MASTER'S THESIS

Application of statistical methods to problems in epidemiological research

Ho, Lai Ping

Date of Award:
2003

Application of Statistical Methods

to Problems in Epidemiological Research

HO Lai Ping

A thesis submitted in partial fulfillment of the requirements

for the degree of

Master of Philosophy

Principal Supervisor: Dr. GOGGINS William B.

Hong Kong Baptist University

March 2003

ABSTRACT

In this study, we apply some statistical methods to three problems in epidemiological research. The first is *Short-term Effects of Air Pollution on Daily Mortality in Hong Kong.* In this problem, we investigate the short-term effect of atmospheric pollutant (NO_2, SO_2, CO, O_3 and RSP) on daily mortality in Hong Kong using Poisson Regression analysis for time series controlling for the confounding effects of time trends and seasonality. Increases in daily concentrations ($\mu g/m^3$) of NO_2 and O_3 were significantly associated with increases of 0.0008 (95% CIs = 1.0001, 1.0016) and 0.001 (95% CIs = 1.0003, 1.0016) respectively deaths due to respiratory diseases. We can conclude the risk of mortality from respiratory diseases seemed to be more affected by air pollution, especially NO_2, CO and O_3 and the identification of target disease and high risk groups would be useful in finding suitable air quality guidelines in environmental health. The second chapter is *Risk Factors for Predicting a Second Primary Kidney Cancers among Kidney Cancer Survivors.* In this chapter, we investigate how the risk among survivors of developing second primary varies with time from diagnosis of kidney cancer by using the Kaplan-Meier Method, and apply the Cox proportional hazard model to determine which specific patient factors will predict the development of a second primary kidney cancer. A higher significantly risk for a second primary kidney cancer for the black kidney cancer survivors, also males, those in middle age, genetics, life habits, living situations and diet habits may be the factors to modulate the risk. The third chapter develops *A Method for Regression Analysis of Failure Time Data with Interval-Censoring on Both the*

Originating Event and Terminating Event. For some types of survival data, both the originating event and the terminating event can be censored into intervals. Since the failure time is the difference between time of originating event and terminating event, the failure time is also interval-censored. In this paper, we consider the case in which both times of infection and disease onset to be interval-censored simultaneously, and also both times of infection and latency that will depend on a covariate by using the Cox Proportional Hazard model. We use a Monte Carlo EM algorithm, and evaluate our method using simulation studies. The proposed method has some advantages over midpoint imputation and right-endpoint imputation; the coverage for midpoint and right-endpoint imputation is less than 95% in some cases, but not for the proposed method. In this part, we also apply to a data set and found that the hazard infection time is higher for patients who were heavily treated, with a hazard ratio of over 2, but whose estimated age at time of infection is greater than 20, with a insignificant hazard ratio under 1.

Contents