 males in 1981 to 989 females: 1,000 males in 2011.

At present, lot of research is focused on disease mapping with the help of Geographical information system (GIS) Elebead et al., 2012 [13] and George et al., 2013[14]. The aim of this study is to focus on the mouth cancer incidence and how it is spatially distributed in Chennai on gender basis for the period of 2004 to 2013. The paper also aims to analyze the relative risk estimation based on the standardized incidence ratio (SIR) method, Poisson Model and Poisson-Gamma Model for men and women.

Spatial data analysis has a distinct application in the field of health Statistics, Walter, 2000 [15], and Bayesian estimation method is used for disease mapping to assess the risk and prediction spatially. Clayton and Kaldor [3] and Besag et al. [1] have studied empirical Bayes Estimates of age standardized relative risk for disease mapping. In Bayesian approach, the major powerful methods are the empirical and fully Bayesian methods. In the Empirical Bayesian (EB) method, the parameters of prior distributions are estimated using observed marginal distributions, but in the Bayesian approach, the prior and posterior distributions are obtained through Markov Chain Monte Carlo (MCMC) computations. Marshal [9] reviews Empirical Bayesian and Fully Bayesian methods for disease mapping. Lawson et al. [5] studied the empirical evaluation of certain disease mapping models.

This study is utilizes the most common method of estimation - the SIR model and Poisson model followed by the use of the Poisson-Gamma model. In this paper, the mouth cancer incidence is used to analyze the spatial distribution for high and low risk areas of different zones in Chennai by gender for the period between the years 2004-2013. The aim of this study is to fit a Poisson Gamma model and to explore the Empirical Bayesian and frequentist approach for disease mapping of mouth cancer incidence for the different zones of Chennai by male and female sex.

2. MATERIALS

Mouth cancer data was collected from the Madras Metropolitian Tumour Registry (MMTR), Cancer Institute (WIA) Adyar, Chennai as it is a Population Based Cancer Registry (PBCR) under the National Cancer Registry Program network in Indian Council of Medical Research. The observed mouth cancer cases on Male and Female for the period of 2004-2013 were used to aggregate on different zones in Chennai namely Tondiarpet, Basinbridge Pulianthoppu, Ayanavaram, Kilpauk, Icehouse Nungambakkam, Kodambakkam, Saidapet, and Adyar. The incidence rate is calculated by direct standardized method and the relative risk estimation is calculated using the WinBUGS software.

3. METHODOLOGY

i. Standardized Incidence/Morbidity Ratio (SIR) Method

The method is most widely used by all researchers to find out the relative risk in disease mapping. In this paper Standardized Incidence Rate compares the observed and expected number of incidence described by Lawson et al. [5]. The area of study is divided into N regions, all regions have observed number of cases O_i and expected number of cases E_i (i=1,2,…,N). The relative risk θ_i , which is the Standardized Incidence Rate of area i, is given by

$\theta_i = \left(\frac{O_i}{E_i}\right)$, where (i=1, 2,…,N). Based on the approach of Lawson et al. [5], SIR is based on the ratio estimator, the mean and variance dependence on E_i . The SIR method is affected by the expected number of cases and if there are no observed cases, then the standardized incidence ratio is to be zero.

ii. Disease Mapping Model Using Maximum Likelihood Estimation

To model the observed cases in region i and area j, the samples are drawn from a Poisson distribution with mean $\theta_i e_{ij}$. i.e., the disease in the area or the region is small; the usual model for the Y_i is the Poisson model, where Y_i is the observation of an event for Poisson model.

Y_i / θ_i ~ Po(E_i θ_i), where θ_i is the true relative risk of disease in the region *i*. The maximum Likelihood Estimate (MLE) of θ_i is,

$\hat{\theta}i \equiv SIR$

$= Y_i / E_i$ (1)

The Standardized Incidence Ratio (SIR) is a simple ratio of the observed number of cases to the expected number of diseased cases.

