

Statistical Aspects of Public Health Problems in South Asian Association for Regional Cooperation (SAARC) Countries

Marina Roshini Sooriyarachchi*, Linali Pathiratne, Peshala Madhubani, Waruni Abeysekera, Udeni de Silva, Gayani Ratnayake, Sulakshana Perera, Nishika Ranatunge

Department of Statistics, Faculty of Science, University of Colombo, Colombo 3, Sri Lanka

Abstract

The use of statistics and analytic techniques in public health research is of vital importance and has led to important applications in the wide area of public health. Some examples of such studies in which we have been involved in include; Economic, social and environmental determinants contributing to life expectancy, Building of an Automated Statistical Information system for diagnosis and determining prognosis of life threatening viral diseases and application to Japanese Encephalitis, Multivariate multilevel modeling of diabetes/cardiovascular diseases and respiratory diseases. These and more studies will be discussed in detail, emphasizing the scope for Statistical modeling in Public Health Data. Different types of statistical models will be discussed briefly explaining the objective and the use of using the particular model. The results given by the model will be interpreted and shown how the results satisfy the objectives of the problem. These studies will be of interest to public health personal, epidemiologists and medical statisticians who are usually involved in the planning, operational and analysis stage of such studies. These studies represent developing countries where the budgets are generally low leading to relatively small sample sizes. The techniques used range from simple to advanced forms of modeling dependent upon the design and type of response/responses in the study.

Keywords

Public Health, Case Studies, Statistical Analysis, WHO

Received: April 11, 2019 / Accepted: June 5, 2019 / Published online: June 14, 2019

@ 2019 The Authors. Published by American Institute of Science. This Open Access article is under the CC BY license.

<http://creativecommons.org/licenses/by/4.0/>

1. Background

According to the World Health Organization (WHO) Public health refers to all organized measures (whether public or private) to prevent disease, promote health, and prolong life among the population as a whole. Its activities aim to provide the conditions in which people can be healthy and focus on entire populations, not on individual patients or diseases [1].

The WHO is the international agency that coordinates and acts on global public health issues. Most countries have their own government public health agencies, sometimes known as

ministries.

The use of statistics and analytic techniques in public health research is of vital importance and has led to important applications in the wide area of public health. Some examples of such studies in which I have been involved in include (1) economic, social and environmental determinants contributing to life expectancy (2) Building of an Automated Statistical Information system for diagnosis and determining prognosis of life-threatening viral diseases and application to Japanese Encephalitis (3) Multivariate multilevel modeling of diabetes/cardiovascular diseases and respiratory diseases.

* Corresponding author
E-mail address: roshini@stat.cmb.ac.lk (M. R. Sooriyarachchi)

2. Some Case Studies from the Sri Lankan and Global Context

2.1. Economic, Social and Environmental Determinants Contributing to Life Expectancy [2]

Objective: To examine the effect of economic, social and environmental indicators on life expectancy using a cross-sectional comprehensive worldwide sample.

Methods: The data corresponds to countries belonging to the United Nations (UN) and was taken from the World Statistics Pocketbook published by the UN in 2005. Using this sample of data, the impact of country level variables on life expectancy (LE) at birth was analyzed. Multiple regression was used to model the LE. Principal Components (PCs) were used as explanatory variables for the model so as to avoid the problem of high correlation between explanatory variables (multi-collinearity).

Findings: The model variants suggested that proxies for economic development, technology, nature conservation, education, healthcare, communication, population density and population growth rate all have a significant effect on average life expectancy.

Conclusion: To increase the LE it is recommended to improve the economy, education level, health facilities, communication facilities and rural life style and to reduce industrialization, pollution, war and inflation. In addition population growth should be encouraged but urbanization should be controlled so as to improve the LE. This analysis provides information required to governments, especially in the developing world as the LE at birth is predicted with high explanatory power by variables that can be influenced through public policy.

