CHAPTER 3

Study results

The following are manuscripts:

(a) Modeling Incidence Rates of Terrorism Injuries in Southern Thailand, publisher is Chiang Mai Journal of Science, Accepted for publish in volume 40, year 2013.

(b) Graphing Incidence Rates over Regions using R and Google Earth, publisher is Journal of Map and Geography Libraries, volume 7, page 211-219, published year 2011.

Chiang Mai J. Sci. 2013; 40(X) : 1-7
http://it.science.cmu.ac.th/ejournal/
Contributed Paper

Modeling Incidence Rates of Terrorism Injuries in Southern Thailand

Sumpunt Khongmark*[a] and Metta Kuning [b]
[a] Faculty of Communication Sciences, Prince of Songkla University, Pattani 94000, Thailand.
[b] Department of Mathematics and Computer Sciences, Faculty of Science and Technology, Prince of Songkla University, Pattani 94000, Thailand.
*Author for correspondence; e-mail: ksumpunt@bunga.pn.psu.ac.th.

Received: xx xxx xxxx
Accepted: xx xxx xxxx

ABSTRACT

We consider methods for modeling and comparing incidence rates of adverse events that vary over space, time and demographic characteristics of subjects including gender and age group. We assume that the adverse events and their corresponding population-at-risk denominators are aggregated into a contingency table whose dimensions correspond to the spatial, temporal and demographic factors. We compare regression models for Poisson and negative binomial generalized linear models with zero-corrected log-transformed linear models. These methods are applied to the terrorism events in regions of Southern Thailand that occurred over the period from 2004 to 2010.

Keywords: Statistical comparison, bubble charts, confidence intervals, Poisson regression model, negative binomial regression model, log-normal model.

1. INTRODUCTION

The threat of death and injury from terrorism has been a serious ongoing problem for residents in the three southernmost provinces of Thailand since January, 2004. Our study focused on the non-Muslim population (mostly Buddhist) because their incidence rates are more than double those of Muslim residents, and also because the two groups have quite different spatial-temporal incidence patterns. It is important to use the most appropriate statistical methods for describing and, if possible, forecasting the incidence rates of injuries from terrorism in the target area.

In the next section we describe three statistical models for incidence rates. The first is the Poisson distribution, the second is the more general negative binomial distribution that allows for over-dispersion due to clustering, and the third assumes that the incidence rates follow a log-normal distribution. Each model provides adjusted incidence rates and confidence intervals for comparing subgroups with an overall mean.

1. Data and Variables

The data used in this study were collected by the Deep South Coordination Centre [1]. The populations were obtained from the 2000 population and housing census of Thailand. We focused on incidence rates per 100,000 population for civilian victims of terrorism classified by gender, age group (<25, 25-44 and 45 or more), district of residence and year (six years from 2004 to 2010 inclusive). The target area comprises 37 districts in southern Thailand. Since the populations in these districts differ substantially, we initially created 23 regions with more equal populations (ranging from 54,039 to 154,634) by aggregating some adjoining districts in the same province.

Preliminary analysis of annual injury rates by region and year (Figure 2) show quite different patterns for Muslim and non-Muslim residents, which makes it difficult to accommodate the entire data set using a single model. Moreover, the incidence rates for non-Muslims are more than double those for Muslim residents, so we restricted our analytic study to non-Muslim residents. But since some of the 23 regions contain quite small non-Muslim populations, it was necessary to further combine some adjoining regions to ensure that population sizes were sufficient to provide valid statistical estimates of relative risks. The 23 regions were thus reduced to 15 by further aggregation of adjoining regions. These divisions are indicated in Table 1.

Table 1 Regions used in analysis of victim injury incidence in Southern Thailand.

Province	Region ID: Districts	Non-Muslim	Population
Songkhla	1: Chana / Thepha	62,621	156,799
	2: Saba Yoi / Na Thawi	62,236	110,507
Pattani	3: Mueang Pattani	41,122	108,271
	4: Kok Pho / Mae Lan	34,812	75,628
	5: Nong Chik / Mayo / Thung Yang Daeng / Kapho / Yaring / Yarang	19,878	303,898
	6: Panare / Sai Buri / Mai Kaen	19,717	108,188
Yala	7: Mueang Yala	75,291	154,634
	8: Betong / Than To	36,706	68,193
	9: Bannang Sata / Krong Pinang / Yaha / Kabang / Raman	17,845	192,710
Narathiwat	10: Mueang Narathiwat	31,950	104,615
	11: Tak Bai	15,376	61,157
	12: Bacho / Yi-ngo / Rueso / Ra-ngae / Si Sakhon / Chanae	23,560	278,922
	13: Sukhirin / Waeng	11,624	63,765
	14: Su-ngai Padi / Cho-airong	13,563	89,251
	15: Su-ngai Kolok	23,323	64,640

3. Statistical Methods

We considered Poisson, negative binomial, and linear regression (based on log-transformed data) models using four factors: gender, age group (AgeGrp) with three levels (under 25 which are studying ages, 25-44 are working ages, and 45 or more are seniors)[7], year (7 levels) and region (15 levels). For brevity the linear model for log-transformed data is referred to simply as the "log-normal" model.