The Variance

Var (SIR_I)= VAR (Y_i / E_i^2) = θ_i / F_i (2)

$$\widehat{Var(SIR_i)} = \frac{\hat{\theta}_i}{E_i} = \frac{Y_i}{E_i 2} \quad (3)$$

The confidence interval for θ_i calculated by using the delta method is given by

$$Var[log(SIR_i)] = \frac{1}{SIR_i^2} Var(SIR_i) = \frac{E_i^2}{Y_i^2} * \frac{Y_i}{E_i^2} = \frac{1}{Y_i} \quad (4)$$

Also, the Standard Error (SE), which indicates the measure of the uncertainty of SIR is given by

SE(SIR) =$\sqrt{Var(SIR)}$

Var(SIR_i)= $\frac{O_i}{E_i}$ (5)

$= \frac{O_i}{E_i 2} = \frac{SIR_i}{E_i}$ (6)

iii. Poisson-Gamma Model

Clayton and Kaldor [3] proposed the Poisson-Gamma Model for the first time, which assumed that the relative risk holds Poisson distribution. The assumption of the Poisson Model is that its mean and variance are the same, but in the spatial context, the data are over dispersed and variance is higher than the mean. A simple way to allow for a higher variance is use to the negative binomial distribution instead of the Poisson distribution. The negative binomial distribution can also be regarded as a mixed model in which the random effect follows a Gamma distribution for each area. This combination is known as the Poisson-Gamma model.

A simple model is,

Y_i/n_i ~ Po(E_i/ n_i), i=(1,2,…) and n_i ~ Gamma(a,b) which denotes gamma distribution with mean $\mu = \frac{a}{b}$ and variance $\sigma^2 = \frac{a}{b^2}$.

From the above mean and variance, the values of *a* and *b* are obtained as

$a = \frac{\mu^2}{\sigma^2}$ and $b = \frac{\mu}{\sigma^2}$

P($\frac{\theta_i}{y_i}$) ~ Gamma ($y_i + E_{ib}$) is the posterior distribution using gamma prior with the Poisson likelihood as Π_i P($\frac{\theta_i}{y_i}$)

The mean of the posterior distribution is given by

$$E\left(\frac{\theta_i}{Y}\right) = E\left(\frac{\theta_i}{y_i}\right)$$

$$= \frac{y_i + a}{E_i + b} = \frac{y_i + \frac{\mu^2}{\sigma^2}}{E_i + \frac{\mu}{\sigma^2}} = \frac{E_i\left(\frac{y_i}{E_i}\right)}{E_i + \frac{\mu}{\sigma^2}} + \frac{\left(\frac{\mu}{\sigma^2}\right)\mu}{E_i + \frac{\mu}{\sigma^2}} \quad (7)$$

$= \omega_i$ SMR_i + (1- ω_i) μ (8)

where, $v_i = E_i \mid \left[E_i + \frac{\mu}{\sigma^2}\right]$.

4. RESULTS

The outcome of the relative risk estimation methods is displayed by gender in tables and maps. The relative risk of SIR, Poisson model using MLE, Bayesian Gamma models are also found and presented in Table 1 and Table 2 respectively for all zones in Chennai. The relative risks are classified as low Relative Risk (RR<1) and high Relative Risk area (RR>1) for Men and Women in all the zones.

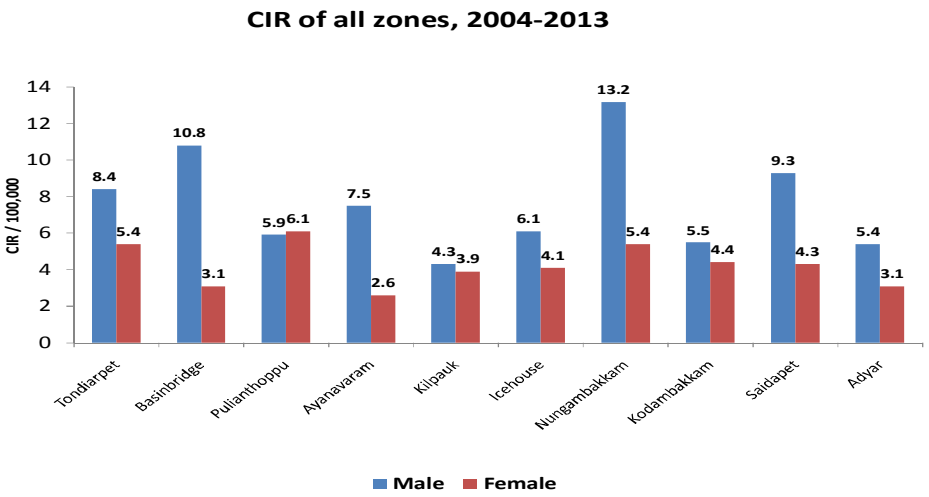


Figure 1: Graphical representation of Crude Incidence Rate (CIR) of Mouth cancer for Men and Women in Chennai

