



ORIGINAL ARTICLE

Spatial and Temporal Distribution of *Plasmodium vivax* Malaria in Korea Estimated with a Hierarchical Generalized Linear Model

Maengseok Noh^a, Youngjo Lee^b, Seungyoung Oh^b, Chaeshin Chu^c,
Jin Gwack^c, Seung-Ki Youn^c, Shin Hyeong Cho^d, Won Ja Lee^d, Sun Huh^{e,*}

^aDepartment of Statistics, Pukyong National University, Busan, Korea.

^bDepartment of Statistics, Seoul National University, Seoul, Korea.

^cDivision of Epidemic Intelligence Service, Korea Centers for Disease Control and Prevention, Osong, Korea.

^dDivision of Malaria and Parasitic Diseases, Korea National Institute of Health, Osong, Korea.

^eDepartment of Parasitology and Institute of Medical Education, College of Medicine, Hallym University, Chuncheon, Korea.

Received: October 19, 2012

Revised: November 5, 2012

Accepted: November 6, 2012

KEYWORDS:

epidemiology,
incidence,
Korea,
linear models,
malaria,
Plasmodium vivax

Abstract

Objectives: The spatial and temporal correlations were estimated to determine *Plasmodium vivax* malarial transmission pattern in Korea from 2001–2011 with the hierarchical generalized linear model.

Methods: Malaria cases reported to the Korea Centers for Disease Control and Prevention from 2001 to 2011 were analyzed with descriptive statistics and the incidence was estimated according to age, sex, and year by the hierarchical generalized linear model. Spatial and temporal correlation was estimated and the best model was selected from nine models. Results were presented as diseases map according to age and sex.

Results: The incidence according to age was highest in the 20–25-year-old group (244.52 infections/100,000). Mean ages of infected males and females were 31.0 years and 45.3 years with incidences 7.8 infections/100,000 and 7.1 infections/100,000 after estimation. The mean month for infection was mid-July with incidence 10.4 infections/100,000. The best-fit model showed that there was a spatial and temporal correlation in the malarial transmission. Incidence was very low or negligible in areas distant from the demilitarized zone between Republic of Korea and Democratic People's Republic of Korea (North Korea) if the 20–29-year-old male group was omitted in the diseases map.

Conclusion: Malarial transmission in a region in Korea was influenced by the incidence in adjacent regions in recent years. Since malaria in Korea mainly originates from mosquitoes from North Korea, there will be continuous decrease if there is no further outbreak in North Korea.

*Corresponding author.

E-mail: shuh@hallym.ac.kr

This is an Open Access article distributed under the terms of the Creative Commons Attribution Non-Commercial License (<http://creativecommons.org/licenses/by-nc/3.0>) which permits unrestricted non-commercial use, distribution, and reproduction in any medium, provided the original work is properly cited.

1. Introduction

Malaria caused by infection with *Plasmodium vivax*, has been a public health concern in Korea (= Republic of Korea) since 1993. At that time, it was suggested that the prevalence would decrease soon since mosquito spread from North Korea was the main cause of re-emerging infection [1]. However, since then, the incidence has not decreased rapidly and there was a fluctuation of incidence even after 10 years due to the fluctuation of the malaria incidence in North Korea [2]. *P. vivax* is the only species that causes malaria in the Korean Peninsula. To estimate the annual pattern of malaria incidence by age and sex in Korea, the hierarchical generalized linear model (HGLM) was adapted. In the model, spatial and temporal correlations were also considered for the best estimation. Results of estimated incidence were represented as a diseases map.

To draw a diseases map through the estimation of incidence from small regions, empirical Bayesian and hierarchical Bayesian estimation methods have been used [3,4]. An alternative method is HGLM [5]. The Bayesian approach requires the assumption of ‘prior to’ parameters, while HGLM presents hierarchical likelihood so that there appears no sensitivity problem of parameter estimation that may occur in the case of wrong prior assumption. Also, Bayesian approaches should penetrate complicated calculation process such as Gibbs sampling procedure for parameter estimation. However, HGLM does not require those complicated procedure but can be easily estimated with statistical packages.

