Available online at www.elixirpublishers.com (Elixir International Journal)

Statistics

Elixir Statistics 62 (2013) 17901-17907

ISSN: 2229-712X

Bayesian Analysis for Epidemiological Study of Child Mortality on the District Level of Uttar Pradesh

G.P. Singh[1,*], A. Tripathi[1], S.K.Singh[2] and Umesh Singh[2]

[1]Department of Community medicine and DST-CIMS, BHU, Varanasi.

[2]Department of Statistics and DST-CIMS, BHU, Varanasi.

ARTICLE INFO

Article history:

Received: 11 July 2013;

Received in revised form: 28 August 2013;

Accepted: 17 September 2013;

Keywords

Child mortality,

Epidemiological,

Bayesian,

Beta-binomial.

ABSTRACT

In this study an attempt has been made to describe the analysis of epidemiological study by Bayesian methods and apply this methodology to district level child mortality of Uttar Pradesh to assign rank to each district for rural and urban separately. The specific objectives of this study are to analysis of epidemiological study by use of fixed effect modeling and random effect modeling in Bayesian setup. To assign rank to each district by this suggest applying strategies to reduce child mortality in those district for those ranks are poor in Uttar Pradesh. The modified retrospective cohort study design used here. For fixed effect modeling beta-binomial modeling approach is used and for random effect modeling logit link function is used. The posterior estimates came in both cases under squared error loss function.

Introduction

The contribution of statisticians to the development of Epidemiological methodology is perhaps the most important contribution that they have made to public health and biomedicine. The central theme of a Epidemiological study is to compare a group of subjects (cases) having the outcome (typically a disease or malfunction) to a control group (not having the malfunction or disease) or In second case compare a group of subjects (Exposed to factor) to a group (not exposed to factor). The Second condition which I discussed above is my interest called cohort study design. The starting point of a cohort study is the recording of healthy individuals with and without exposure to the putative cause being studied. A number of statistical parameter is calculated in this study design to asses our goal. But the most important parameter is the calculation of probability of occurring of event (diseases or malfunction), it is also known as incident risk of occurring event. The event is treated as Bernoulli trial because every subject in follow up has only two outcome either he will have event or not have event, so by use of classical statistical procedure the probability of occurrence of event calculated easily. But in performing such studies various problem arises of sample size its representation of population for drawing authentic result these problems become more dominant in rare event situation so in that situation if we are going to estimate parameter(i. e. Probability),the estimate came are not authentic and reliable. So in this proposed to use a new procedure known as Bayesian inference procedure primarily discovered by Thomas bayes(1763). The Bayesian approach to statistical inference collate all pre existing information, reflecting both evidence based on past studies and current beliefs(as prior belief about parameters).The new evidence from the data collected during the current study is summarized by the likelihood function and the last step in the Bayesian process is to combine the prior distribution with the likelihood function using Bayes' theorem. The result of this process, called the posterior probability distribution, is an updated reflection of our belief about the statistical parameters and has a probabilistic interpretation analogous to the prior distribution. e.g.

If we want to estimate a parameter function $\Phi(\theta)$,Suppose the prior distribution of θ be $p(\theta)$.

After observing the data, suppose the likelihood function be $p(x|\theta)$. Using Bayes rule the posterior distribution of the parameter θ be denoted as $p(\theta|x)$. The Bayes' formula for the posterior distribution of the parameter θ is as follows:

$$\text{Posterior } p(\Theta \mid x) = \frac{\text{prior} \times \text{likelihood}}{\text{m arg inal}} = \frac{p(\Theta) \times p(x \mid \Theta)}{\int p(\Theta) \times p(x \mid \Theta) d\Theta}$$
$$\propto p(\Theta) \times p(x \mid \Theta)$$