Figure 1 shows the Crude Incidence Rate (CIR) of mouth cancer among Men and Women for all ten zones in Chennai . It clearly indicates that Nungambakkam and Basin bridge have the highest mouth cancer incidence whereas Kilpauk has the lowest mouth cancer incidence rate among men. The zones Pulianthoppu and Nungambakkam are found to have the highest incidence rate of mouth cancer among Women.

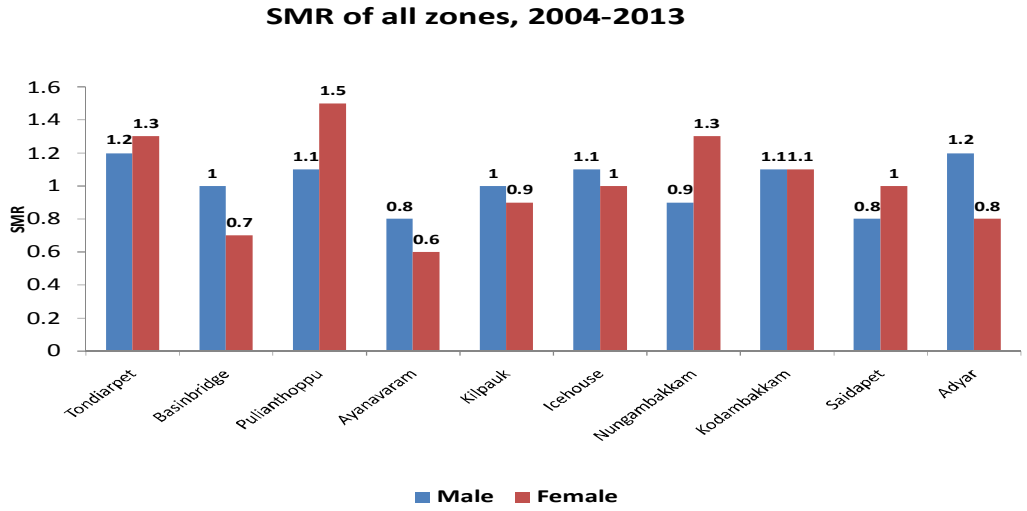


Figure 2: Graphical representation of Standardized Incidence (Morbidity) Ratio of Mouth Cancer for Men and Women in Chennai

Figure 2 presents the Standardized Incidence/Morbidity Ratio (SMR) of mouth cancer among Men and Women for all zones in Chennai. It is found that Tondiarpet and Adyar zones have the highest standardized incidence ratio whereas the Ayanavaram zone has the lowest standardized incidence ratio among men. The zones Pulianthoppu and Tondiarpet are found to be the highest Standardized Incidence/Morbidity Ratio of mouth cancer among Women and the lowest standardized mouth cancer incidence ratio was found in the Ayanavaram zone.

Table 1: Spatial distribution of mouth cancer incidence of Men in Chennai by zone level for the period 2004-2013

Zone	Population	Cases	CIR	SIR	RR - MLE	SE	RR-Poisson Gamma model	Markov Chain Error
Tondiarpet	232089	194	8.4	1.2	0.974	0.0026	1.143	0.0008
Basin bridge	199523	216	10.8	1.0	1.000	0.0000	1.474	0.0011
Pulianthoppu	257830	152	5.9	1.1	0.004	0.0996	0.8109	0.0007
Ayanavaram	276332	207	7.5	0.8	0.680	0.0320	1.028	0.0007
Kilpauk	297049	128	4.3	1.0	0.000	0.1000	0.5985	0.0006
IceHouse	157957	97	6.1	1.1	0.049	0.0951	0.844	0.0010
Nungambakkam	142142	188	13.2	0.9	1.000	0.0000	1.791	0.0015
Kodambakkam	242108	133	5.5	1.1	0.000	0.1000	0.759	0.0006
Saidapet	238455	221	9.3	0.8	1.000	0.0000	1.271	0.0008
Adyar	267308	145	5.4	1.2	0.000	0.1000	0.7513	0.0006

