# INFORMATION TO USERS

This manuscript has been reproduced from the microfilm master. UMI films the text directly from the original or copy submitted. Thus, some thesis and dissertation copies are in typewriter face, while others may be from any type of computer printer.

**The quality of this reproduction is dependent upon the quality of the copy submitted.** Broken or indistinct print, colored or poor quality illustrations and photographs, print bleedthrough, substandard margins, and improper alignment can adversely affect reproduction.

In the unlikely event that the author did not send UMI a complete manuscript and there are missing pages, these will be noted. Also, if unauthorized copyright material had to be removed, a note will indicate the deletion.

Oversize materials (e.g., maps, drawings, charts) are reproduced by sectioning the original, beginning at the upper left-hand corner and continuing from left to right in equal sections with small overlaps.

Photographs included in the original manuscript have been reproduced xerographically in this copy. Higher quality 6" x 9" black and white photographic prints are available for any photographs or illustrations appearing in this copy for an additional charge. Contact UMI directly to order.

Bell & Howell Information and Learning
300 North Zeeb Road, Ann Arbor, MI 48106-1346 USA
800-521-0600

# The case-crossover design: an efficient rate ratio estimator based on prescription times

**Bing Cai**
**Department of Epidemiology and Biostatistics**
**McGill University**
**Montreal, Quebec, Canada**

**July 1999**

A thesis submitted to the Faculty of Graduate Studies and Research in partial fulfilment of the requirements for the degree of Master of Science

# ABSTRACT

The case-crossover design is a new epidemiological method that evolved around binary exposures and the binomial distribution. We develop a new approach of data analysis for this design based on the actual exposure occurrence times, such as those available from computerized prescription databases. Assuming an exponential distribution for the inter-exposure onset times, we derive two new matched-paired estimators of the odds-ratio, one weighted the other unweighted. A simulation study demonstrates that both new estimators based on the exponential distribution are more efficient than the classical estimator based on the binomial distribution and that the unweighted estimator appears to be the most valid. These new estimators of the odds-ratio are also more flexible and amenable to verifying some of the assumptions behind the case-crossover design. We illustrate this approach with data on 54 asthma deaths identified from the Saskatchewan Health databases, to assess the association with the use of inhaled beta-agonists.

# RÉSUMÉ

Le devis cas-chassé-croisé représente une méthode épidémiologique récente s'appuyant sur l'analyse d'expositions binaires et la distribution binomiale. Nous présentons une nouvelle approche d'analyse de données spécifique à ce devis et basée sur les dates exactes d'exposition, telles que celles disponibles dans les bases de données informatisées de santé. En émettant l'hypothèse que les durées des périodes entre les expositions suivent une distribution exponentielle, deux nouveaux estimateurs appariés des rapports de cotes, l'un pondéré et l'autre non-pondéré, ont été développés. Une étude de simulation démontre que ces deux nouveaux estimateurs, issus de la distribution exponentielle, sont plus précis que l'estimateur traditionnel basé sur la distribution binomiale. De plus, l'estimateur pondéré semble être le plus valide. Ces nouveaux estimateurs des rapports de cotes sont également plus flexibles et permettent la vérification de certaines hypothèses inhérentes au devis cas-chassé-croisé. Cette approche est illustrée à l'aide de données provenant des bases de données de l'assurance-santé de la Saskatchewan, afin d'étudier l'association entre 54 décès dus à l'asthme et l'utilisation de bêta-agonistes en inhalation.

**To the memory of my grand parents and my father**

# ACKNOWLEDGEMENTS

First of all, I would like to express my sincere gratitude to my supervisor, Dr. Samy Suissa, who gave me intense supervision from the point of determining the study project throughout the completion of the thesis. I am very proud to have a supervisor who has original views on epidemiology based on his strong statistics background and many years of experience on pharmacoepidemiology. I deeply appreciate his continuing effort to supervise my study--the benefits I derive from it are much more than just finishing this thesis.

Also, I wish to sincerely thank my thesis co-supervisor, Dr. Pierre Ernst. His clinical expertise and deep understanding of epidemiology made his guidance indispensable for my study. Through him, I now understand how important clinical expertise is to an epidemiologist and the knowledge of epidemiology is to a physician.

In addition, I would like to thank Dr. Jean-Francois Boivin and Dr. Lawrence Joseph, who led me to the field of epidemiology and biostatistics and who gave me continued encouragement throughout my study.

I would also like to thank Dr. Michael D. deB. Edwardes who gave me assistance in statistical techniques and Dr. Lucie Blais who shared her knowledge on the Saskatchewan databases.

I am grateful to Ms. Lisa MacMartin and Mr. Serge Benayoun for the help of preparing the manuscript, and to my office mates Ms. Simone Cowan and Ms. Chantal Bourgault for the support and assistance received throughout the research work.

I also thank the Medical Research Council of Canada and the Canadian Drug Manufacturers Association for granting a scholarship for my studies.

Finally, special thanks to my wife, who continues to support my studies and accompany me through the hardest times.

# TABLE OF CONTENTS

# CHAPTER 1 INTRODUCTION

## 1.1 From crossover design to case-crossover design

The randomized crossover design is a special type of controlled clinical trial in which the study subjects are assigned to various treatments at different times during the study. Treatment effects are estimated by comparing the responses to the treatments within a subject. There are several advantages to this study design: First, each subject forms his own control, therefore, except for the distractions of order and time, the control and treatment group are identical. With within-subject comparisons, we can avoid confounding by subject-specific attributes that are constant over time. This is most appealing when some confounding factors are not measurable. Second, for a limited sample size, this study design can give more power than parallel group designs. Other advantages include better control, more reliability and less ethical problems. The limitation of this study design is that the effects of the treatments must be acute and transient otherwise the treatment effects may "carry over" and alter the response to subsequent treatments.

The case-crossover design is an extension of the crossover design to observational studies. It was introduced in 1991 as a new epidemiological technique to examine the transient effects of a brief exposure on the onset of an acute outcome (Maclure 1991). Similar to the crossover design, the study subjects alternate at varying frequencies between exposure and non-exposure to a agent of interest, but the investigator does not decide when and how to assign the study subjects to exposure or non-exposure. Instead, the history of exposure and non-exposure are observed. Instead of observing an entire

cohort, Maclure (1991) demonstrated that the rate ratio can be estimated only from the cases that occur within the cohort. Thus, in the case-crossover design, as its name implies, only the exposure history of cases is needed for the analysis, so it is also called a case design.

To explain the case-crossover design intuitively, we can imagine that we have collected several cases, and some of the cases occurred when they were under the exposure of a specific agent. Then we ask: is the treatment associated with the outcome or do the treatment and outcome happen concurrently just by chance? To answer this question, we need to know: a) the proportion of cases occurring under the exposure and, b) how often each subject is usually exposed to the treatment. One extreme situation may occur when all of the case events happen concurrently with the treatment, and none of the subjects are exposed to the agent when the outcome does not occur. In this scenario, it is strongly suggested that the agent induced the outcome. The other extreme situation is that all the events happened when the subjects were not exposed to the agent, and these subjects were exposed to the agent all the time before or after the outcome occurred, then it is strongly suggested that the agent is protective. In real life, the data we observe can not reveal the association as explicitly as these extreme scenarios. The case-crossover design provides the statistical technique needed to analyze these kinds of data.

The case-crossover design can be interpreted as analogous to a highly stratified case-control study or a highly stratified cohort study, where each subject forms a stratum.

For the case-crossover design based on a cohort study, we can imagine that it is stratified to the point that there is only one subject in each stratum, and each subject forms a 2x2 table. Depending on whether the event occurs concurrently with the exposure, we fill (1,0) or (0, 1) in the first column of the 2x2 table. According to the history of exposure before the event took place, the person-time data can be filled in the second column of the 2x2 table. The person time data are the number of time windows that the subject is under the exposure and the number of time windows that the subject is under non-exposure. Strata with no outcome event contribute no information to the rate ratio estimate, which means the cohort is reduced to a cases-only study. In this way, the analysis of a case-crossover study with n cases may be viewed as a pooled analysis of n retrospective cohort studies, each with a sample size of one. The data can be analyzed using standard Mantel-Haenszel methods for follow-up studies with sparse data in each stratum or with maximum likelihood methods.

For the version of case-crossover design that resembles the case-control study, we can imagine it as a matched case-control design where each subject forms a stratum and each control is the same person as the case, but with exposure data at different time points. As in the classic case-control study, we can apply different strategies of sampling such as one-to-one matching, M:1 matching, or a variable number of controls matched to each case. For instance, if the exposure in the hazard period is defined as the 1-hour period immediately preceding myocardial infraction onset, the control can be chosen as the comparable 1-hour period at the same time of the day, on the day preceding the infarction. For the same study, we can also choose 25 (1-hour) periods preceding

myocardial infarction onset as 25 controls for each case. For the data analysis, the Mantel-Haenszel estimator or a conditional logistic regression is applied.

With this extension of the crossover design to observational studies, the study subjects can be a large population rather than the very limited sample in a clinical trial, so this technique can be used to assess post-marketing drug effects. Compared to other observational studies such as the cohort-study and case-control study, this design inherits the advantages of the crossover design in a clinical trial, that is, the case and its control in a case-crossover study from the same subject, so that the case and control are automatically matched on all characteristics that do not change within individuals, such as genetic characteristics.

Like the classical crossover design, the case-crossover design is also limited to the study of the transient effects of exposure on acute outcomes. If the effect period of the exposure is longer than the assumed time window, the assumption of the study design is violated and the validity of the study may be challenged. Another assumption of the case-crossover design is that the distribution of exposure must remain the same along the whole control period of time, otherwise a bias due to time trends in exposure may occur. To solve this problem, the case-time-control design was proposed (Suissa 1995).

### 1.2 Rationale and objectives

In the case-crossover design, the estimate of the rate ratio and its confidence limits is based on the comparison of the exposure distributions at the time of the outcome

occurrence and during the control period. In practice, the control period is divided into several units of time window consecutively or separately. The distribution of the exposure is assumed to be binomial, which means that the estimate of rate ratio is based on whether the subjects are exposed or not in each time window. This design was first proposed to study the onset of myocardial infarction associated with sexual activity and coffee drinking in which the past exposure is assessed by interview, where it is easier to record exposure in the control period as a proportion of the time rather than to obtain exact time of exposure.

In pharmacoepidemiology, database studies provide more precise data on the time of exposure. For example, we can obtain the exact date that a drug was dispensed so that the time of exposure to a specific drug can be assessed. In this situation, we may be wasting information by dichotomizing the inter-exposure onset times to mold the data into the form required for the binomial distribution, as proposed by Maclure (1991). For instance, using the binomial distribution, we only need to know whether a subject is exposed or not to a drug in a certain time window, but we do not use the information about the frequency of drug use in this time window, which may provide more power of inference.

The objective of this thesis is to propose a new approach of data analysis for the case-crossover design, so that the information from databases about drug usage may be used more efficiently and more flexibly. For this purpose, we will evaluate exposure history in terms of the exact times at which these exposures occurred, and not only whether they occurred or not in each time window. We assume an exponential distribution for the

inter-exposure onset times, and derive several matched-paired estimators of the odds-ratio. For each odd-ratio estimator, we estimate the variance either analytically or numerically so that the confidence interval can be provided.

To illustrate our new methods of data analysis, we use the Saskatchewan Health database to assess the effect of beta-agonists on the outcome of asthma death. There have been several case-control studies, including the Saskatchewan Health database study, showing that the use of beta-agonists is a risk factor for asthma death, but these case-control studies have two limitations: a) the result may be confounded by indication, which means that the beta-agonists appear to increase risk since the patients for whom the medications are prescribed more frequently are more likely to die because they have more severe asthma; b) these case-control studies did not reveal whether the association of beta-agonist and asthma death is an acute or chronic effect.

The object of this thesis is to propose an efficient method of data analysis for the case-crossover design in situations where precise data on exposure time are available. We also apply this new method to the study of acute transient effects of beta-agonist associated with asthma death.

# CHAPTER 2 LITERATURE REVIEW

The case-crossover design (Maclure, 1991) was introduced in 1991 as a new epidemiological technique to assess the effect of transient exposures on the risk of acute events, which is difficult to deal with by conventional epidemiological methods such as cohort or case-control studies (Meittinen, 1989; Guess, 1989). To review the application and development of this new technique, we searched in Medline databases from 1991 to 1999, with the key word 'case-crossover', and collected 30 related publications. Among these publications, 6 studies are about myocardial infarction (Meier, 1998; Millteman 1995; Mittleman, 1997; Muller, 1996; Willich, 1994) or myocardial ischemia (Gullette, 1997); 6 studies are about traffic accident (Barbone, 1998; Redelmeier, 1997; Roberts, 1995) or injury (Petridou, 1998; Burdor, 1997; Vinson, 1995); other outcomes under study are hemorrhagic fever (Dixon,1997), vulvo-vaginal candidiasis (Sturkenboom, 1995) and asthma death (Suissa 1995). There are also two papers applying the technique of the case-crossover design to the study of injury of racehorses (Carrier, 1998; Estberg, 1998). The other publications are all about method development or reviews. We note that the list of publications is not long because the case-crossover design is limited to the study of transient exposure on the risk of acute events, and many study topics are not amenable to this. Nevertheless, the advantage of this technique over the conventional epidemiological design and the ability of this technique to solve the problem which is difficult for conventional epidemiological design make it a very important alternative in epidemiological research.

There is a common point among the case-crossover design, the case-genotype design (Falk, 1987; Self, 1991; Schaid, 1993; Khoury, 1996; flanders,1996) and the case-specular design (Zaffanella, 1998). That is, all these designs use prior theory or assumptions to replace the information supplied by controls in a case-control study. So Greenland (1998) labeled the case-crossover design and these two other designs as case-distribution designs and provided a unified likelihood-based approach to the analysis of the three designs.

The case-crossover design is a novel approach to epidemiological study design, but it is not independent of the conventional methods. Actually, it has borrowed the idea of study design and data analysis from different conventional study designs. In order to reveal this relationship and to compare the strength and limitation of different approaches, we first briefly review conventional study design of pharmacoepidemiology.

## 2.1 Conventional study design and case-crossover design

According to Strom's review (1994), conventional study designs available for pharmaco-epidemiology are summarized as Fig.2.1. The figure shows that case reports are the least costly, but also provide the least evidence of causality. On the other hand, experimental studies are the most convincing designs, but they are the most costly.

Case reports are simply reports of single patients. As used in pharmacoepidemiology, a case report describes a single patient who was exposed to a drug and experienced a particular, usually adverse, event. Case reports are useful for raising hypotheses, but they

usually can not make a statement about causation. An exception is when the treatment causes a change in disease course, but the change is reversible and the patient can receive the same treatment again after the treatment is withdrawn. In this situation, case reports can reveal some information about causation if the change in disease course responds to the treatment. The case-crossover design is based on the same principle as this situation when the exposure can be compared within an individual.

Case series are collections of patients who have exposed to some agent, and whose clinical outcomes are evaluated and described. This kind of study actually reveals the distribution of exposure among the cases. If the predicted distribution of exposures is available, some information about causation can be deduced from this study. For instance, if 80% of cases are male, it reveals the possibility that the outcome is associated with gender, because the predicted distribution proportion of males is 50% provided the procedure of collecting cases is independent of gender. Another example is a Japanese asthma death study (Mutsui, 1996) in which it is found that fenoterol was prescribed in more than half of the asthma deaths at a time when it only had a 18.3 percent market share. The market share of the drug represents the distribution of exposure in controls and the difference in the two distributions reveals the association between the drug and asthma death. This assumes, of course, that there are no confounders, such as asthma severity. However, in most circumstances, the predicted distribution of exposure is not available. For instance, when we collected the cases of asthma death from the Saskatchewan databases, we found that 87% of these cases used beta-agonists in the last 30 days before death, but this information is not sufficient to determine the causation

because we don't know whether this is higher, the same, or lower than would have been expected. As in the case series study, a case-crossover study collects data only on cases, but the information about exposure in different time windows is collected separately. If the exposure is transient and the effect is acute, the distribution of exposure of the cases before the outcome occurs will represent the exposure distribution of the cases provided the outcome didn't occur, thus representing the control distribution.

Analyses of secular trend, also called "ecologic studies", examine trends in an exposure which is a presumed cause and trends in a disease which is a presumed effect, and test whether the coinciding trends can reveal the association of disease and exposure. For instance, the coincidence of an asthma mortality epidemic with the introduction of beta-agonists raised the hypothesis that beta-agonists can induce asthma death (Beasley 1991). This method is useful for rapidly providing evidence for, or against, an hypothesis, but it only studies groups. No data on individual is under the consideration, so it unable to control for confounding variables. In case-crossover design, it is assumed that there is no time trend and the distribution of exposure is stable if outcome does not occur. The data analysis is based on this assumption and some kind of adjustment needs to be made if there is time trend of exposure. For the asthma data, there is an increasing time trend of beta-agonist usage, and this trend leads to overestimation of odds ratio of association of asthma death or near death with beta-agonist usage (Suissa, 1995).