The formulation for the Poisson and negative binomial models is based on two sets of predictors comprising the gender-age group combination, and the year-region combination, respectively, that is,

$$\log (l/P) = \text{factor}(\text{gender.AgeGrp}) + \text{factor}(\text{year.Region}),$$

where l is the mean incidence rate and P is the person-years of exposure to risk per 100,000 resident population for each data cell. The Poisson model has only the single parameter l, and the variance is also l. The negative binomial model generalizes the Poisson model through an additional dispersion parameter q. It has the same mean as the Poisson distribution, but its variance is l(1+l/q). The formulation for the log-normal model is

log (y) = factor(gender.AgeGrp) + factor(year.Region),

where $y = \log(n^+/P)$, and n^+ is defined as the number of injuries in a cell if this number is positive, and 0.5 otherwise. When computing adjusted incidence rates based on the negative binomial and log-normal models, a further correction is needed to ensure that each model gives adjusted incidence rates that preserve the total number of injuries observed. This correction can be implemented simply by rescaling the adjusted incidence rates using the method described by Kongchouy and Sampantarak [2].

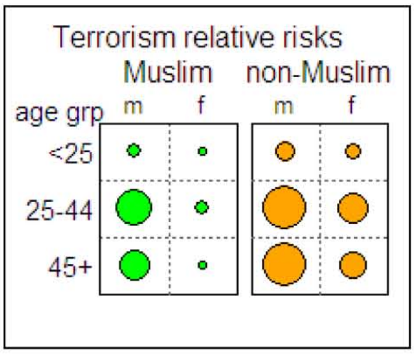
Plots of residuals against normal quantiles are used to assess how well a model fits the data. For Poisson and negative binomial models deviance residuals based on the likelihood are recommended, whereas for linear regression studentized residuals are recommended because simple differences between the observed outcomes and their fitted values tend to overestimate the variance of these residuals (see, for example, section 6.3 of Venables and Ripley [3]). In each case the plotted residuals should follow a straight line on the plot corresponding to normal quantiles. For the Poisson and negative binomial models, there is the added requirement that the deviance should have a chi-squared distribution with the number of degrees of freedom matching that from the model.

Since differences between incidence rates for levels of each factor and the overall mean are of interest when comparing risks, sum contrasts are used for constructing confidence intervals ([3], chapter 6).

We used the free and open-source basic R program [4, 5] for all statistical and graphical analysis.

Preliminary analysis

The overall annual terrorism victim injury rates per 100,000 were 48 for Muslim residents and 121 for non-Muslim residents. Figure 1 shows a table giving incidence rates by gender and age group for Muslim and non-Muslim residents, with a bubble chart [6] to facilitate graphical comparison of these risks. It shows that non-Muslim residents have higher overall risks than Muslim residents, and males have higher risks than females in both groups, particularly at ages 25 or more.



Terrorism relative risks

Age group	Muslim		non-Muslim	
	m	f	m	f
<25	27	8	49	31
25-44	157	24	229	110
45+	109	12	225	89

Figure 1: Comparison of risks of injury from terrorism in southern Thailand for Muslim and non-Muslim residents by gender and age group.

Similar plots for the comparison with respect to region and year are shown in Figure 2. This graph shows clearly that rural regions are more dangerous for non-Muslim residents, with higher rates in rural regions of Pattani, Yala and Narathiwat provinces. Year 2007 had higher injury incidence rates for both Muslim and non-Muslim residents. For Muslim residents, Tak Bai had the greatest number because of what happened there in October 2004, which sparked the ensuing violence, but after 2004 this region had relatively low annual incidence rates.

Given that the non-Muslim incidence rates were much higher than those for Muslim residents, further analysis of these data is restricted to the non-Muslim rates.

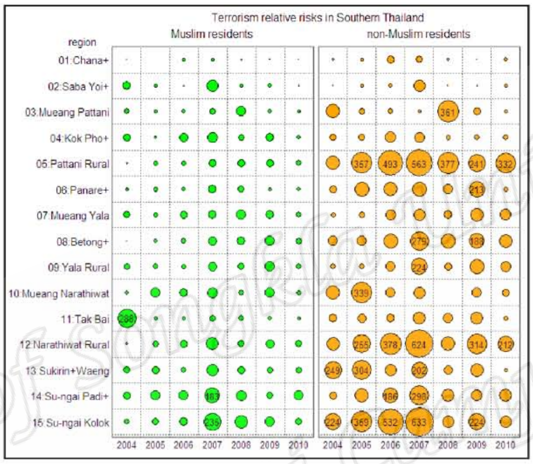


Figure 2: Bubble chart comparing risks of injury from terrorism in Southern Thailand for Muslim and non-Muslim residents by region and year.