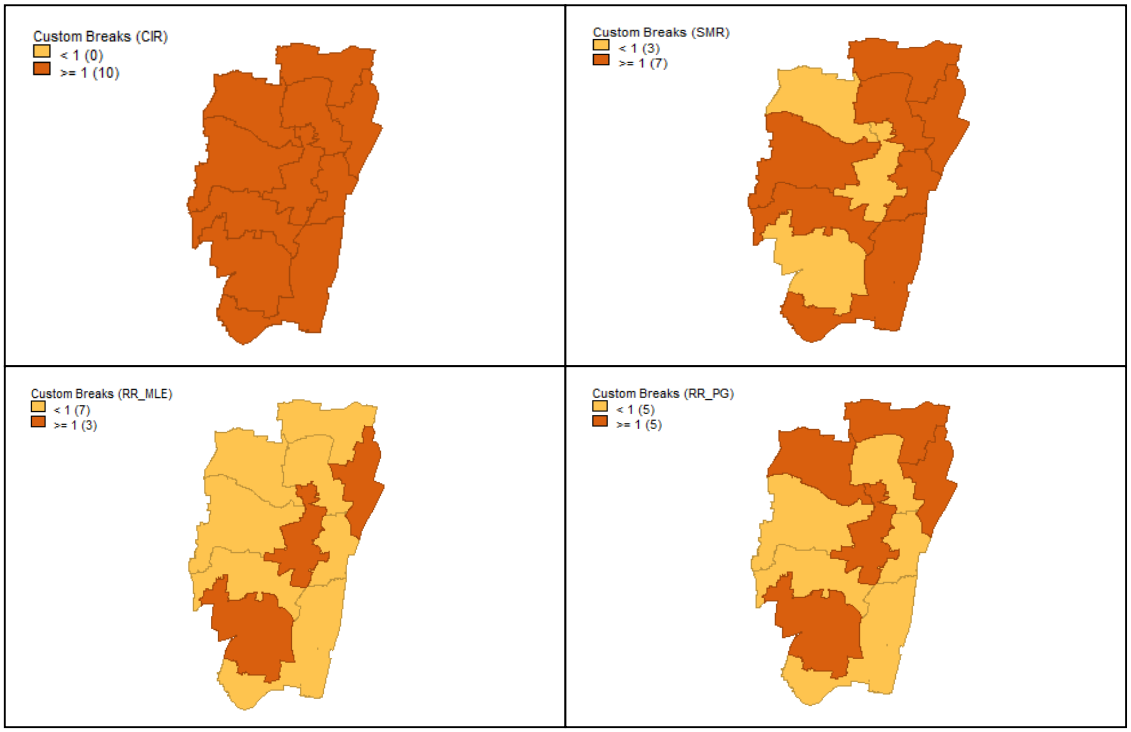
In Table 1, the standardized incidence ratio (SIR) method is found to distinguish the high and low risk of mouth cancer incidence among men at zone level. The maximum likelihood estimation of Standardized Incidence Rate (SIR) is calculated. The high standardized incidence risk ratio is found in Tondiarpet zone where RR is 1.2 and in Adyar zone the RR is 1.2.This method has one drawback, while proceeding with this technique in a region with low population both the observed and the expected cases are small, as a result they tend to present an extreme Standardized Incidence Ratio.

In the frequentist method the SIR of Map 1 is classified in to two groups in which the relative risk is more than 1 and less than 1 by dark and light shades. The zones Tondiarpet, Basinbridge, Ayanavaram, Nungambakkam and Saidapet are classified as the high risk zones as compared to the other remaining zones in the model. The highest Relative Risk among men was found in the Tondiarpet and lowest in the Kilpauk zone . The advantage of this frequentist method is easy to fit but unable to give a spatial smoothing.

