healthline ISSN 2229-337X Volume 2 Issue 2 July-December 2011

Original article

A Bayesian approach to compare the statewise malaria death counts in India

Atanu Bhattacharjee[1], Dibyojyoti Bhattacharjee[2]
[1]Department of Statistics, Gauhati University, Guwahati-14, [2]Department of Business Administration, Assam University, Silchar-11, India.
Correspondence: Atanu Bhattacharjee Email: atanustat@gmail.com

Abstract

Context: In India malaria is a major public health problem as 95 percent of the population is exposed to the danger of being infected. Estimating the true disease burden of malaria in the country is a challenge, as true burden cannot be understood from ordinary count data on the number of deaths or number of infected cases.

Objective: The goal of this study is to compare the Indian states by the true burden due to malaria (*Plasmodium falciparum*) in the year 2010 and rank the states with respect to their performance to deal with malarial death.

Study Design: To deal with count data problems as a representation of death, the Bayesian approach is used. The source of data is the National Malaria Control Report published by NVBDCP of 2009 and 2010 till the month of August. The statewise comparison of death due to malaria has been performed with all computation done using statistical software R.

Main Outcome: Certain states like Orissa, Chhattisgarh, Karnataka, Meghalaya, Jharkhand and Assam had very high incidence of *Plasmodium falciparum* infection. But the mortality per unit exposure due to malarial infection is least in Chhattisgarh.

Results and Conclusion: This statewise malaria death comparison can be helpful to authorize the state specific malaria control programme, as several issues which are not clear from count data.

Key Words: Plasmodium falciparum, Bayesian estimation, Shrikage Estimator, Public health.

Introduction

Malaria is endemic in tropical and subtropical regions and is therefore a disease of hot and humid climates[1] and is caused by *Plasmodium falciparum* infection[2]. Annual malaria morbidity alone is 300-500 million people and has a conservative mortality estimate of 700,000-2.7 million lives[3]. In sub-Saharan Africa, where 90 percent of the world's malaria occurs, about 500 million cases are recorded annually with hundreds of thousands of child deaths[4].

India is one of the remaining nine countries outside sub-Saharan Africa where the incidence of malaria is high. In the country, malaria is a major public health problem as 95% population is exposed with the infection[5]. A steady decline has been observed in malaria infected patients since 1976 to 1996 from 6.47 million cases to 2.5 million cases in India. The number of deaths reported from the country was 1006 and 935 respectively in 2003 and 2008 but in 2006 the number of deaths suddenly increased to 17073. Thirteen National Vector- borne Disease Control Programme (NVBDCP) teams routinely monitor *Plasmodium falciparum* drug sensitivity in the country. These teams are located in various regions so as to cover the entire country. *Plasmodium falciparum* monitoring for drug sensitivity is done using the World Health Organization (WHO) methodology of in vivo (28-day) test procedure for determining the status of resistance to CQ and other anti-malarial drugs in *Plasmodium falciparum*. In 2007, Indian National Vector Borne Disease Control Programme (NVB-DCP) reported cases were 1,502,742 and 1,274 deaths.

The success in the eradication of malaria is much lower in countries where economy grows at a much-reduced rate compared to those without much incidence of malaria[3]. In different independent studies conducted by the Indian Council of Medical Research at different time periods it was proved that malaria incidence and death cases have been hugely under-estimated [6–9]. Estimating the true disease burden of malaria in the country is a challenge task considering its varied epidemiology and dynamics of transmission. The topographical difference, hot and humid living conditions at places, difference in the share of average annual rainfall, disparity in the share of resources, unplanned urbanization and several other factors may be held responsible for the unsuccessful attempts in the uprooting diseases like malaria from the country. It is necessary that the issues of public health be studied at the different regional level, going down to each individual states and if possible to the level of districts, at which the implementation of several government policies including those related to public health initiates. However, to understand the extent and pace to which the different public health policies are implemented at the different states it is necessary to estimate the actual burden of the corresponding health problem, with malaria being no exception.

Considering the severity of malaria in the country and the need of estimating the true burden of malaria at the different states of the country provides the backdrop of such a study. The study is set to understand the disease burden of malaria to which the different states are subjected, based on the national malaria control reports published by NVBDCP up to August 2010.