The following are the main questions in this analysis: Is malaria incidence of one region correlated with that of adjacent regions? Is malaria incidence of 1 year correlated with that of recent previous year? What are the characteristics of malaria incidence according to age and sex? Which HGLM model can best present the spatial and temporal correlation of incidence of malaria?

The results may help the prediction of malaria transmission according to time and geographic location, and may provide strategies for malaria prevention in Korea.

2. Materials and Methods

2.1. Data collection and management

Cases of malaria infection have been reported to the Center for Diseases Control and Prevention (CDC) by medical clinics and hospitals as well as health centers according to the “Law on the prevention and control of infectious diseases” in Korea. All cases reported are *P. vivax* malaria. Data from 2001–2011 were collected by the CDC. Raw data include age, sex, address in city or county, and month and year of occurrence.

Data were manipulated for the analysis as follows: Variables in the data of infected persons were mean age, sex, address, mean month of occurrence, and year of occurrence. Address was classified into 232 cities or counties. Data are grouped according to sex, address, and year. The age of each group is the mean age of infected persons and the month of infection was a mean of infected months. Therefore the number of data inputted for analysis was 5104 (2 sexes \times 232 cities \times 11 years). Incidence in each city or county was standardized based on the number of infection/100,000 individuals. Populations in each city or county in a year were obtained from the webs site of Statistics Korea, available from: <http://kostat.go.kr>.

All the results in the paper can be obtained by using `dhglmfit()` function in the R package `dhglm` at the Comprehensive R Archive Network (CRAN) [6].

2.2. Statistical model

2.2.1. Equation

For response variable, we analyzed observed numbers y_{itk} of malaria cases (i : city or county (1, ..., 232), t : year (2001, ..., 2011), k : sex (1: male, 2: female)). The model used in this paper, the following Poisson HGLM with the log link was considered:

$$y_{itk} | u_i, v_{it} \sim \text{Poisson}(\mu_{itk})$$

$$\eta_{itk} = \log(\mu_{itk}) = \log\left(\frac{N_{itk}}{100,000}\right) + \beta_0 + \beta_1 I(k=2) + b(\text{age}_{ik}) + c(\text{month}_i) + u_i + v_{it}$$

η_{itk} = linear predictor; μ_{itk} = conditional expected numbers of cases; N_{itk} = population of group (city or county) acquainted from Statistics Korea according to sex and year; β_0 = intercept; β_1 = Slope; $I(\cdot)$ = indicator function; age_{ik} = mean age of infection occurrence between 2001 and 2011 out of i_{th} city or county; month_i = mean month of infection occurrence between 2001 and 2011 out of i^{th} city or county; $b(\text{age}_{ik})$ = smoothing spline on the mean age of infection occurrence; $c(\text{month}_i)$ = smoothing spline on the mean month of infection occurrence; u_i = random temporal effects; v_{it} = random spatial effects.

For each temporal and spatial effect, three models were considered to give a total of nine models. Among nine models, the model with least value of conditional Akaike Information Criterion (AIC) was selected as the best fitting model to malaria distribution in Korea from 2001 to 2011.

2.2.2. Spatial models for u_i

Spatial Model 1 is the independent model, where the incidence of a region is assumed to be independent of adjacent regions.

$$u_i \sim i.i.d. N(0, \lambda_1)$$

Spatial Models 2 and 3 assume spatial effects where the incidence of a region is affected through adjacent regions.

Spatial Model 2 is the intrinsic autoregressive model of Besag and Higdon [7], $u_i \sim IAR$ with kernel $\sum_{i \sim j} (u_i - u_j)^2 / \lambda_1$.

Spatial Model 3 is the Markov random-field model [8], $u_i \sim MRF$ with $\text{var}(u)^{-1} = (I - \rho N) / \lambda_1$ where ρ is the spatially correlation and N is the incidence matrix for neighbors of city or county.

