# Bayesian inference for caries prevalence rate in 3-year-old children of Iwate Prefecture

Kazuhiko NARA

Health and Hygiene Division, Department of Health and Welfare, Iwate Prefecture

[Received : December 3, 2003 ; Accepted : February 9, 2004]

**Abstract** : Many regional health projects have been planned or implemented since Health Japan 21 was stipulated. In these projects, many numerical health indicators have been established to evaluate achievement. In small areas, however, numerical indicators are difficult to compare over years or between regions due to data fluctuations. This statistically problematic fluctuation can be seen in the dental caries prevalence rate. Recently, the Bayesian method has received attention as a way to overcome this problem. In this study we attempted to assess the utility of the Bayesian model in estimating caries prevalence in municipalities.

The results show that the Bayesian approach can stabilize the fluctuations in the rates. Moreover, analysis using the ratio of standard error suggests the power of data adjustment by the Bayesian method results from reduction in random error due to population heterogeneity, and the results are clearer in sample sizes of 200 or less. Thus this study confirmed that the Bayesian method is very useful in assessment of dental health.

**Key words** : Bayesian model, estimate, caries prevalence rate, municipalities

## Introduction

Japan's Ministry of Health and Welfare stipulated "Health Japan 21" in March 2000 and then put the "Health Promotion Law" in force in May 2003. In response to this trend, many regional health projects have been planned or implemented at the prefecture, city, town, and village levels. Health Japan 21 establishes seventy health indicators as target values in nine specific areas, which can be used for quantitative evaluation of improvement or achievement[1]. This method of evaluation is also adopted by local governments, so that various health indicators are set in their projects.

In small areas, however, numerical indicators such as mortality rate or standardized mortality ratio (SMR) are difficult to compare over several years or between regions due to irregular data fluctuations[2]. These fluctuations are caused by small population sizes resulting in unstable estimates[3,4]. To overcome this problem, the Bayesian method is applied in estimating SMR and Total Fertility Rate (TFR) in "Vital statistics by health center and municipality 1993-97"[2,5,6]. The statistically problematic fluctuation is observed in dental health indicators. That is,

Bayesian inference for caries prevalence rate in 3-year-old children of Iwate Prefecture
Kazuhiko NARA
Present address : Department of Biological Sciences, University College Chester
Parkgate Road, Chester CH1 4BJ, UK

岩手県盛岡市内丸10-1（〒020-0023），岩手県保健福祉部保健衛生課

Table 1. Average number of 3-year-old dental health examinees (children per municipality) in the administrative health districts.

| | Ave. ± S.D. (Minimum／Maximum) | | | | | |
|---|---|---|---|---|---|---|
| Year | 1997 | 1998 | 1999 | 2000 | 2001 | 2002 |
| Morioka | 389.5±758.6 (46／2,643) | 397.1±742.2 (47／2,600) | 391.8±748.7 (42／2,620) | 380.6±698.0 (41／2,449) | 392.7±741.5 (33／2,583) | 388.0±725.6 (32／2,541) |
| Iwatechubu | 279.6±373.6 (11／951) | 261.7±352.4 (19／902) | 253.0±340.2 (16／840) | 253.4±352.2 (19／902) | 258.1±357.3 (17／867) | 264.6±361.8 (23／920) |
| Tanko | 230.8±212.1 (43／631) | 234.3±207.3 (52／626) | 208.8±204.5 (48／601) | 201.3±174.9 (39／527) | 215.3±205.5 (41／609) | 218.0±217.4 (38／634) |
| Ryouban | 146.8±165.8 (49／580) | 146.4±175.4 (33／605) | 134.6±167.1 (48／575) | 141.9±166.6 (28／575) | 139.1±171.3 (36／587) | 137.4±155.3 (39／542) |
| Kesen | 245.3±197.2 (53／447) | 239.3±203.1 (46／451) | 222.7±186.4 (48／419) | 217.3±191.6 (29／412) | 203.7±174.9 (36／385) | 209.0±175.0 (42／391) |
| Kamaishi | 184.3±112.0 (39／303) | 208.5±159.4 (39／412) | 194.5±142.9 (30／375) | 173.3±120.0 (31／312) | 170.3±111.8 (42／307) | 159.8±107.6 (33／286) |
| Miyako | 147.6±199.0 (25／573) | 139.6±178.5 (22／515) | 133.1±168.4 (26／489) | 131.7±180.8 (23／511) | 128.4±163.3 (23／474) | 124.0±167.4 (11／477) |
| Kuji | 121.2±143.1 (25／402) | 115.3±140.0 (24／385) | 111.0±139.2 (15／384) | 115.5±140.4 (25／390) | 104.0±136.4 (17／370) | 108.3±136.4 (19／378) |
| Ninohe | 119.0±85.4 (32／255) | 118.6±85.4 (50／260) | 107.8±73.9 (35／227) | 110.0±97.5 (33／277) | 98.4±79.9 (30／234) | 100.0±75.5 (33／224) |

indicators such as caries prevalence rate, which are calculated for a single age group, tend to show larger variances in smaller areas. This paper is aimed to assess the utility of the Bayesian approach in estimating caries prevalence among 3-year-old children in the municipalities of Iwate.