Once the posterior distribution of the parameter θ is obtained, we may easily obtain an estimate of $\Phi(\theta)$, any function of θ, under the chosen loss function depending on the nature of decision making. Here we choose squared error loss function, for which posterior mean $E(\Phi(\theta)|x)$ is the Bayes estimator which can be obtained from the following formula:

$$E(\phi(\theta) \mid x) = \int \phi(\theta) p(\theta \mid x) d\theta$$

Apply this methodology to study the child mortality situation prevailing at different district of Uttar Pradesh and find out the rank of various district (rural urban separately) on the basis of probability of experiencing at least one child death by a women during her reproductive life span. Bhattacharyya (2009) perform a Bayesian analysis of child mortality at state level of India.

Tele:

E-mail addresses: ashoka07bhu@yahoo.in

Objective of the study

In this study an attempt has been made to describe the analysis of epidemiological study by Bayesian methods and apply this methodology to district level child mortality of Uttar Pradesh to assign rank to each district. The specific objectives of this study are:

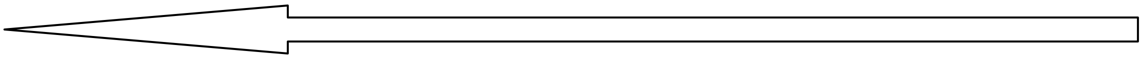
1- The analysis of epidemiological study by use of fixed effect modeling and random effect modeling in Bayesian setup.
2- To assign rank to each district by this suggest applying strategies to reduce child mortality in those district for those ranks are poor in Uttar Pradesh.

Methodology

Modified Retrospective Cohort Study Design:

In this design we select a cohort of females who are completing their fertility span and belonging to the age group of 15-49. The outcomes have occurred before selecting the cohort and we follow the cohort retrospectively and know the probability of occurring of event.

Exposure factor	Child Death Occur (Ot+)	Child Death Not Occur (Ot-)	Total(Cohort)
District[1]	r_1	n_1 -r_1	n_1
District[2]	r_2	n_2 -r_2	n_2
.			
.			
District[i]	r_i	n_i -r_i	n_i
.			
.			
District[70]	r_{70}	n_{70} -r_{70}	n_{70}



Follow Up

Fixed Effect Modeling Process in Bayesian setup :

Let there be k Number of District observed as k groups or classes. Suppose there are n_i Females in i^{th} class divided into two parts- one corresponding to Outcome positive(Ot+) where a Female experiencing child death is available and the other corresponding to Outcome Negative(Ot-) where a Female not experiencing any child death. If p_i is the probability of experiencing child death in the i^{th} class or district, and r_i denotes the number of females who have experiencing child death in this class, the likelihood function (LF) corresponding to i^{th} (i=1, 2,…, k) class can be written as

$$L_i(r_i,n_i \mid p_i) \propto p_i^{r_i}\left(1-p_i\right)^{n_i-r_i} \qquad .....1$$

. A conjugate prior for p_i can be taken to be beta distribution with hyperparameters (a_i, b_i) that is given by

$$g_i(p_i|a_i ,b_i) = p_i^{\,a_i-1} (1- p_i)^{\,b_i-1} / B(a_i ,b_i), \qquad ....2$$

where B(a_i, b_i) is the standard beta function. Obviously, the prior implies that there are a_i females who have experiencing child death out of (a_i+b_i) females who completed her reproductive life span in the i^{th} class (i=1, 2,…, k). Moreover, if both a_i and b_i are taken to be unity, the beta prior given in (2) can be considered as uniform in the range (0, 1). Similarly, if both a_i and b_i are taken to be zero, the beta prior reduces to a vague prior although such choices are strictly outside the parameter space (see Lindley (1965)). These choices for the prior hyperparameters make sense if one is not sure of considering informative priors and rather prefers to work with weak priors.

