Journal of the Faculty of Environmental Science and Technology,
Okayama University
Vol.9, No.1, pp.9-17, February 2004

Small Geographic Area Estimation in WinBUGS with Two Approaches Prediction

Agus SETIAWAN* and Tomoyuki TARUMI**

(Received November 25, 2003)

Small Area Estimation (SAE) is the process of using statistical models to link survey outcome variables to a set of predictor variables known for small domains, in order to predict domain-level estimates. The need for detailed statistics on small area is constantly increasing. Small area estimation is becoming important in survey sampling due to a growing demand for reliable small area statistics from both public and private sectors. Bayesian hierarchical models provide a convenient framework for disease mapping and geographical correlation studies. Computation may be carried out using the freely-available WinBUGS software. Two approaches prediction to estimate total patient in small area i will be presented. For the purpose of this paper, the small area estimation in this context use data of Indonesia's population based on the 2000 census for the population of Jakarta and data of patient diarrhea from District Health Service of Jakarta. We interest to predict total patient of diarrhea as variable of interest and data population as auxiliary data from unsample for each small area.

Key words: *Auxiliary data, Population density, Sample survey, Small area estimation, WinBUGS*

1. INTRODUCTION

Sample surveys are widely used to provide estimates of total, mean and other parameters not only for the total population of interest but also for subpopulation (or domain) such as geographic area. The small area estimations are widely used in practice of sample survey to provide estimates parameter interest of geographic area. Small area typically refers to a geographic area for which very little information is obtained from surveys.

What is meant by a small area? By Murdock and Ellis (1991), small area refer to counties and subcounty areas like cities, census tract, ZIP (Zone Improvement Plan (US post office postal code system)) code areas, and even individual blocks while by J. Cuzick and P. Elliot (1992), this will depend on the context, as it relates to the number of cases (in this case about disease) that are observed.

According to Ulrike Schach (Discussion Paper, 2001), the general idea of a small area estimation in its original sense is an intrapolation of information collected on a large spatial scale to local areas within the study region. A geographic domain (area) is regarded as "large (or major)" if the domain-specific sample is large enough to yield "direct estimates" of adequate precision. A domain is regarded as "small" if the domain-specific sample is not large enough to support direct estimates of adequate precision, J.N.K. Rao (2003).

The term "small area" and "local area" are commonly used to denote a small geographical area. Some other terms used to denote a domain with sample size include "local area", "sub domain", "small subgroup", "subprovince", and "minor domain".

Interest in small area estimation methods has further increased. Some other proceedings of conferences on small area estimation include Platek and Singh (1986), Platek, Rao, Särndall and Singh (1987) and J. N. K. Rao (2001). Papers on small area estimation include Noel J. Purcell (1979), Prasad, N. G. N. and Rao, J. N. K. (1990), Malay Ghosh and Rao J. N. K. (1994), and Malay Ghosh, Kannan Natarajan, T. W. F. Stroud, and Bradley P. Carlin (1998).

For the purpose of this paper, the small area estimation in this context use data of Indonesia's population based on the 2000 census for the population of Jakarta and data of patient diarrhea from District Health Service of Jakarta. We

*Graduate School of Natural Science and Technology, Okayama University 3-1-1 Tsushima-Naka, Okayama, Okayama 700-8530, Japan, i_one_my@ems.okayama-u.ac.jp or i_one_my@lycos.com
**Department of Environmental and Mathematical Sciences, Okayama University 3-1-1 Tsushima-Naka, Okayama, Okayama 700-8530, Japan, t2@ems.okayama-u.ac.jp

interest to estimate total patient of diarrhea as variable of interest and data population as auxiliary data from unsample for each small area.

Jakarta is the capital and largest city of the Republic of Indonesia. Situated between 105° 49′ 5″ east longitude and 06° 10′ 7″ south latitude. The city is centrally located within Indonesia on the north-west coast of the island of Java. The city dominates Indonesia's administrative, economic and cultural activities, and is a major commercial centre and transport hub within Asia.

We would like to describe two approaches of estimation of total patient of diarrhea in Jakarta. The first approach is the predictive approach to domain and the second approach is by pass through approach to data individual.

In the Table 1. is showed that the area of Jakarta is 664.12 square kilometers. The special capital city region of Jakarta is divided among 5 municipalities (*kotamadya*), South Jakarta, East Jakarta, Central Jakarta, West Jakarta, and North Jakarta. These municipalities subdivided into subdistricts (*kecamatan*) and villages (*desa/ kelurahan*). Subdistricts and villages are government at lower levels. The special capital city region of Jakarta comprise 5 municipalities, 43 subdistricts and 265 urban.