## Subjects and Methods

1. Data Sets

Data for 3-year-old dental health examinations from 1997 to 2002 have been used to obtain caries prevalence rate estimates in each of the 58 municipalities of Iwate Prefecture (Table. 1).

2. Bayesian Model

In order to obtain the Bayesian estimates, the Beta-Binomial model[7,8], which was used to calculate TFR in "Vital statistics by health center and municipality 1993-97", was applied[5,6].

Processes of calculation are represented as follows.

(a)Likelihood function of caries prevalence rate.

Let N denote the population of a group and D the corresponding observed number of persons having experienced caries in that group. Then D has a Binomial distribution of Bin (N, q).

Now we can define the caries prevalence rate, q, in the group :

$q = D / N$

Moreover, the probability density function for D is described by the following equation.

$$f\ (D \mid N, q) = {}_{N}C_{D}\, q^{D}\ (1 - q)^{N-D} \qquad (1)$$

with ${}_{N}C_{D} = N!\ (N-D)!\ / D!$

Once the sample data such as N and D have been observed, they are fixed as realized values. With the data sets given, if we treat "q" as an unknown parameter, then the Eq (1) is considered as the likelihood function of q.

$$L\ (q \mid N, D) = {}_{N}C_{D}\, q^{D}\ (1 - q)^{N-D} \qquad (2)$$

(b)Prior distribution

The number of persons who have experienced caries, D, has a Binomial distribution. Hence a Beta distribution is selected as the prior distribution for caries prevalence rate, q. This is because Beta is the conjugate prior for the Binomial distribution which can be expressed as :

$$P(q) = \frac{1}{B(\alpha, \beta)} q^{\alpha-1} (1-q)^{\beta-1} \quad (3)$$

where B ($\alpha$, $\beta$) is the beta function, and expectation and variance are given by :

$$E(q) = \frac{\alpha}{\alpha+\beta} \quad (4)$$

$$V(q) = \frac{\alpha\beta}{(\alpha+\beta)^2(\alpha+\beta+1)} \quad (5)$$

(c)Posterior distribution

According to Bayes theorem[9], we can determine the posterior distribution from the likelihood function, (2), and the prior distribution, (3). Thus, its probability density function is given by :

$$P(q \mid N, D) = \frac{{}_N C_D q^D (1-q)^{N-D} P(q)}{\int_0^1 {}_N C_D q^D (1-q)^{N-D} P(q) dq}$$

$$= \frac{1}{B(\alpha+D, \beta+N-D)} q^{\alpha+D-1}(1-q)^{\beta+N-D-1} \quad (6)$$

which is beta distributed like the prior and has expectation and variance as below.

$$E(q \mid N, D) = \frac{\alpha+D}{\alpha+\beta+N} \quad (7)$$

$$V(q \mid N, D) = \frac{(\alpha+\beta)(\beta+N-B)}{(\alpha+\beta+N)^2(\alpha+\beta+N+1)} \quad (8)$$

(d)Estimation for hyperparameters of the Beta distribution

Following the method of "Vital statistics by health center and municipality 1993-97," it is assumed that the prior expectation and the prior variance for caries prevalence rate in each municipality are equivalent to those in the administrative health district which includes it. In other words, the expectation and the variance of the district are used as moment estimators for its municipalities[2,5,6].

First, let $n_{it}$ and $d_{it}$ ($i=1,2,3,\ldots$ ; $t=1997,\ldots,2002$) denote the number of examinees who underwent 3-year-old dental health examinations and the corresponding observed number of those having experienced caries for each municipality within the health district M during the years 1997 to 2002. Then, $q_{it}=d_{it}/n_{it}$ denotes the raw annual caries prevalence rate among 3-year-olds in the i-th municipality.

