Continuous-time Marginal Structural Models for Adverse Drug Effects in Pharmacoepidemiology

by

Yixing Dong

A thesis submitted in conformity with the requirements
for the degree of Master of Science
Department of Public Health Sciences
University of Toronto

Abstract

Continuous-time Marginal Structural Models for Adverse Drug Effects in Pharmacoepidemiology

Yixing Dong
Master of Science
Department of Public Health Sciences
University of Toronto
2021

Analysis of new user cohort studies of adverse drug effects can be based on either intention-to-treat or as-treated paradigms. In the latter case, controlling for time-dependent confounding is necessary, and can be implemented through inverse probability of treatment (IPT) weighted estimation of marginal structural models (MSM). To develop the weights in the context of pharmacoepidemiological administrative data, besides a time-fixed initial dose model, the subsequent time-varying exposure can be decomposed into modeling the time of the dispensations and the dose at each time through a multinomial or multiordinal logistic model, as a marked point process. We compare different approaches for constructing the weights in a simulation study, simulating data from a multistate model. For estimation of the outcome MSM, we propose IPT weighted case-base sampling, which can reduce the size of long-format data sets needed, especially for rare outcomes. The methods are then demonstrated in a study of the effect of glucocorticoid use on fracture risk.

Acknowledgements

Undertaking this master thesis has been a truly fascinating experience, and this would not have been going smoothly without the help of many people.

I want to express my deepest gratitude to my thesis committee, Dr. Olli Saarela (supervisor) and Dr. Suzanne Cadarette (co-supervisor), and Dr. Zhihui Amy Liu. Thank you for taking me as your student, giving me the freedom to decide the topic, and providing me insightful and patient guidance throughout the entire journey. I have learned numerously from your expertise, and it has been a motivating and inspiring year of working with you. Secondly, I would also like to thank Dr. Ehsan Karim and Dr. Tony Panzarella for making time to review my thesis and attending my thesis defense.

I would like to further thank Dr. Olli Saarela, Dr. Tony Panzarella, Dr. Wendy Lou, and our program and graduate assistants Ryan Rosner and Vinita Krishnan for helping me come up with backup plans regarding pandemic related disruptions. I would like to thank my ICES analyst, Nancy He, for pulling my research data and data dictionary, updating my data sets with lengthy runtime upon request, and patiently checking my plentiful analysis results regarding ICES privacy policy. I would also like to thank Limei Zhou and Dinah Thorpe for all their help in opening remote data access for me. Thirdly, I would like to thank the professors in our Master's program in Biostatistics at the University of Toronto for teaching me to be a better statistician. I also want to express my gratitude to other colleagues/labmates/classmates I have worked with during the past two years.

A big thank you to my labmate, Thai-Son Tang, for the stimulating discussions, for organizing events for the students, and for being an outstanding TA and friend. I would also like to thank my dearest remote friends, Jenny, Aurelien, Hailey, and Bing, for your unconditional support across different time zones.

Last but not least, I must express my profound gratitude to my mother and father for the endless encouragement and being my most beloved listeners.

Table of Contents

Chapter **Page**

List of Abbreviations

ACE	Average Causal Effect
ADR	Adverse Dose Reactions
ALD	Alendronate
BP	Bisphosphonate
COPD	Chronic Obstructive Pulmonary Disease
CTS	Canadian Thoracic Society
DAD	Discharge Abstract Database
DAG	Directed Acyclic Graph
DDD	Defined Daily Dose
EHR	Electronic Health Record
EMA	European Medical Agency
FDA	Food and Drug Administration
FP	Fractional Polynomial
GC	Glucocorticoid
GLM	Generalized Linear Models
HR	Hazard Ratio
ICD	International Classification of Diseases
IPTW	Inverse Probability Treatment Weighting
ITT	Intention-To-Treat
MLE	Maximum Likelihood Estimator
MPP	Marked Point Processes

MSM	Marginal Structural Models
NACRS	National Ambulatory Care Reporting System
ODD	Ontario Diabetes Dataset
OHIP	Ontario Health Insurance Plan
OP	Osteoporosis
PHAC	Public Health Agency of Canada
PH	Proportional Hazard
RCT	Randomized Controlled Trials
RWD	Real-world data
SNM	Structural Nested Models
SRS	Spontaneous Reporting System
TD	Time-Dependent
TIM	Transition Intensity Matrix
WCE	Weighted Cumulative Exposure
WHO	World Health Organization

List of Tables

List of Figures

Chapter 1

Introduction

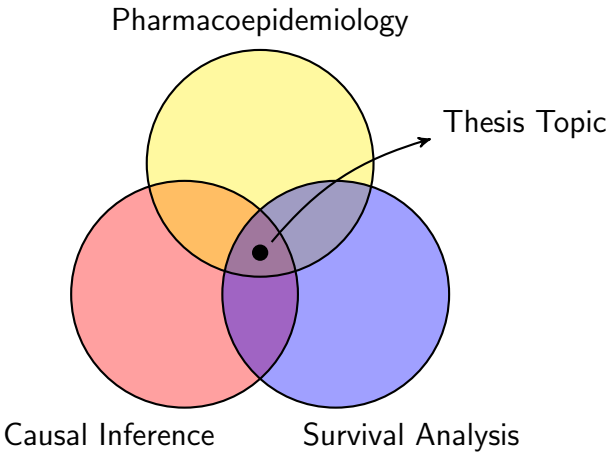


Figure 1.1: This thesis project lies within the intersection of pharmacoepidemiology, causal inference and survival analysis.

In this thesis, the methodological focus is on continuous-time models, however, some of the concepts are more straightforwardly understood in discrete-time. Therefore, before we provide an integration of the longitudinal causal framework under continuous-time counting process expression for time-dependent exposure effects on time-to-event outcome modelling in Chapter 2, we start our introduction by reviewing the three essential areas of our methodology - pharmacoepidemiology, causal inference, and survival analysis Figure 1.1, with some intuitive causal diagrams and assumptions representing discrete-time setting. Thereafter, we summarize the methods previously used for modeling time-dependent drug exposure under time-dependent confounder in different clinical contexts. We then introduce a motivating example on glucocorticoid use and fracture risk, and state our research objectives and their importance.

1.1 Clinical Background

1.1.1 Pharmacoepidemiology and adverse events

Randomized Clinical Trial

With the tremendous progress and breakthrough in pharmaceutical research, especially with the aim for precision medicines and targeted treatment, we have witnessed a large number of new medicines coming to the market (Garbe and Suissa, 2014; Ghassemi et al., 2020). In fact, every approved medicine must have been through the four phases of clinical trials that impose stringent assessments on safety and efficacy. From a target compound in animal studies to clinical trials in human subjects to health authorities submission and eventually, to marketing, the final probability of passing through all these phases is only 11% (Van Norman, 2016). *Randomized Clinical Trials* (RCT) that are double-blinded and placebo-controlled can be served as the gold standard of clinical research; if randomization was done properly, it eliminates potential bias by study design (Burden, 2014; Hernan and Robins, 2020). However, let alone the skyrocketing cost, RCT suffers from limited patient enrollment (Van Norman, 2016). For example, only 3% of the osteoporosis patients visited at a clinic were included in a clinical trial, and even with the least strict inclusion criteria, only 21% were eligible to enter a study (Dowd et al., 2000). Furthermore, it is understandable that patients entering a clinical trial with high hopes for a cure do not want to be placed into the placebo group, merely being served as a comparison to the treatment group but with no alleviation in their disease progression at the end.

Electronic Health Records and Confounding

In recent years, the proliferation of *electronic health records* (EHR) provided alternative opportunities for healthcare. EHR is documented to support daily operational tasks, such as physician billing database for revenue cycle management, real-time tracking of high-frequency biosensor data, and even diagnostic doctor notes (Ghassemi et al., 2020). One subtype of EHR widely used in medical research is the *record linkage databases*, where various sources of data, such as disease registries, drug dispensation, hospitalization, birth and death and visit to clinics entered by general practitioners, are linked through a pseudo-generated unique patient ID. In Canada and Scandinavian countries, this system is set up for the administrative purpose of reimbursement for healthcare providers (Garbe and Suissa, 2014). Some observational administrative databases, also named *real-world data* (RWD) in the pharmaceutical industry, are gaining popularity by expanding the coverage of evidence on top of RCT. For example, *synthetic control methods*, whereas its first proposal falls in the social science sector (Abadie et al., 2010), has been recently adopted to complement RCT results (Beaulieu-Jones et al., 2020; Schmidli et al., 2020). By using meta-analysis on data from historical trials or

sampling patients who have propensity scores (which we will introduce later) matched with the treatment group from an administrative database, some clinical trials can reduce the number of subjects needed for placebo or even replace the placebo arm with a synthetic control arm for comparison purpose (Schmidli et al., 2020). For example, European Medical Agency (EMA) accepted the delivery of a comparison study of the single-arm non-small-cell lung cancer treatment Alecensa versus the standard-of-care cohort from RWD (Beaulieu-Jones et al., 2020; Davies et al., 2018). This hybrid approach partially solves the dilemma, that we are able to preserve the scientific necessity of randomization without leaving patients in the placebo arm untreated. Other than EMAs green light, a pharmaceutical company like Roche has also launched an investment on the RWD curator Flatiron (Petrone, 2018). We have to acknowledge that observational data is revolutionizing the way we study drug safety and efficacy. However, it remains controversial whether the RWD can replace the placebo arm in every RCT completely.

Unlike the experimental RCTs, EHRs are purely observational and sometimes can be collected without a statistical perspective in mind. Most observational studies contain underlying conditions that are not perfectly balanced between the treated versus untreated groups. When conducting statistical analysis, introducing additional covariates might even reverse the estimated direction of the treatment effect, a phenomenon known as Simpsons Paradox (Pearl, 2009). In other words, the association between exposure and outcome is causation in RCT, but not in observational studies. Treatment status that is varying over time are more prone to time-dependent confounding and selection bias (Ali et al., 2016; Burden, 2014; Schmidli et al., 2020). Failing to adjust for the imbalanced variables would lead to exaggerated or neglected risks, especially when it comes to public health policy-making and adverse event reporting and management.

Adverse Event Monitoring

Until the ultimate goal of regulatory approval of a molecule, the entire research and development process usually takes twelve years on average (Van Norman, 2016). When it comes to life-threatening medical needs, health authorities, such as U.S. Food and Drug Administration (FDA), once started to adopt accelerated approval paradigms due to criticism on its rigorous approval process (Van Norman, 2016). However, this effort has been heavily debated with the intention of minimizing the impact of ineffective or unsafe products. One disreputable example that happened in 1982 is the retraction of Oraflex, a nonsteroidal anti-inflammatory agent that leads to the reported death of patients after the drug being released to the market. Furthermore, even after a decade of clinical trials, some *adverse dose reactions* (ADR) do not appear until a sufficient number of users are exposed in the real-world setting. An investiga-

tion of the five drugs removed from the US market between 1997 and 1998 reveals sometimes it takes 200 to 20,000 times of the pre-marketing sample size in the post-marketing use to manifest rare ADRs that lead to drug withdrawal (Garbe and Suissa, 2014).

Compared to RCTs that investigate drug effectiveness, observational pharmacoepidemiological studies usually focus on safety or adverse effects. As introduced by Garbe and Suissa (2014), early in the 1960s, most western countries adopt the *spontaneous reporting system* (SRS) for post-approval adverse event surveillance. SRS collects ADRs reported by physicians and pharmacists, and this information is then passed on to regulatory authorities and pharmaceutical companies. The causality between a drug and its ADRs is determined by trained reviewers. However, this assessment can be ambiguous for ADRs from treatments with a prolonged induction period. Various sources of bias over time provoke challenges for a conclusion on causality. A reliable monitoring system should not only document the ADRs but also actively empower the prediction (Garbe and Suissa, 2014). Therefore, many epidemiological study designs and statistical methods provide guidance on the modeling perspective, as a complement to the SRS.

As a summary of this section and to provide a formal definition, pharmacoepidemiology research usually takes place after the randomized clinical trial. It uses real-world evidence, such as the administrative database, for monitoring adverse effects of post-marketing drugs on large populations that were unselected into clinical trials (Garbe and Suissa, 2014).

1.1.2 Cumulative dose-response curves

The dose-response relationship is used to depict the impact of a stimulus on the environment. In pharmacology, the drug molecules enter the metabolic pathway, bind to the receptors and uptake by the body, and the accumulation of dosage over time triggers treatment or side-effects on the body (Boslaugh, 2007). An illustration of a *dose-response curve* can be served as convincing evidence for causation. To accommodate the large span of dose, pharmacological studies usually plot log-transformed dosage on the X-axis and the risk of the outcome on the Y-axis, and most of the curves can be modeled through a sigmoid function (Motulsky and Christopoulos, 2004). In the untransformed case, the dose-response curve can be any shape, such as linear, U shape, or having a constant risk until a dosage threshold, and then the risk exponentially grows afterwards (Boslaugh, 2007). The shape of a dose-response curve is determined by many factors, such as strength, frequency and duration of drug utilization, and their combined effect of the cumulative dose can be modeled as a continuous variable over time. Before the establishment of generalized linear model, the conventional approach for modeling a continuous variable is by categorizing dosage into ordinal levels and then use

Mantel-Haenszel trend tests, an extension of chi-squared tests, for case and control group odds ratio comparison (Boslaugh, 2007). This trend test approach leaves many uncertainties at the stage of analysis planning, where the number of dose categories is often derived based on previous similar studies, and categorization limits the functional form of the variable (Becher, 2014).

By using regression models, the cumulative dose of A can still be categorized into D levels, and we obtain the estimates for $D-1$ levels compare to the baseline level. We can also monotonically transform A before putting it into a polynomial regression model, although the best transforming strategy is debatable. *Fractional polynomial* (FP) with degree of one $P=1$ models the monotonic effect; it offers a predefined set of transforming powers, $p \in \{-2,-1,-0.5,0,0.5,1,2,3\}$, when determining what transformation to use (Royston and Altman, 1994). When we introduce a second-order FP $P=2$ to allow for non-monotonic effect, A enters the model again with the above set of powers to select. Usually, a FP with a degree $P \leq 2$ is believed to be sufficient to capture the complexity (Becher, 2014). As an alternative flexible approach, *spline regression* is a non-parametric regression technique for the non-linear dose effect (Greenland, 1995). A spline function consists of piecewise functions smoothly joined at prespecified time points called *knots*. The order of a spline is determined by the degree of piecewise functions, and some commonly used splines are quadratic and cubic splines for constructing a basis. Compared to the categorical analysis of dose-effect which assumes the risk is constant over a period of time, spline regression lets the risk vary within and between categories (Becher, 2014).

While spline regression can be constrained at the boundaries of follow-up period to let the risk smoothly go to zero and is less sensitive to local bias compare to FP (Ramsay et al., 1988), they do not provide a definitive answer in epidemiological research (Royston et al., 1999), as the number and position of knots are user-chosen (Hastie and Tibshirani, 1990).

By integrating cubic spline inside the *weighted cumulative exposure* (WCE) proposed by Breslow et al. (1983) and Thomas (1988) and selecting the best set of spline parameters with BIC criteria, Sylvestre (2009) derived the associational functional form of the impact of Flurazepam on fall risk. The effect of Flurazepam is estimated to accumulate for 10 days and then the impact starts declining.

Another medicine also raises concerns about fall risk among the elderly. *Glucocorticoid* (GC) is a type of steroid hormone that exists in vertebral animals, and it is also a life-saving oral medicine prescribed for various inflammatory conditions, including respiratory conditions, asthma, and *chronic obstructive pulmonary disease* (COPD) (Panday et al., 2014). However, an important side effect of oral GC is that it leads to bone mineral density loss and increased fracture risk (Panday et al., 2014). The impact of GC use on fracture risk is often expressed

in a dose and time-dependent manner (Van Staa et al., 2000; Yasir and Sonthalia, 2019). In a recent cohort study, fracture risk increases with cumulative GC exposure, being nearly 2.5 fold higher for elderlies that have a high cumulative dose >5400mg compare to low dose (Balasubramanian et al., 2018). Early in the first year of initiation, GC has the most severe impact on fracture within the first 6 months and is estimated to attenuate over time, followed by further interruption on bone formation with long-term treatment (Amiche et al., 2016; Van Staa et al., 2000).

1.1.3 Intention-to-treat vs. as treated effects

Intention-to-treat (ITT) analysis is usually the primary evidence required by health authorities for clinical trials (Hernan and Hernandez-Diaz, 2012). Similarly, in pharmacoepidemiology, studies on adverse drug effects are often based on the ITT analysis so that the analysis is straightforward. Under this principle, we only need to balance the two treatment cohorts at baseline. We can then assume baseline exchangeability for the assigned treatment strategies, with the intention that patients have full adherence during the follow-up.

However in longitudinal settings, patients usually have a dynamic treatment strategy, such that they change drug dosage, treatment duration, and they might be on or off treatment during the follow-up with their subject-specific prescription pattern. Therefore, when we want to estimate dose-response of the time-dependent exposure effects, rather than comparing two drugs assigned at baseline, a more complicated as-treated analysis is needed. Contrary to the ITT approach, the *as-treated* analysis aims to investigate the dosage actually received (Ten Have et al., 2008). Inverse probability treatment (IPT) weighting and other causal inference methods can reduce the bias introduced due to non-adherence with as-treated analysis (Hernan and Hernandez-Diaz, 2012).

In terms of study design, a lack of adjusting for past treatment usually results in biased estimation for the treatment effect, no matter time-fixed or time-dependent (Hernan and Robins, 2020). To avoid the influence of prior exposure, we consider a *incident new-user cohort*, in which we restrict our study population to those who had not taken a drug-of-interest before (Ray, 2003). Additionally, Suissa et al. (2017) proposed the *prevalent new-user design* that offers a more thorough comparison between two treatments that are different in market access chronologically, in which patients switch from an old drug to a new drug. Depending on the research question, prevalent new-user cohort derives time-conditional propensity scores from either time-interval-matched or prescription-number-matched exposure sets. Compared to the incident new-user cohort design that only studies treatment-naive patients, this design requires systematic and chronological consideration when implementing its exclusion criteria.

Instead of comparing two alternative treatments, the interest herein is in dose-response effect of a single drug of interest, which will still be studied within a new user cohort.

1.2 Causal inference with time-dependent exposures

As said by Aalen et al. (2008), the greatest ambition in statistical inference is to demonstrate causality. Often, we are able to observe an association between two events, but whether one event caused another is difficult to conclude due to the existence of confounders. Consider a time-fixed dichotomous treatment variable A as an indicator for treated versus untreated. Let Y be an observed outcome variable that can be binary (i.e. yes/no fracture) or continuous (i.e. time to fracture), and X be a confounder that influences both A and Y. As a graphical representation, a *directed acyclic graph* (DAG) can illustrate the relationships between these variables. Figure 1.2.

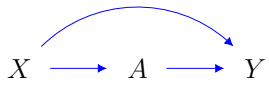


Figure 1.2: Example of a simple DAG

Each individual can be assigned to be either treated or untreated, and the corresponding potential outcomes are generated. We introduce the *potential outcome* notation of Y^a, where the superscript $a \in \{0, 1\}$. For a binary outcome, $Y^{a=1} \in \{0, 1\}$ is the potential outcome of whether a patient would encounter a fracture had he been treated. For a time-to-event outcome, $Y^{a=1}$ means the time until a patient would encounter a fracture if he was treated. In a parallel world paradigm, we would hope to obtain the two versions of potential outcomes Y^1 and Y^0 so that we could derive the individual causal effect contrast as a difference $Y^1 - Y^0$. However, for each person, only one treatment and one potential outcome can be observed, which leads to the fundamental problem of causal inference (Holland, 1986). Instead, we could estimate the causal effect in the population level, for example, the *average causal effect* (ACE) defined by the mean of the difference of the potential outcomes is:

$$ACE = E(Y^1 - Y^0) = E(Y^1) - E(Y^0)$$

where $E(Y^1)$ is the expected potential outcome that we could have been observed, had all patients in this population been assigned to treated group. As long as $E(Y^{a=1}) \neq E(Y^{a=0})$,

we say that the causal effect exists. While mean is the most commonly used ACE metric, there are many forms of population causal contrast, such as median, variance or hazard contrast (Hernan and Robins, 2020). To estimate the population causal effect, we need to make our potential outcome quantities *identifiable*. When the potential outcome can be expressed as a function of the distribution of the observed data, the identifiability of the potential outcome is achieved (Hernan and Robins, 2020). There are three identifiability assumptions, namely, *positivity, consistency* and *conditional exchangeability* (Hernan and Robins, 2020).

When the values of A does not change over time and any confounding effect is solely attributable to the baseline covariate X, we say we have a *time-fixed* setting (Hernan and Robins, 2020), and the above assumptions can be expressed as follow:

- **Positivity**
  The probability of individual i being assigned to each treatment group, under every level of X, must be non-zero, because variability between the causal contrast groups is important for identifiability, i.e. $P(A = a \mid X) > 0$ for all a.

- **Consistency**
  Consistency assumption is the linkage between observed outcome and potential outcome. We assume for any treatment strategy a that is well-defined, the observed outcome under a given treatment assignment is equal to the potential outcome, that is if $A = a$ then $Y^a = Y$.

- **Conditional Exchangeability**
  In a randomized trial, the potential outcome under treated and untreated are exchangeable, $(Y^1, Y^0) \perp\!\!\!\perp A \Rightarrow Y^a \perp\!\!\!\perp A$. In observational studies, we require that within each level of X, the treatment assignment A is random, such that $Y^a \perp\!\!\!\perp A \mid X$. This assumption also only holds when there are *no unmeasured confounders*.

Now consider the longitudinal setting, where we have A and X are measured at multiple time points. Contrary to the previous stationary DAG, the graphical representation now have time-dependent exposure and time-dependent confounder components.

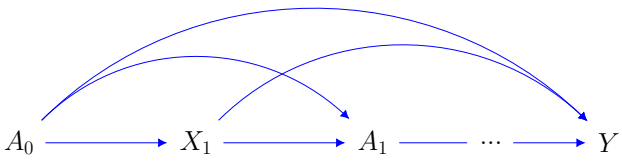


Figure 1.3: Example of a longitudinal DAG

To illustrate the causal assumptions and challenges in the longitudinal setting, we first introduce the notations needed. We use zero-based indexing for treatment time k to be consistent with many publications, where $k = 0, 1, 2, ..., m, ..., K$, and we denote the treatment process $A(t_k) = A_k$ and the covariate process $X(t_k) = X_k$. Let $A_k \in \{0, 1\}$ be a dichotomous indicator for treatment status and let $X_k \in \{0, 1\}$ be the confounder indicator at k^{th} follow-up period or visit in the discrete-time setting. We use an overbar to denote treatment history up to time k, that is, $\bar{A} = \bar{A}_k = (A_0, A_1, ..., A_k)$, and similarly for $\bar{X}_k$. There are two types of treatment strategies, namely *static treatment strategy* and *dynamic treatment strategy*. When current treatment only depends on treatment history $\bar{A}_{k-1}$ but does not depend on covariate histories $\bar{X}_k$, we have a static treatment strategy with $A_k = g(\bar{a}_{k-1})$. Sometimes, the static treatment for each time point k can be determined at baseline, for example, "always treat". On the other hand, we often have dynamic strategies in pharmacoepidemiological research, where we have previous treatment influence current confounder, and current confounder also influence current treatment, a phenomenon called *treatment-confounder feedback* as shown in Figure 1.3. For example, when a physician is prescribing a new cycle of medicine for A, other than referring to previous A use pattern, his judgement might also be influenced by the previous X usage. Meanwhile, the prescription of confounder drug X is also influenced by the target drug A. Therefore, a dynamic strategy has $A_k = g(\bar{a}_{k-1}, \bar{x}_k)$. With these properties, the above time-fixed version assumptions need to be generalized to:

- **Sequential Positivity**
  Positivity holds if at each time point k, the probability of being assigned to each treatment group is never zero, i.e. $f_{A_k|\bar{A}_{k-1},\bar{X}_k}(\bar{a}_k \mid \bar{a}_{k-1}, \bar{x}_k) > 0$, for $\forall\,(\bar{a}_k, \bar{x}_k)$

- **Sequential Consistency**
  For any treatment strategy g, if $A_k = g\left(\bar{A}_{k-1}, \bar{X}_k\right)$ at each time k for an individual i, then the potential outcome under g is equal to the observed outcome $Y^g = Y$.

- **Sequential Conditional Exchangeability**
  Given treatment history up to time $k-1$ and confounder history up to time k Figure 1.3, the potential outcome under strategy g is independent from current treatment assignment A_k, i.e. $Y^g \perp\!\!\!\perp A_k \mid \bar{A}_{k-1}, \bar{X}_k$, for $\forall g$. This assumption is also called *sequential randomization*.

So far, we have provided the formulation of the assumptions in the discrete-time setting, and it is important to lay the intuition of these traditional assumptions before we proceed to the alternative set of assumptions in the continuous-time setting. In continuous-time,

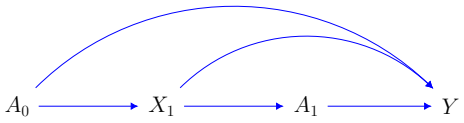
consistency and conditional exchangeability would imply an equivalent causal assumption of *stability* (Dawid et al., 2010). After some background on survival analysis expressed with the counting process notations in Section 1.3, we specify the stochastic causal assumptions that will be used in this thesis later on in Chapter 2.

As mentioned previously, dealing with confounding is a crucial part of observational data analysis. Major causal analysis methods can be categorized into two main fields, namely, *stratification-based methods* and *g-methods*. Both these two categories require the assumption of conditional exchangeability given covariate X. Additionally, there are other methods that do not require conditional exchangeability, such as instrumental variables, but they are more for time-fixed situation and require alternative assumptions (Hernan and Robins, 2020). We summarize the structure of the two major methods as follow:

1. Stratification-based methods:
    - Stratification
    - Matching
    - Outcome regression
2. G-methods
    - IPT weighting with marginal structural models (MSM)
    - Parametric g-formula (standardization)
    - G-estimation of structural nested models (SNM)

Propensity score is defined as the probability of treatment assignment conditional on a confounder level $P(A = a \mid X = x)$ (Rosenbaum, 2002). It is a technique embedded within some of the above methods. For example, among stratification-based methods, stratification and matching are non-parametric methods that rely on propensity score, and outcome regression is a parametric model-based extension of stratification (Hernan and Robins, 2020). Stratification-based methods are traditional in a sense that they adjust the bias by conditioning on confounders along the causal pathway, and return the conditional effect in a subset of the observed population. On the other hand, g-methods simulate the exposure-outcome relation by creating a pseudo-population, where the confounding does not exist. In time-fixed settings, stratification-based methods and g-methods perform equivalently well. However, stratification-based methods fail in the longitudinal setting when we have time-dependent treatment and confounder as well as treatment-confounder feedback. Consider the following scenarios:

(a) Treatment-Confounder Feedback



(b) Null Setting (Collider)

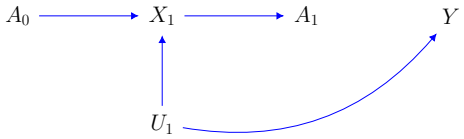


Figure 1.4: Sources of bias in longitudinal causal inference

In Figure 1.4a By conditioning on (remove all arrows attached to) the confounder X_1 that lies on the causal pathway, we dilute the estimate of overall treatment effect on the outcome, because the past treatment effect mediated through a confounder $A_0 \to X_1 \to Y$ is not included. Bias can also arise from conditioning on a collider that opens up a backdoor pathway. For example in the null setting Figure 1.4b, conditioning on X_1 links its parent nodes A and an unmeasured variable U_1 together, and thus we open the pathway $A_0 \to U_1 \to Y$ that introduces treatment effect when in fact, there is no arrow from A to Y. Therefore, in longitudinal studies, conventional methods that require conditioning on elements on the causal pathway fail to handle these challenges.