Table 1. The Number of Area, Subdistrict, Villages and Population Density In Jakarta by Municipalities (2000)

No.	Munici-palities	Area (km^2)	Sub-dis-trict	Village (urban)	Population Density (person / km^2)
1	South Jakarta	145.73	10	65	12,242
2	East Jakarta	188.19	10	65	12,476
3	Central Jakarta	48.08	8	44	18,190
4	West Jakarta	127.11	8	56	14,981
5	North Jakarta	155.01	7	35	9,266
Jakarta		664.12	43	265	12,569

From the aforementioned data, we classified two kinds of data, 25 subdistricts have population density upper 12.569 person per km^2 and 18 subdistricts under 12.569 person per km^2. In this paper, we interest to estimate total patient from subdistrict that have population density upper 12,569 person per km^2. From 25 subdistricts (160 urban area) that have population density upper Jakarta's population density, choosen 75 urban as sample. The population of density in Jakarta Province by subdistricts are presented in Figure 1.

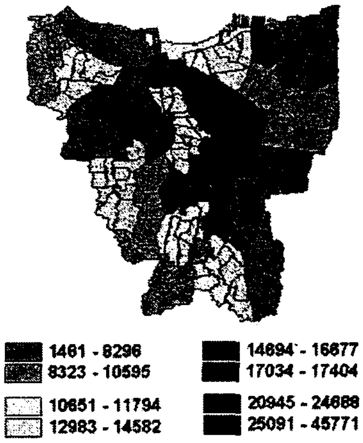

Figure 1. The Population Density of Jakarta Province 2000 Census of Population - Indonesia

This paper focuses on the spatial distribution of the population density of the Jakarta Province at two different levels of scale, densely and rarely populated area. So that, we can see the Jakarta in spatial terms by using different scales of analysis (urban, subdistricts and province).

Figure 1 shows, the most densely populated area signed with the blue color, which have population density between 25,091 – 45,771 person per square kilometers. The most rarely populated area signed with the old brown color, 1,461 – 8,296 person per square kilometers. From Figure 1 also appear that excelsior population density on area then area will be concentrated to center of economic. Because in Jakarta province, the center of economic located in the middle of Jakarta.

The shade of brown colors show areas have belonging population density under 12,569 person per square kilometers, while the shade of green to blue colors show areas have population density ≥ 12,569 person per square kilometers.

Conform by subdistrict that the highest population density (very dense) is Johar Baru (45,771) and Tambora (41,834). Johar Baru with area 2,38 square kilometers is occupied 108,936 person while Tambora with area 5,48 square kilometers is occupied 229,253 person. The lowest population density area (very rare) is Kepulauan Seribu (1,461). Kepulauan Seribu or the thousand islands is a group of small islands scattered in the bay of Jakarta (North Jakarta), actually a total of 110 islands. The population among one island to another is different significantly. In Figure 1 the Kepulauan Seribu is not appear because the position of them are rather scattered.

Table 2. Sample Sizes (n_i). Patient of Diarrhea, and Population Data at Population Density ≥ 12,569 person per km^2 in Jakarta - Indonesia 2000