Secondly, let Qm and Vm be the expectation and the variance for the overall caries prevalence rate in the district M respectively, and they are given by :

$$Qm = \Sigma \frac{d_{it}}{\Sigma n_{it}} \quad (9)$$

$$Vm = \Sigma \frac{n_{it}}{\Sigma n_{it}} (q_{it} - Qm)^2 \quad (10)$$

As mentioned above, E(q) and V(q) are equal to Qm and Vm respectively, so using Eqs (4) and (5), we obtain the following

$$E(q) = \frac{\alpha m}{\alpha m + \beta m} = Qm \quad (11)$$

$$V(q) = \frac{\alpha m \beta m}{(\alpha m + \beta m)^2 (\alpha m + \beta m + 1)} = Vm \quad (12)$$

Eqs (11) and (12) may be set up as follows to solve for parameters $\alpha$m and $\beta$m :

$$\alpha m = Qm\left\{\frac{Qm(1-Qm)}{Vm} - 1\right\} \quad (13)$$

$$\beta m = (1-Qm)\left\{\frac{Qm(1-Qm)}{Vm} - 1\right\} \quad (14)$$

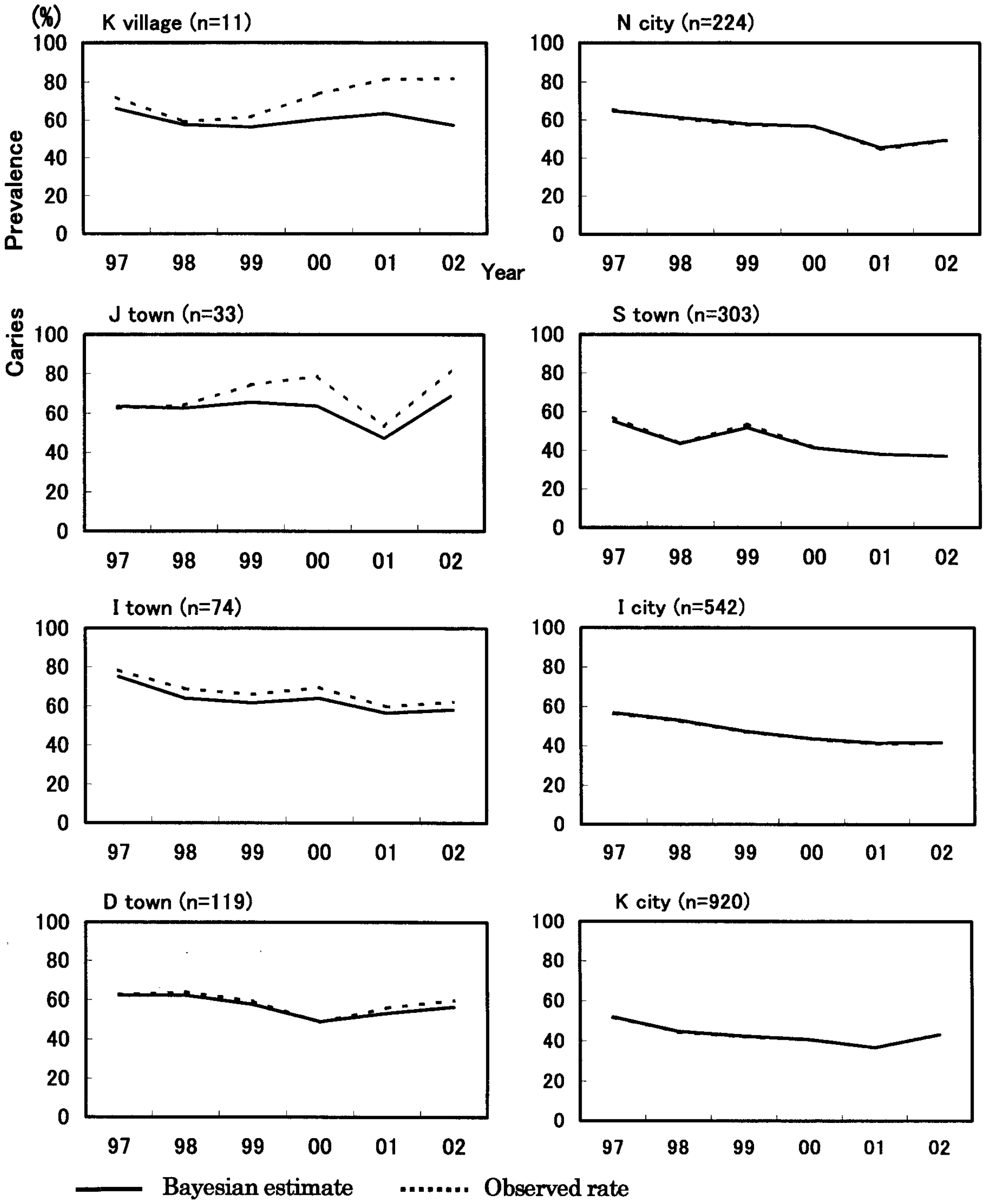


Fig. 1. Comparison of yearly trends between the Bayesian estimates and observed rates (raw observed caries prevalence rate).
n denotes sample size.
Here, sample size means the number of examinees who underwent 3-year-old dental health examination in each municipality in 2002.