The Empirical Bayesian method is also fitted for the model which are exhibited for all ten zones in Table 1 and shown using spatial map in Map 1. In the Map 1, the relative risk was classified into two groups for the area estimates more than one (RR>1) and less than one (RR<1) respectively. From the Table 1, it is revealed that from the Poisson-Gamma model the mean posterior relative risk ranges from 0.59 to 1.79 for all zones among men and the Posterior median ranges to be same as the mean relative risk (Table 3). The overall posterior mean α is 3.72, and the posterior mean β is 050 for men (Table 3). The high risk zones were smoothed towards local mean when it is compared to the frequentist method. In this model five zones namely Nungambakkam, Basin bridge, Tondiarpet, Saidapet and Ayanavaram are classified as high (>1 RR) Relative risk zones among men in Chennai.

In Table 2, the standardized incidence ratio (SIR) method is found to distinguish the high and low risk of mouth cancer incidence among women at zone level. The maximum likelihood estimation of Standardized Incidence Rate (SIR) are calculated. The high standardized incidence risk ratio is found in the Tondiarpet zone and the Nungambakkam zone both with RR as 1.3. This method has one drawback, as mentioned earlier also that while proceeding with this technique in a region with low population both the observed and the expected cases are small as a result which tends to present extreme Standardized Incidence Ratios.

In the frequentist method the SIR of Map 2 is classified into two groups in which the relative risk is more than 1 and less than 1 by dark and light shades. The following zones: Tondiarpet, Pulianthoppu, Nungambakkam, Kodambakkam and Saidapet are classified as high risk zones as compared to remaining zones of this model. The highest Relative Risk among women was found in the Pulianthoppu and lowest in Ayanavaram zone. The advantage of this frequentist method is easy to fit but unable to spatial smoothing, since this method is statistically incorrect.

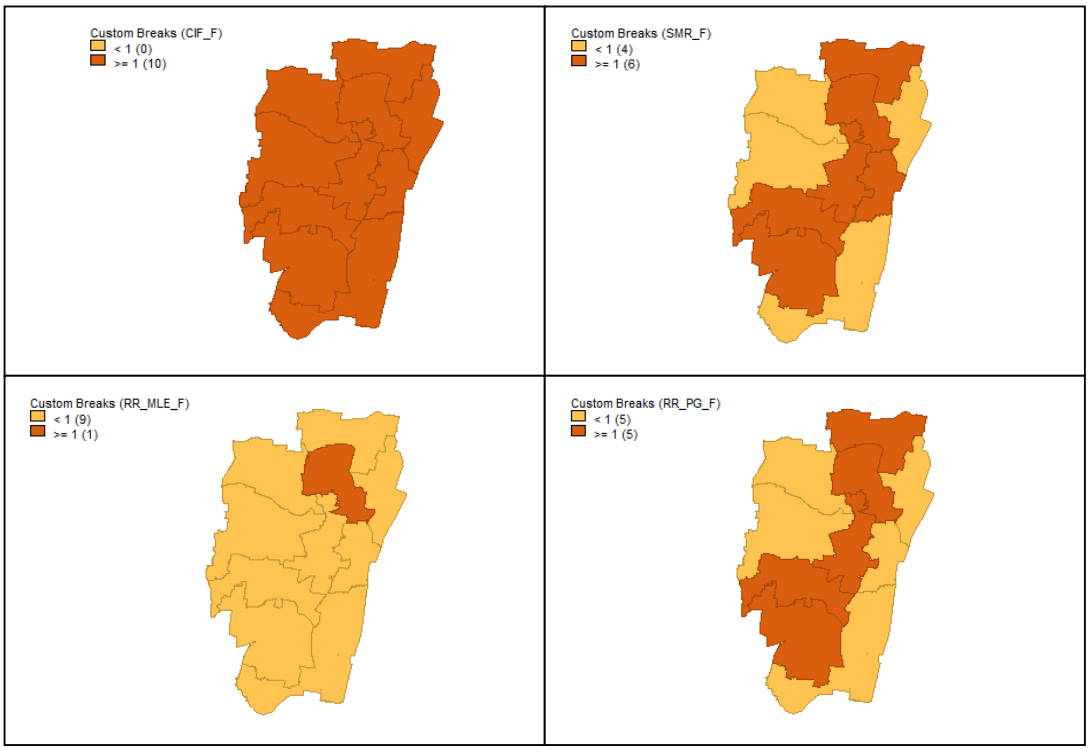


Map 1: Relative risk estimation of male in different models

Table 2: Spatial distribution of mouth cancer incidence of Women in Chennai by zone level for the period 2004-2013