Propensity score can also be used in g-methods for constructing the pseudo-population, such that the covariate distribution between the treated and the untreated are reweighted. With ITT analysis, the treatment assignment is determined at baseline and each patient has a time-fixed propensity score. Usually, a conditional propensity score is parametrized with η through an *exposure model*, i.e. $P(A = a \mid X = x; \eta) = f(x; \eta)$, such as logistic regression for the binary treatment assignment variable, where the coefficient is estimated with maximum likelihood. Thereafter, the weight of each subject is obtained by taking the inverse of the estimated propensity score:

$$w_i(\eta) = \frac{1}{P(A_i = a \mid X_i; \eta)} \tag{1.1}$$

Adding time-dependent (TD) component to (Eq. 1.1), we need to construct a *long-format data set*, in which each ordered person-time contributes to one record. While the time-fixed baseline covariates are replicated throughout an individual's records, the TD indicators, e.g. $A_i(t), X_i(t)$, are updated along the follow-up time. The discrete-time TD weight up to time m is the cumulative product of the weight at each time point k:

$$w_i(t_m; \eta) = \prod_{k=0}^{m} \frac{1}{P\left(A_{ik} \mid \bar{A}_{i,k-1}, \bar{X}_{ik}; \eta\right)} \tag{1.2}$$

The unstabilized TD weights (Eq. 1.2) increase tremendously for later time points, as we are constantly multiplying a number that is smaller than one to the denominator, which makes the weights at later time points fluctuating (Xiao et al., 2010). The stabilized weight is the ratio between the numerator of a marginal propensity score, parametrized with κ in $P(A = a; \kappa) = f(\kappa)$, and the denominator of a conditional propensity score, parametrized with η, with respect to the TD confounder X_{ik}. The stabilized weights are more preferred, as they improve the efficiency of the estimator over the unstabilized weight Cole and Hernan (2008).

$$sw_i(t_m; \kappa, \eta) = \prod_{k=0}^{m} \frac{P\left(A_{ik} \mid \bar{A}_{i,k-1}, \kappa\right)}{P\left(A_{ik} \mid \bar{A}_{i,k-1}, \bar{X}_{ik}; \eta\right)} \tag{1.3}$$

When there exists a vector of baseline covariates Z_i for each individual, we should let A_{ik} also condition on Z_i in both the numerator and the denominator of (Eq. 1.3).

A model that parametrize the distribution of the potential outcome $P(Y_k^{\bar{a}})$ is considered marginal and structural. It is *marginal* because the model is unconditional on individual-level confounders; it is *structural* because it is expressed as a causal quantity rather than through observed data. Under the causal assumptions, the distribution $P(Y_k^{\bar{a}})$ is the same as the weighted observed outcomes:

$$P\left[Y_k^{\bar{a}}\right] = P\Big(sw_i(t_k; \kappa, \eta) \cdot Y_i \cdot \mathbf{1}\left\{\bar{A}_{ik} = \bar{a}_{ik}\right\}\Big) \tag{1.4}$$

The weighted population is now free of confounding and the outcome can be modeled through a marginal structural model $f(\bar{a}; \theta)$ depending on the research question. We call this model the *outcome model*, and the parameters have causal interpretation.

1.3 Statistical models for time-to-event outcomes

In this section, we review fundamental survival analysis concepts for our marginal structural models in Chapter 2. Time-to-event analysis, also called survival analysis, is widely used for modelling the waiting time it takes until the occurrence of an event-of-interest. Let T be a non-negative random variable for time until an event, and let $E \in \{0, 1\}$ denote whether an event-of-interest (e.g. a bone fracture) has happened, and each individual experiences a set of (T, E) during the follow-up.

General Survival Concepts

One unique feature of time-to-event analysis is censoring, which is essentially a missing data problem. Censoring happens when a person's follow-up time is terminated, and we have not observed an event-of-interest yet, but it may or may not happen sometime later had the follow-up not stopped. Therefore, when we follow a study sample of n subjects for a period of τ years, we might get a combination of complete and incomplete observations in terms of the event-of-interest. We introduce the notation of latent event time $T = \min\{\tilde{T}, C\}$, such that in the presence of censoring, the observed event time T is the minimum of latent event time and censoring time. The event or censoring indicator E takes the value 1 if event happens $T = \tilde{T}$, and 0 if censoring happens $T < \tilde{T}$.

The latent event time $\tilde{T}$ can be characterized as a continuous random variable, with probability density function $f(t)$ (Eq. 1.5) and cumulative distribution function $F(t)$ (Eq. 1.6). We use the complement of $F(t)$ to represent the survival function $S(t)$ (Eq. 1.9), where its connection to instantaneous hazard (Eq. 1.7) and cumulative hazard (Eq. 1.8) functions can be specified (Aalen et al., 2008; Rodriguez, 2007):

$$f(t) = \lim_{\Delta t \to 0} \frac{P(t \leq \tilde{T} < t + \Delta t)}{\Delta t} \tag{1.5}$$

$$F(t) = P(\tilde{T} < t) = \int_0^t f(u)\mathrm{d}u \tag{1.6}$$

$$h(t) = \lim_{\Delta t \to 0} \frac{P(t \leq \tilde{T} < t + \Delta t \mid \tilde{T} \geq t)}{\Delta t} \tag{1.7}$$

$$H(t) = \int_0^t h(u)\mathrm{d}u \tag{1.8}$$

$$S(t) = P(\tilde{T} > t) = 1 - F(t) = \exp\left\{-\int_0^t h(u)\mathrm{d}u\right\} = \exp\{-H(t)\} \tag{1.9}$$

Because $\tilde{T}$ is continuous, $S(t) = P(\tilde{T} \geq t)$ is interchangeable with $S(t) = P(\tilde{T} > t)$.

Thereafter, by the definition of conditional probability, we can derive that:

$$\begin{aligned} P(t \leq \tilde{T} < t + \mathrm{d}t) &= P(t \leq \tilde{T} < t + \mathrm{d}t \mid \tilde{T} \geq t) \times P(\tilde{T} \geq t) \\ \iff f(t) &= h(t)S(t) \end{aligned} \tag{1.10}$$

An interpretation for (Eq. 1.10) would be the density function at time t equals to the product of the rate of event occurrence at time t, given no event has happened until time t, and the probability of survival until time t.

When censoring happens due to dropout, loss to follow-up, or administrative cut-off of the follow-up time, we lose the information on the right side of the time scale, and the data is right-censored (Aalen et al., 2008). A right censoring mechanism is said to be *independent censoring* when an individual i still at-risk at time t has the same risk of experiencing the event in the next time interval $[t + \mathrm{d}t)$ as if censoring would have never taken place, such that

$$P(t \leq T_i < t + \mathrm{d}t, E_i = 1 | T_i \geq t) = P(t \leq \tilde{T}_i < t + \mathrm{d}t \mid \tilde{T}_i \geq t) \tag{1.11}$$

In other word, the observed event time is complete for all i in the study population, and we have $\tilde{T} \perp\!\!\!\perp C$ (Aalen et al., 2008). *Non-informative censoring* assumes further that the censoring does not give additional information on the survival of a subject beyond the censoring time, and the survival distribution does not inform censoring distribution (David and Kleinbaum, 2016). For research using administrative data, the length of follow-up is usually pre-determined by study design, and thus non-informative censoring is usually assumed.

Discrete-Time Survival

In discrete time setting, the survival time is split into intervals, with survival time T now a discrete random variable defined at $t_0 < t_1 < ... < t_K$ (Kalbfleisch and Prentice, 2011). The probabilities of survival time takes place at t_k (Eq. 1.12), the hazard experiencing event given survived until at least t_k (Eq. 1.13) and survival function in discrete time (Eq. 1.14) are defined as follow (Kalbfleisch and Prentice, 2011; Rodriguez, 2007):

$$f(t_k) = f_k = P(T = t_k) \tag{1.12}$$

$$h(t_k) = h_k = P(T = t_k \mid T \geq t_k) \tag{1.13}$$

$$S(t_k) = S_k = P(T \geq t_k) = \sum_{k}^{K} f_k \tag{1.14}$$

The parallel connection of the above properties can be drawn as (Eq. 1.10) in the continuous-

time setting now to the discrete-time setting:

$$\begin{aligned} P(T = t_k) &= P(T = t_k \mid T \geq t_k) \times P(T \geq t_k) \\ \iff f(t_k) &= h(t_k)S(t_k) \end{aligned} \tag{1.15}$$

One distinct difference between the hazard function in continuous-time and discrete-time is that the hazard is now a probability rather than a rate (Rodriguez, 2007). Thereafter, the survival probability can be interpreted as a product of surviving through each previous time points until experiencing an event at the t_k:

$$S_k = (1 - h_1)\,(1 - h_2)\ldots(1 - h_{k-1}) \tag{1.16}$$

This formula is comparable to the continuous-time setting where the survival function involves an integral of previous hazard up to time t.

Counting Process Framework

Alternatively, survival analysis can also be expressed from a process point of view. A *stochastic process* is a set of random variables that are indexed with respect to time; it is widely applied in financial and biological research (Ross, 2014). More generally, in a *point process* the points are randomly distributed over some probability space. In survival analysis, when we count the number of events up to a time t, we have a *counting process*.

The concept of *independent increment* requires the number of events in disjoint time intervals to be independent. One example of a counting process that satisfies the independent increment is the *Poisson process*, in which the inter-arrival time between events are exponentially distributed, and thus it has *memoryless* property such that the next state of a process is only dependent on the current state, regardless of the past. Such memoryless property is also called *Markov property*. A *Markov chain* is a discrete and memoryless stochastic model that depicts a sequence of events. In continuous-time, this becomes a *Markov process*, and it is considered *homogeneous* if the transition intensity does not depend on time (Ross, 2014).

On the other hand, when independent increment is not satisfied, the future depends on the past information. With the evolution of a stochastic process, a *filtration* or history $\mathcal{F}_{t^-}$ up to time t is generated as a σ-algebra, a collection of countable sets (Williams, 1991). For example, the intensity process $\lambda(t)$ of a counting process $N(t)$ depends on $\mathcal{F}_{t^-}$.

We can now demonstrate the connection between the above concepts to another stochastic process, the *martingale*. Let $\{N(t), t > 0\}$ be the number of events we have counted until time t. In survival analysis, we count to the first event, therefore the counting process for each individual is defined as $N_i(t) = \mathbf{1}_{\{T_i \leq t, E_i = 1\}}$. Here we denote the counting process jump

as $\mathrm{d}N_i(t) = N_i(t) - N_i(t^-) \in \{0, 1\}$, indicating whether an event happened at time t for individual i. A counting process has local characteristics of a Poisson process (Aalen et al., 2008). Consider a homogeneous Poisson process defined by its intensity λ, characterizing the probability of event occurring in a small time interval $\mathrm{d}t$. Such idea can be extended to the intensity process $\lambda(t)$ of a counting process that adapted to the history $\mathcal{F}_{t^-}$. For individual i:

$$\lambda_i(t)\mathrm{d}t = P(\mathrm{d}N_i(t) = 1 \mid \mathcal{F}_{t^-}) = E(\mathrm{d}N_i(t) \mid \mathcal{F}_{t^-})$$

Because $\mathrm{d}N(t)$ is binary, the probability is expectation. $\lambda(t)\mathrm{d}t$ can be moved inside the conditional expectation because it also depends on the past. A reformulation gives:

$$E(\mathrm{d}N_i(t) - \lambda_i(t)\mathrm{d}t \mid \mathcal{F}_{t^-}) = 0$$

Now we introduce the martingale residual as the "observed" minus the "expected" value of the counting process, i.e. $\mathrm{d}M_i(t) = \mathrm{d}N(t) - \lambda(t)\mathrm{d}t$. We can then obtain the property:

$$E\left[\mathrm{d}M_i(t) \mid \mathcal{F}_{t-}\right] = 0, \tag{1.17}$$

A process with such property (Eq. 1.17) is a *martingale*. Equivalently, a counting process can be decomposed into a predictable process and a martingale.

$$N_i(t) = \Lambda_i(t) + M_i(t), \tag{1.18}$$

where the *cumulative intensity process* is $\Lambda_i(t) = \int_0^t \lambda_i(u)\mathrm{d}u$. (Eq. 1.18) is called the *Doob-Meyer decomposition* in continuous-time.

So far, the history $\mathcal{F}$ is only for one counting process, while in reality, we can consider the history of many covariate processes and external information that are nested as one $\mathcal{F}_t$. We usually make the assumption that martingales of different counting processes are orthogonal, which means no two counting processes can jump simultaneously. We require this assumption for causality (Aalen et al., 2008).

Another important property is the relations between intensity process and the hazard function:

$$\lambda_i(t)\mathrm{d}t = P(t \le T_i < t + \mathrm{d}t, e_i = 1 \mid \mathcal{F}_{t^-}) = \begin{cases} h(t)\mathrm{d}t, & T_i \ge t \\ 0, & T_i < t \end{cases} \tag{1.19}$$

Usually, patients are not observed during the entire follow-up duration, so here we introduce

an at-risk indicator $G_i(t) = \mathbf{1}_{\{T_i \geq t\}}$ that takes value 1 if the individual is still under observation right before time t, or 0 otherwise. We simplify (Eq. 1.19) with the following connection:

$$\lambda_i(t) = G_i(t)h_i(t) \tag{1.20}$$

Let random variables $T_1, T_2, ..., T_n$ be survival time for n individuals. In absolute continuous time, no two T_i are identical, and information on T_i does not interfere another T_l, for $i \neq l$. With this *independent and identical distributed* (I.I.D) survival time assumption, the aggregated process of all individual processes is $N(t) = \sum_{i=1}^{n} N_i(t)$ for counting process, and $G(t) = \sum_{i=1}^{n} G_i(t)$ for at-risk process. When $h_i(t) = h(t)$ for all i, (Eq. 1.20) can be written as $\lambda(t) = G(t)h(t)$.

Regression Models

Often, we want to model the relationship between the survival duplet (T_i, E_i) and some covariate X_i for parametric models, and we need to relax the independent censoring assumption to: given X_i, the latent event time and censoring time are independent, i.e. $\tilde{T}_i \perp\!\!\!\perp C_i | X_i$. One class of the survival analysis models is the *proportional hazard* (PH) model, in which one unit change in the covariate results in multiplicative effect in the hazard. Let η_1 denote the log hazard ratio. The most common PH model was proposed by Cox (1972):

$$h_i(t) = h_0(t) \exp\{\eta_1 X_i\}, \tag{1.21}$$

where the baseline hazard $h_0(t)$ is left unspecified, which makes (Eq. 1.21) a semi-parametric model. The proportionality of hazard can be shown with the following example. Let $h_i(t)$ be the hazard of being diagnosed of lung cancer and $X = 1$ indicates patient i is a smoker, while $h_l(t)$ is the hazard for a non-smoker patient l with $X = 0$. The hazard ratio with respect to these two different smoking status is a constant, regardless of time:

$$HR = \frac{h_i(t)}{h_l(t)} = \frac{h_0(t) \exp\{\eta_1 \cdot 1\}}{h_0(t) \exp\{\eta_1 \cdot 0\}} = \exp\{\eta\} \tag{1.22}$$

The cumulative baseline hazard can be estimated using the Breslow estimator:

$$\hat{H}_0(t) = \sum_{j:t_j \leq t} \frac{\delta_j}{\sum_{l=1}^{n} G_l(t_j) \exp\{\hat{\eta}_1 X_l\}} \tag{1.23}$$

where t_j is the ordered observed event times, and $\delta_j = \sum_i \mathbf{1}_{\{T_i < t_j\}}$ refers to the number of events at each t_j. For any regression model, the likelihood function is the joint probability

distribution across all i. In a survival model like Cox PH regression, the individual probability distribution can be specified with an event indicator. Under independent and non-informative censoring assumptions, if an individual experiences an event at time t_i, the probability contribution to the likelihood is the density function $f_i(t_i)$; otherwise, if censoring happens at time t_i given survival until time t_i, the probability contribution is $S_i(t_i)$ (Kalbfleisch and Prentice, 2011).

$$\begin{aligned} L &= \prod_{i=1}^{n} P(T_i = t_i, E_i = e_i) \\ &= \prod_{i=1}^{n} f_i(t_i)^{e_i} S_i(t_i)^{1-e_i} \end{aligned} \tag{1.24}$$

Under PH assumption, we can now substitute in (Eq. 1.10) to the full likelihood function, and replace the hazard component with our semi-parametric Cox model (Eq. 1.21). The full likelihood function of the Cox model contains two unknown parameters, the baseline hazard that depends on time $h_0(t)$ and the regression coefficient η_1. Cox (1972, 1975) proposed that we can obtain a *partial likelihood function* that is based solely on η_1 by eliminating the nuisance parameter $h_0(t)$, using conditional likelihood or profiling, or a small trick in Appendix. A.1:

$$\begin{aligned} L(\eta_1, h_0(t)) &\overset{\eta}{=} \prod_{i}^{n} h_i(t_i;\eta)^{e_i} S_i(t_i;\eta) \\ L(\eta_1) &\overset{\eta_1}{\propto} \prod_{i}^{n} \left[\frac{\exp\{\eta_1 X_i\}}{\sum_{l \in G(t_i)} \exp\{\eta_1 X_l\}} \right]^{e_i}, \end{aligned} \tag{1.25}$$

where $\eta = (h_0(t), \eta_1)$ is a shorthand notation. Usually, the estimate for our parameter of interest $\hat{\eta}_1$ can be obtained from software output.

In this thesis, except baseline variables and outcome process, all of the other processes are considered TD and we elaborate on the time-dependent version of models in Chapter 2. If we take $X_i = X_i(t)$, we can use Cox regression to estimate the effect of a time-dependent covariate. Further, Aalen et al. (2008) generalized Cox model with counting process notation:

$$\lambda_i(t) = \lambda_0(t) \exp\{\eta_1 X_i(t)\}, \tag{1.26}$$

where $\lambda_i(t)$ is an intensity process. Recall the connection between intensity and hazard, the model (Eq. 1.20) is equivalent to (Eq. 1.26), the *Andersen and Gill generalized Cox model*.

Poisson regression, also called a log-linear model, assumes proportional hazard between different covariate values as well. Rather than leaving the baseline hazard unspecified, we assume a constant value which makes it a fully parametric model. It is closely connected to Cox model in that if we take log on both side of (Eq. 1.21), we get a Poisson model, where

$\eta_0 = \log(h_0(t))$. In addition, the counting process expression of a nonhomogeneous Poisson intensity model has the same form as (Eq. 1.26) (Lawless, 1987). More formally, consider the simple case of time-fixed modeling for count value a_i, let A_i follow a Poisson distribution $A_i \sim \text{Pois}(h_i)$ with mean $E(A_i \mid X_i) = h_i$. We have the Poisson regression:

$$\log(h_i) = \log\left(E(A_i \mid X_i)\right) = \eta_0 + \eta_1 X_i, \tag{1.27}$$

and the likelihood function is the joint probability mass function from all patients:

$$L(\eta) = \prod_{i=1}^{n} \frac{h_i^{a_i} \cdot \exp\{-h_i\}}{a_i!} \tag{1.28}$$

After substituting in (Eq. 1.27), the maximum of the log of the likelihood does not have a closed form solution, but $\hat{\eta}$ can be estimated with gradient descent techniques through the negative log-likelihood functions. On the practical side, these parameters can be easily obtained from the `glm()` output in R (R Core Team, 2013). When using Poisson regression to model a hazard rate, we need to divide the number of counts over a period of time $E(A_i \mid X_i)/t = h_i$, therefore, we adjust model (Eq. 1.27) with an offset term $\log(t)$.

Administrative data are naturally massive due to large sample size and long follow-up time. With long-format splitting, the dimension can increase tremendously. As an alternative partial likelihood for fitting hazard models, *case-base sampling* offers an elegant way of reducing the data size without having a biased estimator (Hanley and Miettinen, 2009). It is also suitable for modeling time-dependent covariates on a continuous-time scale (Saarela, 2016). Case-base sampling includes all person-time coordinates where an event-of-interest takes place, and then we complete the long-format data set with a random sample of person-times from all observed follow-up time (Hanley and Miettinen, 2009). Case-base sampling has an expression of the logistic regression form, which can be fitted with `glm()` in R. Note that this approach will be discussed in more detail later in Section 2.4.2.

Multistate model

Previously, we have presented methods for survival analysis until the first event occurrence. In reality, an individual might encounter events of distinct types, as some of the event types might interfere the event-of-interest, or experiencing recurrent events of the same type. When we only focus on time until the first event, we have a competing risk problem. One approach to estimate the interrelation between event types is using time-dependent covariates (Kalbfleisch and Prentice, 2011; Young et al., 2020). As an extension to competing risk analysis, *multi-state* model allows us to investigate the transitioning between different event types until the

event-of-interest. Let a Markov process $W(t)$ be the state occupied at time t. $\lambda_{qr}(t)$ is the *transition intensity* that an individual jumps from state q at time t to state r at time $t + \mathrm{d}t$, for $q \neq r$ (Kalbfleisch and Prentice, 2011):

$$\lambda_{qr}(t \mid \mathcal{F}_{t^-}) = \lim_{\mathrm{d}t \to 0} \frac{P(W(t + \mathrm{d}t) = r \mid W(t) = q, \mathcal{F}_{t^-})}{\mathrm{d}t}$$
$$\lambda_{qr}(t) = \lim_{\mathrm{d}t \to 0} \frac{P_{qr}(t, t + \mathrm{d}t)}{\mathrm{d}t}$$

In a homogeneous case, $\lambda_{qr}(t) = \lambda_{qr}$. In addition, in the special case of the competing risks model, the transition intensities reduce into cause-specific hazards. The process $W(t)$ is memoryless, such that only current state is needed for deriving the transition intensity. As an analogy, the transition intensity $\lambda_{qr}(t)$ is correspondent to the hazard function $h(t)$ in a survival regression model. One classical example of a multistate model is the illness-death model Figure 1.5. We can specify a *transition intensity matrix* (TIM) for the multistate model, in which each intensity can be parametrized through a regression model. Conventionally, the rows of TIM sum to zero, where the diagonal elements are negative sums of the off-diagonal elements $\lambda_{qq}(t) = -\sum_{q \neq r} \lambda_{qr}(t)$ (Aalen et al., 2008; Kalbfleisch and Prentice, 2011).

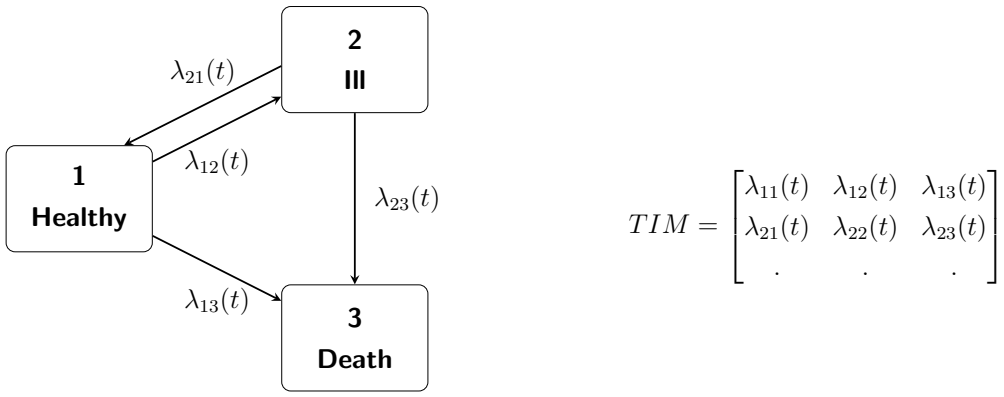


Figure 1.5: Multistate diagram of a health-illness-death model

Recently, by incorporating the counting processes as time-dependent covariates and the concept of causality discussed before, we are able to adapt the graphical representation of a multistate model to dynamic path analysis (Aalen et al., 2008; Fosen et al., 2006). This approach is used in causal mediation analysis (Imai et al., 2010), where treatment happens before confounding Figure 1.6, and the total effect (or causal/marginal effect) of the treatment can be decomposed into direct effect and indirect effect.

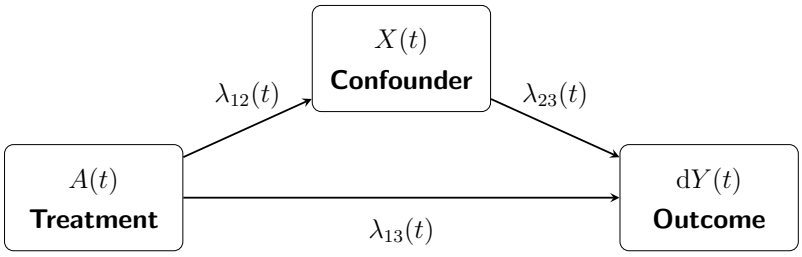


Figure 1.6: Path diagram representing possible transitions of a jump in the counting process

However, in this thesis, the multistate models used were only for data generation in the simulation study, rather than for modeling of the causal relationships. We also do not require treatment and confounder to take place in a causal mediation analysis order as above.

Other than the multistate model, the *marked point processes (MPP)* can also be used to model recurrent events (Mancini and Paganoni, 2019). Nowadays, MPP is widely applied to many fields such as earthquake and medical research. It captures the time of the event (point process) and also the event size (mark), which are predictable and I.I.D conditional on the past $\mathcal{F}_{t^-}$ (Mancini and Paganoni, 2019). Let T and D denote the time and magnitude of a event, respectively.

$$\lambda(t, d \mid \mathcal{F}_{t^-}) = \lambda_V(t \mid \mathcal{F}_{t^-}) f(d \mid \mathcal{F}_{t^-})$$

where $\lambda_V(t \mid \mathcal{F}_{t^-})$ is the intensity of the point process, and $f(d \mid \mathcal{F}_{t^-})$ is a density function of mark D.

1.4 Causal inference in survival analysis

A Cox MSM hazard model that combines many forms of strategies is:

$$h\left[Y_k^{\bar{a}};\theta\right] = h_0(t_k)\cdot\exp\left\{\theta_1 \mathrm{cum}(\bar{a})\right\}, \tag{1.29}$$

where $\mathrm{cum}(\bar{a}) = \sum_{s=0}^{k-1} a_s$. In practice, we can incorporate some spline functions into (Eq. 1.29) to make the modelling flexible (Hernan and Robins, 2020). However, Cox MSM might not yield a causal interpretation due to several reasons. Aalen et al. (2015) argue that after the first event time, the risk sets consist of patients who have not experienced an event-of-interest, which introduces implicit heterogeneity. Furthermore, in a randomized trial, the conditional independence between treatment assignment A and baseline covariates Z among survivors (i.e. $A \perp\!\!\!\perp Z \mid T > t$) is guaranteed if the hazard rate can be decomposed into an additive form:

$$h(t \mid A, Z) = a(t, A) + b(t, Z),$$

which is not satisfied in a conditional Cox model. This property is known as non-collapsibility (Martinussen and Vansteelandt, 2013). In addition, when we marginalize over Z, we lose the proportional hazard. Hence, the individual level treatment effect $\exp\{\theta\}$ in a Cox model cannot be derived from the population treatment effect due to the frailty theories. However, Bartlett (2019) proposed that by viewing hazard as a population distribution of the ratio between the density of the survival time over the survival function, hazard is a valid causal quantity. Nevertheless, the additive models often have more desirable properties as causal quantities compare to the relative risk models (Aalen et al., 2015; Greenland, 1996; Hernan, 2010; Martinussen and Vansteelandt, 2013).