i	n	j	k	y_{ijk}	x_{ijk}	X_{ik}	i	n	j	k	y_{ijk}	x_{ijk}	X_{ik}	i	n	j	k	y_{ijk}	x_{ijk}	X_{ik}
1	2	1	0	715	3367	10676	8	4	28	0	526	3931	20156	18	4	55	0	253	3364	26390
			1	685	30986	98260				1	946	36177	185513				1	294	30961	242897
		2	0	523	3581				29	0	327	2188				56	0	194	2098	
			1	501	32962					1	587	20139					1	225	19313	
2	6	3	0	113	2399	22467			30	0	352	3546				57	0	485	5106	
			1	132	22082	206786				1	635	32638					1	565	46991	
		4	0	104	1669				31	0	187	2265				58	0	397	4221	
			1	120	15361					1	337	20843					1	463	38852	
		5	0	109	1920		9	3	32	0	397	3621	24240	19	3	59	0	328	6543	34011
			1	126	17668					1	477	33325	223105				1	438	60221	313039
		6	0	174	3447				33	0	696	6251				60	0	156	3579	
			1	202	31722					1	835	57539					1	208	32945	
		7	0	116	1920				34	0	566	5803				61	0	119	2996	
			1	135	17673					1	679	53407					1	158	27578	
		8	0	176	2575		10	2	35	0	303	1861	8683	20	4	62	0	300	2212	13392
			1	205	23698					1	359	17127	79915				1	274	20355	123261
3	3	9	0	172	1454	14022			36	0	211	1387				63	0	309	2811	
			1	294	13384	129058				1	250	12764					1	283	25877	
		10	0	221	2801		11	4	37	0	188	2481	22752			64	0	260	2608	
			1	378	25783					1	227	22840	209416				1	239	24005	
		11	0	304	3312				38	0	282	4537				65	0	149	1178	
			1	519	30487					1	341	41760					1	136	10839	
4	5	12	0	267	2121	19121			39	0	200	3098		21	2	66	0	317	3226	39914
			1	249	19523	175992				1	243	28516					1	368	29695	367373
		13	0	314	2267				40	0	209	2878				67	0	542	5926	
			1	292	20868					1	253	26494					1	630	54542	
		14	0	291	2754		12	1	41	0	272	3436	7979	22	2	68	0	67	3205	18182
			1	270	25352					1	430	31624	73442				1	134	29498	167345
		15	0	314	2319		13	1	42	0	113	2508				69	0	97	4862	
			1	292	21348					1	81	23084					1	198	44755	
		16	0	189	1790		14	3	43	0	390	4691	13060	23	2	70	0	203	4017	33201
			1	175	16472					1	532	43173	120207				1	315	36972	305587
5	3	17	0	134	1354	18491			44	0	252	3017				71	0	177	4480	
			1	156	12466	170190				1	345	27770					1	276	41233	
		18	0	142	2413				45	0	160	1399		24	3	72	0	123	3475	25312
			1	165	22210					1	219	12879					1	165	31986	232974
		19	0	130	2124		15	3	46	0	234	3788	25007			73	0	168	4212	
			1	151	19550					1	421	34861	230165				1	226	38766	
6	4	20	0	115	1845	10549			47	0	254	3293				74	0	184	4321	
			1	133	16981	97092				1	455	30307					1	247	39770	
		21	0	117	1613				48	0	205	3394		25	1	75	0	230	3413	20459
			1	137	14844					1	367	31241					1	245	31418	188307
		22	0	71	735		16	3	49	0	154	1795	18730							
			1	83	6760					1	179	16526	172393							
		23	0	60	752				50	0	161	1779								
			1	70	6924					1	187	16369								
7	4	24	0	413	3212	24901			51	0	352	3995								
			1	561	29566	229190				1	410	36775								
		25	0	476	3305		17	3	52	0	324	4026	13423							
			1	649	30423					1	546	37053	123542							
		26	0	484	4549				53	0	178	2172								
			1	660	41869					1	301	19988								
		27	0	166	918				54	0	148	1911								
			1	227	8454					1	248	17594								

i - small area

n - sample size of village in small area i

j - village

k - age group

y_{ijk} - number total of patient in small area *i*, village *j*, and *k*'s age group

x_{ijk} - number total of population in small area *i*, village *j*, and *k*'s age group

X_{ik} - number total of patient in small area *i* and *k*'s age group

2. SAMPLE SELECTED

In sample selection, we just select sample urban that consistent with the population density of the subdistrict. For simplify, please see Table 2 below

Table 3. The Number of Urban that Selected as Sample

Subdistrict		Village (Urban)		Sample Urban
Population Density	Total	Population Density	Total	
≥ 12,569	25	≥ 12,569	135	75
		< 12,569	25	x

For estimating the total patient in a finite population of small area, we grouped the data of sample according to age group, 0 - 4 years and ≥ 5 years. For this purpose, we use software WinBUGS for analysis of bayesian statistical models with use Markov Chain Monte Carlo (MCMC) techniques to obtain estimates of posterior distributions. The software can be down loaded from http://www.mrc-bsu.cam.ac.uk/bugs/.

To estimate the total patient, we use two approaches namely the predictive approach to domain and by pass through approach to data individual. For those data, please see in Table 2. In process computation, we use "ES" as term for explain the result of estimate from calculation by using predictive approach to domain and "EV" is a term for explain the result of estimate that obtained by approach to data individual. Focus of ES is subdistrict while focus of EV is village.

As shown in the Table 3 above, the number of sample village at area subdistricts which have population density ≥ 12,569 person per square kilometers is about 55% from total urban (135). These sample are consistent with the population density of the subdistrict.

Beside on the population data (as auxiliary variable), in this paper we use data patient of diarrhea as characteristic of interest. From the results of the processing data, later on we can compare difference estimation that happened between approach prediction by domain (ES) and approach prediction by data individual (EV).

3. HIERARCHICAL BAYES

For hierarchical Bayes (HB) small area estimation, the estimates of the parameters of interest for these areas can profitably "borrow strength" from other neighboring areas through appropriate models. Model-based inference is becoming increasingly popular in survey sampling, particularly in small area estimation. The focus of this paper is on hierarchical bayes estimators of binary variable.