2.2.3. Temporal model for v_{it}

Temporal Model 1 is the independent model; the incidence of each year does not depend upon the previous years,

$$v_{it} \sim i.i.d. N(0, \lambda_2)$$

Temporal Models 2 and 3 assume temporal effects, where the incidence of each year depends upon the previous year.

Temporal Model 2 is the random walk model, where $v_{it} = v_t \sim RW$ with kernel $\sum_t (v_t - v_{t-1})^2 / \lambda_2$

Temporal Model 3 is the first order autoregressive model, $v_{it} \sim AR(1)$ with kernel $\sum_i \sum_t (v_t - \rho v_{t-1})^2 / \lambda_2$

All script descriptions for each model are in the supplement.

2.3. Diseases mapping

The map of Korea was obtained from the National Geographic Information Institute. The incidence of each region (city or county) and year was inputted and the drawing is presented according to the incidence level in color. MapWizard for Excel was used for drawing of diseases map [9].

3. Results

3.1. The best-fit model

A combination model of Spatial Model 3 and Temporal Model 3 showed the least conditional AIC value. Results of the fitted model are summarized in Table 1.

3.2. Temporal distribution of malaria

Change of malaria incidence in Korea according to year and sex is shown in Figure 1. The total number of malaria cases in Korea from 2001 to 2011 was 17,044, with 14,269 males and 2,775 females. The incidence decreased between 2001 and 2004, but it increased again by 2007. There was a decrease to 2008 and increase to 2010. Monthly incidence shows that July and August were the peak months with incidence of

Table 1. Result of the value of estimate of variables, standard error and t -value of combined model of spatial model 3 (Markov random field model) and temporal model 3 (the first order autoregressive random walk model) for the estimation of incidence of *Plasmodium vivax* malaria reported to the Center for Diseases Control and Prevention, Republic of Korea from 2001–2011 by the hierarchical generalized linear model

	Estimate	Standard Error	t
β_0	3.953	0.055	71.34
β_1	-1.617	0.020	-77.94
$\log(\lambda_1)$	-1.328	0.133	-9.98
$\log(\lambda_2)$	-3.723	0.039	-95.46
ρ	0.7267	0.0038	191.24

8.15 infections/100,000 and 8.53 infections/100,000, respectively (Figure 2). The mean month of malaria incidence was mid-July with 10.2 infections/100,000 after estimation by the best-fit HGLM model (Figure 3).

3.3. Spatial distribution of malaria

The diseases map that shows the annual malaria incidence in each city according to year was presented in Figure 4. There was the highest incidence in the area near the demilitarized zone (DMZ) in the northern part of Korea.

3.4. Incidence according to sex and age

The incidence according to age was highest in 20–25-year-old group, at 244.52/100,000 (Figure 5).



Figure 1. Annual incidence of *Plasmodium vivax* malaria reported to the Center for Diseases Control and Prevention, Republic of Korea from 2001 to 2011 in addition to incidence according to sex.

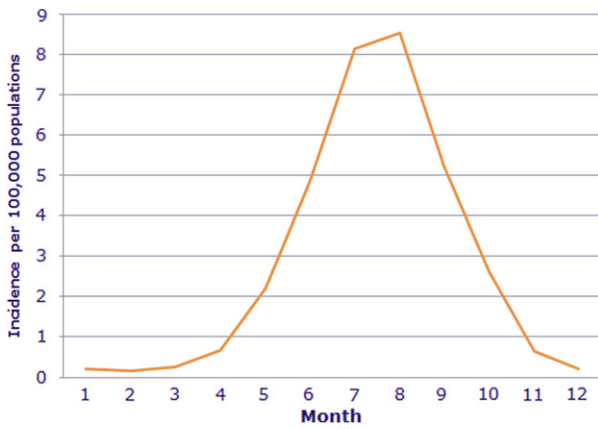


Figure 2. Monthly incidence of *Plasmodium vivax* malaria reported to the Center for Diseases Control and Prevention, Republic of Korea from 2001 to 2011. July and August were peak months with incidence of 8.15 infections/100,000 and 8.53 infections/100,000.