Under aforementioned identifiability assumptions in the high-dimensional setting, an alternative g-method is Robin's g-formula (Robins, 1986), also called standardization in the point treatment setting. With discrete valued covariates X_k, we denote $\bar{a} = \bar{a}_k$ and $\bar{x} = \bar{x}_k$, and a potential hazard outcome for $P(Y_{k+1}^{\bar{a}} = 1 \mid Y_k^{\bar{a}} = 0)$ is modeled through (Young and Tchetgen Tchetgen, 2014):

$$h\left[Y_k^{\bar{a}};\theta,\eta\right] = \sum_{\bar{x}} P\left(Y_{k+1} = 1 \mid \bar{A}_k = \bar{a}, \bar{X}_k = \bar{x}, Y_k = 0;\theta\right)\cdot r_{\bar{a}}\left(k,\bar{x};\theta,\eta\right) / \sum_{\bar{x}} r_{\bar{a}}\left(k,\bar{x};\theta,\eta\right), \tag{1.30}$$

with

$$r_{\bar{a}}\left(k,\bar{x};\theta,\eta\right) = \prod_{s=0}^{k} P\left(Y_s = 0 \mid \bar{a}_{s-1}, \bar{x}_{s-1}, Y_{s-1} = 0;\theta\right) f(x_s \mid \bar{a}_{s-1}, \bar{x}_{s-1}, Y_s = 0;\eta)$$

so that the distribution of the observed outcome given $\bar{a}$ and $\bar{x}$ is standardized to the distribution of confounder history at each time k. When X_k is measured continuously, we can specify an extension of (Eq. 1.30) by replacing the summation with integral. Notice that for this thesis, we focus on continuous-time scale, so the potential outcome expression in (Eq. 1.29) and (Eq. 1.30) is replaced with the randomized trial notation $\mathcal{E}$ that we will introduce in Chapter 2.

However, even though g-formula can be extended to a high-dimensional setting, it is often not used directly due to the computational burden of modeling of the confounder history (Hernan and Robins, 2020). On the other hand, given the connection between g-formula and Cox MSM, for the purpose of simulation studies, we can use (Eq. 1.30) to solve the true Cox MSM parameters from observational data (Young and Tchetgen Tchetgen, 2014). In principle, IPT weighting with MSM, g-formula and g-estimation with SNM (another g-method that we are not going to focus on in thesis) are asymptotically equivalent and are expected to generate ACE that are not significantly different (Hernan and Robins, 2020).

1.5 Recent methodological work in causal inference for pharmacoepidemiology

In longitudinal epidemiological studies, the time-dependent features of patients are followed over time. Although some covariates are naturally continuous in time, the timing of the measurements can force these variables as discrete in time. There are two types of cohorts resulting from discretization. When measurements occur at pre-specified time intervals, irrespective of the evolution of a patient's health condition, we have the classic *interval cohort* that requires a fixed observational plan. On the other hand, when measurements are taken at clinic visits that depend on subject-specific health status, we have a *clinical cohort* that follows a dynamic observational plan (Lau et al., 2007).

The choice of exposure models and simple transformations of IPTWs has been studied in a simulation study with an interval cohort design (Xiao et al., 2010). The authors find that Cox weighted MSM has less biased estimates than pooled logistic weighed MSM (although the difference is small). They also suggest using normalized stabilized weights to reduce the variability of extreme time-dependent weights. Later on, it is shown that instead of normalizing the IPTWs, truncation at 99% or 99.5% of the stabilized weights distribution improves the IPTW estimator, especially for positively skewed weighting distribution (Xiao

et al., 2013).

So far, the studies mentioned previously use only one exposure model for estimating the stabilized IPT weights. Hernan et al. (2009) proposed to separate the IPT weighting into modeling the dynamic observational plan and the treatment assignment in a clinical cohort study. When covariates and outcomes are measured only at clinical visits, missed visits introduce missing data problems. Pullenayegum and Lim (2016) reviewed methods for correcting such problems. One of which is modeling visits as a continuous-time intensity process (Lin et al., 2004). We do not have a similar missing data problem in the current context of pharmacoepidemiology using administrative data, as fracture outcomes are observed regardless of cut-off points such as clinical visits. Also, because the clinical visits have an irregular pattern, we can naturally assume the continuity for drug usage variables, given it is also controversial to find the optimal time length for discretization (Ferreira Guerra et al., 2020).

Historically, the outcome model in an MSM is proposed based on discrete-time with an interval cohort design. Hernán et al. (2000) suggest using weighted pooled logistic MSM to approximate the weighted continuous-time Cox MSM in the outcome model. Since then, most of the MSM applications relied on the follow-up time discretization and used a pooled logistic model or GEE regression for the outcome model (Hernan et al., 2002; Pisani et al., 2015; Xiao et al., 2010; Zhang et al., 2011). However, this approximation is biased with frequent events (Young et al., 2010), which motivates the presentation of the causal framework in continuous-time with counting process notation (Lok et al., 2008).

Roysland et al. (2011) provided a martingale representation of the MSM. The causal effect is defined in terms of a randomized trial measure, where treatment is assumed locally independent of the covariate process. The resulting continuous-time MSM can be estimated using continuous-time weights. In this setting, rather than the causal assumptions introduced in Section 1.2, we adopt causal assumptions in a continuous-time scale (we will introduce in Chapter 2) that emphasize the equivalence of aspects of the observational measure and the randomized trial measure (Dawid et al., 2010). That is, expressing causality without the potential outcome notation (Commenges, 2019).

A few pieces of literature have attempted continuous-time outcome models in MSM. The first continuous-time MSM is the Cox MSM that can be fitted by maximizing the weighted risk sets in a Cox partial likelihood without the approximation from a pooled logistic regression (Xiao et al., 2010). However, they still used discrete-time models to develop the weights.

Given our context, we expect to derive methods that can accommodate variables in full continuous-time, that is, in both exposure and outcome models. Previous research has shown that the estimating function of case-base sampling has a partial likelihood expression and can be applied to continuous-time exposure model to derive IPTW for a continuous-time

Cox MSM (Saarela and Liu, 2016). However, they did not address using case-base sampling for fitting the outcome MSM. The weighted partial likelihood of case-base sampling and its properties have not been derived, which can be an efficient and powerful alternative to the traditional weighted Cox partial likelihood in the continuous-time.

Another recent study also bypassed the causal interpretation dilemma of hazard ratio in a Cox MSM by using additive hazard models (Ryalen et al., 2020). This study also demonstrates the separate modeling of treatment mode, via propensity weights derived from a logistic regression, and time-to-treatment-initiation, through continuous-time weights derived from an additive hazard model in counting process form. Besides, competing risk was handled using a composite endpoint of the earliest event subtype, and the authors use subdistribution hazards to express the weighted cumulative incidences. However, this paper does not consider time-dependent confounders like in our case. Apart from Ryalen et al. (2020), Young et al. (2020) also defined causal quantities under competing risks but in discrete-time, in which we have the total effect of treatment on the outcome or the direct effect where treatment is not mediated through a competing effect.

Another research goal in this paper is to investigate the effect of cumulative exposure on the outcome. Several studies have investigated the modeling of the cumulative dose effect in a causal framework. Bodnar et al. (2004) used data from a randomized trial for estimating the cumulative dose effect of prenatal iron supplementation, under several time-dependent and time-fixed confounders. They categorized the cumulative dose to four categories in both study periods, and they used a logistic MSM model that sums over doses from the two study periods. Xiao et al. (2014) incorporates the cumulative dose estimating methods proposed by Sylvestre and Abrahamowicz (2009) (see Section 1.1.2) into a Cox MSM, where they first use IPTW to remove the confounders to create a pseudo-population and then applied the WCE, in which exposure is further weighted by recency, with a cubic spline function. Nevertheless, these papers assumed that the treatment status occurs at discrete clinical visits and does not express the causal framework in continuous-time with the counting process notation. In our case, a marginal case-base sampling hazard model can also integrate splines for the flexible modeling of the cumulative dose effect.

1.6 Motivation: Glucocorticoid use and fracture risk

Osteoporosis (OP) is a prevalent disease of aging and public health concern that affects over 1.5 million (10%) Canadians aged 40 years and older (PHAC, 2010). Under previously introduced GC-induced osteoporosis in Section 1.1.2, there might be recurrent fractures during a study follow-up period, but the time to first fracture is considered a time-to-event outcome. We define the cause-specific fracture quantity as the event of interest, while death always present as a competing risk.

Bisphosphonate (BP) can be served as a therapy, with 20%-40% reduction in fracture risk in clinical trials (Cranney et al., 2002). Previous evidence has shown that early initiation of BP is beneficial for reducing GC-induced cause-specific fracture hazard (Amiche et al., 2018), and the frequency of COPD prescription is associated with BP use (Chalitsios et al., 2020). In addition, BP persist in bone, which leads to beneficial effects that might last after discontinuation (Burden et al., 2012). Therefore, it is imperative that BP is considered as a time-dependent confounder in modeling the effect of time-dependent GC use on fracture risk. Similarly, other factors that might influence both the exposure of GC and the outcome of fracture, and that may change over time, must be modeled as time-dependent confounders. For example, hospitalization, visits to emergency units and residency in long term care homes, during which period of time, patients might receive high dose of GC and might reduce daily mobility, which might have an impact on both the GC intake and fracture risk.

1.7 Research objectives and importance of our work

For this master thesis, our first objective is to model the exposure as a marked point process, that is to decompose the modelling of cumulative exposure into modelling the time of the dispensations and the dose at each dispensation, specifically in the context of pharmacoepidemiology. Our second objective is to propose IPT weighted case-base sampling for fitting a continuous-time MSM in a causal framework, and we compare its performance with IPT weighted Cox PH model in a simulation study. More specifically, we present the weighted partial likelihood of case-base sampling and derive a variance estimator. The structure of the thesis is as follows. In Chapter 2, we extend the stationary potential outcome framework to a longitudinal setting with counting process-based notations and stochastic assumptions, and we define the formulation of the weighting exposure models and the weighted partial likelihood of the outcome models. In Chapter 3, we use multistate model diagram to repre-

sent longitudinal DAG simulations on four different exposure status scenarios, weighted with different combinations of the estimation methods introduced. In Chapter 4, we select one method from the most relevant simulation scenarios, and illustrate the approach using ICES data with 92,090 GC new user cohort patients. We conclude by discussing our findings and limitations in Chapter 5.

By considering drug dispensations retrieved from record linkage databases as continuous-time counting processes, we avoid discretizing the underlying continuous-time exposure process. Combining the weights derived from the dispensations time model and the daily dose dispensed model, we achieve a closer approximation to the true marginal fracture hazard. We present a simulation study comparing the performance of the continuous-time method to more conventional discretization of the time scale, using a multistate model to generate simulated continuous-time exposure, confounder and outcome data. Furthermore, by extending methodology to incorporate case-base sampling as an outcome model in MSM, we can reduce the size of long-format data set needed, especially for rare outcomes, without having biased estimations. The proposed approach also enables estimation of absolute marginal hazards functions, which can be used to derive predictions under alternative treatment plans.

Chapter 2

Methods

In this section, as an alternative to the potential outcome framework introduced in Section. 1.2 that are commonly used in discrete-time, we present a different causal framework and its assumptions used in continuous-time. In the context of GC exposure and fracture outcome, one single change in the exposure status can be modeled directly as a counting process or decomposed into a marked point process that counts to the first event. We also generalize the single treatment change/discontinuation case into a recurrent MPP to be more closely aligned to the real-world data analysis. For each of the above modelling approaches, we formulate time-dependent IPT weights derived from different survival models introduced in Section. 1.3. Thereafter, we present two weighted partial likelihood functions of Cox and case-base sampling for the continuous-time marginal hazard models.

2.1 Single treatment discontinuation models and weights

2.1.1 Notation and assumptions

Let $A_i(t)$ and $X_i(t)$ denote the exposure and confounder processes, respectively. Here $A_i(t)$ represents categorized dose of the drug of interest at time t. For notational simplicity, we consider there is only one confounder process, while in practice, there could be multiple confounder processes, such as multiple time-dependent health conditions and other drug uses. We follow a cohort of patients $i = 1, ..., n$ from time zero, i.e. the time of the first dispensation where a patient satisfies the follow-up initiation criteria., to a fixed time τ, or death or first fracture, whichever comes first. Let $Y_i(t) \in \{0, 1\}$ be the outcome process, counting until the first fracture, with the at risk process $G_i(t) = 1_{\{Y_i(t^-)=0\}}$ indicating individual i is still under observation at time t. For simplicity, here we let all processes only count to the first event.

Consider a scenario that all patients start on the same dose, but might stop the treatment at some point during the follow-up, i.e. $A_i(t) = 0$ or $1 - A_i(t) = 1$, and stays the same afterwards. In the special case where the exposure process $A_i(t)$ is a counting process, it only counts until the first discontinuation of treatment (the indicator of $A_i(t)$ takes value $1 \to 0$). Here we denote the counting process difference for $A_i(t)$ as $\mathrm{d}A_i(t) = A_i(t + \mathrm{d}t) - A_i(t^-)$, which returns the number of exposure events happened during time $[t, t + \Delta t)$, and similarly for any other counting processes. We also define the baseline covariates Z, but they are suppressed from the notation since all the exposure models are conditional on them. First, the observed information up to time t is given by:

$$\mathcal{F}_t = \sigma\left\{X_i(u), A_i(u), Y_i(u), G_i(t) : i = 1, ..., n, 0 \leq u \leq t\right\}$$

We also introduce the partial observed information up to time t, without the time-dependent confounder $X_i(t)$:

$$\mathcal{F}_t^* = \sigma\left\{A_i(u), Y_i(u), G_i(t) : i = 1, ..., n, 0 \leq u \leq t\right\} \subseteq \mathcal{F}_t$$

The observed treatment process is characterized by the intensity function:

$$\lambda_{A_i}^{\mathcal{O}}(t; \eta)\mathrm{d}t = P^{\mathcal{O}}(\mathrm{d}A_i(t) = -1 \mid \mathcal{F}_{t^-}; \eta), \tag{2.1}$$

parametrized with η. The function (Eq. 2.1) corresponds to the information-based intensity (Arjas et al., 1989), taking value zero when an individual is not at risk $G_i(t) = 0$, or if the treatment discontinuation has already taken place. To express the causal quantities of interest, and the causal assumptions, we introduce the superscript $\mathcal{O}$ and $\mathcal{E}$ for 'observed' and 'experimental' data generating mechanisms (Commenges, 2019; Dawid et al., 2010; Roysland et al., 2011). Under the observed $\mathcal{O}$ setting, the intensity function is conditional on the full history $\mathcal{F}_{t^-}$, while in experimental $\mathcal{E}$ setting, the intensity function is adapted to the partial history $\mathcal{F}_{t^-}^*$, such that it may contain only time-fixed confounding (entirely attributable to baseline covariates) but not the TD confounder process. Under the experimental condition:

$$\lambda_{A_i}^{\mathcal{E}}(t; \kappa)\mathrm{d}t = P^{\mathcal{E}}(\mathrm{d}A_i(t) = -1 \mid \mathcal{F}_{t^-}; \eta) = P^{\mathcal{E}}(\mathrm{d}A_i(t) = -1 \mid \mathcal{F}_{t^-}^*; \kappa), \tag{2.2}$$

is parametrized with κ. We are interested in the marginal outcome hazard under such a 'randomized' dosage assignment. In principle, the target dosage assignment can be chosen. When we choose the marginal assignment probability as the target dosage assignment mechanism, we are able to generate the numerator for the usual kind of stabilized weights. Our

marginal fracture hazard function based on the outcome process is parametrized through our parameters of interest θ.

$$\psi_i(t;\theta)\mathrm{d}t = P^{\mathcal{E}}(\mathrm{d}Y_i(t) = 1 \mid \mathcal{F}^*_{t^-};\theta), \tag{2.3}$$

under randomized setting $\mathcal{E}$. The treatment effect in (Eq. 2.3) can be specified for example through an outcome model $\psi_i(t;\theta) = \exp\{\theta_0(t) + \theta_1(1 - A_i(t))\}$, where $\theta_0(t)$ is a flexibly specified parametric log-baseline hazard function and θ_1 is the treatment effect coefficient and $A_i(t)$ is now a TD variable. We consider (Eq. 2.3) as a marginal structural model. It is marginal because it is unconditional on $X_i(t)$, and structural because under $\mathcal{E}$ setting, we are modeling a causal quantity.

Additionally, for our data generation mechanism (we will introduce in Chapter 3), we specify the conditional fracture hazard :

$$\lambda^{\mathcal{O}}_{Y_i}(t)\mathrm{d}t = P^{\mathcal{O}}(\mathrm{d}Y_i(t) = 1 \mid \mathcal{F}_{t^-}) \tag{2.4}$$

and its 'randomized' version $\lambda^{\mathcal{E}}_{Y_i}(t)$, and the intensity function for TD confounder process:

$$\lambda^{\mathcal{O}}_{X_i}(t)\mathrm{d}t = P^{\mathcal{O}}(\mathrm{d}X_i(t) = 1 \mid \mathcal{F}_{t^-}) \tag{2.5}$$

and its 'randomized' version $\lambda^{\mathcal{E}}_{X_i}(t)$. The functions (Eq. 2.4) and (Eq. 2.5) are not parametrized, but they are needed for expressing the causal assumptions needed for the estimation of (Eq. 2.3).

The causal assumptions for potential outcome framework introduced in Section. 1.2 implies the *stability* assumption, that is, the stochastic way the confounder process evolve, given observed history, is independent of the regime setting (Dawid et al., 2010). Therefore, in our observational and experimental setting, the major causal assumptions are $\lambda^{\mathcal{O}}_{Y_i}(t) = \lambda^{\mathcal{E}}_{Y_i}(t)$ and $\lambda^{\mathcal{O}}_{X_i}(t) = \lambda^{\mathcal{E}}_{X_i}(t)$, that is, the conditional data generating mechanisms under $\mathcal{O}$ and $\mathcal{E}$ are the same, with the exception of the exposure mechanism. The stability assumption is needed to rule out unmeasured confounders. Additionally, we require positivity, or *absolute continuity* of the probability measures under $\mathcal{O}$ and $\mathcal{E}$, that is, $P^{\mathcal{E}} > 0 \Rightarrow P^{\mathcal{O}} > 0$ or $P^{\mathcal{E}} \ll P^{\mathcal{O}}$. In particular, for the dosage assignment mechanism, we require $\lambda^{\mathcal{E}} > 0 \Rightarrow \lambda^{\mathcal{O}} > 0$. In addition, we assume continuous-time setting that no two counting processes can jump simultaneously. We also assume censoring is non-informative with our administrative data.

2.1.2 Estimation of exposure weights

There are three survival models adopted in this thesis, Poisson regression, Cox proportional hazard model and pooled logistic regression, and the corresponding weights generated from these models can be used to weight the partial likelihoods of the outcome models (will be introduced in Section. 2.4). Although Cox regression is developed through a different intuition compare to the other two models, the handling of continuous variables are similar (Becher, 2014). The stabilized weighting function (Eq. 1.3) can be extended with the counting process expression as the likelihood ratio process, which can be further extended for continuous and discrete-time models in the following sections:

$$sw_i(t) = \frac{f_i^{\mathcal{E}}(t)}{f_i^{\mathcal{O}}(t)} = \prod_{u\in[0,t)} \frac{P(\mathrm{d}A_i(u) \mid \mathcal{F}^*_{t^-})}{P(\mathrm{d}A_i(u) \mid \mathcal{F}_{t^-})}, \tag{2.6}$$

The product in the above equation is a product integral, and the other terms in the likelihood ratio cancel out due to the stability assumption. Because we are interested in modeling the treatment effect since the stop of a treatment, the weight is not influenced by the instantaneous hazard before an exposure status change. At the time t_{A_i} when exposure changes, we multiply the survival probability with the instantaneous hazard at time t_{A_i}, and the weight stays the same after the exposure event takes place.

2.1.2.1 Continuous-time exposure models

From (Eq. 2.6), because we are characterizing the exposure as a counting process, it follows that a continuous-time stabilized IPT weight for individual i still at risk at time t can be expressed with treatment assignment intensities under $\mathcal{E}$ and $\mathcal{O}$:

$$sw_i(t) = \exp\left\{-\int_0^t (\lambda^{\mathcal{E}}_{A_i}(u;\kappa) - \lambda^{\mathcal{O}}_{A_i}(u;\eta))\mathrm{d}u\right\} \prod_{u\in[0,t)} \left[\frac{\lambda^{\mathcal{E}}_{A_i}(u;\kappa)}{\lambda^{\mathcal{O}}_{A_i}(u;\eta)}\right]^{\mathrm{d}A_i(u)} \tag{2.7}$$

The above derivation (cf. Roysland et al. (2011), eq. (4.3)) resembles the connection that density function is the product of instantaneous hazard and survival probability in (Eq. 1.10). We can also echo the connection in (Eq. 1.3), where the stabilized weight in continuous time setting can be expressed as the likelihood ratio between the marginal and conditional density functions. In the case of a single treatment change/discontinuation, given the relation between intensity and hazard function, as well as the equivalence of TD regression models and Anderson and Gill generalized intensity models introduced in Section. 1.3, we can instead

model the treatment intensity $\lambda^{\mathcal{E}}_{A_i}(t;\kappa)$ via a marginal hazard model $h_{A_i}(t;\kappa)$, with either Poisson regression or Cox PH model. Similarly, we can obtain the estimate $\hat{\eta}$ from the conditional hazard model $h_{A_i}(t;\eta)$.

Poisson regression

We use Poisson regression to fit a constant hazard model, as they result in the same likelihood expression. Given our current context that survival analysis can be expressed from a process point of view, we use Poisson regression as one of our continuous-time models. We can extended (Eq. 1.27) to TD version

$$\log\left(h_{A_i}(t \mid X_i(t);\eta)\right) = \eta_0 + \eta_1 X_i(t), \tag{2.8}$$

representing the hazard of stopping the treatment, and the marginal model $h_{A_i}(t;\kappa)$ can be obtained by excluding the $X_i(t)$ term. The likelihood functions $L(\eta)$ (Eq. 1.28) and $L(\kappa)$ now have an additional product term over time $u \in [0,t)$. As introduced before, we obtain the estimates $\hat{\eta}_0$, $\hat{\eta}_1$ and $\hat{\kappa}_0$ from the `glm()` output in R. We then specify the TD stabilized weights for the Poisson regression model:

$$\widehat{sw}_i^{\bar{A}}(t;\hat{\kappa},\hat{\eta}) = \begin{cases} \dfrac{\exp\left\{-\int_0^t \exp\left\{\hat{\kappa}_0\right\} \mathrm{d}u\right\}}{\exp\left\{-\int_0^t \exp\left\{\hat{\eta}_0 + \hat{\eta}_1 X_i(u)\right\} \mathrm{d}u\right\}} \cdot 1, & t < t_{A_i} \\ \dfrac{\exp\left\{-\int_0^{t_{A_i}} \exp\left\{\hat{\kappa}_0\right\} \mathrm{d}u\right\}}{\exp\left\{-\int_0^{t_{A_i}} \exp\left\{\hat{\eta}_0 + \hat{\eta}_1 X_i(u)\right\} \mathrm{d}u\right\}} \cdot \dfrac{\exp\left\{\hat{\kappa}_0\right\}}{\exp\left\{\hat{\eta}_0 + \hat{\eta}_1 X_i(t_{A_i})\right\}}, & t \geq t_{A_i} \end{cases} \tag{2.9}$$

Due to the constant hazard assumption in Poisson regression, the instantaneous hazard is given by the exponentiated coefficient estimates. Because we only count until the first the treatment event, when the treatment status changes at t_{A_i}, the weight stays the same afterwards.

Cox Proportional Hazard Model

As an extension to the conventional Cox model (Eq. 1.21), here we specify the time-dependent version for our conditional Cox model for the hazard of treatment termination:

$$h_{A_i}(t \mid X_i(t);\eta) = h_0(t)\exp\left\{\eta_1 X_i(t)\right\}, \tag{2.10}$$

and the marginal model $h_{A_i}(t;\kappa)$ is fitted without the covariate process $X_i(t)$. The likelihood function $L(\eta_1)$ has the same formulation as (Eq. 1.25), except we add the product term over time, replace X with $X_i(t)$, and the event indicator e_i is now the exposure indicator $\mathrm{d}A_i(t)$. We obtain the estimates of η_1 from `coxph()` function in R, and the cumulative baseline hazards $\hat{H}_0^*(t)$ and $\hat{H}_0(t)$ in the marginal and conditional models, respectively, are returned with the Breslow estimator (Eq. 1.23). Then it follows the time-dependent stabilized-weight for individual i is:

$$\widehat{sw}_i^{\bar{A}}(t;\hat{\kappa},\hat{\eta}) = \begin{cases} \dfrac{\exp\left\{-\hat{H}_0^*(t)\right\}}{\exp\left\{-\int_0^t \exp\left\{\hat{\eta}_1 X_i(u)\right\}\mathrm{d}\hat{H}_0(u)\right\}} \cdot 1, & t < t_{A_i} \\ \dfrac{\exp\left\{-\hat{H}_0^*(t_{A_i})\right\}}{\exp\left\{-\int_0^{t_{A_i}} \exp\left\{\hat{\eta}_1 X_i(u)\right\}\mathrm{d}\hat{H}_0(u)\right\}} \cdot \dfrac{\mathrm{d}\hat{H}_0^*(t_{A_i})}{\exp\left\{\hat{\eta}_1 X_i(t_{A_i})\right\}\mathrm{d}\hat{H}_0(t_{A_i})}, & t \geq t_{A_i} \end{cases} \tag{2.11}$$

The instantaneous baseline hazard is obtained from the cumulative baseline hazard by taking $\mathrm{d}\hat{H}_0^*(t)$ and $\mathrm{d}\hat{H}_0(t)$. Similarly, the weight stays the same on and after a change in the treatment status.

2.1.2.2 Discrete-time exposure model

Pooled logistic model

Pooled logistic model, also named sequential logistic model, treats each person-time-interval as a record, and allows for a time-varying intercept term. The transition from continuous-time to discrete-time long format exposure data set requires we cut the follow-up time for each individual into clinically meaningful and predetermined bins. For example, we can choose a monthly (30 days) interval, such that $k = 0, 1, 2, ...$ represents $0^{th}, 30^{th}, 60^{th}, ...$ day of the follow-up. When every subject starts exposed $a_{i0} = 1$ and may stop the treatment at k^{th} interval ($t_{A_i} \in (30(k-1), 30k]$) during the follow-up and stays unexposed afterwards, a_{ik} takes value 1 for $k = 0, ..., m-1$ and value 0 on and after $k = m$. Additionally, we lag the labeling of X_{ik} by one interval for the discrete time model to ensure correct order of the events between exposure and confounder, while in continuous-time, lagging is not needed because by definition no two events can take place simultaneously.

