

Comparison of HIV and AIDS Diseases Mapping in Malaysia Based on Standardized Morbidity Ratio and Poisson-Gamma Model

Perbandingan Pemetaan Penyakit HIV dan AIDS Berdasarkan Standardized Morbidity Ratio dan Model Poisson-Gamma

Sufi Hafawati Ideris & Nor Azah Samat

Department of Mathematics, Faculty of Science and Mathematics, University Pendidikan Sultan Idris,
35900 Tanjung Malim, Perak, Malaysia
e-mail: sufi.hafawati@gmail.com, norazah@fsmt.ups.edu.my

Abstract

Human Immunodeficiency Virus (HIV) is the virus that causes chronic and life-threatening condition called Acquired Immune Deficiency Syndrome (AIDS). There is no currently available vaccine or cure for the disease. Therefore, to control the epidemic from continuing to spread will entirely depend on the control, treatment and prevention strategies. Disease maps have been recognized as one of the important tools in controlling diseases. Better statistical models used to estimate the relative risk will subsequently produce better disease risk maps. Therefore, the aim of this paper is to estimate the relative risk for HIV and AIDS in Malaysia based on the commonest method used in the study of disease mapping which are the Standardized Morbidity Ratio (SMR) and Poisson-Gamma models. In this paper, a review of the SMR method will be given and then will be applied to HIV and AIDS data in Malaysia. This application of SMR is then extended to the next method which is the Poisson-Gamma model and the results are compared in tables, graphs and maps. Result of the analyses shows that the latter model gives better estimation of relative risks compared with the SMR method due to the problem of SMR especially when there is no observed HIV and AIDS case in certain states in Malaysia. Notwithstanding, a major drawback of Poisson-Gamma is that the model does not allow for spatial correlation to exist between risks areas in adjacent areas and this has motivated many researchers to introduce other better alternative methods of relative risk and this estimation for disease mapping.

Keywords HIV and AIDS cases, disease mapping, relative risk, Standardized Morbidity Ratio (SMR), Poisson-Gamma Model

Abstrak

Virus Kurang Daya Tahan Penyakit (HIV) adalah virus yang menyebabkan penyakit kronik dan mengancam nyawa yang dipanggil Sindrom Kurang Daya Tahan Penyakit (AIDS). Buat masa ini, tidak ada vaksin atau penawar bagi penyakit ini. Oleh itu, untuk mengawal wabak ini daripada terus merebak sepenuhnya ia bergantung kepada strategi kawalan, rawatan dan pencegahan. Pemetaan penyakit telah diketahui sebagai salah satu alat penting dalam mengawal penyakit. Model statistik terbaik yang digunakan untuk menganggarkan risiko relatif kemudiannya akan menghasilkan peta risiko penyakit yang lebih baik. Oleh

itu, tujuan kajian ini adalah untuk menganggarkan risiko relatif bagi HIV dan AIDS di Malaysia berdasarkan kaedah paling lazim digunakan dalam kajian pemetaan penyakit iaitu *Standardized Morbidity Ratio* (SMR) dan model Poisson-Gamma. Dalam kajian ini, ulasan terhadap kaedah SMR akan diberikan dan kemudian digunakan terhadap data HIV dan AIDS di Malaysia. Aplikasi SMR ini kemudiannya dilanjutkan kepada kaedah seterusnya iaitu model Poisson-Gamma dan hasilnya dibandingkan dalam bentuk jadual, graf dan peta. Keputusan analisis menunjukkan bahawa model kedua memberikan anggaran risiko relatif yang lebih baik berbanding dengan kaedah SMR disebabkan masalah utama SMR khususnya apabila tiada kes HIV dan AIDS yang dicerap di negeri-negeri tertentu di Malaysia. Walaubagaimana pun, kelemahan utama model Poisson-Gamma ialah model ini tidak membenarkan korelasi ruang antara risiko bagi kawasan yang bersebelahan yang mana ia telah mendorong ramai penyelidik memperkenalkan kaedah alternatif lain yang lebih baik bagi menganggar relatif risiko bagi pemetaan penyakit.

Kata kunci HIV dan AIDS, pemetaan penyakit, risiko relatif, *Standardized Morbidity Ratio* (SMR), Model Poisson-Gamma

