Sains Malaysiana 45(11)(2016): 1741–1745

New Approach to Calculate the Denominator for the Relative Risk Equation

(Pendekatan Baharu untuk Menghitung Pembawah bagi Persamaan Risiko Relatif)

NOR AZAH SAMAT* & SYAFIQAH HUSNA MOHD IMAM MA'AROF

ABSTRACT

Disease frequency is used to measure the situation of the disease with reference to the population size and time period which is in a fractional form. The lower part of the fraction, known as denominator is the important part as it was used to calculate a rate or ratio. Since the disease frequency is based on a ratio estimator, the results are highly dependent upon the value of denominator. Therefore, the main aim of this paper was to propose a new method in calculating the denominator for the relative risk equation with the application to chikungunya disease data from Malaysia. The new method of calculating the denominator of the relative risk equation includes the use of discrete time-space stochastic SIR-SI (susceptible-infective-recovered for human population and susceptible-infective for vector population) disease transmission model instead of the total disease counts. The results of the analysis showed that the estimation of expected disease counts based on total posterior means can overcome the problem of expected counts estimation based on the total number of disease especially when there is no observed disease count in certain regions. The proposed new approach to calculate the denominator for the relative risk equation is suitable for the case of rare disease in which it offers a better method of expected disease counts estimation.

Keywords: Chikungunya disease; disease mapping; relative risk estimation; SIR-SI disease transmission model

ABSTRAK

Frekuensi penyakit digunakan untuk mengukur situasi sesuatu penyakit dengan merujuk kepada saiz populasi dan tempoh masa yang berbentuk pecahan. Bahagian bawah pecahan, yang dikenali sebagai pembawah ialah bahagian yang penting kerana ia digunakan untuk menghitung suatu kadar atau nisbah. Memandangkan frekuensi penyakit adalah berasaskan suatu anggaran nisbah, keputusan anggaran sangat bergantung kepada nilai pembawah tersebut. Oleh itu, matlamat utama kajian ini ialah untuk mencadangkan suatu kaedah baharu dalam mengira pembawah bagi persamaan risiko relatif dengan aplikasi kepada data penyakit chikungunya dari Malaysia. Kaedah baru pengiraan pembawah bagi persamaan risiko relatif mengambil kira penggunaan model jangkitan penyakit stokastik diskrit masa-ruang SIR-SI (rentan-jangkitan-pulih bagi populasi manusia, rentan-jangkitan bagi populasi vektor) dan bukan jumlah bilangan penyakit. Hasil analisis menunjukkan bahawa penganggaran bilangan jangkaan penyakit berdasarkan jumlah posterior min dapat mengatasi masalah penganggaran jumlah jangkaan berdasarkan jumlah bilangan penyakit khususnya apabila tiada penyakit yang diperhatikan dalam sesuatu kawasan. Kaedah baru yang dicadangkan untuk mengira pembawah bagi persamaan risiko relatif adalah sesuai bagi kes penyakit yang jarang berlaku kerana ia menawarkan kaedah yang lebih baik bagi penganggaran bilangan jangkaan penyakit.

Kata kunci: Model jangkitan penyakit SIR-SI; pemetaan penyakit; penganggaran risiko relatif; penyakit chikungunya

INTRODUCTION

Relative risk is one of the methods used to measure the frequency of disease. In epidemiology studies, several other terms that can be used to measure the disease frequency are prevalence, risk, odds and incidence rates, in which all terms are in a fractional form. These four different ways to measure disease frequency were discussed by Bailey et al. (2005). Other research on the measurement of disease frequency can be found in Coutinho et al. (2008), Lindenauer et al. (2010), Stefan et al. (2013) and Szumilas (2010). Many studies used these measurement methods to control health problem (i.e. de la Hunty et al. 2013; Mathenge et al. 2013; McKone et al. 2010). Each measurement will involve two important parts which are the numerator and the denominator. Numerator is the upper part of a fraction and denominator is the lower part of a fraction. The values of numerator and denominator used in a study will obviously affect the value of the disease frequency. For instance, zero value for the denominator will subsequently make the disease frequency difficult to estimate.