Case-control studies are studies that compare cases with a disease to controls without the disease, looking for differences in antecedent exposures. This study design is particularly

useful when one wants to study multiple possible causes of a single disease and when one is studying a relatively rare disease. The main challenge of this study design is to properly select controls, and it is most difficult when some confounding variables can not be easily measured, such as disease severity and socioeconomic status. The case-crossover design can be regarded as a matched case-control design, and the technique for data analysis is the same as case-control study, such as logistic regression and Mantel-Haenszel estimator. The main difference between case-crossover design and case-control design is that the case-crossover study use the cases themselves as their control, but information of exposure is for different time periods, that is, the time period when the outcome occurs (risk period) and the time period when the outcome does not occurs (reference period). The fundamental principle used in selecting controls is that selected controls should be representative of the source population which gave rise to the cases (Rothman, 1986). Because the control that matches a case best is the case itself, the case-crossover design surmounts the difficulty of selecting control as found in the case-control study. Another obstacle in the case-control study is that, when dealing with acute adverse events, the timing of the interview or data collection is crucial, but it is often difficult to get data for the control with good timing in the case-control study. Again, the case-crossover design provides a solution to this problem (Suissa 1994).

Cohort studies are studies that identify subsets of a defined population and follow them over time, looking for differences in their outcome. The main advantages of the cohort study are: a) this design is free of the problem of selecting controls without bias; b) an association demonstrated by a cohort study is more likely to be a causal association than

one demonstrated by a case-control study; c) cohort studies are particularly useful when one is studying multiple possible outcomes from a single exposure, especially a relatively uncommon exposure; d) cohort studies can provide excess risk. The main disadvantage of the cohort study is that it requires extremely large sample sizes to study relatively uncommon outcomes and prospective cohort studies can require a prolonged time period to study delayed adverse effects. With the prospective cohort study, it is very difficult to study acute effect of transient exposure (Schneeweiss,1997), as this kind of study needs to make sure that exposure is correctly recorded for even very short units of time on a routine basis so that when an event occurs at a time not anticipated, the data of exposure shortly before the time of event is available, and that is not practical. One of the approaches of case-crossover design closely resembles a retrospective cohort study that is stratified to the point that there is only one subject in each stratum. Strata with no outcome events contribute no information to the relative risk estimate, which means the cohort is reduced to a cases-only study ( Maclure, 1991). Because the cohort is reduced to only cases, it is easy to obtain the information of exposure shortly before the time of event. Contrary to the convention cohort study, the case-crossover study is very useful in the study of the effect of transient exposure on acute outcome.

Randomized clinical trials (RCT) are experimental studies in which the investigator controls the treatment that is to be received by each participant, and the participants are randomly allocated to the study groups. There is no other method for studying the merits of clinical treatment regimens which can approach the precision of estimating effects and the strength of inference permitted by sound RCTs (Baillar, 1983). It is often impossible,

however, to perform a RCT for ethical or logistic reasons. For instance, it is prohibited to recruit children or pregnant women for RCTs. There are four broad classes of research approaches to the clinical evaluation of treatment: parallel studies, crossover studies, self-controlled studies, and externally controlled studies. In a crossover study each patient receives in succession two or more treatments under evaluation. Self-controlled studies evaluate a single treatment and each subject serves as their own control. Responses to treatment are compared with those measured during periods of time when no treatment was offered to the subject. The case-crossover design closely resembles the above two approaches in that each subject serves as their own control. The distinguishing factor is that the investigator can not control the allocation of the study subjects. Thus, the case-crossover design is considered as observational design that poses fewer ethical concerns than RCTs in the study of such topics as alcohol use and injury (Vison, 1995).

## 2.2 Definition and examples of case-crossover design

In practice, the first step in a case-crossover study is to define a time window and then determine whether the exposure was present during the last time window before outcome. The time window corresponds to the time needed for the exposure to cause the outcome under study. The time window should be defined according to the characteristics of exposure and the outcome. For instance, in the study of alcohol and injury, the time window is defined as 6 hours (Vinson 1995), while in the study about vulvo-vaginal candidiasis associated with acitretin, the time window is defined as 20 days (Sturkenboom, 1995). On the other hand, in the study of urban traffic environment and the risk of child pedestrian injury, the time window is extremely short so that an alternate

method is needed to estimate probability of exposure in the reference period (Roberts, 1995). Sometimes, for the same kind of exposure and same outcome, the definition of the time window may vary according to the assumption about the effect and the way to interpret the results. For instance, in the study of beta-agonists associated with asthma death, a one-year time window is defined to access whether asthma death is associated with increase use of beta-agonists in a year before asthma death. This is analogous to the previous case-control study (Spitzer,1992) and cohort study (Suissa, 1994). The assumption behind the study design is that the effect of beta-agonists on asthma death is from long-term regular excessive use, so it studied the chronic effect or indirect effect of the drug. On the other hand, we can also design a study assuming the time window is 4 hours (Suissa, 1994). This time window is defined according to the elimination half-life of the drug and it is assumed that the drug has an effect on asthma death only before it is metabolized. This design studies the direct acute effect of the drug. With specific data, the longer the time window is defined, the more the result favors the null hypothesis that there is no association between the treatment and outcome.

After the time window is defined and it is determined whether the outcome occurs when under exposure ($C_i$=0 or $C_i$=1), the next step is to measure the 'usual' frequency with which the subject was under exposure ($P_i$). For this step, various strategies of sampling the reference period and methods of measuring the frequency are involved, which will be discussed later.

Maclure proposed that the case-crossover design be taken as a matched case-control design where the controls are the same persons as the cases before the event under study occurred, or a retrospective cohort crossover study that is stratified to the point that there is only one subject in each stratum. With this consideration, an estimator of relative risk and variance of logarithm of relative risk were derived by adoption of the Mantel-Haenszel method (Rothman, 1986).

Marshall proposed a maximum likelihood method to analyse the case-crossover design (Marshall 1993). This method is based on a proportional hazards model (Kalbfleisch, 1980) and it is quite general so that, in principle, it can be used to analyse the joint effects of several transient exposures. The method cannot only be used for binary exposure variables, but also continuous exposure variables. Marshall also discussed a mixed distribution model when the distribution of exposure is mixed, consisting of a discrete probability at zero and a continuous part with a normal distribution. This model is used when an exposure is measured as a continuous variable, but whether this exposure happens or not, is measured as a binary variable.

Marshall showed that when the exposure is measured as a single binary variable, say X, and $P_i=f_i(1)$ is the probability of X=1 for case i, then the same formula proposed by Maclure as the Mantel-Haenszel estimator is derived as an approximation to the maximum likelihood solution. It is important to note that here the $P_i$ represents the prior probability of the event X=1, which is different from the $P_i$ defined by Maclure in estimating the relative risk. For instance, if a case is asked about their frequency of

exposure, say $\lambda_i$ in units of time$^{-1}$, the response of once a day means $\lambda_i$=1/24. Then assuming that occurrences of the exposure happen at random, that is, as a Poisson process with intensity $\lambda_i$, the probability of at least one occurrence in a period $t_0$, is,

$$P_i=1-\exp(-\lambda_i t_0) \qquad (2.1)$$

but the value of Pi used by Maclure is,

$$P_i=\lambda_i t_0 \qquad (2.2)$$

It is easy to see that if $t_0$ is short, the above two equations are approximately the same, but when the $t_0$ is long, the difference between the two equations is significant. Marshall took the data from the Auckland Heart Study (Jackson, 1991) as an example to reveal this difference. The data are based on a survey and it is about alcohol consumption associated with myocardial infarction. Under the assumption of a 24 hr time window, the related risk is 1.87 (95% CI, 1.35-2.58) when $P_i$ is calculated with equation (2.2), but when $P_i$ is calculated with equation (2.1), the relative risk is 0.48 (95% CI, 0.27-0.84). One analysis concluded a harmful effect while the other analysis concluded a protective effect! The difference lies with the 20 cases who said they drink once a day, with an assigned probability of one, when $P_i$ is calculated by (2.2), yet did not drink in the 24 hours before their myocardial infarction. Their contribution outweighs the effects of other cases. But when $P_i$ is calculated using Poisson probabilities, it is only 0.632 as calculated by equation (2.1). The first analysis takes 'once a day' as drinking in every 24-hour period. However, it seems more realistic to assign a $P_i$ of less than 1, on the basis of Poisson probabilities. Perhaps, $P_i$=1 is too high and $P_i$=0.63 is too low. However, this is difficult to judge because the data collection did not intend to elicit $P_i$ directly. Consequently, the author suggested that an instrument to elicit the prior probability distribution be carefully

developed. For instance, instead of letting the respondents choose among 'once a day', 'every 3 or 4 days', 'once a week', 'once a fortnight' categories, they should be asked how many times did they drink in an average fortnight, month or year. From this example we can see the advantage of database studies of pharmacoepidemiology, in which the exposure is based on the record of exact times of prescription, thus avoiding the problem as discussed above.

### 2.3 Study settings and methods of data analysis

Since Maclure introduced the case-crossover design, it has been applied in different fields such as occupational epidemiology, injury epidemiology, environmental epidemiology and pharmacoepidemiology. For these settings, different strategies of design and data analysis have been adapted, though they are all under the principle of the case-crossover design as it was originally proposed. This section discusses some of the related issues that have arisen in the various situations.

#### 2.3.1 Effect modification

In the case-crossover design, all variables such as age, sex and socioeconomic status are automatically matched, so this variable will have no confounding potential. These variables can have effect modification, which is of interest in some studies. In this situation, the study cases need to be grouped by these variables. In the study of triggering acute myocardial infarction onset by episodes of anger (Mittleman, 1995), the cases were grouped by age, sex, clinical history, and medication. The relative risk was then

estimated for each subgroup. The study found that the relative risk was significantly lower among regular users of aspirin compared with non-users ($P<0.05$); the relative risk tended to be lower in men than in women and among regular users of beta-adrenergic antagonists than nonusers, but these differences were not statistically significant. In the study about alcohol and injury (Vinson 1995), the cases were grouped by age, sex, weight, severity of injury, socioeconomic status, etc, but no significant effect modification was found.

### 2.3.2 Multiple risk factor

As in a case-control study, the case-crossover design can be used to study multiple possible causes of an outcome. In the study of transient exposures associated with the risk of childhood injury (Petridou, 1998), several different kinds of activities such as strenuous physical activity, intellectual exertion, involvement in family quarrel, school examination, and pleasing events, are assumed to be risk factors of childhood injury. Using the case-crossover design, a rate riatio was estimated for each of those risk factors. In this study, confounding is possible among correlated transient events, for example, when some individuals experienced during the same time period both strenuous physical activity and intellectual exertion. For this reason, multivariate conditional logistic regression needed to be applied to control for such within-person confounding. In addition, time of day was considered an important potential confounder because exposure to the determinants of interest varies with time of day, as does the risk of injury, so a variable as an indicator of time of day was also included in the conditional logistic models.

### 2.3.3 Continuous exposure

When the case-crossover design was originally proposed, the exposure was considered as binary data, that is, a subject can only be exposed or non-exposed in a certain period of time. In some studies, exposure needs to be measured as a continuous variable, and the dose-response gradient is an important evidence of causation (Hill, 1971; Sackett, 1991). Marshall proposed a technique to estimate the relative risk of a continuous exposure based on the maximum likelihood method (Marshall, 1993). The continuous risk factor can be included in the logistic model directly or be transformed into a categorical variable. In the study of asthma death and beta-agonist use (Suissa, 1995), the quantity of beta-agonists used in a month is defined as a continuous variable in the logistic model and the resulting odds ratio is per canister per month. In the same study, beta-agonist use was also defined as trichotomous and odds ratios of 13-24 vs $\leq$ 12 and >24 vs $\leq$ 12 were obtained separately. In the study of the association of road-traffic accidents with benzodiazepine use, a dose-response relation was studied by classifying each dose prescribed as low, intermediate, or high according to the recommended prescribing dose ranges of the British National Formulary (Barbone, 1998).

### 2.3.4 Discrete period of exposure

In some studies, the period of exposure is discrete and the effect period coincides with the exposure period. For instance, in the study of urban traffic environment and the risk of child pedestrian injury, a high volume traffic road is the risk factor of interest (Roberts, 1995). As a child can only get hit by a car when crossing a street, it is a discrete period of

exposure unlike heavy physical exertion, or alcohol and drug uses, which might be 2 hours, 6 hours, or months. In this case, the time window approach is not applicable. In this study, Roberts calculated the relative risk of high-volume-traffic vs low-volume-traffic with the Mantel-Haenszel estimator, but the expected odds of crossing a high-volume street for a student was determined by the ratio of high-volume-traffic streets to low-volume-traffic streets that he crossed every day, while the observed odds was the odds that the student was injured on a high-volume-traffic street. This setting is different from other studies in which the expected odds was determined by the proportion of time a study subject is exposed. This setting is also useful in studies of occupational injury (Mittleman, 1997).

### 2.3.5 Conditional distribution of exposure.

In the study of cellular-telephone calls and motor vehicle collisions (Redelmeier, 1997a), a conditional distribution is involved. Here, the distributions of cellular phone use provided the user is driving in the risk period and in the reference period must be compared. That is,

$$OR = \frac{P(E|C,D)/P(\bar{E}|C,D)}{P(E|\bar{C},D)/P(\bar{E}|\bar{C},D)}$$

Here, E, C, and D respectively represent exposure, outcome and another conditional variable like driving. Unfortunately, it is difficult to obtain the conditional distribution of cellular phone use provided the user is driving, so a marginal distribution was used instead. This may lead to overestimating exposures immediately prior to the outcome or underestimating exposures during control intervals (Redelmeier, 1997 b). For instance,

overestimating the amount of driving on the reference period would dilute the apparent intensity of cellular activity during the control interval and inflate the measure of relative risk. To control for this kind of bias, two procedures could be done: a) adjust for intermittency of driving with a factor determined by the proportion of subjects who did not drive during the reference period (Breslow, 1980; Efron, 1991); b) recalculate relative risks by limiting the analysis to subjects who can remember his or her driving pattern on both the risk period and the reference period and who were confident that they had driven a motor vehicle during both periods. The problem related to the conditional distribution of exposure also existed in the study of the association of road-traffic accidents with benzodiazepine use and it may exist in other studies about work related injury.

### 2.3.6 The effect-period

In most of the published case-crossover studies, the effect time of exposure is really short and it is easily defined. When this study design is applied to pharmacoepidemiolgy in which the presumed risk factor is some kind of drug, the time window is usually longer and it is more complicated to define. When the case-crossover design was originally proposed, Maclure suggested a model that considers delay of effect, exposed time, duration of effect time, and discounting of overlap (Maclure, 1991). Application of this model should vary according to the nature of exposure and outcome under the study. For the database study in which the information of exposure is solely based on record of prescription, it is more complicated in that the prescribed duration of use of a drug also needs to be estimated and taken into consideration. In the study of vulvo-vaginal candidiasis associated with acitretin, the period is defined as the prescribed duration of

use of acitretin plus 20 days to control for residual effects due to slow elimination. The study also defined a carry over time as the first 30 days after every exposure period, if no subsequent prescription was filled within this period. Then non-exposure time was defined as the reference time subtracting exposure and carry-over time (Sturkenboom, 1995).

### 2.3.7 Different sampling strategy

In the case-crossover design, both case and control information are taken from the same person, but from different time periods. Because no controls need to be sampled, bias in the selection of a control person is impossible, but improper selection of the control time window will still result in bias. Thus, selecting a valid strategy of selecting control periods that can avoid bias yet gain best efficiency is an important task of any case-crossover study. The principle of the case-base paradigm (Miettinen, 1985) should also be followed in the selection of the reference period and usually, periodicity of events should be considered. For instance, in the study of physical exertion as a trigger of acute myocardial infarction, the time window is defined as 1 hour, and the periodicity is considered as 24 hours, so an 1hour time window 24 hours before myocardial infarction is selected as the reference period (Willich, 1993). In the study of the association between cellular-telephone calls and motor vehicle collisions, the periodicity of both driving pattern and cellular phone use are considered as a week, so the day of the work-week preceding the collision is selected as the control period (Redelmeier, 1997). In the study of acute respiratory-tract infections and risk of first-time acute myocardial infarction (AMI), however, the exposure is assumed to vary with seasons, so the date exactly 1 year

before the date of AMI is selected as the reference period (Meier, 1998). In addition to the above method called pair-matched interval approach by Mittleman, a so-called multiple intervals approach and usual frequency approach were also applied to the study of heavy physical exertion and acute myocardial infarction onset (Mittleman, 1995). The multiple interval approach contrasts exposure in the hazard period with a variable number of control periods sampled from the 25 (1-hour) periods preceding myocardial infarction onset, for which exposure data were obtained. This approach is analogous to case-control studies in which a variable number of controls are matched to each case. As in case-control studies, conditional logistic regression can be used in this approach, and according to different assumptions about the effect of time of day on the baseline hazard of infarction onset, the time of day variable enters the logistic regression model in different forms. The usual frequency approach contrasts exposure in the hazard period with the expected exposure, based on each individual's usual frequency of exposure over the entire reference period. This approach is analogous to a highly stratified retrospective cohort study in which each stratum has exactly one case event, and all of the person-time is contributed by a single individual based on cumulated exposed time in the reference period. Mittleman empirically compared the relative efficiency of these three approaches, and the results show that the pair-matched interval approach, with the least information being used, has the lowest relative efficiency; the efficiency of the multiple intervals approach is intermediate and the efficiency greatly increased as the number of control periods sampled increased; the usual frequency approach has the highest efficiency among the three, but it can only be used when within-person confounding by clock-time was negligible. Selection of the sampling strategy should depend upon the length of the

hazard period, the induction time from exposure to outcome onset, the degree of within-person confounding and the quality of the data available. It is a trade-off between precision and potential biases of the estimate.