A Hierarchical Bayes version of the logit-normal model with unit level covariates $\mathbf{x}_{ij}$ with assume that the model holds for the sample $\{(y_{ij}, x_{ij}), j \in s_i; i = 1, \ldots, m\}$ may be expressed as

(a) $y_{ij} \mid p_{ij} \overset{iid}{\sim} \text{Bernoulli}(p_{ij})$

(b) $\xi_{ij} = \text{logit}(p_{ij}) = \mathbf{x}_{ij}^T \beta + \mathbf{v}_i, v_i \overset{iid}{\sim} N(0, \sigma_v^2)$

(c) β and σ_v^2 are mutually independent with $\sigma_v^{-2} \sim G(a,b)$, $a \geq 0, b > 0$ and prior on β, $f(\beta) \propto 1$

Let Y_{ikj} denote a binary random variable for individual *j* in class *k* and small area *i*. Where $i = 1, \ldots, m$, $j = 1, \ldots, N_{ik}$ and $k = 1, 2$. Within small area *i* and class *k*, and conditional on p_{ik}, the Y_{ikj} are assumed to be independent Bernoulli random variables with $\Pr(Y_{ikj} = 1 \mid p_{ik}) = p_{ik}$. Vector of covariates X_k as auxiliary variable and a vector of regression coefficients β_i are given, the general model is $\text{logit}(p_{ik}) = X_k^t \beta_i$. This model is taken from

$$\text{logit}(p_{ik}) = \alpha_i + \beta_i x_k$$

Where $x_k = 0$ if class k corresponds to age group 0 – 4 years and $x_k = 1$ if age group ≥ 5 years. The prior distribution of parameter are

$$\alpha_i \sim N(\alpha_0, \tau_\alpha)$$

$$\beta_i \sim N(\beta_0, \tau_\beta)$$

$\alpha_0, \beta_0, \tau_\alpha, \tau_\beta$ are given independent "non-informative" priors.These parameter is out of *doodle's box* as shown in Figure 3.

For estimating the total patient in a finite population of small area *i* is:

$$\theta_i = \sum_{k=0}^{1} \sum_{j \in s_{ik}} Y_{ikj} + \sum_{k=0}^{1} \sum_{j \in \bar{s}_{ik}} \hat{Y}_{ikj}$$

where $\bar{s}_{ik}$ is the complement of the sample s_{ik}.

3.1 The Predictive Approach to Domain

Let y_s denote the vector of sample observations with assuming that $y_s \sim \mathrm{Bin}(p_{ik}, ns_{ik})$, where ns_{ik} is number of sample auxiliary variable. Because $E(\hat{Y}_{ikj}|p_{ik}) = p_{ik}$, the posterior expected value of Y_i is

$$E(\theta_i|y_s) = \sum_{k=0}^{1}\sum_{j\in s_{ik}} Y_{ikj} + \sum_{k=0}^{1}\sum_{j\in \bar{s}_{ik}} E(p_{ikj}|y_s)$$

$$= \sum_{k=0}^{1}\sum_{j\in s_{ik}} Y_{ikj} + \sum_{k=0}^{1}\left(X_{ik} - \sum_{j\in s_{ik}} x_{ikj}\right)p_{ik}$$

where the total of x_{ikj} for the unsample unit $\bar{s}_{ik}$ is

$$X_{ik} - \sum_{j\in s_{ik}} x_{ikj}$$

where X_{ik} is the total population in subdistrict i, age group k and x_{ikj} is the number of population in each subdistrict i, age group k and village j (see Table 2).

3.2 By Pass Through Approach to Data Individual

In computing estimation of total by pass through approach to data individual, we divide the value of unsample estimate with the number of sample size in each domain,n_i. This respect due to the number of unsample (=total sample – sample) of auxiliary variable had used n_i times in process of computation to obtain total estimate of domain.

We have

$$\theta_{ij} = \sum_{k=0}^{1} Y_{ikj} + \frac{1}{n_i}\sum_{k=0}^{1}\hat{Y}_{ikj}$$

with the assume $y_s \sim \mathrm{Bin}(p_{jk}, ns_{jk})$ the posterior expected value of Y_{ij} is

$$E(\theta_{ij}|y_s) = \sum_{k=0}^{1} Y_{ikj} + \frac{1}{n_i}\sum_{k=0}^{1}\left(X_{ik} - \sum_{j} x_{ikj}\right)p_{jk}$$

Therefore, accumulation of the expected value of Y_{ij} for each j = 1, …, n_i $(j \in s_{ik})$ will obtained value estimate of Y_i is that

$$\hat{Y}_i = \sum_{j=1}^{n_i} E(\theta_{ij}|y_s)$$

$$= \sum_{j=1}^{n_i}\left(\sum_{k=0}^{1} Y_{ikj} + \frac{1}{n_i}\sum_{k=0}^{1}\left(X_{ik} - \sum_{j} x_{ikj}\right)p_{jk}\right)$$

3.3 Relationship between ES and EV

The value of p_{ik} or p_{jk} between 0 and 1, because the logit function has value between 0 and 1, as shown in this figure below