2.2 An Epidemiological Study of Viral Hepatitis in Sri Lanka [3]

Objectives: Viral Hepatitis is one of the main water born diseases and it has been endemic in Sri Lanka for many years. The risk is everywhere and it has emerged as a serious public health threat in the whole country in recent times. Thus it is essential to carry out an island wide research / analyses and recognize possible causal factors for the disease.

Methods: By using statistical and computational techniques an attempt has been made to identify the prognostic factors that affect survival of hepatitis patients, Identify associations between each covariate related to hepatitis, the factors affecting incidence of disease and classify the patients according to two forms of the disease (acute or chronic). The

main study was focused on analyzing information about hepatitis cases reported from all island government and private hospitals during the period 2005-2008. The data was obtained from the Epidemiological Unit, Colombo 8.

Findings: Preliminary analysis reveals the pattern and nature of the disease. In order to identify the factors that affect the incidence of hepatitis, Generalized Linear Models were fitted using Generalized Estimating Equations methodology [3], with a Negative binomial distribution for the responses. The response variable was the number of hepatitis cases recorded from a certain district, in a particular month of the year 2007.

The model reveals that the incidence of hepatitis is influenced by water quality parameters such as conductivity and the amount of fecal coli form in the water. Also, there's a significant association between the month of the year and the incidence. According to the past literature Viral Hepatitis is not affected by weather parameters such as rainfall, humidity and wind. So these were not studied.

A Cox model was fitted to survival data and finally it was pointed out that the risk of death for individuals who used both boiled and unboiled water is higher than for individuals who used boiled water only; an individual residing in Central province has twice the risk of death relative to one in Western province Tamil patients seem to have a higher risk of death than Sinhalese patients. Evidence indicating genetical, geographical and cultural factors have not been found in the literature, but this type of advanced statistical study has rarely been done in Sri Lanka before. This would be the interesting topic for further research.

Log linear model emphasized that all prognostic factors of hepatitis have association with each other and produced informative conclusions. Feed forward procedure of Artificial Neural Networks (Simon Haykin. Feed forward Neural Networks: An Introduction 1998) can be used to classify any patient according to the form of hepatitis that they carry. About 78% accuracy was achieved based on their symptoms.

Conclusion: To reduce the incidence of viral hepatitis it is recommended to improve the water quality by increasing the conductivity and reducing the fecal Colie. To improve the survival of viral hepatitis patients it is recommended to drink boiled water. The classification of viral hepatitis in to acute and chronic can be used to treat the patients accordingly.

2.3. Sexual Behaviors and Knowledge of HIV/Acquired Immuno Deficiency Syndrome (AIDS)/ Sexually Transmitted Diseases (STDs) of the Plantation Workers in Upcountry Estates of Sri Lanka [4]

Objectives: Prevention is better than cure! In the case of

HIV/AIDS, prevention is of course is better than cure since there is still no cure or a vaccine for HIV/AIDS. Thus, effective strategies and interventions are essential in controlling this epidemic. The effectiveness of those strategies and interventions are strongly influenced by leadership at all levels, community attitudes and participation, targeting the right population and such other factors. Apparently, to assess the threat of HIV/AIDS to their country, the leaders must look realistically at the way people behave. For that the following objectives were drafted. To determine the level of knowledge of HIV/AIDS of plantation workers and assess the factors which affect that knowledge, to study the socio-demographic and behavioral characteristics of plantation workers and to determine the knowledge, accessibility and availability of condoms among plantation workers.

Methods: The survey–HIV/AIDS/STD RELATED BASELINE SURVEY AMONG PLANTATION WORKERS IN SRI LNKA-2005–was designed to represent the plantation workers in upcountry estates of Sri Lanka. 594 plantation workers were selected into the sample using a two-stage cluster sampling technique where 5 estates were selected out of 50 estates using the probability proportional to size (PPS) technique in the first stage and 2 divisions were randomly drawn from each selected estate in the second stage. An interviewer-administered, face-to-face interview, using a pre-coded questionnaire was used to collect the data.

The subject coverage by the questionnaire can be summarized as follows.