INTRODUCTION

Research Background and Problem Statement

Recently, disease mapping method has been frequently used in the analysis of public health and has become one of the conventional methods used to study the geographical distribution of disease (Koch, 2005). In the study of disease mapping, the SMR method and Poisson-Gamma model are the most common methods used to estimate the relative risk. In this paper, the focus of attention is to study the geographical distribution of HIV and AIDS in Malaysia based on these common methods. This is because, in Malaysia, most of the studies [see for example A Review of HIV/AIDS Research in Malaysia (2014), Global AIDS Response (2012), Malaysia UNGASS Country Progress Report (2012), MGD Combat HIV/AIDS, Malaria and Other Diseases, National Strategic Pelan on HIV and AIDS 2011-2015 (2011)] have discussed more on risk factors based on behaviour in order to determine the here relative risk of HIV and AIDS such as injected drug users, sex workers, men who have sex with other men and transgender persons. It seems quite elusive to find any study in this country that focuses into high and low risks of HIV and AIDS based on geographical distributions. Therefore, this research would like to discuss and demonstrate the estimation of relative risk for each state in Malaysia based on SMR method and Poisson-Gamma model for HIV and AIDS mappings. These two methods are functioning to estimate relative risk which is not based on the frequency method where a high number of HIV and AIDS cases represent high risk areas, while a low number of HIV and AIDS cases represent low risk areas. In fact, these SMR method and Poisson-Gamma model will estimate the relative risk that take population sizes into consideration.

Research Aim

The purpose of this paper is to propose a risk map for HIV and AIDS cases in Malaysia by showing area of high and low risks of disease based on SMR method and Poisson-Gamma model. In this study, the value for relative risk that is close or equal to one means that there

is no real difference between the conditional probability that a person in the understudy region contracting the disease compared to the conditional probability that a person in the general population contracting the disease.

Next, if the value of relative risk is greater than 1, this indicates that people within the study region are more likely to contract the disease compared to people in the general population. On the contrary, a value of relative risk below 1 indicates that a decrease in likelihood of contracting the disease which means that people in the study region are less likely to contract the disease compared to the people in the general population.

METHODOLOGY

Standardized Morbidity Ratio (SMR)

SMR usually compares the observed incidence to the expected incidence, which has been traditionally used for the analysis of tract count and case events maps as explained in (Lawson et.al, 2003). In this paper, count data is used to examine the occurrences of a disease within a map where the data are aggregated as a count of disease within a sub region of the map (Lawson, 2006). In this tract counts analysis, SMR is the relative risk method used with the ratio of observed cases to expected count within tracts. It is used to estimate the relative risk where the likelihood of a person within a particular region contracting the disease divided by with the probability of person in the population contracting the disease.

In disease mapping, the area of disease map is divided into M mutually exclusive regions where $i = 1, 2, \dots, M$ and each region is assumed to be homogenous with its own number of observe cases O_i and expected number of cases E_i as obtained based on available data which are being used to calculate relative risk, $\hat{\theta}$ for region i , and the SMR is defined as

$$\hat{\theta} = \frac{O_i}{E_i} \quad (1)$$

Although SMR has been used widely as the estimation method of relative risk (Samat and Percy, 2008), it is may be unreliable to measure the relative risk for a small areas such as counties in the US as discussed in (Mark and David, 2010). This is because, the SMR is large when it is applied to a small population size that gives small number of expected cases. Meanwhile, SMR is small when the saiz of population is large as the number of expected cases is also large.

$$E_i = N_i \frac{\sum O_i}{\sum N_j} \quad (2)$$

According to equation (2), N_i is the population of region i and the summation are for $j= 1, 2, \dots, M$. For example, based on (2) when the value of population size of region 1 is 2 000 000 and the total value for observed HIV cases is 98 while the total population for region, N_1 represented by 1 is 23 000, the expected valueis approximatelyequal to 1.127 which is large compared to the expected value of 0.588 for region 2 with a total population of 12,000. This shows that the expected number of cases will directly depend on the total

population for a particular region while the SMR is inversely proportional to the value of expected cases. Therefore, every region with a small population will have a high number of SMR. Thus, SMR is not feasible for small area or counties with a small number of population. Furthermore, when there is no observed case in certain regions, the number of SMR will be zero. For instance, if the number of observed HIV case for region 1 is zero, then this value is divided with the value of expected cases, which will give a zero value for SMR. This indicates that for any region with zero number of observed HIV or AIDS cases will subsequently give no appearance of risk map for the disease.

Poisson-Gamma Model

In Bayesian analysis, the posterior distribution is the product of the likelihood and the prior distribution (lawson et.al, 2003). The posterior distribution can be defined as

$$\text{Posterior} \propto \text{Likelihood} \times \text{prior}$$

Analytically, in Bayes method, likelihood is determined based on the observed data which is proportional to the joint probability distribution function of the sample of data Lawson et.al, 2003). Let assume that θ is the parameter and the observed data are obtained independently and identically as follow;

$$x = x_1, x_2, \dots, x_n$$

Then, the likelihood function is:

$$L(\theta | x) = p(x_1, x_2, \dots, x_n | \theta) = \prod_{i=1}^n p(x_i | \theta) \quad (3)$$