The smoothing spline chart of incidence distribution according to age after estimation is shown in Figure 6. The mean age of incidence of malaria in males was 31.0 infections/100,000 and 45.3 infections/100,000 in females after estimation by the best-fit HGLM model.

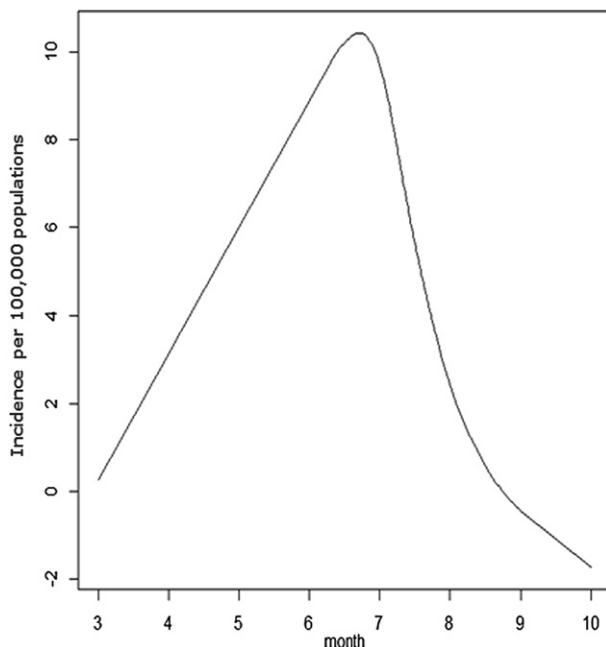


Figure 3. Mean month of incidence of *Plasmodium vivax* malaria from smoothing splined data reported to the Center for Diseases Control and Prevention, Republic of Korea from 2001–2011. The mean month of malaria incidence was mid-July with 10.4 infections/100,000 after estimation by the best-fit hierarchical generalized linear model (HGLM).

3.5. Spatial distribution of malaria for males only, males excluding 20–29-year-olds, and females only in 2011

The diseases map for males only in 2011 was nearly identical to the overall diseases map of malaria in 2011 (Figure 3). If the males aged 20–29 were omitted, the diseases map showed the low or negligible incidence in areas other than northern part of Korea. Cities or counties not adjacent to the DMZ with countable incidence were Uljin, Ichon, Boryeong, Gunsan, and Iksan. The diseases map for females only showed that only two cities or counties are countable apart from the regions adjacent to the DMZ i.e. Jincheon and Hongseong (Figure 7).

4. Discussion

According to the results, it can be said that combination model of Spatial Model 3 (Markov random field model) and Temporal Model 3 (the first order autoregressive random work model) provided the best fit to model of malaria incidence in Korea. According to the best-fit model, it can be seen that there was a significant correlation of malaria incidence between adjacent regions and also significant correlation between the malaria incidence of year and that of recent years.

Since malaria in Korea is primarily transmitted by mosquitoes from North Korea, the distance from the DMZ is important (Figure 1) [1]. Therefore, the incidence rate of a region should be affected by the incidence of adjacent regions. Also, when there is a malaria infection in an area distant from the DMZ, the infected mosquito can move to another adjacent region. The annual incidence is affected by that of recent years since there might be a dormant infection of malaria with a long incubation period [10]. Also the infected mosquito can survive into the next year. The best-fit model can explain these phenomena in Korea very well (Figure 3).

The peak incidence was seen in the age 20–24 years group (Figure 5). This is due to the fact that population most vulnerable to malaria is soldiers who work near the DMZ and discharged soldiers who were infected during military service. As for females, middle-aged females in rural areas work actively on farms where the density of mosquitoes is higher than in urban areas (Figure 6).

