Propensity Score Matching Methods for the Analysis of Recurrent Events

by

© **Yasin Khadem Charvadeh**

A thesis submitted to the School of Graduate Studies in partial fulfillment of the requirements for the degree of Master of Science.

Department of Mathematics and Statistics

Memorial University

August 2019

St. John's, Newfoundland and Labrador, Canada

Abstract

Observational studies are often used to investigate the effects of treatments on a specific outcome. In many observational studies, the event of interest can be of recurrent type, which means that subjects may experience the event of interest more than one time during their follow-up. The lack of random allocation of treatments to subjects in observational studies may induce the selection bias leading to systematic differences in observed and unobserved baseline characteristics between treated and untreated subjects. Propensity score matching is a popular technique to address this issue. It is based on the estimation of conditional probability of treatment assignment given the measured baseline characteristics. The use of the propensity score in the analysis of observational studies with recurrent event outcomes has not been well developed. In this study, we consider three matching methods called propensity score matching, covariate matching and history matching, and compare the accuracy of them to estimate the treatment effects in recurrent event rates through Monte Carlo simulation studies. We consider various scenarios under the settings of time-fixed and time-dependent treatment indicators. A synthetic data set is analyzed to illustrate the methods discussed in the thesis.

To My Family

Acknowledgements

Firstly, I would like to express my sincere gratitude to my supervisor Dr. Candemir Cigsar for his continuous support and enlightening discussions and his critical feedback to my work during my master's studies. I deeply appreciate his immense knowledge, guidance, patience and kindness through this time. I could not have imagined having a better supervisor and mentor for my master's studies.

Besides my supervisor, I would like to thank the rest of my thesis committee for their insightful comments and encouragement. Their hard questions stimulated me to widen my research from various perspectives.

I am also indebted to my parents, Taha Khadem Charvadeh and Elaheh Shahmari Ardehjani, and my brothers and beautiful sister for their unconditionally support and love throughout the years.

Statement of contribution

Dr. Candemir Cigsar proposed the research question that was investigated throughout this thesis. The overall study was jointly designed by Dr. Candemir Cigsar and Yasin Khadem Charvadeh. The algorithms were implemented, the simulation study was conducted and the manuscript was drafted by Yasin Khadem Charvadeh. Dr. Candemir Cigsar supervised the study and contributed to the final manuscript.

Table of contents

List of tables

List of figures

Chapter 1

Introduction

1.1 Propensity Score Methods and Recurrent Events in Observational Studies

In many epidemiological studies, relationships between some explanatory variables (independent variables) and an outcome variable (dependent variable) are of interest. Such relationships are usually investigated through data collected from subjects and their environments. Therefore, data collection is a crucial part of epidemiological studies. It should be carried out in a scientific way. Otherwise, results of a study may lead to wrong conclusions.

Two important general classes of study designs in epidemiology are *experimental* and *observational* studies. Rothman et al. (2008) define an experiment in epidemiology as a study in which investigators deliberately manipulate the exposure or treatment assigned to participants in the study. *Randomized controlled trials* or as sometimes known as *randomized clinical trials* are a type of experimental study where the subjects are humans (Faries et al., 2010). Many times randomized controlled trials are considered as the *gold standard* if the objective of a study is to investigate cause-and-effect type of relationships. An interesting discussion on this subject is given by Grossman and Mackenzie (2005). A major problem with randomized controlled trials is that, since exposures or treatments are assigned to individuals randomly, ethical issues restrict their applications. Observational studies are non-experimental studies in which the main goal is to establish a relationship between exposures or treatments

and outcome variables without conducting an experiment. Investigators simply collect data without manipulating the population. In such studies, researchers do not randomly assign exposures or treatments to individuals. Individuals either self-select them or receive them according to some of their characteristics so that investigators have no control over treatments received by individuals (Rothman et al., 2008). There are many studies compared the randomized controlled trials with the observational studies; e.g., see Faries et al. (2010, Chapter 1) and the references given there. An important disadvantage of randomized controlled trials over observational studies is that the generalizability of the obtained results is limited. This limitation is a result of strict regulations applied to conduct a randomized controlled trial, where individuals are usually followed in well controlled environments. On the other hand, observational studies reflect more realistic situations of individuals because they are not conducted under strict regulations.

The performance of a treatment is usually evaluated by comparing two or more groups. Such a comparison group is called a *control group* if the individuals in this group do not receive an active treatment. Otherwise, it is called a *treatment group*. It should be noted that, if there are only active treatments available in a study, a group which consists of individuals who receive a standard treatment is sometimes called a control group as well. In this case, individuals in a treatment group receive the new or *research* treatment. In randomized controlled trials, individuals are randomly assigned to one of those groups.

Randomization, i.e. random allocation of treatments to experimental units, is an important principle of experimental designs to establish *cause-and-effect* relationships. An important type of bias, called *selection bias*, arises when individuals in study have different probability of being assigned to a treatment or control group (Faries et al., 2010). As a result of this bias, effects of a treatment may confound with the characteristics of individuals in the comparison groups so that the average treatment effects may become incomparable at the end of an experiment. Randomization, at least in theory, allows to form groups similar in all aspects of individuals before the application of treatments so that the average treatment effects obtained from groups become comparable at the end of an experiment (Rothman et al., 2008, Chapter 6). Therefore, randomization allows researchers to deal with the potential selection bias in randomized controlled trials. Unless it is carefully addressed, the observational studies suffer from the selection bias. Because of this issue, it is a difficult task

to establish cause-and-effect relationships in observational studies. The results of most of the observational studies can be interpreted under *association* or *correlation* rather then causation. Other issues in observational studies and a comparison of association against causation in observational studies can be found in Faries et al. (2010, Section 1.2).

Most of the standard methods in statistics depends on random allocation of subjects to a treatment or control groups. The validity of these methods are in question if the randomization is not applied. In the absence of random allocation, it is not technically possible to determine whether the difference in outcome between groups (e.g., between treatment and control groups) is caused by the treatment or some other variables that may effect the outcome of interest. An example of this issue in the context of hypothesis testing for the difference in average treatment effects is discussed by Rothman et al. (2008, Chapter 6). As they noted (Rothman et al., 2008, p. 88),

" *...with random assignment of the treatment allows one to compute the probability of the observed association under various hypotheses about how treatment assignment affects outcome. In particular, if assignment is random and has no effect on the outcome except through treatment, any systematic (nonrandom) variation in outcome with assignment must be attributable to a treatment effect.*"

Because of the lack of random assignment, observational studies are prone to systematic variation and confounding. Therefore, the results of statistical methods should be carefully interpreted in observational studies. Selection bias may create unbalanced groups with respect to certain characteristics of individuals before the application of treatments. Therefore, it may result in confounded average treatment effects as discussed above. The methods in observational studies should address this bias if the goal is to establish causal inference.

As noted in Faries et al. (2010), there are three commonly used statistical tools to deal with the selection bias in observational studies. These are (*i*) *propensity score* (PS), (*ii*) *instrumental variable* (IV) and (*iii*) *marginal structural models and structural nested models*. Comparing with other methods, empirical and theoretical scientific studies have demonstrated that the methods based on PS reduce more bias arising from non-randomization in a study (Austin, 2008). The use of the PS in observational studies has been pioneered by Rosenbaum and Rubin (1983). The PS is basically defined by them as a conditional probability of the treatment assignment

given observed baseline characteristics (i.e., explanatory variables or covariates). A mathematical definition of and more discussion on PS can be found in Chapter 2. As discussed by Rosenbaum and Rubin (1983), a propensity score is a balancing score; that is, the conditional distribution of observed baseline covariates of treated individuals and untreated individuals, given the balancing score, are the same. Furthermore, Rosenbaum and Rubin (1983) showed that the propensity score is the coarsest balancing score.

A model, called the *PS model*, is developed to estimate the PSs. This model functions as a link between the observed baseline covariates and the probability of treatment assignment. Typically, a logistic regression model is used as the PS model in epidemiology and medicine (Austin, 2011a). Once a PS of every individual is estimated, treatment and control groups can be created with the estimated PSs. There are different methods proposed for the creation of these groups such as matching on the PS, stratification on the PS, inverse probability of treatment weighting using the PS, and covariate adjustment using the PS (Austin, 2011a). In particular, the *propensity score matching* (PSM) method has received considerable attention. In their seminal work, Rosenbaum and Rubin (1983) discussed some of these methods including PSM. We explain more details on PSM in Chapter 2. There are various algorithms to apply the PSM method to match treated and untreated individuals. As denoted in Faries et al. (2010), the most common PSM method is the one-to-one or pair matching. In pair matching, a treated individual is matched with an untreated individual with a similar estimated PS. Simple statistical procedures are developed to assess the balance in the baseline characteristics of individuals between treated and untreated groups. When the matched sets are formed, treatment effects can be estimated. Method for the estimation of treatment effect usually depends on the type of the outcome variable.

In epidemiology and medicine, the PS methods have been much discussed when the outcome of interest is of binary (i.e., dichotomous), continuous or time-to-event type. A survey of methods applicable for such outcomes can be respectively found in Agresti and Min (2004), Austin and Laupacis (2011) and Austin (2014). In this study we focus on recurrent events, which is explained next in Section 1.1.1. The PS methods to deal with the recurrent events in observational studies have not been discussed in the literature in depth. As explained later in Section 1.3, this thesis will be a guide in understanding the use of some promising matching methods in

observational studies with recurrent events.

1.1.1 Recurrent Events in Observational Studies

In many observational studies, the event of interest is of recurrent type, which means that subjects may experience an event more than one time during their lifetimes. It is usually assumed that a process generates recurrent events under a random mechanism. Data obtained from such processes are called recurrent event data (Cook and Lawless, 2007). An important objective of analyzing recurrent event data is to investigate the relationships between treatments and other explanatory covariates on event occurrences. Study designs define how data are collected in recurrent event studies. Recurrent event data can be obtained through randomized clinical trials or observational studies in epidemiological studies. In the lack of randomization, statistical methods to analyze recurrent event data may suffer from the selection bias. To analyze data from such recurrent event studies, specific methods and approaches therefore need to be developed and utilized because causal inference cannot be made due to the lack of the random allocation of treatments to subjects.

Analysis of recurrent event data obtained from randomized controlled trials have been well discussed. A survey of statistical methods to deal with such data can be found in Cook and Lawless (2007, Section 8.4). In the case of observational studies, as discussed by Smith and Schaubel (2015), the statistical methods have been developed if the objective of a study is to describe the event generation process. However, there is an important gap in the literature to establish cause-and-effect relations between treatments and event occurrences in particular when the goal of a study is to compare the effectiveness of treatments. Important challenges involved in the analysis of recurrent events include dependencies between event occurrences in the same subject, various censoring mechanisms leading to incomplete data and unexplained heterogeneity in some characteristics of subjects in a population. Examples of recurrent events include occurrence of asthma attacks in infants, infections in renal transplant patients and insurance claims for policy holders. Cook and Lawless (2007) present several examples of recurrent event data.

A key function in modeling of recurrent events is called *the intensity function* of a recurrent event process (Cook and Lawless, 2007, p. 10). We mathematically define the intensity function in the next chapter. The intensity function of a recurrent

event process is very flexible and can be extended to deal with regression problems in recurrent event studies. Therefore, we consider intensity based regression models for recurrent event processes. As discussed in Section 1.3, our main objective in this thesis is to investigate some important matching methods to deal with the selection bias in observational studies when the individuals are subject to recurrent events. Therefore, we consider simple recurrent event models. We discuss more complicated models and how to extend the methods discussed to deal with them in the final chapter of the thesis.

1.2 Literature Review

Researchers have been working on different matching methods in order to be able to establish cause-and-effect relationship using observational data. The importance of these methods can be understood through the vast literature that have been done on this specific subject. The first theoretical basis for matching methods was developed by Cochran and Rubin (1973) and Rubin (1973b). They considered only one covariate and the primary goal of their studies was to estimate the average effect of the treatment on the treated subjects. Althauser and Rubin (1970) discussed some obstacles to the use of matched sampling such as the problems of attrition and incomplete matching, as well as concepts such as how large the control group should be to get good matches, how to define the quality of matches, how to define a "close-enough" match.

One of the popular approaches to reduce the bias in the estimation of causal treatment effects with observational datasets is the PSM method proposed by Rosenbaum and Rubin (1983). Since in observational studies treatment assignment to the treated and control groups is not random, estimation of the effect of a treatment by using approaches used in randomized controlled trials usually suffer from the selection bias. One approach for solving this problem was suggested by Rubin (1977). He proposed that data can be obtained for a set of potential comparison units, which are not necessarily drawn from the same population as the treated units but for whom we observe the same set of pretreatment covariates. With regard to this approach we can estimate the effect of a treatment just by comparing treated individuals with untreated individuals who possess the same pretreatment characteristics. The main drawback

of this method is the high dimensionality of the matching problem that leads to a larger bias due to the fact that many individuals remain unmatched. This issue can be addressed by using the PSM, which substantially reduces the dimensionality of the problem (Rosenbaum and Rubin, 1985a). It should be noted that Rosenbaum and Rubin (1985b), Rubin and Thomas (1996) and Rubin (2001) have found that matching on the linear propensity score can be particularly effective in terms of reducing bias.

There are a few PSM methods proposed by researchers such as the Nearest Neighbor Matching, Caliper and Radius Matching, Stratification and Interval Matching and Kernel and Local Linear Matching. Nearest neighbor matching proposed by Rubin (1973a) is one of the most straightforward and common matching methods, for which we choose an individual from the control or comparison group as the match for the treated individual. In this method, the two matched individuals should have very close propensity scores. Althauser and Rubin (1970), Cochran and Rubin (1973), Rubin (1973a) and Raynor Jr (1983) investigated another matching method, called the caliper matching, which is a variant of the nearest neighbor matching method. This method avoids bad matches, a problem common to the nearest neighbor matching, by imposing a tolerance level on the maximum propensity score distance (i.e., a caliper). In this case an individual from the comparison group is considered as a match for the treated individual if it lies within the caliper. Rosenbaum and Rubin (1985b) discuss the choice of an appropriate caliper width by using results from Cochran and Rubin (1973). Smith and Todd (2005) noted that a drawback of the caliper matching method is that it could be difficult to know a priori what choice for the tolerance level would be reasonable. Dehejia and Wahba (2002) proposed a variant of caliper matching called the *radius matching*. In the radius matching, it is possible to use all comparison group members together with nearest neighbors within each caliper, which in return allows us to use more units as matches, and avoid bad matches.

The work by Rosenbaum and Rubin (1984) is the first solid study on stratification and interval matching based on propensity scores. In this case of propensity score matching, the common support of propensity score is partitioned into a set of intervals. The idea behind this method is to calculate the mean difference in outcomes between treated and controlled observations falling within each strata. This procedure is called the impact within each interval. A weighted average of the interval impact estimates then provides an overall impact estimate.

Kernel and local linear matching are nonparametric matching estimators that use weighted averages of all or, depending on the choice of the kernel function, nearly all individuals in the control group for each observation of the treated group to construct the counterfactual outcome. In this case, the allocation of the weights is based on the propensity score, which means that the closer the propensity score of an individual in the control group to that of the treated individual, the higher the weight would be. The main research on Kernel and local linear matching is done by Heckman et al. (1997, 1998) and Heckman et al. (1998).

Robins et al. (2000) proposed the class of marginal structural models, which is a new class of causal models allowing for improved adjustment of time-dependent confounders when there exist time-varying exposures or treatments. They showed that the parameters of a marginal structural model can be consistently estimated using a new class of estimators called the inverse probability of treatment weighted estimators. Stuart (2010) provided a detailed structure and guidance for researchers interested in using matching methods. Cottone et al. (2019) and Vansteelandt and Daniel (2014) investigated the efficiency and performance of regression adjustment for propensity scores to estimate the average treatment effect in observational studies.

Recurrent events have been of interest for researchers for a long time. The recent history of the statistical analysis of recurrent events through stochastic processes in medical sciences goes back to almost forty years ago. For example, Byar (1980) investigated the effect of instillations of thiotepa on bladder tumors, which could recur during the first two years after transurethral resection. Following Byar's work, Gail et al. (1980) concerned with the comparison of episodic illness data arising from two treatment groups. Lawless and Nadeau (1995) analyzed data on automobile warranty claims, and improved the method discussed by Nelson (1988) for estimating the cumulative mean function of identically distributed processes of recurrent events. Lawless and Nadeau (1995) proposed a robust estimation method based on rate functions of recurrent event processes. Their method can be used with regression models under certain conditions. The gist of their research is that they used point estimates based on Poisson models and developed robust variance estimates, which are still valid even if the assumed model is not a Poisson process.

Over the past two decades, many methodologies such as marginal and conditional methods have been developed to analyze multivariate survival data of recurrent events

((Prentice et al., 1981), (Andersen and Gill, 1982), (Wei et al., 1989), (Lee et al., 1992), (Pepe and Cai, 1993), (Lin et al., 2000)) . Moreover there has been interest in comparing these conditional and marginal methods, which can be found through works done by Cook and Lawless (2002), Cai and Schaubel (2004), Kelly and Lim (2000). Liang et al. (1993) discussed an approach for estimating parameters in a proportional hazards regression (Cox, 1972) type of specification for the recurrent event processes with external covariates. Liang et al. (1995) provided a survey of models and methods for analyzing multivariate failure time data including frailty and marginal models for recurrent events. Lin et al. (2000) proposed a semi-parametric regression for the mean and rate functions of recurrent events providing rigorous justification through modern empirical process theory. An important assumption of the above methods is independent censoring or as sometimes known as the conditionally independent censoring; see, Cook and Lawless (2007, Section 2.6) for more details.

Lawless et al. (1997) studied the mean and rate functions of recurrent events among survivors at certain time points. They suggested joint rate/mean function models for recurrent and terminal events by modeling marginal distribution of failure times and the rate function for the recurrent events conditional on the failure time. The objective of their paper was to present fairly simple methods for assessing the effects of treatments or covariates on recurrent event rates when other terminal events inducing the dependent censoring are present. Chen and Cook (2004) described methods for testing for differences in mean functions between treatment groups when each particular event process is ultimately terminated by death as a terminating event. They showed that the methods based on the assumption that the recurrent event process is independently terminated as a regular censoring may not be a valid assumption. There has been so many different models and methods for the statistical analysis of recurrent events and special cases, which can be found in the books by Daley and Vere-Jones (2003) and Cook and Lawless (2007) and the references given in them.

1.3 The Goal of The Thesis

The propensity score matching (PSM) is a statistical matching technique that attempts to estimate the effect of a treatment in the analysis of observational data, policy or another type of intervention by accounting for covariates which predict

whether receiving a treatment (Rosenbaum and Rubin, 1983). In other words, researchers intend to mimic the properties of randomized experimental designs with the propensity score (PS) techniques by trying to make the treatment and control groups similar on covariates that are believed to interfere with the correct estimation of treatment effect. After applying the PSM, the only difference between treatment and control groups would be the treatment in theory (Rosenbaum and Rubin, 1983).

The use of PSM in univariate survival analysis has been recently studied, but there has not been too much research on the estimation of treatment effects in the presence of recurrent events. In this thesis, we consider observational studies in which individuals are subject to recurrent events, and receive a certain type of treatment according to some characteristics of them. Furthermore, we investigate relationships between explanatory factors and an outcome, and also their incorporation in modeling PSM for recurrent events. We consider simple recurrent event models to investigate the effects of different matching methods in more detail. This allows us to get rid of the complexity added by the event generation models. We focus on a simple "treated" (i.e., treatment) versus "untreated" (i.e., control) groups case. In some settings, we discuss the situations where individuals switch from an existing treatment to a new treatment regimen.

We consider three matching methods; (i) propensity score matching, (ii) covariate matching, and (iii) history matching. Among these methods, the history matching is a respectively new matching method that can be applied only in event history settings, which includes recurrent events as a special case. This technique has not been extensively discussed in the literature. To our knowledge, it has only been applied in a restricted setting by Smith and Schaubel (2015) and Smith et al. (2018). We discussed this technique in two different settings. The first setting includes a time-fixed treatment assigned after the start of the follow-up of individuals in the study. In the second setting, we consider a time-varying treatment in a sense that the treatment is assigned at the start of their follow-up and may change at some point during their follow-up. In each setting, we conducted simulation studies with various scenarios. The studies and results are explained in Chapters 3 and 4. In Chapter 5, we present an illustrative analysis of a synthetic data set generated to mimic data sets obtained from studies of recurrent epileptic seizures in adults.

Our main objective with this thesis is to investigate the effects of these three different matching methods on the accuracy of the estimation of treatment effects in observational studies with recurrent events. The novelty of the study is the use of history information to match the treated and untreated subjects in the cohort. In other words, we investigate the information obtained from the past event occurrences experienced by individuals in observational studies to balance the baseline characteristics between treated and untreated groups in a cohort of individuals. Furthermore, we compare the accuracy of history matching method in the estimation of treatment effects with that of two popular matching methods.

Chapter 2

Propensity Score Models and Methods for the Analysis of Recurrent Events

In most biomedical and epidemiological studies, subjects on which measurements are taken are individuals. In this thesis, we consider that an individual is observed over a pre-specified time window. Occurrence times of a well defined event are recorded along with the value of other explanatory variables believed to affect the probabilistic characteristics of event occurrences. In this chapter, we set up the notation and introduce the models and methods used in this thesis.