Combining the LF (1) with the prior (2) via the Bayes theorem yields the posterior distribution of p_i as Beta distribution, that is,

$$P_i(p_i \mid r_i) \propto \quad p_i^{a_i+r_i-1}.(1-p_i)^{b_i+n_i-r_i-1} \qquad .....3$$

Thus inferences on p_i can be easily drawn from (3). Suppose, one assumes the squared error loss function, the Bayes estimator of p_i which is the posterior mean can be easily written as

$$\hat{p}_i = \frac{a_i + r_i}{a_i + b_i + n_i} \qquad ......4$$

The posterior risk of this estimator corresponding to squared error loss function can also be obtained in close form and this can be written as

$$P_R(\hat{p}_i \mid r_i; a_i, b_i) = \frac{(a_i + r_i)(b_i + n_i - r_i)}{(a_i + b_i + n_i)^2 (a_i + b_i + n_i + 1)} \quad ....5$$

Since (5) denotes the posterior expected loss in the estimation of p_i, the objective includes to minimize its value. A possible way includes the choice of hyperparameters a_i, and b_i in such a way that these minimize (5). It is to be noted that if the variation in the values of a_i, and b_i does not have any major effect on (5), one can go with any choice of a_i, and b_i. So in analysis of the data first we determine the value of a_i and b_i .

Random Effect Modelling in Bayesian Setup :

From eqn.1 we take p_i which is show the probability of experiencing child death by a female in districts i. we consider p_i has some random fluctuation for taking p_i in real line scale we take logit link function and taken model as follows:

$$\text{logit}(p_i) = \log\left[\frac{p_i}{1 - p_i}\right] = b_i$$

Now b_i can take any real value and hence we may assume that b_i is normally distributed with mean μ and σ^2 . It is also says as follows

$$\text{logit}(p_i) = b_i \sim N(\mu, \sigma^2)$$

Non-informative priors were selected for these parameters. The probability is calculated under squared error loss function with use of Monte Carlo Markov chain process. Here Gibbs sampling procedure used for this purpose.

Data

The Data taken District Level Household and Facility Survey (DLHS-3) is third in the series preceded by DLHS-1 in 1998-99 and DLHS-2 in 2002-2004. DLHS-3(2007-2008) is designed to provide estimates on maternal and child health, family planning and other reproductive health indicators. The Data was collected from 7,20,320 households from 34 states and union territories of india(excluding Nagaland). From these households, 6,43,944 ever married women aged 15-49 years and 1,66,260 unmarried women aged 15-24 years are interviewed DLHS-3 adopted a multi-stage stratified probability proportion to size sampling design. The data according our methodology is taken from given data. The description of used data is given as following table. The data taken from state Of Uttar Pradesh at each district level and on the basis of locality of female it categorize in two parts as rural and urban.

Table-1: District and locality wise classification of females(contain Ot+ and Ot-).

District of Uttar Pradesh	Type of locality						Final Total
	Rural			Urban			
	Ot+	Ot-	Total	Ot+	Ot-	Total	
Saharanpur	91	270	361	38	92	130	491
Muzaffarnagar	144	285	429	37	139	176	605
Bijnor	120	262	382	47	84	131	513
Moradabad	212	282	494	49	169	218	712
Rampur	254	350	604	55	102	157	761
Jyotiba phule nagar	198	365	563	64	105	169	732
Meerut	74	140	214	39	156	195	409
Baghpat	157	338	495	38	104	142	637
Ghaziabad	79	139	218	89	251	340	558
Gautam buddha nagar	93	176	269	57	130	187	456
Bulandshahar	182	274	456	33	101	134	590
Aligarh	169	235	404	59	129	188	592
Hathras	123	227	350	17	69	86	436
Mathura	130	193	323	47	87	134	457
Agra	86	136	222	60	125	185	407
Firozabad	292	287	579	53	151	204	783
Etah	252	279	531	58	75	133	664
Mainpuri	321	431	752	27	69	96	848
Budaun	316	311	627	44	75	119	746
Bareilly	220	254	474	93	175	268	742
Pilibhit	239	319	558	36	98	134	692
Shahjahanpur	275	270	545	47	94	141	686
Kheri	251	279	530	14	39	53	583
Sitapur	211	269	480	29	48	77	557
Hardoi	307	324	631	44	72	116	747
Unnao	160	225	385	25	43	68	453
Lucknow	41	66	107	58	155	213	320

