

MASTER'S THESIS

Application of statistical methods to problems in epidemiological research

Ho, Lai Ping

Date of Award:
2003

[Link to publication](#)

General rights

Copyright and intellectual property rights for the publications made accessible in HKBU Scholars are retained by the authors and/or other copyright owners. In addition to the restrictions prescribed by the Copyright Ordinance of Hong Kong, all users and readers must also observe the following terms of use:

- Users may download and print one copy of any publication from HKBU Scholars for the purpose of private study or research
- Users cannot further distribute the material or use it for any profit-making activity or commercial gain
- To share publications in HKBU Scholars with others, users are welcome to freely distribute the permanent URL assigned to the publication

**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 (NO₂, SO₂, CO, O₃ 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 ($\mu\text{g}/\text{m}^3$) of NO₂ and O₃ 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 NO₂, CO and O₃ 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

Declaration	i
Abstract	ii
Acknowledgements	iv
Table of Contents	v
List of Tables	viii
List of Figures	xi
Chapter 1 The Short-term Effects of Air Pollution on Daily Mortality in Hong Kong	1
1.1 Introduction	2
1.2 Methodology	3
1.2.1 Data Set	3
1.2.2 Statistical Analysis	6
1.2.3 Statistical Modelling	7
1.2.4 Single Pollutant Models	11

1.2.5	Interactions Between Pollutants	12
1.3	Results	12
1.3.1	Descriptive Statistics	12
1.3.2	Statistical Modelling	13
1.3.3	Single-Pollutant Model	17
1.3.4	Interactions Between Pollutants	20
1.4	Discussion and Conclusions	24

Chapter 2 Risk Factors for Predicting a Second Primary Kidney Can-

cer among Kidney Cancer Survivors 27

2.1	Introduction	28
2.2	Material and Methodology	28
2.2.1	Data Set	28
2.2.2	Methods	29
2.3	Result	31
2.3.1	Descriptive Statistics	31
2.3.2	Kaplan-Meier Method	31
2.3.3	Cox Proportional Hazard Models	31
2.3.4	Checking the Proportional Hazards Assumption	41
2.3.5	Adjusted Cox Proportional Hazard Model	48
2.4	Discussion and Conclusion	49

Chapter 3 A Method for Regression Analysis of Failure Time Data

with Interval-Censoring on Both the Originating Event

and Terminating Event	52
3.1 Introduction	53
3.2 Methodology	54
3.3 Estimation of Variance for Proposed Method	57
3.4 Simulation Study	59
3.4.1 Simulation Methods	59
3.4.2 Simulation Results	61
3.5 Application: Hemophilia Data	63
3.6 Discussion	65
Bibliography	66
Appendix I	73
Appendix II	77
Appendix III	78
Appendix IV	79
Curriculum Vitae	88