Finally, by substituting $\alpha m$ and $\beta m$ for Eq (7) and Eq (8) we can estimate the unknown posterior expectation (posterior variance) for caries prevalence rate in i-th municipality. That is :

$$E\ (q \mid n_{it}, d_{it}) = \frac{\alpha m + d_{it}}{\alpha m + \beta m + n_{it}} \tag{15}$$

$$V\ (q \mid n_{it}, d_{it}) = \frac{(\alpha m + \beta m)(\beta m + n_{it} - d_{it})}{(\alpha m + \beta m + n_{it})^2 (\alpha m + \beta m + n_{it} + 1)} \tag{16}$$

## Results

1. Comparison of yearly trends

Fig. 1 shows illustrations for several sample sizes to compare the yearly trends of the Bayesian estimates and observed caries

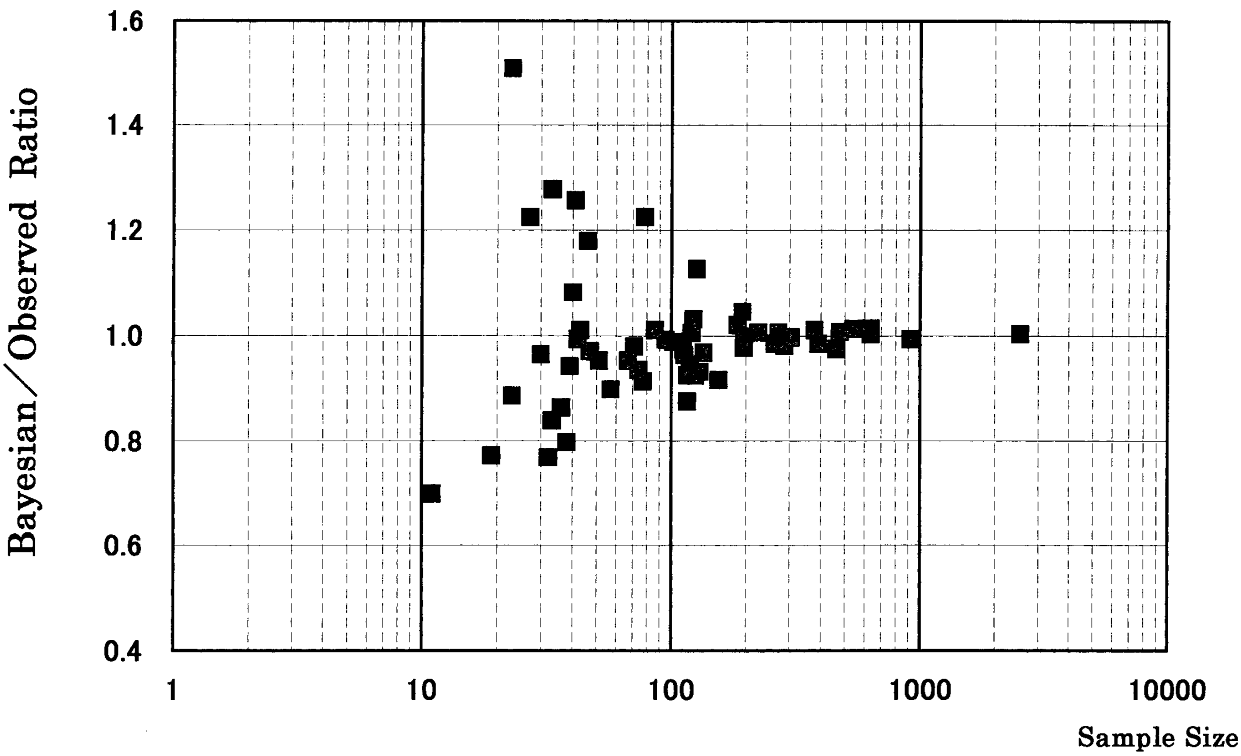


Fig. 2. The ratio of the Bayesian estimates and observed rates.
The ratio is plotted against the sample sizes in logarithmic scale.
The meaning of sample size is the same as in Figure 1.

prevalence rates (hereafter simply referred to as observed rates). Throughout this article, the term "sample size" is used to denote the raw annual number of examinees who underwent 3-year-old dental health examinations in each municipality. In other words, the number of examinees is presumed to be an observed value, and the number of those having experienced caries as realizations from independent Binomial random variables. As can be seen in Fig. 1, observed rates fluctuate larger in smaller sample sizes, whereas the Bayesian estimates are stable in any sample size.

2. Relation between adjustment and sample sizes

Fig. 2 plots the ratio of the Bayesian estimates to observed rates against sample sizes. In sample sizes larger than 200, the ratio is approaching 1, i.e. the Bayesian estimates and observed rates are substantially equivalent. In sample sizes of 200 or smaller, however, the ratio fluctuates above and below 1.