4. Results

In one region, there were no injuries recorded (Su-ngai Padi/Cho-airong in 2008), and Poisson and negative binomial models fail to converge in such cases. To overcome this problem, we recoded the year to 2008 for one of the five women aged 25-44 who were victims in this region in 2007, and then refitted the model. Figure 3 shows residuals plots for each of the three models. The Poisson model (left panel) is not satisfactory because the residual deviance (982.9) is more than twice its degrees of freedom (445). The negative binomial model with q = 4 gives a better fit (middle panel), and the chi-squared test indicates that the fit is statistically acceptable. For the log-normal model (right panel), the fit is extremely good.

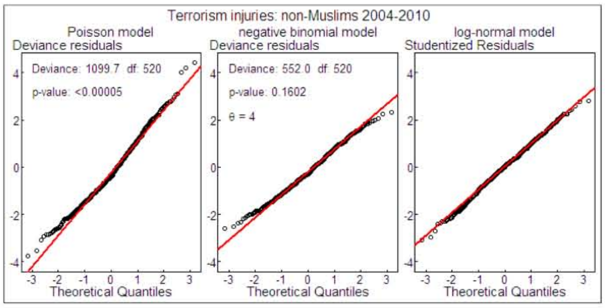


Figure 3: Residuals plots for Poisson, negative binomial and log-normal models.

Figure 4 shows adjusted incidence rates with 95% confidence intervals based on the negative binomial model, with corresponding adjusted incidence rates for comparing the region-year effects based on the log-normal model superimposed.

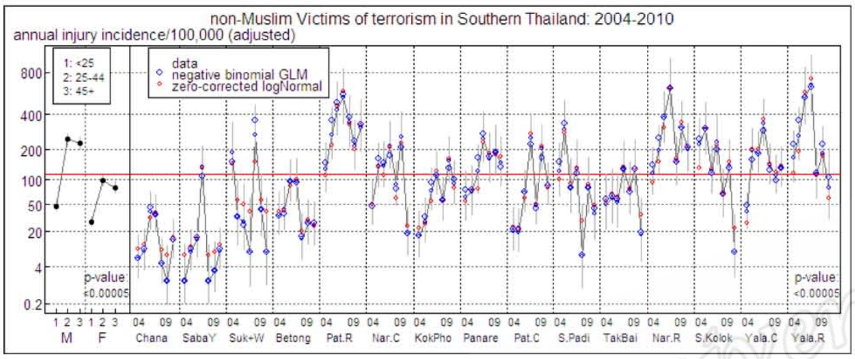


Figure 4: Adjusted incidence rates with confidence intervals from negative binomial model.

Although the adjusted rates are scaled to have the same overall mean, the log-normal model gives higher estimates than the negative binomial model when the rates themselves are lower. Figure 5 compares these rates for the two models. There were some outliers due to the lower incidence rates given by the log-normal model (left panel).

The two models give different confidence intervals. Since the design is balanced, the log-normal model gives confidence intervals of equal size when plotted on a log scale. However, for the negative binomial model the confidence intervals vary according to the number of observed events. In the right panel of Figure 4 confidence intervals for the log-normal model cover a wider range than those for the negative binomial model, but most of them are shorter.

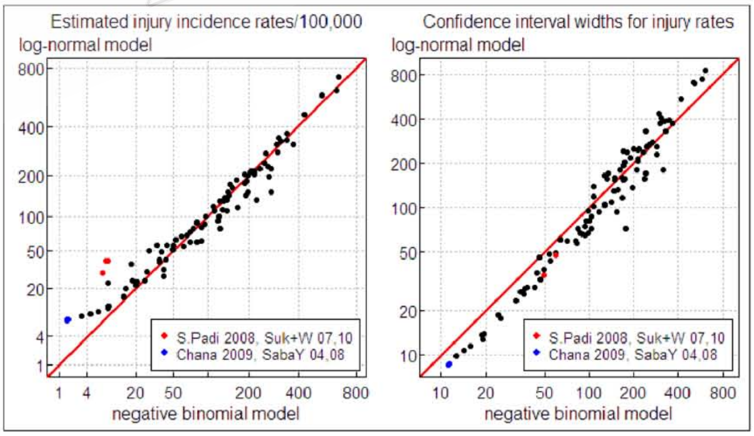


Figure 5: Plots of incidence rates comparing negative binomial and log-normal models (left panel) and corresponding confidence interval widths (right panel).

5. Discussion

Consider non-Muslim residents in the bubble chart of Figure 2. Incidence rates in the rural regions increased from year 2004 to 2010. Year 2007 had highest incidence rates. Moreover, the other regions increased as well as the rural areas but with lower increases. In Sukirin and Waeng districts of Narathiwat province, the incidence rates decreased until year 2007, and then rose sharply. But the opposite pattern occurred in rural districts of Narathiwat and Yala provinces, which the incidence rates increased at first and then suddenly decreased in year 2008. Other regions of note are Saba Yoi/Na Thawi and Su-ngai Padi/Cho-airong, where injury rates decreased to almost zero in year 2008. (Su-ngai Padi/Cho-airong had no adverse events at all reported in that year.) These patterns illustrate the essential unpredictability of these events.