Zone	Popn	Cases	CIR	SIR	RR - MLE	SE	RR-Poisson Gamma model	Markov Chain Error
Tondiarpet	229599	123	5.4	1.3	0.997	0.0003	1.269	0.0007
Basin bridge	191440	60	3.1	0.7	0.012	0.0988	0.7617	0.0006
Pulianthoppu	254165	154	6.1	1.5	1.000	0.0000	1.434	0.0008
Ayanavaram	276360	73	2.6	0.6	0.000	0.1000	0.6482	0.0005
Kilpauk	292962	114	3.9	0.9	0.240	0.0760	0.9357	0.0006
IceHouse	150230	61	4.1	1.0	0.445	0.0555	0.9689	0.0009
Nungambakkam	143174	77	5.4	1.3	0.986	0.0014	1.261	0.0010
Kodambakkam	240082	106	4.4	1.1	0.735	0.0265	1.057	0.0007
Saidapet	225739	97	4.3	1.0	0.634	0.0366	1.029	0.0007
Adyar	267724	84	3.1	0.8	0.004	0.0996	0.7585	0.0005

In the Empirical Bayesian method, the relative risk map was classified in to two groups in which the estimates are more than one (RR>1) and less than one (RR<1). The dark shades in Map 2 are classified as the higher relative risk areas whereas those with the light shades are having less relative risks. Here, the result of the Poisson-Gamma model revealed that the mean posterior relative risk ranges for all zones from 0.64 to 1.43 among women. The Posterior median ranges found the same as mean relative risk. The overall posterior mean α is 4.23 and the posterior mean β is 0.06 for women (Table 3). The high risk zones when smoothed towards the local mean standard error are also not found when compared to frequentist method. In this model five zones namely Pulianthoppu, Tondiarpet, Nungambakkam, Kodambakkam, and Saidapet were classified as high relative risk zones in Chennai among women (RR >1). The highest relative risk is 1.434 found in Pulianthoppu and the lowest relative risk RR is 0.648 in Ayanavaram zone in Chennai among women.



Map 2: Relative risk estimation of females in different models

Table 3: Comparison for Estimate of Frequentist and Empirical Bayesian Relative Risk

	Frequentist Model		Empirical Bayesian Model	
	Men	**Women**	**Men**	**Women**
SIR/Mean Relative Risk	0.8 – 1.2	0.6 - 1.5	0.59 – 1.79	0.64 – 1.43
Median Relative Risk	-		0.59 – 1.79	0.64 – 1.43
Standard Error	1.4 – 2.2	1.1 - 2.0	0.0003 – 0.0009	0.008 – 0.001
Mean Alpha			3.72	4.23
Mean Beta			0.050	0.06

From table 3, it is inferred that the Standardized Incidence Ratio of mouth cancer ranges among men from 0.8 to 1.2 as compared to the corresponding values of 0.59 to 1.79 on Empirical Bayesian method and the standard error ranges from 1.4 to 2.2 as compared to 0.0003-0.009. Also the standardized incidence rate of mouth cancer among women ranges 0.6 to 1.5 as compared to 0.64 to 1.4 on Empirical Bayesian method and the local standard error ranges from 1.1 to 2.0 as compared to the value 0.008-0.001 with Empirical Bayes method. The highest relative risk is reduced towards the local mean and the standard error is also reduced in the Empirical Bayesian method.