Cox (1972) proposed that Cox PH model can be relaxed to a logistic form in discrete-time:

$$\begin{aligned} \frac{h_i\left(t \mid X_i(t); \eta\right)}{1-h_i\left(t \mid X_i(t); \eta\right)} &= \frac{h_0\left(t\right)}{1-h_0\left(t\right)} \exp\left\{\eta_1 X_i(t)\right\} \\ \iff \operatorname{logit}\left(h_i\left(t_k \mid X_{ik}; \eta\right)\right) &= \eta_0 + \eta_1 X_{ik} \end{aligned} \tag{2.12}$$

where $\eta_0 = \operatorname{logit} h_0(t_k)$. Recall from Section. 1.3, in discrete-time setting, the hazard $h_{A_i}(t_k \mid X_i(t); \eta)$ is a probability $P(A_{ik} = 0 \mid X_{ik}; \eta)$, also denoted as p_{ik}, rather than a rate. At each time interval, an individual's exposure status is a Bernoulli random variable: $A_{ik} \sim \text{Bern}(p_{ik})$, with mean equal to p_{ik}. We can then derive the logistic model for each fixed-length interval k:

$$\operatorname{logit}\left(P(A_{ik} = 0 \mid X_{ik}; \eta)\right) = \eta_0 + \eta_1 X_{ik}, \tag{2.13}$$

for modeling the hazard of treatment discontinuation, and the marginal model $h_{A_i}(t; \kappa)$ can be fitted without the $X_i(t)$ term. We define the inverse of logit as expit:

$$p_{ik} = \operatorname{expit}(\eta_0 + \eta_1 X_{ik}) = \frac{\exp(\eta_0 + \eta_1 X_{ik})}{1 + \exp(\eta_0 + \eta_1 X_{ik})} \tag{2.14}$$

The estimates of $\eta = (\eta_0, \eta_1)$ can be obtained by maximizing the likelihood via fitting the model with `glm()`:

$$\begin{aligned} L(\eta) &= \prod_{i=1}^{n} \prod_{s=0}^{k} P\left(A_{is} = a_{is} \mid X_{is} = x_{is}; \eta\right) \\ &= \prod_{i=1}^{n} \prod_{s=0}^{k} \operatorname{expit}(\eta_0 + \eta_1 x_{is})^{a_{is}} \cdot \left[1 - \operatorname{expit}(\eta_0 + \eta_1 x_{is})\right]^{1-a_{is}} \end{aligned}$$

and similarly solving $L(\kappa)$ for $\hat{\kappa}$. Unlike continuous-time models that require decomposition of the probability density function $f_i(t)$ into a survival probability and an instantaneous hazard of stopping the treatment, the probabilities $\hat{p}_{ik}$ of stopping the treatment at interval k for individual i, given observed information history, can be obtained by using the `predict()` function in R. These marginal and conditional probabilities are needed for the stabilized weighting function (Eq. 2.6). We pool these probabilities sequentially over k time points from each time-specific logistic model. Before a treatment discontinuation, the probability of being on treatment at $k \in \{1, ..., m-1\}$ is the cumulative product of living through each of these intervals without experiencing an exposure event of interest, that is, the cumulative product of $1 - \hat{p}_{ik}$. At the interval of $k = m$, the probability of stopping the treatment is

directly $\hat{p}_{im}$. When we consider the simple scenario that each counting process only counts to the first event, the probability of treatment termination remains unchanged after the time point of m. Similarly, we get $\hat{p}^*_{ik}$ from a marginal pooled logistic model $h_{A_i}(t_k;\kappa)$ coupled with (Eq. 2.13) (Hernan, et al., 2000):

$$\widehat{sw}_i^{\bar{A}}(t_m;\hat{\kappa},\hat{\eta}) = \begin{cases} \dfrac{\prod_{k=1}^{m-1} 1-\hat{p}^*_{ik}}{\prod_{k=1}^{m-1} 1-\hat{p}_{ik}}, & t < t_{A_i} \\ \dfrac{\hat{p}^*_{im}\cdot\prod_{k=1}^{m-1} 1-\hat{p}^*_{ik}}{\hat{p}_{im}\cdot\prod_{k=1}^{m-1} 1-\hat{p}_{ik}}, & t \geq t_{A_i} \end{cases} \tag{2.15}$$

2.2 Single treatment change models and weights

2.2.1 Notation for MPP

To achieve a closer approximation of our exposure process, instead of modelling the exposure process $A_i(t)$ directly, we can model the explicit visiting process $V_i(t)$ that induces a potential change in daily dose level $D_i(t)$, and thus also a potential change in the exposure status $A_i(t)$. At any given time, $A_i(t)$ is determined by the history of dispensations (visits) counting process $V_i(t) \in \{0, 1, ...\}$ and the corresponding dosage assignment (dose) $D_i(t) \in \{0, 1, 2, ...\}$, defined at the times when $\mathrm{d}V_i(t) = 1$. Together $(V_i(t), D_i(t))$ is a marked point process. Let Z_i represent a collection of baseline covariates, and $D_{i0} \in \{1, 2, ...\}$ as the initial dosage value. With MPP, the observed information up to time t for $\mathcal{F}_{t^-}$ and $\mathcal{F}^*_{t^-}$ contain the MPP $(V_i(t), D_i(t))$ and the baseline covariates Z_i and D_{i0}, except $A_i(t)$ is no longer a process but an exposure indicator. The causal assumptions remain unchanged as stated in Section. 2.1.1. In the single treatment change setting, we assume the process $V_i(t)$ only counts to the first visit time (the indicator of $V_i(t)$ takes value $0 \to 1$), and the exposure indicator $A_i(t)$ only documents the first potential stop of treatment (the indicator of $A_i(t)$ takes value $1 \to 0$), as introduced in the previous section.

2.2.2 Visit process exposure weights

Our visiting process models follow similar forms as (Eq. 2.8), (Eq. 2.10), and (Eq. 2.13) in Section. 2.1.2, but rather than treatment discontinuation $A_i(t) = 0$, we are modeling a visit initiation $V_i(t) = 1$. Additionally, all notations on $A_i(t)$ and t_{A_i} becomes $V_i(t)$ and t_{Vi}, whereas the marginal and conditional models are still parametrized with κ and η, respectively.

$$\log\left(h_{V_i}(t \mid X_i(t); \eta)\right) = \eta_0 + \eta_1 X_i(t) \tag{2.16}$$

$$h_{Vi}(t \mid X_i(t); \eta) = h_0(t) \exp\left\{\eta_1 X_i(t)\right\} \tag{2.17}$$

$$\operatorname{logit}\left(P(V_{ik} = 1 \mid X_{ik}; \eta)\right) = \eta_0 + \eta_1 X_{ik} \tag{2.18}$$

The corresponding stabilized weighting functions follow from (Eq. 2.9), (Eq. 2.11), and (Eq. 2.15), but with respect to history $\bar{V}$. The contribution of the weighting effect in an MPP can be isolated by fitting only a visit model or only a dosage model. Note that both the current visit process $V_i(t)$ and the exposure process $A_i(t)$ in Section.2.1.2 are counting processes.

2.2.3 Dosage level exposure weights

A multinomial or multiordinal logistic regression dosage level model can be used to model the dispensed daily dose categories $D_i(t)$; it is the dosage assignment model for our MPP.

Logistic Model

A simplified case of a multinomial logistic regression is a logistic regression, where everyone in the new user cohort starts at the same dose, but at some point t during the follow-up, the treatment can stop or a new dispensation takes place at the same dose, $d \in \{0, 1\}$. This logistic model can be specified similarly as (Eq. 2.13) and (Eq. 2.14) from the discrete-time exposure model, but parametrized with $\phi_d = (\phi_0, \phi_1)$ for the binary dosage model that takes $d = 0$ as the reference level:

$$P(D_i(t) = d \mid X_i(t); \phi_d) = \begin{cases} \dfrac{1}{1 + \exp\{\phi_0 + \phi_1 X_i(t)\}}, & d = 0 \\ \dfrac{\exp\{\phi_0 + \phi_1 X_i(t)\}}{1 + \exp\{\phi_0 + \phi_1 X_i(t)\}}, & d = 1 \end{cases}, \tag{2.19}$$

while the marginal version of (Eq. 2.19) excludes the confounder process $X_i(t)$. With this dosage model, everyone starts at the same dosage, so they are indicated as on treatment $A_i(0) = 1$ to start with. Therefore, before a visit/dispensation change, the stabilized weight of dosage D takes value 1 and does not have an impact on the overall exposure weight. On and after a visit, the weight of the dosage model becomes the ratio of the treatment probabilities from marginal and conditional models at the time t_{Vi}, which gives the below set of weighting function:

$$\widehat{sw}_i^{\bar{D}}(t; \hat{\phi}^*, \hat{\phi}) = \begin{cases} 1, & t < t_{Vi} \\ \dfrac{\hat{p}^{*d}_{t_{Vi}}}{\hat{p}^{d}_{t_{Vi}}} = \dfrac{P(D_i(t_{Vi}) = d; \hat{\phi}_0^*)}{P(D_i(t_{Vi}) = d \mid X_i(t_{Vi}); \hat{\phi}_d)}, & t \geq t_{Vi} \end{cases} \tag{2.20}$$

In discrete-time setting, t_{Vi} falls in the k^{th} interval, but in order to make a distinction with the $\hat{p}_{ik}$ in the pooled logistic model (Eq. 2.15), we name this dosage probaility as $\hat{p}^d_{ik}$ for later reference in Chapter 3. In addition, all notations for $X_i(t)$ is X_{ik}, $X_i(t_{Vi})$ notation is changed to X_{im} and similarly for process $D_i(t_{Vi})$ to D_{im}.

Multinomial Logistic Model

We can also extend the dosage from binary to multiple levels, where $D_i(t)$ takes value $d \in \{0, 1, 2\}$. Multiordinal model can be an alternative to the multinomial model, if we decide to make some additional assumptions, such as the levels in dosage effect $D_i(t)$ are equidistance. A patient can stop the treatment that takes $d = 0$ or can continue on treatment with dosage level that represents, for example, $d = 1$: (0, 5mg] and $d = 2$: (5, 20mg]. For more than two dosage levels, we can again choose level $d = 0$ as the reference group, with the model extension from (Eq. 2.19), we have:

$$P(D_i(t) = d \mid X_i(t); D_{i0}; \phi_d) = \begin{cases} \dfrac{1}{1 + \sum_{\ell \in \{1,2\}} \exp\{\phi_{0\ell} + \phi_{1\ell} X_i(t) + \phi_{d_0} D_{i0}\}}, & d = 0 \\ \dfrac{\exp\{\phi_{0d} + \phi_{1d} X_i(t) + \phi_{d_0} D_{i0}\}}{1 + \sum_{\ell \in \{1,2\}} \exp\{\phi_{0\ell} + \phi_{1\ell} X_i(t) + \phi_{d_0} D_{i0}\}}, & d \neq 0 \end{cases}, \tag{2.21}$$

where ϕ_{d_0} is the effect of initial dose on the decision of current dose, since we assume patients tend to stay on the same prescription. Let $\phi_d = (\phi_{0d}, \phi_{1d}, \phi_{d_0})$ be a shorthand notation. Similarly, the marginal model of (Eq. 2.21) can be fitted without the confounder process $X_i(t)$, so it is parametrized through $\phi_d^* = (\phi_{0d}, \phi_{d_0})$. We can fit a logistic model for the initial dosage $d_0 \in \{1, 2\}$ sampling, and a multinomial logistic model for after a new dispensation takes place $\mathrm{d}V_i(t) = 1$, where $d \in \{0, 1, 2\}$. The weighting function in (Eq. 2.20) then becomes:

$$\widehat{sw}_i^{\bar{D}}(t; \hat{\phi}^*, \hat{\phi}) = \begin{cases} 1, & t < t_{Vi} \\ \dfrac{\hat{p}^{*d}_{t_{Vi}}}{\hat{p}^{d}_{t_{Vi}}} = \dfrac{P(D_i(t_{Vi}) = d \mid D_{i0}; \hat{\phi}_d^*)}{P(D_i(t_{Vi}) = d \mid X_i(t_{Vi}); D_{i0}; \hat{\phi}_d)}, & t \geq t_{Vi} \end{cases} \tag{2.22}$$

The initial dose weights are not needed, because we adjust for d_0 in the outcome model, so the stabilized weight stays 1 before such event takes place.

2.2.4 Combined stabilized weights

The stabilized weighting function for exposure indicator $A_i(t)$ is a product of stabilized weights for $V_i(t)$ and $D_i(t)$, which can be expressed as:

$$\hat{sw}_i^{\bar{A}}(t) = \hat{sw}_i^{\bar{V}}(t) \times \hat{sw}_i^{\bar{D}}(t) \tag{2.23}$$

Consider the simple case that we use a logistic model for the dosage assignment model. With Poisson$^{\bar{V}}$-Logistic$^{\bar{D}}$ weighting, the combined weights can be expressed as a product of (Eq. 2.9) and (Eq. 2.20), where $\hat{\omega} = \left(\hat{\kappa}, \hat{\eta}, \hat{\phi}^*, \hat{\phi}\right)$ is a collection of parameters involved:

$$\widehat{sw}_i^{\bar{A}}(t;\hat{\omega}) = \begin{cases} \dfrac{\exp\left\{-\int_0^t \exp\left\{\hat{\kappa}_0\right\} \mathrm{d}u\right\}}{\exp\left\{-\int_0^t \exp\left\{\hat{\eta}_0 + \hat{\eta}_1 X_i(u)\right\} \mathrm{d}u\right\}} \times 1, & t < t_{Vi} \\ \dfrac{\exp\left\{-\int_0^{t_{Vi}} \exp\left\{\hat{\kappa}_0\right\} \mathrm{d}u\right\} \cdot \exp\left\{\hat{\kappa}_0\right\}}{\exp\left\{-\int_0^{t_{Vi}} \exp\left\{\hat{\eta}_0 + \hat{\eta}_1 X_i(u)\right\} \mathrm{d}u\right\} \cdot \exp\left\{\hat{\eta}_0 + \hat{\eta}_1 X_i(t_{Vi})\right\}} \times \dfrac{\hat{p}^{*d}_{t_{Vi}}}{\hat{p}^{d}_{t_{Vi}}}, & t \geq t_{Vi} \end{cases} \tag{2.24}$$

With Cox$^{\bar{V}}$-Logistic$^{\bar{D}}$ weighting, we can merge (Eq. 2.11) and (Eq. 2.20):

$$\begin{aligned} &\widehat{sw}_i^{\bar{A}}(t;\hat{\omega}) \\ &= \begin{cases} \dfrac{\exp\left\{-\hat{H}_0^*(t)\right\}}{\exp\left\{-\int_0^t \exp\left\{\hat{\eta}_1 X_i(u)\right\} \mathrm{d}\hat{H}_0(u)\right\}} \times 1, & t < t_{Vi} \\ \dfrac{\exp\left\{-\hat{H}_0^*(t_{Vi})\right\} \cdot \mathrm{d}\hat{H}_0^*(t_{Vi})}{\exp\left\{-\int_0^{t_{Vi}} \exp\left\{\hat{\eta}_1 X_i(u)\right\} \mathrm{d}\hat{H}_0(u)\right\} \cdot \exp\left\{\hat{\eta}_1 X_i(t_{Vi})\right\} \mathrm{d}\hat{H}_0(t_{Vi})} \times \dfrac{\hat{p}^{*d}_{t_{Vi}}}{\hat{p}^{d}_{t_{Vi}}}, & t \geq t_{Vi} \end{cases} \end{aligned} \tag{2.25}$$

With Pooled Logistic$^{\bar{V}}$-Logistic$^{\bar{D}}$ weighting, we have the product of (Eq. 2.15) and (Eq. 2.20):

$$\widehat{sw}_i^{\bar{A}}(t_m;\hat{\omega}) = \begin{cases} \dfrac{\prod_{k=1}^{m-1} 1 - \hat{p}^*_{ik}}{\prod_{k=1}^{m-1} 1 - \hat{p}_{ik}} \times 1, & t < t_{Vi} \\ \dfrac{\hat{p}^*_{im} \cdot \prod_{k=1}^{m-1} 1 - \hat{p}^*_{ik}}{\hat{p}_{im} \cdot \prod_{k=1}^{m-1} 1 - \hat{p}_{ik}} \times \dfrac{\hat{p}^{*d}_{im}}{\hat{p}^{d}_{im}}, & t \geq t_{Vi} \end{cases} \tag{2.26}$$

The similar combinations for a multinomial logistic dosage assignment $D_i(t)$ model can be specified by replacing (Eq. 2.20) with (Eq. 2.22).

2.3 Multiple visit MPP models and weights

Previously, we had the simplified case that the visiting process $V_i(t)$ only counts to 1. However, in the real world setting, a patient might stop, continue or change the daily dosage level of treatment several times sequentially throughout the follow-up period. Therefore, the MPP $(V_i(t), D_i(t))$ are defined at multiple time points, and we need to use Andersen and Gill generalized Cox model for recurrent exposure.

As introduced previously, the visiting process can be modeled continuously given the context of administrative data. However, given the large sample size in administrative data, the long-format data set split based on frequent person-times usually result in large data dimensions. As an alternative, the discrete pooled logistic model can be used to approximate the visiting process intensity in the continuous-time, but only at some equidistance time intervals, which can possibly result in a smaller long-format data size.

In the current context, "visit" is used to refer to a time of a potential treatment change, that is, a documented new dispensation or an inferred stop of treatment. The stop of treatment time is the start of the gap time between two treatment episodes. Everyone starts with their initial dosage that might span over multiple discrete time points k until a new subsequent visit happens. Given the lack of data on adherence and the treatment is crucial for the life-threatening condition, we can make a reasonable assumption that patients adhere to their prescriptions such that their subsequent dosages remain the same in between two visiting events.

We impose a Markov assumption for our MPP, that a patient's current visit intensity and dosage assignment only depends on the dosage at the previous one time point $D_{i,k-1}$ rather than the patient's entire history of dosage assignment $D_{i0}, D_{i1}, ..., D_{i,k-2}$. Based on the pooled logistic visiting process model (Eq. 2.18), we further include the baseline covariates Z_i, the cumulative dose of GC $\text{cum}(\bar{a}) = \sum_{s=0}^{k-1} a_s$ up to time $k-1$, and a term for continuous time effect k. The conditional probability of visit initiation for individual i is:

$$\text{logit}\, P(V_{ik} = 1 \mid \Omega; \eta) = \eta_0 + \eta_1 Z_i + \eta_2 X_{ik} + \sum_{\ell \in \Delta} \eta_{3\ell} \mathbf{1}\{D_{i,k-1} = \ell\} + \eta_4 \text{cum}(\bar{a}) + \eta_5 k, \quad (2.27)$$

where $P(V_{ik} = 1 \mid \Omega; \eta) = p_{ik}$, and Ω is a collection of the covariates on the right hand side of the equation. Note that A_i is no longer a counting process, so $\bar{a}$ denotes the dose assignment history up to time k. We use the lowest dose level $d = 1$ as the reference so $\Delta = D^{\otimes} = \{0, 2, ..., 5\}$.

We also fit the D_{ik} models coupled with our V_{ik} model. D_{ik} follows the same form as

(Eq. 2.21), except with the same extended set of covariates Ω in the visit model:

$$p_{ik}^{d} = P(D_{ik} = d \mid \Omega; \phi_d) = \begin{cases} \dfrac{1}{1 + \sum_{\ell \in \Delta} \exp\{\Upsilon\}}, & d = c \\ \dfrac{\exp\{\Upsilon_{\ell=d}\}}{1 + \sum_{\ell \in \Delta} \exp\{\Upsilon\}}, & d \neq c \end{cases}, \tag{2.28}$$

where Υ has the same form as the right hand side of (Eq. 2.27) but parametrized with ϕ, and c is the reference level not in Δ. However, although both $D_{i,k-1}$ and D_{ik} have the same number of levels, where $d \in \{0, ..., 5\}$, we have a violation of positivity assumption by definition at the $D_{i,k-1} = 0$ to $D_{ik} = 0$ transition, because the dosage model is defined at a visit, and a patient cannot visit for no drug. Therefore, we separate the modeling of D_{ik} based on the previous dose level. If a patient has been unexposed at the previous time point, we model with the first equation of (Eq. 2.29), where the current dose D_{ik} can take values other than $d = 0$. Otherwise, if a patient has been exposed with some dose level at the previous time point, the current dose can take any value of d.

$$\Upsilon = \begin{cases} \sum_{\ell \in D^{\odot}} \phi_{0\ell} + \phi_1 Z_i + \phi_2 X_{ik} + \phi_4 \text{cum}(\bar{a}) + \phi_5 k, & D_{i,k-1} = 0 \\ \sum_{\ell \in D^{\oplus}} \phi_{0\ell} + \phi_1 Z_i + \phi_2 X_{ik} + \sum_{\ell \in D^{\oplus}} \phi_{3\ell} \mathbf{1}\{D_{i,k-1} = \ell\} + \phi_4 \text{cum}(\bar{a}) + \phi_5 k, & D_{i,k-1} > 0 \end{cases}, \tag{2.29}$$

where $D^{\odot} = \{2, ..., 5\}$ with $d = 1$ as the reference level and $D^{\oplus} = \{1, ..., 5\}$ with $d = 0$ as the reference level. Alternatively, the multiordinal logistic regression has much fewer parameters compare to multinomial logistic regression. For the consideration of small covariate categories, we can switch the D_{ik} model to below:

$$\text{logit}\, P(D_{ik} \leq d \mid \Omega; \phi) = \log \left[\frac{P(D_{ik} \leq d \mid \Omega; \phi)}{1 - P(D_{ik} \leq d \mid \Omega; \phi)} \right] = \Upsilon \tag{2.30}$$

Unlike previous sections, where the effect of baseline (time-fixed) confounders on the initial dosage assignment is not modeled, we now introduce another multinomial logistic model for

the initial dosage D_{i0}:

$$p_{i0}^{d_0} = P(D_{i0} = d_0 \mid Z_i; \xi_{d_0}) = \begin{cases} \dfrac{1}{1 + \sum_{\ell_0 \in D^{\odot}} \exp\{\xi_{0\ell_0} + \xi_1 Z_i\}}, & d_0 = 1 \\ \dfrac{\exp\{\xi_{0d_0} + \xi_1 Z_i\}}{1 + \sum_{\ell_0 \in D^{\odot}} \exp\{\xi_{0\ell_0} + \xi_1 Z_i\}}, & d_0 \neq 1 \end{cases}, \tag{2.31}$$

where $\xi_{d_0} = (\xi_{0d_0}, \xi_{1d_0})$. Since we are modeling a new user cohort, no patient can start with unexposed, so D_{i0} model (Eq. 2.31) has one less category such that $d_0 \neq 0$. We can again specify the lowest dose level $d_0 = 1$ as the reference level. Similarly, (Eq. 2.31) is exchangeable with a multiordinal logistic model for the initial dose D_{i0} model.

Previously in our stabilized weights, we have the time points $k \in \{1, 2, ..., m, ..., K\}$. We can now bring in the probability of dosage assignment at $k = 0$ modeled through (Eq. 2.31) to the combined weighting function for multiple MPP:

$$\begin{aligned} \widehat{sw}_i^{\bar{A}}(t_m; \hat{\omega}) &= \widehat{sw}_i^{\bar{D}_0}(t_m; \hat{\xi}^*, \hat{\xi}) \times \widehat{sw}_i^{\bar{V}}(t_m; \hat{\kappa}, \hat{\eta}) \times \widehat{sw}_i^{\bar{D}}(t_m; \hat{\phi}^*, \hat{\phi}) \\ &= \frac{\hat{p}_{i0}^{*d_0}}{\hat{p}_{i0}^{d_0}} \cdot \frac{\prod_{k=1}^{m} (1 - \hat{p}_{ik}^*)^{\mathbf{1}\{V_{ik}=0\}} \cdot (\hat{p}_{ik}^*)^{\mathbf{1}\{V_{ik}=1\}}}{\prod_{k=1}^{m} (1 - \hat{p}_{ik})^{\mathbf{1}\{V_{ik}=0\}} \cdot (\hat{p}_{ik})^{\mathbf{1}\{V_{ik}=1\}}} \cdot \left(\frac{\hat{p}_{ik}^{*d}}{\hat{p}_{ik}^{d}}\right)^{\mathbf{1}\{V_{ik}=1\}}, \end{aligned} \tag{2.32}$$

where $\hat{\omega} = \left(\hat{\xi}^*, \hat{\xi}, \hat{\kappa}, \hat{\eta}, \hat{\phi}^*, \hat{\phi}\right)$ is a collection of marginal and conditional parameters in the three models for D_{i0}, V_{ik}, D_{ik}. Because the two D-regression models above are based on separate data sets and thus are mutually exclusive, the weights derived from them are combined together as $\widehat{sw}_i^{\bar{D}}(t_k; \hat{\phi}^*, \hat{\phi})$. At different time points, depending on the visiting status, some parts of the stabilized weights are equal to one in the cumulative products of visiting model and the corresponding dosage model, but the baseline components always exist.

2.4 Estimation of flexible hazard models for exposure-outcome relationships

Our marginal hazard model is on continuous-time scale. The same model (Eq. 2.33) can be fitted by both weighted Cox partial likelihood and the weighted case-base sampling partial likelihood. The log-baseline hazard function $\theta_0(t)$ is left unspecified in Cox model, but it is fitted through a spline function in the case-base estimation. Consider the case where we are modeling the marginal effect, parametrized with θ_1, of not being on a treatment, i.e. $A_i(t) = 0$ or $1 - A_i(t) = 1$. The marginal hazard is a function of time and the effect may not be proportional, and thus we can use the time interaction of treatment discontinuation and time since stop of the treatment, parametrize with θ_2, to model the dependency on time. Thus, the outcome model for (Eq. 2.3) can be extended as:

$$\psi_i(t;\theta) = \exp\left\{\theta_0(t) + \theta_1\left(1 - A_i(t)\right) + \theta_2\left(1 - A_i(t)\right)\left(t - t_{A_i}\right)\right\} \tag{2.33}$$

2.4.1 Weighted Cox partial likelihood

The standard Cox partial likelihood for our outcome model has the form:

$$L(\theta) = \prod_{i=1}^{n}\prod_{u\in[0,t)}\left(\frac{\exp(\theta_1 A_i(u) + \theta_2 A_i(u)(u - t_{A_i}))}{\sum_{l=1}^{n} G_l(u)\exp(\theta_1 A_l(u) + \theta_2 A_l(u)(u - t_{A_i}))}\right)^{\mathrm{d}Y_i(u)} \tag{2.34}$$

Given the stabilized weights $sw_i(t)$ derived from exposure models in the previous section, the weighted Cox partial likelihood is:

$$\tilde{L}(\theta) = \prod_{i=1}^{n}\prod_{u\in[0,t)}\left(\frac{sw_i(u)\exp\{\theta A_i(u) + \theta_2 A_i(u)(u - t_{A_i})\}}{\sum_{l=1}^{n} G_l(u) sw_l(u)\exp\{\theta A_l(u) + \theta_2 A_l(u)(u - t_{A_i})\}}\right)^{sw_i(u)\mathrm{d}Y_i(u)} \tag{2.35}$$

The weighted partial likelihood does not have the usual likelihood properties (equivalence of score covariance and information), so the naive standard errors based on inverting the observed information matrix are not valid. Therefore, we use a robust sandwich estimator for the variance of θ by specifying the '`cluster`' variable to each patient ID in the `coxph()` function in R. Internally, this step is achieved by using the grouped jacknife techniques (Therneau and Grambsch, 2000). We obtain $\hat{\theta}_{-i}$ as the maximum likelihood estimator (MLE) excluding all person-time observations of person i. J is a matrix, where in its i^{th} row, we document the difference between each $\hat{\theta}_{-i}$ and the MLE estimator $\hat{\theta}$. Let $\bar{J}$ be the column means of J. The estimated variance for our parameters of interest $\theta = (\theta_1, \theta_2)$ can be calculated as (Therneau

and Grambsch, 2000):

$$\widehat{Var}(\hat{\theta}) = \frac{n-1}{n}(J - \bar{J})'(J - \bar{J})$$

2.4.2 Weighted case-base sampling partial likelihood

In the case-base sampling method by Hanley and Miettinen (2009), the follow-up experience is considered a collection of person-time coordinates (i, t). A finite number of these can be sampled for the analysis purposes, with negligible loss of information. We keep all person-times where an event happened $\mathrm{d}Y_i(t) = 1$, and pair this 'case series' with a random sample of person-times from all follow-up experience as a 'base series'. The implementation details for base series sampling follows from the two-step mechanism in Hanley and Miettinen (2009): depending on how much a person contributes to the person-time, the number of base time points a patient has is sampled from a multinomial distribution; afterwards, the base time points are sampled uniformly within the patient's follow-up period. Based on the simulation results of Saarela (2016), a base series coefficient with a value larger than 100 is shown to have no lost in efficiency. In current thesis, we define the size of the base series as 200 times the number of outcome events in the case series to accommodate for the large sample size.