The region-year incidence rates in Figure 4 show three clear patterns. The first pattern is a rise followed by a fall, and is followed in Chana/Thepha, rural districts of Pattani, and Mueang Yala). The second pattern is a rise followed by a fall and another rise, and is followed in SabaYoi/NaThawi, Betong/Than To, Mueang Narathiwat, Kok Pho/Mae Lan, Panare/Saiburi/MaiKaen, Mueang Pattani, Su-ngai Padi/Cho-airong, Tak Bai, Narathiwat rural districts, Su-ngai Kolok and Yala rural districts. The last pattern is different from the others where the incidence rates fell, then rose, and fell again. This pattern is followed in Sukirin and Waeng districts in Narathiwat province.

These varying patterns of incidence rates over the whole region indicate an important feature of guerilla warfare practiced by terrorists. If such incidence rates could be described simply by additive linear models, the task of authorities attempting to counter terrorism would be much easier, because an additive model is largely predictable, where all regions have the same time series pattern. However, such additive models do not fit these data anywhere near as well as the model we have fitted, which involves combining the 7 years and 15 regions into a single factor with 105 levels.

From Figure 3, fitting a model for incidence rates over three provinces of Thailand including the four eastern districts of Songkhla province, the Poisson regression model fits poorly compared to the other two models. However, the results from negative binomial and log-normal are similar, different only in regions with small incidence rates. In general, the choice of model will depend on (a) the number of cells into what the data are aggregated and (b) the number of events in these cells. However for this data set, log-normal are fitted better than the other two models. Further study could focus on comparison of these models based on data from other studies.

ACKNOWLEDGEMENT

We thank Professor Don McNeil for supervision.

REFERENCES

[1] Deep South Coordination Center, Available (in Thai) from: http://medipe2.psu.ac.th/~dscc/web.

[2] Kongchouy N. and Sampantarak U., Confidence intervals for adjusted proportions using logistic regression., Modern Applied Science., 2010; **4(6):** 2-6.

[3] Venables W.N. and Ripley B.D., Modern Applied Statistics with S., 4th Edition. Springer, 2002.

[4] R Development Core Team, R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria, Available

from: http://www.R-project.org.

[5] Venables W. And Smith D., The R Development Core Team. An introduction to R: notes on R: a programming environment for data analysis and graphics version 2.12.2 (2011-02-25). Available from: http://cran.r-project.org/doc/manuals/R-intro.pdf.

[6] Wikipedia, The free encyclopedia, Bubble chart, 2011,http://en.wikipedia.org/wiki/Bubble_chart.

[7] Komolmalai W., Kuning M. and McNeil D., Muslim Victims of Terrorism Violence in Southern Thailand, International Journal of Business and Social Science., 2012; **3(13):** 114-119.

[8] Marohapout P., Choonpradub C. and Kuning M., Terrorism Risk Modeling in Southern Border Provinces o Thailand during 2004 to 2005., Songklanakarin Journal of Social Sciences & Humanities., 2009; **15(6):** 883-895.

Journal of Map And Geography Libraries, 7:211–219, 2011

ISSN: 1542-0353 print / 1542-0361 online
DOI: 10.1080/15420353.2011.566843

Routledge
Taylor & Francis Group

Graphing Incidence Rates over Regions using R and Google Earth

SUMPUNT KHONGMARK, METTA KUNING,
and ATTACHAI UERANANTASUN
Prince of Songkla University, Pattani, Thailand

A method for modeling and graphically comparing incidence rates of adverse events that vary over geographical regions is shown and discussed. To achieve our goal, we used a statistical model to compare incidence rates and showed how informative three-dimensional graphs of such incidence rates can be created dynamically using R and interactively controlled using Google Earth with Keyhole Markup Language (KML). These methods are applied to the terrorism events in regions of Southern Thailand that occurred from 2004 to 2009.

KEYWORDS *Geographical mapping, three-dimensional plots, dynamic KML, statistical comparison, confidence intervals, R software, Google Earth*

INTRODUCTION

Displaying a quantity that varies by geographic region presents a cartographic challenge because two spatial coordinates are already needed to specify latitude and longitude, so a place must be found to show the statistical variation. Ideally, this variation should include not just the data variation but also the extent to which differences between values at different locations are real, or simply a result of chance variation. A method is thus needed for displaying both the data and the reality of differences on a graphic information system (GIS) map. The purpose of this paper is to provide an appropriate

We would like to thank our colleagues Don McNeil, Somporn Chuai-aree, and Patarapan Odton for their help and suggestions.
Address correspondence to Sumpunt Khongmark, Department of Communication Sciences, Prince of Songkla University, Pattani Campus, 181 Chareonpradit Road, Rusamilae, Pattani, Thailand 94000. E-mail: ksumpunt@bunga.pn.psu.ac.th

method of doing this using software that is relatively easy to use and freely available.