### 2.4 Time trend

The principle of selecting controls in conventional case-control studies is that controls should be selected independently of their exposure status. That is, the controls should be representative of the source population with respect to exposure (Rothman, 1998; Wacholder 1992). This principle is followed well in the case-crossover design, as the best representatives of the source population that produced the cases would be the cases themselves. However, a strong assumption behind this design is that neither exposure nor confounders are changing over time in a systematic way, which means the distributions of exposure in the risk period and reference periods should be the same under the null hypothesis. If the case and control time windows are very long, or if they are short but far apart, this assumption may be challenged because the distribution of exposure may change with time regardless of outcome. Suissa first raised this question with an example of a case-crossover analysis of beta-agonist use associated with asthma mortality. In that study, the time window is defined as 1 year because of the strong seasonal variations in disease and drug use. The reference period was taken to be the year immediately preceding the 1-year current period. Over a period of 2 years, there may be a "natural increase " of drug use over time because of changing medical practice, greater recognition of the drug's benefits, more assurance with prescription of the drug, a wider spectrum of indications, increasing patient reliance on the drug, and aggressive marketing

(Suissa, 1995). When this time trend occurs and we do not adjust for it in our analysis, we may conclude that a drug is a risk factor of the outcome, when in fact there is no association between the drug and the outcome, apart from time trends. In a limited number of papers about case-crossover studies that have been published to date, three methods of controlling the time trend bias have been used and are listed below.

*Case-time control design.* This study design was proposed by Suissa in 1995 and was analyzed with a proportional-hazards model by Greenland in 1998. The original purpose of this design was to use within-subject comparisons to adjust for confounders that are not measurable, such as severity of disease, and which result in confounding by indication of drugs. Compared with the case-crossover design, this new design used a non-case control group to adjust for the control-time selection bias in the case-crossover design, which includes the time trend bias. With the case-time control design, the odds ratio from the case-crossover analysis can be divided into two portions – one portion due to time trends and the other portion due to the drug effect. The resulting odds ratio is adjusted for time trend and controls for between-person confounding. If there is no natural time trends in drug use exposure, the coefficient for the external control-time becomes 0 and the design is reduced to the case-crossover design.

*Population time trend control.* In the study on the association between road-traffic accidents and benzodiazepine use (Barbone, 1998), a strategy that is similar to the case-time-control design is used to control the time trend bias. In the analysis, a term is added to the logistic model of the case-crossover design to control for the time trend. However,

this term is not for external control selected to match each case, as in a case-time control design. It is for quarterly drug-utilization patterns in the study population. This term adjusted for the likelihood that a doctor would prescribe such a drug in the population, thus, it adjusted for the time trend of drug use. This strategy does not need to select external controls, but it is under an assumption that all subjects have the same drug-utilization pattern.

*Two-direction sampling control.* In all case-crossover studies we have discussed above, the control time could only be a time period that preceded the event because the study outcome was likely to affect subsequent exposures. So, sampling control time after the event could result in reverse-causality bias. However, in some studies such as that of the effect of environmental rather than behavioral exposures, the outcome does not affect subsequent exposure. In this situation, select the control time period after outcome is applicable. Navidi (1998) proposed and analyzed a so-called bidirectioal case-crossover design in which the control times were taken both before and after the outcome. With a simulation study, Navidi showed that relative risk estimates by this bidirectional case-crossover design are resistant to the time-trend confounding. Unfortunately, in most of the pharmacoepidemiology studies, the outcome will strongly affect subsequent exposures, so that the strategy is not applicable.

## 2.5 Summary

In this review, we compared the case-crossover design with conventional epidemiological study designs. The case-crossover design is based on the same principle of other

observational studies such as the cohort and case-control study, and the techniques of data analysis of these conventional studies are applied. However, this novel approach borrowed the advantages of the case-series study with respect to cost and feasibility, and as well as the advantages of the randomized crossover clinical trial with respect to the power to control for variations among subjects. The main advantages of the case-crossover design can be summarized as follows:

1) No separate group of control subjects needs to be identified. This will save a lot of time and expense and no case needs be dropped from a study because an adequate control could not be found. Also, there is no possibility of bias in the selection of control subjects, although bias could occur in the selection of control periods.
2) Case and control data are from the same subject. This within-subject comparison automatically matches on all characteristics that do not change within individuals. As a result, proper (matched) analysis of case-crossover data will control for all such non-varying (fixed) confounders. This advantage solves the vexing difficulty of controlling for confounders that are unknown or not measurable in the conventional case-control study.

The case-crossover design, on the other hand, has the following limitations:

1) The exposure must vary over time within individuals rather than stay constant. If the exposure does not vary within an individual, then there is no basis for comparing exposed and unexposed time periods of risk within the individual. For instance, if

blood type is assumed to be associated with some kind of outcome, this assumption can not be tested by the case-crossover design because the exposure is constant for every individual, and in this case, external controls must be sampled for the study. For the same reason, in pharmacoepidemiology, the case-crossover design is not applicable to drugs with regular patterns of use that vary only minimally within an individual.

2) Time window or effect period needs to be well defined. This is the main reason that the case-crossover design has been applied to a limited number of studies in which the effect time of exposure is very short and is easily defined. If there are carry-over effects and the time of these carry-over effects are unknown, the case-crossover design is not applicable because of the difficulty of defining the time window. The application of the case-crossover design can be extended from short transient exposures to longer chronic exposures as long as the time window can be reasonably defined. In this situation, a time trend effect may confound the result and if it occurs, case-time-control design is an option.
3) Within-individual confounding threatens the validity of the design. Using the subjects as their own controls eliminates confounding by subject characteristics that remain constant, but not by characteristics that change over time. Time trend effect is also a within-individual confounding that happens when the time window is long.

In conclusion, the case-crossover design has provided a new technique to study the effect of transient exposures on acute outcome. Its application is restricted to a narrow range of

scientific questions, but when it is applicable, it can ingeniously solve some problems that are challenging by conventional design.

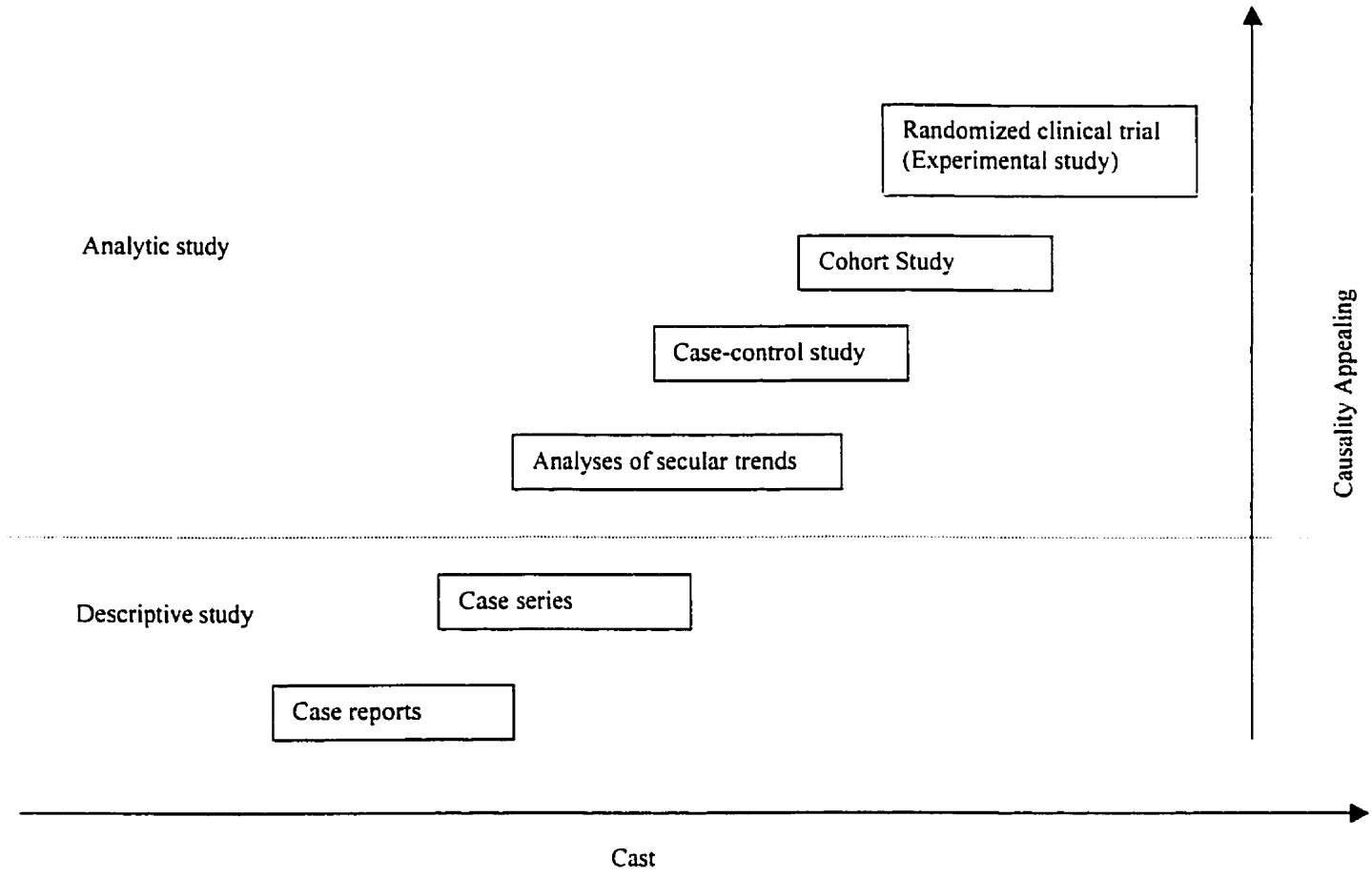


Fig 2.1 Summary of conventional study designs in pharmacoepidemiology.

# CHAPTER 3 DATA FOR ILLUSTRATION

### 3.1 Saskatchewan Health Databases

In this study, the data of exposure history and other information for each case of asthma death came from the computerized databases of Saskatchewan Health.

Saskatchewan is one of 10 provinces in Canada. It has a population of about one million people, which is 4% of the total population of Canada. For more than 20 years, the Saskatchewan Department of Health has been accumulating a very large amount of heath care information in computerized databases. These databases record all kinds of health services for each individual such as, hospitalization, outpatient drug prescription, and physician services. In addition, the Vital Statistics Division of Saskatchewan Health also provides death information such as the time and the course of death.

The main strength of the Saskatchewan Health databases is the fact that 95 percent of the population in the province is covered by the universal health program. This fact, plus the more than 20 years of accumulation of data, makes the Saskatchewan Health databases a unique model for pharmacoepidemiologic study in the entire population. Another strength of the Saskatchewan Health databases is the fact that individuals are identified by an unique Health Service Number (HSN). This number is used to code almost all healthcare services and hence can be used to electronically link data from any of the computerized databases.

The databases of Saskatchewan Health used in this study are the Health Insurance Registration File, the Prescription Drug Data and the Hospital Inpatient Data.

Health Insurance Registration File (HIRF) is a registry of all Saskatchewan residents who have registered with Saskatchewan Health for a Health Services Card. This file contains sociodemographic information for 95 percent of the Saskatchewan population such as name, address, sex, date of birth, dates of effective coverage and family status. The HSN of this file allows the linkage of this file with all the other computerized files.

Prescription Drug Data records all prescriptions of medications dispensed out of hospitals to residents who have a valid Health Services Card, thus eligible for benefits under the Prescription Drug Plan, with the exception of about 6% who have their prescription costs paid for by another government agency. The prescription drug data from September 1975 to June 1987, and from January 1989 to date, includes the information of each individual's HSN, the identification number (DIN) of drug dispensed, the quantity of the drug dispensed, and the date of dispensing. Data between July 1,1987 and December 31, 1998 are not available on an individual basis. The HSN can link this data file to the other health data files and DIN can be matched to information about the brand, strength, pharmaceutical preparation, and manufacturer through linkage with a "drug rules" file.

The Hospital Services Data includes the HSN, date of birth, date of discharge, length of stay, and diagnostic and treatment information (e.g., discharge diagnoses and primary surgical procedure). Discharge diagnostic data are coded for both primary and secondary diagnoses by using the International Classification of Disease (ICD).

As the information about history of drug exposure is from the Prescription Drug Data and this database only records the outpatient prescription, the database can not provide valid data for the patients hospitalized for a long period of time, so we used the Hospital Serviced Data for checking the hospitalization history of each subject and the subjects with long time hospitalization should be excluded from the study.

### 3.2 Source population

The source population is defined as all beneficiaries aged between 5 and 44 years old who received at least one prescription of an anti-asthma medication between September 1975 and December 1991. The anti-asthma medications include all anti-asthma medications covered by the health insurance plan, such as beclomethasone, budesonide, triamcinolone acetate, flunisolide, sodium cromoglycate, ketotifen, nedocromil, salbutamol, fenoterol, terbutaline, isoproterenol, metaproterenol, procaterol, epinephrine bitartrate, ipratropium bromide, and any compound of theophylline. Oral corticosteriods were not included in the list because the prescription of this drug alone may be given to patients with some disease other than asthma. In the Prescription Drug Data, each of the drugs listed above has a specific code. By linking this database with the HIRF file, 31,307 subjects were identified to form the source population.

### 3.3 Case Identification

The source population was followed from 1977 to 1993, or until age 55, and a total of 467deaths of all causes occurred during this follow up. Death certificates for 427 deaths

were obtained from Saskatchewan Health and were reviewed blindly and independently by two respirologists to determine the cause of death. To determine whether a death was caused by asthma, a procedure was carefully predefined and the main criteria were: a) whether asthma was the cause or a direct antecedent cause as listed in Part 1 of the death certificate, or the course of death listed in Part 1 suggested that the patient died of some disease other than asthma; b) whether Part 2 of the death certificate mentioned asthma as another significant condition while the cause listed in Part 1 appeared to be a direct consequence of asthma; c) whether the course of death was suggestive of processes which may accompany an acute exacerbation of asthma. In addition, other evidence such as clinical note, coroner's assessment, macroscopy and histological descriptions were also reviewed if it was necessary to confirm the classification of the cause of death. Any discrepancy was resolved on the basis of consensus.

After the subjects who died of asthma were identified, further including criteria were then established as: a) information about drug prescriptions of at least 365 days before death was available and b) there was at least one prescription of beta-agonists in the 365 days before death. With these criteria, we excluded the cases that did not have sufficient exposure data and the cases that did not provide any information for the estimate.

### 3.4 Exposure history of the cases

For each death identified above, the HSN was obtained and the Prescription Drug Data was linked by this number to obtain all records of drug prescriptions for these cases. A smaller data set was then built by including only the records of prescriptions dated within

365 days before death and including only the records of prescriptions of beta-agonists. The beta-agonists include terbuterline, isoproterenol, metaproteranol, salbutamol, fenoterol, and procaterol. As the data between July 1,1987 and December 31, 1998 were not available on individual basis in this Prescription Drug Data, The death date of the study cases must have been before, or one year after, this "dark period".

The HSN number for each cases was also used to link Hospital Services Data and the Health Insurance Registration File to obtain hospitalization information and sociodemographic information for each case.