Demographic characteristics including Marriage and live in relationships, Sexual history, Condom usage and knowledge, Knowledge and experience in STDs, Knowledge and Attitudes towards HIV/AIDS, Exposure to mass media

The study consisted of a preliminary analysis and two different modeling approaches, namely ordinal logistic regression for modeling the response-knowledge of HIV/AIDS and log-linear models that illustrate the significant associations between the socio-demographic and behavioral characteristics of the respondents [5]. Analyses were performed using statistical packages SAS and S-Plus.

Findings and Conclusions: The males in the plantation sector are more unstable regarding their sexual relationships than females). However the males seem more knowledgeable on the diseases like HIV/AIDS and STDs and even more aware of the use of condoms regarding their safety at sex. One reason for this could be the high exposure to mass media of males when compared with females. The young plantation workers (<30) who do not use condoms are more educated and knowledgeable than older age groups on the facts of interest of the study. The condom use of plantation workers

was not so satisfactory since it was *not* found that the people with higher sexual practices (can even word as highly vulnerable) tend to use condoms more frequently compared to others. The respondents with higher level of education have a greater knowledge of STD relative to those with lower levels of education. Also, the odds of having a higher knowledge is around 5 times higher in plantation workers with a higher exposure to mass media relative to those with lower exposure to mass media.

2.4. Generalized Linear Multilevel Models for Ordinal Categorical Responses: Methods and Application to Medical Data [6]

Background and Objectives: This study focuses on analysing the factors that affect the severity of respiratory infections diagnosed in patients. The study is based on clinical data collected at 14 family practices around the island. The data consisted of individual patient records, clustered within the practices (clinics) and thus required a multilevel Modeling approach [7]. The response variable, *severity* of the disease diagnosed, is an ordinal categorical variable. Respiratory infections are one of the most common forms of disease contracted by humans. More than 20 different types of respiratory infections exist and identifying the correct treatment plan depends chiefly on the accurate and timely diagnosis. Hence the major objective of this study is to identify the effect of different factors at different levels, on the response variable which is the severity of the respiratory infection diagnosed.

Methods: The initial phase of analysis was done using an array of graphical techniques, the most prominent of which were the Extended Association Plots. These plots are a graphing technique which enables the visualisation of Pearson residuals, for testing the association between two categorical variables. However owing to certain characteristics of the data, relevant literature suggested that the Pearson's Chi-Squared tests of association would fail in the current situation, and thus a more suitable univariate test had to be carried out in order to obtain an accurate overview of the factors. The Generalized Cochran-Mantel-Haenzsel test for correlated categorical was carried out in order to assess the individual effect of each factor on the response. The test revealed the patients symptom and the duration of symptoms to be the two factors with the highest effect on the response.

A forward selection procedure was followed in selecting variables to the final model and a non-proportional odds generalized linear multilevel model was fitted for the data, using the most significant factors.

Findings and Conclusions: Symptoms, duration of symptoms

and age of the patient proved to be significant in the model whilst the gender of the patient was seen to be insignificant. The residual analysis on the level-2 residuals (practice-level residuals) confirmed the model to be adequate.

2.5. Automated Statistical Information System (ASIS) for Diagnosis and Prognosis of Life-threatening Viral Diseases [8]

Background and Objectives: Diagnosis of life-threatening viral diseases, such as Meningitis, Viral Hepatitis, Japanese Encephalitis, Dengue, Leptospirosis (Rat Fever) to name a few, is extremely challenging particularly in low-resource settings, because the clinical presentation of such diseases cannot accurately be differentiated from that of other types of viral fever and laboratory tests need to be done to confirm the diagnosis. Due to limitations on cost or availability of diagnostics, or lack of access to laboratory facilities for specimen testing, it may not be possible to conduct diagnostic testing nationwide on all recorded suspected disease cases. Therefore epidemiologists will select a subset of such suspected cases for further investigation based on a rule of thumb. Thus a classification rule is vital to assist doctors in order to do this selection. In addition to diagnosis, it is also important to determine the prognosis of such patients as the concern is on life threatening diseases. Determining diagnosis and prognosis is often further complicated by the presence of missing values. The major objective of this study was to develop a user friendly Automated Statistical Information System (ASIS) that will output the diagnosis and prognosis of the patient when details regarding risk factors are given.