Our method involves first fitting an appropriate linear statistical model, and then combining the graphs produced by this model as screen overlays in a GIS with dynamic three-dimensional histograms produced by Google Earth software, where colors are used to distinguish real differences. An analysis and model-fitting part can be achieved by using R programming language, widely used in statistics research due to its structure. Users with little programming background can easily start to write their own codes in R. In addition, it is powerful enough to provide a broad range of analyzing commands, and it can work with sizeable relational databases. In terms of data visualization, R programming can produce a graphic output using a series of graphics functions from its library. These outputs can be presented as graphs and images and transferred for use in Google Earth.

All software used to develop the system is freely downloadable from the Internet. The system uses Keyhole Markup Language (KML) and the basic Google Earth program. The KML source code is created dynamically using R (R Development Core Team 2010). We also used R to fit the statistical model and to create Figures 1 and 2 (below).

In the following sections, the data and statistical method are briefly described. Further sections describe how a GIS map can be created by using KML source code in basic Google Earth software. In the Discussion section we share some extensions of the method to more general dynamic graphics applications.

ILLUSTRATIVE DATA: VICTIMS OF VIOLENCE IN SOUTHERN THAILAND

For our model, we considered incidence rates per hundred thousand population of terrorism events classified by gender, age-group (<25, 25–44, or 45 or more), district of residence (15 levels as shown in Table 1), and year (six levels from 2004 to 2009 inclusive). Each adverse outcome corresponds to a civilian victim suffering injury or death as a result of a defined violent terrorism event in the target area. In this case, the target area is defined as all districts in the three southernmost provinces of Thailand (Pattani, Yala, and Naratiwat) as well as four districts on the eastern side of Songkla Province. These data were retrieved from a database maintained by the Deep South Coordination Centre, Thailand (http://medipe2.psu.ac.th/~dscc/). The population denominators were obtained from the 2000 population and housing census of Thailand.

Because the overall victim incidence rates for Muslims were very much lower than those for other residents of the terrorist target area, we restricted the study to non-Muslim victims. For purposes of analysis, we first aggregated the thirty-seven districts within the target area into twenty-three regions with

TABLE 1 Regions used in analysis of victim violence incidence in Deep South of Thailand

Province	RegionID: Districts	Non-Muslim Pop
Songkla	1: Chana/TePa	62,621
	2: SabaYoi/NaTawi	62,236
Pattani	3: Pattani City	41,122
	4: KoPho/MaeLan	34,812
	5: NongChik/Mayo/ThungYangDang/Kapo/Yaring/Yarang	19,878
	6: Panare/Saiburi/MaiKaen	19,717
Yala	7: Yala City	75,291
	8: Betong/ThanTo	36,706
	9: BannangSata/KrongPinang/Yaha/Kabang/Raman	17,845
Naratiwat	10: Naratiwat City	31,950
	11: TakBai	15,376
	12: Bacho/YinGo/Rueso/Rangae/SaSikon/Chanae	23,560
	13: Sukirin/Waeng	11,624
	14: SungaiPadi/Cho-airong	13,563
	15: SungaiKolok	23,323

populations ranging from 54,039 to 154,634. However, some of these regions contained fewer than 5,000 non-Muslim residents, so to provide acceptably stable estimates of incidence rates, we further aggregated the regions with smaller non-Muslim populations and thus reduced the number of regions to fifteen as listed in Table 1.

STATISTICAL MODEL

A regression model for incidence rates was established using the four factors noted above. To remove skewedness and thus satisfy statistical assumptions, the incidence rates were log-transformed. There were 109 zeroes in the 540 cells in the contingency table of victim counts; the incidence rates in these cells were inflated by replacing these zeros by 0.5 to enable log-transformed incidence rates to be calculated for all cells.

To assess the risks for subgroups of residents, we fitted an additive model with three factors: age-group, year, and gender-region, using sum contrasts to obtain confidence intervals for comparing incidence rates for each level of each factor, with the overall mean after adjusting for other factors (see Venables and Ripley 2002, chap. 6).

Figure 1 gives confidence interval plots for the incidence rates with respect to gender-region after adjusting for age-group and year. The incidence rates for females was substantially lower than those for males in all regions except Naratiwat City and Sukirin/Waeng.

