INTRODUCTION

Background

There is a long history of geographical surveillance of disease by publishing disease atlases. If there are areas with exceptionally high rates, they may give the clues to the etiology of the disease, indicating areas where health care needs to be improved, or to be targeted for preventive measures. Implementing scan statistic, we can do the surveillance of particular disease on particular area. This study applied retrospective space-time permutation scan statistic to construct measles disease surveillance in West Java.

According to the World Health Organization (WHO), vaccination rate has been high enough to make measles relatively uncommon in developed countries, but in developing countries it is still common. Globally, measles deaths went down 60 percent, from an estimated 873,000 deaths in 1999 to 345,000 in 2005. Africa has been the most success area, with annual measles deaths falling by 75 percent in just 5 years, from an estimated 506,000 to 126,000 (UNICEF World Press Release in Wikipedia, 2007).

The ability in detecting measles outbreaks early is important in order to minimize morbidity and mortality through timely implementation of measles prevention and control measures. A scan statistic can be used widely in any field to recognize any significant hotspot in terms to find any spatial areas that have elevated risk than their surroundings.

In space-time, the scan statistic can provide early warning of disease outbreaks and can monitor their spatial spread. This study using a retrospective space-time permutation scan statistic for detecting measles disease hotspot in West Java that utilized only case numbers, with no need for population-at-risk data, where these data are very difficult or irrelevant to be obtained.

This method was applied on annual data of measles disease cases in West Java in 2003 and 2006 periods.

Objective

The objective of this study is to detect measles disease hotspots in West Java in order to reveal its outbreaks using historical data of Podes 2003 and 2006, where the detected hotspots indicated to be taken care due to the condition of health care and preventive measures action must be conducted.

DEFINITIONS

Measles

Measles, also called rubeola, is a highly contagious but rare respiratory infection that is caused by a virus. It causes a total-body skin rash and flu-like symptoms, including a fever, cough, and runny nose.

Measles is spread through respiration (contact with fluids from an infected person's nose and mouth, either directly or through aerosol transmission). About 90% of people without immunity sharing a house with an infected person will catch it. Airborne precautions should be taken for all suspected cases of measles. The incubation period usually lasts for 4–12 days (during which there are no symptoms). Infected people remain contagious from the appearance of the first symptoms until 3–5 days after the rash appears (Wikipedia, 2007).

Hotspot

Hotspot is defined as something unusual, anomaly, aberration, outbreak, elevated cluster, critical area, etc (Patil and Taillie, 2004). Hotspot clusters were generated by setting the relative risk in some counties to be larger than one (Song and Kulldorff, 2003).

Hotspots are locations or regions that have consistently high levels of disease and may have characteristics unlike those of surrounding areas (Haran, Molineros, & Patil, 2006 in Septiani, 2006).

Scan Statistic

First studied by Naus in 1965, the scan statistic is an elegant way to solve problems of multiple testing when there are closely overlapping spatial areas and/or time intervals being evaluated. Temporal, spatial, and space–time scan statistic are now commonly used for disease cluster detection and evaluation, for a wide variety of diseases.

The basic idea is that there is a scanning window that moves across space and/or time. For each location and size of the window, the number of observed and expected cases is counted. Among these, the most “unusual” excess of observed cases is noted. The statistical significance of this cluster is then evaluated taking into account the multiple testing stemming from the many potential cluster locations and sizes evaluated (Kulldorff et al, 2005).
Space-Time Permutation Scan Statistic

The space-time permutation scan statistic utilizes thousands or millions of overlapping cylinders to define the scanning window, each being a possible candidate for an outbreak. The circular base represents the geographical area of the potential outbreak.

A typical approach is to first iterate over a finite number geographical grid points and then gradually increase the circle radius from zero to some maximum value defined by the user, iterating over the areas in the order in which they enter the circle. In this way, both small and large circles are considered, all of which overlap with many other circles.

The height of the cylinder represents the temporal length. Only alive clusters, clusters that reach all the way to current time as defined by the study period end date, are then searched for. This study using the retrospective space-time scan statistic since the research only done once in West Java Province in 2003 and 2006.

The space-time scan statistic may be used for either a single retrospective analysis or for time-periodic prospective surveillance. In a retrospective analysis, the analysis is done only once for a fixed geographical region and a fixed study period, evaluating both ‘alive’ clusters, lasting until the study period and date, as well as ‘historic clusters’ that ceased to exist before the study period end date. The prospective option is used for the early detection of disease outbreaks, when analyses are repeated every day, week, month or year. Only alive clusters, clusters that reach all the way to current time as defined by the study period end date, are then searched for.

What is new with the space–time permutation scan statistic is the probability model. Since we do not have population-at-risk data, the expected must be calculated using only the cases. Suppose we have annually case counts for z areas, where \( c_{zd} \) is the observed number of cases in area \( z \) during time \( d \). The total number of observed cases \( (C) \) is

\[
C = \sum_z \sum_d c_{zd} \quad \ldots (1)
\]

For each area and time, we calculate the expected number of cases \( \mu_{zd} \) conditioning on the observed marginals:

\[
\mu_{zd} = \frac{1}{C} \left( \sum_z c_{zd} \right) \left( \sum_d c_{zd} \right) \quad \ldots (2)
\]

In words, this is the proportion of all cases that occurred in area \( z \) times the total number of cases during time \( d \). The expected number of cases \( \mu_A \) in a particular cylinder \( A \) is the summation of these expectations over all the area-time within that cylinder:

\[
\mu_A = \sum_{(z,d) \in A} \mu_{zd} \quad \ldots (3)
\]

The underlying assumption when calculating these expected numbers is that the probability of a case being in area \( z \), given that it was observed on time \( d \), is the same for all times \( d \).