3. Comparison of the ratio of standard error

The ratio of standard error (SE-ratio) is a percentage of standard error (SE) to estimate. Since SE corresponds to standard deviation (SD) for estimates, the SE-ratio can be used to indicate the precision of estimates[2]. Fig. 3 compares the SE-ratio of Bayesian estimates and observed rates.

The SE-ratio for observed rates tends to become higher as sample sizes become smaller; the highest value is 35.1%. Although a similar tendency is observed in the Bayesian estimates, the highest value of 16.0% is much lower than that in observed rates. On the other hand, the difference in the SE-ratio between observed rates and

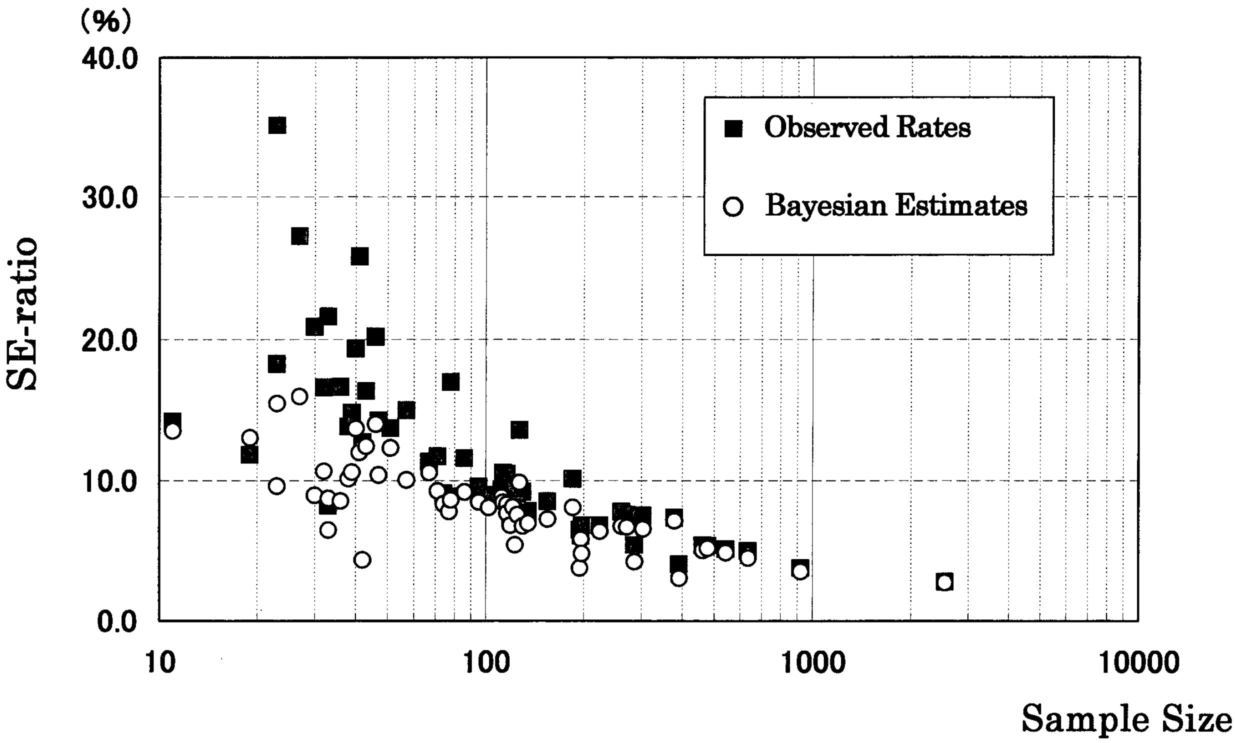


Fig.3. The ratio of standard error (SE-ratio) of the Bayesian estimates and observed rates by sample size.
The ratio is plotted against the sample sizes in logarithmic scale.
The meaning of sample size is the same as in Figure 1.

Bayesian estimates becomes smaller in sample sizes over 100, and becomes negligible in sample sizes over 200. These results confirm that the Bayesian method is effective enough to be used to adjust large variances in processed statistics like caries prevalence rate, which is obtained from extremely small populations.

Next, the proportion of the SE-ratio for the Bayesian estimates to that for observed rates is plotted against sample sizes, and then compared between the administrative health districts (Fig. 4). As mentioned above, there is a tendency for the SE-ratio of Bayesian estimates to be smaller than that for observed rates in the municipalities with small sample sizes (Fig. 3). For that reason, as the sample sizes decrease, the proportion of the SE-ratio falls further from 1, while when the sample sizes increase, it approaches 1 (Fig. 4). However, in the municipalities with the smallest sample sizes in the districts of Miyako, Kuji, and Ninohe, the proportion of SE-ratio is close to or over 1 (indicated in Fig. 4 with an arrow). This indicates that the adjustment of variances by the Bayesian approach could be deemed barely effective or ineffective in these areas.