Meanwhile, in Bayesian framework, the prior distribution is prior assertions or assumptions of a prior distribution of the parameter, θ represents as $g(\theta)$ (Samat and Percy, 2014). Thus, the posterior distribution is determined as follow;

$$P(\theta | y) \propto P(y | \theta)g(\theta) \quad (4)$$

Therefore in this model, y_{ij} is the number of new infectives of HIV and AIDS for state i at time j for $i = 1, 2, \dots, M$ study region and $j = 1, 2, \dots, T$ time period with mean of Poisson distribution is $e_{ij}\theta_{ij}$ where e_{ij} represents the expected number for the new infectives while θ_{ij} is the relative risk. y_{ij} is assumed to follow a Poisson Distribution as below:

$$y_{ij} | e_{ij}, \theta_{ij} \sim \text{Poisson}(e_{ij}\theta_{ij}) \quad (5)$$

While the relative risk θ_{ij} is assumed to have a Gamma distribution by using parameters α and β . In this case, the prior gamma is:

$$\theta_{ij} \sim \text{Gamma}(\alpha, \beta) \quad (6)$$

This Poisson-Gamma model will give output of posterior expected relative risks for all regions, i and for all time periods, j .

Application of Standardized Morbidity Ratio and Poisson-Gamma Model to HIV and AIDS Data from Malaysia

This section discusses the two applications of relative risk using the commonest methods in the study of disease mapping which are SMR method and Poisson-Gamma model using the observed count of HIV and AIDS reported cases data from 2012 until 2013. The data set are analysed by using WinBUGS software and the disease map will be displayed by using ArcGIS software. Then, all results will be compared and presented in table, graph and maps to reveal the suitable method for estimating relative risk of HIV and AIDS in Malaysia.

The Data Set

The data used in these analyses were provided by the Disease Control Department from Ministry of Health (MOH) Malaysia which involves all states and federal territories in Malaysia. Furthermore, the data counts used refer to the observed new infective of HIV and AIDS cases from the year 2012 until 2013. Malaysia consists of thirteen states which are Perlis, Kedah, Penang, Perak, Selangor, Negeri Sembilan, Melaka, Johor, Pahang, Terengganu, Kelantan, Sabah and Sarawak and three federal territories which are Kuala Lumpur, Putrajaya and Labuan. In this analysis, number of cases for both Federal Territories of Kuala Lumpur and Putrajaya are combined. This is because of the data collections provided by Disease Control Department from MOH has no separated cases for Kuala Lumpur and Putrajaya from the year 2001 until today.

RESULT AND DISCUSSION

The Result

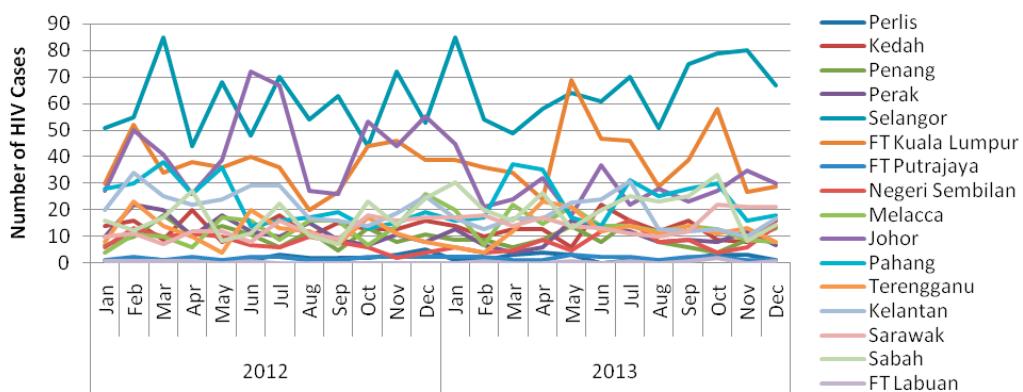


Figure 1 Time series plot of HIV cases for all states in Malaysia

This section discusses the result for the analysis of SMR method and Poisson-Gamma model by using HIV and AIDS data for all states in Malaysia. Figures 1 and 2 depict the time series plots of HIV and AIDS cases reported from January to December in the year 2012 and 2013. Based on Figure 1, it can be seen that the state of Selangor has the highest number of HIV cases while Federal Territory of Labuan has the lowest frequency of HIV cases. The frequency of HIV cases is from zero to 30 for all states with the exception of Selangor, Federal Territory of Kuala Lumpur and Johor where the number of cases for all epidemiology from 2012 to 2013 are about 30 to 85 cases reported. Therefore, based on