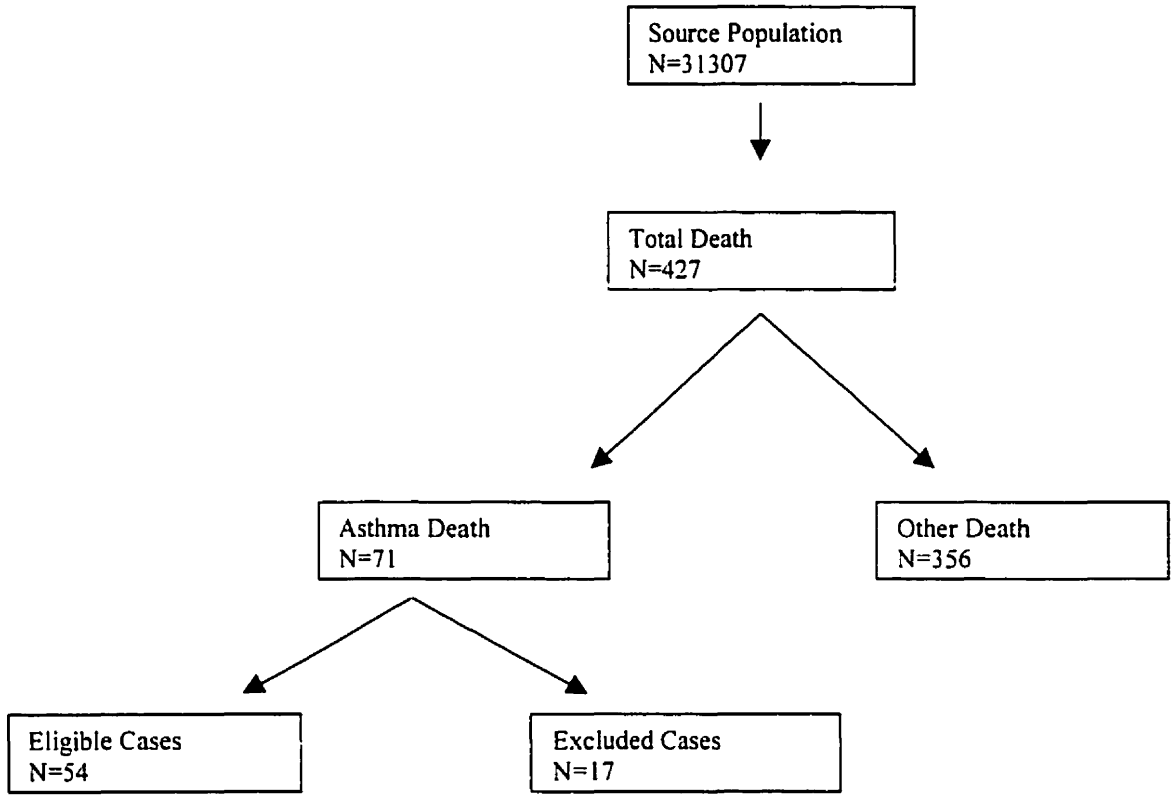


**Fig 3.1** Selection of the study cases.

# CHAPTER 4 METHOD

## 4.1 New estimation based on Poisson distribution

In general, the odds ratio can be written as,

$$OR = \frac{P(E|Case)/P(\bar{E}|Case)}{P(E|Control)/P(\bar{E}|Control)}$$

$$= \frac{\left[1 - P(\bar{E}|Case)\right]/P(\bar{E}|Case)}{\left[1 - P(\bar{E}|Control)\right]/P(\bar{E}|Control)} \qquad (4.1)$$

From the above equation we can see that the main task of a case-control study is to compare the exposure distribution of the cases to that of the controls. Under the null hypothesis, the distribution of exposure for the cases and for the controls is the same, so that OR is 1. OR>1 means the probability of exposure is larger in cases than in controls, implying that the drug is harmful. On the contrary, OR<1 means the treatment is protective. As discussed in Chapter 1 and Chapter 2, the case-crossover study design uses the cases themselves as controls and compares the distribution of exposure of the subjects when the event occurs to the distribution of exposure of the same subject when the event did not happen to them.

In classical case-crossover design, the statistical technique provided to estimate the case-crossover rate ratio and its confidence limits is based on the binomial distribution for exposure. Thus, exposure data in each pre-specified time window is dichotomized as exposure or non-exposure. Considering that a database study in pharmacoepidemiology can provide precise data on the actual exposure times, it is assumed that some

information may be wasted by simply reducing exposure data into the dichotomized form required for the binomial distribution.

### 4.1.1 Estimation using exposure times

With the database providing precise dates of drug prescriptions, we can obtain data of precise inter-exposure onset times. We assume that the number of prescriptions during a given time interval follows a Poisson process, that is, the distribution of inter-exposure onset time is an exponential distribution.

Suppose that the distribution of inter-exposure onset time for cases is exponential ($\lambda_1$) and the distribution of inter-exposure onset times for the controls is exponential ($\lambda_0$), then following 4.1, and the fact that $P(E) = 1 - P(\bar{E}) = 1 - e^{-\lambda t}$, we have,

$$OR = \frac{(1 - e^{-\lambda_1 t}) / e^{-\lambda_1 t}}{(1 - e^{-\lambda_0 t}) / e^{-\lambda_0 t}}$$

$$= \frac{e^{\lambda_1 t} - 1}{e^{\lambda_0 t} - 1} \qquad (4.2)$$

Applying the Delta method, the estimator of the variance of OR can be derived as follow,

$$\frac{\partial OR}{\partial \lambda_1} = \frac{t e^{\lambda_1 t}}{e^{\lambda_0 t} - 1} \qquad (4.3)$$

$$\frac{\partial OR}{\partial \lambda_0} = -t\, e^{\lambda_0 t}\frac{e^{\lambda_1 t}-1}{\left(e^{\lambda_0 t}-1\right)^2} = -\frac{t\, e^{\lambda_0 t}}{e^{\lambda_0 t}-1}OR \tag{4.4}$$

$$Var(OR) \approx \left(\frac{\partial OR}{\partial \lambda}\right)\sum\left(\frac{\partial OR}{\partial \lambda}\right)'$$

where,
$$\sum = \begin{pmatrix} Var\,\hat{\lambda}_1 & Cov \\ Cov & Var\,\hat{\lambda}_0 \end{pmatrix}$$

Therefore,

$$Var(\hat{OR}) = t^2\left(\frac{e^{\lambda_1 t}}{e^{\lambda_0 t}-1}\right)^2 Var\,\hat{\lambda}_1 + \frac{t^2 e^{2\lambda_1 t}}{\left(e^{\lambda_0 t}-1\right)^2}OR^2 Var\,\hat{\lambda}_0$$

$$-\frac{2t^2 e^{2\lambda_0 t}}{\left(e^{\lambda_0 t}-1\right)^2}ORCov \tag{4.5}$$

Assuming cov=0, we have,

$$Var(\hat{OR}) \approx \frac{t^2}{\left(e^{\lambda_0 t}-1\right)^2}\left(e^{2\lambda_1 t}Var\,\hat{\lambda}_1 + e^{2\lambda_0 t}OR^2 Var\,\hat{\lambda}_0\right) \tag{4.6}$$

Moreever, using the logarithmic transformation for the ratio to avoid problems of asymmetry, we have,

$$Var(\log \hat{OR}) \approx \left(\frac{\partial \log OR}{\partial OR}\right)^2 Var\hat{OR}$$

$$= \frac{t^2}{\left(e^{\lambda_1 t}-1\right)^2}\left(e^{2\lambda_1 t}Var\,\hat{\lambda}_1 + e^{2\lambda_0 t}OR^2 Var\,\hat{\lambda}_0\right) \tag{4.7}$$

For Poisson model, we have,

$$Var\,\hat{\lambda}_j = N_j/T_j^2 = \lambda_j/T_j \qquad (4.8)$$

Where j=0 for controls and 1 for cases, $T_j$ is the total duration of time and $N_j$ is the total number of prescriptions in $T_j$ time period.

### 4.1.2 Variance-weighted average estimator

We can consider each case as a stratum where the controls are taken from a different time period. Thus, we have n strata corresponding to the n cases along with their own control times. To obtain an overall estimate of OR, we can either use standardization or pooling. Here we used pooling in which the weights were determined solely by the data and the homogeneity assumption. To assign the weights that reflect the amount of information in each stratum, and to minimize the variance of the overall weighted average without introducing bias, the weights should be assigned to the stratum specific values as inversely proportional to the estimated variance of each stratum-specific estimator. Thus, we pooled the n log (OR), by a weighted average, weighting by the inverse of the variance of log OR.

Let,

$$\text{Var}\,(\log(\,OR_i\,)) = V_i\,, \qquad (4.9)$$

$$\widehat{\log OR} = \frac{\sum_{i=1}^{n} \frac{1}{V_i} \log OR_i}{\sum_{i=1}^{n} \frac{1}{V_i}} \qquad (4.10)$$

Then we have,

$$\widehat{\mathrm{Var}(\log OR)} = \left( \frac{1}{\sum_{i=1}^{n} \frac{1}{V_i}} \right)^2 \sum_{i=1}^{n} \frac{1}{V_i^2} \mathrm{Varlog}\, OR_i$$

$$= \left( \frac{1}{\sum_{i=1}^{n} \frac{1}{V_i}} \right)^2 \sum_{i=1}^{n} \frac{1}{V_i^2} V_i$$

$$= \frac{1}{\sum_{i=1}^{n} \frac{1}{V_i}} \qquad (4.11)$$

Here, n is the number of cases.

For the equation above, we can use the number of prescriptions in the last time window period divided by the length of the time window ($T_1$) as an estimator of $\lambda 1$, and the number of prescriptions before the last time window divided by the length of control time ($T_0$) as an estimator of $\lambda_0$. That is,

$$\hat{\lambda}_0 = \frac{\text{\# of Rx befor the time window period}}{T_0} \qquad (4.12)$$

$$\hat{\lambda}_1 = \frac{\text{\# of Rx during the last time window period}}{T_1} \qquad (4.13)$$

For the study of the association of beta-agonists with asthma death, it is assumed that the effect time window is 30 days. For each case, we divided the one year period before asthma death into case periods and control periods. The case period is from 30 days before death to the date of death, that is, the last time window period. The control period is from 365 days before death to the date of 30 days before death. The number of beta-agonist prescriptions in the case-period and the control period were counted and $\lambda_0$ and $\lambda_1$ were estimated by equation (4.12) and (4.13) respectively. In the equation (4.6) and (4.7). $\lambda_0$ and $\lambda_1$ cannot be 0, but in practice, it happens that some patients do not have any prescriptions either in the last time window period or in the control period, which makes $\hat{\lambda}_1$ and $\hat{\lambda}_0$ equal to 0 and the estimate of log(OR) and Var(logOR) infinite. To avoid introducing bias by discarding these cases, we used the reciprocal of time duration from the date of last prescription to the date of death as an estimate of $\lambda_1$ if the number of prescriptions in the last time window period was 0. We used the reciprocal of time duration from the date of 365 days before death to the date of first prescription as an estimate of $\lambda_0$ if the number of prescriptions was 0 in the control period.

### 4.1.3 Unweighted estimator

Another way to estimate the overall odds ratio is to separately pool $\lambda_0$ and $\lambda_1$ with equal weight across subjects and to use the overall $\lambda_0$and $\lambda_1$ to estimate OR:

$$\hat{\lambda}_j = \frac{\sum_{i=1}^{n} \hat{\lambda}_{ij}}{n} \qquad (4.14)$$

$$Var\,\lambda_j = Var\frac{\sum_{i=1}^{n}\hat{\lambda}_{ij}}{n}$$

$$\approx \frac{1}{n^2}\sum_{i=1}^{n} Var\hat{\lambda}_{ij}$$

$$= \frac{1}{n^2}\sum_{i=1}^{n}\frac{\hat{\lambda}_{ij}}{T_i} \qquad (4.15)$$

Then the overall estimate of odds ratio is,

$$OR = \frac{e\hat{\lambda}_1 t - 1}{e\hat{\lambda}_0 t - 1} \qquad (4.16)$$

The VarlogOR can be derived by equation (4.7) and (4.15). For this method, $\hat{\lambda}_1$ and $\hat{\lambda}_0$, for some subjects were allowed to be 0, so that the adjustment in the previous method is not necessary.

### 4.2 Classic estimation based on binomial distribution

To compare our new method with the classical method based on the binomial distribution, we also calculated the OR and Variance of OR in the same way as originally proposed for the case-crossover design (Maclure, 1991). In this method, there are two elements for the estimation of RR and its variance, $P_{xi}$ and $C_{xi}$.

$P_{xi}$ is the probability that the ith subject was in a state of altered risk due to exposure x. There are two ways to calculate $P_{xi}$. First, the reference period is divided into time

windows, and $P_{xi}$ is calculated as the number of time windows the subject i is exposed divided by the total number of time windows in the reference period for the subject i. For the second way, the $P_{xi}$ is calculated as the proportion of time the subject is exposed to the total time of the reference period.

$C_{xi}$ is an indication of whether a case are classified as exposed ($C_{xi}$=1) or not ($C_{xi}$=0) during the assumed effect-period.

It is assumed that the time window is 30 days, which means that if a patient was given a prescription of beta-agonists, the patient was exposed to the drug for 30 days following the prescription. This effect-period is rather long as compared with the frequency of prescriptions, so there are overlaps among the effect-periods. As proposed by Mclure, this overlap should be discounted.

Occasionally a patient may have more than one prescription of beta-agnoists on the same day. If this happens and we calculate the $P_{xj}$ as above, the effect-period of one prescription on one day will be the same as two or three prescriptions on that same day, which is not reasonable. To adjust for this, we assumed that if there is more than one prescription on the same day, these prescriptions are evenly distributed in the period between this day and next the prescription day.

To determine the value of $C_{xi}$, we just need to know the date of the last prescription of beta-agonists. If the date of last prescription of beta-agonists is less then 30 days before the date of death, then the case is defined as exposed and $C_{xi}$=1. Otherwise, the case is

defined as not-exposed and $C_{xi}$=0. The way of calculating $P_{xi}$ and $C_{xi}$ is illustrated in Fig. 4.1.

When the $P_{xj}$ and $C_{xj}$ are defined, the Mantel-Haenszel estimator of rate ratio can be calculated by the formula,

$$RR_{MH} = \frac{\sum_{i=1}^{n} C_{xi}(1 - P_{xi})}{\sum_{i=1}^{n} (1 - C_{xi}) P_{xi}} \tag{4.17}$$

The variance of the $RR_{MH}$ can be calculated by the formula,

$$Var\left[\log(RR_{MH})\right] = \frac{\sum_{i=1}^{n} P_{xi}(1 - P_{xi})}{\left[\sum_{i=1}^{n} C_{xi}(1 - P_{xi})\right]\left[\sum_{i=1}^{n} (1 - C_{xi}) P_{xi}\right]} \tag{4.18}$$

A confidence interval can be estimated according to this estimate of variance under the assumption that $\ln(RR_{MH})$ is normal distributed.

### 4.3 Simulation

To assess the performance of the different estimators mentioned above and to compare the efficiency of the new estimators with the classical estimator of the case-crossover design, a simulation study using Monte Carlo methods was conducted.

In the simulation study, the events were generated under the following assumptions:

1) The events of prescriptions in the case periods and the control periods are Poisson distributed with rates $\lambda_1$ and $\lambda_0$ respectively, that is, the lengths of time between the events follow exponential distributions.
2) The rate of exposure during the control period varies among subjects, but all cases have a common odds ratio. That is, the rate of exposure for the case time-window is adjusted accordingly.

With these two assumptions, the following steps were conducted:

1) The number of cases, n, and $E(\lambda_0)$ were fixed as a priori, and then n $\lambda_0$s were generated from a gamma distribution.
2) With a predetermined odds ratio, $\lambda_1$ for each case was determined by its $\lambda_0$ and the predetermined OR.
3) For the case period, an event was generated with the time interval between the event and death to be exponential ($\lambda_1$). Then another event was successively generated, with the time interval between this event and the previous event to be exponential ($\lambda_1$), until the cumulated time reached the predetermined time window, and the number of events were counted. Events in the control period were also generated by the same process with exponential ($\lambda_0$) distribution.
4) After a set of data with n cases was generated, odds ratios were estimated from the different estimators given above.

Suppose R is an estimator of the odds ratio, $R_k = f(\lambda_k)$, where $\lambda_k$ is the generated matrix of $\lambda$, and k=1,2,... ,N represents the N simulations. Then we have,

$$\overline{R} = \frac{1}{N} \sum_{k=1}^{N} \hat{R}_k \tag{4.19}$$

By the Law of Large Numbers,

$$\overline{R} \xrightarrow{a.s.} E(R) \tag{4.20}$$

The standard error of $\overline{R}$ is thus

$$\sigma_{\overline{R}} = \sigma / \sqrt{N} \tag{4.21}$$

As we don't know the variance of R, it was estimated by the formula,

$$S^2 = \frac{1}{N-1} \sum_{k=1}^{N} \left( R_k - \overline{R} \right)^2 \tag{4.22}$$

To compare the efficiency of the two estimator Ra and Rb, by definition, the relative efficiency is,

$$\frac{Var\left(\hat{R}_a\right)}{Var\left(\hat{R}_b\right)} \approx \frac{{S_a}^2}{{S_b}^2} \tag{4.23}$$

The simulations study was performed with SAS macro language. We created three macros to fulfill a simulation process. The first macro was used to generate a data set as above. In this macro, the presumed time window, odds ratio (OR), $\lambda 0$, and seed for generating random variables were passed as macro parameters, so these variables could be easily changed when recalling this macro. The second macro was to calculate

log(OR), OR and Variance of log(OR) for a data set generated by the first macro. The name of data set was passed as parameter, so this macro can not only used for this simulation study, but also can be used independently to calculate OR and Variance for a specific real data set. The third macro was to iterate recalling the first two macros and to generate a data set by concatenating the results from recalling the second macro. For each iterating, the seed was automatically incremented by 1 so that different data sets were generated for each iterating.

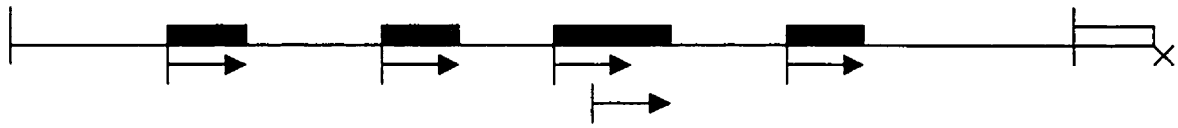

**Fig 4.1** Calculation $P_{xi}$ and $C_{xi}$ in Mantel-Haenszel estimator. The arrow represents the effect period of each prescription; the filled block represents the time the subject is at altered risk due to the exposure; the empty block represents the last time window; the X represents event, and the $C_{xi}$ is determined by whether or not there is a prescription of beta-agonists in this time window.