Let the sampling mechanism for base series be a Poisson process $R_i(t) \in \{0, 1, ...\}$. Let $\rho_i(t)$ be the intensity function of $R_i(t)$:

$$\rho_i(t)\mathrm{d}t = P^{\mathcal{O}}(\mathrm{d}R_i(t) = 1 \mid \mathcal{F}_{t^-}) = P^{\mathcal{O}}(\mathrm{d}R_i(t) = 1 \mid G_i(t))$$

We take $\rho_i(t) = \rho_i^{\mathcal{O}}(t) = \rho_i^{\mathcal{E}}(t)$ to be consistent with the stability assumption. We define $Q_i(t) = Y_i(t) + R_i(t)$ as the counting process for all sampled person-moments contributed by individual i. Then the intensity function for $Q_i(t)$ can be expressed as the sum of intensities from case and base series:

$$P^{\mathcal{E}}(\mathrm{d}Q_i(t) = 1 \mid \mathcal{F}^*_{t^-}) = (\psi_i(t;\theta) + \rho_i(t))\,\mathrm{d}t$$

and the conditional probability follows:

$$\begin{aligned} &P^{\mathcal{E}}(\mathrm{d}Y_i(t) = 1 \mid \mathrm{d}Q_i(t) = 1, \mathcal{F}^*_{t^-}) \\ &\quad = \frac{P^{\mathcal{E}}(\mathrm{d}Y_i(t) = 1 \mid \mathcal{F}^*_{t^-})}{P^{\mathcal{E}}(\mathrm{d}Q_i(t) = 1 \mid \mathcal{F}^*_{t^-})} = \frac{\psi_i(t;\theta)}{\psi_i(t;\theta) + \rho_i(t)} \end{aligned}$$

The above function is crucial for the derivation of the weighted partial likelihood for our marginal hazard function. An interpretation would be: if a person-time was sampled, then the probability of it being an outcome event is the above ratio. As shown in Saarela (2016),

the two counting processes under $\mathcal{E}$ has Doob-Meyer decomposition:

$$\begin{aligned} \mathrm{d}Y_i(t) &= \psi_i^{\mathcal{E}}(t;\theta)\mathrm{d}t + \mathrm{d}M_i(t) \\ \mathrm{d}Q_i(t) &= \mathrm{d}Y_i(t) + \mathrm{d}R_i(t) = \psi_i^{\mathcal{E}}(t;\theta)\mathrm{d}t + \rho_i(t)\mathrm{d}t + \mathrm{d}M_i(t) + \mathrm{d}M_i^{'}(t) \end{aligned}$$

where $M_i(t)$ and $M_i^{'}(t)$ are orthogonal martingales adapted to $\mathcal{F}_{t^-}^*$, with $E^{\mathcal{E}}(\mathrm{d}M_i(t) \mid \mathcal{F}_{t^-}^*) = 0$. However, due to confounding, we don't have the same martingale property for $E^{\mathcal{O}}(\mathrm{d}M_i(t) \mid \mathcal{F}_{t^-}^*)$. Similarly to Cox partial likelihood, exposure weights can be introduced to remove the confounding.

The partial likelihood for case-base sampling is the joint probability contributions of each individual:

$$L(\theta) = \prod_{i=1}^{n} \prod_{u\in[0,t)} \left(\frac{\psi_i(u;\theta)^{\mathrm{d}Y_i(u)}}{\psi_i(u;\theta) + \rho_i(u)} \right)^{\mathrm{d}Q_i(u)} \tag{2.36}$$

This is of the logistic regression form with an offset term $\log(1/\rho_i(t))$. With the stabilized weights $sw_i(t)$ from the previous section, we now have the weighted partial likelihood for case-base sampling:

$$\tilde{L}(\theta) = \prod_{i=1}^{n} \prod_{u\in[0,t)} \left(\frac{\psi_i(u;\theta)^{\mathrm{d}Y_i(u)}}{\psi_i(u;\theta) + \rho_i(u)} \right)^{sw_i(u)\mathrm{d}Q_i(u)} \tag{2.37}$$

In order to maximize the weighted case-base sampling partial likelihood, we fit the marginal hazard model with a logistic regression form using `glm()`, adjusting for an offset term for time. Score function is the first order derivative of the natural log of a likelihood function (Therneau and Grambsch, 2000). In case-base sampling model, all outcome events are sampled, such that $\mathrm{d}Y_i(t)\mathrm{d}Q_i(t) = \mathrm{d}Y_i(t)$. The pseudo-score process corresponding to (Equation. 2.37) is given by $U(t;\theta) = \sum_{i=1}^{n} U_i(t;\theta)$, where

$$\begin{aligned} U_i(t;\theta) = &\int_0^t sw_i(u) \frac{\partial}{\partial\theta} \log \psi_i(u;\theta)\mathrm{d}Y_i(u) \\ &- \int_0^t sw_i(u) \frac{\partial}{\partial\theta} \log[\psi_i(u;\theta) + \rho_i(u)]\mathrm{d}Q_i(u), \end{aligned} \tag{2.38}$$

with the observed pseudo-information process given by $J(t;\theta) = \sum_{i=1}^{n} J_i(t;\theta)$, where

$$\begin{aligned} J_i(t;\theta) = &- \int_0^t sw_i(u) \frac{\partial^2}{\partial\theta\partial\theta'} \log \psi_i(u;\theta)\mathrm{d}Y_i(u) \\ &+ \int_0^t sw_i(u) \frac{\partial^2}{\partial\theta\partial\theta'} \log[\psi_i(u;\theta) + \rho_i(u)]\mathrm{d}Q_i(u), \end{aligned} \tag{2.39}$$

We denote $U(t;\theta) = U(\theta)$ and $J(t;\theta) = J(\theta)$. Although $U(\theta)$ does not have a direct martingale representation as the unweighted one shown in Saarela (2016), the score process with the likelihood ratio weights is asymptotically unbiased (Roysland et al., 2011). The unweighted partial likelihood (Equation. 2.36) has the information equality property, where the variance estimation can be achieved by inverting the observed information matrix at the maximum likelihood point. However, the weighted version (Equation. 2.37) does not have this property, and we need to use a sandwich estimator:

$$\mathrm{cov}[\hat{\theta}] \approx J(\theta_0)^{-1}\mathrm{cov}[U(\theta_0)]J(\theta_0)^{-1},$$

where the pseudo-score covariance is estimated by the empirical variance:

$$\mathrm{cov}[U(\theta_0)] \approx \sum_{i=1}^{n} U_i(\hat{\theta})' U_i(\theta) \tag{2.40}$$

and the observed information by $J(\theta_0) \approx J(\hat{\theta})$. This sandwich form is easily obtained as we have I.I.D patients in our study cohort. However, in the long-format (or counting process format), each individual contributes a cluster of person-times. Therefore, the covariance in (Equation. 2.40) is first calculated within individual and then summed across individual. In practice, the cluster sandwich estimator for generalized linear models is implemented in the `infjack.glm` R function (Lumley and Heagerty, 1999), or the sandwich package (Zeileis, 2006; Zeileis et al., 2020).

Chapter 3

Simulation Study

To assess the performance of different IPT weighting and MSM combinations presented in the previous chapter, we conduct four simulation studies. In Section 3.1, we present the conditional data generation mechanisms needed for all of our counting processes and variables. In Section 3.2, we use DAG to illustrate different simulation scenarios, and we then incorporate conditional data generation mechanisms to multistate diagrams with the corresponding transition intensity matrices. In Section 3.3 we present the log marginal hazard rates graphs and simulation result tables.

We expect to see that under dynamic treatment, the marginal hazard models have complex functional forms, so we allow the treatment and time interaction term. We aim to show that the weighted models should have less bias for estimating the treatment effect compared to the unweighted model. When we separate the exposure process into the visit process and dose level, the combined weights should result in the least bias. Furthermore, we expect case-base sampling MSM to produce similar results as Cox MSM, but case-base sampling requires much less run time due to smaller data set sizes.

3.1 Simulation algorithms

Havercroft and Didelez (2012) proposed generating data from a known Cox MSM and a known treatment assignment model. Young and Tchetgen Tchetgen (2014) suggested that we can derive the true parameter values of a Cox MSM from observed data by using the connection of Cox MSM and g-formula. With our data generation mechanism adapted from these previous works, we ensure the exposure models and outcome MSM models are correctly specified. However, these authors assumed discrete-time models for the time-dependent treatment and covariate processes, while only allowing the outcome model to be on a continuous-time scale. Although the pooled logistic model is one of our exposure models in the previous chapter,

the discretization is indeed a result of manually splitting the data from a continuous-time scale, whereas our data generation is based on a continuous-time model. In the simulation, we focus on a simplified setting where all the counting processes only count to the first event, as in this case the true marginal hazard can be calculated as a function of the conditional data generating mechanism. Therefore, in this setting, the marginal hazard function can be expressed as Robin's g-formula:

$$\psi_i(t;\theta) = \frac{\int \lambda^{\mathcal{E}}_{Y_i}(t) \prod_{u\in[0,t)} \lambda^{\mathcal{E}}_{X_i}(u)^{\mathrm{d}x_i(u)} \exp\left\{-\int_0^t \left[\lambda^{\mathcal{E}}_{X_i}(u) + \lambda^{\mathcal{E}}_{Y_i}(u)\right] \mathrm{d}u\right\}}{\int \prod_{u\in[0,u)} \lambda^{\mathcal{E}}_{X_i}(u)^{\mathrm{d}x_i(u)} \exp\left\{-\int_0^t \left[\lambda^{\mathcal{E}}_{X_i}(u) + \lambda^{\mathcal{E}}_{Y_i}(u)\right] \mathrm{d}u\right\}}, \tag{3.1}$$

where we integrate over the density of the time when the TD confounder process jumps.

In fact, (Eq. 3.1) would have a closed-form solution only in special cases, such as when the outcome process only depends on the current value of a TD confounder process, which makes the expression memoryless (Young and Tchetgen Tchetgen, 2014). However, the above memoryless assumption might not be reasonable. Moreover, in the scenario tailored to our case, even under the special case of only counting to the first event for both the outcome and the confounder processes, (Eq. 3.1) does not have a closed-form solution. As an illustration, we let the confounder process $X_i(t) \in \{0,1\}$ only count to the first event, and let the conditional mortality hazard function depends only on $X_i(t^-)$. Then, we define the conditional mortality hazard as $\lambda^{\mathcal{E}}_{Y_i}(t; x_i(u^-))$. The denominator of (Eq. 3.1) can be expressed as:

$$\begin{aligned}
&\int \prod_{u\in[0,t)} \lambda^{\mathcal{E}}_{X_i}(u)^{\mathrm{d}x_i(u)} \exp\left\{-\int_0^t \left[\lambda^{\mathcal{E}}_{X_i}(u) + \lambda^{\mathcal{E}}_{Y_i}(u; x_i(u^-))\right] \mathrm{d}u\right\} \\
&\quad = \exp\left\{-\int_0^t \lambda^{\mathcal{E}}_{Y_i}(u;0)\mathrm{d}u\right\} \exp\left\{-\int_0^t \lambda^{\mathcal{E}}_{X_i}(u)\mathrm{d}u\right\} \\
&\qquad + \int_0^t \left[\exp\left\{-\int_0^v \lambda^{\mathcal{E}}_{Y_i}(u;0)\mathrm{d}u - \int_v^t \lambda^{\mathcal{E}}_{Y_i}(u;1)\mathrm{d}u\right\} \times \lambda^{\mathcal{E}}_{X_i}(v) \exp\left\{-\int_0^v \lambda^{\mathcal{E}}_{X_i}(u)\mathrm{d}u\right\}\right] \mathrm{d}v,
\end{aligned}$$

Similarly, the numerator of (Eq. 3.1) can be expressed as:

$$\begin{aligned}
&\int \lambda^{\mathcal{E}}_{Y_i}(t; x_i(t^-)) \prod_{u\in[0,t)} \lambda^{\mathcal{E}}_{X_i}(u)^{\mathrm{d}x_i(u)} \exp\left\{-\int_0^t \left[\lambda^{\mathcal{E}}_{X_i}(u) + \lambda^{\mathcal{E}}_{Y_i}(u; x_i(u^-))\right] \mathrm{d}u\right\} \\
&\quad = \lambda^{\mathcal{E}}_{Y_i}(t;0) \exp\left\{-\int_0^t \lambda^{\mathcal{E}}_{Y_i}(u;0)\mathrm{d}u\right\} \exp\left\{-\int_0^t \lambda^{\mathcal{E}}_{X_i}(u)\mathrm{d}u\right\} \\
&\qquad + \lambda^{\mathcal{E}}_{Y_i}(t;1) \int_0^t \left[\exp\left\{-\int_0^v \lambda^{\mathcal{E}}_{Y_i}(u;0)\mathrm{d}u - \int_v^t \lambda^{\mathcal{E}}_{Y_i}(u;1)\mathrm{d}u\right\} \times \lambda^{\mathcal{E}}_{X_i}(v) \exp\left\{-\int_0^v \lambda^{\mathcal{E}}_{X_i}(u)\mathrm{d}u\right\}\right] \mathrm{d}v,
\end{aligned}$$

where the above numerator and denominator do not have a closed form solution.

Therefore, we suggest using the below conditional data generation mechanisms, and we derive the true treatment effect of (Eq. 3.1) using numerical integration (Saarela and Liu, 2016). We assume the intercept term to be constant over time, and instead of modeling the exponentiated intercept parameters of $\alpha'_Y(t)$ and $\alpha'_X(t)$, we simplify the notation to $\exp(\alpha'_Y(t)) = \exp(\alpha'_Y) = \alpha_Y$ and $\exp(\alpha'_X(t)) = \exp(\alpha'_X) = \alpha_X$.

$$\lambda^{\mathcal{E}}_{Y_i}(t) = \alpha_Y \cdot \exp\left\{\beta_Y A_i(t) + \gamma_Y X_i(t^-)\right\} \tag{3.2}$$

$$\lambda^{\mathcal{E}}_{X_i}(t) = \alpha_X \cdot \exp\left\{\beta_X A_i(t)\right\} \tag{3.3}$$

Combined with causal assumptions of $\lambda^{\mathcal{E}}_{Y_i}(t) = \lambda^{\mathcal{O}}_{Y_i}(t)$ and $\lambda^{\mathcal{E}}_{X_i}(t) = \lambda^{\mathcal{O}}_{X_i}(t)$. The parameters α_Y, β_Y, γ_Y, α_X and β_X are presented in Table 3.4 depending on the simulation scenario. Then we can model the marginal hazard under: (1) being treated or (2) not being treated. The resulting marginal log-hazard for (1) and (2) are plotted in Figure 3.5 and Figure 3.6 respectively. For a multinomial logistic model we have the extension in Figure 3.7. In our next section about simulation design, (1) is estimated with simulation 1, (2) is estimated with simulation 2 and simulation 3, and the extension on multinomial model is through simulation 4. Because our clinical context is a new-user cohort design, we put more focus on the case that everyone starts exposed, but can stop or change treatment during the follow-up period.

We complete our conditional data generating mechanism by specifying the exposure mechanism under $\mathcal{O}$:

$$\lambda^{\mathcal{O}}_{A_i}(t) = \alpha_A \cdot \exp\left\{\beta_A X_i(t^-)\right\} \tag{3.4}$$

similarly we define $\exp(\alpha'_A(t)) = \exp(\alpha'_A) = \alpha_A$ and $\lambda^{\mathcal{O}}_{A_i}(t) = \lambda^{\mathcal{E}}_{A_i}(t)$. The data generating mechanism now has the feature of treatment-confounder feedback, where the exposure process is influenced by the current value of the confounder process, which in itself is modified by exposure. We assume type I censoring due to the end of the follow-up period and did not simulate baseline covariates (although these can be easily incorporated in all the processes). The target marginal treatment intensity $\lambda^{\mathcal{E}}_{A_i}(t;\kappa)$ can be expressed as a function of (Eq. 3.2), (Eq. 3.3), and (Eq. 3.4), obtained by integrating out $X_i(t)$. In practice, we approximate this by fitting a marginal model $\lambda^{\mathcal{E}}_{A_i}(t;\kappa) = \exp\{\kappa_0\}$, resulting in the counterparts needed for stabilized weights in discrete-time setting. The parameters α_A and β_A are presented in Table 3.4 with respect to different simulation scenarios. If we separate the modeling of $A_i(t)$ as $(V_i(t), D_i(t))$, we also provide parameters used for the dosage assignment models for $D_i(t)$. In all our simulation scenarios, we assume our exposure/visit, confounder, and outcome processes only count to the their first event.

3.2 Simulation design

We use the discrete-time DAG Figure 3.1 to present our simulation scenarios, while the actual data generation was done using multistate model for the continuous-time setting.

(a) Simulation 1 and 2 DAG (b) Simulation 3 DAG (c) Simulation 4 DAG

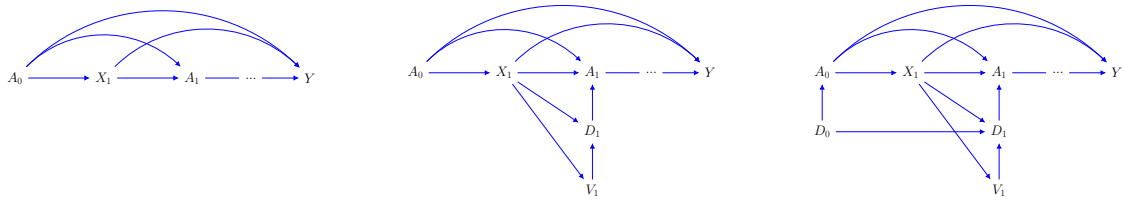


Figure 3.1: Simulation designs with DAG representation

1. **Simulation 1** Figure 3.1a:
   A scenario that patients all start unexposed, and then start the treatment at some individual-specific time point. $A_i(t)$ can be interpreted as a counting process for treatment initiation. Note that baseline covariates Z are omitted from the simulation studies.

2. **Simulation 2** Figure 3.1a:
   In scenario 2, we simulate a reversed situation of scenario 1. With the new-user design described in Section 1.1.3, we have $A_i(0) = a_{i0} = 1$ so that everyone starts exposed, and then a change in visit status results in a stop of treatment. $A_i(t)$ is a complement of a counting process that counts to the treatment discontinuation event.

3. **Simulation 3** Figure 3.1b:
   Similar to the second scenario, all patients start with being on treatment, but we separate the modeling of exposure process $A_i(t)$ into a visit process $V_i(t)$ that induces the corresponding dosage assignment indicator $D_i(t)$, where the dosage level is binary, indicating treatment stopping or continuation on the same treatment.

4. **Simulation 4** Figure 3.1c:
   Similar to the second and third scenario, all patients start on treatment, with some initial dose D_{i0} sampled from a logistic model. We also separate the modelling of $A_i(t)$ into $V_i(t)$ and $D_i(t)$, but we extend the dosage level to three categories, indicating stop treatment, or continue on the same treatment or change to a higher dose treatment. We use a multinomial model for the modeling of $D_i(t)$.

As a summary of Section 2.1 and Section 2.2, the different combinations of exposure and outcome models can be specified as the below cross tabulations. When we model the exposure

status $A_i(t)$ directly in simulation 1 and 2, we use model specifications in Table 3.1. On the other hand, when we separate the modeling of $A_i(t)$ into a MPP $(V_i(t), D_i(t))$ in simulation 3 and 4, we can i) examine the contribution of each model in the MPP separately, by using the model specifications when fitting a visit model alone Table 3.1 or a dosage assignment model alone Table 3.2, ii) use the combined weights of MPP, with the model specifications listed in Table 3.3.

Table 3.1: Model specifications when weighted with an exposure process or a visiting process model alone

Fracture Outcome Model \ Exposure Model		Continuous Time		Discrete Time	(Either)
		Poisson	Cox	Pooled Logistic	Unweighted
Outcome Models	Cox	(I)	(II)	(III)	(IV)
	Case-base	(V)	(VI)	(VII)	(VIII)

Table 3.2: Model specifications when weighted with a dosage assignment model alone

Fracture Outcome Model \ Dosage Assignment Model		Continuous Time	Discrete Time	(Either)
		Logistic		Unweighted
Outcome Models	Cox	(IX)	(X)	(IV)
	Case-base	(XI)	(XII)	(VIII)

Table 3.3: Model specifications when weighted with the combination of a visiting process model and a dosage assignment model

Fracture Outcome Model \ Dosage Assignment Model \ Visit Model		Continuous Time		Discrete Time	(Either)
		Poisson	Cox	Pooled Logistic	Unweighted
		Logistic			Unweighted
Outcome Models	Cox	(I†)	(II†)	(III†)	(IV†)
	Case-base	(V†)	(VI†)	(VII†)	(VIII†)

We specify the parameters used in our data generating mechanism under different scenarios as Table 3.4, where θ is the true value of our parameter of interest in the marginal hazard model. We then introduce our scenario-specific multistate diagrams, transition matrix, exposure and outcome models.

Table 3.4: Theoretical parameter values used in four simulation scenarios

Data Generation	Parameters	Simulation 1	Simulation 2	Simulation 3	Simulation 4
		$\theta_1 = -1$	$\theta_1 = 1$	$\theta_1 = 1$	$\theta_1 = 1$
					$\theta_2 = -1$
		(Eq. 3.5)	(Eq. 2.33)	(Eq. 2.33)	(Eq. 3.7)
Conditional Mechanism	α_X	0.05	0.05	0.04	0.04
	$\beta_{X_d}^{\S}$	-1.5	1.5	1.5	1.5
	$\alpha_A^{\Diamond}$	0.1	0.2	0.08	0.08
	$\beta_A^{\Diamond}$	1.5	1.5	1.5	1.5
	α_Y	0.02	0.02	0.02	0.02
	$\beta_{Y_d}^{\S}$	-1	-1	-1	-1 -2
	γ_Y	1.5	1.5	1.5	1.5
Binomial Dose $d \in \{0,1\}^{\ddagger,\triangleright}$	ϕ_0 ϕ_1	-	-	-0.38 -1	0
Multinomial Dose	ϕ_{0_1} ϕ_{1_1}	-	-	-	-1.5 1
$d \in \{0,1,2\}^{\triangleright}$	ϕ_{0_2} ϕ_{1_2}				-2 0

§ β_{X_d}, β_{Y_d} are β_X, β_Y for simulation 1 and 2. β_{X_d}, β_{Y_d} has $d \in \{0,1\}$ in simulation 3 and $d \in \{0,1,2\}$ in simulation 4, in which $\beta_{X_1} = \beta_{X_2} = \beta_{X_d}$. In both simulation 3 and 4, $\beta_{X_0} = \beta_{Y_0} = 0$.

$^{\Diamond}$ In simulation 3 and 4, α_A, β_A are α_V, β_V.

‡ In simulation 4, rather than after visit dose d sampling, we use this for $d_0 \in \{1,2\}$ sampling, (ϕ_0, ϕ_1) is now ϕ_2. The effect of d_0 on multinomial regression for d is $\phi_{d_0} = 0.5$.

$^{\triangleright}$ $d = 0$ is the reference level for data generation, while $d = 1$ is the reference level for parameter estimation.

Each simulated cohort has $n = 1,000$ patients, with event prevalence rate around 10% and an administrative censoring time of $\tau = 5$ years. We repeat the Monte Carlo simulation experiment $1,000$ times in each scenario. For all simulation scenarios, the continuous-time exposure data for person i is constructed by taking all person-time records until the exposure indicator changes or until the administrative censoring if process $A_i(t)$ never happens. The discrete-time exposure data set further split the follow-up period with 30-day intervals. For all simulation scenarios, the continuous-time outcome data set needed for the weighted Cox partial likelihood is constructed by taking all the person-times in the population, while we sample a much smaller data set with a base series coefficient of 200 in case-base sampling. For example, with 100 events out of a $1,000$ patients cohort, we have a long-format data size of $100 + 100 \times 200 = 20,100$ rows for case-base versus around $95,000$ rows for Cox.

We then detail each simulation scenario by presenting a multistate diagram, the corresponding data generation equations, and the exposure and outcome models used.

Simulation 1

In the first scenario, everyone starts unexposed and confounder value at 0 at time 0. A multistate model for simulation 1 is shown in Figure 3.2. Everyone starts at state 1 and then move to other states depending on the order of the treatment, confounder and fracture outcome counting processes they encounter. For example, a patient can initiate a treatment process $A_i(t)$ and then initiate another confounder drug represented by process $X_i(t)$, which makes them take the path of states $1 \rightarrow 2 \rightarrow 5$. Once a patient encounters a fracture process $Y_i(t)$, we stop the follow-up and call the resulting nodes *end states* (state 4, 6, 7 and 8). Administrative censoring can take place before a patient reaches one of the below states, and we denote it as an invisible state 0. The simulation algorithm works by drawing the time of the next event from exponential distribution, and then the next state is randomly chosen within the row of the current state in the transition matrix, with probabilities of each row intensity entries normalized with the current row sum of intensities. For example, a patient currently in state 2 can land in state 5 or state 6 with probabilities of $(\frac{\lambda_{25}}{\lambda_{25}+\lambda_{26}}, \frac{\lambda_{26}}{\lambda_{25}+\lambda_{26}})$, where $\lambda_{25} = \alpha_X \cdot e^{\beta_X} = 0.05 \cdot e^{-1.5} = 0.011$ and $\lambda_{26} = \alpha_Y \cdot e^{\beta_Y} = 0.02 \cdot e^{-1} = 0.007$.