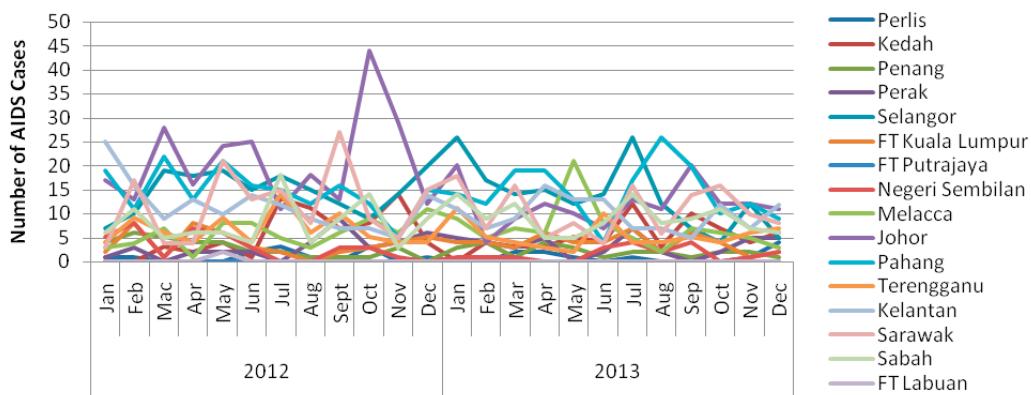
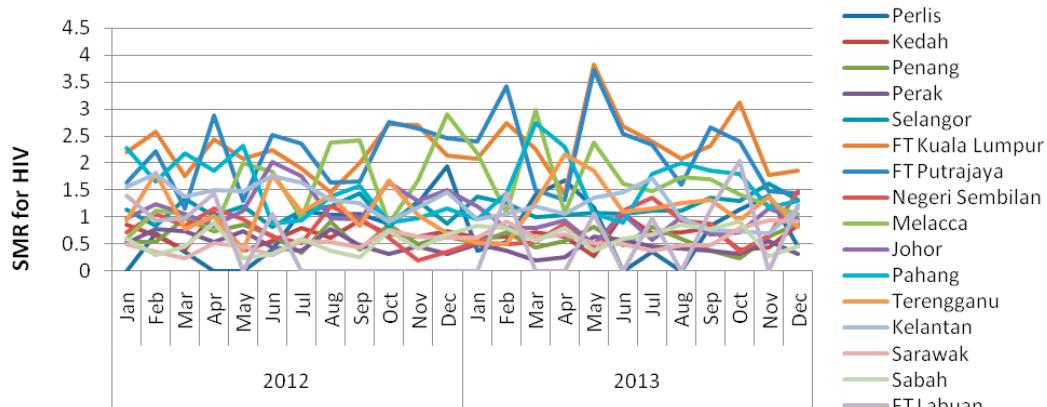


Figure 2 Time series plot of AIDS cases for all states in Malaysia

frequency reported, Selangor, Kuala Lumpur and Johor have highest number of risks for HIV and AIDS. Figure 2 shows a time series plot of AIDS cases for all states in Malaysia from January to December for 2012 and 2013. Based on Figure 2, most cases reported are below 25 cases for every epidemiology month. It can be seen clearly the state of Selangor has the highest number of HIV cases recorded while federal territory of Labuan has the lowest number of HIV cases recorded.

Next, the outcomes of relative risk of HIV and AIDS by using SMR method from the epidemiology months for 2012 and 2013 are displayed in Figures 3 and 4 respectively.



Both graphs show that most of the states have relative risk below 1 for all epidemiology months. This means that people within that regions are less likely to contract HIV and AIDS compared to people in the overall population. Exceptionally, the Federal Territories of Kuala Lumpur & Putrajaya, Malacca and Pahang have relative risks of more than one for HIV cases for all epidemiology months. This leads to a general conclusion that people within the regions are more likely to contract HIV compared to people in the general population. For AIDS cases, the states of Malacca, Pahang, Kelantan and Terengganu have the highest numbers of SMR which is greater than 1.5. This indicates that, people in these states have high risk of contracting AIDS compared with the overall population in Malaysia.

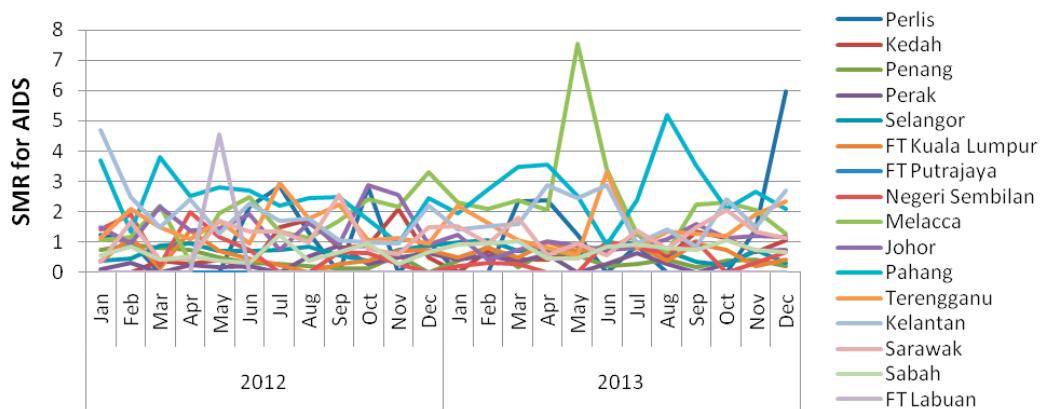


Figure 4 Time series plot of the relative risk for AIDS based on SMR method

Figures 5 and 6 present the time series plot for the relative risk estimation of HIV and AIDS based on Poisson-Gamma model for all states in Malaysia from the epidemiology months January to December as of year 2012 and 2013. From Figure 5, the highest relative risk calculated is Federal Territory of Kuala Lumpur with a relative risk of 3.2 for May 2013. This is followed by, the states of Malacca and Pahang. Therefore, people in Federal