For the analysis purposes, there are typically two structural forms of recurrent event data; times of event occurrences and waiting times between successive events, called the gap times. The former is usually applied when individuals frequently experience the events of interest, and the events are incidental, which means that their occurrences do not substantially alter the process under observation. Examples of incidental events include mild epileptic seizures or asthma attacks in humans. The analysis of gap times is usually conducted when events are relatively infrequent. The use of this type of data is common when the dependency between event occurrences is of interest.

After defining the event of interest, the individuals are selected to form the study cohort. Data can be obtained through a prospective or retrospective study. In a

prospective study, the selected individuals are longitudinally followed, and the event of interest occurring during their follow-up are recorded, while in a retrospective study the data is available for analysis purposes prior to the study design. Cook and Lawless (2007) provide many examples of recurrent event data arising from various research fields.

2.1 Basic Notation and Fundamental Concepts

In this section, we introduce the notation frequently used in the remaining parts of the thesis and some fundamental concepts. Recurrent event data are usually analyzed under the point process framework, where a process may undergo some sort of events repeatedly over time. Rigorous probabilistic treatment of point processes can be found in point process textbooks; e.g., in Daley and Vere-Jones (2003, 2007). We adapted a standard counting process notation given by Cook and Lawless (2007).

A *stochastic process* $\{W(t); t \in T\}$ is a family of random variables that is indexed by the element t in the index set T. In this thesis, the index t denotes the time so that $t \geq 0$. Therefore, $W(t)$ is a random variable representing the observable value of $w(t)$ at time t, where $t \in T$. A stochastic process is called a discrete-time process if the set T is finite or countable; otherwise, it is a continuous-time process (Daley and Vere-Jones, 2003). A *point process* is a probabilistic model for random scatterings of points on some space S often assumed to be a subset of R^d for some $d > 0$. Oftentimes, point processes describe the time or space occurrences of random events, in which the occurrences are revealed one-by-one as time evolves (Jacobsen, 2006). A *counting process*, denoted by $\{N(t); t \geq 0\}$, is a stochastic process, where $N(t)$ represents the cumulative number of events occurred over the time interval $(0, t]$ with the following properties: $N(0) = 0$, $N(t)$ is a positive integer, and if $s \leq t$ then $N(s) \leq N(t)$. If $s < t$, the notation $N(s, t)$ represents the number of event occurrences in the interval $(s, t]$; that is, $N(s, t) = N(t) - N(s)$.

Two important and commonly used counting processes are *Poisson processes* (PPs) and *renewal processes* (RPs). In many studies, the interest is in modeling either the mean or rate functions of a counting process. The *mean function* of a counting process

$\{N(t); t \geq 0\}$ is defined as

$$\mu(t) = E\{N(t)\}, \tag{2.1}$$

and the associated *rate function* $\rho(t)$ is the derivative of the mean function; that is,

$$\rho(t) = \mu'(t) = \frac{d}{dt}\mu(t), \tag{2.2}$$

where we assume that the expectation in (2.1) and derivative in (2.2) exist.

Let T be a non-negative and continuous random variable. The *cumulative distribution function* (c.d.f.) and *probability density function* (p.d.f.) of the random variable T are defined as $F(t) = \Pr(T \leq t)$ and $f(t) = (d/dt)F(t)$, respectively. The complement of the c.d.f. is called, the *survival function* $S(t)$, which gives the probability that an event has not occurred up to time t. Thus, we have

$$S(t) = \Pr(T \geq t) = 1 - F(t) = \int_t^\infty f(x)dx, \qquad t \geq 0. \tag{2.3}$$

Another non-negative function which can be used to characterize the distribution of T is the *hazard function* $h(t)$, which gives the instantaneous rate of an event occurrence at time t, given that the event has not been occurred up to time t. It is mathematically defined as

$$h(t) = \lim_{dt \to\ 0} \frac{\Pr\{t \leq T < t + dt \,|\, T \geq t\}}{dt}, \qquad t > 0. \tag{2.4}$$

It can be shown that $h(t) = f(t)/S(t)$, $t > 0$.

For events occurring in continuous time, we assume that two events cannot simultaneously occur. This assumption is sometimes called the *orderliness of a stochastic process* in the point process literature. A process possesses the orderliness property is called an *orderly process* (Cox and Isham, 1980, pp. 25-26). From now on, unless otherwise stated, we assume throughout the thesis that the index t represents the continuous time and all the processes are orderly.

A very important function to model the recurrent event processes is called *the intensity function*. Let $\{N(t); t \geq 0\}$ be a counting process. The associated intensity function is then denoted by $\lambda(t \,|\, H(t))$, where $H(t) = \{N(s); 0 \leq s < t\}$ is called *the history* of the process. The intensity function gives the instantaneous probability of an event occurring at time t, conditional on the process history $H(t)$, and is

mathematically defined as

$$\lambda(t \mid H(t)) = \lim_{\Delta t \to 0} \frac{\Pr\{\triangle N(t) = 1 \mid H(t)\}}{\triangle t}, \tag{2.5}$$

where $\triangle N(t) = N(t + \triangle t^-) - N(t^-)$ represents the number of events in the interval $[t, t + \triangle t)$. Note that the history $H(t) = \{N(s) : 0 \leq s < t\}$ records all information on event occurrences of the counting process $N(t); t \geq 0\}$ over the time interval $[0, t)$, which includes event occurrence times over $[0, t)$. The intensity function is important in the analysis of recurrent events because it completely defines an orderly counting process (Cook and Lawless, 2007, p. 10). Therefore, we use intensity functions to generate event times of recurrent event processes in our simulation studies in this thesis. Details of the use of the intensity function in simulation studies can be found in Section 2.4 in this chapter.

2.1.1 Covariates

Covariates play an important role in modeling recurrent events and estimating propensity scores (PSs). In the case of recurrent events, covariates could affect the probabilistic characteristics, that is the intensity function of a counting process. In probability score matching (PSM), covariates are crucial mainly because they are used to match individuals in a control group with individuals from a treatment group.

Covariates can be observed, unobserved, and are basically classified as external or internal (Kalbfleisch and Prentice, 2002, pp. 197-200). An *internal covariate* is one where the change of the covariate over time is related to the behavior of the individual, meaning that any change in a covariate is in sensible relationship with the condition of the individual. Examples of internal covariates include disease complications, blood pressure, etc. In contrast, an *external covariate* is one whose value is external to the individual. In other words, individuals under study cannot affect the value of external covariates, but external covariates may cause some specific change in individual's physical or mental health. For example, levels of air or water pollution can be classified as external covariates. Furthermore, a covariate is called *time-dependent* if its value changes over time or called *time-fixed* otherwise. Note that fixed covariates are naturally external.

We use the notation x or z to represent the value of a fixed covariate, and $x(t)$

or $z(t)$ to represent the value of a time-varying covariate. As discussed by Cook and Lawless (2007, Section 2.2), in recurrent events settings covariates can be included in the model as follows. Suppose that there are p covariates of interest in a study. We denote a vector of covariates by $\boldsymbol{x}(t) = (x_1(t), ..., x_p(t))'$ and the history of covariates up to time t by $x^{(t)} = \{\boldsymbol{x}(u); 0 \leq u \leq t\}$. The history of a counting process is then extended to include the information on covariates so that the intensity function depends on covariates through the history of the process; that is, we define

$$\lambda(t \,|\, H(t)) = \lim_{\Delta t \to 0} \frac{\Pr\{\triangle N(t) = 1 \,|\, H(t)\}}{\triangle t},$$

where $H(t) = \{N(s), \boldsymbol{x}(u);\ 0 \leq s < t,\ 0 \leq u \leq t\}$. Note that the history $H(t)$ includes information on the counting process $\{N(t); t \geq 0\}$ over $[0, t)$ but information on covariates over $[0, t]$, which means that the value of the covariate process $\boldsymbol{x}(t)$ is known in the intensity function at time t. More discussion on the extended history functions to include covariates can be found in Daley and Vere-Jones (2003) and Cook and Lawless (2007).

2.2 Fundamental Models

In this section, we describe the basic families for recurrent event processes such as PPs and RPs that will be used in subsequent chapters for describing and analyzing data.

2.2.1 Poisson Processes

The *Poisson process* (PP) is one of the most widely-used counting processes. It is usually used in scenarios where we count the occurrences of certain events that appear to happen at a certain rate, but completely at random; that is, without a certain structure. For example, suppose that from the past data, we know that heart attacks happen to an individual with a rate of two per year. Other than this information, the timings of heart attacks seem to be completely random. In such a case, the PP might be a good model for making inference on the rate of heart attacks.

In modeling recurrent event processes, a PP describes a situation, in which events

occur randomly in such a way that the number of events in non-overlapping time intervals are independent. PPs are also suitable when there are external covariates which affect occurrence of events. It is worth mentioning that PPs or other models which are based on counts are appropriate for incidental events where their occurrence does not change the process itself.

There are various equivalent ways of defining a PP. One way of defining a PP is through its intensity function (Cook and Lawless, 2007, Section 2.1.1). Let $\{N(t); t \geq 0\}$ be a counting process with the intensity function $\lambda(t|H(t))$. Then, $\{N(t); t \geq 0\}$ is called a PP if the intensity function is of the form

$$\lambda(t \,|\, H(t)) = \rho(t), \qquad t > 0. \tag{2.6}$$

It is obvious that the Poisson process intensity function (2.6) does not depend on the history of the process $H(t)$, meaning that in the absence of covariates, intensity is specified only by t. This fact is a result of the independent increment property of the PPs, which shows that the PPs possess the Markov property (Cook and Lawless, 2007, p. 32). For the special case, in which $\rho(t) = \rho > 0$ (i.e. a positive constant), the process $\{N(t); t \geq 0\}$ is called a *homogeneous Poisson process* (HPP); otherwise, it is called a *non-homogeneous Poisson process* (NHPP). It should be noted that a HPP $\{N(t); t \geq 0\}$ with the rate function $\rho > 0$ has the following properties:

- $N(0) = 0$,
- $\{N(t); t \geq 0\}$ has independent increments,
- the number of events in any interval of length s is a Poisson random variable with the mean function $\mu(s) = \rho s$, $s \geq 0$.

A proof of this result can be found in Daley and Vere-Jones (2003). For any PP $\{N(t);\ t \geq 0\}$, the following results can be obtained from the intensity function given in (2.6) (Daley and Vere-Jones, 2003).

(i) N(0) = 0.

(ii) $N(s,t)$ has a Poisson distribution with mean $\mu(s,t) = \mu(t) - \mu(s)$, for $0 \leq s < t$.

(iii) Let $(s_1, t_1]$ and $(s_2, t_2]$ be any two non-overlapping intervals, then $N(s_1, t_1)$ and $N(s_2, t_2)$ are independent random variables.

The following result is the key to generate realizations of a NHPP through the Monte Carlo simulations. A proof of it can be found in Daley and Vere-Jones (2003).

Proposition 2.2.1. *Let* $\{N(t); t \geq 0\}$ *be a NHPP with the mean function* $\mu(t)$*. Then,* $\{N^*(s); s \geq 0\}$ *is a HPP with the rate function* $\rho^*(s) = 1$ *if we define* $s = \mu(t)$ *and*

$$N^*(s) = N(\mu^{-1}(s)), \qquad s > 0.$$

Therefore, by generating event times of a HPP with rate function $\rho^*(s) = 1$, we can consequently generate event times of a NHPP using the relation $t = \mu^{-1}(s)$.

The external covariates affecting the event occurrence rate can be easily incorporated in PP models through the intensity function (2.6). These covariates can be involved in a PP by redefining the history of the associated intensity (i.e., rate) function to include covariate information. As discussed above, the intensity function of a PP at time t depends only on t, and is not a function of the past of the process; i.e., the history $H(t)$. Covariates in PPs can be included in the intensity function as follows. Let $\boldsymbol{x}(t)$ be a p-dimensional vector of time-fixed and/or time varying covariates. We define $\boldsymbol{z}(t) = (z_1(t), ..., z_q(t))^{'}$, $q \geq p$, as a q-dimensional vector of covariates, whose elements include $\boldsymbol{x}(t)$, as well as functions of t in the case if the model depends on that. The intensity function can be defined then as

$$\lambda(t \,|\, H(t)) = \rho(t \,|\, x^{(t)}) = \rho_0(t) \exp(\boldsymbol{z}^{'}(t)\, \boldsymbol{\beta}), \tag{2.7}$$

where $\boldsymbol{\beta}$ is a q-dimensional vector of parameters and $\rho_0(t)$ is called *the baseline rate function* of the process $\{N(t); t \geq 0\}$. The model (2.7) is usually called the *multiplicative model*, in which the effect of covariates $\boldsymbol{z}(t)$ on the rate function is assumed to be of log-linear form. The multiplicative model is the most common family of regression models for recurrent events. Therefore, we consider only the multiplicative models in this thesis. However, if the log-linearity assumption of the multiplicative model (2.7) is not valid, additive or time transform models can be used for regression in recurrent events as well. The multiplicative model (2.7) is fully parametric if the rate function including both the baseline rate function and the exponential function is determined parametrically. If the baseline rate function is free of parameters but the exponential function in (2.7) is parametrically specified, the model is semi-parametric, which is sometimes called the *the Andersen-Gill model* (Cook and Lawless, 2007).

It should be noted that the intensity function (2.7) can represent a PP if and only if the covariates are external. The model including internal covariates is no longer a PP, but can be specified as a general intensity-based process. These models are useful in particular when there is a need for modeling the past of a process. Since we only focus on PPs in this thesis, we do not consider the general intensity-based models. However, the methods discussed in the following chapters can be extended to deal with such models. The general intensity-based models are discussed by Cook and Lawless (2007, Chapter 5).

2.2.1.1 Mixed Poisson Processes

Poisson models are useful in some settings and applications, but the main drawback of using them to model recurrent events is that usually real-life data sets are overdispersed and exhibit variability in the number of event occurrences beyond the amount predicted by Poisson models. This situation usually occurs whenever there is heterogeneity among subjects due to some unmeasured factors or subject specific effects that influence event rates (Cook and Lawless, 2007, p. 35). Such a heterogeneity is called the *unexplained heterogeneity*. In such situations, even after conditioning on observed covariates, $Var\{N(t)\}$ appears to be substantially larger than $E\{N(t)\}$. Since under a Poisson model the mean and the variance of $N(t)$ need to be equal, the use of Poisson models is therefore no longer plausible when unexplained variability is present in a given data set.

This issue can be addressed by incorporating unobservable random effects. To explain this, we now consider a cohort of m individuals, and introduce the index i to denote the ith individual process, where $i = 1, \ldots, m$. Following the notation given in Cook and Lawless (2007, Section 2.2.3), we let u_i denote the unobserved random effect for the ith individual, $i = 1, \ldots, m$. For simplicity we assume that $\boldsymbol{z}_i$ denotes a p-dimensional vector of time-fixed covariates. The results in this section can be extended to the external time-varying covariates case as well. Conditional on covariates $\boldsymbol{z}_i$ and the random effect u_i, the mixed Poisson model of the process $\{N_i(t); t \geq 0\}$ is then given with the intensity function

$$\rho(t \mid \boldsymbol{z}_i, u_i) = u_i \, \rho_0(t) \, \exp(\boldsymbol{z}_i' \boldsymbol{\beta}), \qquad t > 0, \tag{2.8}$$

where the u_i are i.i.d. random variables following a distribution function $G(u)$ with a finite mean. It should be noted that, even tough the model given in (2.8) is a Poisson process for the given value of u_i, the marginal process $\{N_i(t); t \geq 0\}$ is not a PP in general.

We may assume without loss of generality that $E(u_i) = 1$ and $Var(u_i) = \phi$, where $\phi > 0$. Any c.d.f. under these assumptions can be used to model the random effects u_i. The most commonly used distribution for the u_i is however the gamma distribution as it would make the multiplicative mixed Poisson model (2.8) mathematically more convenient to work with. In this case, the u_i have a gamma distribution with mean 1 and variance ϕ, and the p.d.f. of the form

$$g(u; \phi) = \frac{u^{\phi^{-1}-1} \exp(-u/\phi)}{\phi^{\phi^{-1}} \Gamma(\phi^{-1})}, \qquad u > 0. \tag{2.9}$$

Let $\mu_i(s,t)$ denote the expected number of events in $\{N_i(t); t \geq 0\}$ over the time interval $(s,t]$, where $0 < s < t$; that is, $\mu_i(s,t) = E\{N_i(s,t)\} = E\{N_i(t) - N_i(s)\}$. Then, by definition, for $i = 1, \ldots, m$,

$$\mu_i(s,t) = \int_s^t \rho_0(v) \exp(\boldsymbol{z}_i' \boldsymbol{\beta})\, dv = \mu_0(s,t) \exp(\boldsymbol{z}_i' \boldsymbol{\beta}). \tag{2.10}$$

Given $\boldsymbol{z}_i$ and u_i, the random variable $N_i(s,t)$ follows a Poisson distribution with mean function $\int_s^t \rho(t \mid \boldsymbol{z}_i, u_i) = u_i\, \mu_i(s,t)$. Note that, given only $\boldsymbol{z}_i$, the distribution of $N_i(s,t)$ is no longer Poisson but is negative binomial with probability function of the form

$$\begin{aligned} \Pr(N_i(s,t) = n \mid \boldsymbol{z}_i) &= \int_0^\infty \frac{[u\mu_i(s,t)]^n}{n!} \exp\{-u\mu_i(s,t)\}\, g(u;\phi)\, du, \\ &= \frac{\Gamma(n+\phi^{-1})}{\Gamma(\phi^{-1})} \frac{[\phi\, \mu_i(s,t)]^n}{[1+\phi\, \mu_i(s,t)]^{n+\phi^{-1}}}, \qquad n = 0, 1, 2, ... \end{aligned} \tag{2.11}$$

Note that the limit as $\phi \to 0$ gives the Poisson distribution (Cook and Lawless, 2007, p. 36). Therefore, the model converges to a Poisson process in the limit when $\phi \to 0$. However, the case $\phi > 0$ represents overdispersion for the Poisson model, and the process becomes a negative binomial process for which the intensity function at time

t can be expressed as

$$\lambda_i(t \mid H_i(t)) = \frac{(1 + \phi\, N_i(t^-))}{1 + \phi\, \mu_i(t)}\, \rho_i(t), \qquad t \geq 0, \tag{2.12}$$

(Cook and Lawless, 2007, p. 37). The level of the overdispersion in the observed event counts are defined by the parameter ϕ. A high value of of ϕ represents a more pronounced overdispersion (i.e., unexplained heterogeneity) in the event counts across individual processes. Because of this reason, the parameter ϕ is sometimes called *the heterogeneity parameter* of the mixed Poisson process.

We now represent the expressions for the marginal mean and variance of $N_i(s,t)$ based on the random effects model (2.8). It is easy to see that the marginal mean is given by

$$\begin{aligned} E\{N_i(s,t)\} &= E\{E[N_i(s,t)|u_i, \boldsymbol{z}_i]\}, \\ &= E\{u_i\, \mu_i(s,t)\}, \\ &= \mu_i(s,t), \end{aligned} \tag{2.13}$$

and the marginal variance is given by

$$\begin{aligned} Var\{N_i(s,t)\} &= E\{Var[N_i(s,t) \mid u_i, \boldsymbol{z}_i]\} + Var\{E[N_i(s,t) \mid u_i, \boldsymbol{z}_i]\}, \\ &= E\{u_i \mu_i(s,t)\} + Var\{u_i \mu_i(s,t)\}, \\ &= \mu_i(s,t) + \phi\, \mu_i(s,t)^2. \end{aligned} \tag{2.14}$$

Moreover, the marginal covariance for event counts over non-overlapping intervals can be written as

$$Cov\{N_i(s_1,t_1), N_i(s_2,t_2)\} = \phi\, \mu_i(s_1,t_1)\, \mu_i(s_2,t_2). \tag{2.15}$$

It is worth mentioning that relationships (2.13), (2.14) and (2.15) hold for any distribution function for the u_i.

2.2.2 Renewal Processes

A *renewal process* (RP) is a stochastic process model for recurrent events that randomly occur in time and are subject to some sort of "renewal" after each event occurrence. As defined in this section, RPs have a very strict conditions by definition, which limits their use for many applications. However, they can be modified for building more realistic models. In this section, we introduce only some basic RP models and a few extensions of them. More details on RPs can be found in Daley and Vere-Jones (2003) and Cook and Lawless (2007).

Let T_j, $j = 1, 2, \ldots$, be the occurrence time of the jth event, which is usually called the jth *arrival time*, of the counting process $\{N(t); t \geq 0\}$ with the associated intensity function $\lambda(t \,|\, H(t))$, and let $T_0 = 0$. Then, $W_j = T_j - T_{j-1}$, $j = 1, 2, \ldots$, is called the jth gap time; that is, the time between the $(j-1)$st and jth events. RPs are defined as stochastic processes in which the gap times between successive events are independent and identically distributed. The definition of the RPs is analogous to the case where the intensity function (2.5) is of the form

$$\lambda(t \,|\, H(t)) = h\left(t - T_{N(t^-)}\right), \tag{2.16}$$

where $t - T_{N(t^-)}$ is clled *the backward recurrence time*; that is, the elapsed time since the most recent event before time t, and $h(\cdot)$ is the hazard function of the gap times W_j as defined in (2.4).