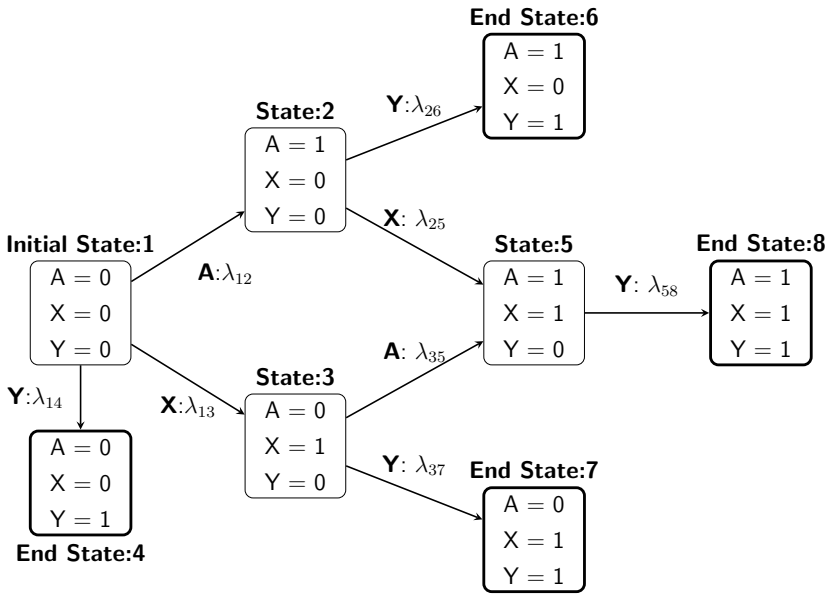


Figure 3.2: Simulation 1 multistate model for data generation, patients start unexposed with no confounder

The outcome, confounder and treatment counting processes are defined as (Eq. 3.2), (Eq. 3.3), and (Eq. 3.4). The corresponding transition intensity matrix for the first simulation scenario is:

$$TIM_1 = \begin{bmatrix} . & \lambda_{12} & \lambda_{13} & \lambda_{14} & . & . & . & . \\ . & . & . & . & \lambda_{25} & \lambda_{26} & . & . \\ . & . & . & . & \lambda_{35} & . & \lambda_{37} & . \\ . & . & . & . & . & . & . & . \\ . & . & . & . & . & . & . & \lambda_{58} \\ . & . & . & . & . & . & . & . \\ . & . & . & . & . & . & . & . \\ . & . & . & . & . & . & . & . \end{bmatrix}$$

$$= \begin{bmatrix} . & \alpha_Z & \alpha_X & \alpha_Y & . & . & . & . \\ . & . & . & . & \alpha_X \cdot e^{\beta_X} & \alpha_Y \cdot e^{\beta_Y} & . & . \\ . & . & . & . & \alpha_Z \cdot e^{\beta_Z} & . & \alpha_Y \cdot e^{\gamma_Y} & . \\ . & . & . & . & . & . & . & . \\ . & . & . & . & . & . & . & \alpha_Y \cdot e^{(\beta_Y + \gamma_Y)} \\ . & . & . & . & . & . & . & . \\ . & . & . & . & . & . & . & . \\ . & . & . & . & . & . & . & . \end{bmatrix}$$

Our data generating mechanisms are based on homogeneous Poisson processes and their intensity functions. Therefore, Poisson regression offers an intuitive examination to benchmark the correctness of our data generation mechanisms. We have verified that our data are generated correctly, by fitting the correctly specified conditional models to the simulated data.

In simulation 1, we are modeling the effect of being on treatment coded by $A_i(t)$ directly. Therefore, the pooled logistic model (Eq. 2.13) has a small modification to model $A_{ik} = 1$:

$$\text{logit}\left(P\left(A_{ik} = 1 \mid X_{i,k-1}; \eta\right)\right) = \eta_0 + \eta_1 X_{i,k-1}$$

Although the marginal and conditional exposure models for Poisson regression (Eq. 2.8) and Cox PH model (Eq. 2.10) have the functional forms unaltered, we need to change the interpretation to the hazard of starting the treatment.

Thereafter, the corresponding stabilized weighting functions of these exposure models have the same functional form as (Eq. 2.9), (Eq. 2.11) and (Eq. 2.15), except the interpretation for t_{A_i} is changed from treatment discontinuation time to treatment starting time for person i. The outcome model for the first scenario is a derivation of (Eq. 2.33), such that we are

modeling the marginal causal effect of treatment initiation $A_i(t) = 1$:

$$\psi_i(t;\theta) = \exp\left\{\theta_0(t) + \theta_1 A_i(t) + \theta_2 A_i(t)\,(t - t_A)\right\} \tag{3.5}$$

Our parameter of interest is θ_1, which can be interpreted as the treatment effect at the time of treatment initiation. The true value of θ_1 can be obtained from the coefficient of the effect of treatment on the outcome $\beta_Y = -1$ in the conditional data generating mechanism of $Y_i(t)$ (Eq. 3.2), and it can also be read from the log hazard ratio plot in Figure 3.5, where the coefficient change at treatment starting time $t = 2$ is -1.

Simulation 2

Even with a different modeling purpose, simulation 2 shares the same data generating mechanism with simulation 1. However, the initial state in the multistate model Figure 3.3 is now changed to state 2, that is, everyone starts exposed with no confounder and no death event. With some small variations at the direction of transitioning for state 2 & state 1, and state 5 & state 3, the TIM in the second scenario can be derived as below.

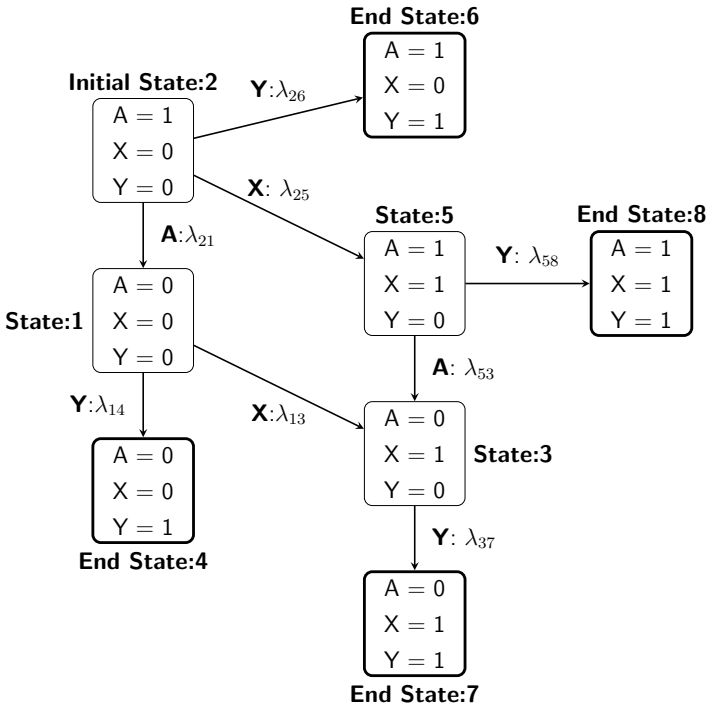


Figure 3.3: Simulation 2 multistate model for data generation, a new-user cohort starting with no confounder

$$TIM_2 = \begin{bmatrix} . & . & \lambda_{13} & \lambda_{14} & . & . & . & . \\ \lambda_{21} & . & . & . & \lambda_{25} & \lambda_{26} & . & . \\ . & . & . & . & . & . & \lambda_{37} & . \\ . & . & . & . & . & . & . & . \\ . & . & \lambda_{53} & . & . & . & . & \lambda_{58} \\ . & . & . & . & . & . & . & . \\ . & . & . & . & . & . & . & . \\ . & . & . & . & . & . & . & . \end{bmatrix}$$

$$= \begin{bmatrix} . & . & \alpha_X & \alpha_Y & . & . & . & . \\ \alpha_Z & . & . & . & \alpha_X \cdot e^{\beta_X} & \alpha_Y \cdot e^{\beta_Y} & . & . \\ . & . & . & . & . & . & \alpha_Y \cdot e^{\gamma_Y} & . \\ . & . & . & . & . & . & . & . \\ . & . & \alpha_Z \cdot e^{\beta_Z} & . & . & . & . & \alpha_Y \cdot e^{(\beta_Y + \gamma_Y)} \\ . & . & . & . & . & . & . & . \\ . & . & . & . & . & . & . & . \\ . & . & . & . & . & . & . & . \end{bmatrix}$$

As a new user cohort, we are modeling the effect of not being on treatment, coded by $1 - A_i(t)$. We construct the long-format exposure data set similarly as stated in simulation 1. The conditional exposure models - Poisson regression (Eq. 2.8), Cox PH model (Eq. 2.8), and pooled logistic model (Eq. 2.13) - are the same as previously introduced. The corresponding stabilized weighting functions of these exposure models (Eq. 2.9), (Eq. 2.11) and (Eq. 2.15) are unaltered, as we are modeling the exposure process directly. Besides, the marginal outcome model should also be modeling $1 - A_i(t)$, as presented in (Eq. 2.33).

Simulation 3

Similar to simulation 2, we are modeling a new-user cohort. However, from now on, we separate the modeling of the exposure process $A_i(t)$ into an MPP $(V_i(t), D_i(t))$. In simulation 3, we fit a logistic model for the dosage assignment model. Depending on the value of $D_i(t) = d$, we would have more states than the current eight states in Figure 3.3. However, instead of introducing more states, we decide to generate one TIM under each value of d and we switch between these matrices when a visit happens and we encounter a new value of d. Simulation 3 has the generalized multistate model Figure 3.4. We assume the initial dose $d = 1$, such that if there was no visit process happening, the d value in state 2, 5, 6

and 8 stays 1. For the other states, we have $d \in \{0, 1\}$. The data generating mechanisms for outcome and confounder processes remain the same as (Eq. 3.2) and (Eq. 3.3), but since $A_i(t)$ is a latent process induced by a change in the visiting process $V_i(t)$, we change the generation of $\lambda^{\mathcal{E}}_{A_i}(t)$ to $\lambda^{\mathcal{E}}_{V_i}(t)$.

$$\lambda^{\mathcal{E}}_{V_i}(t) = \alpha_V \cdot \exp\{\beta_V X_i(t^-)\} \tag{3.6}$$

The value of d is sampled with a logistic regression (Eq. 2.19) at time when $\mathrm{d}V_i(t) = 1$, with parameter values listed in Table 3.4. Thereafter, the aggregated two TIMs of simulation 3 can be specified as the below, and the parameter values can be found in Table 3.4.

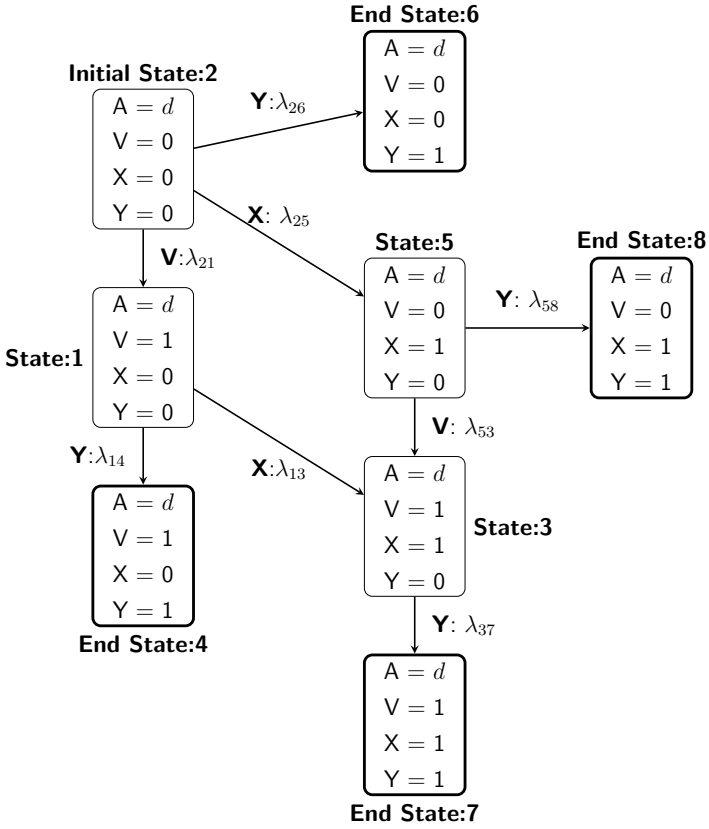


Figure 3.4: Simulation 3 multistate model for data generation. A new-user cohort starting with no confounder, and treatment A can change from exposed to stop or continue on the same dosage after a visit process V.

$$TIM_{3,D=d} =$$

$$\begin{bmatrix}
. & . & \alpha_X \cdot e^{\beta_{X_d}} & \alpha_Y \cdot e^{\beta_{Y_d}} & . & . & . & . \\
\alpha_V & . & . & . & \alpha_X \cdot e^{\beta_{X_d}} & \alpha_Y \cdot e^{\beta_{Y_d}} & . & . \\
. & . & . & . & . & . & \alpha_Y \cdot e^{\beta_{Y_d}+\gamma_Y} & . \\
. & . & . & . & . & . & . & . \\
. & . & \alpha_V \cdot e^{\beta_V} & . & . & . & . & \alpha_Y \cdot e^{(\beta_{Y_d}+\gamma_Y)} \\
. & . & . & . & . & . & . & . \\
. & . & . & . & . & . & . & . \\
. & . & . & . & . & . & . & .
\end{bmatrix}$$

After plugging in different values of d, we can see that the two TIMs for scenario 3 are different at the intensities λ_{13}, λ_{14} and λ_{37}.

The exposure status $A_i(t)$ starts with on treatment, but we might have a new prescription indicating treatment discontinuation with dose $d = 0$ at $\mathrm{d}V_i(t) = 1$, which forces $A_i(t) = 0$. Therefore, we are modeling a visit initiation in the visiting process, but not being on treatment in the outcome model.

The long-format visiting data set is constructed with the same idea as the long-format exposure data set described in simulation 1 and 2, such that we stop collecting the person-time record until $V_i(t) = 1$ or administrative censoring, which ever comes first. The conditional visiting models - the Poisson regression (Eq. 2.16), the Cox PH model (Eq. 2.17), and the pooled logistic model (Eq. 2.18) - are the same as introduced in Section 2.2.2.

In addition to the visiting models, we complete the MPP model by fitting a logistic dosage assignment model for the dosage level $D_i(t)$ (Eq. 2.19). The data set for fitting th dose is not in long-format, because we only take the person-time record at $\mathrm{d}V_i(t) = 1$.

The combined stabilized weights $\hat{sw}_i^{\bar{A}}(t)$ are directly (Eq. 2.24), (Eq. 2.25), and (Eq. 2.26) introduced in Section 2.2.4. To observe the impact of the different components in the weights, we also used stabilized weights $\hat{sw}_i^{\bar{V}}(t)$ or $\hat{sw}_i^{\bar{D}}(t)$ from fitting either a $V_i(t)$ model or a $D_i(t)$ model only.

The outcome model for simulation 3 is the same as (Eq. 2.33) for simulation 2, with a small modification that t_{A_i} is now changed to t_{V_i}.

Simulation 4

In simulation 4, we introduce an initial dosage variable $d_0 \in \{1, 2\}$ and expand the logistic model for d in the previous simulation to a multinomial logistic model, where $d \in \{0, 1, 2\}$.

Simulation 4 shares the same multistate model diagram, Figure 3.4, as simulation 3, but due to an increase in the number of d levels, we have three $TIM_4, D = d$. Because the initial dose d_0 is never zero, the d value in states 2, 5, 6 and 8 keeps consistent with d_0 before a visit process takes place, which means we do not switch to the $TIM_4, D = 0$ for these states. The outcome and confounder counting processes we used to generate the TIMs are:

$$\lambda^{\mathcal{E}}_{Y_i}(t) = \alpha_Y \cdot \exp\left\{\sum_d \beta_{Y_d} \cdot \mathbf{1}\{A_i(t) = d\} + \gamma_Y X_i(t^-)\right\}$$
$$\lambda^{\mathcal{E}}_{X_i}(t) = \alpha_X \cdot \exp\left\{\sum_d \beta_{X_d} \cdot \mathbf{1}\{A_i(t) = d\}\right\},$$

and the conditional generating mechanism for the visiting process stays the same as (Eq. 3.6) in simulation 3. The initial dose D_{i0} is sampled with a logistic model that is independent from $X_i(0) = 0$, and we assume the initial dose sampling has equal probabilities for the two d_0 levels:

$$\begin{aligned} &\text{logit}(P(D_{i0} = 2)) = \phi_2 \\ \iff &P(D_{i0} = 2) = \text{expit}(0) = 0.5 \end{aligned}$$

The generation for an after-visit dose level d is from a multinomial distribution (Eq. 2.21), with the corresponding parameter values in Table 3.4.

The long-format visiting data set and the dose data set are created as simulation 3. The visiting model functions can be adapted from simulation 3, but we add in D_{i0} as a baseline covariate to both the marginal and the conditional models. The combined stabilized weights under their corresponding model specifications have similar expression as (Eq. 2.24), (Eq. 2.25) and (Eq. 2.26) in simulation 3, except we add in $\kappa_{d_0} D_{i0}$ and $\eta_{d_0} D_{i0}$ to the numerator and the denominator, respectively, and we replace the logistic dosage model with the multinomial logistic model (Eq. 2.22) for simulation 4.

Previously in simulation 2 and 3, we take $1 - A_i(t) = 1$ to model the effect of not being on treatment, such that $A_i(t) = 1$ is the reference level. Because now we have more treatment levels of interest, the outcome model has an additional term for $A_i(t) = 2$, but we keep the reference level as $A_i(t) = 1$. However, the flat line after a change in treatment in Figure 3.7 indicates there is no need for a time-interaction term for $A_i(t) = 2$. We specify our outcome model simulation 4 as:

$$\begin{aligned} \psi_i(t;\theta) = \exp\{&\theta_0(t) + \theta_1 \mathbf{1}\{A_i(t) = 0\} + \theta_2 \mathbf{1}\{A_i(t) = 2\} \\ &+ \theta_3 \mathbf{1}\{A_i(t) = 0\}(t - t_{V_i})\} \end{aligned} \tag{3.7}$$

3.3 Simulation results

For simulation 1, the resulting marginal log-hazard graphs under two different treatment assignments are shown in Figure 3.5. We show the causal contrast between never treated versus start treatment at time $t = 2$. While the conditional models (dashed lines) shows proportional hazards, with constant baseline hazards, the marginal hazard (solid lines) shows complex functional form that is neither constant over time, nor proportional. In practice, parametric marginal hazard models can be made sufficiently flexible to characterize time-varying baseline hazards and treatment effects by introducing splines and interaction terms with time. However, for the purposes of the simulation study, we focus on a single parameter of θ_1 that characterizes the treatment effect at the time of treatment initiation/discontinuation. Because the true value of θ_1 is known and can be captured by the specified parametric model, it allows the comparison between weighted Cox and case-base sampling partial likelihoods.

The true value of our parameter of interest can be obtained for each scenario. In simulation 1, θ_1 in (Eq. 3.5) represents the log hazard ratio of starting treatment at some time $t = 2$ during the follow-up versus never treated can be read from the bottom graph of Figure 3.5, where the log hazard ratio drop at $t = 2$ is -1. For simulation 2 and simulation 3, a similar pattern of marginal versus conditional hazards can be observed in Figure 3.6; except in these scenarios, we model the stop of the treatment, and the true value for θ_1 in (Eq. 2.33) is now 1. For simulation 4, we assume the effect of treatment on the confounder process is the same regardless of the dosage level to avoid complex time interaction term in the marginal hazard function, that is $\beta_{X_1} = \beta_{X_2}$. The top graph of Figure 3.7 shows the same log hazard ratio trend as those in simulation 2 and 3 (the bottom graph of Figure 3.6), and thus the true value for θ_1 that models the stop of the treatment in (Eq. 3.7) is 1. Additionally, we have θ_2 that represents a change of treatment to another dose level after a visit takes place, and the true value $\theta_2 = -1$ can be read from the bottom graph of Figure 3.7.

The parameter estimation results are presented in the following tables; Table 3.5 for simulation 1, Table 3.6 for simulation 2, Table 3.7 for simulation 3, and Table 3.8 and Table 3.9 for simulation 4. The inferential statistics of interest are presented in the columns of these tables, in the order of Mean, Bias, Monte Carlo Standard Deviation (SD), Mean Standard Error (Mean SE), Root-Mean-Square Error (RMSE), 100 $\times$ Monte Carlo Error (100 $\times$ MCE), Coverage, and Power. The simulation results are not strictly rounded with respect to their MCE due to the small differences in the estimates. The SD and Mean SE produce approximately similar values, which indicates a valid simulation setup.

We observe several consistent patterns from all scenarios: i) The Cox and the case-base sampling partial likelihoods have similar performance in terms of efficiency properties.

However, due to the smaller long-format data dimension required in case-base sampling, the run time is approximately five times less compared to that of the Cox MSM. For example, in simulation 4, it took 10.86 seconds to complete with case-base sampling, while it took 45.90 seconds by using Cox. ii) The small undercoverage in the confidence interval observed in both Cox and case-base sampling MSM is likely due to the downward bias in the sandwich estimators (Fay and Graubard, 2001). iii) Regardless of continuous or discretized weights, the weighted estimates are similar, which likely results from sufficiently fine discretization of time. If the discretization is coarse, the discrete exposure weights might fail to approximate the continuous-time weights. iv) The combined weighted models give larger variance estimates than that of the single weighted models, and weighted models always have larger variance estimates compare to unweighted. v) However, the weighted models have a smaller and more favorable RMSE, a statistic as the sum of measured bias and standard deviation, compared to the unweighted models for model selection. We provide coverage and power as additional inferential statistics, but these are not of primary interest in the current simulation study. If either the point estimate or the SE estimate is biased, the coverage will be away from nominal probability of 0.95. Our coverage results show that MPP weightings always achieve the closest approximation to 0.95. Power is a function of efficiency, so we expect to see case-base and Cox having similar power. However, we should not compare power for methods that are biased, as bias away from the null always increase the power.

We also observe some characteristics that are specific to each simulation study. In Table 3.5 and Table 3.6, and we can see that the weighted estimates always have smaller bias compared to the unweighted estimates. For simulation 3, we separate the modeling of exposure into a visit model and a dosage assignment model, and we also examine the weighting contribution of each of these models separately. Therefore, we include all the model specifications introduced in the previous chapter. In Table 3.7, we can see that the visit or the dosage assignment model alone corrects some bias from the unweighted estimates, but the combined weights of these two exposure models always give the least biased estimates. The best model combination is using combined weights from a pooled logistic visit model and a logistic dosage assignment model, with a case-base estimation of MSM (model specification (VII†)). We observe the same overall results in simulation 4 that once again, the combined weights give the least biased estimates. Note that the dosage assignment model alone in Table 3.8 gives a more biased estimate than that of the unweighted, which is likely a coincidence. Because the biases from using dosage assignment alone or from visit model alone are to the opposite directions, they cancel out to some degree.

From the simulation studies, we have verified that when modeling exposure as a marked point process, we need to used the combined weights for the outcome model. Because dose

weights are independent from the outcome model form, the same kind of weights can also be used to model cumulative exposures. While we keep the outcome model in simulation simple to know the true parameter, this separation in the exposure modeling also naturally leads to our discussion about cumulative exposure in the real data analysis in the next chapter. We also learned that we can save significant amount of computational time by using case-base sampling.

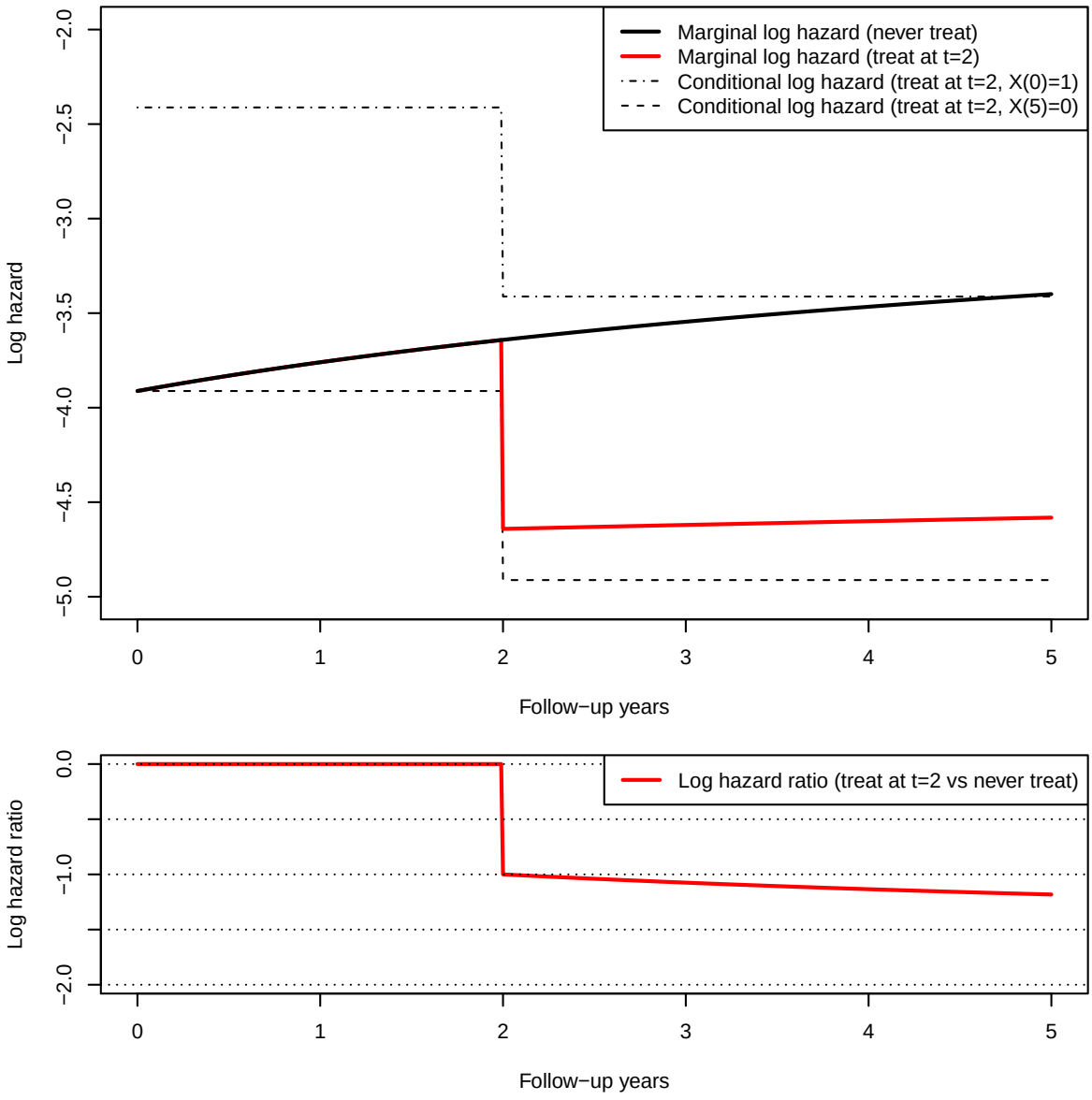


Figure 3.5: Simulation 1 log hazard rate graphs, where $A = 0$ has true value $\theta_1 = -1$