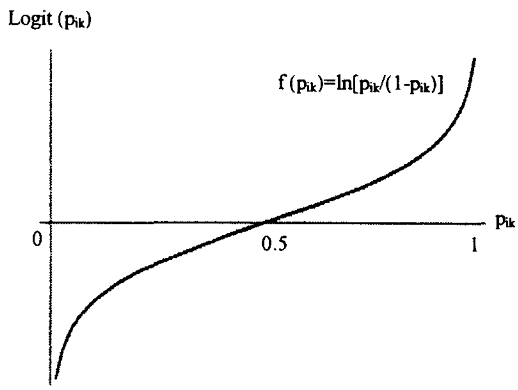

Figure 2. The Curve of Logit Function

To simplify the notation, we write $0 < p_{ik} < 1$ for approach to domain and $0 < p_{jk} < 1$ for approach to data individual.

We compare total estimate of patient diarrhea for each subdistrict i that obtained via ES and EV then we will obtain a relationship between

$$\frac{1}{n_i}\sum_{j=1}^{n_i} p_{jk} \text{ with } p_{ik}$$

with some feasibility that is

(a) EV > ES if $\frac{1}{n_i}\sum_{j=1}^{n_i} p_{jk} > p_{ik}$

In inequality is $0 < p_{ik} < \frac{1}{n_i}\sum_{j=1}^{n_i} p_{jk} < 1$

(b) EV < ES if $\frac{1}{n_i}\sum_{j=1}^{n_i} p_{jk} < p_{ik}$

In inequality is $0 < \frac{1}{n_i}\sum_{j=1}^{n_i} p_{jk} < p_{ik} < 1$

(c) EV = ES if $\frac{1}{n_i}\sum_{j=1}^{n_i} p_{jk} = p_{ik}$

In inequality is $0 < p_{ik} = \frac{1}{n_i}\sum_{j=1}^{n_i} p_{jk} < 1$

It can be happen if $n_i = 1$.

In these computation, we use software WinBUGS for analysis of Bayesian Statistical Models with use Markov Chain Monte Carlo (MCMC) techniques to obtain estimates of posterior distributions.

4. WinBUGS

WinBUGS (**Windows** **B**ayesian inference **U**sing **G**ibbs **S**ampling), a Bayesian MCMC package, is distributed freely. The Gibbs sampling methodology is described by Gilks et al. (1994) and Spiegelhalter et al (1994). The WinBUGS is the result of many years of development by a team of statisticians and programmers at the Medical research Council Biostatistics Research Unit in Cambridge (http://www.mrc-bsu.cam.ac.uk/bugs/). Models are represented by a flexible language, and there is also a graphical feature, DOODLEBUGS, that allows users to specify their model in terms of a directed graph.

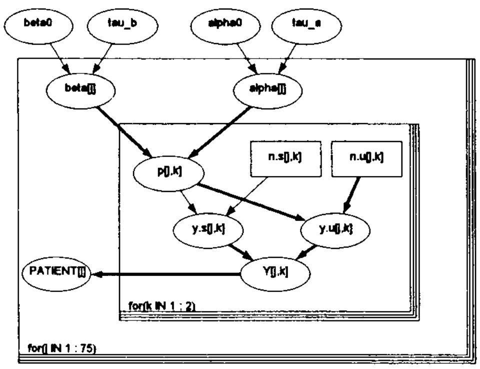

Figure 3. Graphical Model (DoodleBUGS) for the Patient Data

Five thousand MCMC simulations were conducted of which one thousand simulations were "burn in" to stabilize the MCMC output. The results are summarized in Table 4.

The WinBUGS software uses diffuse (or vague) proper priors of the form $\mu \sim N(0,\sigma_0^2)$, $\sigma_v^{-2} \sim G(a,b)$ and $\sigma_e^{-2} \sim G(a,b)$ as default priors, where σ_0^2 is choosen very large (say 5,000 or 10,000) and $a_0(>0)$ very small (say 0.001) to reflect lack of prior information on μ, σ_v^2 and σ_e^2.

The MCMC simulation for the total estimate of patient for all of subdistrict are displayed in the Figure 4.