There are endless studies that are concentrated on issues related to the concerned topic. Bose[10] highlighted the issues of high incidence of malaria cases in the desert state of Rajasthan, especially after the monsoons despite the impressive control programs. Amrith[2] express how the political intervention in different public health related issues, including malaria, make the states initiative narrowly targeted. Deb Roy[11] discusses the ambiguity of practicing physicians approach to the treatment of malaria following development of several conflicting theories relating to understanding the disease especially after the malarial epidemic in the third quarter of the previous century. Das Gupta[12] raised issues of sever neglect in the status of public health in the country and holds such negligence responsible for spread of diseases like malaria. Mahajan and others[13] tried to estimate the underlying preferences and use the model to study the adoption of bednets among poor households in rural Orissa from number of reported cases of malaria occurrence based on NVBDCP, 2008 data. Singh and others[14] expressed the effectiveness of rapid malaria diagnostic test over traditional tests in India. However, our search did not enrich any study related to the estimation of true burden of the disease at the different states of the country. The true burden of malaria cannot be understood from the death count data due to reasons explained in the later part of the work.

The primary goal of this study is to compare the states by their number of death reported due to malaria (*Plasmodium falciparum*) in the year 2010. To evaluate the death rate of the states, we use the death report published by NVBDCP, in 2009. It is true that death count in many states due to malaria come down near zero, yet still there are some states where the death count is very high. To increase the strength of estimation, we prefer to use hierarchical Bayesian approach with the help of prior information by the size of mortality. The hierarchical Bayesian models are used when it is believed that the observations are statistically dependent. This generally happens when subjects drawn from the same cluster are more similar to each other compared to subjects from different clusters. For further reading on hierarchical Bayesian approach readers are requested to refer to Lynch[15] or Gill[16].

Using Bayesian approach in identifying risk factor of diseases in a confined population is frequently encountered in several health related studies. Some pioneering works in this regard includes Diggle, Elliott, Morris and Shaddick[17]; Elliott[18]; Volinsky, Magigan, Raftery and Kronmal[19]; Wakefield and Morris[20] etc.

Material and Methods

The data required for the study is based on the National Malaria Control Report published by NVBDCP of 2009 and 2010 till the month of August. The report provides statewise total blood slides examinations, malaria cases, *Plasmodium falcifarum* cases and death in each year. The reports used to publish in each year based on information of the previous years. The statewise comparison of death due to malaria has been performed by comparing the mortality rate per unit exposure rate. The estimate of which for each of the states is obtained through a Bayesian analysis. It is practical to assume the true rates are similar in size that generates the dependency between the parameter. The presence of one states true mortality rate can influence the image of other states true mortality rate. To deal with such problem it is good practice to call a hyper-parameter to reduce the dependency between parameters. The whole process produces the system of a hierarchical prior guiding us to use hierarchical Bayesian approach. All the relevant calculations are performed in the statistical software R(The Software can be freely downloaded from http://cran.r-project.org)

The total number of exposed person due to malaria has been denoted by e. The estimate of mortality rate per unit of exposure rate assumed by λ. It has been assumed that the death count Y follow Poisson distribution with mean $e\lambda$. The standard estimate of λ is, $\hat{\lambda} = y/e$.

The comparison due to death rate among the states in 2010, has been performed by the prior information of mortality rate generated by reports of 2009 death counts. In the annual state report many state's death count is zero or nearer to zero and many states are having higher number of death counts. The prior information about death has been obtained from 10 randomly selected states as a representative of malaria endemic and epidemic area.

Here, Z_i represents the number of deaths in the i^{th} state and O_i represents the number of people exposure due to *Plasmodium falciparum* in the year 2009. It is assumed that Z_i will follow Poisson distribution of mean λ. Initially λ assigns as a standard non-informative prior by, $P(\lambda) = \lambda^{-1}$ and the distribution for λ, given the form of 35 states becomes $P(\lambda) \propto \lambda^{-1} exp(-\beta\lambda)$, $\lambda > 0$. The gamma (α, β) prior for λ has been use by $\alpha = \sum_{i=1}^{35} Z_i$ and $\beta = \sum_{i=1}^{35} O_i$

In 2009 data, $\sum_{i=1}^{35} Z_i$ = 336 and $\beta = \sum_{i=1}^{35} O_i$ = 346272. We assign, gamma prior for λ with parameters (336, 346272).

The observed number of deaths due to the *Plasmodium falciparum* is denoted by y and it is supposed that for the particular state with exposure of e the distribution will be Poisson($e\lambda$).