GIS THEMATIC MAP

Figure 2 shows a simple thematic map of the target area containing a pair of colored square boxes for the two sexes in each of the fifteen regions. The

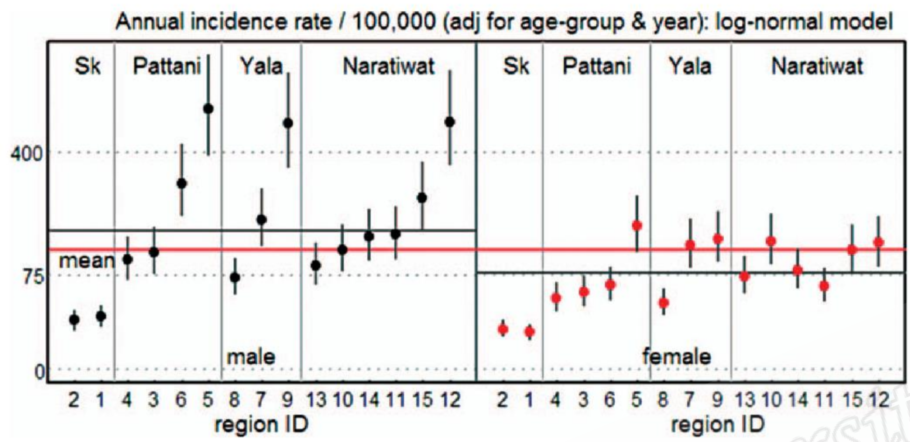


FIGURE 1 Estimates of annual incidence rates per 100,000 population of injury to civilian non-Muslim residents of the terrorism target area in Southern Thailand, classified by gender and region after adjusting for year (2004–2009) and age group (<15, 15–44, 45+). The horizontal red line denotes the overall mean incidence rate and the grey horizontal lines denote the rates for males and females, respectively. The vertical lines denote 95% confidence intervals for differences between the incidence rates and the overall mean.

color codes are based on the locations of the confidence intervals in Figure 1 with respect to the overall mean incidence rate. A box is colored red, blue, or green according to whether its corresponding confidence interval is entirely above the mean, entirely below the mean, or crosses the mean.

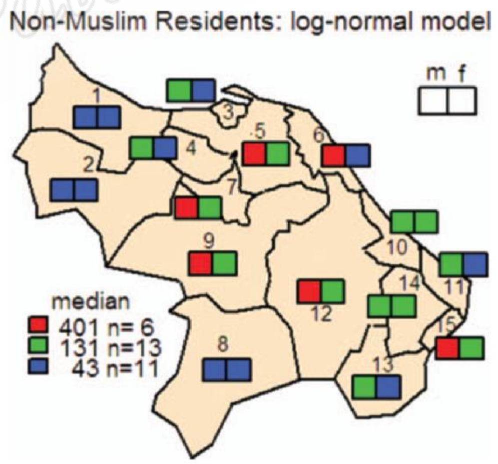


FIGURE 2 Simple thematic map of the 15 regions in the Southern Thailand terrorism target area, using three colors to classify the terrorism risk for males (left boxes) and females (right boxes) as above average (red), below average (blue), or not evidently different from average (green). The numbers in the legend denote the median incidence rates per 100,000 populations in each of the three groups.

Although colors of the boxes on the thematic map are useful for classifying the risk with respect to the overall mean, they say little about the magnitude of the risk. However, thematic maps can be improved to incorporate these magnitudes by replacing the colored boxes with colored bar charts. This alternative has the drawback of taking up valuable space on the map. In the next section we consider how to augment the thematic plot by moving into three dimensions.

INTERACTIVE GIS GRAPHICS SYSTEM

In this section we describe how to create an interactive graphics system for displaying and controlling the position of graphs and maps similar to those shown in Figures 1 and 2. This system also creates three-dimensional histograms as shown in Figure 3. (See this sample of our work and how it can be interacted with a viewer at http://scitech.sat.psu.ac.th/GEDSV/).

KML is an extended development of XML, a readable text file used by Google Earth for manipulating and displaying geographic information. Although R does not yet have a built-in library or function to create a KML source code file, it can handle text files and perform database functions, so additional software is not essential. However, a library for R that can facilitate KML source code creation is freely available on the Internet (XML for S-PLUS Core Team 2009).

For our application, the R source code reads the following data tables:

a. A contingency table of events classified by gender, age-group, region, and year;
b. a table of population denominators classified by gender, age-group and region;
c. shape files containing the longitudes and latitudes of the boundary arcs for each region; and
d. longitude and latitude coordinates specifying the locations within each region for placing the three-dimensional histograms.

The first part of the R program fits the models, computes the confidence intervals, and creates the graphs corresponding to the confidence interval plot and thematic map shown in Figures 1 and 2, which are stored as JPG image files. The second part of the R program dynamically creates the KML text file, comprising header and body components.