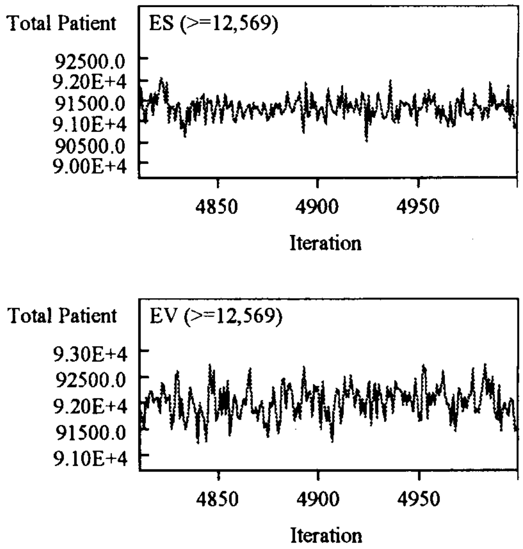

Figure 4. Total Patient Estimate for ≥ 12,569

In Figure 4 the estimation for total patient at level on the population density of Jakarta (≥ 12,569 person per km^2) for EV is appear more high than estimation with using ES. Figure 5 shows the kernel density estimation for the total patient

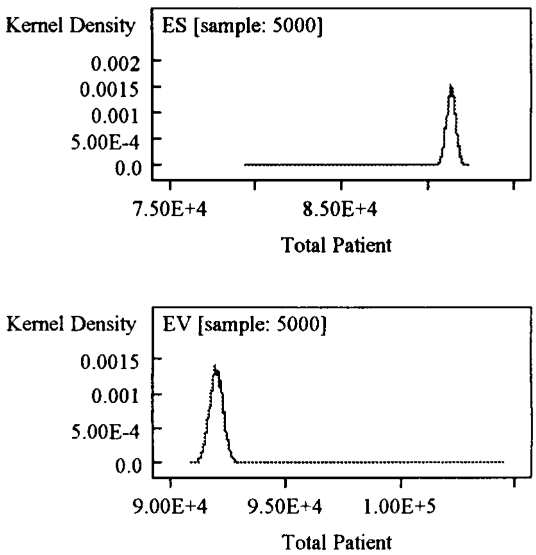

Figure 5. Density Plots for MCMC results

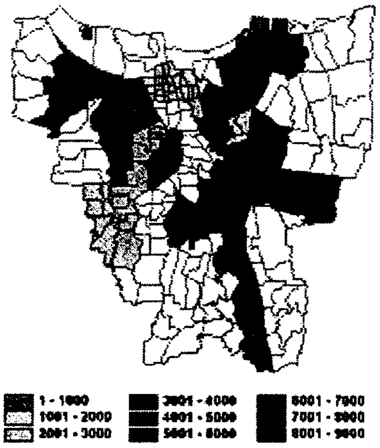

Figure 6. Jakarta Map Representing the True Patient of Diarrhea

Figure 6 and 7 show the total patient of diarrhea per subdistrict for the 25 subdistrict. It can also be seen from Figure 7 that the performance of the patient of diarrhea with use approaches ES and EV there are little under estimate from the true patient. If total patient from true data compared ES, then there is difference about 4.9% while if compared EV there is difference about 4.2%. See in Table 5.

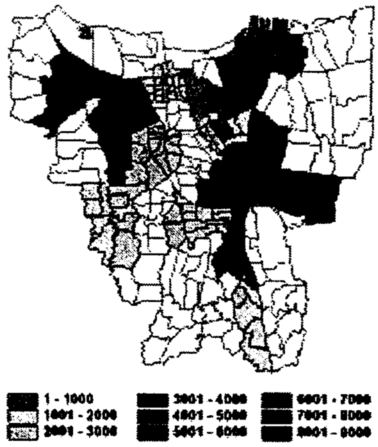

Figure 7. Jakarta MapRrepresenting the Small Area Population Density by the Posterior Total Values obtained from the ES or EV

Because the area sample map for ES and EV have the same interval so that the color for each prediction approach are also same color. In this simulation is obtained that the interval for each subdistrict have same interval both ES and EV. Figure 6 shows that there is no subdistrict which have interval 1 – 1000 but in the Figure 7 appear that the interval located at the subdistrict Sawah Besar.

Table 4. The Probability of ES $\{p_{k}\}$ and EV $\{p_{jk}\}$ for Population Density ≥ 12,569 by sudistrict