## Discussion

This paper attempts to assess the usefulness of the Beta-Binomial Bayesian model in estimating caries prevalence rate among 3-year-old children by municipality.

First, the yearly trends of the Bayesian estimates were compared with those of observed rates (Fig. 1). The result shows that observed rates fluctuate more in smaller sample sizes, whereas the Bayesian

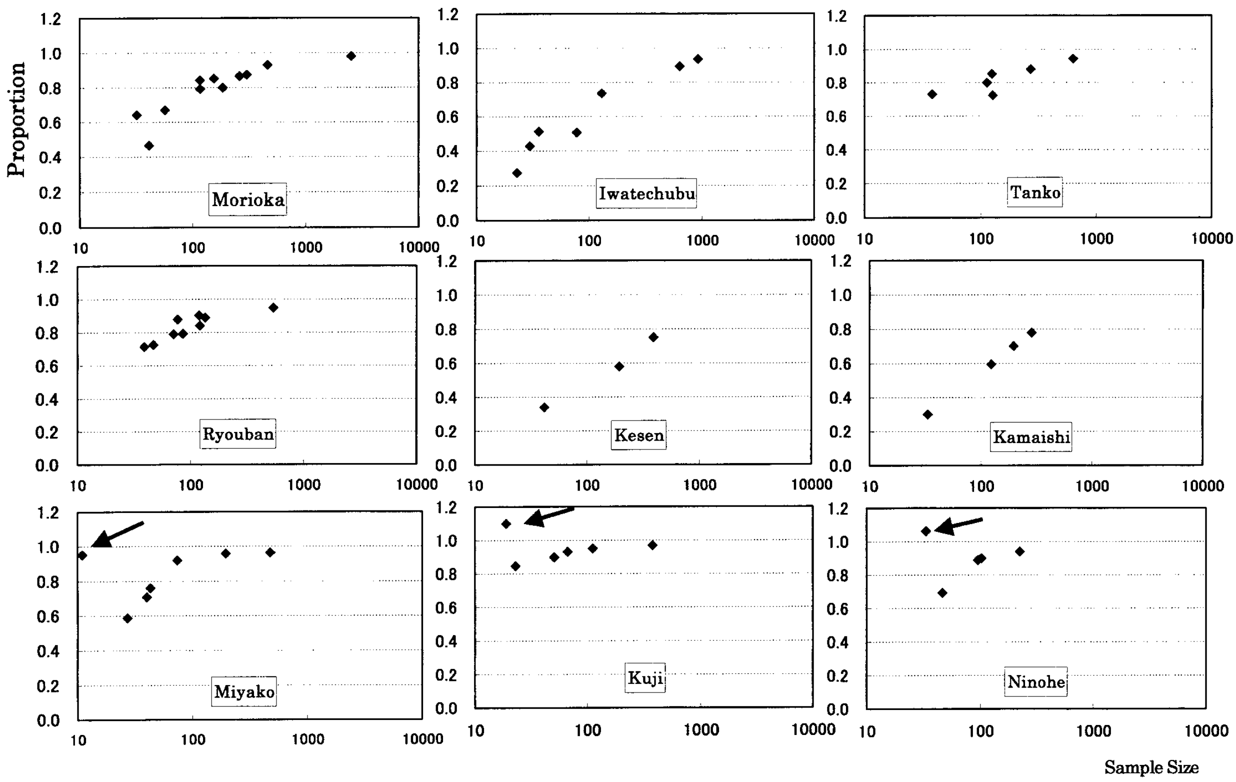


Fig. 4. The proportion of the rate of standard error of the Bayesian estimates and observed rates by administrative health district.
The ratio is plotted against the sample sizes in logarithmic scale.
The meaning of sample size is the same as in Figure 1.

estimates are stable in any sample size. This suggests that the Bayesian method can be used also for smoothing data series like caries prevalence rates.

As can be seen in Fig. 1, stabilization of caries prevalence rates by the Bayesian method is much clearer in smaller sample sizes. The ratio of the Bayesian estimates to observed rates is plotted against the sample sizes to investigate the relation between adjustment and sample sizes. The results show that adjustment is effective in sample sizes of 200 or smaller (Fig. 2).

The posterior expectation of the Bayesian estimates is computed using Eq (15), which is denoted in Subjects and Methods. To further examine the properties of the estimator, let us rewrite the Eq as :

$$\frac{n_{it}}{\alpha m+\beta m+n_{it}}\,\frac{d_{it}}{n_{it}}+\frac{\alpha m+\beta m}{\alpha m+\beta m+n_{it}}\,\frac{\alpha m}{\alpha m+\beta m} \quad (17)$$

From this, we can see that the Bayesian estimate is adjusted closer to "$d_{it}/n_{it}$" (observed rate) when sample size, $n_{it}$ is large enough, and that it is adjusted closer to "$\alpha m/(\alpha m+\beta m)$" when sample size is small enough. In this case, it appears that the point at $n_{it} \approx 200$ is the critical point.