Rae bareli	175	256	431	18	30	48	479
Farrukhabad	284	317	601	61	104	165	766
Kannauj	280	360	640	40	83	123	763
Etawah	155	247	402	45	105	150	552
Auraiya	194	298	492	21	61	82	574
Kanpur dehat	142	350	492	13	29	42	534
Kanpur nagar	33	82	115	71	216	287	402
Jalaun	83	153	236	25	72	97	333
Jhansi	27	67	94	27	94	121	215
Lalitpur	142	147	289	7	51	58	347
Hamirpur	97	151	248	16	45	61	309
Mahoba	83	120	203	24	59	83	286
Banda	203	253	456	21	56	77	533
Chitrakoot	204	205	409	13	32	45	454
Fatehpur	172	243	415	16	30	46	461
Pratapgarh	140	245	385	3	21	24	409
Kaushambi	224	230	454	15	34	49	503
Allahabad	148	166	314	23	87	110	424
Barabanki	165	237	402	24	32	56	458
Faizabad	145	225	370	12	35	47	417
Ambedaker nagar	152	296	448	10	31	41	489
Sultanpur	194	310	504	7	31	38	542
Bahraich	216	275	491	19	57	76	567
Shrawasti	233	289	522	3	7	10	532
Balrampur	245	280	525	15	30	45	570
Gonda	155	259	414	11	25	36	450
Siddharthnagar	218	271	489	9	14	23	512
Basti	241	405	646	9	30	39	685
Sant kabir nagar	169	330	499	5	42	47	546
Maharajganj	175	275	450	6	22	28	478
Gorakhpur	121	253	374	19	79	98	472
Kushinagar	140	258	398	3	5	8	406
Deoria	91	324	415	10	32	42	457
Azamgarh	161	339	500	11	29	40	540
Mau	117	337	454	37	65	102	556
Ballia	133	297	430	13	35	48	478
Jaunpur	123	271	394	8	27	35	429
Ghazipur	155	310	465	11	17	28	493
Chandauli	103	201	304	12	33	45	349
Varanasi	67	137	204	35	128	163	367
Sant ravidas nagar	211	265	476	26	41	67	543
Mirzapur	129	169	298	22	38	60	358
Sonbhadra	113	153	266	10	42	52	318
Total (Uttar Pradesh)	11755	17642	29397	2063	5073	7136	36533

Result

At the first step of analysis we find the value of hyper parameters a_i and b_i we plot the value of a and b with posterior risk at fixing one to vary other hyper parameter for the rural and urban separately in following Figures where each district show a line in Figure, posterior risk taken on y axis and hyperparameters value taken on x axis.

Figure-1: Post risk for the varying value of b at constant value a(for all a_i) =2(rural)