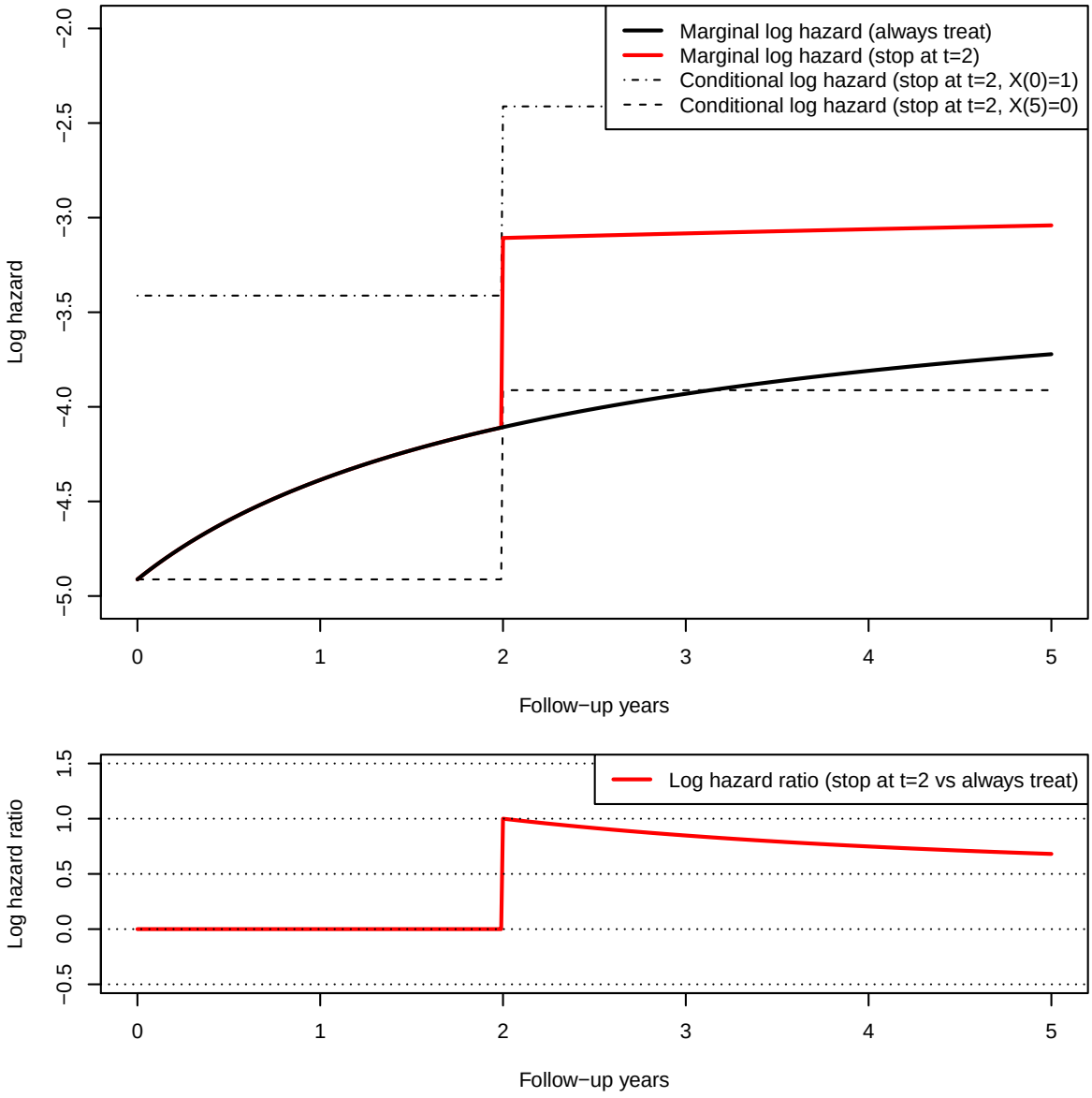


Figure 3.6: Simulation 2 and 3 log hazard rate graphs, where $A = 0$ has true value $\theta_1 = 1$

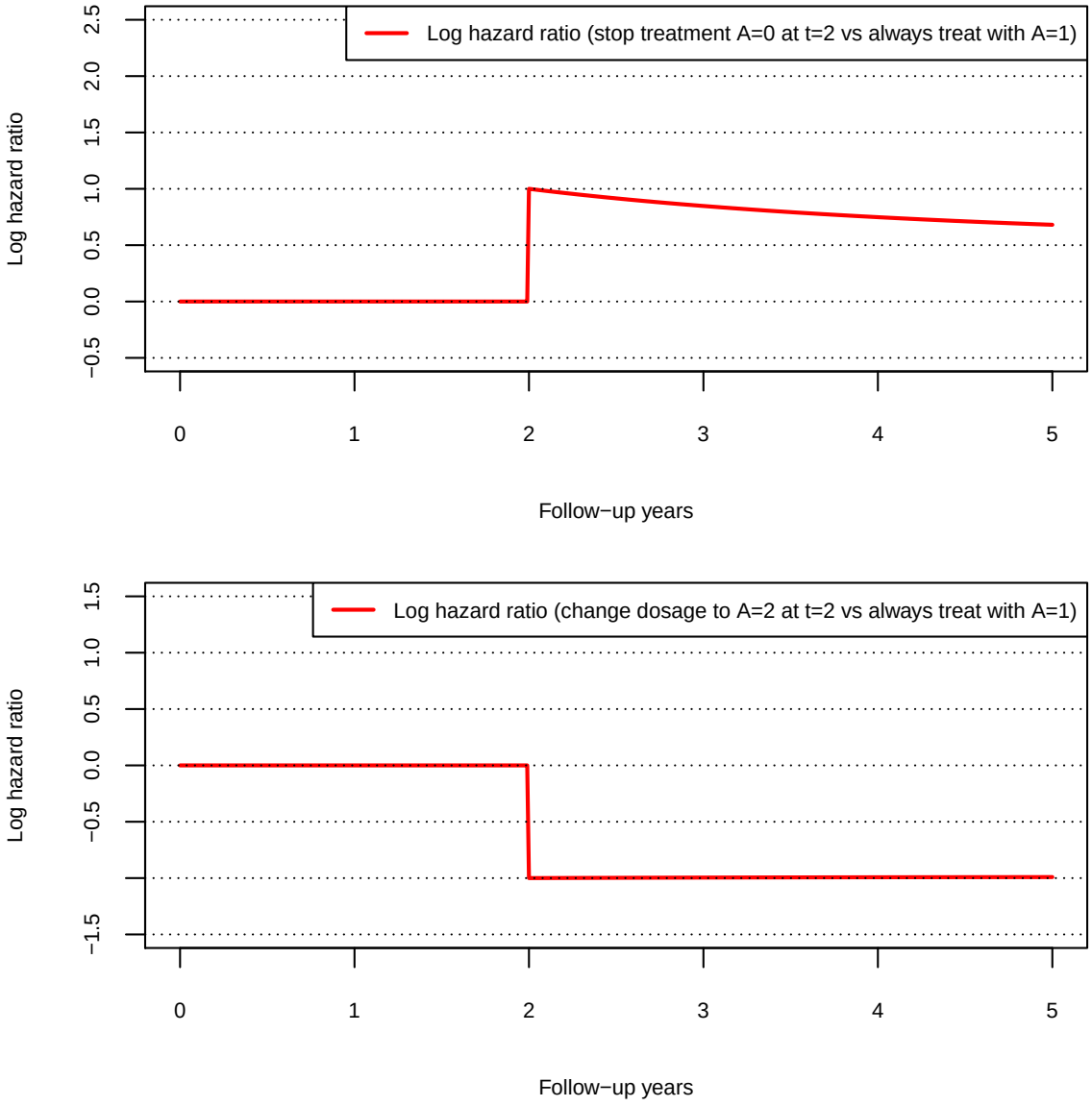


Figure 3.7: Simulation 4 log hazard ratio graphs, where $A = 0$ has true value $\theta_1 = 1$ and $A = 2$ has true value $\theta_2 = -1$

Table 3.5: Simulation 1 result: treatment initiation by modeling exposure as a counting process

Outcome Model Estimation Method	Exposure		Combo	Mean	Bias	SD	Mean SE	RMSE	100xMCE	Coverage	Power
	Weights	Model									
Cox	Continuous	Poisson	(I)	-1.063	-0.063	0.589	0.547	0.592	1.729	0.928	0.480
		Cox	(II)	-1.065	-0.065	0.583	0.544	0.587	1.722	0.932	0.485
	Discrete	Logistic	(III)	-1.062	-0.062	0.586	0.545	0.589	1.722	0.928	0.482
		Unweighted	(IV)	-0.654	0.346	0.508	0.485	0.615	1.534	0.829	0.217
Case-base	Continuous	Poisson	(V)	-1.063	-0.063	0.590	0.547	0.593	1.729	0.926	0.489
		Cox	(VI)	-1.064	-0.064	0.585	0.544	0.588	1.722	0.930	0.489
	Discrete	Logistic	(VII)	-1.061	-0.061	0.587	0.545	0.590	1.722	0.924	0.491
		Unweighted	(VIII)	-0.653	0.347	0.510	0.498	0.617	1.574	0.837	0.201

Table 3.6: Simulation 2 result: treatment discontinuation by modeling exposure as a counting process

Outcome Model Estimation Method	Exposure		Combo	Mean	Bias	SD	Mean SE	RMSE	100xMCE	Coverage	Power
	Weights	Model									
Cox	Continuous	Poisson	(I)	1.009	0.009	0.318	0.313	0.318	0.991	0.941	0.900
		Cox	(II)	1.010	0.010	0.321	0.315	0.321	0.996	0.941	0.889
	Discrete	Logistic	(III)	1.005	0.005	0.313	0.309	0.313	0.978	0.942	0.904
		Unweighted	(IV)	1.494	0.494	0.265	0.255	0.560	0.807	0.515	1.000
Case-base	Continuous	Poisson	(V)	1.011	0.011	0.319	0.315	0.319	0.995	0.945	0.896
		Cox	(VI)	1.012	0.012	0.323	0.316	0.323	1.000	0.936	0.885
	Discrete	Logistic	(VII)	1.007	0.007	0.314	0.310	0.314	0.981	0.947	0.901
		Unweighted	(VIII)	1.497	0.497	0.266	0.258	0.564	0.816	0.511	1.000

Table 3.7: Simulation 3 result: single treatment change with a logistic dose model in the MPP

Outcome Model Estimation Method	Exposure			Combo	Mean	Bias	SD	Mean SE	RMSE	100xMCE	Coverage	Power
	Weights	Visit Model	Dosage Model									
Cox	Continuous	Poisson	-	(I)	1.100	0.100	0.331	0.324	0.346	1.025	0.929	0.886
		Cox		(II)	1.108	0.108	0.323	0.318	0.340	1.007	0.928	0.903
		-	Logistic	(IX)	1.417	0.417	0.283	0.281	0.504	0.889	0.665	0.997
		Poisson	Logistic	(I†)	0.980	-0.020	0.356	0.347	0.357	1.096	0.943	0.795
		Cox		(II†)	0.987	-0.013	0.344	0.337	0.344	1.067	0.949	0.811
	Discrete	Pooled Logistic	-	(III)	1.101	0.101	0.329	0.323	0.344	1.020	0.928	0.893
		-	Logistic	(X)	1.418	0.418	0.283	0.281	0.505	0.889	0.663	0.997
		Pooled Logistic	Logistic	(III†)	0.984	-0.016	0.353	0.344	0.354	1.089	0.945	0.802
	Unweighted			(IV†)	1.509	0.509	0.280	0.278	0.581	0.879	0.529	0.999
Case-base	Continuous	Poisson	-	(V)	1.101	0.101	0.334	0.327	0.349	1.305	0.931	0.882
		Cox		(VI)	1.108	0.108	0.325	0.321	0.343	1.016	0.932	0.902
		-	Logistic	(XI)	1.417	0.417	0.286	0.284	0.505	0.898	0.666	0.995
		Poisson	Logistic	(V†)	0.979	-0.021	0.360	0.350	0.360	1.107	0.939	0.790
		Cox		(VI†)	0.986	-0.014	0.347	0.341	0.347	1.077	0.945	0.797
	Discrete	Pooled Logistic	-	(VII)	1.103	0.103	0.332	0.326	0.347	1.030	0.929	0.887
		-	Logistic	(XI)	1.418	0.418	0.285	0.284	0.506	0.897	0.664	0.995
		Pooled Logistic	Logistic	(VII†)	0.983	-0.017	0.357	0.348	0.357	1.100	0.937	0.793
	Unweighted			(VIII†)	1.510	0.510	0.282	0.283	0.583	0.894	0.540	0.998

Table 3.8: Simulation 4 result: single treatment change to discontinuation $A = 0$ with a multinomial logistic dose model in the MPP

Outcome Model Estimation Method	Exposure			Combo	Mean	Bias	SD	Mean SE	RMSE	100xMCE	Coverage	Power
	Weights	Visit Model	Dosage Model									
Cox	Continuous	Poisson	-	(I)	0.812	-0.188	0.436	0.424	0.475	1.339	0.921	0.524
		Cox		(II)	0.813	-0.187	0.430	0.414	0.468	1.309	0.928	0.542
		-	Logistic	(IX)	1.326	0.326	0.372	0.369	0.495	1.168	0.836	0.926
		Poisson	Logistic	(I†)	0.964	-0.036	0.424	0.415	0.425	1.314	0.947	0.647
		Cox		(II†)	0.968	-0.032	0.417	0.407	0.418	1.286	0.943	0.669
	Discrete	Pooled Logistic	-	(III)	0.815	-0.185	0.433	0.421	0.471	1.333	0.925	0.535
		-	Logistic	(X)	1.327	0.327	0.372	0.369	0.496	1.168	0.835	0.926
		Pooled Logistic	Logistic	(III†)	0.968	-0.032	0.421	0.413	0.422	1.308	0.946	0.662
	Unweighted			(IV†)	1.189	0.189	0.373	0.368	0.418	1.165	0.909	0.871
Case-base	Continuous	Poisson	-	(V)	0.815	-0.185	0.439	0.427	0.476	1.351	0.932	0.519
		Cox		(VI)	0.814	-0.186	0.432	0.418	0.470	1.321	0.931	0.532
		-	Logistic	(XI)	1.332	0.332	0.377	0.373	0.502	1.178	0.841	0.925
		Poisson	Logistic	(V†)	0.971	-0.029	0.427	0.419	0.428	1.324	0.951	0.650
		Cox		(VI†)	0.974	-0.026	0.421	0.410	0.422	1.297	0.945	0.671
	Discrete	Pooled Logistic	-	(VII)	0.818	-0.182	0.435	0.425	0.472	1.345	0.931	0.529
		-	Logistic	(XI)	1.333	0.333	0.377	0.373	0.503	1.179	0.839	0.925
		Pooled Logistic	Logistic	(VII†)	0.976	-0.024	0.424	0.417	0.424	1.318	0.950	0.654
	Unweighted			(VIII†)	1.192	0.192	0.376	0.368	0.422	1.164	0.905	0.870

Table 3.9: Simulation 4 result: single treatment change to a higher dose $A = 2$ with a multinomial logistic dose model in the MPP

Outcome Model Estimation Method	Exposure			Combo	Mean	Bias	SD	Mean SE	RMSE	100xMCE	Coverage	Power
	Weights	Visit Model	Dosage Model									
Cox	Continuous	Poisson	-	(I)	-1.111	-0.111	0.510	0.499	0.522	1.579	0.937	0.626
		Cox		(II)	-1.114	-0.114	0.505	0.492	0.518	1.557	0.935	0.627
		-	Logistic	(IX)	-1.146	-0.146	0.475	0.469	0.497	1.483	0.940	0.712
		Poisson	Logistic	(I†)	-1.029	-0.029	0.518	0.507	0.519	1.604	0.932	0.527
		Cox		(II†)	-1.030	-0.030	0.512	0.498	0.513	1.576	0.939	0.541
	Discrete	Pooled Logistic	-	(III)	-1.110	-0.110	0.507	0.498	0.520	1.574	0.937	0.633
		-	Logistic	(X)	-1.146	-0.146	0.475	0.469	0.497	1.483	0.940	0.713
		Pooled Logistic	Logistic	(III†)	-1.027	-0.027	0.516	0.506	0.517	1.600	0.933	0.538
	Unweighted			(IV†)	-1.231	-0.231	0.472	0.466	0.525	1.475	0.934	0.773
Case-base	Continuous	Poisson	-	(V)	-1.113	-0.113	0.511	0.501	0.523	1.586	0.936	0.626
		Cox		(VI)	-1.115	-0.115	0.506	0.494	0.519	1.563	0.936	0.632
		-	Logistic	(XI)	-1.143	-0.143	0.476	0.471	0.497	1.489	0.947	0.700
		Poisson	Logistic	(V†)	-1.025	-0.025	0.518	0.510	0.519	1.611	0.938	0.533
		Cox		(VI†)	-1.026	-0.026	0.512	0.501	0.512	1.583	0.940	0.543
	Discrete	Pooled Logistic	-	(VII)	-1.111	-0.111	0.509	0.500	0.521	1.581	0.936	0.629
		-	Logistic	(XI)	-1.143	-0.143	0.476	0.471	0.497	1.489	0.947	0.700
		Pooled Logistic	Logistic	(VII†)	-1.023	-0.023	0.516	0.508	0.516	1.607	0.937	0.534
	Unweighted			(VIII†)	-1.232	-0.232	0.473	0.458	0.527	1.450	0.926	0.797

Chapter 4

Data Analysis

This chapter demonstrates the marginal structural modeling of adverse drug effects in a real-world data application. We first present the chronic GC new-user cohort study design. Then, we derive data cleaning steps and definitions for our baseline covariates, GC exposure, time-dependent confounders, and fracture outcome. Corresponding descriptive statistics are also provided. Thereafter, the exposure models and the combined weighting function are adapted from Section 2.3. Subsequently, outcome models for each dose level and cumulative dose are fitted on the weighted population where time-dependent confounders are removed. We also generate the potential fracture hazard of always treated under each dose level.

4.1 Study design and patient cohort

We leveraged an existing cohort of older adults ($66 \leq$ age < 100) with respiratory conditions who were newly initiating chronic oral glucocorticoid (GC) users between January 1, 1998 and September 30, 2014 in Ontario, Canada (Amiche et al., 2018). The cohort was created using healthcare administrative data housed at ICES and cohort entry date t_0 was the date of first chronic oral GC claim. Patients were excluded previously before our cohort creation, if they have resided in a long-term-care facility at cohort entry or had co-morbidities (Pagets disease, osteomalacia, chronic renal disease, organ transplant, or malignancy other than skin), during the one-year look-back before cohort entry, see Appendix Figure A.1. We define t_1 as the start of follow-up date for each patient after the patient satisfies the chronic user criteria. Baseline risk factors were collected within one year prior to their t_1, except diabetes and previous bisphosphonate (BP) variables had more than one-year look back window, to allow for more complete documentation of diabetes diagnostic codes or the delayed effect due to BP could persist in bone after treatment discontinuation. Once a patient satisfied the chronic user criteria (with ≥ 2 GC prescriptions and a cumulative dose of ≥ 450 mg within

six months after t_0), we followed the patient until death, fracture, or one-year administrative censoring, whichever happened first. Patients with an extreme daily oral GC dose (>100mg) were censored on the dispensation date. The study design is presented in Figure 4.1.

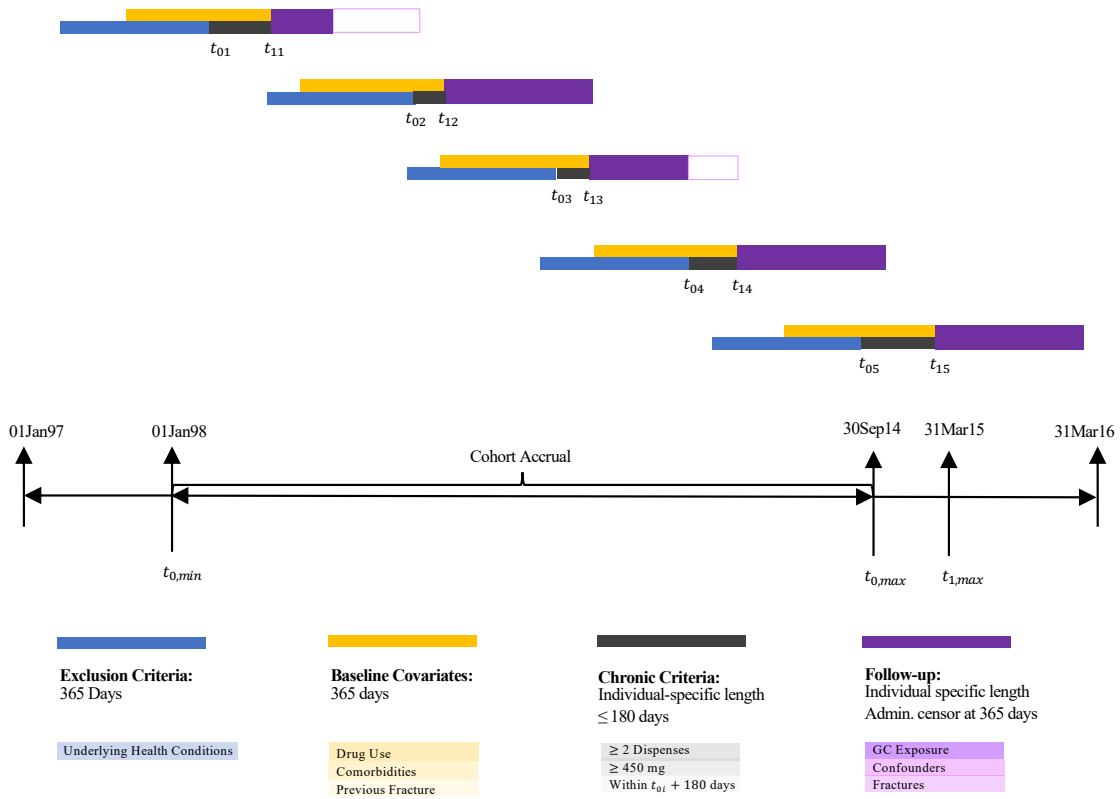


Figure 4.1: Study design diagram. $t_{0,\min}$: begin of cohort accrual; $t_{0,\max}$: end of cohort accrual; $t_{1,\max}$: the latest start of follow-up date possible; t_{0i} and t_{1i}, $i = 1...5$, are patient specific index date and start of follow-up date. All patients had their first claim t_0 taken place between $t_{0,\min}$ and $t_{0,\max}$, which ranges from January 01, 1998 to September 30, 2014, and the entire study data range from January 01, 1997 to March 31, 2016.

4.2 Data cleaning and definitions

4.2.1 Baseline covariates

To derive baseline covariates, we considered chronic comorbidities, drug use history, and previous fracture of each patient one year before the start of follow-up time t_1, except past diabetes diagnosis and past BP use had a longer look-back window of more than one year. We included age and adjusted for time since t_1 in a flexible and non-parametric way, and stratified the analysis by sex.

Table 4.1: Male patients characteristics by their initial daily dose categories

	Levels	(0,5mg]	(5,10mg]	(10,30mg)	[30,50mg)	≥50mg
Sample size		2007	4351	9864	11792	12606
Age (mean (sd))		76.94 (7.11)	76.36 (6.60)	75.59 (6.51)	74.80 (6.38)	74.40 (6.29)
Fall related drugs use (%)	**No**	513 (25.6)	1241 (28.5)	2881 (29.2)	3474 (29.5)	3662 (29.0)
	Yes	1494 (74.4)	3110 (71.5)	6983 (70.8)	8318 (70.5)	8944 (71.0)
Sex hormone (%)	**No**	1981 (98.7)	4283 (98.4)	9740 (98.7)	11652 (98.8)	12454 (98.8)
	Yes	26 (1.3)	68 (1.6)	124 (1.3)	140 (1.2)	152 (1.2)
Prev. BP duration (%)	**0 Days**	1762 (87.8)	3828 (88.0)	8843 (89.6)	10651 (90.3)	11540 (91.5)
	0-1 Year	147 (7.3)	324 (7.4)	568 (5.8)	643 (5.5)	520 (4.1)
	1-3 Years	61 (3.0)	102 (2.3)	230 (2.3)	276 (2.3)	295 (2.3)
	3-5 Years	18 (0.9)	50 (1.1)	122 (1.2)	101 (0.9)	127 (1.0)
	≥5 Years	19 (0.9)	47 (1.1)	101 (1.0)	121 (1.0)	124 (1.0)
Inhale GC (%)	**No**	1058 (52.7)	2340 (53.8)	4435 (45.0)	4888 (41.5)	5310 (42.1)
	Yes	949 (47.3)	2011 (46.2)	5429 (55.0)	6904 (58.5)	7296 (57.9)
Bronchodilators (%)	**No**	790 (39.4)	1852 (42.6)	3326 (33.7)	3424 (29.0)	3589 (28.5)
	Yes	1217 (60.6)	2499 (57.4)	6538 (66.3)	8368 (71.0)	9017 (71.5)
Thiazide (%)	**No**	1219 (60.7)	2878 (66.1)	6767 (68.6)	8275 (70.2)	9061 (71.9)
	Yes	788 (39.3)	1473 (33.9)	3097 (31.4)	3517 (29.8)	3545 (28.1)
Arthritis (%)	**No**	1599 (79.7)	3530 (81.1)	8590 (87.1)	10794 (91.5)	11761 (93.3)
	Yes	408 (20.3)	821 (18.9)	1274 (12.9)	998 (8.5)	845 (6.7)
Fall related conditions (%)	**No**	1248 (62.2)	2864 (65.8)	6595 (66.9)	8201 (69.5)	9026 (71.6)
	Yes	759 (37.8)	1487 (34.2)	3269 (33.1)	3591 (30.5)	3580 (28.4)
Emergency or	**No**	626 (31.2)	1288 (29.6)	2627 (26.6)	3145 (26.7)	2918 (23.1)
Hospitalization (%)	**Yes**	1381 (68.8)	3063 (70.4)	7237 (73.4)	8647 (73.3)	9688 (76.9)
Diabetes (%)	**No**	1480 (73.7)	3293 (75.7)	7499 (76.0)	8978 (76.1)	9319 (73.9)
	Yes	527 (26.3)	1058 (24.3)	2365 (24.0)	2814 (23.9)	3287 (26.1)
Previous fracture (%)	**No**	1961 (97.7)	4284 (98.5)	9731 (98.7)	11645 (98.8)	12449 (98.8)
	Yes	46 (2.3)	67 (1.5)	133 (1.3)	147 (1.2)	157 (1.2)
Fracture (%)	**No**	1960 (97.7)	4248 (97.6)	9641 (97.7)	11577 (98.2)	12390 (98.3)
	Yes	47 (2.3)	103 (2.4)	223 (2.3)	215 (1.8)	216 (1.7)

Previous raloxifene or calcitonin, denosumab, and anti-sex hormone are excluded due to small counts.

There were nine baseline drug use history variables selected into our model. We grouped psychiatric drugs, Parkinsons disease drugs, Alzheimers disease drugs, hypertension drugs, benzodiazepines and antiepileptics as a falls related drug use binary variable. We grouped estrogen and androgen drugs as a sexual hormone indicator variable, and other anti-estrogenic and anti-androgenic therapies as a hormone antagonists and related agents indicator variable. We also included the total days treated with ALD equivalent BP previous as a continuous variable (primarily categorized as 0 days, < 365 days, $1 - 3$ years, $3 - 5$ years, and > 5 years). We grouped previous raloxifene and calcitonin as an other OP binary indicator. Because denosumab is a second line of OP therapy drug, we derived a baseline denosumab variable. We included previous use of other respiratory medicines, such as inhaled GC indicator and bronchodilators indicator. We also include thiazide diuretics indicator because it is associated with increase in bone mineral density (Lim et al., 2005).