5. CONCLUSION

The identification of the high risk zone for mouth cancer incidence in Chennai is important for promoting the preventive measures. The mouth cancer incidence data for the ten zones in Chennai is analyzed using the Bayesian gamma model which gives the stable estimates and lower standard error as compared with the conventional methods. The conventional disease mapping using maximum likelihood estimation of standardized incidence ratio (SIR) and frequentist method of Poisson models are also fitted but cannot do the smoothing because the higher value of the estimates of relative risk is based on less value only with small population. Based on the result of the spatial models, we conclude that the Poisson-Gamma model provides better estimation of risk than SIR and CIR. These findings suggest that the Poisson-Gamma model is a good fit for estimating the relative risks. The zones Tondiarpet, Basinbridge, Ayanavaram, Nungambakkam and Saidapet for

males and zones Tondiarpet, Pulianthoppu, Nungambakkam, Kodambakkam and Saidapet for females are found to be high risk areas of Mouth cancer incidence in Chennai. Moreover, the zones like Tondiarpet, Nungambakkam and Saidapet were high risk areas for both the genders together.

6. ACKNOWLEDGMENTS

All the authors of this research paper would like to thank the staff members in the Division of Epidemiology and Cancer Registry at the Cancer Institute (WIA), Chennai for Data facilitation and their tireless efforts since the very beginning of this study.

REFERENCES

[1]. Besag, J., York, J. and Mollie, A. (1991). Bayesian image restoration with two applications in spatial statistics, *Ann Inst Stat Math*, 43, 1-59.

[2]. Breslow, N. and Clayton, D.G. (1993). Approximate inference in generalized linear mixed Models, Journal *of the American Statistical Association,* 88, 9-25.

[3]. Clayton, D. and Kaldor, J. (1987). Empirical Bayes estimates of age standardized relative risk for use in disease mapping, *Biometrics*, 43, 671- 681.

[4]. Devine, O.J. and Louis, T.A. (1994). A Constrained empirical Bayes estimator for incidence rates in areas with small populations, *Stat Med*, 13, 1119-1133.

[5]. Lawson, A.B., Biggeri, A., Boehning, D., Leasaffre, E., Viel, J.F., Clark, A., Schlattmann, P. and Divino, F. (2000). Disease mapping models: an empirical evaluation, *Statistics in Medicine* (Special issue: Disease Mapping with Emphasis on valuation of Methods), 19, 2217-2242.

[6]. Lawson, A.B. (2009). Bayesian Disease Mapping: Hierarchical Modeling in Spatial Epidemiology, *Int Stat Rev,* 77, 325-326.

[7]. Meza, J.L. (2003). Empirical Bayes Estimation smoothing of relative risks in disease mapping, *J Stat Plan Inference,* 112, 43- 62.

[8]. Manton, K., Woodbury, M. and Stallard, E. (1981). A variance components approach to categorical data models with heterogeneous mortality rates in North Carolina Counties, *Biometrics*, 37, 259-269.

[9]. Marshall, R.J. (1991). Mapping disease and mortality rates using empirical Bayes Estimators, *J R Stat Soc Ser C Appl Stat*, 40, 283-294.

[10]. Mollie, A. and Richardson, S. (1991). Empirical Bayes estimates of cancer mortality rates using spatial models, *Stat Med*, 10, 95-112.

[11]. Myint, L. and Ye, H. (1991). Study of the malaria situation in forested foothills and nearby plain areas of Myanmar, *Southeast Asian Jr. Trop Med Public Health*, 22, 509-514.

[12]. Boyle, P. and Parkin, D.M. Statistical methods for registries. IARC Sci. Publ. 1991; (95): 126-58, Lyon Cedex 08, France

[13]. Elebead, F.M., Hamid, A, Hilmi, H.S. and Galal, H. (2012). Mapping Cancer disease using Geographical Information System (GIS) in Gezira state – Sudan, *J. Community Health*, 2012 Aug; 37(4): 830-9, doi: 10.1007/ S 10900-011-9517-9

[14]. George, J.M., Po- Huang, C., Tyler, S. et al (2013), Uses of GIS mapping as a public health tool from cholera to cancer, *Health Serv Insights*, 6, 111-16.

[15]. Walter, S. D. (2000). Disease mapping; a historical perspective, In: Elliott, P., Wake field, J.C., Best, N.G. and Briggs, D. (editors), Spatial Epidemiology: Methods and Applications, Oxford University Press, 223-239.