The KML header contains style definitions, including the solid and transparent colors used in the ground-overlay maps, as well as the method for executing the rollovers from solid to transparent colors when a region or extruded polygon is cursor-selected. These extruded polygons have specified altitudes (in meters) and are necessarily anchored at ground level (or sea level, or ocean-floor level, if preferred), but it is not possible in KML to

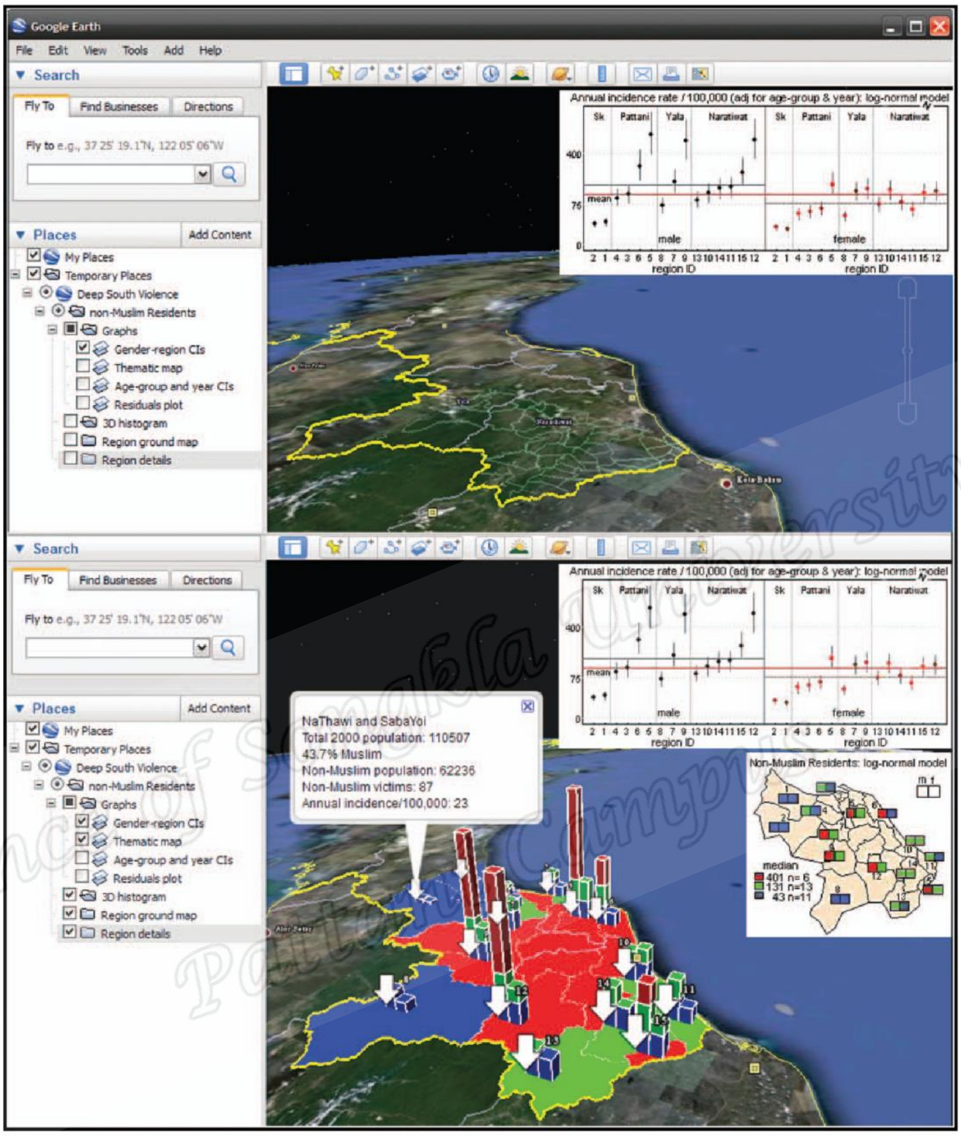


FIGURE 3 Google Earth displays of GIS map of terrorism incidence rates to non-Muslim residents in regions of Southern Thailand. The upper panel shows the location of the target area in the Malay Peninsula viewed from a specified location above the Earth's surface, with the graphical user interface in the sidebar to the left of the map that appears when a user double-clicks on the Google Earth icon, and with the graph in Figure 1 appearing as a screen overlay. The lower panel shows additional features that appear when the user selects further menu options, including the simple thematic map in Figure 2, and corresponding color-coded extruded polygons with ground overlays and pop-up boxes giving further details about region characteristics.

anchor them at a specified altitude. However, you can circumvent this problem and thus create "floating" extruded polygons by putting them inside a sleeve defined as a similar polygon anchored at ground level and then coloring the sleeve with a transparent color having 100% opacity.

The body of KML consists of the code to create the 3-D histograms and the ground cover regions on Google Earth, based on the fitted model. It also requires reading the longitude and latitude coordinates of the boundary arcs of the regions, stored in a preprocessed text file.

When the resulting KML file is compiled, Google Earth will execute it immediately. The user can operate the system by selecting menu options from a sidebar and using the mouse to zoom, move, and rotate objects displayed in the Google Earth viewing window.