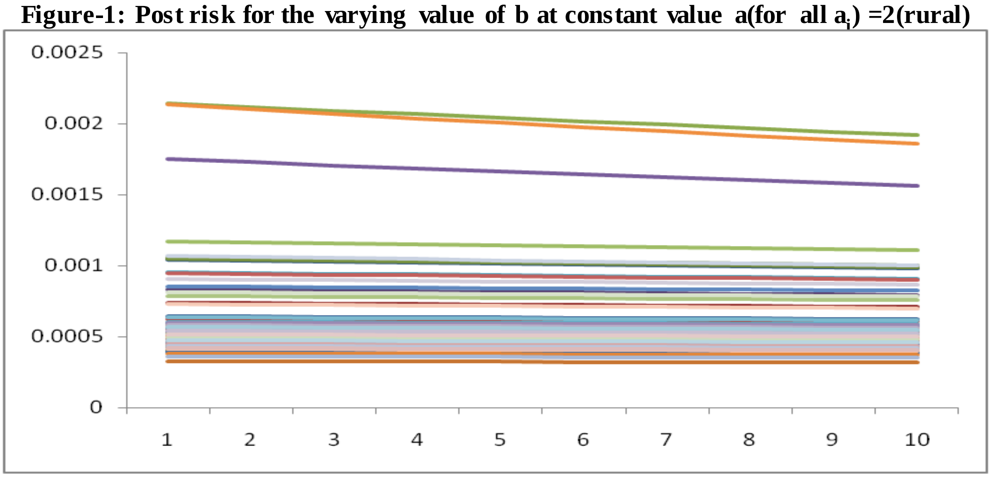


Figure-2:Post risk for the varying value of b at constant value a(for all a_i) =2(urban)

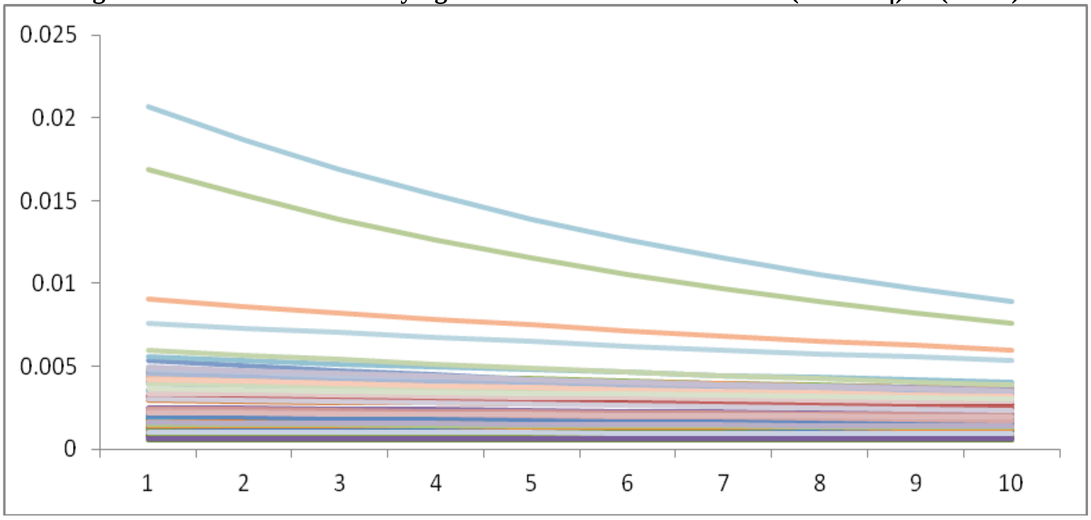


Figure-3 : Post risk for the varying value of a at constant value b(for all b_i) =2(rural).