PROBLEM STATEMENT

In the study of disease mapping, relative risk estimation is the main object of the analysis. A study by Samat and

Percy (2012) showed that the formula that can be used to estimate the relative risk, r_{ij}, for i regions and j time periods is as follows:

$$r_{ij} = \frac{o_{ij}}{e_{ij}}, \tag{1}$$

where o_{ij} is the observed disease count, e_{ij} is the expected disease count, both for i regions; and j is the time periods. The most common method used to estimate the expected disease count or the denominator is based on the population standardization as follows,

$$e_{ij} = N_i \frac{\sum_{j=1}^{53} o_{ij}}{\sum_{i=1}^{16}\sum_{j=1}^{53} N_{ij}}. \tag{2}$$

From (2), the summation of the numerator represents the total number of observed disease count for j time periods, meanwhile the summation for the denominator represents the overall total population. Here the N_i represents the total population for i regions.

Equation (1) has been improved by Samat and Percy (2012) in order to overcome the problem when there is no observed disease count in certain regions. They incorporate the space-time factor and the disease transmission model in the calculation of the relative risk specifically for the numerator of the equation that is based on the discrete time-space stochastic SIR-SI model which can be written as follows,

$$r_{ij} = \frac{\tilde{\lambda}_{ij}}{e_{ij}}, \tag{3}$$

where $\tilde{\lambda}_{ij}$ is the posterior mean of disease count; and e_{ij} is the same definition as explained earlier and the calculation of it is similar as in (2). Unfortunately (2) has exactly the same problem as (1) especially when there is no observed disease count in certain regions. For instance, in (2), when the summation of o_j is zero, then the e_{ij} will subsequently become zero. This problem is not previously discovered by Samat and Percy (2012). This is because in their study, dengue data are used to demonstrate the model and dengue in Malaysia is considered as non-rare disease. Hence, no zero observed dengue disease occurred weekly for each region. Therefore, in this study, we would like to propose a method to overcome the problem in (2).

Research Objectives

The objectives of this study were: To investigate and propose a new method of expected disease count estimation for the denominator of relative risk equation and to demonstrate the advantage of the proposed method of expected disease count estimation using chikungunya data from Malaysia.

Methods

In this study, the equation for the expected disease count is shown in (2) which have been improvised based on the same approach proposed by Samat and Percy (2012). Here, the disease count is recognized as the number of new infective human. For ($i = 1, 2,..., M$) study regions and ($j = 1, 2,..., T$) time periods, a pseudo-random sample of observations which is for $k = 1, 2,..., n$ is generated from the posterior distribution for the mean number of infective human (λ_{ij}). From this sample, the posterior expected mean number of infective human can be approximated using the unbiased sample as follows:

$$\tilde{\lambda}_{ij} = \frac{1}{n}\sum_{k=1}^{n} \lambda_{ijk}. \tag{4}$$

While the posterior expected number of new infective for human ($\tilde{e}_{ij}$) can be defined as,

$$\tilde{e}_{ij} = N_i \frac{\sum_{j=1}^{53} \tilde{\lambda}_{ij}}{\sum_{i=1}^{16}\sum_{j=1}^{53} N_{ij}}, \tag{5}$$

where N_i is the population of region; i is the human population and the summations are for $i= 1, 2, ..., M$ study regions; and j is 1, 2, ..., T time periods. Therefore, the relative risk can be estimated using (6),

$$\tilde{r}_{ij} = \frac{\tilde{\lambda}_{ij}}{\tilde{e}_{ij}}. \tag{6}$$

From (2), it can be seen that the numerator of the equation is based on the total number of observed cases. Conversely, (5) shows that the estimation of the expected disease counts consider a discrete time-space stochastic SIR-SI disease transmission model in order to give the posterior number of new infected cases in the numerator of the equation. Here, the stochastic SIR-SI model is initially comes from the model proposed by Samat and Percy (2012), and details explanation of the disease transmission model can be found in their study.

Results

The Data Set

Data used in the analysis of this study were provided by the Ministry of Health, Malaysia. The data involved are the chikungunya disease data for every 16 states in Malaysia from epidemiology week 1 to epidemiology week 52 in the year of 2013. The states involve Perlis, Kedah, Pulau Pinang, Perak, Kelantan, Terengganu, Pahang, Selangor, Kuala Lumpur, Putrajaya, Negeri Sembilan, Melaka, Johor, Sarawak, Labuan and Sabah.