The distribution of counts $N(s,t)$ in a RP is often of interest. When the W_j are exponentially distributed, the corresponding counting process $\{N(t); t \geq 0\}$ is equivalent to a HPP, and thus $N(s,t)$ follows a Poisson distribution with the mean $\mu(s,t)$. It is however not easy to obtain the distribution of counts in other cases. The following relation can be useful to obtain the distribution of $N(t)$ in some cases.

$$\Pr(N(t) \geq n) = \Pr(T_n \leq t), \qquad n = 0, 1, \ldots, \tag{2.17}$$

where $T_n = W_1 + \cdots + W_n$ is a sum of n i.i.d. random variables. Using (2.17), it can be shown that $\Pr(N(t) = n) = \Pr(T_n \leq t) - \Pr(T_{n+1} \leq t)$, and consequently

$$\mu(t) = E\{N(t)\} = \sum_{n=1}^{\infty} F_n(t), \tag{2.18}$$

where $F_n(t)$ is the distribution function of T_n.

Covariates can be incorporated in RPs in a similar way explained in Section 2.2.1. If there are time-fixed covariates $\boldsymbol{z}$, which are believed to affect the RP, we can let the distribution of the gap times W_j depend on the covariates $\boldsymbol{z}$. Since the gap times W_j are positive valued, it is possible to apply regression models used in connection with lifetime data (e.g. Lawless (2003)). For doing so, there are two well-known regression models; (i) the proportional hazard model, where the hazard function of W_j conditional on $\boldsymbol{z}$ is given by

$$h(w \mid \boldsymbol{z}) = h_0(w)\, \exp(\boldsymbol{z}^{'}\boldsymbol{\beta}), \qquad w > 0, \tag{2.19}$$

and (ii) the accelerated failure time model, where the hazard function of W_j given $\boldsymbol{z}$ is

$$h(w \mid \boldsymbol{z}) = h_0(w \exp(\boldsymbol{z}^{'}\boldsymbol{\beta}))\, \exp(\boldsymbol{z}^{'}\boldsymbol{\beta}), \qquad w > 0. \tag{2.20}$$

In (2.19) and (2.20), the function $h_0(w)$ is called the baseline hazard function of W_j.

If the external time-varying covariates are of interest, they can be included in a RP with the intensity function

$$\lambda(t \mid H(t)) = h(t - T_{N(t^-)} | \boldsymbol{z}(t)) \tag{2.21}$$

This can be done in a similar way to the case where we incorporate time-varying covariates to the hazard function of W_j; that is,

$$h(w \mid \boldsymbol{z}(t)) = h_0(w)\, \exp(\boldsymbol{z}^{'}(t)\, \beta), \tag{2.22}$$

where $t = w + t_{N(t^-)}$. Since we mainly focus on the PPs, we do not discuss the RPs in detail. More information on regression models of recurrent events and beyond can be found in Cook and Lawless (2007, Chapter 4).

2.3 Propensity Score Matching

In this section, we first discuss the treatment evaluation and some examples, and then, introduce the propensity score (PS) methodology. Propensity scoring is used to properly analyze data obtained from observational studies. In such studies, researchers

do not conduct randomized controlled trials to make causal inference, instead some pretreatment characteristics of individuals are used to find the propensity score.

In many fields of study, the primary goal is to evaluate the effectiveness of a program, which typically means the comparison of the effects of that program on the outcome of interest with the effects of another program or a placebo. Examples of treatment evaluation can be the effect of a new medicine on epileptic seizures, effect of training programs on job performance or government programs targeted to help school and their effect on student performance. Note that in these studies, unlike lab experiments, individuals decide whether to participate in the program or not. Since individuals who decide to participate are different in terms of various characteristics from individuals who do not participate, it is statistically imprudent to directly compare the outcome of interest. Therefore, we need to balance the observed and unobserved outcome-related covariates between treatment and control groups and then compare their outcomes. Below are the assumptions and the procedure required for conducting the propensity score matching (PSM) technique initiated by Rosenbaum and Rubin (1983).

Let y_0 and y_1 be the potential outcomes for the control group and treatment group, respectively. There exists a set x of observable covariates such that after controlling for these covariates, the potential outcomes are independent of treatment assignment; that is, in notation,

$$y_0, y_1 \perp Trt \,|\, x,$$

where Trt is a binary variable such that $Trt = 1$ corresponds to the treated observations and $Trt = 0$ corresponds to the control observations. This assumption is known as conditional independence and requires that all variables relevant to the probability of receiving the treatment may be observed (Rosenbaum and Rubin, 1983). As a result, it allows the untreated units to be used to construct an unbiased counterfactual for the treatment group.

Another assumption required to apply the PSM methods is that, for each value of x, there must be both treated and control observations. In the other words, for each treated observation, there is a matched control observation with similar x values.

This assumption is known as common support and is given by

$$0 < \Pr(Trt = 1 \,|\, x) < 1.$$

The assumption of common support ensures that there is sufficient overlap in the characteristics of treated and untreated units to find adequate matches. When these assumptions are satisfied, the treatment assignment is said to be strongly ignorable in the terminology of Rosenbaum and Rubin (1983).

The procedure for estimating the effect of a treatment can be divided into three steps:

- Step 1: Estimate the PS of individuals in a cohort.
- Step 2: Choose a matching algorithm that will use the estimated PSs to match untreated units with treated units.
- Step 3: Estimate the effect of the treatment with the matched sample and calculate standard errors.

The PS can be statistically defined as

$$p(x) = E(Trt \,|\, x) = \Pr(Trt = 1 \,|\, x). \tag{2.23}$$

A binary outcome model is usually employed to estimate the PS given in (2.23) for each subject under a study. Logit and probit models are the commonly used binary outcome models in developing PS methods. These models are used to estimate the probability of receiving a treatment conditional on the observed pretreatment measurements. It is essential that a flexible functional form be used to allow for possible nonlinearities in the participation model.

After defining a suitable binary outcome model and estimating the propensity scores for each subject, we need to apply a matching algorithm to match subjects in the treatment group with subjects in the control group so that we may be able to calculate the treatment effect in an observational study. Note that, here our goal is to find a match or matches for each subject in the treatment group not for the subjects in the control group. Figure 2.1 gives a visual representation of how the PSM methods works. In this figure, the y-axis shows the estimated propensity scores

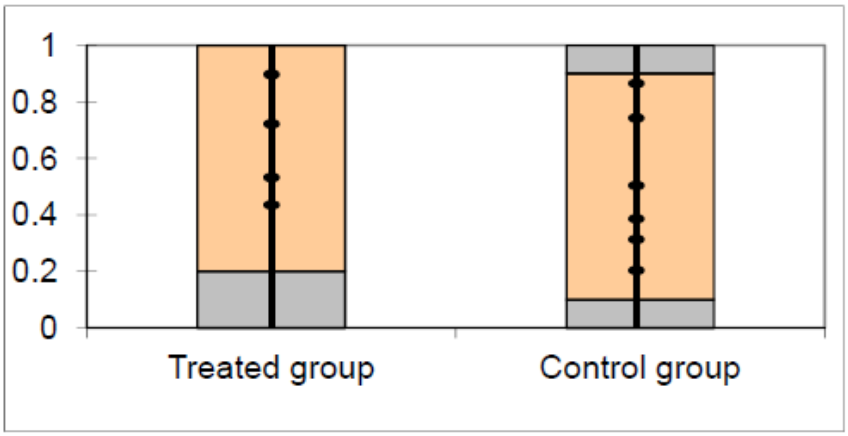

Figure 2.1: Predicted probabilities or propensity scores of subjects in the treated and control groups.

of four individuals in the treated group and six individuals in the control group. There are many matching methods available for different situations including kernel matching, nearest neighbor, radius (or caliper) and stratification, which are briefly explained below by using the example given in Figure 2.1.

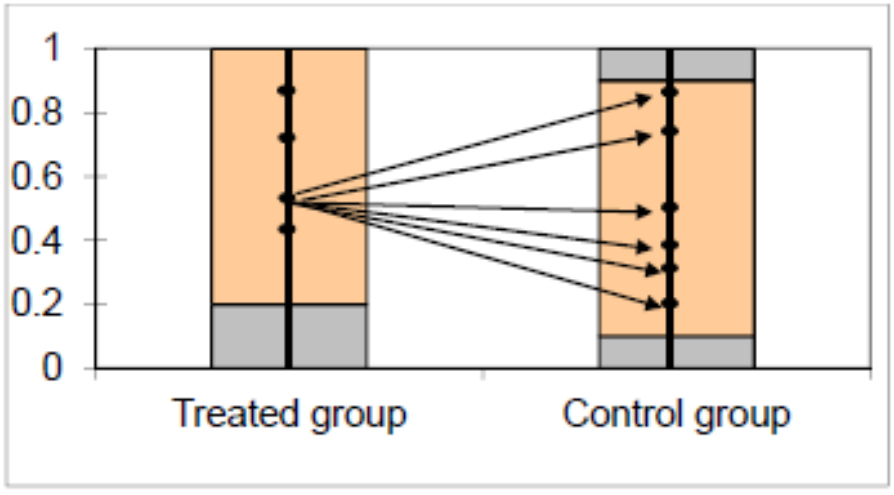

Figure 2.2: Illustrative figure of the kernel matching method.

In the *kernel matching* method, each subject from the treatment group is matched with the weighted average of all the control subjects. In this matching method, we need to weigh each individual in the control group based on their PSs, where the individual with closest PS to the one of the treated subject get the highest weight,

and so on. In other words, the weights are inversely proportional to the distance between the treatment and control group's PSs. Figure 2.2 shows how the kernel matching method works. In this method, a weight for the treated subject i and control subject j, denoted as $w(i,j)$, is defined by

$$w(i,j) = \frac{K\left(\frac{p_j - p_i}{h}\right)}{\sum_{j=1}^{n_C} K\left(\frac{p_j - p_i}{h}\right)}, \tag{2.24}$$

where $K(.)$ is a prespecified kernel function which in fact is a weighting function used in non-parametric estimation techniques, h is bandwidth parameter and n_C denotes the number of individuals in the matched control group. A difficulty linked to kernel matching method is selecting an appropriate bandwidth parameter which can affect the bias and variance directly (Imbens, 2004).

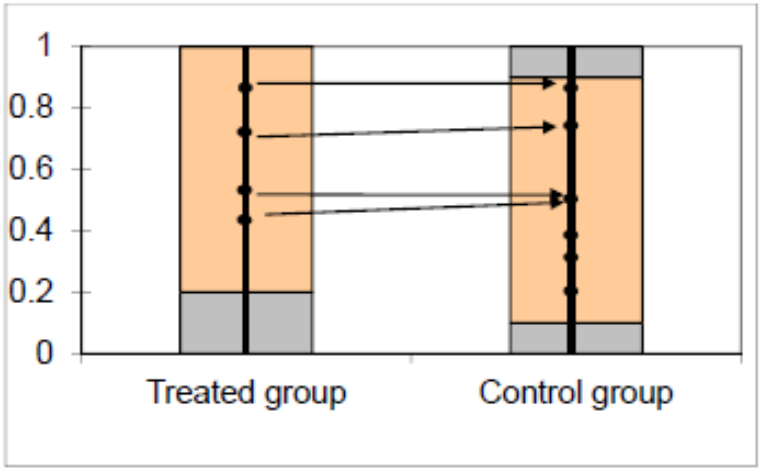

Figure 2.3: Illustrative figure of the nearest neighbor matching method.

Another matching method is called the *nearest neighbor matching*, for which we match a subject from treatment group with a subject from control group whose PS, in comparison to others, is closest in value to the one for the treated subject. It should be noted that, although not common, the PSM methods can be applied with replacement; that is, if we are using PSM with replacement, it is possible to use an untreated individual more than one time as a match. The nearest neighbor matching method is easy to implement and understand. However, one of the major issues

involved in this matching method is that it may result in some bad matches if the PSs of the matched subjects are far from each other. Let p_i and p_j be the PSs for two observations from treatment and control groups respectively, then

$$min\|p_i - p_j\|, \tag{2.25}$$

determines the match, where $\|.\|$ denote the absolute-value norm. Figure 2.3 illustrates the nearest neighbor matching method.

In the *radius matching* method, we only need to put a certain radius, and choose all the control observations that fall within the radius. In this method, matches are based on the inequality

$$\|p_i - p_j\| < r, \tag{2.26}$$

where r is a pre-specified radius. Figure 2.4 illustrates the radius matching method. As shown in this figure, all the control subjects that fall inside the circle can be used as matches for the selected treated individual. The main advantage of using radius matching is that it is possible to use all the observations in the control group, which results in an increase in the estimation precision. In the case of having poor matches when PSs are not close enough, we can use the radius (or caliper) matching method as an alternative to nearest neighbor method (Rosenbaum and Rubin, 1985b).

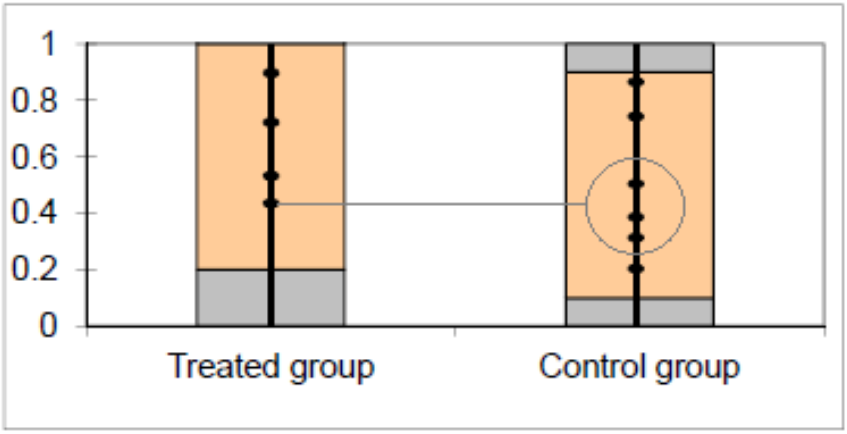

Figure 2.4: Illustrative figure of the radius matching method.

Finally for the *stratification matching* method, we need to divide the observations into blocks based on the estimated PSs and for observations that fall in a certain

block we use the individuals in the matching block and the difference estimated as the average of within-stratum effects. Theoretical and empirical results indicate that the popular version of stratification via estimated propensity scores based on within-stratum sample mean differences and a fixed number of strata may lead to biased inference due to residual confounding and this bias leads to more misleading results as sample size increases, therefore caution must be taken in stratifying on quintiles (Lunceford and Davidian, 2004).

After choosing an appropriate matching method and defining matches, we need to calculate the effect of the treatment. The common way to calculate treatment effect is through the following formula.

$$ATE = E(Y_1 \,|\, x, Trt = 1) - E(Y_0 \,|\, x, Trt = 0), \tag{2.27}$$

where ATE stands for average treatment effect. ATE is suitable for randomized experiments, where there are usually little differences between observations in treatment and control groups. Therefore, we need to calculate the average treatment effect on the treated (ATET), which is the difference between the outcomes of the treated observations and the outcomes of the treated observations if they were not treated; this is, in notation,

$$ATET = E(Y_1 \,|\, x, Trt = 1) - E(Y_0 \,|\, x, Trt = 1). \tag{2.28}$$

The second term in (2.28) cannot be calculated as it is not possible to observe the outcome y_0 for observations who receive the treatment ($Trt = 1$). In this situation, we can apply PSM using which we can estimate the treatment effect by comparing the outcomes of the matched control subjects with the outcomes of the matched treated subjects.

$$ATET = E(Y_1 \,|\, p(x), Trt = 1) - E(Y_0 \,|\, p(x), Trt = 0). \tag{2.29}$$

The empirical estimate of the treatment effect is equal to

$$ATET = \frac{1}{n_{Trt}} \sum_{i \in \{Trt=1\}} \left[y_{1,i} - \sum_j w(i,j) y_{0,j} \right], \tag{2.30}$$

where the $w(i,j)$ represent the weights, and n_{Trt} denotes the number of individuals in the matched treated group. Note that, if no weighting methods are used, then the

$w(i,j)$ are equal to 1.

After estimating the treatment effect, it is recommended to verify whether the treatment assignment is independent of the observed measurments x, given the associated propensity score. This can be statistically shown as

$$Trt \perp x|p(x).$$

This assumption is known as balancing condition, and is testable (Senn, 1994). Balancing tests consider whether the estimated propensity score adequately balances characteristics between the treatment and control group units.

2.4 Simulation Procedures for Recurrent Event Processes

In this section, we introduce simulation methods used for generating realizations of a recurrent event process with a general intensity function. The generated event times will be used later for assessing the effects of new programs or treatments applied on some groups of people. Let $\{N(t);\ t \geq 0\}$ denote a counting process with intensity function (2.5), then the probability density of n events occurred at times $0 < t_1 < t_2 < \cdots < t_n$ over the determined interval $[\tau_0, \tau]$ conditional on the history $H(\tau_0)$ is

$$\prod_{j=1}^{n} \lambda(t_j|H(t_j)) \exp\left\{-\int_{\tau_0}^{\tau} \lambda(u|H(u))du\right\}. \tag{2.31}$$

A derivation of the above result can be found in Cook and Lawless (2007, Section 2.1). The differences between successive events T_j generated by the counting process $\{N(t); t \geq 0\}$ result in waiting times $W_j = T_j - T_{j-1}$, $(j = 1, 2, ...)$, where $T_0 = 0$. The survival function of W_j, the waiting time between $(j-1)$st and jth events, conditional on $H(t_{j-1})$ and t_{j-1} is given by (Cook and Lawless, 2007, Section 2.1),

$$\Pr\{W_j > w \,|\, T_{j-1} = t_{j-1}, H(t_{j-1})\} = \exp\left\{-\int_{t_{j-1}}^{t_{j-1}+w} \lambda(u|H(u))du\right\}. \tag{2.32}$$

Using the result given in (2.32) and the fact that any continuous and strictly increasing

c.d.f. of a random variable follows a standard uniform distribution, it can be easily shown that given t_{j-1} and $H(t_{j-1})$ the random variable

$$E_j = \int_{t_{j-1}}^{t_{j-1}+W_j} \lambda(u \,|\, H(u))\, du, \tag{2.33}$$

follows a standard exponential distribution. Therefore, we can generate event times $t_j = t_{j-1} + w_j;\ j = 1, 2, 3, ...,$ by generating E_j and solving the equation (2.33) for each W_j. Note that here $t_0 = 0$.

In the case of a PP with the rate function $\rho(t)$, the result (2.32) is equal to

$$\Pr\{W_j > w \,|\, T_{j-1} = t_{j-1}, H(t_{j-1})\} = \exp\{-\mu(t_{j-1}, t_{j-1} + w)\}, \qquad j = 1, 2, ..., \tag{2.34}$$

where $\mu(t_{j-1}, t_{j-1} + w) = \int_{t_{j-1}}^{t_{j-1}+w} \rho(s)\, ds$. Following this result, for a HPP with the rate function ρ, we obtain

$$\Pr\{W_j > w \,|\, T_{j-1} = t_{j-1}, H(t_{j-1})\} = \exp(-\rho w), \qquad w > 0. \tag{2.35}$$

Using the result in (2.35), we can simulate a HPP, which can be used for simulating a NHPP as explained in Proposition 2.2.1. This simulation method is useful when (2.33) cannot be easily solved.

Following steps elaborate the computer simulation procedure for generating event times of a given intensity function over the time interval $[0, \tau]$.

1. Set $j = 1$ and $t_{j-1} = 0$.
2. Generate E_j from a standard exponential distribution.
3. Replace E_j in (2.33) with the generated value obtained from the second step.
4. Solve the equation $E_j = \int_{tj-1}^{t_{j-1}+W_j} \lambda(u|H(u))du$ by solving nonlinear equations in order to find the waiting time W_j.
5. Calculate the jth event time $T_j = t_{j-1} + W_j$.
6. If T_j is less than the upper bound τ, then set $j = j + 1$, $t_{j-1} = T_{j-1}$ and return to the second step. Otherwise, break the loop and the calculated values $t_1, t_2, ..., t_{j-1}$ are the recurrent event times.

If there are external covariates that are of interest, the intensity function $\lambda(t \mid H(t))$ can be extended with covariates in the above algorithm. For a more detailed explanation regarding simulation methods refer to Cook and Lawless (2007, pp. 44-45 and Problem 2.2).

2.5 Construction of the Likelihood Function

Suppose that there are m independent counting processes under observation. The ith process, $i = 1, \ldots, m$, is observed over the observation window $[\tau_{i0}, \tau_i]$, where τ_{i0} and τ_i are, respectively, the starting and end of the follow-up times of the ith process. Let $t_{i1} < t_{i2} < \cdots < t_{in_i}$, $i = 1, \ldots, m$, denote the n_i event times experienced by the ith process. Then, the contribution of the ith process to the likelihood function $L(\boldsymbol{\theta})$ can be expressed as

$$L_i(\boldsymbol{\theta}) = \prod_{j=1}^{n_i} \lambda_i(t_{ij}|H_i(t_{ij})) \exp\left\{ - \int_{\tau_{i0}}^{\tau_i} \lambda_i(u|H_i(u))du \right\}, \tag{2.36}$$

where $\boldsymbol{\theta}$ is a parameter vector specifying the intensity function. The likelihood function for the m independent processes is the product of such terms, which is

$$L(\boldsymbol{\theta}) = \prod_{i=1}^{m} L_i(\boldsymbol{\theta}) = \prod_{i=1}^{m} \prod_{j=1}^{n_i} \lambda_i(t_{ij}|H_i(t_{ij})) \exp\left\{ - \int_{\tau_{i0}}^{\tau_i} \lambda_i(u|H_i(u))\, du \right\}. \tag{2.37}$$

The derivation of the above likelihood function can be found in Cook and Lawless (2007, Section 2.6). In the case of mixed Poisson processes with random effects, where the random effects u_i follows a gamma distribution with mean 1 and variance ϕ, the likelihood function for m independent processes is of the form

$$L(\boldsymbol{\theta}, \phi) = \prod_{i=1}^{m} \left\{ \prod_{j=1}^{n_i} \frac{\rho_i(t_{ij})}{\mu_i(\tau_i)} \right\} \frac{\Gamma(n_i + \phi^{-1})}{\Gamma(\phi^{-1})} \frac{(\phi\mu_i(\tau_i))^{n_i}}{(1 + \phi\mu_i(\tau_i))^{n_i + \phi^{-1}}}, \tag{2.38}$$

where $\mu_i(t) = \int_{\tau_{i0}}^{t} \rho_i(s)ds$. This result is given in Cook and Lawless (2007, p. 36).