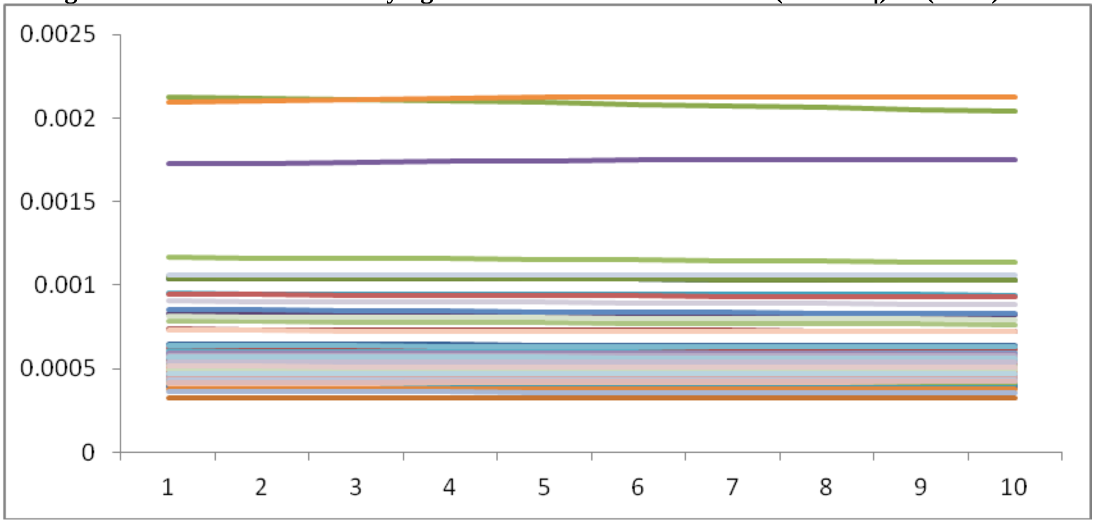
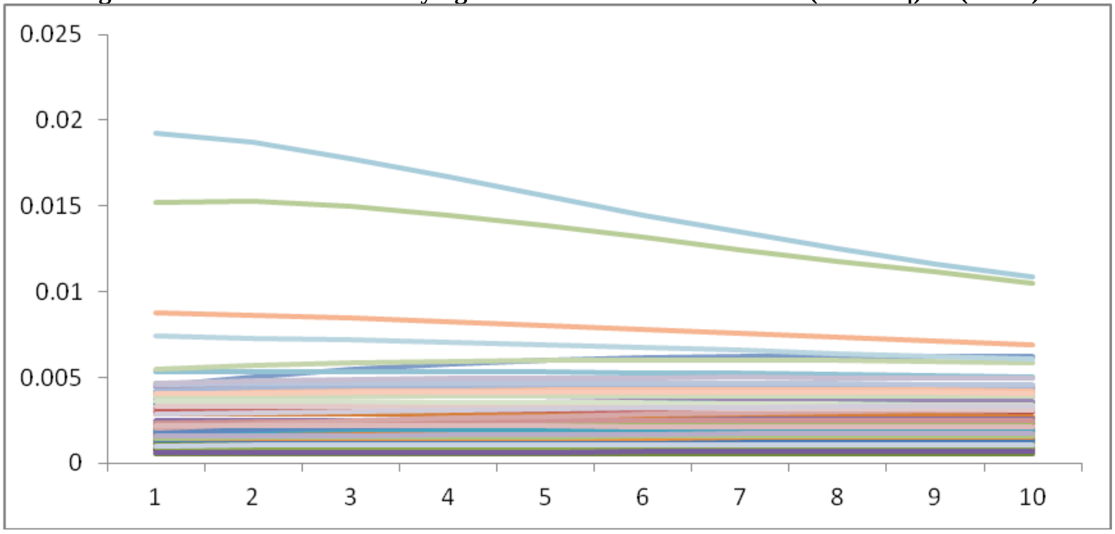


Figure-4: Post risk for the varying value of a at constant value b(for all b_i) =2(urban)



It is obvious from the Figures that the posterior risks, in general, are quite small (of the order 10^{-3} to 10^{-4}) and there is no significant variation in the values of posterior risks for the changing values of either a_i or b_i. An overall recommendation for the choice can be considered as a small value for a_i and moderately large value for b_i in order to have the small values of posterior risks. Our final results obtained to take a_i =1.5 and b_i =8.0 for both rural and urban where these values of a_i and b_i have been approximately assessed on the basis of figure 1-4.

Now we consider the data of table -1 and analyze it according to the given methodology the result obtained from the data is given in table-2. In table -2 probability of child death with assign rank for each district by given methodology for rural and urban separately.