In order to obtain stable health indicators, it is crucial to obtain less biased estimates. It is said that the Bayesian estimator is a slightly better estimator than the maximum likelihood estimator (MLE), because the mean square error (MSE) of the former is much smaller than that of the latter [10]. A simulation study carried out by Nielsen and Lewy [11] demonstrates that the MSE of the Bayesian estimator can be less than 5% of that of the MLE in some cases. In our study, though the MLE could not be obtained the precision of the Bayesian estimates was evaluated in comparison with that of observed rates, using the ratio of standard error (SE-ratio). The relation to the sample sizes was also compared. It is evident from the above studies that the SE-ratio for the Bayesian estimates is smaller than that for observed rates in the small sample sizes, and that the difference between them becomes larger as sample sizes become smaller (Fig. 3). The largest difference of SE-ratio between the Bayesian estimate and observed rate (35.1% and 9.6% respectively) in the same municipality is 25.5%. These observations reaffirm that the nature of data adjustment by the Bayesian method comes as a result of reduction in the random error arising from population fluctuation or heterogeneity[3,4].

In addition, the proportion of the SE-ratio for the Bayesian estimates to that for observed rates is plotted against sample sizes and compared between administrative health districts (Fig. 4). Because of the abovementioned adjustment by the Bayesian method (Fig. 3), the proportion of SE-ratio approaches 1 as the sample size increases, and decreases further from 1 as the sample size decreases. But the proportions of the SE-ratio of the municipalities with smallest sample sizes were 0.95, 1.10, 1.06 in the districts of Miyako, Kuji, and Ninohe respectively. This result suggests that the adjustment of biases by the Bayesian approach could be deemed barely effective or ineffective in these cases. However, municipalities with nearly the same sample size in other districts did not exhibit this problem, suggesting that it is not caused by sample size. In order to obtain Bayesian estimates for caries prevalence rate in municipalities, the expectation and the variance of the district are adopted as the moment estimators for its municipalities[2,5,6]. This is based upon the presumption that regional rates are drawn from superpopulation rates[4]. Additionally, it appears that administrative health districts are most appropriate as superpopulations. Thus, these estimates assume that municipalities in the same district will exhibit similar caries prevalence rates. (This assumption follows the methodology in " Vital statistics by health center and municipality 1993-97"[2,5,6].) In Iwate, however, the administrative health districts have some issues, including:

- They have heterogeneous populations (Table 1).
- They do not correspond fully to wider living or cultural areas.

Considering these issues, there is a possibility that the above three municipalities might have very different qualities than others in the same districts.

Caries prevalence rate is an indicator calculated for a single age group, such as 18 months or 3 years old. Furthermore, when mortality rate and SMR are calculated, the data are adjusted to obtain higher stability by calculating averages or summation over approximately five years[6]. In calculating caries prevalence rates, however, this method of adjustment is almost meaningless because the rate is calculated for a single age group, of which the smallest municipal sample size is only 11 individuals (Table 1). For that reason, estimates were computed from single-year data. In spite of such disadvantageous conditions, the Bayesian approach could successfully smooth even the data from extremely small sample-sized areas; it has been confirmed that the method is particularly useful in municipalities in which the sample size is 200 or less. Because Bayesian smoothing does not require as much data as other approaches[8,12], it might be much more useful in small towns and villages than other smoothing techniques. It is expected that the Bayesian method could become a powerful tool in dental health. On the other hand, as issues in the establishment of superpopulations to estimate the prior expectation and variance were recognized, further studies appear necessary.

## Conclusion

The usefulness of the Bayesian method in estimating caries prevalence rate of 3-year-old children in municipalities was assessed.

1. Year-to-year changes of the Bayesian estimates are stable in any sample sizes, whereas observed rates fluctuate more in smaller sample sizes.
2. Plotting the ratio of the Bayesian estimates to observed rates against sample sizes shows adjustment by the Bayesian method works in sample sizes of 200 or smaller.
3. The ratio of standard error (SE-ratio) for the Bayesian estimates is smaller than that for observed rates in the small sample sizes, and the difference becomes larger as sample sizes decrease, with a maximum difference of 25.5%. This reaffirms that the Bayesian method reduces random error due to fluctuation in or heterogeneity of the surveyed population.
4. Moreover, the proportion of the SE-ratio for the Bayesian estimates to that for observed rates shows that the Bayesian approach is largely ineffective in the smallest sample-sized municipalities of the Miyako, Kuji, and Ninohe. The particular reason for this remains unclear, but it might be due to issues with the method for estimating the prior expectation (the prior variance).
5. From the above results, the Bayesian method can be used for estimating caries prevalence in municipalities. However, it is necessary to further consider the process of estimating the prior expectation (variance).