According to the pattern of annual incidence, the incidence of malaria in Korea might decrease year by year. The epidemiological pattern in Korea is typical unstable malaria: transmission or incidence is variable from year to year and sometimes there are outbreaks [2]. The re-emergence is dependent on the malaria infection status in North Korea [11]. If the public

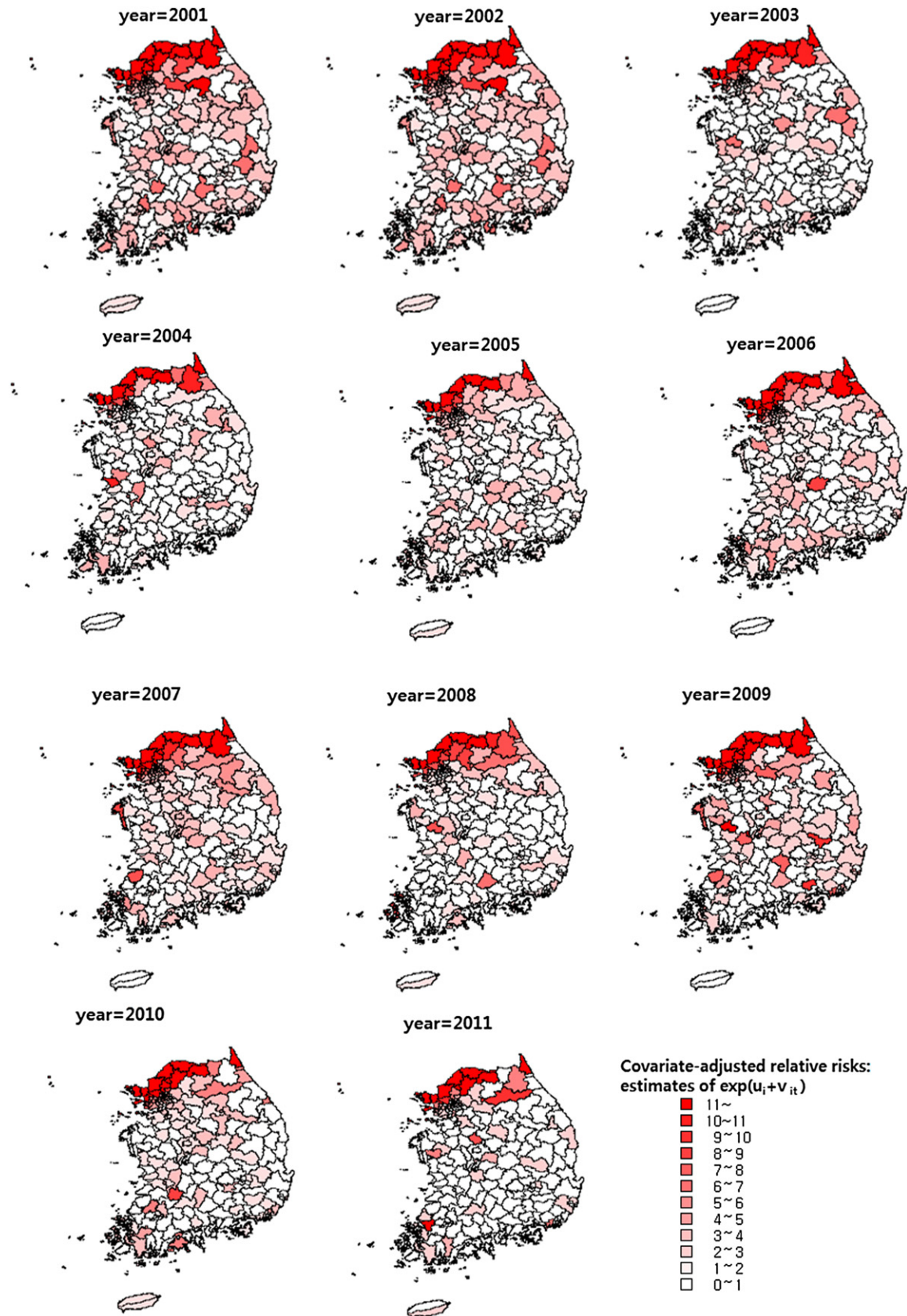


Figure 4. Disease maps of incidence of *Plasmodium vivax* malaria reported to the Center for Diseases Control and Prevention, Republic of Korea from 2001–2011. Incidence is presented as number of infections/100,000 after estimation by the best-fit hierarchical generalized linear model (HGLM).