It is find from table -2 that both the model give nearly same result but REM show slighter higher values of probability of child death. most of district shows the probability of child death higher in rural areas as compared to urban areas but it is quite appreciable that six district name “Saharanpur, bijnor, jyotiba phule nagar,deoria, mau, ghazipur”(“Saharanpur, Bijnor, Deoria, Mau” in case of REM) shows lower risk of child death in rural areas as compared to urban areas. In rural areas deoria has lowest probability of child death while district budaun have highest. In urban areas sant kabir nagar(Lalitpur in case of REM) has lowest probability of child death while district etah have highest. The rank assign to each district according to lowest to highest probability.

Table-2 : District and locality wise classification of Probability and rank for rural, urban separately

District	Random Effect Model (REM)				Fixed Effect Model (FEM)			
	Rural		Urban		Rural		Urban	
	Probability	Rank	Probability	Rank	Probability	Rank	Probability	Rank
Saharanpur	0.273	2	0.29	40	0.250	2	0.283	44
Muzaffarnagar	0.343	14	0.238	4	0.332	15	0.208	11
Bijnor	0.325	8	0.33	62	0.310	9	0.345	61
Moradabad	0.425	49	0.244	7	0.424	49	0.222	16
Rampur	0.418	45	0.328	59	0.416	46	0.339	58
Jyotiba phule nagar	0.356	20	0.348	69	0.348	24	0.367	68
Meerut	0.356	21	0.23	3	0.338	19	0.198	8
Baghpat	0.325	9	0.275	26	0.314	10	0.261	35
Ghaziabad	0.369	26	0.267	16	0.354	25	0.259	33
Gautam buddha nagar	0.354	19	0.299	46	0.339	20	0.298	46
Bulandshahar	0.398	37	0.263	15	0.394	37	0.240	22
Aligarh	0.415	44	0.305	48	0.412	44	0.306	48
Hathras	0.357	23	0.245	8	0.346	22	0.194	7
Mathura	0.401	38	0.325	58	0.395	38	0.338	55
Agra	0.389	32	0.312	52	0.378	30	0.316	51
Firozabad	0.494	69	0.268	17	0.499	69	0.255	31
Etah	0.467	63	0.378	70	0.469	63	0.418	70
Mainpuri	0.424	47	0.284	36	0.424	47	0.270	39
Budaun	0.495	70	0.335	65	0.499	70	0.354	64
Bareilly	0.456	58	0.332	64	0.458	58	0.341	59
Pilibhit	0.425	48	0.277	27	0.424	48	0.261	36
Shahjahanpur	0.494	68	0.316	54	0.499	68	0.322	52
Kheri	0.466	61	0.28	31	0.468	62	0.248	26
Sitapur	0.435	52	0.33	61	0.434	52	0.353	63
Hardoi	0.479	65	0.34	66	0.482	65	0.363	67
Unnao	0.412	43	0.323	57	0.409	43	0.342	60
Lucknow	0.386	30	0.277	28	0.365	27	0.267	38
Rae bareli	0.404	39	0.32	55	0.401	40	0.339	57
Farrukhabad	0.466	61	0.341	67	0.468	61	0.358	65
Kannauj	0.434	51	0.309	50	0.433	51	0.313	49
Etawah	0.386	31	0.295	44	0.380	31	0.292	45
Auraiya	0.394	36	0.273	24	0.390	36	0.246	24
Kanpur dehat	0.3	4	0.295	43	0.286	6	0.282	42
Kanpur nagar	0.324	7	0.258	13	0.277	5	0.245	23
Jalaun	0.36	24	0.272	22	0.344	21	0.249	28
Jhansi	0.329	10	0.251	10	0.275	4	0.218	15
Lalitpur	0.475	64	0.225	1	0.481	64	0.126	2
Hamirpur	0.391	34	0.278	30	0.383	33	0.248	27
Mahoba	0.405	40	0.288	37	0.398	39	0.276	41
Banda	0.439	55	0.281	32	0.439	55	0.260	34
Chitrakoot	0.486	67	0.288	37	0.491	67	0.266	37
Fatehpur	0.412	42	0.309	49	0.409	42	0.315	50
Pratapgarh	0.368	25	0.255	12	0.359	26	0.134	3
Kaushambi	0.482	66	0.295	42	0.487	66	0.282	43
Allahabad	0.459	59	0.246	9	0.462	60	0.205	10
Barabanki	0.408	41	0.344	68	0.405	41	0.389	69
Faizabad	0.392	35	0.277	29	0.386	35	0.239	20
Ambedaker nagar	0.346	17	0.274	25	0.336	18	0.228	18
Sultanpur	0.386	29	0.258	14	0.381	32	0.179	4
Bahraich	0.435	53	0.271	20	0.435	53	0.240	21
Shrawasti	0.441	57	0.29	39	0.441	57	0.231	19
Balrampur	0.459	60	0.304	47	0.461	59	0.303	47
Gonda	0.377	28	0.293	41	0.370	28	0.275	40