ESTIMATION OF EXPECTED DISEASE COUNTS USING CHIKUNYA DATA OF MALAYSIA

Figure 1 represents the number of cases for every 16 states during 2013. Based on the bar graph, it can be seen that Sabah has the highest number of chikungunya cases which was 2 cases during epidemiology week 12. Perak has recorded one case during epidemiology weeks 8, 13 and 48. The total observed disease count of cases at Perak during 2013 is 3. Meanwhile there is one chikungunya disease case at Sarawak during epidemiology week 3 and at Selangor during epidemiology week 19.

Figure 2 depicts the line graph for the estimated expected disease count which is based on the observed disease cases. From Figure 1, it can be seen that there are chikungunya cases occurred in epidemiology weeks 3, 8, 12, 13, 19 and 48. Therefore, the value of estimated expected disease count shown in Figure 2 only appears in the states and epidemiology weeks in which the disease cases are exist. Meanwhile, other states remain zero. From Figure 2, the highest estimated expected disease count is at Selangor during epidemiology week 12 with the value of 0.38636. The second highest estimated expected disease count is still at epidemiology week 12 but the state is Sabah with the value of 0.23352. The remaining states have estimated expected disease count below 0.2.

Figure 3 depicts the estimated expected disease count which is based on the posterior means that has been proposed (5). This method can overcome the weakness of calculation of expected disease count based on observed disease count as shown in (2) especially when there is no observed disease count in certain region and time period. In addition, the time series plot shows that the line graph does not touch x-axis, which means that all estimated relative risks are more than zero for all states and all epidemiology weeks. The highest estimated expected disease count is at Selangor during epidemiology week 3 with 0.35298. All states have high estimation during the early epidemiology weeks of the year and went down through the end of the year.

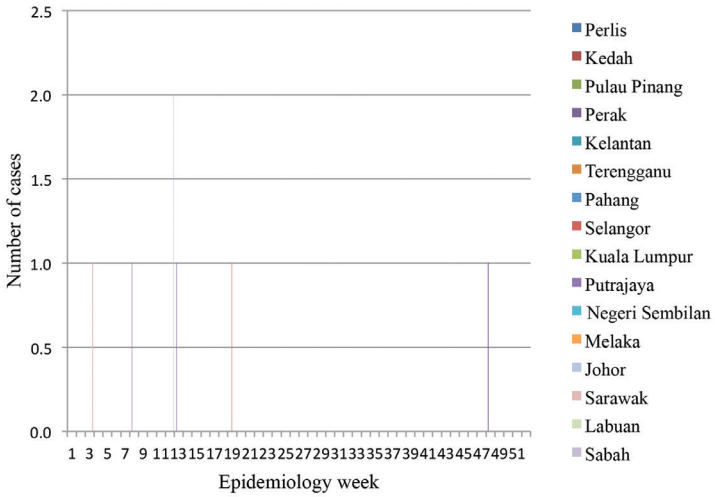

FIGURE 1. Chikungunya cases for all states in Malaysia during 2013

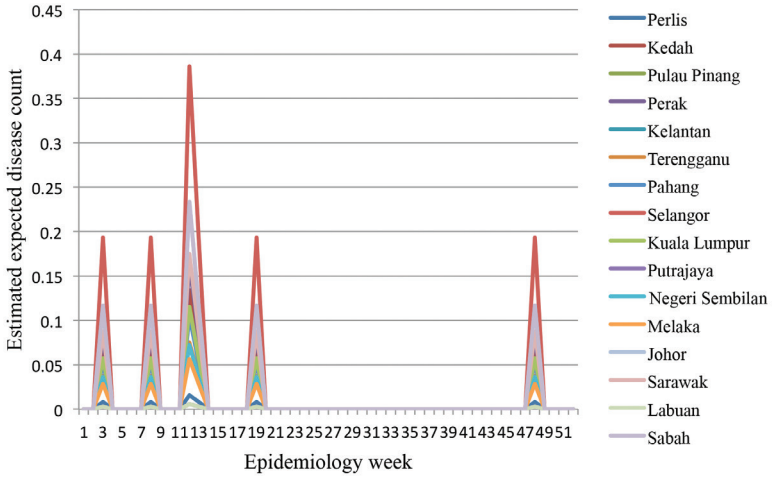

FIGURE 2. Estimated expected disease count based on total number of observed cases