# CHAPTER 5 RESULTS

## 5.1 General description of data

In this study, the source population was all beneficiaries of the Saskatchewan Health Program aged between 5 and 44 years old who received at least one prescription of an anti-asthma medication between September 1975 and December 1991. This source population was followed from July 1977 to December 1993 or until age 55. The subjects were censored at age 55 because chronic obstructive pulmonary disease usually becomes symptomatic around this age, which makes the diagnosis of asthma uncertain. During this follow-up, a total of 467 deaths of all causes were identified by the Health Insurance Registration File (HIRF). The death certificates of 427of them (91.4%) were reviewed and 71 cases were identified as due to asthma death (16.9%). Because the Prescription Drug Data were not available on an individual basis between July 1, 1987 and December 31, 1988, the follow up time was divided into 4 calendar periods and the number of cases of asthma death in each period was shown in Fig 5.1.

The 14 cases in periods 2 and 3 were excluded because there was less than one year of prescription data for these cases. Among the remaining 57 cases, we excluded 3 cases who did not have any prescriptions of beta-agonists in the year before death and thus would not provide any information for the estimate of the odds ratio. As a result, our final study case series was 54 asthma deaths.

Table 5.1 summarizes the distribution of the deaths of all causes, of the total asthma deaths and of the 54 study cases. The data show that the proportion of males in the 427 deaths of all causes and in the 71 asthma deaths were 61.4 and 57.1 respectively. The

mean age at death for asthma was 30.1 and the mean age of death for all causes was 52.7. The proportion of young people (age<15) who died of asthma and died of all causes was 16.9% and 6.4% respectively. The time of asthma onset was defined as the time that subjects entered the asthma cohort by the three prescription criteria, that is, the time they received three asthma drug prescriptions in a consecutive 365 days after the first prescription of asthma drugs. The mean age at onset for the people who died of asthma was 23.3, and for the people who died of all causes, it was 30.0.

As we have mentioned before, we could only obtain information on outpatient prescriptions from the Saskatchewan Health databases. Thus, we needed to obtain the hospitalization information for the study cases to check the validity of the data for the drug use. The Hospital Service Data provided the date of discharge, length of stay and ICD-9 code for diagnostic and treatment information for our study cases. Table 5.2 summarizes this data for the 365 days before death. The table shows that patients had 1.7 hospitalizations on average in the year preceding deaths and most of the hospitalizations were for asthma (88.2%). The average duration of hospital stay for the 54 cases was 8.4 days in a year, which was a very small portion (2.3%) of total exposure time. However, the maximum hospital duration in a year was 57 days, which was 15.6% of total exposure time, and there were 5 of the 54 cases who had hospitalization duration over 30 days. Thus, the analysis was repeated after excluding these 5 cases.

We identified all inhaled beta-agonist prescriptions for the study cases. During the 365 days before death, the 54 study subjects received a total of 959 beta-agonist prescriptions.

and these 959 records accounted for 1558 canisters of beta-agonists (1.6 canisters/prescription). It was found that the majority of beta-agonist prescriptions for these 54 study cases in the year before death were for salbutamol and fenoterol (93.4%). Only one patient used metaproteranol as well as salbutamol. There were 6 people who used both salbutamol and fenoterol. No other beta-agonists, such as terbutaline, isoproterenol and procaterol were used by these 54 study cases (Table 5.3). It is important to note that fenoterol was prescribed in 38.9% of patients who died asthma, but the market share of this drug is only about 10%, which suggests that fenoterol may be associated with asthma death.

## 5.2 Estimate of OR and confidence interval

### 5.2.1 New Method: Poisson distribution estimation

Under the Poisson distribution assumption, we considered each case as a stratum and used two different approaches to obtain the pooled estimate of the odds ratio and its confidence interval. The first approach was to average the log(OR), weighted by the inverse of the variance of the log(OR). The estimate of the log(OR) and the variance of log(OR) are given by equations 4.5-4.11. The second approach was to pool $\hat{\lambda}_1$s and $\hat{\lambda}_0$s with equal weight to estimate the overall $\lambda_1$ and $\lambda_0$ (equation 4.14-4.16).

Table 5.4a provides the estimates of $\lambda_1$ and $\lambda_0$ under different time window assumptions.

It shows that $\hat{\lambda}_0$ remains stable when the time window changes, but $\hat{\lambda}_1$ increases as the length of the time window decreases. It also shows that $\hat{\lambda}_1$ is larger than $\hat{\lambda}_0$. Table 5.4b lists the same data as Table 5.4a, except that the adjustment was done to avoid $\hat{\lambda}_{11}$ and $\hat{\lambda}_{10}$ being 0, as described in the method section. In Table 5.4b, $\hat{\lambda}_0$ also remains stable when the time window changes and it is very close to that in Table 5.4a, and $\hat{\lambda}_1$ also increase as the length of time window decreases. The $\hat{\lambda}_1$ in Table 5.4b is larger than $\hat{\lambda}_1$ in Table 5.4a and this difference became bigger as the size of the time window decreased. This is because more cases were needed to adjust $\hat{\lambda}_1$ in order to avoid it being 0 when the size of the time window decreases.

The odds ratios and the confidence intervals were calculated as shown in Table 5.5. When the time window was assumed to be 30 day, the OR was 1.58 (95% CI: 0.98-2.57) for the weighted method and 1.54 (95% CI: 0.99-2.41) for the unweighted method. When the time window was assumed to be 25 days, the OR increased to 1.92 (95% CI: 1.19-3.09) for the weighted method and 1.81 (95% CI: 1.17-2.81) for the unweighted method, while the 95% CIs no longer included 1. When the assumed time window decreased further, the odds ratio monotonely increased for both methods, and the highest estimate of odds ratio was obtained when the time window was 10 days, the smallest time window tested. The highest estimate of OR obtained was 2.90 (0.95 CI 1.81-4.67) for the weighted method and 2.46 (0.95 CI 1.60-3.79) for the unweighted method.

The five cases with hospitalization durations longer than 30 days in the year before death may challenge the validity of our estimation. We repeated the same analyses when these 5 cases were removed and the results are shown in Table 5.6. After removing the 5 cases, no major difference was found for the estimate of OR under any time window assumption. However, we can see that the estimation of OR was uniformly higher when the 5 cases were excluded. This is because the patients tended to have more hospitalizations just before death, so that the number of prescriptions of any drug will be underestimated in the period just before death.

### 5.2.2 Classic method: Binomial distribution

We analyzed the same data to estimate the RR using Maclure's method and the results are shown in Table 5.7. In this analysis, the estimates were given by equations (4.17) and (4.18). In Table 5.7, Methods 1 and 2 refer to different ways of calculating the $P_{xi}$. For Method 1, $P_{xi}$s were calculated as proportion of alerted risk time to the whole reference period of time. For Method 2, the $P_{xi}$s were calculated as the proportion of the number of alerted risk time windows to the total number of time windows in the reference period of time, as described in Chapter 4.

In Method 1, a RR of 4.57 (95% CI: 1.60-13.02) was obtained when the time window was assumed to be 30 days. This is much higher than the estimate with the Poisson approach, and the confidence interval is much wider. When the time window decreases to 10 days, the estimate of OR is 2.62 (0.95 CI 1.45-4.70), an estimate very close to the result of the Poisson approach. The estimates of variance are also close for the two

approaches when the time window decreases to 10 days. We can see from Table 5.7 that the two methods of calculating $P_{xi}$ give very similar results.

### 5.3 Simulation

Using the approach described in the Methods section, we simulated data based on a one-year exposure, a time window of 30 days, and mean $\lambda_0$ of 1/30day, 1.5/30day, and 2/30day. The true OR varied from 0.5 to 5.0, increasing with step of 0.5, and we considered two sample sizes, n=60 and n=300. All simulations were based on 1000 replications.

Table 5.8 lists the simulation results when the number of cases was set as 60. In this table, Mean log(OR) represents E(logOR) and Mean VarlogOR represents E(VarlogOR), as demonstrated by equation (4.20). Var(logOR) is $S^2$ as in equations (4.22) and (4.23). In the binomial distribution method, $P_{xi}$ was calculated as the proportion of alerted risk time to the whole reference period of time. The Poisson distribution method 1 is to pool log(OR)s by the weight of the inverse of the variance of the log(OR) and the Poisson distribution method 2 is to pool $\hat{\lambda}_{i1}$ and $\hat{\lambda}_{i0}$ with equal weights.

We can evaluate the accuracy of the estimation of log(OR) by comparing Mean log(OR) with True log(OR) and evaluate the accuracy of the estimation of the variance of the log(OR) by comparing Mean Varlog(OR) with Varlog(OR).

The data in Table 5.8 show that for the binomial distribution method and the unweighted Poisson distribution method, the estimates of the variance of log(OR) are very close to the "true" variance of log(OR). On the other hand, for the weighted Poisson distribution method, the estimates of the variance of log(OR) are overestimated.

For the binomial distribution method, the estimate of log(OR) is uniformly higher than the true log(OR) and the bias increases as $\lambda_0$ increases from 1/30day to 2/30day. For the Poisson distribution methods, the direction of bias is more complicated and it varies according to both $\lambda 0$ and true OR. In order to reveal the accuracy of the different methods of estimation, we plotted the estimated log(OR) to the true log(OR). Fig 5.2 shows the plot when $\lambda_0$ is 1.5/30day. We can see from this graph that when the true OR is 0.5-2.5, the Unweighted Poisson method uniformly gives the best estimate. The estimate of log(OR) by weighted Poisson method is severely biased towards the null when the true OR is less than 1.

RE1 and RE2 in Table 5.8 are the relative efficiencies of the Poisson distribution methods 1 and 2 to the binomial distribution. It was calculated by equation (4.23) as the ratio of Var(logOR). The relative efficiency can be simply interpreted as the proportion of the sample size needed for the two estimators if the variances of the two estimators are equal (Suissa, 1991). The table shows that the relative efficiency for the weighted Poisson method is 0.16-0.3, which means less than one third of the sample size is needed by this method to achieve the same precision as the binomial distribution method. For the unweighted Poisson method, the relative efficiency is 0.2 –0.50, which means less than

half of the sample size is needed for this method to achieve the same precision as the binomial distribution method.

Fig 5.3 plots the variance of log(OR) against the true OR when λ0 is 1.5/30days. It shows that the variance of log(OR) for the binomial distribution method increases as the true OR increases from 0.5 to 5.0, but the variance of log(OR) for the Poisson method is stable as the true OR changes. It also shows that the weighted Poisson method has the smallest variance of logOR.

To illustrate the performance of the binomial method estimator compared to that of the Poisson method, we compared the histograms of OR estimates from the two approaches. Fig 5.4 gives an example comparing the binomial method with the weighted and unweighted Poisson methods when the true OR=2 (logOR=0.69), λ0 is 1.5/30days, and with 1000 simulations. Compared to the Poisson distribution method, the distribution of the estimated log(OR) by the binomial method more dispersed and skewed.

To test the effect of sample size on the performance of these estimators, we conducted another simulation with the sample size increased by five-fold, that is, the sample size was increased from 60 to 300. The results are listed in Table 5.9. It shows that when the sample size was increased by five-fold, the variance of the estimated log(OR) decreased proportionately for all three methods and that the relative findings of all three methods remained the same.

Table 5.1. Population distribution of deaths

| | All deaths N=427 | Asthma deaths N=71 | Eligible cases N=54 |
|---|---|---|---|
| | Number (percent) | | |
| Male | 262 (61.4) | 41 (57.7) | 33 (61.1) |
| Age>15 | 403 (94.6) | 59 (83.1) | 43 (79.6) |
| Age at Onset >15 | 340 (79.6) | 45 (63.4) | 35 (64.8) |

Table 5.2 Number of records and duration of hospitalization.

| | Number of Records | | | Duration (day) | | |
|---|---|---|---|---|---|---|
| | mean | min | Max | mean | min | max |
| Total Hospitalization | 1.7 | 0 | 10 | 8.4 | 0 | 57 |
| Asthma Hospitalization* | 1.5 | 0 | 10 | 7.1 | 0 | 46 |
| Other Hospitalization | 0.2 | 0 | 3 | 1.4 | 0 | 27 |

* Identified by ICD-9 code.

Table 5.3 Distribution of the use of different beta-agonists .

| | Total Number | | |
|---|---|---|---|
| | Prescriptions | Canisters | Users* |
| Total | 959 | 1558 | 54 |
| Terbutaline | 0 | 0 | 0 |
| Isoproterenol | 0 | 0 | 0 |
| Metaproteranol | 63 | 109 | 1 |
| Salbutamol | 418 | 662 | 39 |
| Fenoterol | 478 | 787 | 21 |
| Procaterol | 0 | 0 | 0 |

* There were 6 people use both salbutamol and fenoterol, the the patient using metaproteranol was also a salbutamol user.

Table 5.4a Estimate of number of beta-agonists prescriptions per 100 days for different time window assumption *

| | Reference period | | | Risk period | | |
|---|---|---|---|---|---|---|
| Time Window (days) | Mean | Min | Max | Mean | Min | Max |
| 30 | 4.78 | 0 | 20.89 | 5.80 | 0 | 20.00 |
| 25 | 4.77 | 0 | 20.59 | 6.22 | 0 | 20.00 |
| 20 | 4.77 | 0 | 20.29 | 6.57 | 0 | 20.00 |
| 15 | 4.80 | 0 | 20.00 | 6.42 | 0 | 20.00 |
| 10 | 4.81 | 0 | 19.72 | 6.85 | 0 | 30.00 |

* No adjustment was made, allow the rate to be 0.

Table 5.4b Estimate of number of beta-agonist prescriptions per 100 days for different time window assumption*

| | Reference period | | | Risk period | | |
|---|---|---|---|---|---|---|
| Time Window (days) | Mean | Min | Max | Mean | Min | Max |
| 30 | 4.79 | 0.27 | 20.89 | 5.94 | 0.28 | 20.00 |
| 25 | 4.77 | 0.27 | 20.59 | 6.56 | 0.28 | 20.00 |
| 20 | 4.77 | 0.27 | 20.28 | 7.31 | 0.28 | 20.00 |
| 15 | 4.80 | 0.27 | 20.00 | 8.26 | 0.28 | 25.00 |
| 10 | 4.81 | 0.27 | 19.72 | 9.25 | 0.28 | 30.00 |

* The adjustment was made so that the rate can not be 0.

Table 5.5 Estimate of odds ratio (OR) for the asthma death associated with the use of beta-agonists under the Poisson distribution assumption.

| | Weighted * | | | Unweighted** | | |
|---|---|---|---|---|---|---|
| Time window | OR | 95%CI | Var(logOR) | OR | 95%CI | Var(logOR) |
| 30 | 1.58 | 0.98-2.57 | 0.06 | 1.54 | 0.99-2.41 | 0.05 |
| 25 | 1.92 | 1.19-3.09 | 0.059 | 1.81 | 1.17-2.81 | 0.05 |
| 20 | 2.24 | 1.40-3.60 | 0.058 | 2.01 | 1.35-3.20 | 0.048 |
| 15 | 2.57 | 1.61-4.12 | 0.058 | 2.32 | 1.51-3.56 | 0.047 |
| 10 | 2.90 | 1.81-4.67 | 0.058 | 2.46 | 1.60-3.79 | 0.049 |

* logOR pooled by 1/VarlogOR

** $\lambda_{ij}$ was pooled with equal weight for each cases

Table 5.6 Sensitivity study: Estimate OR under the Poisson distribution assumption when the five cases were removed.

| | Weighted * | | | Unweighted ** | | |
|---|---|---|---|---|---|---|
| Time window | OR | 95%CI | Var(logOR) | OR | 95%CI | Var(logOR) |
| 30 | 1.71 | 1.04-2.81 | 0.07 | 1.65 | 1.05-2.58 | 0.05 |
| 25 | 2.05 | 1.25-3.36 | 0.06 | 1.93 | 1.24-3.00 | 0.05 |
| 20 | 2.38 | 1.46-3.87 | 0.06 | 2.20 | 1.42-3.39 | 0.05 |
| 15 | 2.77 | 1.71-4.49 | 0.06 | 2.50 | 1.63-3.83 | 0.05 |
| 10 | 3.06 | 1.89-4.97 | 0.06 | 2.60 | 1.69-4.00 | 0.05 |

* logOR pooled by 1/VarlogOR

** $\lambda_{ij}$ was pooled with equal weight for each cases

Table 5.7 Estimate of odds ratio (OR) for asthma death associated with the use of beta-agonists under the binomial distribution assumption

| | Method 1* | | | Method 2** | | |
|---|---|---|---|---|---|---|
| Time window | OR | 95%CI | Var(logOR) | OR | 95%CI | Var(logOR) |
| 30 | 4.57 | 1.60-13.02 | 0.29 | 4.38 | 1.68-11.43 | 0.24 |
| 25 | 3.23 | 1.44-7.24 | 0.17 | 2.48 | 1.22-5.08 | 0.13 |
| 20 | 2.59 | 1.34-5.02 | 0.11 | 2.21 | 1.18-4.16 | 0.10 |
| 15 | 2.46 | 1.32-4.58 | 0.10 | 2.27 | 1.23-4.18 | 0.10 |
| 10 | 2.62 | 1.45-4.70 | 0.09 | 2.41 | 1.33-4.37 | 0.09 |

* The Px is calculated as proportion of alerted risk time to the whole reference period of time.

** The Px is calculated as the ratio of number of alerted risk time windows to the total number of time window in reference period of time.