No.	Subdistrict	n	0 - 4 years							≥ 5 years						
			ES $\{p_k\}$	EV $\{p_{jk}\}$						ES $\{p_k\}$	EV $\{p_{jk}\}$					
				$j=1$	$j=2$	$j=3$	$j=4$	$j=5$	$j=6$		$j=1$	$j=2$	$j=3$	$j=4$	$j=5$	$j=6$
1	Johar Baru	2	0.1766	0.2082	0.1433					0.0186	0.0224	0.0154				
2	Tambora	6	0.0569	0.0477	0.0624	0.0570	0.0507	0.0605	0.0682	0.0072	0.0060	0.0079	0.0072	0.0064	0.0077	0.0087
3	Matraman	3	0.0927	0.1206	0.0809	0.0935				0.0170	0.0216	0.0144	0.0168			
4	Kemayoran	5	0.1214	0.1217	0.1344	0.1029	0.1312	0.1019		0.0124	0.0131	0.0143	0.0109	0.0140	0.0110	
5	Palmerah	3	0.0688	0.0974	0.0589	0.0614				0.0087	0.0126	0.0075	0.0078			
6	Taman Sari	4	0.0731	0.0626	0.0724	0.0941	0.0785			0.0093	0.0079	0.0093	0.0124	0.0103		
7	Jatinegara	4	0.1282	0.1280	0.1437	0.1063	0.1770			0.0190	0.0189	0.0213	0.0157	0.0268		
8	Tebet	4	0.1169	0.1348	0.1514	0.1009	0.0859			0.0228	0.0260	0.0288	0.0193	0.0158		
9	Koja	3	0.1056	0.1089	0.1110	0.0971				0.0138	0.0144	0.0145	0.0127			
10	Senen	2	0.1562	0.1598	0.1488					0.0205	0.0211	0.0198				
11	Kramat Jati	4	0.0678	0.0756	0.0623	0.0647	0.0727			0.0089	0.0100	0.0082	0.0085	0.0096		
12	Cempaka Putih	1	0.0801	0.0807						0.0135	0.0135					
13	Sawah Besar	1	0.0432	0.0430						0.0038	0.0039					
14	Mampang Prapatan	3	0.0881	0.0835	0.0842	0.1138				0.0131	0.0123	0.0124	0.0169			
15	Pulo Gadung	3	0.0669	0.0638	0.0790	0.0629				0.0128	0.0119	0.0148	0.0115			
16	Grogol Petamburan	3	0.0879	0.0850	0.0895	0.0876				0.0112	0.0109	0.0115	0.0112			
17	Pancoran	3	0.0807	0.0820	0.0845	0.0803				0.0146	0.0146	0.0148	0.0138			
18	Kebon Jeruk	4	0.0897	0.0749	0.0915	0.0943	0.0937			0.0114	0.0095	0.0117	0.0121	0.0120		
19	Duren Sawit	3	0.0462	0.0507	0.0447	0.0411				0.0066	0.0072	0.0063	0.0057			
20	Tanah Abang	4	0.1145	0.1316	0.1069	0.0972	0.1199			0.0116	0.0138	0.0112	0.0102	0.0131		
21	Cengkareng	2	0.0936	0.0972	0.0910					0.0119	0.0125	0.0116				
22	Pesanggrahan	2	0.0219	0.0247	0.0229					0.0044	0.0043	0.0042				
23	Tanjung Priok	2	0.0454	0.0522	0.0412					0.0075	0.0084	0.0066				
24	Kebayoran Lama	3	0.0399	0.0370	0.0410	0.0438				0.0058	0.0051	0.0058	0.0062			
25	Ciracas	1	0.0668	0.0667						0.0079	0.0079					

Table 5. The Total Patient of Diarrhea by The True, ES and EV

No.	Subdistrict	Total		
		TRUE	ES	EV
1	Johar Baru	3435	3722	3728
2	Tambora	2964	2762	2781
3	Matraman	3525	3498	3568
4	Kemayoran	4330	4507	4502
5	Palmerah	2868	2757	2871
6	Taman Sari	1646	1678	1731
7	Jatinegara	7870	7551	7889
8	Tebet	6322	6582	6571
9	Koja	6658	5643	5649
10	Senen	3464	2995	2984
11	Kramat Jati	3546	3403	3429
12	Cempaka Putih	2034	1629	1631
13	Sawah Besar	1210	821	824
14	Mampang Prapatan	3244	2721	2773
15	Pulo Gadung	5154	4624	4634
16	Grogol Petamburan	3454	3570	3570
17	Pancoran	3068	2889	2886
18	Kebon Jeruk	4936	5131	5114
19	Duren Sawit	3724	3650	3590
20	Tanah Abang	3425	2963	2981
21	Cengkareng	7860	8097	8154
22	Pesanggrahan	1205	1124	1135
23	Tanjung Priok	4437	3799	3829
24	Kebayoran Lama	2511	2349	2350
25	Ciracas	3134	2852	2853
Total		96024	91317	92027

FAS Moura and HS Migon (2002) were written about logit function. The random variables y_{ij} are assumed to be independent Bernoulli-distrubuted with parameters π_{ij}, within the ith small area as

$$\operatorname{logit}(\pi_{ij}) = \ln\frac{\pi_{ij}}{1-\pi_{ij}} = \mathbf{x}_{ij}'\boldsymbol{\beta}_i$$

where the vector of $p+1$ covariates $x_{ij}' = (1, x_{1,ij}, \ldots, x_{p.ij})$.