In studies where the subjects are intermittently observed or cease to be at risk temporarily it is useful to denote when an individual or process is under observation and at risk of an event. This can be done with the at-risk indicator $Y(t)$. For example,

if the ith subject is observed over the interval $[\tau_{i0}, \tau_i]$ and under risk of having an event over the observation window, the at-risk indicator is $Y_i(t) = I(\tau_{i0} \leq t \leq \tau_i)$.

Sometimes it is more convenient to write down the likelihood function by using the at-risk indicator $Y(t)$. Following the notation given by Cook and Lawless (2007, Section 2.6), the observed part of the counting process $\{N(t); t \geq 0\}$, called the observable process, can be written as $\bar{N}(t) = \int_{\tau_0}^{t} Y(u)dN(u)$ with the intensity function

$$\bar{\lambda}(t \mid \bar{H}(t)) = \lim_{\Delta t \downarrow 0} \frac{\Pr(\Delta \bar{N}(t) = 1 \mid \bar{H}(t))}{\Delta t}, \qquad t \geq \tau_0, \tag{2.39}$$

where $\bar{H}(t) = \{\bar{N}(s), Y(u); \tau_0 \leq s < t, \tau_0 \leq u \leq t\}$ is the history of the observable process. If $\Delta N(t)$ and $Y(t)$ are conditionally independent given $H(t)$, then $\bar{\lambda}(t|\bar{H}(t)) = Y(t)\lambda(t|H(t))$ (Cook and Lawless, 2007, Section 2.6) and the complete likelihood function for m independent processes can be written as

$$L(\boldsymbol{\theta}) = \prod_{i=1}^{m} L_i(\boldsymbol{\theta}) = \prod_{i=1}^{m} \prod_{j=1}^{n_i} \lambda_i(t_{ij}|H_i(t_{ij})) \exp\left\{ - \int_0^{\infty} Y_i(u)\lambda_i(u|H_i(u))du \right\}. \tag{2.40}$$

The likelihood function (2.40) is not only valid for the case where an individual process is intermittently observed but can also be used when starting and end of follow-up times are random as stopping times (Cook and Lawless, 2007, Section 2.6).

Chapter 3

Estimation of Time-Fixed Treatment Effects

Estimating efficacy and effectiveness of treatments has been an appealing subject to medical and health specialists. Despite the fact that there have been several methods and approaches developed by researchers to draw rigorous causal inference, caution should be exercised when using those techniques as they may lead to biased inference if not applied properly. In this chapter, we investigate the capability and accuracy of different propensity score matching (*PSM*) models, history matching (*HM*) and crude matching on observed covariates, what we refer to as covariate matching (*CM*), methods in estimation of time-fixed treatment effects. Moreover, we discuss the advantages and disadvantages of these matching techniques and provide some guidelines on how to improve them.

3.1 Models and Methods

In this section, we introduce the models used in our Monte Carlo simulation studies to examine the bias arised from different *PSM* models, *HM* and *CM* methods. Our discussion includes a detailed explanation of the methods used for developing causal connection based on the conditions of the occurrence of an effect. In Chapter 2, we review some widely used models in analyzing and describing recurrent events such as renewal processes (RPs) and Poisson processes (PPs). PPs can be divided into two

general classes; (*i*) homogeneous Poisson processes (HPPs) and (*ii*) non-homogeneous Poisson processes (NHPPs). For the sake of simplicity in interpretation, we choose simple processes under two settings. In the first setting, we use a HPP to generate event times so that there is no overdispersion involved in the data generation, while in the second setting our model construction is based on the presence of overdispersion.

The major goal of our Monte Carlo simulation study is to determine the impact of different matching methods in the estimation of treatment effects. Therefore, as mentioned above, we consider three matching methods listed below:

1. *PSM*: In this matching method, we use seven different models to obtain propensity scores and match the subjects to balance the observed covariates between treated and untreated subjects.

2. *CM*: This is the most basic matching method, in which we try to find subjects with similar values on outcome-related covariates. Unlike the *PSM* method, in the *CM* method we match each of the pre-treatment measurements separately.

3. *HM*: This method is based on the rate of events observed in the past of individuals. In the *HM* method, we use the previous number of events experienced by each subject prior to the experimental treatment initiation to match treated and untreated subjects.

Each of the matching methods mentioned above has their own advantages and disadvantages. When there are a few covariates on which subjects need to be matched, *CM* is one of the most powerful matching techniques as it allows us to match the subjects on covariates directly so that we can find the best matched subjects. Methods based on *PSM* can be more practical compared to *CM* when there are too many covariates involved in the matching process. In such cases, it might be technically hard to use *CM* to match the subjects on each of the covariates separately because of the high dimensionality of the covariates. As a result, we may end up with too many treated subjects being excluded from the study. In contrast, by applying the *PSM* method, it is possible to summarize all covariates in a single value (i.e., the estimated propensity score) and use it for matching.

On the other hand, the *HM* method is simple to implement as it does not require the explanatory variables to be known. This would make the study much easier

since the researchers are no longer in need for identification of the key covariates that are used for matching subjects. The *HM* method could be powerful in cases, where the information provided by the history is sufficient in order to be able to match the subjects on their history. This condition may require subjects to experience enough number of observed events in a fixed follow-up period before the experimental treatment assignment. If this is not possible, the history data can be extended by additional information on subjects such as the addition of some explanatory variables at the baseline. We briefly discuss this issue later on this section. Note that the history information used in the matching process may vary in a sense that one may want to match the subjects on something other than the rate or the number of events observed in the past of subjects. For example, it is also possible to match the subjects based on the gap times between successive events experienced by subjects prior to the treatment assignment.

We now introduce the setup of our simulation study. In order to represent a general case, we consider different type of explanatory variables. More specifically, we use ten binary variables, a continuous variable and a count variable. The association of these explanatory variables with the outcome or treatment selection can be strong, medium or weak. We let $x_1, x_2, x_3, x_4, x_5, x_6, x_7, x_8, x_9, x_{10}, x_{11}, x_{12}$ represent the explanatory variables used in simulations. Table 3.1 presents their association with the treatment selection and outcome.

Table 3.1: Explanatory variables used in simulations.

	Associated with treatment	Not associated with treatment
Associated with outcome	$x_1, x_2, x_4, x_5, x_{11}$	x_3, x_6
Not associated with outcome	x_7, x_8, x_{10}, x_{12}	x_9

The nine variables $x_1, x_2, x_4, x_5, x_7, x_8, x_{10}, x_{11}, x_{12}$ in Table 3.1 are associated with the treatment selection and the variables $x_1, x_2, x_3, x_4, x_5, x_6, x_{11}$ are associated with the outcome. The variable x_9 is associated with neither treatment selection nor outcome. In an epidemiological terminology, the five variables $x_1, x_2, x_4, x_5, x_{11}$ are sometimes referred to as *true confounders*, which means that they are associated with both treatment selection and the outcome (Rothman et al., 2008). The other two covariates

x_3 and x_6 can be considered as *potential confounders* although they are theoretically not associated with the treatment selection. This is because for any given realization of a data set, there can be a small relation because of chance between the covariates and the treatment selection. If those covariates are also related to the outcome, then they are *empirical confounders* (or potential confounders) for that specific data set (Brookhart et al., 2006). In this chapter, we assume that the covariates x_3 and x_6 are unobserved. We use them in some of the models to be able to demonstrate the bias and improvements resulted from excluding and including them in the models.

As mentioned above, we use different levels of association between covariates and treatment selection, as well as between covariates and the outcome. The covariates x_1, x_4, x_7 and x_{10} are strongly associated with the treatment selection. The covariates x_2, x_5, x_8 and x_{12} are moderately associated with the treatment selection. The covariate x_{11} is weakly associated with the treatment selection. We also let the strength of association between outcome and the covariates vary so that we can have a good understanding of how well aforementioned matching methods balance the covariates between treated and untreated subjects. In our simulations, association between the covariates x_1, x_3, x_4 and outcome is strong, while covariates x_2, x_5 and x_6 are moderately associated with the outcome. Finally, the covariate x_{11} is weakly associated with the outcome. Table 3.2 gives the relations of covariates with the treatment assignment and outcome.

Table 3.2: The levels of association between the explanatory variables and outcome / treatment selection.

	Outcome	Treatment
Strong Association	x_1, x_3, x_4	x_1, x_4, x_7, x_{10}
Moderate Association	x_2, x_5, x_6	x_2, x_5, x_8, x_{12}
Weak Association	x_{11}	x_{11}

The outcome of interest could be the rate of the event occurrences if the follow-up times for individuals vary. Let $\underset{\sim}{x}$ denote the vector of selected covariates given in Table 3.1. For convenience, we consider $\{N_i(t);\ t \geq 0\}$ continuously observed over the interval $[0, \tau]$ for all $i = 1, 2, \ldots, m$. Note that we take $\tau_i = \tau$ for all $i = 1, 2, ..., m$ for the sake of simplicity in interpretation of the results. However, the results can be extended to the case, in which τ_i values vary as well. When $\tau_i = \tau$ for all individuals, we can equivalently focus on the expected number of events over the interval $[s, \tau]$

instead of focusing on the rate of event occurrences; that is,

$$E\{N_i(\tau)|\underset{\sim}{x}\} = \int_s^\tau \lambda_i(v|\underset{\sim}{x})dv, \quad i = 1, 2, ..., m, \tag{3.1}$$

for the model with no overdispersion, and

$$E\{N_i(\tau)|\underset{\sim}{x}, u_i\} = \int_s^\tau u_i\lambda_i(v|\underset{\sim}{x})dv, \quad i = 1, 2, ..., m, \tag{3.2}$$

for the random effects model, where u_i follows a gamma distribution with mean 0 and variance ϕ. The lower limits of the integrals s given in (3.1) and (3.2) represent the time of the experimental treatment initiation, which is equal to 5 years in our study. We represent the outcome of the matching methods in two forms; theoretical estimate (T.E.) and empirical estimate (E.E.) of the treatment effect. The theoretical estimate can be obtained by calculating

$$\frac{E\{N_1(\tau)|\underset{\sim}{x}\}}{E\{N_0(\tau)|\underset{\sim}{x}\}},$$

where $N_1(\tau)$ and $N_0(\tau)$ correspond to the treated and untreated matched subjects, respectively. Moreover, empirical estimate is the total number of post-treatment events for matched treated subject divided by the total number of post-treatment events for matched untreated subject.

We consider the following propensity score models, each differing in the choice of explanatory variables entering the model:

- PS 1: This model contains all variables associated with the treatment selection.
- PS 2: The model PS 2 contains all variables associated with the treatment selection as well as previous number of events experienced by each subject prior to the treatment selection.
- PS 3: This model includes all the true confounding variables that are associated with both the treatment selection and outcome.
- PS 4: This model includes all the true confounding variables and previous number of events experienced by each subject prior to the treatment selection.

- PS 5: In this model we obtain propensity scores using the true confounders with an additional adjustment for variable representing the history of the subjects.
- PS 6: All twelve variables are included in the propensity score model.
- PS 7: All observed and unobserved variables associated with outcome are included in the model.

As recommended by Cochran and Rubin (1973), in this study we apply caliper matching where calipers of width of 0.2 of the standard deviation of the propensity scores are used. For PS 5, control subjects that have the same number of pre-treatment events as treated subjects and share similar propensity scores are considered as matches. In other words, in addition to matching subjects on their propensity scores, we use exact matching of subjects based on their previous number of events. From a statistical point of view, the model PS 5 is analogous to blocking in a randomized study.

For *CM*, we consider the following four cases:

- CM 1: In this case, we match the subjects on the variables that are associated with the treatment selection.
- CM 2: All the variables associated with the treatment selection as well as previous number of events experienced by each subject prior to the treatment selection are considered for matching subjects.
- CM 3: In this case, we use the true confounders x_1, x_2, x_4, x_5 and x_{11} to match the subjects.
- CM 4: We match the subjects on the true confounders and previous number of events experienced by each of them prior to the treatment selection.

Note that in *CM* we do not match the subjects based on their propensity scores. Instead of this, we directly match them on binary covariates and history. In other words, we use exact matching for binary covariates and the number of events occurred before the treatment assignment. For continuous and count variables, we apply a caliper of width of 0.2 of the standard deviation of the corresponding covariate. Finally, for *HM*, we consider an untreated subject as a match for a treated subject if it has the same pre-treatment number of events as treated subject.

Table 3.3: Coefficients used to obtain propensity scores.

$\beta_{0,trt}$	β_1	β_2	β_3	β_4	β_5	β_6
-3.5	log(5)	log(2)	log(5)	log(2)	log(5)	log(2)

Table 3.4: Coefficients used to generate event times.

ρ_0	β_{trt}	α_1	α_2	α_3	α_4	α_5	α_6
0.3	-1.099	0.389	0.148	0.389	0.389	0.148	0.148

Following steps are used to generate non-overdispersed data throughout simulations discussed in this chapter.

- Step 1: We considered m (=500, 1000) independent subjects. For each of them, the binary covariates $x_1 - x_9$ and x_{12} were generated from independent Bernoulli distributions with parameters 0.5 and 0.92, respectively. Other two covariates x_{10} and x_{11}, which respectively represent the continuous and count variables, were generated from the standard Normal distribution and the negative binomial distribution $NB(r = 60,\ p = 0.56)$.

- Step 2: We then assigned each subject to the treatment or control group by using the binary model

$$Trt_i \sim Bernoulli(p_{i,trt}), \qquad i = 1, 2, ..., m, \tag{3.3}$$

  where the propensity score for each of the subjects can be obtained by using the logistic regression model

$$\begin{aligned} logit(p_{i,trt}) = \beta_{0,trt} + \beta_1 x_1 + \beta_2 x_2 + \beta_3 x_4 + \beta_4 x_5 + \beta_5 x_7 \\ + \beta_6 x_8 + \log(0.1) x_{10} + \log(1.03) x_{11} + \log(0.45) x_{12}. \end{aligned} \tag{3.4}$$

  The values of the parameters in the model (3.4) are given in Table 3.3.

- Step 3: Let $\{N_i(t);\ t \geq 0\}$ be a PP with the associated intensity function

$$\begin{aligned} \lambda_i \left(t \,|\, \underset{\sim}{x} \right) = \rho_0 \exp\{\beta_{trt} Trt_i + \alpha_1 x_1 + \alpha_2 x_2 + \alpha_3 x_3 + \alpha_4 x_4 \\ + \alpha_5 x_5 + \alpha_6 x_6 + \log(1.05) x_{11}\}, \end{aligned} \tag{3.5}$$

where ρ_0 indicates the baseline rate function and Trt_i is a binary variable defining whether the ith subject receives the treatment or not. We used the model given in (3.5) to generate event times for m individual processes over 10 years of follow-up. The procedures used to generate events in this setup are given in Section 2.4. In our case, none of the subjects received the experimental treatment during their first 5 years of follow-up period. The values of the parameters in the model (3.5) are given in Table 3.4.

- Step 4: Total number of events experienced by each subject during the first five years (i.e., during the pre-treatment period) is recorded. This information is used in the *HM* method.

- Step 5: We matched the subjects in the treatment group with subjects in the control group using the proposed methods and models as previously discussed in this section.

- Step 6: For the matched sample obtained from the previous step, we calculate the mean of the empirical and theoretical estimates of the treatment effect resulted from all matched subjects.

- Step 7: We repeat Steps 1 to 6 B (=1000) times. Finally, the Monte Carlo estimate of the treatment effect is obtained by averaging over the 1000 estimates resulted from simulated data sets.

We next give the steps used to generate data in the presence of overdispersion. Some of the steps below are the same as the ones given above, but for the sake of completeness we report them again.

- Step 1*: Like Step 1, we consider m (=500, 1000) independent subjects. We generate ten binary covariates $x_1 - x_9$ and x_{12} for each of m subjects. The nine covariates $x_1 - x_9$ are drawn from independent Bernoulli distributions, each with parameter 0.5. The other covariate x_{12} was drawn from a Bernoulli distribution with the value of the success probability parameter 0.92. The continuous and count covariates, x_{10} and x_{11}, were respectively generated from the standard Normal distribution and the negative binomial distribution $NB(r = 60,\ p = 0.56)$.

- Step 2*: We generated a treatment status for each of the m subjects by using the following binary model

$$Trt_i \sim Bernoulli(p_{i,trt}) \qquad i = 1, 2, ..., m, \tag{3.6}$$

where the propensity score model is defined as

$$\begin{aligned} logit(p_{i,trt}) = \beta_{0,trt} + \beta_1 x_1 + \beta_2 x_2 + \beta_3 x_4 + \beta_4 x_5 + \beta_5 x_7 \\ + \beta_6 x_8 + \log(0.1) x_{10} + \log(1.03) x_{11} + \log(0.45) x_{12}. \end{aligned} \tag{3.7}$$

- Step 3*: We then generated event times for each subject using the random effect Poisson model

$$\begin{aligned} \lambda_i \left(t \,|\, \underset{\sim}{x}, u_i \right) = u_i \rho_0 \exp\{\beta_{trt} Trt_i + \alpha_1 x_1 + \alpha_2 x_2 + \alpha_3 x_3 + \alpha_4 x_4 \\ + \alpha_5 x_5 + \alpha_6 x_6 + \log(1.05) x_{11}\}, \end{aligned} \tag{3.8}$$

where u_i follows a gamma distribution with mean 1 and variance $\phi(= 0.3$ and $0.6)$. Note that, Trt_i $(i = 1, ..., m)$ equals zero during the first 5 years of subject's follow-up period. Parameters used in formulas (3.7) and (3.8) are the same as those used in the formulas (3.4) and (3.5).

- Step 4*: We recorded the total number of events that each subject experienced prior to the time of experimental treatment initiation, and then used that for HM and improving the performance of other matching methods.

- Step 5*: We next used the aforementioned matching methods and models to match the subjects so that we can estimate the treatment effect.

- Step 6*: Using the matched sample obtained from previous step we calculate the mean of theoretical and empirical estimates of treatment effect.

- Step 7*: We repeat the Steps 1* to 6* $B(= 1000)$ times, each of size m and finaly the Monte Carlo estimate of the treatment effect is obtained by averaging over the 1000 estimates resulted from simulated data sets.

The estimates obtained from the Step 7 and Step 7* are compared to $\exp\{\beta_{trt}\} = 0.33$ where β_{trt} is the true treatment effect. Results of the Monte Carlo simulations

are reported in Tables 3.5 – 3.8. Table 3.6 and Table 3.8 represent the results of the matching methods when covariates x_3 and x_6 are strongly associated with outcome. In this case, we set $\alpha_3 = 0.9$ and $\alpha_6 = 0.55$ to see how well the proposed matching methods work. We next summarize the results of the simultion studies.

3.2 Monte Carlo Simulations: Summary and Results

Our primary goal of using *PSM*, *HM* and *CM* is to balance all outcome-related covariates involved in the process to obtain an accurate treatment effect measure. It is important to note that in a real life situation there may exist some unobserved outcome related covariates not measured due to the lack of enough understanding of the process. Unlike randomization, *PSM* methods do not gaurantee the balance of unmeasured covariates (Rubin and Thomas, 2000). As a result, a bias in the estimate of the treatment effect may occur. Tables 3.5 – 3.8 can help to indicate the accuracy of the suggested matching methods in estimating a treatment effect under various settings.

We summarize our findings as follows. First, we found that *CM* and *HM* resulted in the least biased estimators of the treatment effect, while matching on propensity scores resulted in a more pronounced degree of bias. For example, in Table 3.5 the empirical estimate of PS 4 in the absence of overdispersion is 18 per cent more biased comparing to the result of CM 4. In particular, the propensity score model PS 1 including all covariates associated with the treatment assignment resulted in the greatest biased results. Whereas, including only the confounders x_1, x_2, x_4, x_5 and x_{11} in the propensity score model PS 3 resulted in a greater percision in the estimation of the treatment effect. This result supports the fact that the goal of propensity score methods is to efficiently balance the outcome related covariates between treated and untreated subjects, not to predict the probability of receiving the treatment (Brookhart et al., 2006). The results of our simulation study reveal that if variables unrelated to the outcome but related to the exposure are added to the propensity score model, the bias might be more pronounced as a result of not well-balanced matches or decreased number of matched subjects. This statement can be supported by the estimates resulted from PS 1 and PS 6. For example, in Table 3.5,

Table 3.5: Theoretical estimates (T.E.'s) and empirical estimates (E.E.'s) of the treatment effect resulted from the matching methods ($m = 1000$).