In the prior model λ assigned with gamma (α, β) and the posterior distribution becomes in the form of gamma($\alpha + y$, $\beta + e$) The predictive density of y is,

$$f(y) = \frac{f(y \mid \lambda) g(\lambda)}{g(\lambda \mid y)} \quad \ldots(1)$$

for $f(y|\lambda) \sim$ Poi(eλ). In the sampling $g(\lambda)$, $g(\lambda|y)$ is the prior and posterior density of λ.

In the first step, the death rate λ_i assumed to be generate from gamma(α, α/μ) distribution with the mean μ and μ^2/α.

$$G(\lambda \mid \alpha, \mu) = \frac{(\alpha/\mu)^{\alpha} \lambda^{\alpha-1} \exp(-\alpha\lambda/\mu)}{\Gamma(\alpha)}, \ \alpha, \lambda > 0 \quad \ldots(2)$$

In the second step, μ and α are assumed to be independent and μ is in gamma prior by gamma(a,b) and α has a density function of $g(\alpha) = \frac{V}{(\alpha + v_0)^2}$, $\alpha > 0$.

V is the median value of α. As α tends to infinity rate λ_i's will concentrated in the same line by $\lambda_1 = \lambda_2 = \cdots = \lambda_{35}$

The posterior distribution of λ_i is gamma ($y_i + \alpha$, $e_i + \alpha/\mu$). The posterior mean of λ_i can be express by

$$E(\lambda_i \mid y, \alpha, \mu) = \frac{y_i + \alpha}{e_i + \alpha/\mu} \quad \ldots(3)$$

The shrinkage estimator B_i can be useful in place of λ_i to know the true posterior mean.

The shrinkage estimator can be replaced in equation (1) by,

$$E(\lambda_i \mid y, \alpha, \mu) = \frac{y_i + \alpha}{e_i + \alpha/\mu} = (1 - B_i)\frac{y_i}{e_i} + B_i\mu \quad \ldots(4)$$

where $B_i = \frac{\alpha}{\alpha + e_i}$. This estimator is useful to improving the estimation by reducing the mean squared error towards zero. Shrinkage is implicit in Bayesian inference.

The number of deaths in one year for *Plasmodium falcifarum* has been reported for each of the 35 states. Let y_i and e_i is the number of deaths and exposure for the i^{th} state. We assumed that the number of deaths y_i follows a Poisson distribution with mean $e_i\lambda_i$ and the objective is to estimate the mortality rate per unit exposure $e_i\lambda_i$. The fraction y_i/e_i is the number of deaths per unit exposure and can be viewed as an estimate of the death rate for the i^{th} state. We plot the ratios y_i/e_i against the logarithms of the exposures $log(e_i)$ for all states where each point is labeled by the number of observed deaths y_i. The estimated rates are highly variable, especially for programs with small exposures. The states experiencing no deaths (a plotting label of 0) also are primarily associated with small exposures. Suppose we are interested in simultaneously estimating the true mortality rates λ_i for all states. One option is to simply estimate the true rates by the individual death rates $y_1/e_1, \cdots, y_{35}/e_{35}$.

Unfortunately, these individual rates can be poor estimates, especially for the states with small exposures. We saw that some of these states did not experience any deaths and the individual death rate $y_i/e_i = 0$ would likely underestimate the states' true risk of mortality. Also it is found that the rates for the states with small exposures have high variability.

Since the individual death rates are not reliable estimates of the actual situation, so it seems desirable to combine the individual estimates in some way to obtain improved estimates.

Suppose we assume that the true mortality rates are equal across states; that is, $\lambda_1 = \cdot\cdot\cdot = \lambda_{35}$. Under this "equal-means" Poisson model, the estimate of the mortality rate for the i^{th} state would be the pooled estimate $\sum_{i=1}^{35} y_i / \sum_{i=1}^{35} e_i$. But this pooled estimate is based on the strong assumption that the true mortality rate is the same across states. This is questionable since one would expect some variation in the true rates. We have discussed two possible estimates for the mortality rate of the i^{th} states: the individual estimate y_i/e_i and the pooled estimate $\sum_{i=1}^{35} y_i / \sum_{i=1}^{35} e_i$.