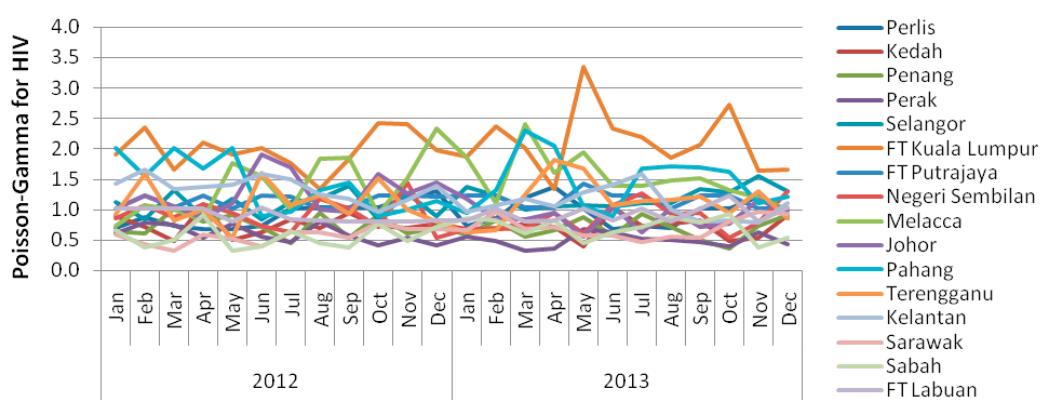


Figure 5 Time series plot the relative risk for HIV based on Poisson-Gamma Model

Territory of Kuala Lumpur, Malacca and Pahang are more prone to contract HIV compared to other people in Malaysia. Meanwhile, most of the relative risks for the rest of the country are below 1 which mean that people from these states are less likely to contract HIV.

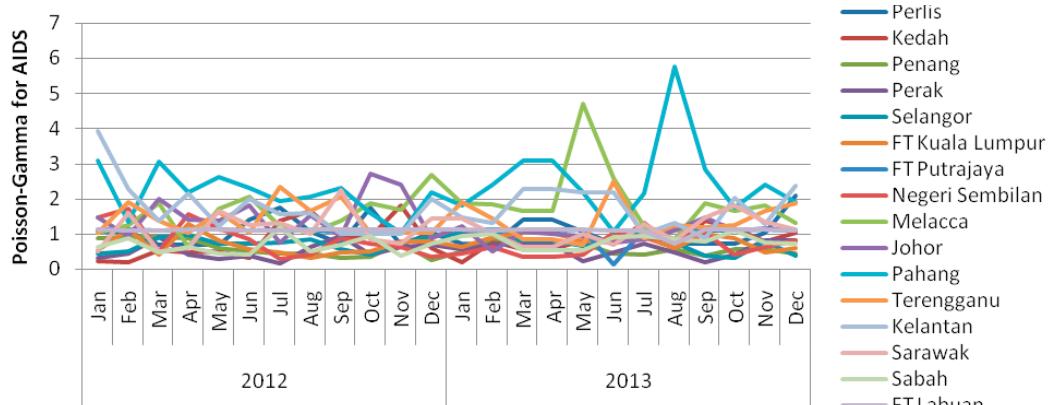


Figure 6 Times series plot of the relative risk for AIDS based on Poisson-Gamma Model

Based on Figure 6, Pahang has the highest number of relative risk calculated for AIDS which is 5.8 as of August in 2012. This means that people from Pahang have higher probability to contract AIDS compared to the overall population in Malaysia. While the

Table 1 Comparison of the relative risk estimation for HIV in epidemiology month **June 2013**

States	Number of Observed Case	Relative Risk based on SMR Method	Relative Risk based on Poisson-Gamma Model
Perlis	0	0	0.696
Kedah	22	1.0561	1.048
Penang	8	0.4792	0.576
Perak	14	0.5587	0.6302
Selangor	61	1.059	1.048
Federal Territory of Kuala Lumpur	47	2.6758	2.334
Federal Territory of Putrajaya	2	2.5407	1.252
Negeri Sembilan	12	1.1142	1.074
Malacca	14	1.6232	1.401
Johor	37	1.0514	1.061
Pahang	14	0.8856	0.9092
Terengganu	12	1.0872	1.077
Kelantan	24	1.4557	1.418
Sarawak	13	0.4998	0.5706
Sabah	20	0.5864	0.6363
Federal Territory of Labuan	0	0	0.8337
Mean	18.75	1.0421	1.0353
Standard Deviation	16.2923	0.7344	0.4287