Details of R commands needed to fit the statistical models are given in chapters 6 and 7 of Venables and Ripley (2002), and comprehensive details of R commands needed to create statistical graphs are described in chapter 3 of Murrell (2006). Comprehensive information on KML can be found on Google's KML Code Web site (http://code.google.com/intl/th/apis/kml/).

DISCUSSION

Presenting experimental results in digital formats to the public without having to install a complicated tool is a goal of some scientists and statisticians. Google Earth is freely available for downloading both as a stand-alone program and as a plug-in for a Web browser in almost any computer and operating system. Google Earth also provides customizability and interactiveness. A layer can be added or removed, depending on the viewer's request, while the ability to zoom, rotate, and tilt a global map aids viewers by adjusting an angle to compare spatial data of their own preference. Additionally, spatial outcomes from an experiment or analysis can be imported to Google Earth in the form of KML files and made available to the public via the Internet. Apart from our study, Wood et al. (2007) have demonstrated an application for using the Google Earth–displayed tag approach of data sets and embedding KML files in a server.

Dodsworth (2008, 67) shows examples of using Google Earth with historical maps, but some applications include purchased software such as ArcGIS or Excel to create KML files. This is not a practical choice when purchasing propriety software is not an option. Alternatively, R is developed under GNU general public license and can be freely used. This means that using R combined with Google Earth is a virtually free approach opening possibilities for conveying research with less, or no, cost involved.

Open-source software and free software have become common in GIS research for some time now. GRASS GIS (http://grass.fbk.eu/) is among the most preferred tools due to its open-source nature and functionalities, and its combination with R has been successfully demonstrated (Bivand and Neteler 2000). Although this combination is effective, it demands a high learning curve and extra programming skill to develop interactive and widely accessed maps compared with using the combination R and Google Earth.

CONCLUSION

The prototype system we have developed takes advantage of ongoing developments in computing and communications technology to provide an improved paradigm for statistical graphics, following pioneering work by Sandvik (2008). The ability to combine screen overlays of statistical graphs with an interactive geographical information system having the ability to range over the earth's surface and zoom into local areas provides a very powerful tool for the scientist. The terrorism victim application clearly shows the importance of looking at both the statistical graphs and the raw data at the same time. Viewing the confidence intervals alone is very useful for making valid statistical comparisons of risk rates, but transforming these rates to remove skewedness, and thus satisfy statistical assumptions, can obscure their real magnitudes. Showing the untransformed histograms on a map together with the statistical graphs restores the balance, giving the viewer a more complete and informative picture.

This system can be developed much further, not only for assisting social scientists to gain a better understanding of terrorism study, but also for other studies with a similar data structure, such as providing informative graphs of death rates by age-group, gender, district of residence and year, with the object of identifying environmental factors associated with mortality (Odton 2010). In studies wherein region can be separated from other factors that include interactions, it is of interest to graph incidence rates for these factors after adjusting for region, and such graphs can still be created as ground overlays using KML, even though they are not connected to a particular geographical location. The method can also be applied to outcomes other than incidence rates, such as means and proportions. A further application of current interest is the pattern of global temperature changes on the earth's surface in recent decades. There are several GIS software systems that have been used to analyze data and display outcomes on a map. However, a combination of R and Google Earth has some key benefits compared with other software.

REFERENCES

Bivand, R. S., and Markus Neteler. 2000. Open source geocomputation: Using the R data analysis language integrated with GRASS GIS and PostgreSQL data base systems. Paper presented at the 5th Conference on GeoComputation, University of Greenwich, U.K., August 23–25.

Dodsworth, E. 2008. Historical mapping using Google Earth. *Cartographic Perspectives* 61: 63–69.

Murrell, P. 2006. *R Graphics*. Boca Raton, FL: Chapman and Hall/CRC, Taylor and Francis Group.

Odton, P. 2010. Trend and levels of liver cancer mortality in the upper north-eastern region of Thailand: A case study of using Google Earth in public health. Paper presented at the annual National Statistical Conference in Chiang Mai, Thailand, May 27–28.

R Development Core Team. 2010. R: A language and environment for statistical computing. Vienna: R Foundation for Statistical Computing. http://www.R-project.org (accessed September 10, 2010).

Sandvik, B. 2008. "Thematic mapping at Googleplex." *Thematic mapping blog*. http://blog.thematicmapping.org/2008/12/thematic-mapping-at-googleplex.html (last modified December 13, 2008).

Venables, W. N., and B. D. Ripley. 2002. *Modern Applied Statistics with S*. 4th ed. New York: Springer.

Wood, J., J. Dykes, A. Slingsby, and K. Clarke. 2007. Interactive visual exploration of a large spatio-temporal dataset: Reflections on a geovisualization mashup. *IEEE Transactions on Visualization and Computer Graphics* 13(6): 1176–1183.

XML for S-PLUS Core Team. 2010. XML: Tools for parsing and generating XML within R and S-Plus. In *The Comprehensive R Archive Network*, http://cran.r-project.org/web/packages/XML/index.html (accessed December 14, 2010).