Table 5.8 Accuracy of estimation of the different methods for n=60 cases.

| | | | Binomial Distribuiton | | | Poisson Distribution Method 1** | | | Poisson Distribution Method 2*** | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| lamda0 | True OR | True logOR | Mean logOR | Var logOR | Mean VarlogOR | Mean logOR | Var logOR | Mean VarlogOR | Mean logOR | Var logOR | Mean VarlogOR | RE1 | RE2 |
| 0.033 | 0.5 | -0.69 | -0.60 | 0.12 | 0.10 | -0.05 | 0.02 | 0.06 | -0.60 | 0.06 | 0.06 | 0.20 | 0.50 |
| (1/30) | 1.0 | 0.00 | 0.12 | 0.11 | 0.09 | 0.31 | 0.03 | 0.06 | 0.01 | 0.05 | 0.05 | 0.24 | 0.48 |
| | 1.5 | 0.41 | 0.52 | 0.12 | 0.10 | 0.57 | 0.03 | 0.06 | 0.35 | 0.05 | 0.05 | 0.25 | 0.42 |
| | 2.0 | 0.69 | 0.81 | 0.11 | 0.11 | 0.76 | 0.03 | 0.06 | 0.60 | 0.04 | 0.05 | 0.29 | 0.37 |
| | 2.5 | 0.92 | 1.04 | 0.13 | 0.11 | 0.92 | 0.03 | 0.06 | 0.79 | 0.05 | 0.05 | 0.28 | 0.37 |
| | 3.0 | 1.10 | 1.23 | 0.14 | 0.12 | 1.05 | 0.04 | 0.06 | 0.95 | 0.05 | 0.05 | 0.26 | 0.35 |
| | 3.5 | 1.25 | 1.39 | 0.15 | 0.13 | 1.16 | 0.04 | 0.06 | 1.09 | 0.05 | 0.05 | 0.25 | 0.34 |
| | 4.0 | 1.39 | 1.53 | 0.17 | 0.14 | 1.26 | 0.04 | 0.06 | 1.20 | 0.05 | 0.05 | 0.23 | 0.28 |
| | 4.5 | 1.50 | 1.68 | 0.17 | 0.15 | 1.35 | 0.04 | 0.06 | 1.31 | 0.05 | 0.05 | 0.23 | 0.29 |
| | 5.0 | 1.61 | 1.77 | 0.19 | 0.16 | 1.43 | 0.04 | 0.06 | 1.40 | 0.05 | 0.05 | 0.21 | 0.28 |
| 0.050 | 0.5 | -0.69 | -0.56 | 0.10 | 0.10 | -0.27 | 0.03 | 0.06 | -0.57 | 0.05 | 0.05 | 0.26 | 0.48 |
| (1.5/30) | 1.0 | 0.00 | 0.14 | 0.12 | 0.11 | 0.15 | 0.03 | 0.06 | 0.00 | 0.05 | 0.05 | 0.28 | 0.41 |
| | 1.5 | 0.41 | 0.55 | 0.14 | 0.12 | 0.43 | 0.04 | 0.06 | 0.34 | 0.05 | 0.05 | 0.27 | 0.36 |
| | 2.0 | 0.69 | 0.86 | 0.16 | 0.13 | 0.64 | 0.04 | 0.06 | 0.59 | 0.05 | 0.05 | 0.23 | 0.30 |
| | 2.5 | 0.92 | 1.09 | 0.18 | 0.15 | 0.81 | 0.04 | 0.06 | 0.79 | 0.05 | 0.05 | 0.22 | 0.31 |
| | 3.0 | 1.10 | 1.28 | 0.20 | 0.16 | 0.94 | 0.04 | 0.06 | 0.95 | 0.05 | 0.05 | 0.20 | 0.26 |
| | 3.5 | 1.25 | 1.42 | 0.19 | 0.17 | 1.06 | 0.04 | 0.06 | 1.08 | 0.05 | 0.05 | 0.21 | 0.27 |
| | 4.0 | 1.39 | 1.56 | 0.24 | 0.19 | 1.16 | 0.04 | 0.07 | 1.19 | 0.06 | 0.56 | 0.18 | 0.23 |
| | 4.5 | 1.50 | 1.68 | 0.25 | 0.20 | 1.25 | 0.05 | 0.07 | 1.29 | 0.06 | 0.06 | 0.19 | 0.23 |
| | 5.0 | 1.61 | 1.80 | 0.26 | 0.22 | 1.34 | 0.05 | 0.07 | 1.39 | 0.05 | 0.06 | 0.18 | 0.21 |
| 0.067 | 0.5 | -0.69 | -0.55 | 0.13 | 0.11 | -0.40 | 0.03 | 0.06 | -0.55 | 0.04 | 0.05 | 0.23 | 0.34 |
| (2/30) | 1.0 | 0.00 | 0.16 | 0.14 | 0.13 | 0.04 | 0.04 | 0.06 | 0.00 | 0.05 | 0.05 | 0.27 | 0.37 |
| | 1.5 | 0.41 | 0.56 | 0.02 | 0.14 | 0.33 | 0.04 | 0.06 | 0.34 | 0.05 | 0.05 | 2.50 | 3.37 |
| | 2.0 | 0.69 | 0.87 | 0.18 | 0.16 | 0.55 | 0.04 | 0.06 | 0.58 | 0.05 | 0.05 | 0.23 | 0.31 |
| | 2.5 | 0.92 | 1.11 | 0.21 | 0.18 | 0.71 | 0.04 | 0.07 | 0.77 | 0.05 | 0.06 | 0.19 | 0.25 |
| | 3.0 | 1.10 | 1.29 | 0.23 | 0.20 | 0.86 | 0.04 | 0.07 | 0.92 | 0.06 | 0.06 | 0.19 | 0.25 |
| | 3.5 | 1.25 | 1.45 | 0.15 | 0.22 | 0.98 | 0.05 | 0.07 | 1.06 | 0.06 | 0.06 | 0.31 | 0.41 |
| | 4.0 | 1.39 | 1.59 | 0.27 | 0.24 | 1.09 | 0.05 | 0.07 | 1.18 | 0.06 | 0.06 | 0.18 | 0.23 |
| | 4.5 | 1.50 | 1.71 | 0.29 | 0.26 | 1.18 | 0.05 | 0.07 | 1.30 | 0.06 | 0.06 | 0.17 | 0.22 |
| | 5.0 | 1.61 | 1.83 | 0.32 | 0.28 | 1.28 | 0.05 | 0.07 | 1.40 | 0.06 | 0.06 | 0.16 | 0.20 |

** Pool logOR by the weight of inverse of VarlogOR

*** Pool lamda0 and lamda1 by equal weight for each cases.

Table 5.9 Accuracy of estimation of the different methods for n=300 cases

| | | | Binomial Distribuiton | | | Poisson Distribution Method 1** | | | Poisson Distribution Method 2*** | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| lamda0 | True OR | True logOR | Mean logOR | Var logOR | Mean VarlogOR | Mean logOR | Var logOR | Mean VarlogOR | Mean logOR | Var logOR | Mean VarlogOR | RE1 | RE2 |
| 0.033 | 0.5 | -0.69 | -0.59 | 0.02 | 0.02 | -0.05 | 0.00 | 0.01 | -0.59 | 0.01 | 0.01 | 0.21 | 0.50 |
| (1/30) | 1.0 | 0.00 | 0.01 | 0.02 | 0.02 | 0.31 | 0.01 | 0.01 | 0.00 | 0.01 | 0.01 | 0.28 | 0.52 |
| | 1.5 | 0.41 | 0.51 | 0.02 | 0.02 | 0.57 | 0.01 | 0.01 | 0.35 | 0.01 | 0.01 | 0.28 | 0.44 |
| | 2.0 | 0.69 | 0.79 | 0.02 | 0.02 | 0.76 | 0.01 | 0.01 | 0.60 | 0.01 | 0.01 | 0.28 | 0.44 |
| | 2.5 | 0.92 | 1.01 | 0.02 | 0.02 | 0.92 | 0.01 | 0.01 | 0.79 | 0.01 | 0.01 | 0.31 | 0.39 |
| | 3.0 | 1.10 | 1.20 | 0.03 | 0.02 | 1.05 | 0.01 | 0.01 | 0.96 | 0.01 | 0.01 | 0.29 | 0.39 |
| | 3.5 | 1.25 | 1.35 | 0.03 | 0.02 | 1.16 | 0.01 | 0.01 | 1.09 | 0.01 | 0.01 | 0.28 | 0.38 |
| | 4.0 | 1.39 | 1.48 | 0.03 | 0.02 | 1.26 | 0.01 | 0.01 | 1.21 | 0.01 | 0.01 | 0.29 | 0.36 |
| | 4.5 | 1.50 | 1.61 | 0.03 | 0.02 | 1.35 | 0.01 | 0.01 | 1.32 | 0.01 | 0.01 | 0.24 | 0.32 |
| | 5.0 | 1.61 | 1.72 | 0.04 | 0.03 | 1.44 | 0.01 | 0.01 | 1.41 | 0.01 | 0.01 | 0.21 | 0.31 |
| 0.050 | 0.5 | -0.69 | -0.56 | 0.02 | 0.02 | -0.27 | 0.01 | 0.01 | -0.55 | 0.01 | 0.01 | 0.25 | 0.46 |
| (1.5/30) | 1.0 | 0.00 | 0.13 | 0.02 | 0.02 | 0.14 | 0.01 | 0.01 | 0.00 | 0.01 | 0.01 | 0.29 | 0.45 |
| | 1.5 | 0.41 | 0.53 | 0.03 | 0.02 | 0.42 | 0.01 | 0.01 | 0.34 | 0.01 | 0.01 | 0.27 | 0.39 |
| | 2.0 | 0.69 | 0.82 | 0.03 | 0.02 | 0.62 | 0.01 | 0.01 | 0.59 | 0.01 | 0.01 | 0.29 | 0.36 |
| | 2.5 | 0.92 | 1.04 | 0.03 | 0.03 | 0.78 | 0.01 | 0.01 | 0.78 | 0.01 | 0.01 | 0.26 | 0.36 |
| | 3.0 | 1.10 | 1.22 | 0.03 | 0.03 | 0.92 | 0.01 | 0.01 | 0.95 | 0.01 | 0.01 | 0.29 | 0.37 |
| | 3.5 | 1.25 | 1.37 | 0.03 | 0.03 | 1.04 | 0.01 | 0.01 | 1.08 | 0.01 | 0.01 | 0.24 | 0.29 |
| | 4.0 | 1.39 | 1.50 | 0.04 | 0.03 | 1.15 | 0.01 | 0.01 | 1.20 | 0.01 | 0.01 | 0.25 | 0.30 |
| | 4.5 | 1.50 | 1.62 | 0.04 | 0.03 | 1.24 | 0.01 | 0.01 | 1.31 | 0.01 | 0.01 | 0.26 | 0.29 |
| | 5.0 | 1.61 | 1.74 | 0.04 | 0.04 | 1.32 | 0.01 | 0.01 | 1.40 | 0.01 | 0.01 | 0.21 | 0.25 |
| 0.067 | 0.5 | -0.69 | -0.54 | 0.02 | 0.02 | -0.42 | 0.01 | 0.01 | -0.55 | 0.01 | 0.01 | 0.27 | 0.40 |
| (2/30) | 1.0 | 0.00 | 0.15 | 0.03 | 0.02 | 0.02 | 0.01 | 0.01 | 0.00 | 0.01 | 0.01 | 0.29 | 0.39 |
| | 1.5 | 0.41 | 0.54 | 0.03 | 0.03 | 0.31 | 0.01 | 0.01 | 0.34 | 0.01 | 0.01 | 0.30 | 0.41 |
| | 2.0 | 0.69 | 0.84 | 0.03 | 0.03 | 0.52 | 0.01 | 0.01 | 0.06 | 0.01 | 0.01 | 0.24 | 0.31 |
| | 2.5 | 0.92 | 1.06 | 0.03 | 0.03 | 0.69 | 0.01 | 0.01 | 0.78 | 0.01 | 0.01 | 0.25 | 0.35 |
| | 3.0 | 1.10 | 1.23 | 0.04 | 0.03 | 0.83 | 0.01 | 0.01 | 0.94 | 0.01 | 0.01 | 0.24 | 0.29 |
| | 3.5 | 1.25 | 1.38 | 0.04 | 0.04 | 0.96 | 0.01 | 0.01 | 1.08 | 0.01 | 0.01 | 0.21 | 0.28 |
| | 4.0 | 1.39 | 1.53 | 0.05 | 0.04 | 1.06 | 0.01 | 0.01 | 1.20 | 0.01 | 0.01 | 0.21 | 0.27 |
| | 4.5 | 1.50 | 1.65 | 0.05 | 0.04 | 1.16 | 0.01 | 0.01 | 1.31 | 0.01 | 0.01 | 0.19 | 0.24 |
| | 5.0 | 1.61 | 1.76 | 0.05 | 0.05 | 1.25 | 0.01 | 0.01 | 1.40 | 0.01 | 0.01 | 0.20 | 0.25 |

** Pool logOR by the weight of inverse of VarlogOR

*** Pool lamda0 and lamda1 by equal weight for each cases.

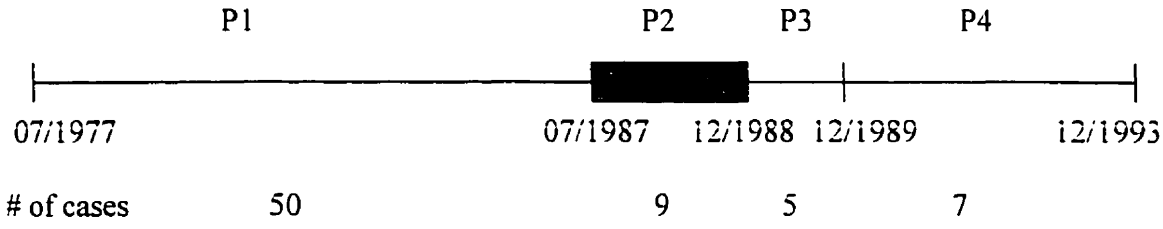


Fig. 5.1 Calendar time of follow up and number of asthma deaths in each calendar time period.