Let \( c_A \) be the observed number of cases in the cylinder. Conditioned on the marginals, and when there is no space–time interaction, \( c_A \) is distributed according to the hypergeometric distribution with mean \( \mu_A \) and probability function (Kulldorff, 2005):

\[
P(C_A) = \frac{\left( \sum_{z \in A} c_{zd} \right) \left( C - \sum_{z \in A} c_{zd} \right)}{\sum_{d \in A} c_{zd} - C} \quad \ldots (4)
\]

The relative risk, any non-negative number, representing how much more common an event (case) is in this location compared to the baseline. Setting a value of one is equivalent of not doing any adjustments. A value of greater than one is used to adjust for an increased risk and a value of less than one to adjust for lower risk. A relative risk of zero is used to adjust for missing data for that particular location. Relative risk is calculated simply by dividing observed number to its expected (Kulldorff, 2006).

Relative Risk

For each location and size of the scanning window, the alternative hypothesis is that there is an elevated risk within the window as compared to outside. When both \( \sum_{z \in A} c_{zd} \) and \( \sum_{d \in A} c_{zd} \) are small compared to \( C \), \( c_A \) is approximately Poisson distributed with mean \( \mu_A \). Based on this approximation, we use the Likelihood Ratio Test.
Poisson generalized likelihood ratio (GLR) as a measure of the evidence that cylinder A contains an outbreak:

\[
\left( \frac{c_A}{\mu_A} \right)^{c_A} \left( \frac{C-c_A}{C-\mu_A} \right)^{C-c_A} I(\cdot) \ldots (5)
\]

In words, this is the observed divided by the expected to the power of the observed inside the cylinder, multiplied by the observed divided by the expected to the power of the observed outside the cylinder. I(\cdot) is an indicator function. When the objective of this study is set to scan only for clusters with high rates, I(\cdot) is equal to 1 when the window has more cases than expected under the null-hypothesis, and 0 otherwise. The opposite is true when the objective of the study is set to scan only for clusters with low rates. When the study scans for clusters with either high or low rates, then I(\cdot)=1 for all windows.

The likelihood function is maximized over all window locations and sizes, and the one with the maximum likelihood constitutes the most likely cluster. This is the cluster that is least likely to have occurred by chance. The likelihood ratio for this window constitutes the maximum likelihood ratio test statistic (Kulldorff, 2006).

Monte Carlo Hypothesis Testing

Once the value of test statistic calculated, it is easy to do the inference. We cannot expect to find the distribution of test statistic in closed analytical form. Instead we rely on Monte Carlo hypothesis testing.

With Monte Carlo hypothesis test, the statistical significance of an observed test statistic calculated from a set of data that assessed by comparing it with a distribution obtained by generating alternative set of data from some assumed model. If the assumed model implies that all data orderings are equally likely then this amounts to a randomization test with random sampling of the randomization distribution.

Monte Carlo hypothesis testing using in scan statistic consist of four steps procedure:

1. Calculate the value of the test statistic for the real data.
2. Create a large number of random data sets under the null hypothesis.
3. Calculate the value of test statistic for each random replication.
4. Sort the value of test statistic in step 3 (simulation data), then compare it with the value of test statistic in step 1 (real data). If value of test statistic for real data ranked in the highest α percent of value of test statistic for simulation data, then reject the null hypothesis at a percent significance level.

On the other hand, p-value can be denoted by \( p-value = \frac{rank}{1 + \text{number of simulation generated}} \) (Kulldorff, 1999).

MATERIALS AND METHODS

Data Sources

This study used secondary data of measles disease cases in West Java obtained from 2003 and 2006 of Potensi Desa (Podes) surveys by Statistics Indonesia.

Methods

During the completion of this study, several steps were done as mentioned below:

2. Generating any possible scan windows with the maximum spatial cluster size was 10% of population and maximum temporal cluster size was 1 year.
3. Calculating the relative risk of each scanned window mentioned on step 2, and then discarding the window with relative risk less than 1.
4. Calculating the likelihood function value of each scanned window obtained at step 3.
5. Obtaining the statistical significant value for each hotspot candidate using Monte Carlo simulation with 9999 replications.
6. Interpreting the results and presenting the hotspot area on West Java map using MapInfo Professional 7.8 SCP.

This study use SaTScan 7.0 to perform step 2 to 5.

The size of spatial and temporal window can vary up to 50% of the total study area or time. Subjectively, the value of the spatial window was set up to 10% of the total West Java area and the value of temporal window was set up to 50% of the total two-year period of study, 2003 and 2006. Before the value of the spatial window was set to 10%, this research had tried the different maximum spatial windows, of 5%, 15%, 25%, and 50%. The value of maximum temporal window cannot be changed, because this research only involved two points of time.

The maximum spatial and temporal windows could not exceed 50% of the total...