	Without overdispersion ($\phi = 0$)		With overdispersion ($\phi = 0.3$)		With overdispersion ($\phi = 0.6$)	
	T.E.	E.E.	T.E.	E.E.	T.E.	E.E.
PS 1	0.4631	0.5027	0.6737	0.7489	1.2025	1.0490
PS 2	0.4678	0.5073	0.6641	0.7396	1.1682	1.0339
PS 3	0.3824	0.4096	0.5442	0.6042	0.9463	0.8566
PS 4	0.4015	0.4307	0.5639	0.6285	0.9673	0.8717
PS 5	0.3466	0.3712	0.3577	0.3956	0.3811	0.4079
PS 6	0.4697	0.5099	0.6664	0.7423	1.0955	1.0232
PS 7	0.4014	0.4306	0.5733	0.6380	1.0342	0.9087
CM 1	0.3489	0.3701	0.4954	0.5382	0.8606	0.8123
CM 2	0.3452	0.3615	0.3551	0.3969	0.3962	0.4511
CM 3	0.3483	0.3700	0.4967	0.5484	0.8982	0.8007
CM 4	0.3421	0.3697	0.3583	0.4009	0.3882	0.4169
HM	0.3568	0.3814	0.3624	0.3970	0.3818	0.4029

Table 3.6: Theoretical estimates (T.E.'s) and empirical estimates (E.E.'s) of the treatment effect resulted from the matching methods when unobserved covariates are strongly associated with outcome ($m = 1000$).

	Without overdispersion ($\phi = 0$)		With overdispersion ($\phi = 0.3$)		With overdispersion ($\phi = 0.6$)	
	T.E.	E.E.	T.E.	E.E.	T.E.	E.E.
PS 1	0.5872	0.6287	0.8652	0.9645	1.4029	1.3209
PS 2	0.5860	0.6279	0.8476	0.9462	1.4804	1.3227
PS 3	0.4800	0.5088	0.6854	0.7498	1.2007	1.0983
PS 4	0.4970	0.5290	0.7027	0.7683	1.2166	1.1169
PS 5	0.3461	0.3659	0.3526	0.3819	0.3697	0.3959
PS 6	0.5861	0.6315	0.8336	0.9313	1.5069	1.2974
PS 7	0.4941	0.5245	0.7039	0.7757	1.2519	1.1341
CM 1	0.4371	0.4622	0.6266	0.6931	1.1074	0.9999
CM 2	0.3512	0.3733	0.3613	0.3730	0.3522	0.3742
CM 3	0.4367	0.4604	0.6260	0.6826	1.0887	1.0143
CM 4	0.3454	0.3702	0.3539	0.3836	0.3753	0.4035
HM	0.3495	0.3671	0.3551	0.3789	0.3681	0.3886

Table 3.7: Theoretical estimates (T.E.'s) and empirical estimates (E.E.'s) of the treatment effect resulted from the matching methods ($m = 500$).

	Without overdispersion ($\phi = 0$)		With overdispersion ($\phi = 0.3$)		With overdispersion ($\phi = 0.6$)	
	T.E.	E.E.	T.E.	E.E.	T.E.	E.E.
PS 1	0.4726	0.5126	0.6869	0.7604	1.2144	1.0625
PS 2	0.4794	0.5289	0.6775	0.7406	1.2117	1.0515
PS 3	0.3875	0.4143	0.5490	0.6070	0.9565	0.8775
PS 4	0.4056	0.4343	0.5717	0.6311	1.0884	0.8989
PS 5	0.3476	0.3715	0.3572	0.3946	0.3792	0.4099
PS 6	0.4738	0.5128	0.6728	0.7534	1.1684	1.0475
PS 7	0.4094	0.4386	0.5842	0.6468	1.0306	0.9165
CM 1	0.3482	0.3690	0.4989	0.5449	0.8077	0.7984
CM 2	0.3359	0.3659	0.3418	0.3858	0.3598	0.3761
CM 3	0.3487	0.3688	0.5011	0.5516	0.8523	0.7971
CM 4	0.3423	0.3683	0.3583	0.3958	0.3809	0.4103
HM	0.3573	0.3815	0.3618	0.3954	0.3778	0.4026

Table 3.8: Theoretical estimates (T.E.'s) and empirical estimates (E.E.'s) of the treatment effect resulted from the matching methods when unobserved covariates are strongly associated with outcome ($m = 500$).

	Without overdispersion ($\phi = 0$)		With overdispersion ($\phi = 0.3$)		With overdispersion ($\phi = 0.6$)	
	T.E.	E.E.	T.E.	E.E.	T.E.	E.E.
PS 1	0.6024	0.6415	0.8376	0.9379	1.6833	1.3271
PS 2	0.5922	0.6294	0.8559	0.9529	1.4931	1.3606
PS 3	0.4890	0.5163	0.6930	0.7554	1.2142	1.1127
PS 4	0.5004	0.5301	0.7009	0.7636	1.2145	1.1115
PS 5	0.3460	0.3650	0.3539	0.3833	0.3691	0.3977
PS 6	0.5921	0.6463	0.8642	0.9545	1.4659	1.3363
PS 7	0.5036	0.5349	0.7157	0.7815	1.2128	1.1473
CM 1	0.4335	0.4528	0.6208	0.6814	1.0679	1.0657
CM 2	0.3342	0.3543	0.3573	0.3967	0.3602	0.4131
CM 3	0.4385	0.4595	0.6256	0.6783	1.1062	1.0187
CM 4	0.3435	0.3646	0.3531	0.3807	0.3691	0.4039
HM	0.3496	0.3679	0.3556	0.3811	0.3686	0.3898

the theoretical estimate resulted from the model PS 1 is equal to 0.4631 which is more biased in comparison to 0.3824, the theoretical estimate resulted from PS 3.

Another finding of our simulation study is that matching on the history of the subjects or just adding the history to the matching model significantly reduces the bias in the estimates. For example, in Table 3.5 the estimates resulted from the model PS 5 and HM supports this idea as they are close to the true value 0.33. This is because of the fact that the history of a subject is a direct result of the observed and unobserved covariates. Therefore, using it to match the subjects not only increases the precision but also accounts for unobserved covariates in some cases. Moreover, it is worth mentioning that, based on our findings in Table 3.5 and Table 3.6, the HM method is quite robust to some changes that make the estimation of treatment effect complicated. In particular, it can be seen in Table 3.5 that, when we incorporate overdispersion in the model, the results of HM did not deviate noticeably from the target value 0.33. Similarly, in Table 3.6 when we increase the effects of unobserved covariates x_3 and x_6, the resulted estimates are more precise comparing to the models that do not include the history in the matching process.

Another interesting result is that propensity score matching using the models PS 2 and PS 4 has not increased the precision in the estimation of the treatment effect, even though we have included the history of the process in the model. The reason for this result is that propensity score matching does not perform exact matching on the history of the process and just balance it on the average so that we may end up with matches that have different history.

Tables 3.5 and 3.6 show that propensity score matching using the model PS 7 does not improve the results as expected. This model includes all the observed and unobserved outcome related covariates. Although after matching the covariates are balanced on average, the results are still biased. We conducted a Monte Carlo simulation study to specify the root cause of this result, and found that balancing some key covariates on average is not a good idea since it has a profound impact on the event rate. In our study, the covariate x_{11} can make a noticeable difference if matched subjects differ on this covariate. We recommend that researchers should thoughtfully identify the pivotal covariates and make the necessary adjustments before conducting the PSM.

Tables 3.7 and 3.8 represent the results of the simulation studies when the population size is reduced to 500. In this case, we realized that reducing the population size increases the amount of the bias in the estimate of treatment effect arised from using the *PSM* comparing to the bias in estimates when the the population size is 1000. For example, in Table 3.8, the empirical estimate of the model PS 1 in the absence of overdispersion is equal to 0.6415 which is 3.9 per cent more biased compared to the corresponding result given in Table 3.6. Furthermore, we observe that reducing sample size does not greatly affect the results obtained by *CM* and *HM*. This result is expected because in our settings *CM* and *HM* result in good matched samples where matched units share similar covariates. As a result, in this case if the size of the matched samples decreases, the estimates are not affected.

There are many recommendations made for researchers who use propensity score methods to make causal inference in observational studies. One of the key points in using propensity score methods is to include all important true and potential confounding variables in a propensity score model. Any failure in doing so may result in the excluded variables being imbalanced between treated and untreated subjects, which eventually may lead to biased estimation of the treatment effect (Austin et al., 2007). In a real life situation it is usually common to have unobserved or unobservable covariates. In such cases, efficient and capable approaches to address this issue are needed. In the case where subjects under study provide a relatively good history, we suggest that this information should be included in the model as it significantly reduces the bias in the estimation of the treatment effects.

Another interesting conclusion that can be made from the simulation study is that if the data is overdispersed, regular propensity score matching will result in baised estimates. Overdispersion has been a challenging problem for researchers since it is hard to pinpoint the real reason why the data are overdispersed. Our findings show that the more the overdispersion, the worse the estimates. Based on the results given in Tables 3.5 - 3.8 it can be concluded that this issue can be addressed by using HM or PSM methods with some adjustment on the history of the possible matched subjects such as the propensity score model PS 5.

Chapter 4

Estimation of Time-Varying Treatment Effects

In Chapter 3, we discussed a setting in which subjects receive a time-fixed treatment at a certain time during their follow-up times. In a real life situation subjects may change their treatments while they are under follow-up. This situation usually takes place when the first treatment is not desirable or affordable or maybe another treatment has become available in the hospital. As a result, patients may switch to a new treatment. In this chapter, we consider two different scenarios. In the first scenario we assume that a standard treatment is available for all the individuals and they can either choose to receive that treatment or not. After a while, a new treatment becomes available for those individuals who received the standard treatment, and hence some of them may change the treatment. In such a case, the estimation of the new treatment effect and whether it is more effective comparing to the standard treatment become undoubtedly important. In the following section, we try to develop and describe the models and methods required for evaluating the treatment effects in such situations. In the second scenario, we discuss the situation where there are two different treatments available for individuals and at most one of them can be selected and received sometime during their follow-up.

4.1 Models and Methods

In this section, we introduce the models and methods used in the simulation study. Our primary goal is to examine the effectiveness of three different matching methods in various settings. The compared matching methods are propensity score matching (PSM), covariate matching (CM) and history matching (HM). We consider a data-generating process in the presence and absence of overdispersion. In the absence of overdispersion, we generate event times for individuals from a homogeneous Poisson process (HPP). We use a mixed Poisson model to generate event times for individuals when overdispersion is present.

We now introduce the setup of our simulation study. Similar to the setup given in the previous chapter, we consider twelve explanatory variables among which ten are binary variables, one is a continuous variable and another one is a count variable. We let x_1, x_2, x_3, x_4, x_5, x_6, x_7, x_8, x_9, x_{10}, x_{11}, x_{12} denote the values of the explanatory variables. The variables x_1, x_2, x_3, x_4, x_5, x_6, x_7, x_8, x_9, x_{12} are binary, the variables x_{10} and x_{11} are continuous and count, respectively. We choose to assign different levels of association with the outcome and treatment to these variables so that we can argue the effectiveness of different matching methods in a more general case. A strong association between a variable and either outcome or treatment selection causes a profound impact on the rate of the event occurrences or on the likelihood of selecting the treatment if that variable is present. We consider the strength of association between a variable and either outcome or treatment selection as moderate if the presence of that variable is not as impactful as a covariate with strong association, but helps to predict the probability of treatment selection or increase the event rate to a reasonable degree. Finally, the association between a variable and either outcome or treatment selection is defined as weak if the presence of that variable does not help in predicting the dependent variable or has a slight effect on the event rate which is ignorable in many situations. The presence or absence of association of the explanatory variables with the outcome or treatment selection are presented in Table 4.1.

Table 4.1: Explanatory variables used in simulations.

	Associated with treatment	Not associated with treatment
Associated with outcome	$x_1, x_2, x_4, x_5, x_{11}$	x_3, x_6
Not associated with outcome	x_7, x_8, x_{10}, x_{12}	x_9

As shown in Table 4.1, the variables x_1, x_2, x_4, x_5, x_7, x_8, x_{10}, x_{11}, x_{12} are associated with the treatment selection and the variables x_1, x_2, x_3, x_4, x_5, x_6, x_{11} are associated with the outcome. The variables x_1, x_2, x_4, x_5, x_{11} are called true confounders as they are associated with both the treatment selection and the outcome. In the current study, the two variables x_3 and x_6 are considered as potential confounders. We assume that these two variables are unobserved, but we include them in some of the *PSM* models to indicate the degree of bias resulted from excluding them in other models.

Table 4.2: The level of association of the explanatory variables.

	Outcome Variable	Treatment Selection
Strong Association	x_1, x_3, x_4	x_1, x_4, x_7, x_{10}
Moderate Association	x_2, x_5, x_6	x_2, x_5, x_8, x_{12}
Weak Association	x_{11}	x_{11}

The levels of association of the explanatory variables are described in Table 4.2. The covariates x_1, x_4, x_7 and x_{10} are strongly associated with the treatment selection. The covariates x_2, x_5, x_8 and x_{12} are moderately associated with the treatment selection. The covariate x_{11} is weakly associated with the treatment selection. We also consider different levels of association between the outcome and covariates. This allows us to understand the strength of different matching methods in balancing different type of covariates with different levels of association. In our simulation study, the association between the covariates x_1, x_3, x_4 and the outcome variable is strong, while the covariates x_2, x_5 and x_6 are moderately associated with the outcome variable. Finally, the covariate x_{11} is weakly associated with the outcome variable. We next briefly explain the first scenario of our simulation study.

Scenario 1: In the first scenario, subjects can change their treatment during their follow-up times. We assume that the standard treatment is available at the beginning

of the follow-up time. Individuals either choose to receive the standard treatment or not. Selection of the new treatment depends on whether the individual has received the standard treatment and on the availability of the new treatment before the end of the follow-up time. We assume a Bernoulli distribution with the success parameter 0.3 in order to indicate if the new treatment is available for an individual, and then generate the time of the new treatment initiation using a Weibull distribution with the shape parameter 2.3 and scale parameter 5.5. The parameters of Weibull distribution are chosen so that we have a reasonable follow-up time before and after the new treatment initiation. We generate the event times for individuals who receive the both treatments (treated group) as well for those who do not receive any treatment (control group), and then use the aforementioned matching methods to estimate and compare the effects of the standard and new treatments.

Figure 4.1: Event history of a matched pair.

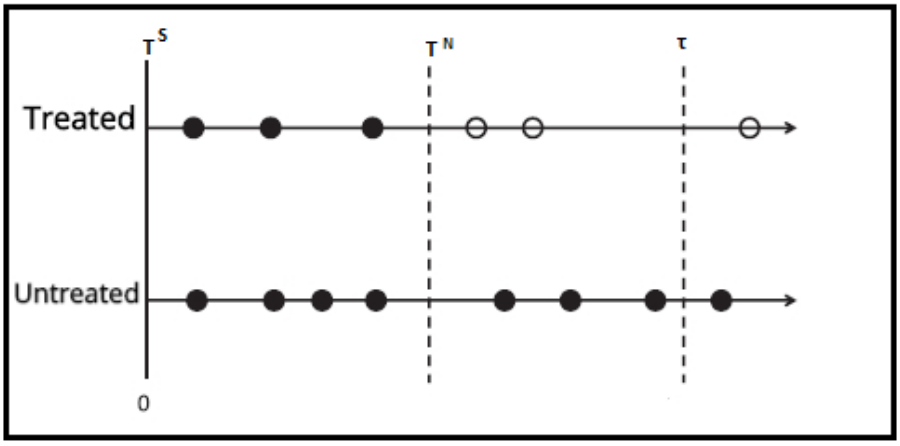

Figure 4.1 shows the event histories of two matched individuals. In the top line, the individual receives the standard treatment at time T^S (time of the standard treatment initiation), and then switch to the new treatment at time T^N (time of the new treatment initiation). In the bottom line, the individual does not receive any treatment.

In order to evaluate the effectiveness of matching methods in estimating the efficacy of the new treatment over the standard treatment, we present the outcome in two different forms, theoretical and empirical estimates of the standard treatment effect compared to the new treatment effect. Let $E(N_1(t^N))$ and $E(N_0(t^N))$ denote the expected number of events for a matched pair treated and control individuals over

the time interval $[0, T^N)$, respectively. Moreover, let $E(N_1(t^N, \tau))$ and $E(N_0(t^N, \tau))$ denote the expected number of events for the same matched treated and control individuals over the time interval $[T^N, \tau]$, respectively. Then, the theoretical estimate $(T.E.)$ of the matched sample can be obtained by calculating

$$T.E. = \frac{E(N_1(t^N))/E(N_0(t^N))}{E(N_1(t^N, \tau))/E(N_0(t^N, \tau))} \tag{4.1}$$

for all the matched individuals in the matched sample and taking their average.

Empirical estimate $(E.E.)$ can be obtained by following a similar idea. Let $EM_1(t^N)$ and $EM_0(t^N)$ denote the observed number of events for a matched pair treated and control individuals over the time interval $[0, T^N)$, respectively. Furthermore, let $EM_1(t^N, \tau)$ and $EM_0(t^N, \tau)$ denote the observed number of events for the same matched treated and control individuals over the time interval $[T^N, \tau]$, respectively. Then, we can estimate the effect of the standard treatment by calculating

$$\frac{EM_1(t^N)}{EM_0(t^N)}$$

for all matched individuals in a matched sample and taking their average. The effect of the new treatment can be estimated by calculating

$$\frac{EM_1(t^N, \tau)}{EM_0(t^N, \tau)}$$

for all the matched individuals and taking their average. Finally, the empirical estimate of the standard treatment effect compared to the new treatment effect for the matched sample can be obtained by dividing the resulted estimate of standard treatment by resulted estimate of new treatment. We next explain the second scenario for our simulation study where there are two different treatments available for subjects to receive.

Scenario 2: In this scenario, we assume that there exist two treatments; Treatment A and Treatment B. We assign individuals to either a treatment or a control group using a binary distribution. The treatment group consists of individuals who

receive Treatment A, and the control group consists of individuals who receive Treatment B. The time to receive Treatment A or B is generated from a Weibull distribution with the shape parameter 2.1 and scale parameter 3.2. We generate the event times for those who receive Treatment A as well as for those who receive Treatment B. Then, we use the *PSM*, *CM* and *HM* methods to estimate and compare the effects of Treatments A and B.

Let $t^{trt.A}$ and $t^{trt.B}$ denote the times of receiving Treatment A and Treatment B, respectively. We use $E(N_A(t^{trt.A}, \tau))$ and $E(N_B(t^{trt.B}, \tau))$ to respectively denote the expected number of events for a matched pair treated and control individuals over the time intervals $[t^{trt.A}, \tau)$ and $[t^{trt.B}, \tau)$. Then, the theoretical estimate of the effect of Treatment A compared to the effect of Treatment B can be shown as

$$T.E. = \frac{E(N_A(t^{trt.A}, \tau))/(\tau - t^{trt.A})}{E(N_B(t^{trt.B}, \tau))/(\tau - t^{trt.B})}. \tag{4.2}$$

In order to calculate the empirical estimate, the expected number of events in the formula (4.2) is replaced by the corresponding observed number of events.

We propose the following *PSM* and *CM* models, each differing in the choice of variables entering the model:

- PSM 1: This model contains all variables associated with the treatment selection.
- PSM 2: This model contains all variables associated with the treatment selection as well as rate of events experienced by each subject prior to the treatment selection.
- PSM 3: This model includes all the true confounding variables that are associated with both the treatment selection and outcome.
- PSM 4: This model includes all the true confounding variables and rate of events experienced by each subject prior to the treatment selection.
- PSM 5: In this model, we obtain propensity scores using the true confounders with an additional adjustment for variable representing the history of the subjects.

- PSM 6: All twelve variables are included in the propensity score model.
- PSM 7: All observed and unobserved variables associated with the outcome are included in the model.

As recommended by Cochran and Rubin (1973), in all simulation studies in this chapter, we apply caliper matching, where calipers of width of 0.2 of the standard deviation of propensity scores are used. For PSM 5, we simultaneously match the individuals on their propensity scores and histories. In this model, we adjust the histories of the potential matched individuals so that the difference between their pre-treatment event rates falls within 0.2 standard deviation of the history variable.

For *CM*, we consider the following four cases:

- CM 1: In this case, we match the subjects on the variables that are associated with the treatment selection.
- CM 2: All the variables associated with the treatment selection as well as rate of events experienced by each subject prior to the treatment selection are considered for matching subjects.
- CM 3: In this case, we use the true confounders x_1, x_2, x_4, x_5 and x_{11} to match the subjects.
- CM 4: We match the subjects on the true confounders and rate of events experienced by each of them prior to the treatment selection.

In *CM* method, we use exact matching for the binary covariates x_1, x_2, x_3, x_4, x_5, x_6, x_7, x_8, x_9, x_{12}. For continuous and count variables, x_{10} and x_{11}, respectively, we apply a caliper of width of 0.2 of the standard deviation of the associated covariate. In order to match the subjects on their histories, we apply a caliper of width of 0.2 of the standard deviation of the variable representing the history of the subjects under study. The reason why we use the rate of pre-treatment events rather than the previous number of events experienced by subjects is that the pre-treatment follow-up times for subjects vary.