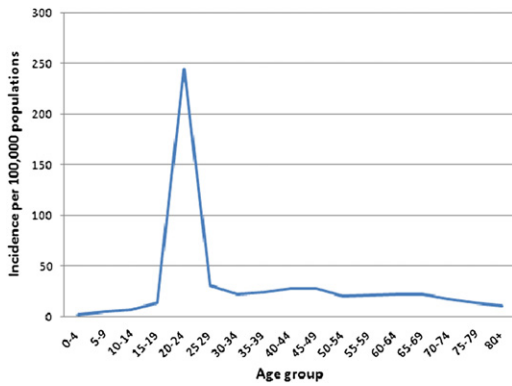


Figure 5. Incidence of *Plasmodium vivax* malaria reported to the Center for Diseases Control and Prevention, Republic of Korea from 2001 to 2011 according to age. The peak was the 20–25-year-old group (244.52 infections/100,000).

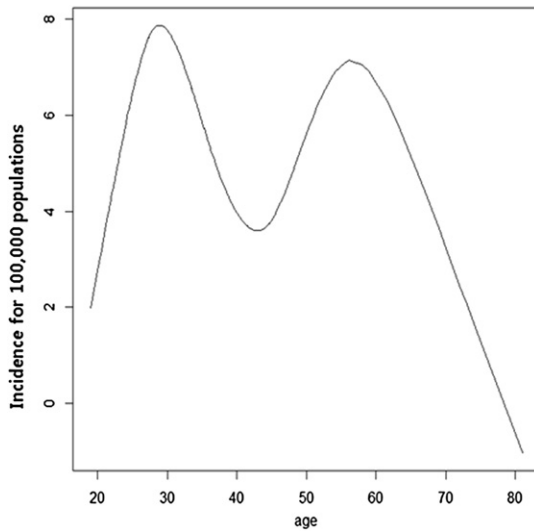


Figure 6. Mean age of male and female infected with *Plasmodium vivax* from smoothing splined data reported to the Center for Diseases Control and Prevention, Republic of Korea from 2001 to 2011. Mean age of males and females was 31.0 years and 45.3 years with incidence 7.8 infections/100,000 and 7.1 infections/100,000, respectively, after estimation by the best-fit hierarchical generalized linear model (HGLM).

health status of North Korea does not decrease rapidly, the malaria incidence in North Korea will decrease. Although there is a regional cycle of malaria infection in a few regions in Korea, the incidence is very low except in areas adjacent to the DMZ. All malaria cases are now reported to the CDC of the Republic of Korea, and can be treated well with a standard regimen of chloroquine and primaquine. A few cases of chloroquine-resistant strains have been reported [12], but this is not a basis for the persistence of incidence. It is said that malaria is endemic in Korea since there is a regional cycle of malaria infection in a few regions as well as a rapid dissemination of newly introduced genotypes [13]. However, since the malaria incidence is limited to regions adjacent to the DMZ, continuous transmission of malaria is difficult to anticipate throughout the whole country. These findings are well presented in the diseases map of Figure 7, where the map of males only, excluding the 20–29-year-old group, and the map of females only showed very low or negligible incidence except in a few regions. The incidence of infection in men is nearly same as the incidence of both sexes. The estimation of the incidence in the following year can be done with further one of regression model in the range of HGLM.

In conclusion, to estimate the malaria incidence in Korea from 2001–2011, the combination of the spatial and temporal effects gives the best fitting model. According to the model, incidence at a region was influenced by adjacent regions. Also the incidence of a year was affected by that of recent previous year. There is a decreasing trend of malaria incidence in Korea over the recent several years if there is no further outbreak of malaria in North Korea.

Supplement. R script for the estimation of *Plasmodium vivax* malaria from data reported to the Korea Centers for Disease Control and Prevention, from 2001–2011.