Table 2 Comparison of the Relative Risk Estimation for AIDS in Epidemiology Month **June 2013**

States	Number of Observed Case	Relative Risk based on SMR Method	Relative Risk based on Poisson-Gamma Model
Perlis	0	0	0.7134
Kedah	4	0.7026	0.7654
Penang	1	0.2192	0.4531
Perak	2	0.292	0.4523
Selangor	14	0.8892	0.9033
Federal Territory of Kuala Lumpur	4	0.8332	0.8799
Federal Territory of Putrajaya	0	0	1.093
Negeri Sembilan	3	1.0191	1.029
Malacca	8	3.3937	2.604
Johor	7	0.7278	0.7614
Pahang	4	0.9258	1.025
Terengganu	10	3.3149	2.472
Kelantan	13	2.885	2.186
Sarawak	4	0.5627	0.681
Sabah	8	0.8582	0.9218
Federal Territory of Labuan	0	0	1.082
Mean	5.125	1.039	1.029
Standard Deviation	4.3283	1.09397	0.6607

rest of the states have relative risk close to 1 which indicates that there is no real difference between people from all those states contracting AIDS compared to overall population in Malaysia.

Based on graphs of SMR method and Poisson-gamma model for both HIV and AIDS cases, it can be seen clearly when there is no observed cases recorded, the SMR will be zero. Meanwhile for Poisson-gamma model, there are values of relative risk when there is no observed cases reported in certain states.

Tables 1 and 2 show the comparison of the relative risk estimation based on SMR method and Poisson-Gamma model for epidemiology month of June for the year 2013 for HIV and AIDS respectively. Based on Table 1, SMR method shows that federal territory of Kuala Lumpur has the highest number of relative risk of contracting HIV similar with relative risk based on Poisson-Gamma model which are 2.6758 and 2.334 respectively. Meanwhile Perlis and Federal Territory of Labuan have the lowest relative risk with a relative risk of zero which is different with the estimation of relative risk by using Poisson-Gamma model whereas the state of Sarawak has the lowest number of relative risk with a relative risk of 0.5706.

In Table 2, Malacca has the highest number of relative risk of AIDS based on SMR method and Poisson-Gamma model which are 3.3937 and 2.604 respectively. This indicates that people from Malacca has the highest probability of contracting AIDS compared to people in the overall population. While Perlis and Federal Territories of Putrajaya and Labuan show a relative risk of zero for AIDS according to relative risk value of SMR

method. In contrast, based on Poisson-Gamma model, the state of Perak has the lowest relative risk of contracting AIDS compared to people in overall population which is 0.4531 value of relative risk. Similar with HIV cases, Poisson-Gamma has the lowest number of mean and standard deviation which are 1.029 and 0.6607 respectively. Thus, Poisson-Gamma has better estimation of relative risk for AIDS cases compared to frequency and SMR methods.

Moreover, test of goodness-of-fit based on mean and standard deviation show that Poisson-Gamma has the lowest value with mean and standard deviation of 1.0353 and 0.4287 respectively. This indicates that Poisson-Gamma has the better estimation of relative risk for HIV cases compared with frequency and SMR methods.

As discussed before, based on the table above when the number of cases recorded for HIV and AIDS are zero, the SMR will give zero relative risk estimation for both cases. This would give problem for the states with zero number of cases such as Perlis and federal territory of Labuan. Nevertheless, Poisson-Gamma model overcomes this drawback by giving the value for relative risk estimation when there is no observed case recorded in certain area.

Maps of the Relative Risk Estimations for HIV and AIDS

In this paper, disease map is used to indicate the high and low risk regions of HIV and AIDS for all states in Malaysia including the three federal territories of Malaysia. Figures 9 and 10 represent the risk maps for HIV and AIDS based on SMR method while Figures 11 and 12 depict the risk maps of HIV and AIDS based on Poisson-Gamma model. Based on the map, the data of relative risks are divided into five different levels of risk which are very low risk, low risk, medium risk, high risk and very high risk with respective interval of [<0.5), [$0.5 - 1.0$), [$1.0 - 1.5$), [$1.5 - 2.0$) and [$2.0 >$) where the darkest region shaded represents the highest risk while the lightest shaded region indicates the lowest risk.