Siddharthnagar	0.44	56	0.31	51	0.440	56	0.323	53
Basti	0.375	27	0.271	18	0.370	29	0.216	14
Sant kabir nagar	0.344	16	0.228	2	0.335	17	0.115	1
Maharajganj	0.389	33	0.271	19	0.384	34	0.200	9
Gorakhpur	0.333	12	0.24	5	0.319	12	0.191	6
Kushinagar	0.357	22	0.296	45	0.347	23	0.257	32
Deoria	0.243	1	0.273	23	0.218	1	0.223	17
Azamgarh	0.329	11	0.284	35	0.319	11	0.253	30
Mau	0.274	3	0.328	60	0.256	3	0.345	62
Ballia	0.32	5	0.282	34	0.306	7	0.252	29
Jaunpur	0.323	6	0.271	21	0.309	8	0.213	13
Ghazipur	0.34	13	0.314	53	0.330	14	0.333	54
Chandauli	0.348	18	0.281	33	0.333	16	0.248	25
Varanasi	0.343	15	0.242	6	0.321	13	0.212	12
Sant ravidas nagar	0.438	54	0.332	63	0.438	54	0.359	66
Mirzapur	0.426	50	0.32	56	0.424	50	0.338	56
Sonbhadra	0.419	46	0.254	11	0.416	45	0.187	5

Conclusion & Recommendations:

The paper provides Bayesian analysis for epidemiological study design. Probabilities of child death for female during her reproductive life span" is calculated for the each district of Uttar Pradesh with rural and urban areas separately and assign rank each district accordingly by use of both model.

A number of interpretations have been drawn from the result on the basis of that we have give following main Suggestions:

- Government should intensify efforts at providing facilities by which the child mortality controlled in rural areas because majority of population lives in rural area.
- It also suggested for Uttar Pradesh government that the resource by which the health facilities improve are distributed in each district with its condition of requirement.

Acknowledgement: Present work is a part of research grant obtained from Indian Council of Medical Research (ICMR), New Delhi, India under JRF award No. 3/1/3/JRF-2011/HRD. Authors also give thanks to Mr. Mayank Srivastava for reading manuscript critically.

References

1. Bayes T, Price R (1763). An Essay Towards Solving a Problem in the Doctrine of Chances. By the late Rev. Mr. Bayes, communicated by Mr. Price, in a letter to John Canton, MA.and F.R.S." Philosophical Transactions of the Royal Society of London, 53, 370-418.

2. Bhattacharyya R & Rao A S R S (2009) " Estimation and analysis of child mortality for Indian states through a Bayesian approach",(Webpage),http://www.isical.ac.in/~arni

3. DLHS-3 Distric level health Survey (2007-2008),IIPS Mumbai, India.

4. Lindley, D., V., (1965). Introduction to Probability and Statistics from a Bayesian Viewpoint, Part 2 – Inference, Cambridge, Cambridge University Press.