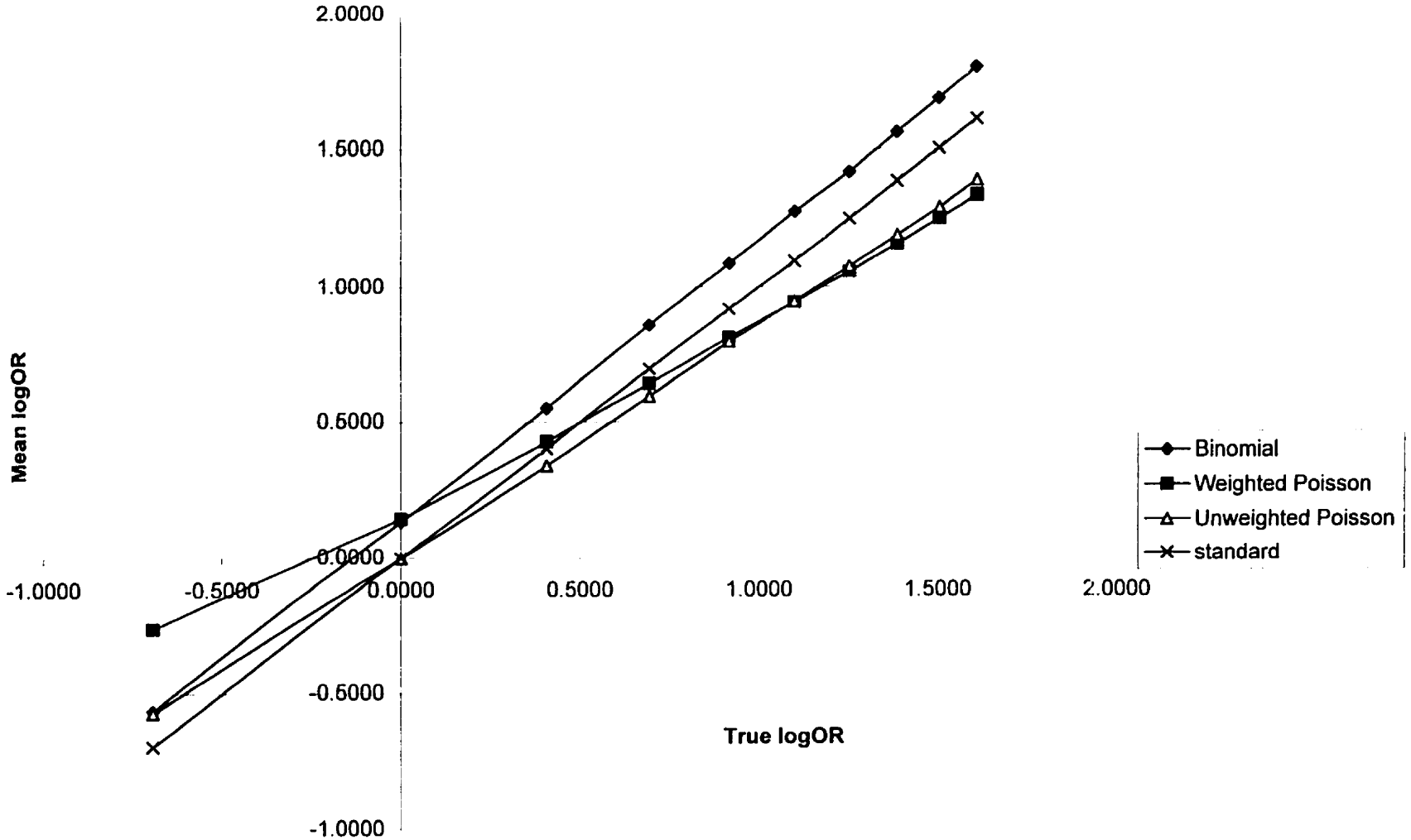


**Fig 5.2 Accurary of different methods of estimation**

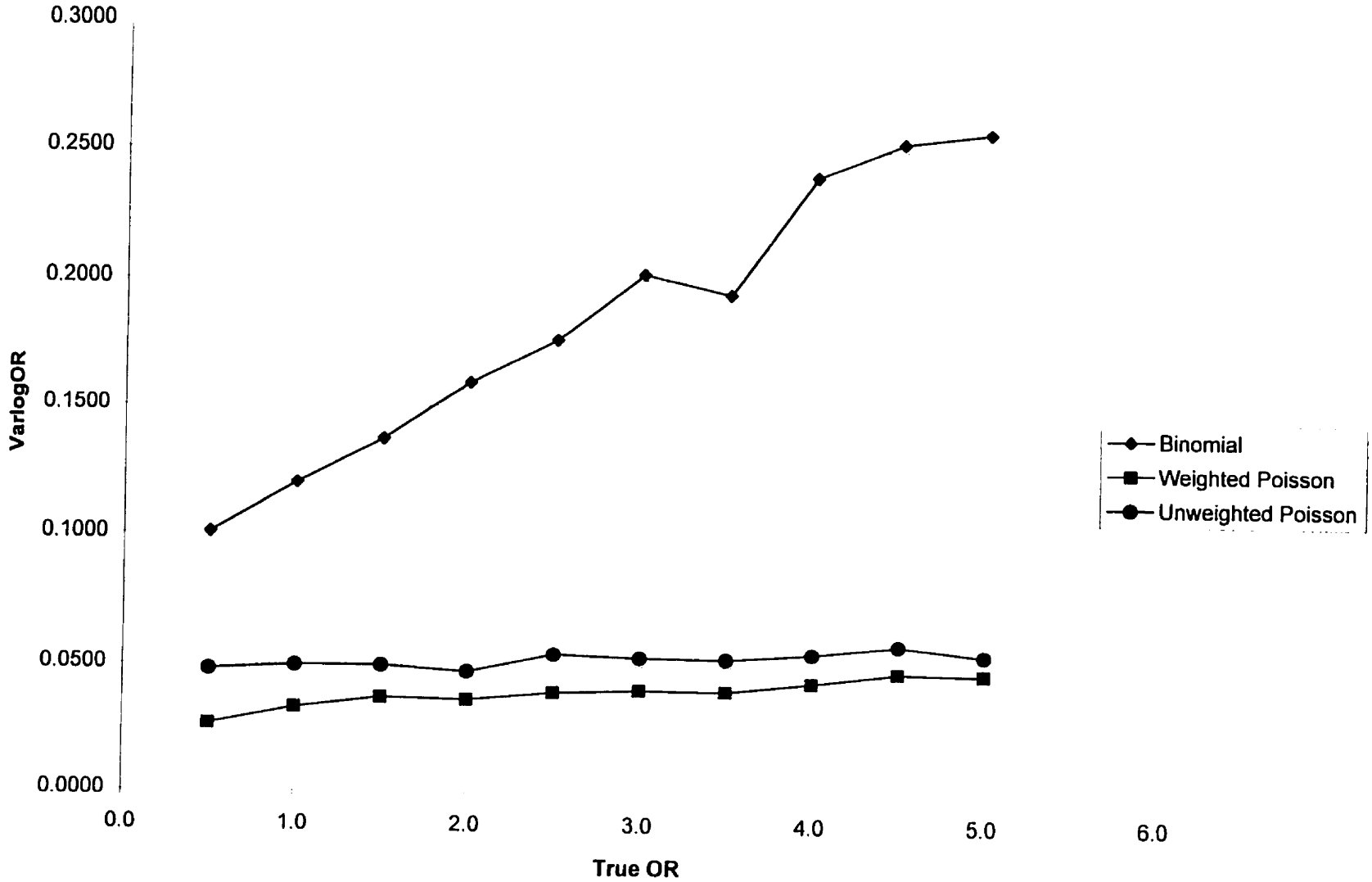


**Fig 5.3 Precision of estimation for different methods**

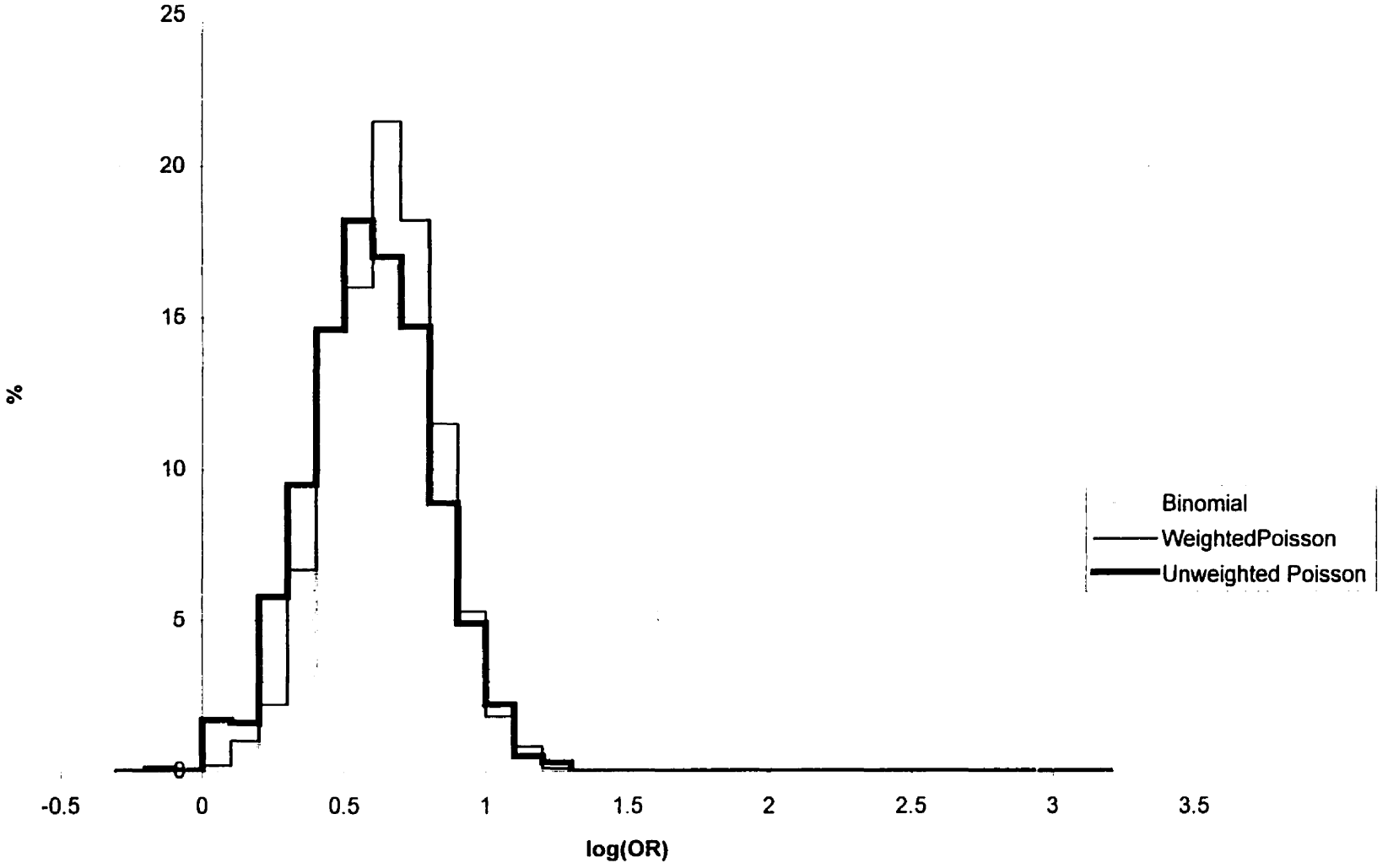


**Fig 5.4 Comparison of Binomial and Poisson method: An example when the true OR=2 (logOR=0.69)**

# CHAPTER 6 DISCUSSION AND CONCLUSION

## 6.1 Efficiency and flexibility

The case-crossover design was first proposed to study the onset of myocardial infarction associated with sexual activity and coffee drinking (Maclure 1991). In that study, the time window was assumed to be 2 hours. Compared to the 1 year period used as a reference time in the present study, that time window is very small, and the number of time windows in the reference period is large (n=4380). In that situation, the estimator of OR with the binomial distribution and the Poisson distribution are very close, as in the limit when $n \to \infty, \theta \to 0$, and $n\theta = \lambda$ remains constant, the binomial distribution converges to the Poisson distribution (Freund 1992). In pharmacoepidemiology, the time window is defined according to the effect time of the drugs, patterns of drug prescriptions and drug use. The time window can be rather long compared to the whole reference time. In this case, the binomial and Poisson distributions may differ significantly because the number of time windows in the reference period is small.

In studies where the data is obtained by survey, it is more objective to collect exposure data in dichotomous form, particularly when the exposure is drug use. In database studies, however, databases provide more precise data on the actual exposure times. In this situation, we may be wasting information if we try to mold the data into the dichotomous form required for the binomial distribution.

In this thesis, we proposed an estimator of OR and its confidence interval for the case-crossover design assuming a Poisson distribution for the inter-exposure times. We used simulations to compare this new approach with the classic binomial approach. The result

shows that the new approach is more efficient, in that its variance is smaller and the confidence limits are tighter. In the simulation study, the data were generated according to the real data used for illustration, that is, the time window was 30 days and the total observation time was one year. In this situation, the relative efficiency was 50% or less.

For the classic case-crossover study analysis that we reviewed in Chapter 2, the risk factor was dichotomized as exposed and non-exposed. This is a 'single hit' model. This 'single hit' model means that an exposure needs to happen only once for a subject to be considered "exposed". Whether the exposure occurs once or more than once does not change the exposure. The method we have proposed is also a "single" hit model, but uses all the information to arrive at its estimate. It can also be amended for other models. For instance, for a 'double hit' model, the equation (4.1) can be amended as,

$$\begin{aligned} OR &= \frac{\left[1 - P(E < 2|Case)\right]/P(E < 2|Case)}{\left[1 - P(E < 2|Control)\right]/P(E < 2|Control)} \\ &= \frac{\left[1-\left(e^{-\lambda_1 t} + \lambda_1 t e^{-\lambda_1 t}\right)\right]\left(e^{-\lambda_1 t} + \lambda_1 t e^{-\lambda_1 t}\right)}{\left[1-\left(e^{-\lambda_0 t} + \lambda_0 t e^{-\lambda_0 t}\right)\right]\left(e^{-\lambda_0 t} + \lambda_0 t e^{-\lambda_0 t}\right)} \end{aligned} \tag{6.1}$$

For the same model as above, the binomial distribution approach can dichotomize the exposure as less than 2 times and greater or equal to 2 times, but this dichotomization also wastes some information which may lead to lower efficiency.

## 6.2 Exposure time trend

An important assumption behind the case-crossover study is that the distribution of exposure does not change with time during the reference period. In both the binomial and the Poisson approaches, the distribution of exposure in the reference period was estimated under this assumption. Only if this assumption is valid, can we assume that this estimate represents the reference period and that the estimate of OR is valid. In case there is a time trend in the exposure, as certain drugs may be more frequently prescribed than before, the basic assumption of the case-crossover design is challenged. The case-time-control design (Suissa 1995) provided a model to control the time trend, but an external control group was needed. Compared to the binomial distribution approach, which needs to tediously calculate the proportion of alerted time while considering the overlapping of effective time of drug use, the Poisson approach's estimate of the OR are functions of only the rate of drug use. This makes the new approach not only convenient but also more flexible and amenable to verifying some of the assumptions behind the validity of the case-crossover design. For the time trend problem, we can fit $\lambda_0$ as a function of time using a regression model. This form of time data is easily amendable to a wide variety of models, including splines, which can identify heterogeneity of exposure over time during the reference period.

It can be predicted that if exposure increase with time, ignoring the time trend in data analysis will make $\lambda_0$ to be less estimated, which lead to the overestimation of OR and on the other hand, ignoring decreasing time trend will lead to underestimation of OR.

## 6.3 Independence of $\hat{\lambda}_1$ and $\hat{\lambda}_0$

In deriving the variance for equation (4.6) and (4.7), we assumed that the $\hat{\lambda}_1$ and $\hat{\lambda}_0$ were independent. This assumption is based on the consideration that for each individual, the drug use in one period is independent of the drug use in another period. The fact that a patient who has a higher rate of drug use in the reference period may also have a higher rate of drug use in the risk period does not imply that this assumption is invalid. This is confirmed by the result of our simulations. In the simulation, we first randomly generated $\lambda_0$, but $\lambda_1$ was determined by fixing OR and $\lambda_0$. Thus, the subject with a higher $\lambda_0$ will also have a higher $\lambda_1$. When we randomly generated the exposures, however, we generated them independently for the risk and the reference periods. This implies that an exposure in the reference period will have nothing to do with whether an exposure happened in the risk period, except from the higher rate. This suggests independence of $\hat{\lambda}_1$ and $\hat{\lambda}_0$, which is confirmed by the unbiased variance estimators that are derived under the assumption of independence.

## 6.4 Limitation of the Poisson distribution method

The first limitation of the method is that it can only be applied to the study of a randomly occurring exposure. In pharmacoepidemiologic studies, the drug under evaluation must be used irregularly(e.g. use as needed) in order to satisfy the Poisson distribution assumption. In addition, the Poisson distribution method cannot be applied to case-crossover studies in which the time-window of the ecposure is extremely short. For instance, in a study of the risk of child pedestrian injury in an urban traffic environment

(Roberts et. al. 1995), the exposure was defined as to whether a child was crossing a high traffic street or not when the accident occurred. In this case, only the Bernoulli distribution is appropriate. Similarly, the Poisson distribution method is inappropriate to use in the analysis of survey data, when the exact time of exposure is not available, as the basic requirement of the Poisson distribution approach cannot be met. Thus, this method is most useful in pharmacoepidemiological database studies where the drug is taken as needed, the time window is sufficiently long to allow for repeated exposure and the exact time of drug prescription is available.

### 6.5 Discussion about asthma data

We illustrated this new method based on the Poisson distribution with data from a study of asthma deaths associated with the use of beta-agonists and compared this result with that of the classical case-crossover design. When the time window is assumed to be 10 days, the results from the binomial and Poisson distribution approaches are very close, but when the time window is assumed to be 30 days, the results from the two approaches are clearly different. The binomial approach gave odds-ratios of 4.57 or 4.38, while the Poisson distribution approach gave ORs of 1.58 or 1.54, which are clearly smaller. We can first explain this difference from the simulation analyses, which show that the binomial distribution method gives higher estimates of the odds-ratio. Second, we can see that the confidence interval (CI) given by the binomial distribution approach is wide, and the 95% CIs from the two methods are overlapping, which implies that the difference between the two methods could be due to random error. To further explain the difference

between the two types of estimators, we need to check the data in detail, which are listed below.

| ID | W 1 | W 2 | W 3 | W 4 | W 5 | W 6 | W 7 | W 8 | W 9 | W 10 | W 11 | W 12 | Numerator | Denominator |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| 12092 | 1 | 2 | 1 | 0 | 2 | 3 | 1 | 1 | 3 | 0 | 3 | 0 | 0.2727 | 0.0000 |
| 16385 | 1 | 2 | 1 | 2 | 1 | 1 | 1 | 1 | 1 | 0 | 1 | 1 | 0.0909 | 0.0000 |
| 17804 | 1 | 1 | 1 | 1 | 1 | 1 | 2 | 0 | 1 | 1 | 1 | 2 | 0.0909 | 0.0000 |
| 22198 | 1 | 1 | 1 | 1 | 1 | 0 | 1 | 2 | 1 | 1 | 1 | 2 | 0.0909 | 0.0000 |
| 37059 | 1 | 1 | 2 | 1 | 1 | 1 | 0 | 2 | 1 | 2 | 1 | 0 | 0.1818 | 0.0000 |
| 14120 | 5 | 2 | 3 | 2 | 2 | 2 | 2 | 1 | 3 | 4 | 2 | 3 | 0.0000 | 0.0000 |

In the above table, W1-W12 are the 30-day time window counted back from the date of death, so that W1 is the risk period, and W2-W12 are reference period. The table lists the number of prescriptions in each time window. Numerator and denominator are respectively the contribution of the subjects to the numerator and denominator of RR as in equation (4.17). We can see that the first 5 subjects all have the numerator greater than 0 and the denominator of 0. This means that these subjects favor the hypothesis that RR>1. For the Poisson distribution method, however, we can see that all these subjects favor the hypothesis that RR<1, because, for these subjects, the rate of prescription in the risk period is higher than in the reference period. Subject 14120 is a contrasting example. For the Poisson distribution method, this subject favors the hypothesis that RR>1 but for the binomial distribution method this subject favors the hypothesis that RR=1. For our 54 study cases there are 19 cases that are clearly in the first situation and 6 cases in the contrasting situation. By its nature, it will not happen that for the Poisson distribution approach a subject will favor the hypothesis that RR>1 and the hypothesis that RR<1 for the binomial distribution method. All these are reason for the estimator of RR from the

binomial distribution method to be higher than the RR from the Poisson distribution method.