Finally, for *HM*, we consider a control subject as a match for a treated subject if the absolute difference between their pre-treatment event rates is equal to or within 0.2 standard deviation of the variable representing the history.

We now present the steps used for the data-generating process in the absence of overdispersion for the first scenario.

Table 4.3: Coefficients used to obtain the propensity scores.

$\beta_{0,treatment}$	β_1	β_2	β_3	β_4	β_5	β_6
-3.5	log(5)	log(2)	log(5)	log(2)	log(5)	log(2)

Table 4.4: Coefficients used to generate event times.

ρ_0	β_S	β_N	β_A	β_B	α_1	α_2	α_3	α_4	α_5	α_6
0.3	-1.099	-1.15	-1.099	-1.25	0.389	0.148	0.389	0.389	0.148	0.148

- Step 1: We considered $m(= 500$ and $1000)$ independent subjects. For each of them, the binary covariates $x_1 - x_9$ and x_{12} were generated from independent Bernoulli distributions with parameters 0.5 and 0.92, respectively. Remaining two covariates x_{10} and x_{11}, which respectively represent the continuous and count variables, were generated from the standard Normal distribution and the negative binomial distribution $NB(r = 60,\ p = 0.56)$.

- Step 2: We then assigned the standard treatment to subjects by using the binary model

$$Trt_{i,S} \sim Bernoulli(p_{i,treatment}), \qquad i = 1, 2, ..., m, \tag{4.3}$$

where the propensity score for each of the subjects is obtained by the logistic regression model

$$\begin{aligned} logit(p_{i,treatment}) = \beta_{0,treatment} + \beta_1 x_1 + \beta_2 x_2 + \beta_3 x_4 + \beta_4 x_5 + \beta_5 x_7 \\ + \beta_6 x_8 + \log(0.1) x_{10} + \log(1.03) x_{11} + \log(0.45) x_{12}. \end{aligned} \tag{4.4}$$

The values of the parameters in the model (4.4) are given in Table 4.3.

- Step 3: We assigned the new treatment to those subjects who received the standard treatment by using a binary outcome distribution

$$Trt_{i,N} \sim Bernoulli(0.3), \qquad i = 1, 2, ..., m. \tag{4.5}$$

Note that, if $Trt_{i,S} = 0$ then $Trt_{i,N} = 0$.

- Step 4: The time of new treatment initiation was generated from a Weibull distribution with the shape parameter 2.3 and scale parameter 5.5.

- Step 5: The event times over the time intervals $[0, T_i^N)$ and $[T_i^N, \tau]$ for those individuals who received both the standard treatment and the new treatment were generated by using the following intensity functions:

$$\begin{aligned}\lambda_i\left(t \,|\, \underset{\sim}{x}\right) = \rho_0 \exp\{&\beta_S Trt_{i,S} + \alpha_1 x_1 + \alpha_2 x_2 + \alpha_3 x_3 + \alpha_4 x_4 \\ &+ \alpha_5 x_5 + \alpha_6 x_6 + \log(1.05) x_{11}\},\end{aligned} \tag{4.6}$$

  and

$$\begin{aligned}\lambda_i\left(t \,|\, \underset{\sim}{x}\right) = \rho_0 \exp\{&\beta_N Trt_{i,N} + \alpha_1 x_1 + \alpha_2 x_2 + \alpha_3 x_3 + \alpha_4 x_4 \\ &+ \alpha_5 x_5 + \alpha_6 x_6 + \log(1.05) x_{11}\},\end{aligned} \tag{4.7}$$

  respectively.

- Step 6: The event times over the time interval $[0, \tau]$ for those individuals who did not receive any treatment were generated using the following intensity function.

$$\begin{aligned}\lambda_i\left(t \,|\, \underset{\sim}{x}\right) = \rho_0 \exp\{&\alpha_1 x_1 + \alpha_2 x_2 + \alpha_3 x_3 + \alpha_4 x_4 \\ &+ \alpha_5 x_5 + \alpha_6 x_6 + \log(1.05) x_{11}\},\end{aligned} \tag{4.8}$$

  where ρ_0 in the formulas (4.6), (4.7) and (4.8) indicates the baseline rate function. We set $\tau = 10$ for all the individuals under study. The procedures used to generate events in this setup are given in Section 2.4. The values of the parameters in the models (4.6), (4.7) and (4.8) are given in Table 4.4.

- Step 7: We matched the subjects in the treatment group with subjects in the control group using some of *PSM* and *CM* models.

- Step 8: For the matched sample obtained from the previous step, we calcualte theoretical estimate (4.1) and empirical estimate.

- Step 9: We repeat Steps 1 to 8 $B(= 1000)$ times. Finaly the Monte Carlo estimate of the compared treatment effects is obtained by averaging over the 1000 means resulted from simulated data sets.

The steps used to generate data in the presence of overdispersion in the first scenario are given below.

- Step 1: We considered $m(= 500$ and $1000)$ independent subjects. The binary covariates $x_1 - x_9$ and x_{12} were generated for each of the subjects from independent Bernoulli distributions with parameters 0.5 and 0.92, respectively. Remaining two covariates x_{10} and x_{11}, which respectively represent the continuous and count variables, were generated from the standard Normal distribution and the negative binomial distribution, that is $NB(r = 60,\ p = 0.56)$.

- Step 2: We then assigned the standard treatment to subjects by using the binary model

$$Trt_{i,S} \sim Bernoulli(p_{i,treatment}), \qquad i = 1, 2, ..., m, \tag{4.9}$$

  where the propensity score for each of the subjects can be obtained by

$$\begin{aligned} logit(p_{i,treatment}) = \beta_{0,treatment} + \beta_1 x_1 + \beta_2 x_2 + \beta_3 x_4 + \beta_4 x_5 + \beta_5 x_7 \\ + \beta_6 x_8 + \log(0.1) x_{10} + \log(1.03) x_{11} + \log(0.45) x_{12}. \end{aligned} \tag{4.10}$$

- Step 3: The new treatment was then assigned to those subjects who received the standard treatment by using a bianry outcome distribution

$$Trt_{i,N} \sim Bernoulli(0.3), \qquad i = 1, 2, ..., m. \tag{4.11}$$

  Note that, if $Trt_{i,S} = 0$ then $Trt_{i,N} = 0$.

- Step 4: The time of new treatment initiation was generated from a Weibull distribution with the shape parameter 2.3 and scale parameter 5.5.

- Step 5: The event times over the time intervals $[0, T_i^N)$ and $[T_i^N, \tau]$ for those individuals who received both the standard and new treatments were generated by using the following intensity functions:

$$\begin{aligned} \lambda_i \left(t \,|\, \underset{\sim}{x}, u_i \right) = u_i \rho_0 \exp\{\beta_S Trt_{i,S} + \alpha_1 x_1 + \alpha_2 x_2 + \alpha_3 x_3 + \alpha_4 x_4 \\ + \alpha_5 x_5 + \alpha_6 x_6 + \log(1.05) x_{11}\}, \end{aligned} \tag{4.12}$$

and

$$\lambda_i\left(t \mid \underset{\sim}{x}, u_i\right) = u_i\rho_0 \exp\{\beta_N Trt_{i,N} + \alpha_1 x_1 + \alpha_2 x_2 + \alpha_3 x_3 + \alpha_4 x_4 + \alpha_5 x_5 + \alpha_6 x_6 + \log(1.05)x_{11}\}, \tag{4.13}$$

respectively.

- Step 6: The event times over the time interval $[0, \tau]$ for those individuals who did not receive any treatment were generated using the following intensity function.

$$\lambda_i\left(t \mid \underset{\sim}{x}, u_i\right) = u_i\rho_0 \exp\{\alpha_1 x_1 + \alpha_2 x_2 + \alpha_3 x_3 + \alpha_4 x_4 + \alpha_5 x_5 + \alpha_6 x_6 + \log(1.05)x_{11}\}, \tag{4.14}$$

  where u_i in (4.12), (4.13) and (4.14) follows a gamma distribution with mean 1 and variance $\phi(= 0.3$ and $0.6)$. Note that $\tau = 10$ in all the subjects under study.

- Step 7: We matched the subjects in the treatment group with subjects in the control group using some of *PSM* and *CM* models.

- Step 8: For the matched sample obtained from the previous step, we calculate theoretical estimate (4.1) and empirical estimate.

- Step 9: We repeat Steps 1 to 8 $B(= 1000)$ times. Finaly the Monte Carlo estimate of the compared treatment effects is obtained by averaging over the 1000 means resulted from simulated data sets.

It should be noted that, in the first scenario of our simulation study, we cannot use the *HM* method or any other models that include pre-treatment event rates due to the lack of available history of individuals. We now give the steps required for the data-generating process in the absence of overdispersion for the second scenario.

- Step 1*: Like Step 1, we consider $m(= 500$ and $1000)$ independent subjects. We generate ten binary covariates $x_1 - x_9$ and x_{12} for each of m subjects. The nine covariates $x_1 - x_9$ are drawn from independent Bernoulli distributions, each with parameter 0.5. The other covariate x_{12} was drawn from a Bernoulli distribution with the value of the success probability parameter 0.92. The continuous

and count covariates, x_{10} and x_{11}, were respectively generated from the standard Normal distribution and the negative binomial distribution, with notation $NB(r = 60,\ p = 0.56)$.

- Step 2*: We generated a treatment status for each of the m subjects by using the following binary model

$$Trt_i \sim Bernoulli(p_{i,treatment}) \qquad i = 1, 2, ..., m, \tag{4.15}$$

where the propensity score model is defined as

$$\begin{aligned} logit(p_{i,treatment}) = \beta_{0,treatment} + \beta_1 x_1 + \beta_2 x_2 + \beta_3 x_4 + \beta_4 x_5 + \beta_5 x_7 \\ + \beta_6 x_8 + \log(0.1) x_{10} + \log(1.03) x_{11} + \log(0.45) x_{12}. \end{aligned} \tag{4.16}$$

Note that, here if $Trt_i = 1$ then the subject receives Treatment A; otherwise, Treatment B. The values of parameters in formula (4.13) is given in Table 4.3.

- Step 3*: We assigned subjects that received treatment A to treatment group, and those that received treatment B to control group.

- Step 4*: The time of treatment assignment was generated from a Weibull distribution with the shape parameter 2.1 and scale parameter 3.2.

- Step 5*: We generated pre-treatment event times for all subjects using the following intensity function:

$$\lambda_i \left(t \,|\, \underset{\sim}{x} \right) = \rho_0 \exp\{\alpha_1 x_1 + \alpha_2 x_2 + \alpha_3 x_3 + \alpha_4 x_4 + \alpha_5 x_5 + \alpha_6 x_6 + \log(1.05) x_{11}\}. \tag{4.17}$$

- Step 6*: We then generated post-treatment event times for subjects in treatment and control groups using the following intensity function.

$$\begin{aligned} \lambda_i \left(t \,|\, \underset{\sim}{x} \right) = \rho_0 \exp\{\beta_A Trt_{i,A} + \beta_B Trt_{i,B} + \alpha_1 x_1 + \alpha_2 x_2 + \alpha_3 x_3 + \alpha_4 x_4 \\ + \alpha_5 x_5 + \alpha_6 x_6 + \log(1.05) x_{11}\}. \end{aligned} \tag{4.18}$$

Note that, at most one of $Trt_{i,A}$ and $Trt_{i,B}$ can be equal to one. The values of parameters in formulas (4.14) and (4.15) are given in Table 4.4.

- Step 7*: We recorded the rate of the events that each subject experienced prior to the time of experimental treatments initiation, and then used that for *HM*, *CM* and *PSM*.

- Step 8*: We next used the proposed matching methods and models to match the subjects so that we can estimate the treatment effects.

- Step 9*: Using the matched sample obtained from previous step we calculate theoretical estimate (4.2) and empirical estimate for each of the matched individuals in a matched sample and then calculate the mean of the resulted estimates.

- Step 10*: We repeat the Steps 1* to 9* $B(= 1000)$ times, each of size m and finaly the Monte Carlo estimate of the compared treatment effects is obtained by averaging over the 1000 means resulted from simulated data sets.

In the case of overdispersed data, the steps of the data-generating process are given below.

- Step 1*: We consider m (=500 and 1000) independent subjects and generate ten binary covariates x_1 – x_9 and x_{12} for each of m subjects. The nine covariates x_1 – x_9 are drawn from independent Bernoulli distributions, each with parameter 0.5. The other covariate x_{12} was drawn from a Bernoulli distribution with the value of the success probability parameter 0.92. The continuous and count covariates, x_{10} and x_{11}, were respectively generated from the standard Normal distribution and the negative binomial distribution $NB(r = 60,\ p = 0.56)$.

- Step 2*: A treatment status was generated for each of the m subjects by using the following binary model

$$Trt_i \sim Bernoulli(p_{i,treatment}) \qquad i = 1, 2, ..., m, \tag{4.19}$$

where the propensity score model is defined as

$$\begin{aligned} logit(p_{i,treatment}) = \beta_{0,treatment} + \beta_1 x_1 + \beta_2 x_2 + \beta_3 x_4 + \beta_4 x_5 + \beta_5 x_7 \\ + \beta_6 x_8 + \log(0.1) x_{10} + \log(1.03) x_{11} + \log(0.45) x_{12}. \end{aligned} \tag{4.20}$$

Note that, here if $Trt_i = 1$ then subject receive treatment A; otherwise treatment B.

- Step 3*: We assigned subjects that received treatment A to treatment group, and those that received treatment B to control group.

- Step 4*: The time of treatment assignment was generated from a Weibull distribution with the shape parameter 2.1 and the scale parameter 3.2.

- Step 5*: We generated pre-treatment event times for all subjects using the following intensity function:

$$\begin{aligned}\lambda_i\left(t \,|\, \underset{\sim}{x}, u_i\right) = u_i\rho_0 \exp\{&\alpha_1 x_1 + \alpha_2 x_2 + \alpha_3 x_3 + \alpha_4 x_4 \\ &+ \alpha_5 x_5 + \alpha_6 x_6 + \log(1.05) x_{11}\}.\end{aligned} \tag{4.21}$$

- Step 6*: We then generated post-treatment event times for subjects in treatment and control groups using the following intensity function.

$$\begin{aligned}\lambda_i\left(t \,|\, \underset{\sim}{x}, u_i\right) = u_i\rho_0 \exp\{&\beta_A Trt_{i,A} + \beta_B Trt_{i,B} + \alpha_1 x_1 + \alpha_2 x_2 + \alpha_3 x_3 + \alpha_4 x_4 \\ &+ \alpha_5 x_5 + \alpha_6 x_6 + \log(1.05) x_{11}\}.\end{aligned} \tag{4.22}$$

Note that, at most one of $Trt_{i,A}$ and $Trt_{i,B}$ can be equal to one, and u_i follows a gamma distribution with mean 1 and variance $\phi(= 0.3$ and $0.6)$.

- Step 7*: We recorded the rate of the events that each subject experienced prior to the time of experimental treatments initiation, and then used that for *HM*, *CM* and *PSM*.

- Step 8*: We next used the proposed matching methods and models to match the subjects so that we can estimate the treatment effects.

- Step 9*: Using the matched sample obtained from previous step we calculate theoretical estimate (4.2) and empirical estimate for each of the matched individuals in a matched sample and then calculate the mean of the resulted estimates.

- Step 10^*: We repeat the Steps 1^* to 9^* $B(= 1000)$ times, each of size m and finally the Monte Carlo estimate of the compared treatment effects is obtained by averaging over the 1000 means resulted from simulated data sets.

The estimates obtained from Step 9 and Step 10^* are respectively compared to $\exp\{\beta_S - \beta_N\} = 1.053$ and $\exp\{\beta_A - \beta_B\} = 1.16$, where β_S, β_N, β_A and β_B are the true treatment effects. Results of the Monte Carlo simulations are reported in Tables 4.5 – 4.12. Tables 4.6, 4.8, 4.10 and 4.12 represent the results of the matching methods when covariates x_3 and x_6 are strongly associated with the outcome. In this case, we set $\alpha_3 = 0.9$ and $\alpha_6 = 0.55$ to see how well the proposed matching methods work. Note that, we did not report the *T.E.* for the first scenario since the formula (4.1) results in the true estimate in all the settings. We next summarize the results of the simulation studies.

4.2 Monte Carlo Simulations: Summary and Results

In Chapter 3 our main goal was to assess the strength of *PSM*, *CM* and *HM* methods in eliminating the bias in estimation of time-fixed treatment effect in observational studies. In this chapter, we assumed that individuals under study can change their treatment or choose to receive a different treatment instead of receiving no treatment, the case in Chapter 3. We used the same matching techniques here to estimate and compare the treatment effects. We summarize our findings as follows.

First, based on the results given in Tables 4.5 and 4.6, we found that except the model CM 1 the other matching models resulted in estimates with very small degree of bias. For example, in Table 4.5 the estimate resulted from the model PSM 3 in the presence of overdispersion ($\phi = 0.6$) is equal to 1.0712 which is close to the true value, 1.053. The reason why in the first scenario of our study different matching models resulted in precise estimates is that we used multiplicative intensity function to generate event times and estimate the treatment effects. As a result, the bias arised from a matching model cancels out when we compare the estimated treatment effects. It is important to note that the size of the matched sample should be large enough to be able to eliminate the bias when comparing the treatment effects. For example, in

Table 4.5: Empirical estimates (E.E.'s) resulted from the matching methods in the first scenario (m=1000).

	Without overdispersion ($\phi = 0$)	With overdispersion ($\phi = 0.3$)	With overdispersion ($\phi = 0.6$)
	E.E.	E.E.	E.E.
PSM 1	1.0826	1.0801	1.0839
PSM 3	1.0665	1.0717	1.0712
PSM 6	1.0773	1.0758	1.0693
PSM 7	1.0673	1.0709	1.0584
CM 1	1.1465	1.1736	1.4264
CM 3	1.0631	1.0654	1.0492

Table 4.6: Empirical estimates (E.E.'s) resulted from the matching methods in the first scenario when $\alpha_3 = 0.9$ and $\alpha_6 = 0.55$ (m=1000).

	Without overdispersion ($\phi = 0$)	With overdispersion ($\phi = 0.3$)	With overdispersion ($\phi = 0.6$)
	E.E.	E.E.	E.E.
PSM 1	1.0771	1.0901	1.0715
PSM 3	1.0663	1.0624	1.0529
PSM 6	1.0699	1.0819	1.0837
PSM 7	1.0662	1.0669	1.0769
CM 1	1.1484	1.1688	1.2656
CM 3	1.0643	1.0536	1.0577

Table 4.7: Theoretical estimates (T.E.'s) and empirical estimates (E.E.'s) resulted from the matching methods in the second scenario (m=1000).

	Without overdispersion ($\phi = 0$)		With overdispersion ($\phi = 0.3$)		With overdispersion ($\phi = 0.6$)	
	T.E.	E.E.	T.E.	E.E.	T.E.	E.E.
PSM 1	1.6249	1.9185	2.3153	2.5131	3.9776	2.9648
PSM 2	1.6340	1.9072	2.3309	2.4908	4.8024	2.9238
PSM 3	1.3353	1.5576	1.9144	2.1062	3.4185	2.5947
PSM 4	1.3980	1.6417	1.9694	2.1548	3.4806	2.6193
PSM 5	1.2411	1.4459	1.3245	1.5036	1.5085	1.4871
PSM 6	1.6333	1.9304	2.3170	2.4805	4.0505	2.9068
PSM 7	1.3942	1.6290	2.0046	2.1803	3.4601	2.6741
CM 1	1.2164	1.4069	1.7477	1.9346	3.0295	2.3377
CM 2	1.1948	1.4354	1.3458	1.4983	1.6154	1.5384
CM 3	1.2159	1.4099	1.7422	1.9384	3.0485	2.4227
CM 4	1.2013	1.4223	1.3371	1.5275	1.5899	1.5336
HM	1.3453	1.5737	1.3910	1.5625	1.5654	1.5181

Table 4.8: Theoretical estimates (T.E.'s) and empirical estimates (E.E.'s) resulted from the matching methods in the second scenario when $\alpha_3 = 0.9$ and $\alpha_3 = 0.55$ (m=1000).

	Without overdispersion ($\phi = 0$)		With overdispersion ($\phi = 0.3$)		With overdispersion ($\phi = 0.6$)	
	T.E.	E.E.	T.E.	E.E.	T.E.	E.E.
PSM 1	2.0409	2.3907	2.9549	3.1997	4.9536	3.9628
PSM 2	2.0787	2.3594	2.8895	3.1049	4.9940	3.8969
PSM 3	1.6805	1.9219	2.3918	2.6073	4.1530	3.3444
PSM 4	1.7373	1.9879	2.4357	2.6649	4.1419	3.3157
PSM 5	1.2451	1.3942	1.2982	1.4402	1.4413	1.4497
PSM 6	2.0486	2.3761	2.9788	3.1367	5.1439	3.9538
PSM 7	1.7248	1.9671	2.4583	2.6949	4.2988	3.4169
CM 1	1.5305	1.7459	2.2154	2.4571	3.8006	3.0774
CM 2	1.2385	1.4247	1.3598	1.5335	1.6953	1.6036
CM 3	1.5256	1.7296	2.1733	2.3936	3.8345	3.1151
CM 4	1.2530	1.4379	1.3451	1.5210	1.6009	1.5583
HM	1.2864	1.4383	1.3279	1.4590	1.4558	1.4511

Table 4.9: Empirical estimates (E.E.'s) resulted from the matching methods in the first scenario (m=500).