*Acknowledgement*

The author is grateful to Prof. M. Yonemitsu, Head of the Department of Preventive Dentistry, Iwate Medical University, for critical reading of the manuscript.

The author also thanks Dr. T. Sato, the Standing Director of the Iwate Dental Association, for his useful instruction.

## References

1) Health Japan 21 Plan Study Committee, Health Japan 21 Plan Development Committee : Health Japan 21 report , Japan Health Promotion and Fitness Foundation, Tokyo, 2000. (in Japanese)
2) Hirako, T., Saeki, N., Tanaka, T. : Jinko dotai shichoson-betsu tokei e no beizu-tokei no oyo ni tsuite 〈1〉 Hyoujunka shibo hi e no oyo (About the application of Bayesian method to vital statistics by health center and municipality. (1) An application to SMR.). *J. Health and Welfare Statistics* 46 : 3 -11, 1999. (in Japanese)
3) Hill, E. G., Allen, A. S., and Waller, L. A. : A comparison of focused score tests and Bayesian hierarchical models for detecting spatial disease clustering. *J. Natl. Inst. Public Health* 48 : 102-112, 1999.
4) Waller, L. A. : Hierarchical models for disease mapping. In : The encyclopedia of environmetrics, Vol.2, eds. by El-Shaarawi, A. H., and Piegorsch, W. W., John Wiley and Sons Ltd., Chichester, pp1004-1007, 2002.
5) Japan, Ministry of Health and Welfare : Vital

statistics by health center and municipality 1993-97 Special report on vital statistics. Statistics and Information Department, Ministry of Health and Welfare, Tokyo, 1999. (in Japanese)

6) Saeki, N., Hirako, T., and Tanaka, T. : Jinko dotai shichoson-betsu tokei e no beizu-tokei no oyo ni tsuite 〈2〉 Gokei tokushu shusshou ritsu e no oyo (About the application of Bayesian method to vital statistics by health center and municipality. (2) An application to TFR.). *J. Health and Welfare Statistics* 46 : 3-10, 1999. (in Japanese)

7) Lee, J. C., and Sabavala, D. J. : Bayesian estimation and prediction for the beta-binomial model. *Journal of Business and Economic Statistics* 5 : 357-367, 1987.

8) Lee, J. C., and Lio, Y. L. : A note on Bayesian estimation and prediction for the beta-binomial model. *InterStat* (Online journal : http://interstat.stat.vt.edu/InterStat/index/Mar97.html) : 1997.

9) Petrie, A., Bulman, J. S., and Osborn, J. F. : Further statistics in dentistry Part 9 : Bayesian statistics. *Br. Dent. J.* 194 : 129-134, 2003.

10) Efron, B., and Morris, C. : Stein's estimation rule and its competitors- an empirical Bayes approach. *J. Amer. Statist. Assoc.* 68 : 117-130, 1973.

11) Nielsen, A, and Lewy, P. : Comparison of the frequentist properties of Bayes and maximum likelihood estimates in an age-structured fish stock assessment model. *Can. J. Fish. Aquat. Sci.* 59 : 136-143, 2002.

12) Petrie, A., Bulman, J. S., and Osborn, J. F. : Further statistics in dentistry Part 6 : Multiple linear regression. *Br. Dent. J.* 193 : 675-682, 2002.

# 岩手県市町村別３歳児う歯有病者率へのベイズ推計の応用

奈良 一彦
岩手県保健福祉部保健衛生課

（受付：2003年12月３日）
（受理：2004年２月９日）

**抄録**：健康日本21が示されて以来、多くの地域において健康づくり計画の策定が進んでいる。これらの計画には，達成度評価のために多種の数値指標が設定されている。しかし，小地域においては，数値指標が不規則に変動するため，経年的動向を見たり，地域間比較が困難である。う歯有病者率も同様な問題を抱えている。近年，ベイズ推定法がその解決策として注目されてきている。本研究では，市町村のう歯有病者率推定にベイズ・モデルを応用し，その有効性を検証した。
その結果，ベイズ推定法がう歯有病者率においても数値の変動を抑制できることが示された。また，標準誤差率による分析から，その誤差調整作用が人口の不均等によって生じる偶然誤差の抑制によるものであり，200人以下の標本サイズで顕著であることが分かった。以上から，ベイズ推定法が歯科保健分野においても有効な手法であることが確認された。

キーワード：ベイズ・モデル，推定値，う歯有病者率，市町村