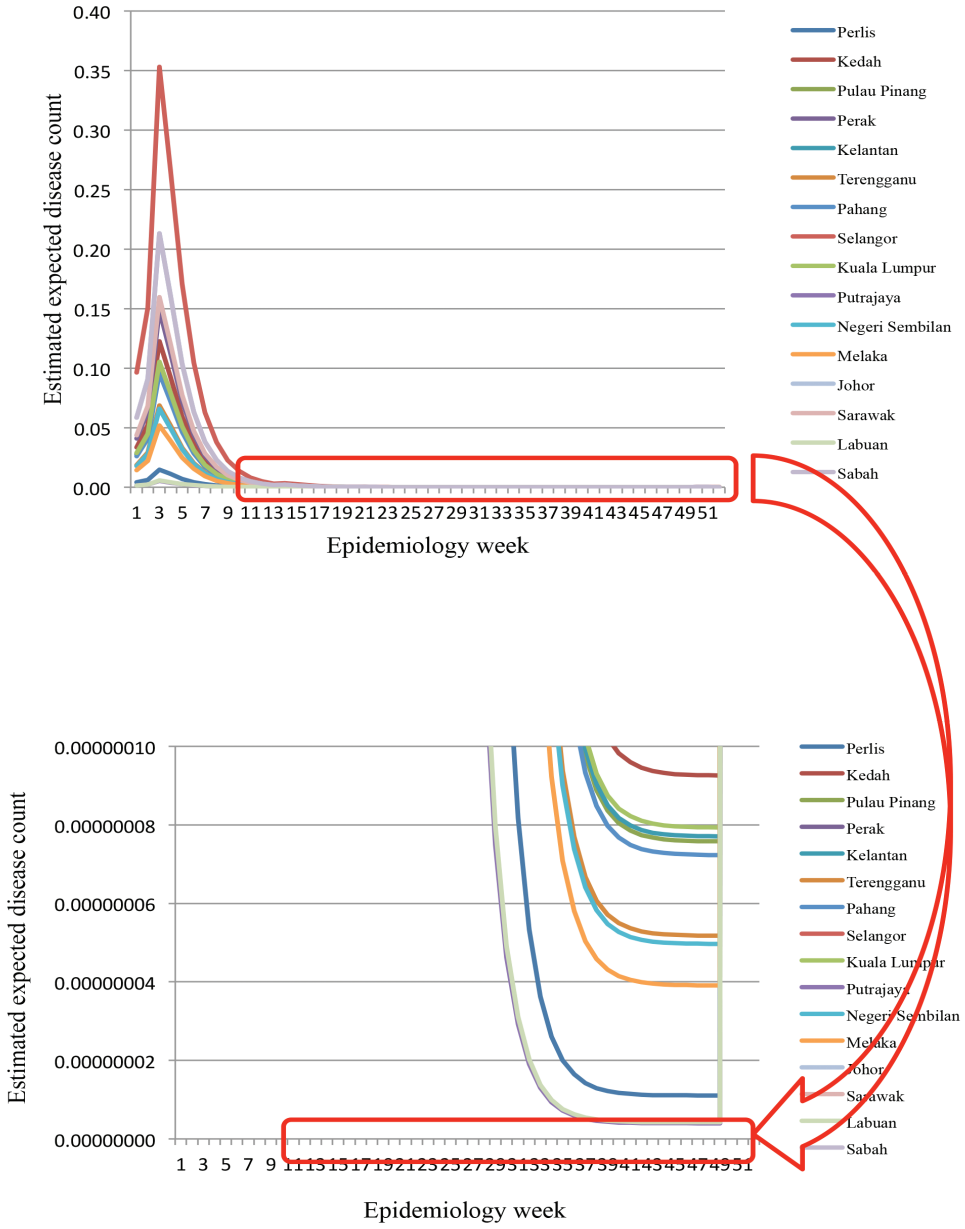

FIGURE 3. Estimated expected disease count based on total posterior means

Table 1 is the comparison of the estimated expected disease count by using total number of observed cases and the use of total posterior means which is based on discrete time-space stochastic SIR-SI model during epidemiology week 11. From the table, it can be seen that the used of total posterior means (5) can overcome the drawback of the used of observed disease count (2). The latter method gives zero value of expected disease count for the case when there is no observed disease count in the regions. Meanwhile, the first method has been demonstrated did not suffer from this problem. In reality, people in the whole population are at risk of contracting any diseases, either low or high risks. No observed case or zero case of disease occurrence does not means that people in the area are not at risk. Therefore, it is very important to have the risk value even if in reality no such case is seen. For epidemiology week 11, the highest estimated expected disease count is at Selangor with 0.00819. This was probably due to Selangor has the highest number of population among other states during 2013. The second highest estimated expected disease count is 0.00495 at Johor and Sabah. Both states have similar estimated value since the numbers of population are similar during 2013.