Based on Figure 9, Federal Territories of Kuala Lumpur and Putrajaya represent states with a very high risk. It also can be seen that the states of Kedah, Selangor, Negeri Sembilan, Johor and Kelantan have medium risk for HIV. While, Perlis, Sarawak and

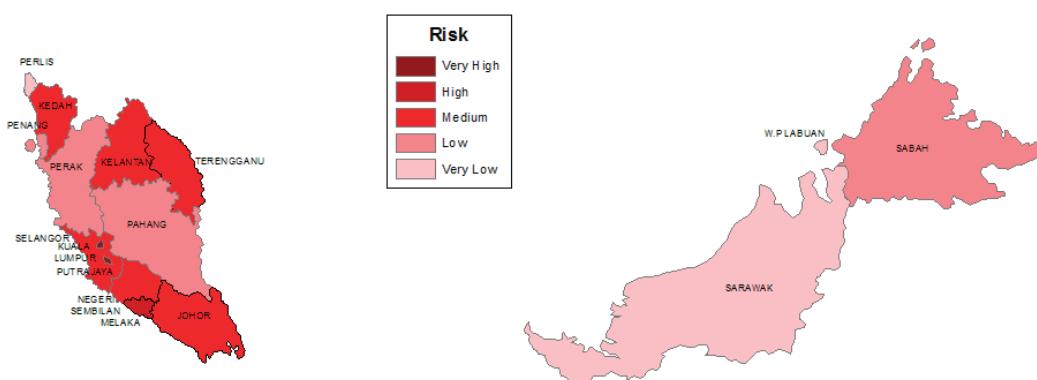


Figure 9 Disease map of estimated relative risk for HIV based on SMR method in epidemiology month June, 2013

Federal Territory of Labuan show a very low risk for HIV. The Poisson-Gamma model map shown in Figure 10 shows that Federal Territory of Kuala Lumpur is the only state with a very high risks for HIV while the rest of the states have medium and low risk of relative risks for HIV.

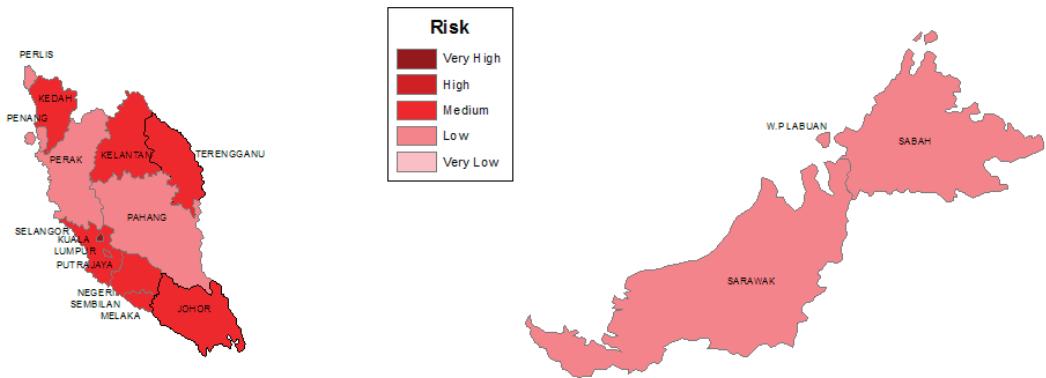


Figure 10 Disease map of estimated relative Risk for HIV based on Poisson-Gamma Model in epidemiology month June, 2013

Referring to Figure 11, it can be seen that the states with very high risk of AIDS include the states of Malacca, Terengganu and Kelantan, while Negeri Sembilan has medium risk of AIDS. This is followed by the state of Kedah, Selangor, Federal Territory of Kuala Lumpur, Johor, Pahang, Sarawak and Sabah which are recorded as low risk for AIDS. The rest of the states are recorded as very low risk for AIDS. Similarly Figure 12, depicts the analysis based on Poisson-Gamma model that shows Malacca, Terengganu and Kelantan have a very high risk for AIDS. Meanwhile, Federal Territory of Putrajaya, Negeri Sembilan, Pahang and Federal Territory of Labuan are recorded as medium risk for AIDS and the other states have low and very low risk for AIDS.

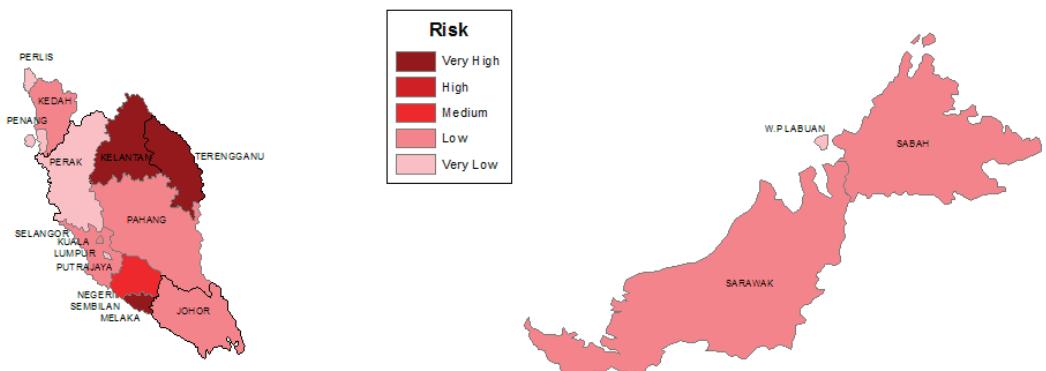


Figure 11 Disease map of estimated relative risk for AIDS based on SMR method in epidemiology month June, 2013