A third likelihood is the cooperate estimate,

$$(1-\gamma)\frac{y_i}{e_i} + \gamma\frac{\sum_{i=1}^{35} y_i}{\sum_{i=1}^{35} e_i} \quad \ldots(5)$$

This estimate shrinks or moves the individual estimate y_i/e_i toward the pooled estimate $\sum_{i=1}^{35} y_i / \sum_{i=1}^{35} e_i$, where the parameter $0 < \gamma < 1$ determines the size of the shrinkage. We use the posterior mean B_i as a representation of the i^{th} state shrinkage.

Results

We find the best states by using the smallest estimated mortality of states. The posterior mean of the mortality has been computed from the expectation of equation (1). We observed that Chhattisgarh as the one with smallest followed by Madhya Pradesh and Arunachal Pradesh. Kerala is the state with the highest true mortality status followed by Manipur and Maharashtra. To compare the best state "Chattisgarh" with the remaining states, we displayed the statewise rank in Table 1. Table 1 gives the probability P($\lambda_{beststate} < \lambda_i$) for all i, where $\lambda_{beststate}$ represents the rate for "Chattisgarh" in this exercise. We have shown the probabilities for all the 35 states in column 4 of Table 1. The state "Chattisgarh" is better than most of these states as it's posterior probability is close to zero.

Table1. Malaria Status with respect to best state

State	PF cases	Death	$P(\lambda_{beststate} < \lambda_i) \times 10^6$	Rank
Andhra Pradesh	**12510**	**11**	**901.95**	**10**
Arunachal Pradesh	**2314**	**0**	**196.95**	**3**
Assam	**30817**	**28**	**918.65**	**11**
Bihar	**291**	**1**	**2834.32**	**29**
Chhattisgarh	**32522**	**2**	**80.73**	**1**
Goa	**162**	**1**	**3935.32**	**31**
Gujarat	**1340**	**1**	**943.61**	**12**
Haryana	**13**	**0**	**2089.35**	**18**
Himachal Pradesh	**0**	**0**	**2354.8**	**27**
Jammu and Kashmir	**11**	**0**	**2238.77**	**24**
Jharkhand	**28861**	**8**	**296.35**	**5**
Karnataka	**3368**	**1**	**432.85**	**6**
Kerala	**126**	**3**	**9989.18**	**35**
Madhya Pradesh	**4799**	**0**	**113.58**	**2**
Maharashtra	**9390**	**60**	**6291.26**	**33**
Manipur	**217**	**4**	**9737.24**	**34**
Meghalaya	**24175**	**52**	**2146.12**	**20**
Mizoram	**7363**	**12**	**1619.88**	**16**
Nagaland	**963**	**1**	**1285.25**	**13**
Orissa	**166459**	**120**	**724.93**	**9**
Punjab	**4**	**0**	**2132.16**	**19**
Rajasthan	**149**	**0**	**1348.29**	**14**
Sikkim	**8**	**0**	**2225.16**	**23**
TamilNadu	**248**	**1**	**3120.46**	**30**
Tripura	**12868**	**2**	**197.48**	**4**
Uttarakhand	**21**	**0**	**2053.21**	**17**
UttarPradesh	**59**	**0**	**1617.04**	**15**
WestBengal	**6076**	**28**	**4438.71**	**32**
A.N.Islands	**624**	**0**	**618.12**	**7**
Chandigarh	**0**	**0**	**2305.65**	**25**
DNHaveli	**503**	**0**	**710.90**	**8**
Daman andDiu	**10**	**0**	**2197.08**	**22**
Delhi	**1**	**0**	**233.65**	**26**
Lakshadweep	**0**	**0**	**2427.01**	**28**
Puducherry	**0**	**0**	**2196.74**	**21**

Discussion

Poor economic condition and deplorable conditions of living of the people in the country is a hindrance in the control of Malaria in spite of several efforts from the government and NGOs. The statewise ranking based on malaria status can be computed in each year based on its status report presented by NVBDCP. This statewise malaria death comparison can be helpful in providing necessary guidelines for planning the course of action for the state specific malaria control programme.

In India, maximum malaria cases are recorded in Orissa (Table 1). Similarly, in the other states inhabited by ethnic tribes mainly in the forest ecosystems, meso- to hyper-endemic conditions of malaria exist with the preponderance of *Plasmodium falciparum* to the extent of 90% or even more. During August 2009 to July, 2010 of resurgence of malaria, certain states in India like Orissa, Chhattisgarh, Karnataka, Meghalaya, Jharkhand and Assam are found to have high incidence of *Plasmodium falciparum* infection. At the same time period Kerala, Maharashtra, Manipur and Orissa performed worst due to high amount of malaria deaths. Among all the states Punjab, Sikkim and Jammu and Kashmir performed best recording zero number of death and lowest number of malaria cases. However, when the true burden of malaria is considered via the posterior expectation of λ the ranking of the states showed several changes as evident from Table-1.