In the case control study on which these data are based in part (Spitzer et. al. 1992), the exposure was defined as the use of beta-agonists during the 12 months before the index date, which means that the time window is defined as 12 months. When the analysis is based on continuous exposure of any inhaled beta-agonists, the odds ratio is 2.6 (95% CI 1.7-3.9) per canister per month. Later, Suissa developped a new technique named the case-time-control design (Suissa 1995) and applied this method to the same study with the same data, to better control for the confounding by indication present under the case-control design. His approach showed that for the continuous forms of beta-agonist use, the odds ratio is 1.7 (95% CI=0.9-3.0) per canister of beta-agonist per month. This odds ratio is lower than that from the case-control study because the latter is based on a better control for confounding by indication and time trends in drug use.

In the present study, we obtained an odds ratio of 1.54 (95% CI=0.99-2.41) for the Poisson distribution approach and a relative risk of 4.57 (95% CI=1.60-13.02) for the binomial distribution approach when the time window is defined as 30 days. According to the previous discussion, the Poisson distribution approach is assumed to produce better estimation. These results and the results from the previous studies are related but they are not comparable. First, the measure of exposure is different. In the previous studies, the odds ratio compares the risk of using 1 additional canister of beta-agonist per month. In the present study, we try to compare the risk of patients who use beta-agonists in a time

window of 30 days, with the risk of asthma death for the patient who does not use beta-agonist in this time window. Second, the present study and the previous studies are based on different assumptions and they are trying to answer different questions. In the previous studies, the time window is defined as 12 months, so it studied a chronic effect of these drugs. An assumption behind this design is that the effect of beta-agonist use to asthma death will last for one year and the use of a beta-agonist at any time during a year has the same effect on asthma death. In the present studies, the time window is defined as 30 days, so it studied a transient effect. The assumption behind this study is that the effect of beta-agonist can and only can last for 30 days, so the drug use 30 days before asthma death do not have such a transient effect to asthma death. Third, the two kinds of studies are related. There may be different explanations for the result of previous studies. A) All the excess risk from the previous studies is due to the increased beta-agonist use in the last month before asthma death; the cases and controls have the same distribution of beta-agonist use between one month before asthma death and one year before asthma death, so there is only transient effect and no chronic effect. B) There is only chronic effect and no acute effect. In that case, the distribution of beta-agonist use in the risk period is the same as the reference period of case-crossover study, and for the whole year, the use of the drug is at a higher level than the external control. C) The effect of beta-agonists to asthma death is a combination of chronic and transient effects. Our present study suggested that there is a transient effect of beta-agonists to asthma death, so that situation B) is excluded.

## 6.6 Strength and limitation of the study

This is the first study to reveal the transient effect of beta-agonists to asthma death. Compared with the previous study from the same cohort, this study has 6 more years of follow-up and the sample size of asthma deaths increased from 44 to 54, a 20% increase. This study is based on the Saskatchewan health databases and thus inherits the advantages of a database study over a field study. First, the Saskatchewan database is population-based and it includes health care record of more than 95% of the population of the province. With this database, we collected all cases of asthma death in a well-defined cohort that was followed up for 16 years. Since all the cases that meet the criteria are collected, this study is immune from selection bias. Second, in this database study, we can obtain precise information about the date of drug prescription, quantity and category of drug, date and stay of hospitalization, date of death etc, so the outcome and exposure are precisely measured and the study is immune from recall bias.

The main limitation of this study is that the drug usage is completely based on the record of prescription, and the information about compliance is not available. Also, the time widow is defined based on drug prescriptions, that is, after the date of drug prescription, a patient will be exposed to the drug for a certain number of days. It is not easy to validate this assumption.

## 6.7 Conclusion

The case-crossover design is a new epidemiological technique to assess the transient

effects of a brief exposure on the onset of an acute outcome. Two apparent advantages make it a prospective technique in pharmacoepidemiology studies. First, this study design compares the exposure within subject, so that some confounders that are not measurable can be easily controlled. This advantage is very helpful in pharmacoepidemiology studies in which confounding by indication is a challenge. Second, this study design only needs information from cases, and no controls need to be sampled, so it reduces effort and expense and eliminates problems arising from control selection.

In this thesis, we discussed the application of the case-crossover design to pharmacoepidemiology database studies, where the time window is usually rather long, and multiple exposure may happen within a time window. On the other hand, in database studies, precise information about time and quantity of drug prescriptions are often available. In this situation, if we still use the case-crossover design based on the binomial distribution and ignore the number and times of prescriptions in the time windows, we may waste information and lose efficiency. For this reason, we proposed a new approach of data analysis for the case-crossover design based on exponential distribution for the inter-exposure onset times. We derived an estimator of the odds ratio and the variance of the logarithm of odds ratio so that a confidence interval of odds ratio can be calculated. We developed a SAS macro to carry out a simulation study to compare this new approach with the conventional case-crossover study. The results show that the new approach is indeed more efficient in that its confidence limits are tighter, than the conventional case-crossover design.

We also illustrated this approach with data on 54 asthma death identified from the Saskatchewan Health databases, and this is the first study to assess the transient effect of inhaled beta-agonists on asthma death. Our result demonstrates that asthma death is associated with an acute increase of usage of inhaled beta-agonists in the 30 days before death.

The main assumption behind the case-crossover design is that the distribution of exposure remains stable over time during the reference period, so time trends of exposure are the main challenge of validity of this method. We have discussed the possibility of generalizing the Poisson distribution model to adjust for the time trend. It is believed that the Poisson distribution model is more flexible to verify the assumptions behind the validity of the case-crossover design, but further research is needed.

# BIBLIOGRAPHY


1. **Bailar, J.C.** Introduction. In: *Clinical Trials*, edited by S.H. Shapiro and T.A. Louis. New York and Basel: Marcel Dekkeer,INC., 1983, p. 1-12.

2. **Barbone, F., A.D. McMahon, P.G. Davey, A.D. Morris, I.C. Reid, D.G. McDevitt, and T.M. MacDonald.** Association of road-traffic accidents with benzodiazepine use. *Lancet* 352: 1331-1336, 1998.

3. **Beasley, R., N. Pearce, J. Crane, H. Windom, and C. Burgess.** Asthma mortality and inhaled beta agonist therapy. *Australian & New Zealand Journal of Medicine* 21: 753-763, 1991.

4. **Breslow, N.E. and N.E. Day.** *Statistical methods in cancer research. Vol. 1. The analysis of case-control studies*. Lyon, France: Internnational Agency for Research on Cancer, 1980, p. 162-189.

5. **Burdorf, A., G.S. Sorock, R.F. Herrick, and T.K. Courtney.** Advancing epidemiologic studies of occupational injury--approaches and future directions. *American Journal of Industrial Medicine* 32: 180-183, 1997.

6. **Carrier, T.K., L. Estberg, S.M. Stover, I.A. Gardner, B.J. Johnson, D.H. Read, and A.A. Ardans.** Association between long periods without high-speed workouts and risk of complete humeral or pelvic fracture in thoroughbred racehorses: 54 cases (1991-1994). *Journal of the American Veterinary Medical Association* 212: 1582-1587, 1998.

7. **Dixon, K.E.** A comparison of case-crossover and case-control designs in a study of risk factors for hemorrhagic fever with renal syndrome. *Epidemiology* 8: 243-246, 1997.

8. **Efron, B. and R. Tibshirani.** Statistical data analysis in the computer age. *Science, vol* 253, no.5018, 26 July 1991, pp.390-5. USA.: 1899.

9. **Estberg, L., I.A. Gardner, S.M. Stover, and B.J. Johnson.** A case-crossover study of intensive racing and training schedules and risk of catastrophic musculoskeletal injury and lay-up in California thoroughbred racehorses. *Preventive Veterinary Medicine* 33: 159-170, 1998.

10. **Flanders, W.D. and M.J. Khoury.** Analysis of case-parental control studies: method for the study of associations between disease and genetic markers [published erratum appears in Am J Epidemiol 1997 Mar 1;145(5):477-8]. *American Journal of Epidemiology* 144: 696-703, 1996.

11. **Freund J.E.** *Mathematical Statistics*. Englewood Cliffs: A Simon & Schuster Company, 1992,

12. **Greenland, S.** A unified approach to the analysis of case-distribution (case-only) studies. *Statistics in Medicine* 18: 1-15, 1999.

13. **Guess, H.A.** Behavior of the exposure odds ratio in a case-control study when the hazard function is not constant over time. *Journal of Clinical Epidemiology* 42: 1179-1184, 1989.

14. **Gullette, E.C., J.A. Blumenthal, M. Babyak, W. Jiang, R.A. Waugh, D.J. Frid, C.M. O'Connor, J.J. Morris, and D.S. Krantz.** Effects of mental stress on myocardial ischemia during daily life. *JAMA* 277: 1521-1526, 1997.

15. **Hill, A.B.** *Principles of Medical Statistics*. London: Lancet, 1971, p. 312-320.

16. **Jackson, R., R. Scragg, and R. Beaglehole.** Coronary heart disease risk associated with habitual alcohol consumption in men and women. *British Medical Journal* 303: 211-216, 1991.

17. **Kalbfleisch, J.D. and R.L. Prentice.** *The Statistical Analysis of Failure Time Data*. 1980, p. 32

18. **Khoury, M.J. and W.D. Flanders.** Nontraditional epidemiologic approaches in the analysis of gene-environment interaction: case-control studies with no controls:. [Review] [41 refs]. *American Journal of Epidemiology* 144: 207-213, 1996.

19. **Maclure, M.** The case-crossover design: a method for studying transient effects on the risk of acute events. *American Journal of Epidemiology* 133: 144-153, 1991.

20. **Marshall, R.J. and R.T. Jackson.** Analysis of case-crossover designs. *Statistics in Medicine* 12: 2333-2341, 1993.

21. **Matsui T.** Asthma death and $\beta_2$-agonists. In: *Currennt advances in pediatric allergy and clinical immunology. Selected proceedings from the 32nd annual meeting of the japanese society of pediatric allergy and clinical immunology*, edited by Shinomiya K. Tokyo, Japan: Churchill Livingstone, 1996, p. 161-164.

22. **Meier, C.R., S.S. Jick, L.E. Derby, C. Vasilakis, and H. Jick.** Acute respiratory-tract infections and risk of first-time acute myocardial infarction. *Lancet* 351: 1467-1471, 1998.

23. **Miettinen, O.S.** *Theoreticcal Epidemiology: Principles of Occurence Research in Medicine*. New York: John Wiley & Sons, 1985.

24. **Miettinen, O.S. and J.J. Caro.** Principles of nonexperimental assessment of excess risk, with special reference to adverse drug reactions. *Journal of Clinical Epidemiology* 42: 325-331, 1989.

25. **Mittleman, M.A., M. Maclure, J.B. Sherwood, R.P. Mulry, G.H. Tofler, S.C. Jacobs, R. Friedman, H. Benson, and J.E. Muller.** Triggering of acute myocardial infarction onset by episodes of anger. Determinants of Myocardial Infarction Onset Study Investigators. *Circulation* 92: 1720-1725, 1995.

26. **Mittleman, M.A., M. Maclure, and J.M. Robins.** Control sampling strategies for case-crossover studies: an assessment of relative efficiency. *American Journal of Epidemiology* 142: 91-98, 1995.

27. **Mittleman, M.A., M. Maclure, M. Nachnani, J.B. Sherwood, and J.E. Muller.** Educational attainment, anger, and the risk of triggering myocardial infarction onset. The Determinants of Myocardial Infarction Onset Study Investigators. *Archives of Internal Medicine* 157: 769-775, 1997.

28. **Mittleman, M.A., G. Maldonado, S.G. Gerberich, G.S. Smith, and G.S. Sorock.** Alternative approaches to analytical designs in occupational injury epidemiology. *American Journal of Industrial Medicine* 32: 129-141, 1997.

29. **Muller, J.E., A. Mittleman, M. Maclure, J.B. Sherwood, and G.H. Tofler.** Triggering myocardial infarction by sexual activity. Low absolute risk and prevention by regular physical exertion. Determinants of Myocardial Infarction Onset Study Investigators. *JAMA* 275: 1405-1409, 1996.

30. **Navidi, W.** Bidirectional case-crossover designs for exposures with time trends. *Biometrics* 54: 596-605, 1998.

31. **Petridou, E., M.A. Mittleman, D. Trohanis, N. Dessypris, T. Karpathios, and D. Trichopoulos.** Transient exposures and the risk of childhood injury: a case-crossover study in Greece. *Epidemiology* 9: 622-625, 1998.

32. **Redelmeier, D.A. and R.J. Tibshirani.** Interpretation and bias in case-crossover studies. *Journal of Clinical Epidemiology* 50: 1281-1287, 1997.

33. **Redelmeier, D.A. and R.J. Tibshirani.** Association between cellular-telephone calls and motor vehicle collisions. *New England Journal of Medicine* 336: 453-458, 1997.

34. **Roberts, I., R. Marshall, and T. Lee-Joe.** The urban traffic environment and the risk of child pedestrian injury: a case-crossover approach. *Epidemiology* 6: 169-171, 1995.

35. **Rothman, K.J.** *Modern Epidemiology*. Toronto: Little Brown and Company, 1986,

36. **Rothman, K.J. and S. Greenland.** *Modern Epidemiology*. Philadelphia: Lippincott-Raven Publishers, 1998,

37. **Schaid, D.J. and S.S. Sommer.** Genotype relative risks: methods for design and analysis of candidate-gene association studies. *American Journal of Human Genetics* 53: 1114-1126, 1993.

38. **Sckett, D.L., R.B. Haynes, G.H. Guyatt, and P. Tugwell.** *Clinical Epidemiology: A Basic Science for Clinical Medicine*. Toronto: Little Brown and Co., 1991, p. 283-302.

39. **Self, S.G., G. Longton, K.J. Kopecky, and K.Y. Liang.** On estimating HLA/disease association with application to a study of aplastic anemia. *Biometrics* 47: 53-61, 1991.

40. **Spitzer, W.O., P. Ernst, S. Suissa, J.F. Boivin, R.I. Horwitz, B. Habbick, D. Cockcroft, M. McNutt, and A.S. Buist.** Fenoterol and death from asthma. *Medical Journal of Australia* 157: 567-568, 1992.

41. **Strom, B.L.** Study designs available for pharmacoepidemiology studies. In: *Pharmacoepidemiology*, edited by B.L. Strom. Toronto: John Wiley & Sons, 1994, p. 15-27.

42. **Sturkenboom, M.C., A. Middelbeek, de Jong van den Berg LT, P.B. van den Berg, B.H. Stricker, and H. Wesseling.** Vulvo-vaginal candidiasis associated with acitretin. *Journal of Clinical Epidemiology* 48: 991-997, 1995.

43. **Suissa, S.** Novel approaches to pharmacoepidemiology study design and statistical analysis. In: *Pharmacoepidemiology*, edited by B.L. Strom. Toronto: John Wiley & Sons, 1994, p. 629-645.

44. **Suissa, S., P. Ernst, J.F. Boivin, R.I. Horwitz, B. Habbick, D. Cockroft, L. Blais, M. McNutt, A.S. Buist, and W.O. Spitzer.** A cohort analysis of excess mortality in asthma and the use of inhaled beta-agonists. *American Journal of Respiratory & Critical Care Medicine* 149: 604-610, 1994.

45. **Suissa, S.** The case-time-control design. *Epidemiology* 6: 248-253, 1995.

46. **Vinson, D.C., N. Mabe, L.L. Leonard, J. Alexander, J. Becker, J. Boyer, and J. Moll.** Alcohol and injury. A case-crossover study. *Archives of Family Medicine* 4: 505-511, 1995.

47. **Wacholder, S., J.K. McLaughlin, D.T. Silverman, and J.S. Mandel.** Selection of controls in case-control studies. I. Principles. *American Journal of Epidemiology* 135: 1019-1028, 1992.

48. **Wacholder, S., D.T. Silverman, J.K. McLaughlin, and J.S. Mandel.** Selection of controls in case-control studies. II. Types of controls. *American Journal of Epidemiology* 135: 1029-1041, 1992.

49. **Wacholder, S., D.T. Silverman, J.K. McLaughlin, and J.S. Mandel.** Selection of controls in case-control studies. III. Design options. *American Journal of Epidemiology* 135: 1042-1050, 1992.

50. **Willich, S.N., M. Lewis, H. Lowel, H.R. Arntz, F. Schubert, and R. Schroder.** Physical exertion as a trigger of acute myocardial infarction. Triggers and Mechanisms of Myocardial Infarction Study Group. *New England Journal of Medicine* 329: 1684-1690, 1993.

51. **Willich, S.N., M. Lewis, H.R. Arntz, H. Lowel, F. Schubert, R. Stern, and Schroder.** Stress factors in acute myocardial infarct: the role of physical exertion and unusual life events. [German]. *Zeitschrift fur Kardiologie* 83: 423-430, 1994.

52. **Zaffanella, L.E., D.A. Savitz, S. Greenland, and K.L. Ebi.** The residential case-specular method to study wire codes, magnetic fields, and disease. *Epidemiology* 9: 16-20, 1998.