CONCLUSION

Better method used to estimate the expected disease count or the denominator of relative risk equation will subsequently give better relative risk estimation. In this study, findings of the research has shown that the proposed new approach to calculate the denominator for the relative risk equation offers better method of expected disease

TABLE 1. Comparison of the estimated expected disease count for epidemiology week 11

State	Estimated expected disease count based on total number of observed cases	Estimated expected disease count based on total posterior means
Perlis	0	0.00034
Kedah	0	0.00285
Pulau Pinang	0	0.00233
Perak	0	0.00346
Kelantan	0	0.00237
Terengganu	0	0.00159
Pahang	0	0.00223
Selangor	0	0.00819
Kuala Lumpur	0	0.00244
Putrajaya	0	0.00012
Negeri Sembilan	0	0.00153
Melaka	0	0.00120
Johor	0	0.00495
Sarawak	0	0.00371
Labuan	0	0.00013
Sabah	0	0.00495

count estimation especially when there is no observed disease case in certain regions or rare disease.

ACKNOWLEDGEMENTS

The authors acknowledge Universiti Pendidikan Sultan Idris and the Ministry of Higher Education (MOHE), Malaysia (RAGS code: 2013-0006-102-72) for their financial support in respect of this study, and the Ministry of Health (MOH), Malaysia for the data of mosquito-borne infectious disease provided.

REFERENCES

Bailey, L., Vardulaki, K., Langham, J. & Chandramohan, D. 2005. *Introduction to Epidemiology*. London: Open University Press.

Coutinho, L.M.S., Scazufca, M. & Menezes, P.R. 2008. Methods for estimating prevalence ratios in cross-sectional studies. *Revista de Saúde Pública* 42(6): 992-998.

de la Hunty, A., Gibson, S. & Ashwell, M. 2013. Does regular breakfast cereal consumption help children and adolescents stay slimmer? A systematic review and meta-analysis. *Obes Facts* 6(1): 70-85.

Lindenauer, P.K., Bernheim, S.M., Grady, J.N., Lin, Z., Wang, Y., Merrill, A.R., Han, L.F., Rapp, M.T., Drye, E.E., Normand, S.L. & Krumholz, H.M. 2010. The performance of US hospitals as reflected in risk-standardized 30-day mortality and readmission rates for medicare beneficiaries with pneumonia. *Journal of Hospital Medicine* 5(6): E12-18.

Mathenge, W., Bastawrous, A., Peto, T., Leung, I., Foster, A. & Kuper, H. 2013. Prevalence of age related macular degeneration in Nakuru, Kenya: A cross-sectional population-based study. *PLoS Medicine* 10(2): e1001393.

McKone, E., Aimola Davies, A., Fernando, D., Aalders, R., Leung, H., Wickramariyaratne, T. & Platow, M.J. 2010. Asia has the global advantage: Race and visual attention. *Vision Research* 50(16): 1540-1549.

Samat, N.A. & Percy, D.F. 2012. Vector-borne infectious disease mapping with stochastic difference equations: An analysis of dengue disease in Malaysia. *Journal of Applied Statistics* 39(9): 2029-2046.

Stefan, M.S., Pekow, P.S., Nsa, W., Priya, A., Miller, L.E., Bratzler, D.W., Rothberg, M.B., Goldberg, R.J., Baus, K. & Lindenauer, P.K. 2013. Hospital performance measures and 30-day readmission rates. *Journal of General Internal Medicine* 28(3): 377-385.

Szumilas, M. 2010. Explaining odds ratios. *Journal of the Canadian Academy of Child and Adolescent Psychiatry* 19(3): 227-229.

Department of Mathematics
Faculty of Science and Mathematics
Universiti Pendidikan Sultan Idris
35900 Tanjong Malim, Perak Darul Ridzuan
Malaysia

*Corresponding author; email: norazah@fsmt.upsi.edu.my

Received: 21 May 2015
Accepted: 24 March 2016