Methods: In order to satisfy each of these objectives logistic modeling, survival modeling and Missing value imputation was used [9]. Once the appropriate models were fitted these models were combined using a Hierarchical Statistical Decision model (HSDM) to aid in developing the ASIS. The methodology developed was illustrated on a dataset of Acute Encephalitis Syndrome (AES) patients [10, 11].

Findings and Conclusions: The developed ASIS is applicable for any life threatening viral disease and it will help the epidemiologist to make quick decisions particularly in low income settings where there are low funds for sophisticated diagnostics.

2.6. Factors Associated with the Fertility of Sri Lankan Women [12]

Background and Objectives: Human Reproduction is a purely biological process. It is affected by numerous biological as well as demographic, socio-economic, cultural and some other factors. Some of these factors can be more significant

in this scenario. The mechanism through which they influence may be quite complex [13]. Along with other developing countries, Sri Lanka is also experiencing declining trend of fertility during last five decades due to various Demographic, Socio economic and Cultural factors. The average number of live births during the mother's reproductive life span has been decreased from 5.4 live births in 1953 to 1.9 live births in 2000 and the total fertility rate has approximately become stable during 2000 to 2006/07.

The aim of this study was to study the relationships between fertility and demographic, socio economic and cultural factors of women, To identify the most significant factors which affect the fertility of women and to find a suitable model for the fertility using selected significant factors to identify how these factors behave in the process of child bearing.

Methods: Data set of the Demographic and Health Survey – 2006/07 is being used for this study. The questionnaire which is used for the survey contains two main sections, namely household section and a section on women and children. A stratified two-stage cluster sample design was used to select the sample. A sample of 21,060 housing units has been selected for the survey. 19,862 households and 14,909 ever married women aged 15-49 yrs were interviewed. Graphs and tabulations were used for descriptive statistics. As several observations were collected from a single household the data were correlated. Thus the Generalized Estimating Equations (GEE) method was used for modeling.

Findings and Conclusions:

The results showed positive association between fertility and current age, contraceptive use, changing the type of the occupation from Managerial and Professional grade to the lower level employments and current unions.

There is negative association between fertility and age at first union and educational level.

The Generalized Pearson's Chi square statistic was 7474.4144 on 14901 degrees of freedom resulting in a p-value of 1.0. Since the p value is far greater than 0.05, the model fits the data quite well.

2.7. Multivariate Multilevel Modeling of Age Related Diseases as Reported by Who in 2013 [14]

Objectives: This study aimed to examine the emerging role of modeling multivariate multilevel data in the context of analyzing the risk factors for the severity of Cardiovascular Disease and Diabetes (CDD) and Chronic Respiratory Conditions (CRC).

Methods: The data set consists of worldwide mortality rates

of adults in 195 countries which are clustered within the continents geographically. The two response variables are the mortality rates of CDD and CRC, which are binary variables. Hence a multivariate multilevel probit regression model was adopted using the Bayesian Markov Chain Monte Carlo (MCMC) method [7].

Findings and Conclusions: When determining the risk of diseases, Water showed a significant impact [15]. The modeling phase results drew up some important interaction terms between blood glucose, blood pressure, obesity, smoking and alcohol to the mortality rates. It further established that North America and Europe contributed to a lower risk of having CDD and CRC while Asia and Oceania vice versa.

3. Lessons to be Learnt

- i. The Use of Proper Statistical Tools is Vital for the Analysis of Public Health Data
- ii. Descriptive statistics in the form of tables and graphs is important for the layman. These should be backed up by univariate tests.
- iii. Appropriate modeling techniques are very important to draw out the hidden patterns in the data.
- iv. Goodness of fit tests and diagnostics should be used to test the fit of the models.
- v. Models should be internally and externally validated.
- vi. It is important to determine whether the study findings tally with what is known.
- vii. All the case studies discussed are relevant for SAARC countries.