	Without overdispersion ($\phi = 0$)	With overdispersion ($\phi = 0.3$)	With overdispersion ($\phi = 0.6$)
	E.E.	E.E.	E.E.
PSM 1	1.1245	1.1687	1.1601
PSM 3	1.0915	1.1005	1.0975
PSM 6	1.1308	1.1399	1.1247
PSM 7	1.0812	1.0806	1.0799
CM 1	1.3520	1.4123	1.5009
CM 3	1.2988	1.3378	1.4810

Table 4.10: Empirical estimates (E.E.'s) resulted from the matching methods in the first scenario when $\alpha_3 = 0.9$ and $\alpha_6 = 0.55$ (m=500).

	Without overdispersion ($\phi = 0$)	With overdispersion ($\phi = 0.3$)	With overdispersion ($\phi = 0.6$)
	E.E.	E.E.	E.E.
PSM 1	1.1296	1.1301	1.1341
PSM 3	1.1001	1.0978	1.0120
PSM 6	1.1397	1.1450	1.1401
PSM 7	1.1119	1.1121	1.1080
CM 1	1.2908	1.3589	1.4091
CM 3	1.2703	1.3152	1.3840

Table 4.11: Theoretical estimates (T.E.'s) and empirical estimates (E.E.'s) resulted from the matching methods in the second scenario (m=500).

	Without overdispersion ($\phi = 0$)		With overdispersion ($\phi = 0.3$)		With overdispersion ($\phi = 0.6$)	
	T.E.	E.E.	T.E.	E.E.	T.E.	E.E.
PSM 1	1.6950	1.9880	2.3590	2.5822	4.0103	3.1623
PSM 2	1.6987	1.9801	2.3827	2.5521	4.1561	3.2010
PSM 3	1.3909	1.5800	1.9593	2.1667	3.5210	2.6197
PSM 4	1.4333	1.5054	1.9997	2.4702	3.8966	2.7090
PSM 5	1.2609	1.4219	1.3320	1.5970	1.5997	1.5733
PSM 6	1.7102	1.9822	2.4450	2.5001	4.2509	2.9729
PSM 7	1.5580	1.7610	2.1240	2.2967	3.4517	2.9003
CM 1	1.3919	1.4820	1.8912	1.9901	3.5402	2.7879
CM 2	1.3026	1.4645	1.8083	2.1298	2.7539	2.4448
CM 3	1.2569	1.4513	1.8102	1.9833	3.2918	2.6032
CM 4	1.3073	1.4800	1.5905	1.5945	2.2190	2.1304
HM	1.3625	1.5918	1.5647	1.6010	2.1756	1.9983

Table 4.12: Theoretical estimates (T.E.'s) and empirical estimates (E.E.'s) resulted from the matching methods in the second scenario when $\alpha_3 = 0.9$ and $\alpha_3 = 0.55$ (m=500).

	Without overdispersion ($\phi = 0$)		With overdispersion ($\phi = 0.3$)		With overdispersion ($\phi = 0.6$)	
	T.E.	E.E.	T.E.	E.E.	T.E.	E.E.
PSM 1	2.1392	2.5103	2.9992	3.2417	4.9810	4.3190
PSM 2	2.1400	2.5035	3.0920	3.2025	5.0016	4.1980
PSM 3	1.7890	1.9867	2.7071	2.9120	4.4330	3.5502
PSM 4	1.7929	1.9815	2.6923	2.9304	4.4519	3.3603
PSM 5	1.3217	1.4329	1.4064	1.4827	1.5160	1.5279
PSM 6	2.4528	2.5439	3.1776	3.2215	5.8963	4.2660
PSM 7	1.8995	2.0933	2.9009	3.0198	4.3360	3.8720
CM 1	1.6517	1.9123	2.1866	2.6778	3.9910	3.4492
CM 2	1.4526	1.4689	1.4993	1.5823	1.8841	1.8353
CM 3	1.6277	1.8415	2.0223	2.5419	3.9005	3.2255
CM 4	1.3630	1.4790	1.4890	1.5790	1.6903	1.6217
HM	1.3101	1.4596	1.3728	1.4991	1.4944	1.5007

the model CM 1 where we match the individuals on the covariates that are associated with treatment selection, the estimates are relatively biased due to the small matched number of observations.

Second, we demonstrated that failure to include some important confounders in the *PSM* or *CM* models can result in a higher degree of bias. For example, in Table 4.8 the results of the model PSM 1 are more biased comparing to the results of PSM 7. It can be concluded that including the covariates that are solely associated with treatment selection does not improve the accuracy of the estimates and in some cases may even increase the bias due to decreased number of matched subjects. For example, in Table 4.7 the estimate under the model PSM 1 in the absence of overdispersion is equal to 1.6249, while the estimate under of the model PSM 3 equals to 1.3353 which is much closer to the true estimate, 1.16. The same conclusion can be made based on the estimates resulted from the model PSM 6 in Table 4.7. Based on this, we can conclude that in developing a *PSM* model, true confounders play the most crucial role.

Third, we observed that among all the *PSM* models, the model PSM 5 has the lowest degree of bias which led us to the conclusion that beside other confounding variables, matching individuals on their history balance out noises caused by unmeasured covariates. Furthermore, it is worth mentioning that in the models in which the history of the possible matched subjects is adjusted, the estimates are relatively robust regardless of the degree of overdispersion. For example, in Tables 4.7 and 4.8 the estimates resulted from the models PSM 5, CM 2, CM 4 and *HM* support this statement. On the other hand, the other models resulted in more biased estimates when the degree of overdispersion increased. We recommend that in the case where researchers cannot or miss to measure some important covariates, including history in the matching model can account for those unmeasured covariates to some degree depending on how informative the history is.

Fourth, when we decreased the population size to 500, the degree of bias increased for all the settings. In particular, *CM* and *PSM* models resulted in more degree of bias. For example, in Table 4.11 the model CM 1 in the absence of overdispersion resulted in an estimate with a bias of 0.322, while the same model in Table 4.7 resulted in an estimate with a bias of 0.247. The reason of this result can be linked to the fact that *CM* is generally useful when there are a few explanatory variables in the

matching model, and in particular if the population size is small, it may result in small matched sample and therefore lead to a more biased estimate. It is worth noting that the results of *CM* method could be worse if the number of treated subjects in the population is small. In this case, *HM* method also resulted in more biased estimates comparing to the corresponding estimates when $m = 1000$. For example, in Table 4.12, the empirical estimate resulted from *HM* method in the presence of overdispersion ($\phi = 0.3$) is equal to 1.4991 which is 3.5 per cent more biased in comparison to the corresponding estimate in Table 4.8. The reason why *HM* method did not cause any noticeable degree of bias is that in *HM* method we only use one variable to match the individuals and therefore it is easier to find comparison unit(s) for treated units.

Chapter 5

Analysis of an Illustrative Data Set

In this chapter, we consider a real world example from an epilepsy study to illustrate the matching methods used in the previous chapters. To this end, we generated a synthetic data set based on the information obtained from the recently published literature on recurrent epileptic seizures in adults. The generated data set includes variety of explanatory variables that help to assess the capability of applied matching methods in details. Our primary goal in this chapter is to determine how well propensity score matching (*PSM*) works when count and continuous explanatory variables are available along with binary explanatory variables in a given data set.

5.1 Epileptic Seizures in Adults

Seizures are caused by some abnormal activities in the brain due to a central nervous system (neurological) disorder. Anyone can develop epilepsy and it affects both males and females of all races, ethnic backgrounds and ages. The onset of epilepsy is most common in children and older adults, but the condition can occur at any age. For majority of people with epilepsy there are a few ways to control seizures, including treatment with medication and surgery. In many scientific and research papers, it has been shown that patients with epilepsy may undergo several seizure attacks in a weak, month or year (e.g., Moran et al. (2004); Hoppe et al. (2007); and Viteva (2014)). This allows us to have informative histories for patients under study which in return make it possible for us to apply the history matching (*HM*) method.

There are many things that make seizures more likely for some people with epilepsy. These are often called *triggers*. Below, we mention some of the seizure triggers that have been reported by people with epilepsy:

1. Lack of enough sleep.
2. Stress.
3. Alcohol and recreational drugs.
4. Living in urban and industrial areas.

The reasons why sleep deprivation can trigger seizures are not clearly known, but seizure specialists believe that changes in the brain's electrical and hormonal activity occurring during sleep can be related to why lack of sleep can provoke seizures. Stress is another trigger because the areas of brain responding to stress overlap with the areas important for seizures. Moreover, stress also causes sleep disorders, which may provoke seizures. Other triggers such as drinking alcohol, using recreational drugs and living in mega-cities with excessive noise may negatively affect brain activities and cause stress, which eventually lead to seizure attacks. We use these triggers as explanatory variables that affect the outcome of interest; that is, a seizure attack. In addition to these variables, we consider two other explanatory variables, age and gender, which play an important role in having seizure attacks.

Similar to the second scenario considered in Chapter 4, in this study we assume that the follow-up time for each of the individuals under study is 10 years, and there are two different treatments; Treatment A and Treatment B. Individuals under study can only choose to receive one of these treatment. Our secondary goal is to estimate and compare the effects of these two treatments.

5.2 Matching Methods and Models

In this section, we develop the propensity score models using some observed baseline covariates. Covariates that we use in the *PSM* models are gender, stress, living in urban and industrial areas, age and years of schooling. Table 5.1 shows the association between explanatory variables and outcome or treatment selection.

Table 5.1: The presence or absence of association of explanatory variables with the treatment selection and the outcome considered in data generation.

	Associated with trt. selection	Not associated with trt. selection
Associated with outcome	St.[1], G.[2], U.I.[3], Age	H.S.[4], Al.[5], R.D.[6]
Not associated with outcome	Sc.[7]	-

In Table 5.1, the variables G., U.I., Al. and R.D. represent the binary variables and Age, H.S., Sc. and St., represent the continuous and count variables, respectively. Following *PSM* models are used for estimating propensity scores for individuals under study.

- PSM 1: This model contains all variables associated with the treatment selection.
- PSM 2: This model includes all the true confounding variables that are associated with both the treatment selection and outcome.
- PSM 3: In this model, we obtain propensity scores using the true confounders followed by an additional adjustment for variable representing the history of the subjects.
- PSM 4: All explanatory variables are included in the propensity score model.
- PSM 5: All outcome-related covariates are included in the model.
- PSM 6: In this model, we obtain propensity scores using all outcome-related covariates followed by an additional adjustment for variable representing the history of the subjects.

Based on the findings in Austin (2009b), we apply caliper matching, where calipers of width of 0.2 of the standard deviation of the logit of the estimated propensity scores are used. Austin (2009b) showed that matching on the logit of propensity score, using

[1]Stress
[2]Gender
[3]Urban and Industrial Areas
[4]Hours of Sleep
[5]Alcohol
[6]Recreational Drugs
[7]Years of Schooling

calipers of width 0.2 of the standard deviation of the logit of the propensity score, tended to have superior performance for estimating treatment effects compared with other competing methods that are used in the medical literature. For PSM 3 and PSM 6, we adjust the history of the potential matched individuals so that they have the same number of pre-treatment events.

For *CM*, we consider the following four cases:

- CM 1: In this case, we match the subjects on the variables that are associated with the treatment selection.
- CM 2: All the outcome-related covariates are included in this model.
- CM 3: In this case, we use the true confounders to match the subjects.
- CM 4: We match the subjects on the true confounders and histories of subjects.

In *CM* method, we use exact matching for the binary covariates. For continuous and count variables we apply a caliper of width of 0.2 of the standard deviation of the associated covariate. In order to match the subjects on their histories, we find the subjects with the same number of pre-treatment events.

Since the pre-treatment follow-up time is the same for all the individuals, we used the previous number of events experienced by individuals prior to the time of treatment initiation to represent the history of individuals. Therefore, we considered an untreated individual as a match for a treated individual in the *HM* method if they have the same number of pre-treatment events.

5.3 Data-generating Process

We now present the steps used for the data-generating process in the absence and presence of overdispersion.

Table 5.2: Parameters used to estimate the propensity scores.

$\beta_{0,treatment}$	β_1	β_2	β_3	β_4	β_5
-0.6	log(5)	log(1.2)	log(5)	log(1.18)	log(0.5)

Table 5.3: Parameters used to generate event times.

ρ_0	β_A	β_B	α_1	α_2	α_3	α_4	α_5	α_6	α_7
1.25	-1.25	-1.75	0.15	-0.08	0.2	0.2	0.2	0.05	0.05

- We considered 20,000 patients with epilepsy. For each of them, we generated covariates given in Table 5.1 as follows:

  1. We generated the variable G. from a Bernoulli distribution with the probability of success 0.57. Note that if G.=1, then the individual is male; otherwise female.
  2. The variable Age was generated from a Normal distribution with parameters $\mu = 39.5$ and $\sigma = 12.3$.
  3. The variable U.I. was generated from a Bernoulli distribution with probability of success 0.75.
  4. The variable Al. was generated from a Bernoulli distribution with probability of success 0.8.
  5. The variable Dr. was generated from a Bernoulli distribution with probability of success 0.3.
  6. The variable H.S. was generated from a Normal distribution with with parameters $\mu = 7.5$ and $\sigma = 2$.
  7. The variable St. was generated from a Negative Binomial distribution with notation $NB(r = 20, p = 0.56)$.
  8. The variable Sc. was generated form a Normal distribution with parameters $\mu = 16$ and $\sigma = 4$.

- We generated a treatment status for each of the individuals by using the following binary model

$$Trt_i \sim Bernoulli(p_{i,treatment}) \qquad i = 1, 2, \ldots, 20{,}000, \tag{5.1}$$

where the propensity score model is defined as

$$logit(p_{i,treatment}) = \beta_{0,treatment} + \beta_1 G._i + \beta_2 St._i + \beta_3 U.I._i + \beta_4 Age_i + \beta_5 Sc._i. \tag{5.2}$$

Note that, here if $Trt_i = 1$ then subject receive treatment A; otherwise treatment B. The values of parameters in the logistic regression model (5.2) is given in Table 5.2.

- We assigned the individuals to either treatment group or control group. Treatment group consists of individuals who received treatment B and control group consists of individuals who received treatment A.

- We generated pre-treatment event times for the first two years of follow-up time for all subjects using the following intensity function.

$$\begin{aligned}\lambda_i\left(t \mid \underset{\sim}{x}\right) = \rho_0 \exp\{&\alpha_1 St._i + \alpha_2 Sl._i + \alpha_3 U.I._i + \alpha_4 Al._i \\ &+ \alpha_5 Dr._i + \alpha_6 Age_i + \alpha_7 G._i\}, \qquad i = 1, 2, \ldots, 20{,}000.\end{aligned} \tag{5.3}$$

- We then generated post-treatment event times for subjects in treatment and control groups using the following intensity function.

$$\begin{aligned}\lambda_i\left(t \mid \underset{\sim}{x}\right) = \rho_0 \exp\{&\beta_A Trt_{i,A} + \beta_B Trt_{i,B} + \alpha_1 St._i + \alpha_2 Sl._i + \alpha_3 U.I._i + \alpha_4 Al._i \\ &+ \alpha_5 Dr._i + \alpha_6 Age_i + \alpha_7 G._i\}, \qquad i = 1, 2, \ldots, 20{,}000.\end{aligned} \tag{5.4}$$

Parameters used in the intensity functions (5.3) and (5.4) and their values are given in Table 5.3.

- We matched the individuals using the proposed matching methods and models.

- For the matched sample obtained from the previous step, we compared the effects of two treatments in two forms:

  1. Theoretical estimate

$$T.E. = \frac{E(N_A(t))/(\tau - 2)}{E(N_B(t))/(\tau - 2)}, \tag{5.5}$$

  where $E(N_A(t))$ and $E(N_B(t))$ represent the expected number of post-treatment events for the matched individuals in the control and treatment groups, respectively, and $\tau = 10$ which represents the end of the follow-up time.

2. Empirical estimate
   In order to calculate the empirical estimate, the expected number of events in T.E. given in (5.5) is replaced with the corresponding observed number of events.

- Finally estimates resulted from previous step were compared to $\exp\{\beta_A - \beta_B\} = 1.65$

Steps required for generating data in the presence of overdispersion is the same as the steps given above. The only difference is that, instead of the intensity functions (5.3) and (5.4), we used the following event generating models, respectively.

$$\begin{aligned}\lambda_i\left(t \,|\, \underset{\sim}{x}, u_i\right) = u_i\rho_0 \exp\{&\alpha_1 St._i + \alpha_2 Sl._i + \alpha_3 U.I._i + \alpha_4 Al._i \\ &+ \alpha_5 Dr._i + \alpha_6 Age_i + \alpha_7 G._i\}, \qquad i = 1, 2, \ldots, 20{,}000,\end{aligned} \tag{5.6}$$

and

$$\begin{aligned}\lambda_i\left(t \,|\, \underset{\sim}{x}, u_i\right) = u_i\rho_0 \exp\{&\beta_A Trt_{i,A} + \beta_B Trt_{i,B} + \alpha_1 St._i + \alpha_2 Sl._i + \alpha_3 U.I._i + \alpha_4 Al._i \\ &+ \alpha_5 Dr._i + \alpha_6 Age_i + \alpha_7 G._i\}, \qquad i = 1, 2, \ldots, 20{,}000,\end{aligned} \tag{5.7}$$

Table 5.4: Estimates resulted from different matching methods and models

	Without Overdispersion ($\phi = 0$)		With Overdispersion ($\phi = 0.3$)	
	T.E.	E.E.	T.E.	E.E.
PSM 1	3.5799	3.9651	4.8265	5.5907
PSM 2	2.3235	2.4115	3.2023	3.4519
PSM 3	1.6644	1.7294	1.6765	1.7762
PSM 4	3.8927	3.9412	5.6948	5.6437
PSM 5	2.4434	2.5219	3.3241	3.5064
PSM 6	1.6584	1.7182	1.6627	1.7547
CM 1	1.7457	1.8079	2.5776	2.7427
CM 2	1.6727	1.7375	2.4378	2.5925
CM 3	1.7241	1.7775	2.4004	2.5169
CM 4	1.6678	1.7387	1.6866	1.8188
HM	1.6744	1.7312	1.6793	1.7670

where the random effect u_i follows a gamma distribution with mean 1 and variance $\phi = 0.3$, representing a moderate amount of heterogeneity commonly seen in such studies.

5.4 Results and Balancing Test based on the Generated Data Set

Based on the results given in Table 5.4, it can be concluded that in our settings *PSM* resulted in a higher degree of bias comparing to *CM* and *HM*. The reason for this is that *PSM* failed to perfectly balance some key covariates between treated and untreated subjects. For example, the variables Age and Stress have profound effects on the event rate and may result in a higher degree of bias if they are not sufficiently balanced between treatment and control groups.

One of the commonly used numerical balance diagnostics is the standardized mean difference. The standardized mean difference can be used to compare balance in baseline covariates between treated and untreated units in a matched sample. It can also be used to evaluate the propensity score balance. The concept and formulas for standardized mean difference has been thoroughly discussed by Austin (2009a) and Stuart (2010). There is no consensus as to what value of a standardized difference would denote important residual imbalance between treated and untreated subjects in the matched sample (Austin, 2009a). However, it is recommended that the standardized mean difference should be close to zero for propensity scores (Austin, 2011b), and for continuous covariates, the standardized mean difference should be less than 0.25 standard deviation units (Stuart (2010); Clearinghouse (2014)). Finally, the standardized mean difference for categorical covariates should be less than 0.10 (Austin, 2009a). It should be noted that, for any type of covariate, the closer the standardized mean difference to zero, the better the matched sample is. Therefore, regardless of the mentioned recommendation, researchers should carefully identify key covariates that are prognostically important to apply the matching method.

The standardized mean differences for each of the covariates in the *PSM* models are reported in Tables 5.5 and 5.6. A few conclusions can be made based on the absolute value of the standardized mean difference.

First, if a covariate is not a predictive of treatment assignment, then the standardized mean difference of that covariate is more likely to increase after matching. This may be particularly true for covariates with small differences between the means of treatment and control groups before matching (Stuart, 2010). In our simulation study, the covariate H.S. is not associated with the treatment selection. As shown in Tables 5.5 and 5.6, the standardized mean difference has increased for the models PSM 4 and PSM 5. This issue was addressed by adjusting the history of the possible matched subjects (model PSM 6).

Another finding of our simulation study was that if a covariate is weakly associated with treatment assignment, but has a relatively strong effect on the outcome, then the *PSM* method cannot fairly balance that covariate between treated and untreated units. This is especially true for count and continuous covariates with big values since they can make a noticeable difference in the event rate. For example, the covariates St. and Age are weakly associated with the treatment assignment. On the other hand, they can significantly increase the event rate. Based on the results reported in Tables 5.5 and 5.6, except the models PSM 3 and PSM 6, the propensity score models did not adequately balance out these covariates between the treatment and control groups as they resulted in a higher degree of bias. In contrast, the models PSM 3 and PSM 6 resulted in lower standardized mean difference and less degree of bias in estimation of treatment effects. Therefore, it can be concluded that, in the case of having informative history, some history adjustment may help to reduce the residual imbalance between the treated and untreated subjects.