In this paper, the value of p_{ik} or p_{jk} in the logit function depend on the age group, x_k. In using WinBUGS , we use

$$\operatorname{logit}(p_{ik}) = \alpha_i + \beta_i x_k \begin{cases} x_k = 0 \text{ if } 0\text{ - }4 \text{ years} \\ x_k = 1 \text{ if } \geq 5 \text{ years} \end{cases}$$

if $x_k = 0$ then $\operatorname{logit}(p_{i0}) = \alpha_i$ while $x_k = 1$ then $\operatorname{logit}(p_{i1}) = \alpha_i + \beta_i$. Thus

$$\operatorname{logit}(p_{i0}) = \ln\frac{1}{1-p_{i0}} = \alpha_i$$

and

$$\operatorname{logit}(p_{i1}) = \ln\frac{1}{1-p_{i0}} = \alpha_i + \beta_i$$

Using these values of α_i and β_i, we calculate p_{i0} and p_{j1}. and we will obtain the value of p_{i0} and p_{j1} respectively

$$p_{i0} = \frac{e^{\alpha_i}}{1+e^{\alpha_i}}$$

and

$$p_{i1} = \frac{e^{\alpha_i+\beta_i}}{1+e^{\alpha_i+\beta_i}}$$

In Table 4 shows that of $p_{i0} \geq p_{i1}$. It is caused the proportion y_{ij0} / x_{ij0} is more than y_{ij1} / x_{ij1}.

To obtain estimation with using WinBUGS needed a hyperparameter that it avail as prior distribution. Even when non-informative priors were used.

For example the hyperparameter that we used to estimate total patient by approach domain in WinBUGS is

```
for (i in 1:25)
{
alpha[i] ~ dnorm(alpha0, tau.a)
beta[i] ~ dnorm(beta0, tau.b)
}

alpha0 ~ dnorm(0.0,0.001)
beta0 ~ dnorm(0.0,0.001)
tau.a ~ dgamma(0.01,0.01)
tau.b ~ dgamma(0.01,0.01)

list(
alpha0=0,
beta0=0,
tau.a=1,
tau.b=1,
alpha=c(0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0,
0, 0, 0, 0, 0, 0, 0, 0, 0),
beta=c(0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0,
0, 0, 0, 0, 0, 0, 0, 0, 0))
```

5. CONCLUSION

BUGS is a power tool bayesian analysis which can be used to do simulation tool and make graphical models with Doodle BUGS and also simple to representation of model.

The estimate of variable interest implemented using WinBUGS software to analysis Small Area Estimation (SAE) like ES and EV are give the result enough approach to true data.

In this research, we have not much simulation sample to compare between the one sample and the other sample. Nevertheless, we may decide in this sample with the computational in running 5000 iterations that EV is better than ES with refer to difference between estimator and true data. This is not general conclusion. We must over examine again.

REFERENCES

[1] Agus Setiawan and Tomoyuki Tarumi (2003), The Utilization Small Geographic Area Estimation to Build Model Based on Population Density, 2003年度 統計関連学会連合大会講演報告集, 531-532.

[2] FAS Moura and HS Migon (2002), Bayesian Spatial Models for Small Area Estimation of Proportion, *Statistical Modelling*, **2**, 183 – 201.

[3] Ghosh, M., Kannan Natarajan, Stroud, T.W.F., and Bradley P. Carlin (1998), General Linear Model for Small Area Estimation, *Journal of the American Statistical Association,*, **93**, 273-282.

[4] Gilks, W.R., Thomas, A. and Spiegelhalter, D.J (1994), A language and program for complex Bayesian Modelling. *The Statistician*, **43**, 169-178.

[5] Kaja Sostra. "Small area estimation of the number of unemployed" *http://www.ms.ut.ee /ained/SoomeMCMC/Comp10.doc.*

[6] M. Ghosh and J. N. K. Rao (1994), Small Area Estimation: An Appraisal, *Statistical Science*, **9**, 55-93.

[7] Murdock, Steve H. and David R. Ellis. 1991. *Applied Demography*. Boulder: Westview Press.

[8] Noel J. Purcell (1979), Estimation for Small Domains, *Biometrics*, **35**, 365-384.

[9] P. Elliot, J. Cuzick, D. English, and R. Stern (1992), *Geographical and Environmental Epidemiology Methods for Small-Area Studies*, Oxford University Press

[10] Platek, R., Rao, J.N.K., Sarndal, C.E., and Singh, M.P., eds. (1987), Small Area Statistics, John Wiley and Sons.

[11] Rao, J.N.K. (2001), On Measuring the Quality of Indirect Small Area Estimates in Proceedings of Statistics Canada Symposium, Ottawa, Canada.

[12] Rao, J.N.K. (2003), *Small Area Estimation*, Canada: John Wiley and Sons.

[13] Ulrike Schach. "Modeling approaches to a spatio-temporal small area estimation." *http://www.sfb475.uni-dortmund.de/berichte /tr19-01.ps*