4. Conclusions

Public Health personal and medical doctors without a proper background in Statistics have a problem of using appropriate statistical techniques to analyze their studies. The design of the study and the response type should determine the type of analysis. In all stages of the study a statistician should be consulted in order to have a valid study. It is recommended to have a study team for any study which includes a statistician. This advice is important in the publication of study results. Almost all the studies used in this paper have been published on their own. Each study model has been tested for the validity of the used model using some form of diagnostics. This is important to have a model that can be generalized to other data of similar structure. The individual conclusions for each study are reported under each study.

References

- [1] Workshop on Statistics in Public Health, Kolkota, India (2016) <https://mail.google.com/mail/u/0/#inbox/14c34c9cf18a281e>.
- [2] Linali Pathiratne and M. R. Sooriyarachchi, Factors Affecting Life Expectancy: A Global Perspective. *Journal of Environment Protection and Sustainable Development*. In Press.
- [3] Peshala Madhubani and M. R. Sooriyarachchi (2013), Model to Identify the Affect of Water Quality Parameters on Incidence of Hepatitis. *International Journal of Earth Sciences and Engineering* 6 (4): 751 – 756.
- [4] W. W. M. Abeysekera, Roshini Sooriyarachchi (2008), A NOVEL METHOD FOR TESTING GOODNESS OF FIT OF A PROPORTIONAL ODDS MODEL: AN APPLICATION TO AN AIDS STUDY 36 (2): 125 – 135.
- [5] Collett, D. (2003). *Modeling Survival Data in Medical Research*. New York: Chapman & Hall/CRC.
- [6] De Silva, D. and Sooriyarachchi, M. R., 2012. Generalized Linear Multilevel Models for Ordinal Categorical Responses: Methods and Application to Medical Data. *Sri Lankan Journal of Applied Statistics*, 12, pp. 83–99
- [7] Browne, W. J. 2009 *MCMC Estimation in MLwiN v2.10*. Centre for Multilevel Modeling, University of Bristol, Bristol, UK.
- [8] Rathnayake, G. I. and Sooriyarachchi, M. R., 2014. Automated Statistical Information System (ASIS) for Diagnosis and Prognosis of Life-threatening Viral Diseases. *Sri Lankan Journal of Applied Statistics*, 15 (3), pp. 185–210. DOI: <http://doi.org/10.4038/sljastats.v15i3.7796>.
- [9] Raghunathan, T. E., Lepkowski, J. M., Van Hoewyk, J., & Solenberger, P. (2001). A Multivariate Technique for Multiply Imputing Missing Values Using a Sequence of Regression Models. *Survey Methodology, Statistics Canada*, 27 (1), 85-95.
- [10] Peiris, J. S., Amerasinghe, F. P., Amerasinghe, P. H., Rathnayake, C. B., Karunaratne, S. H., & Tsai, T. F. (1992). Japanese encephalitis in Sri Lanka--the study of an epidemic: vector incrimination, porcine infection and human disease. *PubMed*, 86 (3), 307-313.
- [11] *Water Sanitation and Health (WSH)*. (2011). Retrieved 2011, from WHO: http://www.who.int/http://www.who.int/water_sanitation_health/diseases/encephalitis/en//diseases/encephalitis/en/.
- [12] Sulakshana Perera (2011). Reproduction and child mortality. Unpublished B.Sc. Thesis.
- [13] K. M. D. S. D. Karunaratne and M. R. Sooriyarachchi (2013). Factors Associated with the fertility of Sri Lankan women. Annual Research Symposium, 2013 Faculty of Science, University of Colombo.
- [14] KapurugeNishikaOshadiniRanathunga and M. R. Sooriyarachchi (2017), A study of the effect of water quality related variables on some age related diseases adjusted for other well-known risk factors: a multivariate multilevel study. *JWSHD*, Open Access.
- [15] www.who.int/mediacentre/news/releases/2016/deaths-attributable-to-unhealthy-environments/en/. Retrieved on 19.02.2017.