Table 5.5: The standardized mean differences for the *PSM* models ($\phi = 0$). B.M. and A.M. stand for before matching and after matching, respectively.

	G.		U.I.		Al.		R.D.		St.		Age		H.S.		Sc.	
	B.M.	A.M.	B.M.	A.M.	B.M.	A.M.	B.M.	A.M.	B.M.	A.M.	B.M.	A.M.	B.M.	A.M.	B.M.	A.M.
PSM 1	32.33	-7.0198	32.204	-6.3583	-	-	-	-	35.817	3.9121	83.204	16.408	-	-	-127.49	10.357
PSM 2	32.426	-2.1588	29.1	-0.4493	-	-	-	-	34.788	-3.8224	85.502	3.0608	-	-	-	-
PSM 3	30.301	1.6882	30.099	0.18902	-	-	-	-	38.964	-1.0214	85.954	0.65254	-	-	-	-
PSM 4	34.746	-4.5056	30.062	2.2322	-0.0512	-3.165	0.88127	-5.1074	33.872	1.3372	86.537	7.7623	-1.1891	10.159	-130.26	5.4993
PSM 5	33.657	-0.8822	31.538	2.1759	0.4369	2.7067	-1.0407	1.2858	39.322	-3.4443	84.689	1.5956	0.2904	-2.4814	-	-
PSM 6	33.579	0.7828	26.738	1.1534	-2.1351	1.0842	-0.4510	0.1039	36.365	-1.1808	85.383	0.6717	0.9783	-0.0558	-	-

Table 5.6: The standardized mean differences for the *PSM* models ($\phi = 0.3$). B.M. and A.M. stand for before matching and after matching, respectively.

	G.		U.I.		Al.		R.D.		St.		Age		H.S.		Sc.	
	B.M.	A.M.	B.M.	A.M.	B.M.	A.M.	B.M.	A.M.	B.M.	A.M.	B.M.	A.M.	B.M.	A.M.	B.M.	A.M.
PSM 1	31.09	9.9671	29.068	-1.177	-	-	-	-	36.581	5.1628	82.911	8.2901	-	-	-126.92	11.431
PSM 2	35.147	-1.1342	30.017	2.7045	-	-	-	-	33.64	-1.4961	87.543	0.9508	-	-	-	-
PSM 3	33.541	-1.2745	30.776	0.9117	-	-	-	-	34.999	1.4024	85.452	0.6389	-	-	-	-
PSM 4	32.552	-9.9704	30.132	-7.1505	0.0177	-4.576	1.6846	5.0825	36.695	2.3153	83.64	-4.6997	-0.70781	3.0791	-129.08	-9.8582
PSM 5	31.16	2.3614	30.568	-1.3408	0.68087	2.458	-2.3757	2.74	37.541	-1.7449	86.147	0.5912	0.41593	1.4922	-	-
PSM 6	32.762	-0.37258	27.342	-0.0398	0.9519	-0.6526	1.1821	-1.9093	36.601	1.7554	83.358	0.1236	0.25641	2.8605	-	-

Chapter 6

Summary and Future Work

In this thesis, we considered the estimation of treatment effects in observational studies with recurrent event outcomes. In particular, we assessed the accuracy and effectiveness of three matching methods called propensity score matching (PSM), covariate matching (CM) and history matching (HM) in the estimation of treatment effects in observational health studies. In this chapter, we present a summary and conclusion of our study. Future work is given in the last section.

6.1 Summary and Conclusions

Experimental studies are considered as gold standard in epidemiology and health research to investigate the effects of treatments. These studies are typically called randomized controlled trials if the subjects are human. Many standard models and methods have been developed to analyze data arising from such studies. An important aspect of randomized controlled trials is that the assignment of subjects to a treatment or control group is randomly conducted by using a chance mechanism so that the treatment or control groups may become comparable. Randomization may not be possible in some studies. In this case, methods developed for the analysis of randomized controlled trials should be carefully applied because they may lead to wrong conclusions.

Another important class of study designs used to measure the treatment effects is the observational studies. Contrary to experimental studies, typical observational

studies do not include randomization as a design principle. The main goal with the observational studies is to develop cause-and-effect type of relationships between explanatory variables and outcome variables when randomized controlled trial is not feasible. There are many studies compared these two important classes of designs with advantages and disadvantages.

In this study, we mainly focused on the observational studies and discussed the estimation of treatment effects in rucurrent events. To this end, we considered *PSM*, *CM* and *HM* methods for matching in observational studies. It should be noted that the use of *PSM* and *CM* methods has been relatively well-documented when the outcome of interest is not of recurrent type. As noted by Smith and Schaubel (2015), the matching methods have not been discussed in detail especially when there is an interest to assess the treatment effects. To fill this gap, we considered commonly used *PSM* and *CM* methods, as well as a relatively new matching method called *HM*, which is applicable in recurrent event studies. *HM* has been applied by Smith and Schaubel (2015) in a recurrent event setting. To our knowledge, this has not been discussed extensively by others. In order to make more general conclusions, we used *PSM*, *CM* and *HM* techniques in various settings commonly seen in epidemiology studies. We considered time-fixed treatment effects in Chapter 3 and time-varying treatment effects in Chapter 4 through Monte Carlo simulations. In the simulations, we focused on the bias in estimation of the treatment effects and did not discuss the variance because most of the observational studies matching procedures are population based and usually the standard errors are negligible. The results of our study can be summarized as follows.

First, we demonstrated that *HM* or any other model in which the history is adjusted provide the best matched sample. When an outcome-related covariate was omitted from the matching model, we showed that including the history in that model can greatly decrease the bias due to excluded covariate. This result was expected since the history is caused by all measured and unmeasured outcome-related covariates, and therefore, can be used as an alternative to them. It should be noted that the use of the history as an alternative approach to deal with unmeasured or unobserved covariates should not be regarded as a panacea as this conclusion thoroughly depends on how informative the history is. Furthermore, we observed that estimates resulted from *HM* or the models in which histories of potential matched subjects are adjusted are relatively robust to overdispersion.

Second, we demonstrated that covariates that are associated with the treatment selection but not the outcome should not be included in the *PSM* or *CM* models. Their inclusion can potentially increase the number of bad matches and in most cases decrease the size of the matched sample, which eventually leads to a higher degree of bias. Based on this point, it can be concluded that, in developing and applying the *PSM* and *CM* methods, one should only include true and potential confounders in the matching model. This will help to maximize the number of matched subjects, which in return increases the accuracy of the estimation of treatment effects.

Third, *CM* may result in precise estimates if there are a few covariates on which subjects need to be matched. In contrast, if there are too many covariates, models based on *CM* may result in many treated subjects being excluded from the matched sample. As a result, the bias may increase. In such cases, methods based on *PSM* are preferred as they summarize all covariates in a single quantity. Furthermore, we showed in Chapter 4 that it is better to avoid models based on *CM* when the population size is small.

Fourth, if a confounding variable has a noticeable impact on the outcome, then it is better to follow the idea of the randomized block design, and adjust that variable between treated and untreated subjects prior to conducting any matching method. This helps to improve the performance of matching methods by reducing the bias, which could be a result of balancing the key prognostic covariates on average.

Fifth, based on the simulation studies presented in Chapters 3 and 4, it can be concluded that in our settings the importance of true confounders in developing *PSM* models is more than the importance of potential confounders. For example, in Table 3.5 results obtained from the model PS 7 is more biased compared to the results of the model PS 3. The model PS 7 includes all true and potential confounders, while the model PS 3 includes only true confounders. This result may run counter to intuition for many as the goal of observational matching analysis is to balance the true and potential confounders between treatment and control groups, but the results of the model PS 7 are unexpectedly more biased in spite of the fact that all the outcome-related covariates are included in the model. This issue was addressed by using the model PS 5 where potential confounders are replaced by the history, but future work should thoroughly examine the reasons why the model PS 7 did not result in more precise estimates.

6.2 Future Work

In this section, we suggest some future extensions to our work in this thesis. We aim with the future work to address some of the shortcomings of the current simulations and extending the ideas explored here to deal with several other relevant situations in the causal inference for recurrent events data.

First, we showed in Chapters 3 and 4 that matching subjects on their history can be helpful in increasing the accuracy of estimation of treatment effects. This is only true when the history of the subjects under study is reasonably informative so that it can be used as an alternative to other outcome-related covariates. Therefore, it could be interesting to establish some criteria on when one can use the history of subjects or *HM* method in estimation of treatment effects in recurrent events settings. Moreover, for the *HM* method instead of pre-treatment number of events, pre-treatment gap times between successive events can be used as the history of subjects.

Second, all covariates as well as their effects considered in this thesis are time-fixed. We aim to consider the situation where covariates and their effects vary over time as a future work. Such an extension would be very useful because time-varying covariates and their effects are of interest in many epidemiology and public health studies with recurrent events. For example, in recurrent events analysis the occurrence of a new event usually depends on the previous event occurrences. Therefore, more complex recurrent event models based on event intensity functions can be considered. Furthermore, our study in this thesis can be extended to the case where the effect of a treatment can change over time or when estimation of a new treatment effect in the presence of old treatments effects is of interest. For example, in Chapter 4 we assumed that the effect of the standard treatment does not interfere with the effect of the new treatment. This assumption is not realistic in many real life situations.

Third, throughout the thesis we considered only the multiplicative type of intensity functions to generate event times. It can be useful to redevelop and evaluate the performance of the matching methods when the intensity function is of additive form. The intensity function can also be generalized to include trend component in the baseline rate function.

Fourth, in this study we used three different matching methods to estimate the treatment effects. There are other matching techniques such as stratification on the

propensity score, inverse probability of treatment weighting using the propensity score, mahalanobis distance matching and coarsened exact matching which can be applied in future studies.

Bibliography

Agresti, A. and Min, Y. (2004). Effects and non-effects of paired identical observations in comparing proportions with binary matched-pairs data. *Statistics in Medicine*, 23(1):65–75.

Althauser, R. P. and Rubin, D. (1970). The computerized construction of a matched sample. *American Journal of Sociology*, 76(2):325–346.

Andersen, P. K. and Gill, R. D. (1982). Cox's regression model for counting processes: a large sample study. *The Annals of Statistics*, 10(4):1100–1120.

Austin, P. C. (2008). A critical appraisal of propensity-score matching in the medical literature between 1996 and 2003. *Statistics in Medicine*, 27(12):2037–2049.

Austin, P. C. (2009a). Balance diagnostics for comparing the distribution of baseline covariates between treatment groups in propensity-score matched samples. *Statistics in Medicine*, 28(25):3083–3107.

Austin, P. C. (2009b). Some methods of propensity-score matching had superior performance to others: results of an empirical investigation and monte carlo simulations. *Biometrical Journal: Journal of Mathematical Methods in Biosciences*, 51(1):171–184.

Austin, P. C. (2011a). An introduction to propensity score methods for reducing the effects of confounding in observational studies. *Multivariate Behavioral Research*, 46(3):399–424.

Austin, P. C. (2011b). Optimal caliper widths for propensity-score matching when estimating differences in means and differences in proportions in observational studies. *Pharmaceutical Statistics*, 10(2):150–161.

Austin, P. C. (2014). A comparison of 12 algorithms for matching on the propensity score. *Statistics in Medicine*, 33(6):1057–1069.

Austin, P. C., Grootendorst, P., and Anderson, G. M. (2007). A comparison of the ability of different propensity score models to balance measured variables between treated and untreated subjects: a monte carlo study. *Statistics in Medicine*, 26(4):734–753.

Austin, P. C. and Laupacis, A. (2011). A tutorial on methods to estimating clinically and policy-meaningful measures of treatment effects in prospective observational studies: a review. *The International Journal of Biostatistics*, 7(1):1–32.

Brookhart, M. A., Schneeweiss, S., Rothman, K. J., Glynn, R. J., Avorn, J., and Stürmer, T. (2006). Variable selection for propensity score models. *American Journal of Epidemiology*, 163(12):1149–1156.

Byar, D. (1980). The veterans administration study of chemoprophylaxis for recurrent stage i bladder tumours: comparisons of placebo, pyridoxine and topical thiotepa. In *Bladder Tumors and Other Topics in Urological Oncology*, pages 363–370. Plenum, New York.

Cai, J. and Schaubel, D. E. (2004). Marginal means/rates models for multiple type recurrent event data. *Lifetime Data Analysis*, 10(2):121–138.

Chen, B. E. and Cook, R. J. (2004). Tests for multivariate recurrent events in the presence of a terminal event. *Biostatistics*, 5(1):129–143.

Clearinghouse, W. W. (2014). Procedures and standards handbook (version 3.0).

Cochran, W. G. and Rubin, D. B. (1973). Controlling bias in observational studies: A review. *Sankhyā: The Indian Journal of Statistics, Series A*, 35(4):417–446.

Cook, R. J. and Lawless, J. (2007). *The statistical analysis of recurrent events*. Springer Science & Business Media.

Cook, R. J. and Lawless, J. F. (2002). Analysis of repeated events. *Statistical Methods in Medical Research*, 11(2):141–166.

Cottone, F., Anota, A., Bonnetain, F., Collins, G. S., and Efficace, F. (2019). Propensity score methods and regression adjustment for analysis of nonrandomized studies with health-related quality of life outcomes. *Pharmacoepidemiology and Drug Safety*, 28(5):690–699.

Cox, D. R. (1972). Regression models and life-tables. *Journal of the Royal Statistical Society: Series B (Methodological)*, 34(2):187–202.

Cox, D. R. and Isham, V. (1980). *Point processes*, volume 12. CRC Press.

Daley, D. J. and Vere-Jones, D. (2003). *An introduction to the theory of point processes. Vol. I. Probability and its Applications*. Springer-Verlag, New York.

Daley, D. J. and Vere-Jones, D. (2007). *An introduction to the theory of point processes: volume II: general theory and structure*. Springer Science & Business Media.

Dehejia, R. H. and Wahba, S. (2002). Propensity score-matching methods for nonexperimental causal studies. *Review of Economics and Statistics*, 84(1):151–161.

Faries, D. E., Obenchain, R., Haro, J. M., and Leon, A. C. (2010). *Analysis of observational health care data using SAS*. SAS Institute.

Gail, M., Santner, T., and Brown, C. (1980). An analysis of comparative carcinogenesis experiments based on multiple times to tumor. *Biometrics*, 36(2):255–266.

Grossman, J. and Mackenzie, F. J. (2005). The randomized controlled trial: gold standard, or merely standard? *Perspectives in Biology and Medicine*, 48(4):516–534.

Heckman, J., Ichimura, H., Smith, J., and Todd, P. (1998). Characterizing selection bias using experimental data. *Econometrica*, 66(5):1017–1098.

Heckman, J. J., Ichimura, H., and Todd, P. E. (1997). Matching as an econometric evaluation estimator: Evidence from evaluating a job training programme. *The Review of Economic Studies*, 64(4):605–654.

Hoppe, C., Poepel, A., and Elger, C. E. (2007). Epilepsy: accuracy of patient seizure counts. *Archives of Neurology*, 64(11):1595–1599.

Imbens, G. W. (2004). Nonparametric estimation of average treatment effects under exogeneity: A review. *Review of Economics and Statistics*, 86(1):4–29.

Jacobsen, M. (2006). *Point process theory and applications: marked point and piecewise deterministic processes*. Springer Science & Business Media.

Kalbfleisch, J. and Prentice, R. (2002). The statistical analysis of failure time data. john wiley & sons. *Inc., Hoboken, New Jersey*.

Kelly, P. J. and Lim, L. L.-Y. (2000). Survival analysis for recurrent event data: an application to childhood infectious diseases. *Statistics in Medicine*, 19(1):13–33.

Lawless, J. F. (2003). *Statistical models and methods for lifetime data*, volume 362. John Wiley & Sons.

Lawless, J. F. and Nadeau, C. (1995). Some simple robust methods for the analysis of recurrent events. *Technometrics*, 37(2):158–168.

Lawless, J. F., Nadeau, C., and Cook, R. J. (1997). Analysis of mean and rate functions for recurrent events. In *Proceedings of the First Seattle Symposium in Biostatistics*, pages 37–49. Springer.

Lee, E. W., Wei, L., Amato, D. A., and Leurgans, S. (1992). Cox-type regression analysis for large numbers of small groups of correlated failure time observations. In *Survival analysis: state of the art*, pages 237–247. Springer.

Liang, K.-Y., Self, S. G., Bandeen-Roche, K. J., and Zeger, S. L. (1995). Some recent developments for regression analysis of multivariate failure time data. *Lifetime Data Analysis*, 1(4):403–415.

Liang, K.-Y., Self, S. G., and Chang, Y.-C. (1993). Modelling marginal hazards in multivariate failure time data. *Journal of the Royal Statistical Society: Series B (Methodological)*, 55(2):441–453.

Lin, D. Y., Wei, L.-J., Yang, I., and Ying, Z. (2000). Semiparametric regression for the mean and rate functions of recurrent events. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, 62(4):711–730.

Lunceford, J. K. and Davidian, M. (2004). Stratification and weighting via the propensity score in estimation of causal treatment effects: a comparative study. *Statistics in Medicine*, 23(19):2937–2960.

Moran, N., Poole, K., Bell, G., Solomon, J., Kendall, S., McCarthy, M., McCormick, D., Nashef, L., Sander, J., and Shorvon, S. (2004). Epilepsy in the united kingdom: seizure frequency and severity, anti-epileptic drug utilization and impact on life in 1652 people with epilepsy. *Seizure*, 13(6):425–433.

Nelson, W. (1988). Graphical analysis of system repair data. *Journal of Quality Technology*, 20(1):24–35.

Pepe, M. S. and Cai, J. (1993). Some graphical displays and marginal regression analyses for recurrent failure times and time dependent covariates. *Journal of the American Statistical Association*, 88(423):811–820.

Prentice, R. L., Williams, B. J., and Peterson, A. V. (1981). On the regression analysis of multivariate failure time data. *Biometrika*, 68(2):373–379.

Raynor Jr, W. J. (1983). Caliper pair-matching on a continuous variable in case-control studies. *Communications in Statistics-Theory and Methods*, 12(13):1499–1509.

Robins, J. M., Hernan, M. A., and Brumback, B. (2000). Marginal structural models and causal inference in epidemiology. *Epidemiology*, 11(5):550–560.

Rosenbaum, P. R. and Rubin, D. B. (1983). The central role of the propensity score in observational studies for causal effects. *Biometrika*, 70(1):41–55.

Rosenbaum, P. R. and Rubin, D. B. (1984). Reducing bias in observational studies using subclassification on the propensity score. *Journal of the American Statistical Association*, 79(387):516–524.

Rosenbaum, P. R. and Rubin, D. B. (1985a). The bias due to incomplete matching. *Biometrics*, 41(1):103–116.

Rosenbaum, P. R. and Rubin, D. B. (1985b). Constructing a control group using multivariate matched sampling methods that incorporate the propensity score. *The American Statistician*, 39(1):33–38.

Rothman, K. J., Greenland, S., Lash, T. L., et al. (2008). *Modern epidemiology*, volume 3. Wolters Kluwer Health/Lippincott Williams & Wilkins Philadelphia.

Rubin, D. B. (1973a). Matching to remove bias in observational studies. *Biometrics*, 29(1):159–183.

Rubin, D. B. (1973b). The use of matched sampling and regression adjustment to remove bias in observational studies. *Biometrics*, 29(1):185–203.

Rubin, D. B. (1977). Assignment to treatment group on the basis of a covariate. *Journal of Educational Statistics*, 2(1):1–26.

Rubin, D. B. (2001). Using propensity scores to help design observational studies: application to the tobacco litigation. *Health Services and Outcomes Research Methodology*, 2(3-4):169–188.

Rubin, D. B. and Thomas, N. (1996). Matching using estimated propensity scores: relating theory to practice. *Biometrics*, 52(1):249–264.

Rubin, D. B. and Thomas, N. (2000). Combining propensity score matching with additional adjustments for prognostic covariates. *Journal of the American Statistical Association*, 95(450):573–585.

Senn, S. (1994). Testing for baseline balance in clinical trials. *Statistics in medicine*, 13(17):1715–1726.

Smith, A. R. and Schaubel, D. E. (2015). Time-dependent prognostic score matching for recurrent event analysis to evaluate a treatment assigned during follow-up. *Biometrics*, 71(4):950–959.

Smith, A. R., Zhu, D., Goodrich, N. P., Merion, R. M., and Schaubel, D. E. (2018). Estimating the effect of a rare time-dependent treatment on the recurrent event rate. *Statistics in Medicine*, 37(12):1986–1996.

Smith, J. A. and Todd, P. E. (2005). Does matching overcome lalonde's critique of nonexperimental estimators? *Journal of Econometrics*, 125(1-2):305–353.

Stuart, E. A. (2010). Matching methods for causal inference: A review and a look forward. *Statistical science: a review journal of the Institute of Mathematical Statistics*, 25(1):1–21.

Vansteelandt, S. and Daniel, R. M. (2014). On regression adjustment for the propensity score. *Statistics in Medicine*, 33(23):4053–4072.

Viteva, E. I. (2014). Seizure frequency and severity: how really important are they for the quality of life of patients with refractory epilepsy. *Annals of Indian Academy of Neurology*, 17(1):35–42.

Wei, L.-J., Lin, D. Y., and Weissfeld, L. (1989). Regression analysis of multivariate incomplete failure time data by modeling marginal distributions. *Journal of the American Statistical Association*, 84(408):1065–1073.