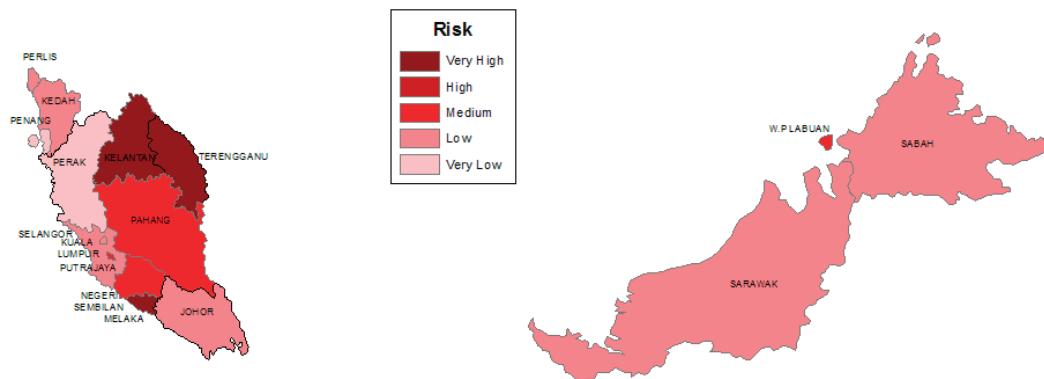


Figure 12 Disease map of estimated relative risk for AIDS based on Poisson-Gamma Model in epidemiology month June, 2013

CONCLUSION

Based on the analysis of HIV cases by using the frequency method both SMR method and Poisson-Gamma model indicate that the lowest value of goodness-of-fit test reveals the best estimation of relative risk. In this case, Poisson-Gamma has the lowest number for the goodness-of-fit test. Furthermore the Poisson-Gamma model can overcome the problem of SMR when the value of observed disease for HIV case is zero for a particular area. Similarly for AIDS cases, the Poisson-Gamma model gives the best estimation compared to the frequency method and the SMR method. As a conclusion, Poisson-Gamma model gives better estimation of relative risk for HIV and AIDS cases compared to the frequency method and the SMR method. This is because Poisson-Gamma model is based on calculation that consists of prior estimation. However, this model does not take into account the geographical location of the region where the model does not cope with spatial correlation (Lawson et.al, 2003). Therefore, the extension of this work should include the estimation of relative risk based on stochastic SIR models for HIV and AIDS as suggested by (Samat and Percy, 2012).

REFERENCES

- Koch, T. (2005). *Cartographies of disease: Maps, mapping, and medicine*. Redlands, CA: ESRI Press.
- A Review of HIV/AIDS Research in Malaysia. (2014). Retrieved November 30, 2014. From <http://www.e-mjm.org/2014/supplement-A/HIV-AIDS>
- Global AIDS Response. (2012). *Global AIDS response country progress report*, 116, pp. 8-13.
- Malaysia UNGASS Country Progress Report. (2012). *AIDS/STD section disease control division*. Ministry of Health, Government of Malaysia. Retrieved September 11, 2014 from <http://www.unaids.org>
- MGD Combat HIV/AIDS, Malaria and Other Diseases. Retrieved December 4, 2014 from http://www.epu.gov.my/c/document_library/get_file?uuid=32a98b8c-af93-4cf2-8450-e447f66335f5&groupId=283545
- National Strategic Pelan on HIV and AIDS 2011-2015. (2011). *The Malaysia HIV Epidemic*.
- A.B.Lawson, W.J.Browne, and C.L.Vidal Rodeiro. (2003). *Disease mapping with WinBUGS and*

- MLwiN. England: John Wiley & Sons.
- A.B.Lawson. (2006). *Statistical method in spatial epidemiology* (2nd Ed.). England: John Wiley and Sons.
- N.A.Samat and D.F.Percy. (2008). Standardize mortality and morbidity ratios and their application to Dengue disease mapping in Malaysia, *Proceeding of the Salford*.
- J.L.Meza. (2002). Emperical bayes estimation smoothing of relative risk in disease mapping. *Journal of Statistical Planning and Inference*, 112 pp.43 – 62.
- Mark, E.G and David, A.V. (2010). *Basic Bayesian Methods*, volume 404 of Lecturer Notes. Totowa, New Jersey, United States pp. 319-338.
- N.A.Samat and D.F.Percy. (2012). Vector-borne infectious disease mapping with stochastic difference equations: an analysis of dengue disease in Malaysia *Journal of Applied Statistics*, vol.39(9), pp. 2029 – 2046. DOI: 10.1080/02664763.2012.700450.