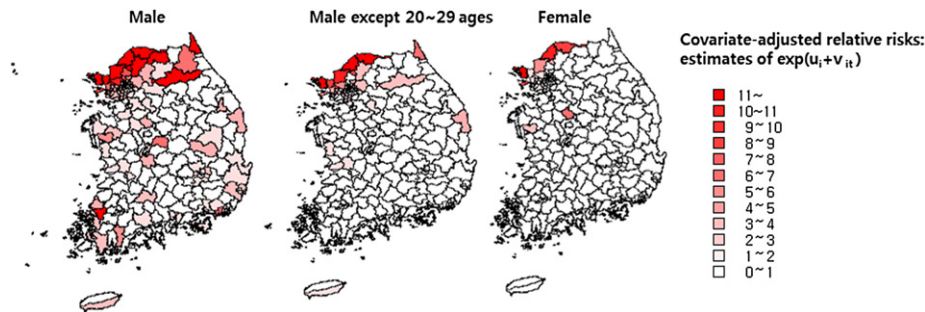


Figure 7. Disease maps of incidence of *Plasmodium vivax* malaria according to sex and the map of male excluding the 20–29-year-old group reported to the Center for Diseases Control and Prevention, Republic of Korea for 2011. Incidence is presented as number of infections/100,000 population after estimation by the best-fit hierarchical generalized linear model (HGLM).

Acknowledgments

This work was supported by the research grant of Center for Diseases Control and Prevention, Republic of Korea in 2012 (2012E2400100).

References

1. Lee JS, Lee WJ, Cho SH, et al. Outbreak of vivax malaria in areas adjacent to the demilitarized zone, South Korea, 1998. *Am J Trop Med Hyg* 2002 Jan;66(1):13–7.
2. Park JW, Jun G, Yeom JS. *Plasmodium vivax* malaria: status in the Republic of Korea following reemergence. *Korean J Parasitol* 2009 Oct;47(Suppl.):S39–50.
3. Clayton DG, Kaldor J. Empirical Bayes estimates of age-standardized relative risks for use in disease mapping. *Biometrics* 1987;43(3):671–81.
4. Ghosh M, Natarajan K, Stroud TWF, Carlin BP. Generalized linear models for small-area estimation. *J Am Stat Assoc* 1998; 93(1):273–82.
5. Lee Y, Jang M, Lee W. Prediction interval for disease mapping using hierarchical likelihood. *Comput Stat* 2011 March;26(1):159–79.
6. Noh M, Lee Y. dhglm: Double hierarchical generalized linear models. R package version 1.0 [Internet]. Vienna (Austria): The R Foundation for Statistical Computing; 2012 [cited 2012 Nov 1]. Available from: <http://CRAN.R-project.org/package=dhglm>.
7. Besag JE, Higdon DM. Bayesian analysis of agricultural field experiments. (with Discussion). *J R Stat Soc B* 1999;61(4): 691–746.
8. Cressie N. *Statistics for Spatial Data*. Revised ed. Wiley-Interscience; 1993. p. 557.
9. MapWizard for Excel, Version 1.0. Seoul: TasTech Corp; 2006.
10. Nishiura H, Lee HW, Cho SH, et al. Estimates of short- and long-term incubation periods of *Plasmodium vivax* malaria in the Republic of Korea. *Trans R Soc Trop Med Hyg* 2007 Apr;101(4): 338–43.
11. Iwagami M, Fukumoto M, Hwang SY, et al. Population structure and transmission dynamics of *Plasmodium vivax* in the Republic of Korea based on microsatellite DNA analysis. *PLoS Negl Trop Dis* 2012;6(4):e1592.
12. Lee KS, Kim TH, Kim ES, et al. Short report: chloroquine-resistant *Plasmodium vivax* in the Republic of Korea. *Am J Trop Med Hyg* 2009 Feb;80(2):215–7.
13. Choi YK, Choi KM, Park MH, et al. Rapid dissemination of newly introduced *Plasmodium vivax* genotypes in South Korea. *Am J Trop Med Hyg* 2010 March;82(3):426–32.