It is interesting to note that Kerela, with a high human development index, highest literacy rate, and even less number of *Plasmodium falciparum* malaria cases has a high mortality rate due to malaria and is worst compared to the other states. This may be attributed to some chance cause occurring in a particular year or may be due to some hidden reasons. But this definitely calls for serious concern in subsequent years.

References

1. Alumanah EO, Onyeneke EC, Okpoba NA and Okonkwo CJ. Serum protein index in human malaria, *Cylobios* 2000; *103*: *61-4*.
2. Amrith S. Political Culture of Health in India. *Economic and Political Weekly* 2007; *42*(2): 114-121.
3. Breman JG, Egan A and Keusch GT. Multilateral Malaria Initiative (MIM). The Intolerable Burden of Malaria: A New Look at the Numbers, *Supplement to Am J Trop Med Hyg*. 2001; *64*: iv-vii.
4. Asagba SO, Eriyamremu GE and George BO and Okoro I. Biochemical indices of severity in human malaria, *Journal of Medical Science* 2010; *10(4)*: 87-92.
5. National Vector Borne Disease Control Programme.http://nvbdcp.gov.in/malaria-new.html
6. Mrazek MF and Mossialos E. Stimulating pharmaceutical research and development for neglected diseases, *Health Policy,* 2003; *64*(1)*:* 75-88.
7. Snow RW, Guerra CA, Noor AM, Myint HY and Hay SI. The global distribution of clinical episodes of Plasmodium falciparum malaria. *Nature,* 2005; *434*: 214-7.
8. Choudhury DS, Malhotra MS, Shukla RP, Ghosh SK, Sharma VP. Resurgence of malaria in Gadarpur PHC district Nainital, Uttar Pradesh. *Indian J Malariol*. 1983; *20*: 49-58.
9. Yadav RS, Bhatt RM, Kohli VK, Sharma VP. The burden of malaria in Ahmedabad city, India - a retrospective analysis of reported cases and deaths. *Ann Trop Med Parasit*.; 2003; *97*: 793-802.
10. Bose A. Malaria deaths in Rajasthan desert. *Economic and Political Weekly*, 2004; *June 19*: 2557-2559.
11. Deb Roy R. Secluded and Proximate Illiterates among Couples. *Economic and Political Weekly*, 2007; *42*(2): 62-70.
12. Das Gupta M. Mal-areas of health. *Economic and Political Weekly*, 2005; *40*(49): 5159-5166.
13. Mahajan A., Tarozzi A., Yoong J. and Blackburn B. Bednets, Information and Malaria in Orissa.2009; June 9 Available at SSRN: http://ssrn.com/abstract=1533690.
14. Singh N, Shukla MM, Shukla MK, Mehra RK, Sharma S, Bharti PK, Singh M, Singh A and Gunasekar A. Field and laboratory comparative evaluation of rapid malaria diagnostic tests versus traditional and molecular techniques in India. *Malaria Journal*, 2010; *9*: 191-203.
15. Lynch SM. *Introduction to Applied Bayesian Statistics and Estimation for Social Scientists.* Springer Science+Business Media, LCC, 2007; 231-237.
16. Gill J. *Bayesian Methods: A Social and Behavioral Sciences Approach.* Second Edition, Chapman & Hall/CRC Press, Boca Raton, 2006; 396-400.
17. Diggle PJ, Elliott P, Morris SE and Shaddick G. Regression modeling of disease risk in relation to point sources. *Journal of Royal Statistical Society,* 1997; Ser. A, 160: 491-505.
18. Elliott, P. Small-Area Studies in *Environmental Epidemiology: Exposure and Disease,* Eds. R. Bertollini and M. Lebowitz, Lewis Publishers, 1996; 187-199.
19. Volinsky CT, Magigan D, Raftery AE and Kronmal RA. Bayesian model assessing in proportional hazard model: Assessing the risk of stroke. *Applied Statistics*, 1997: *46*(4); 433-448.
20. Wakefield JC and Morris SE. The Bayesian modeling of disease risk in relation to a point source. *Journal of American Statistical Association,* 2001: *96*(453); 77-91.