There were four chronic comorbidity variables selected into our model. We grouped all inflammatory arthritis subtypes - rheumatoid arthritis, polymyalgia, systemic lupus erythematosus, gouty arthritis or other rheumatic disease - as an arthritis condition binary variable. We grouped all falls related conditions into a binary variable, including comorbidities of dementia and Alzheimers disease, stroke, cardiovascular conditions, psychiatric disease and syncope orthostatic hypotension. We coded hospitalizations or visits to emergency during baseline period as a healthcare use binary variable. We include the diabetes binary variable for whether there is a diagnosis of diabetes from the Ontario Diabetes Dataset (ODD) prior to the follow-up date.

Table 4.2: Female patients characteristics by their initial daily dose categories

	Levels	(0,5mg]	(5,10mg]	(10,30mg)	[30,50mg)	≥50mg
Sample size		2876	5141	11986	13223	12996
Age (mean (sd))		77.50 (7.37)	76.33 (6.93)	75.56 (6.82)	74.66 (6.66)	74.14 (6.51)
Fall related drugs use (%)	**No**	619 (21.5)	1113 (21.6)	2871 (24.0)	3183 (24.1)	3230 (24.9)
	Yes	2257 (78.5)	4028 (78.4)	9115 (76.0)	10040 (75.9)	9766 (75.1)
Sex hormone (%)	**No**	2623 (91.2)	4648 (90.4)	10759 (89.8)	11947 (90.4)	11903 (91.6)
	Yes	253 (8.8)	493 (9.6)	1227 (10.2)	1276 (9.6)	1093 (8.4)
Anti-sex hormone (%)	**No**	2852 (99.2)	5080 (98.8)	11867 (99.0)	13110 (99.1)	12844 (98.8)
	Yes	24 (0.8)	61 (1.2)	119 (1.0)	113 (0.9)	152 (1.2)
Prev. BP duration (%)	**0 Days**	1712 (59.5)	3151 (61.3)	7723 (64.4)	8518 (64.4)	8653 (66.6)
	0-1 Year	435 (15.1)	842 (16.4)	1629 (13.6)	1629 (12.3)	1386 (10.7)
	1-3 Years	331 (11.5)	456 (8.9)	1089 (9.1)	1301 (9.8)	1222 (9.4)
	3-5 Years	151 (5.3)	304 (5.9)	678 (5.7)	791 (6.0)	686 (5.3)
	≥5 Years	247 (8.6)	388 (7.5)	867 (7.2)	984 (7.4)	1049 (8.1)
Prev. Raloxifene	**No**	2848 (99.0)	5106 (99.3)	11909 (99.4)	13139 (99.4)	12919 (99.4)
or Calcitonin (%)	**Yes**	28 (1.0)	35 (0.7)	77 (0.6)	84 (0.6)	77 (0.6)
Prev. Denosumab (%)	**No**	2870 (99.8)	5135 (99.9)	11968 (99.8)	13200 (99.8)	12967 (99.8)
	Yes	6 (0.2)	6 (0.1)	18 (0.2)	23 (0.2)	29 (0.2)
Inhale GC (%)	**No**	1559 (54.2)	2772 (53.9)	5505 (45.9)	5311 (40.2)	5161 (39.7)
	Yes	1317 (45.8)	2369 (46.1)	6481 (54.1)	7912 (59.8)	7835 (60.3)
Bronchodilators (%)	**No**	1338 (46.5)	2411 (46.9)	4543 (37.9)	4280 (32.4)	3849 (29.6)
	Yes	1538 (53.5)	2730 (53.1)	7443 (62.1)	8943 (67.6)	9147 (70.4)
Thiazide (%)	**No**	1654 (57.5)	3141 (61.1)	7516 (62.7)	8635 (65.3)	8636 (66.5)
	Yes	1222 (42.5)	2000 (38.9)	4470 (37.3)	4588 (34.7)	4360 (33.5)
Arthritis (%)	**No**	2028 (70.5)	3847 (74.8)	10008 (83.5)	11777 (89.1)	11974 (92.1)
	Yes	848 (29.5)	1294 (25.2)	1978 (16.5)	1446 (10.9)	1022 (7.9)
Fall related conditions (%)	**No**	1859 (64.6)	3402 (66.2)	8241 (68.8)	9301 (70.3)	9517 (73.2)
	Yes	1017 (35.4)	1739 (33.8)	3745 (31.2)	3922 (29.7)	3479 (26.8)
Emergency or	**No**	931 (32.4)	1659 (32.3)	3454 (28.8)	3644 (27.6)	3175 (24.4)
Hospitalization (%)	**Yes**	1945 (67.6)	3482 (67.7)	8532 (71.2)	9579 (72.4)	9821 (75.6)
Diabetes (%)	**No**	2271 (79.0)	3991 (77.6)	9430 (78.7)	10500 (79.4)	10163 (78.2)
	Yes	605 (21.0)	1150 (22.4)	2556 (21.3)	2723 (20.6)	2833 (21.8)
Previous fracture (%)	**No**	2765 (96.1)	4984 (96.9)	11647 (97.2)	12865 (97.3)	12703 (97.7)
	Yes	111 (3.9)	157 (3.1)	339 (2.8)	358 (2.7)	293 (2.3)
Fracture (%)	**No**	2741 (95.3)	4892 (95.2)	11488 (95.8)	12737 (96.3)	12529 (96.4)
	Yes	135 (4.7)	249 (4.8)	498 (4.2)	486 (3.7)	467 (3.6)

We grouped all fracture subtypes - hip, vertebral, radius, humerus - as a previous fracture binary variable. The patient characteristic tables for male and female patients are shown in Table 4.1 and Table 4.2, respectively.

During the one year follow-up period, we had nearly equal number of male (46.7%) and female (53.2%) patients in our final cohort. We see a higher fracture incidence rate among female (4%) than male (2%). There are several consistent patterns from both characteristic tables. High dose patients have a younger mean age at the start of follow-up t_1 than that of the low dose patients. Across the columns of both tables, while the daily dose levels increase from low to high, we see a decreasing trend in percentages for arthritis, fall related conditions, and previous fracture. An interpretation for arthritis condition, for example, is that less proportion of high initial GC daily dose patients had arthritis condition in the year before t_1, compared to the proportions in the lower initial GC daily dose groups. On the other hand, the percentages of had visited an emergency department or had been hospitalized, or had used inhale GC during baseline period are positively correlated with the initial dose level. For female patients only, the percentage of previous fall related drugs use tends to decrease as the initial daily dose level rises.

4.2.2 GC Exposure

For each GC dispensation of a patient, due to the various pharmacokinetics of each GC subtype, we converted different GC strength into prednisone equivalent strength Table 4.3 and multiplied prednisone equivalent strength by the quantity dispensed to get the total dose of one dispensation. Due to lack of data on patient adherence, we then assumed patients have full adherence so that days-supply is an accurate measurement of treatment duration. Therefore, we inferred the average daily dose from a patients one prescription by dividing the total dose (strength $\times$ quantity) in prednisone equivalence with days-supplied.

In order to determine a reasonable average daily dose for our cohort, we first pulled Canadian Thoracic Society (CTS) guidelines on steroid treatment for asthma and Chronic Obstructive Pulmonary Disease (COPD). As indicated in Table 4.4, the guideline on daily dose and duration of prednisone treatment changes over the years for acute exacerbation for both indications. In addition, both asthma and COPD guidelines note that the actual dose and duration should be individualized based on previous or current response. There are also some differences in guidelines for asthma and COPD. While the COPD guidelines do not suggest long-term treatment and highlight that there is only evidence for short-term GC treatment benefits within 30 days after exacerbation, long-term oral GC is required for difficult to control asthma but with osteoporosis prevention being addressed (Lougheed et al.,

2010; Odonnell et al., 2007). Based on these guidelines, we can primarily conclude that the common GC daily dose for respiratory disease exacerbation treatment is 25-60mg for 7-14 days as a single prescription, but the exact GC daily dose and duration should be individualized in our data analysis. However, after checking the distribution of our data (Appendix Figure A.2 and Appendix Figure A.3), we confirmed that there was no consistent treatment daily dose, or duration, or an obvious tapering regime among our cohort, so we decided to use the individualized calculation of average daily dose.

Table 4.3: GC subtypes conversion table (Album, 2014; Edsbacker and Andersson, 2004)

Drugs	5mg Prednisone Equivalent	Conversion Factor
Budesonide	1.125	4.44
Cortisone	25	0.20
Dexamethasone	0.75	6.67
Hydrocortisone	20	0.25
Methylprednisone	4	1.25
Prednisolone	5	1
Prednisone	5	1
Triamcinolone	4	1.25

Table 4.4: Canadian Thoracic Society guidelines on prednisone equivalent treatment on asthma and COPD for acute exacerbation (Lougheed et al., 2010, 2012; Odonnell et al., 2007, 2008)

Asthma	**2010**	**2012**
	40-60 mg per day for 7 to 14 days.	30-50 mg per day for at least 5 days.
COPD	**2007**	**2008**
	30-40 mg per day for 10-14 days.	25-50 mg per day for 7-14 days.

Another challenging point was the overlapping of two dispensation duration dates. Treatment episode construction is one of the four common pharmacoepidemiological problems (Pazzagli et al., 2018). In reality, patients may collect new medications before finishing all pills from the previous visit, due to personal reasons that might make them have to reschedule the next visit, and thus introducing overlaps in the administration data record (Pazzagli et al., 2018). The different methods used to construct the treatment episode might lead to various estimates on drug efficacy and safety. Current methods suggested by Pazzagli

et al. (2018) on dealing with overlaps are 1) adding the overlap days to the end of the next dispensation, which prolongs the overall duration of treatment or 2) ignoring the overlap by assuming shorter duration of the previous dispensation, which might affect the drug effect outcome estimation. We define our treatment episode for GC using a combination of these approaches based on pharmacological and clinical considerations.

Some patients had two dispensations on the same date, and the smaller dosage could be a tapering treatment after the major treatment. We combine these two dispensations as one by appending the smaller daily dose dispensation after the major prescription ended, and calculate the daily dose as the sum of total dose divided by sum of total days supplied. If this combined dispensation has a days-supply exceeds the maximum allowed value of 100, we cut off at 100 days and truncate the total dose accordingly.

In order to correct the overlapping days from different dispensations, we use the preliminary evidence from Appendix 4. to make an assumption that if a patient gets the next dispensation within 30 days earlier than expected, then we would push the next dispensation forward until the previous dispensation is finished and start the second treatment later. All subsequent dispensations would be pushed forward accordingly. On the other hand, we assumed that if a patient gets the next dispensation more than 30 days earlier, the patient might need to renew a dispensation, so we truncate the current days-supply and switch to the new dispensation.

After the data cleaning stated above, 97.5% of the daily dose distribution of all dispensations in our data Appendix Figure A.2 was within 100mg, which incorporates the suggested daily dose provided by CTS Table 4.4. We then censored the patients who ever had $>$ 100mg extreme daily dose in their oral GC dispensations at that extreme dispensation date. The resulting new distribution of daily dose from all dispensations is shown in Appendix Figure A.5.

We determined the start of follow-up date using the criteria of $\geq$ 2 oral GC dispensations and $\geq$ 450mg cumulative dose. If a patient reached 450mg at the first dispense, then we start the follow-up at the second dispense date; if a patient did not reach 450mg at the first dispense, then from the second dispense on, we obtain the cumulative dose by summing total dose of each dispense of a patient, until the date of a dispense that the cumulative dose reached 450mg. According to Appendix Figure A.6 $< 1\%$ patients have t_1 beyond one and half years of t_0, we exclude patients with extremely late follow-up dates to ensure a more homogeneous study cohort. Our study follow-up length was determined to be one year with effective GC treatment length consideration (Amiche et al., 2016, 2018; De Vries et al., 2007; Steinbuch et al., 2004), so we calculated the administrative censoring date as 365 days after the start of follow-up date. We excluded patients who have no more GC exposure after the start of follow-up date.

4.2.3 Fracture Outcome

Our fracture data are pulled from three databases - Ontario Health Insurance Plan (OHIP), Discharge Abstract Database (DAD) and National Ambulatory Care Reporting System (NACRS). OHIP is for outpatient records, DAD contains hospitalization records, and NACRS documents visits to emergency department (ICES, 2017, 2020a,b). DAD and NACRS contain ICD-9 and ICD-10 diagnosis codes listed in Appendix Table A.1 that we used to code our fracture variable. The fracture outcome variable is coded if any of the fracture subtypes - hip, vertebral, radius, or humerus happened during the follow-up.

A washout period is the length of days between recurrent fracture claims (Folkestad et al., 2017). If a patient has two records of the same fracture type within a washout period, we ruled out the second record as a physician check-up for the first fracture instead of a unique second fracture. Previous literature has shown that different definitions of a fracture event might lead to inconsistent estimates in osteoporosis administrative data research (ODonnell et al., 2013). Therefore, we explored the length of different washout periods to see if there was a significant reduction in number of outcomes identified. According to Appendix Table A.4, the length of washout period (0 day, 90 days, and 120 days) does not affect the case numbers of hip fracture, because it is unlikely to have recurrent events for hip fracture, and thus we did not use a washout period for our definition of hip fracture. On the other hand, for vertebral, radius or ulna, and humerus fractures, there was a small decrease in number of outcomes identified when using a 120 day washout period compared to a 90 day washout period. Our data and literature reviews suggest that a 90 day washout period would be sufficient (Lix et al., 2012; ODonnell et al., 2013), so we used a 90-day washout period for defining vertebral, radius, or humerus fracture. The outcome of interest is the first fracture of any fracture subtype. We removed patients who died before their start of follow-up date and censored at death time for patients who died before a fracture event, which might due to data entry error. With the above data cleaning steps from GC exposure and fracture outcome presented in Figure 4.2, we finalized our study cohort.

In our study, patients who are non-administratively censored can experience either death or fracture. Due to death being a competing risk for our event of interest fracture, the fracture hazards are cause-specific hazards. As descriptive analyses show in Figure 4.3, the distribution of density curves are similar between female and male for either death or fracture subgroup. The peak of death time is around the 30^{th} day of the follow-up period. The impact of GC on fracture also appears to manifest early during the one-year follow-up (female at around 30^{th} day and male around 80^{th} day).

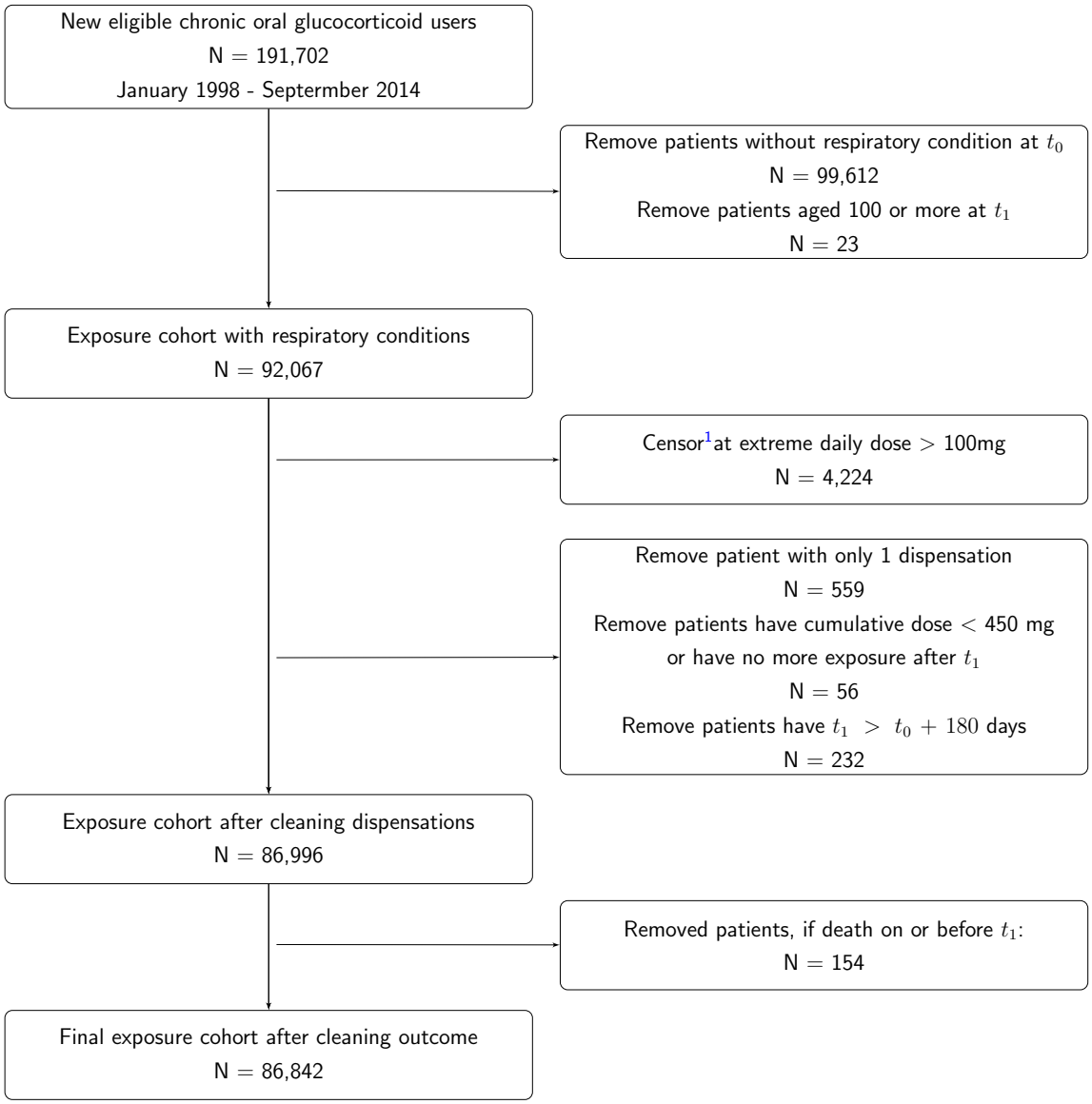


Figure 4.2: Study flow diagram. t_0 is the index date of first dispensation and t_1 is the start of follow-up date after a patient satisfy the chronic user criteria; both t_0 and t_1 are subject-specific. Among the final cohort, a total of 2,639 (3.0%) patients experienced an osteoporosis fracture; 11,858 (13.7%) patients died without a fracture; 72,345 patients reached the end of follow-up and experienced an administrative censoring. In terms of TD confounder, 19,331 patients (666 had a fracture) ever had an BP prescription and 2,069 patients (68 had a fracture) ever entered long-term care during the one year follow-up.

[1] All patients with an extreme daily dose had it at the first dispensation, so censoring is deletion.

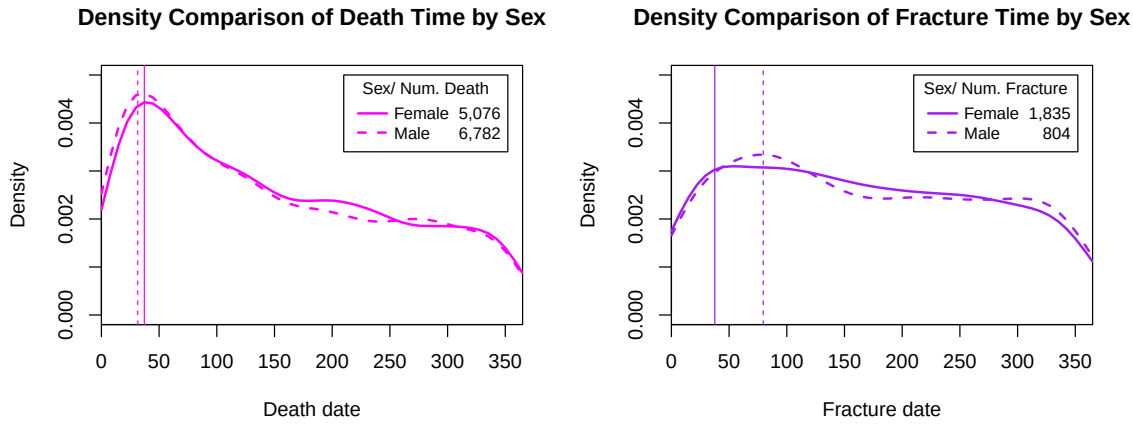


Figure 4.3: Event time density comparison by sex for non-administratively censored patients.

After determining each patient's start and end of follow-up dates, we can visualize their exposure pattern with different daily dose categories and outcome types. Due to ICES privacy protection guidelines, we simulated data for exposure and outcome of five patients at the individual level for illustration Figure 4.4.

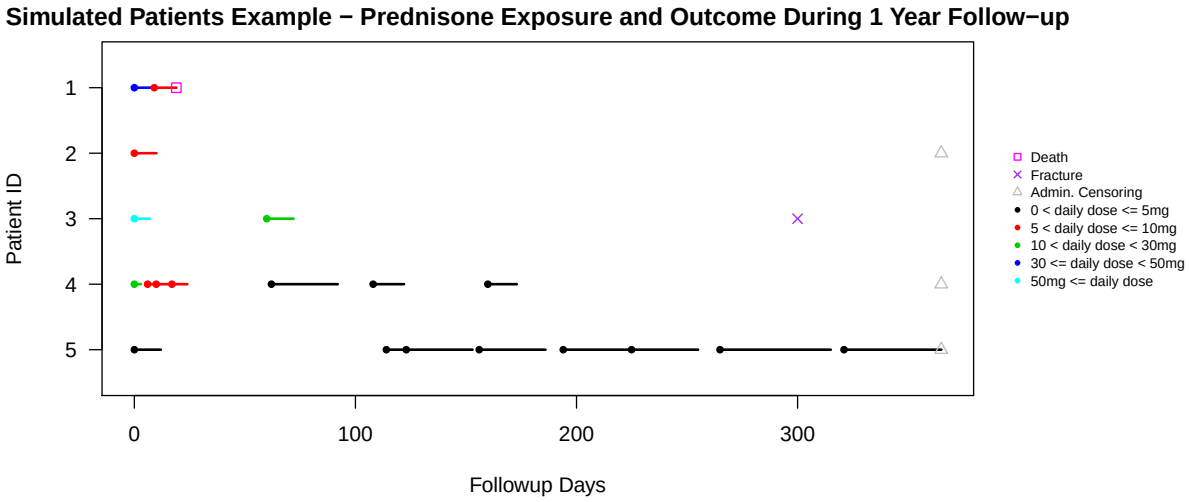


Figure 4.4: Simulated patient GC exposure and outcome during 1 year follow-up. We put the start of follow-up date t_1 of the 5 patients in Figure 4.1 onto the same 0-365 days scale.

With the aim of controlling for data dimension while maintaining sufficient clinical information, we discretize our exposure data set into long-format with an interval of 5 days, which corresponds to the first peak of our treatment duration distribution from all dispensations Appendix Figure A.3. In addition, the density comparisons for dose-specific treatment duration is presented in Figure 4.5, from which we validated that high dose groups have short

days-supplied values, while low dose groups are for maintenance treatments that can have longer therapy cycles.

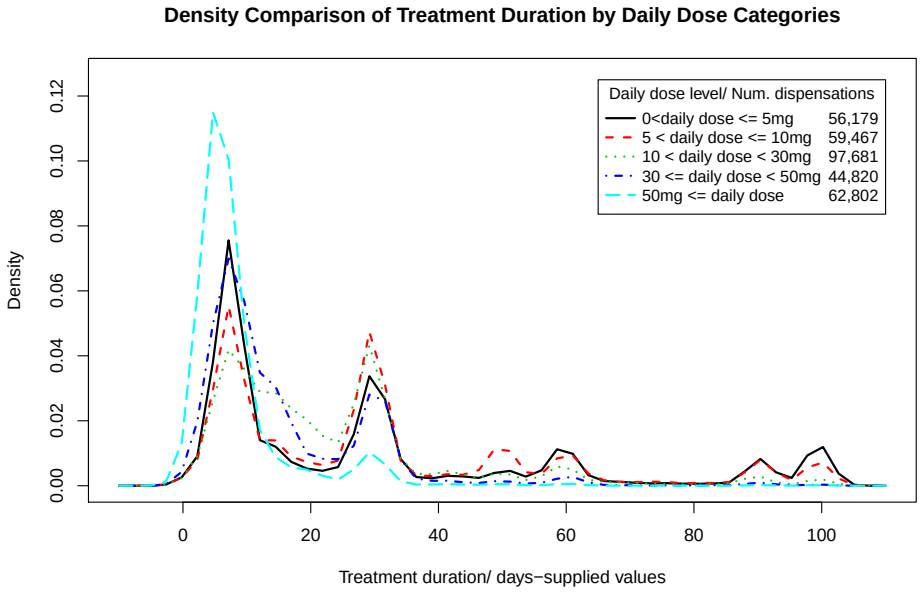


Figure 4.5: Days-supplied value for different dosage categories.

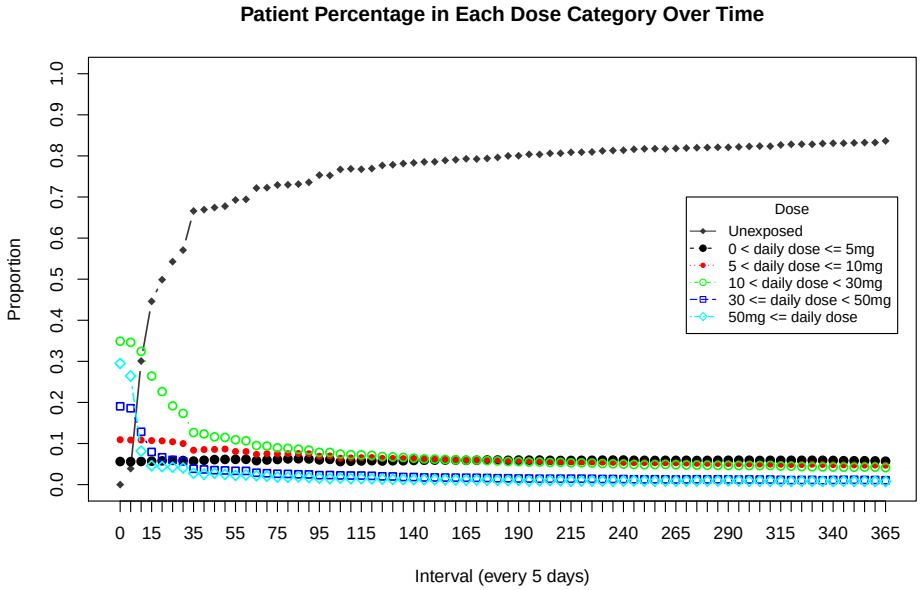


Figure 4.6: The percentages of patients prescribed with each daily dose category in the riskset over time.

During the study length of 365 days, a patient can have a maximum of 73 time intervals k

if the follow-up is terminated by administrative censoring. With the long-format GC exposure data set, a descriptive visualization of GC dispensation patterns over the follow-up period is presented in Figure 4.6. The gray line represents the percentage of unexposed patients in our cohort, which starts from zero at t_1, reached a plateau at day 105, and eventually landed around 80% for those reached the end of the study period. On the other hand, high daily dose ($>$ 10mg) prescriptions are nearly 20% more common than the low daily dose ($\leq$ 10mg) prescriptions at t_1, and this high-dose dominance pattern gradually faded beyond the first three months.

4.2.4 TD Confounders

Standard methods, such as Cox outcome regression, cannot deal with time-dependent confounders in the presence of a treatment-confounder feedback loop, in which case MSMs are needed. Our time-dependent confounders are long term care during the follow-up, and OP drugs, which can be divided into classes of BP and other OP drugs.

BPs can be categorized into two subtypes: nitrogen-containing and non-nitrogen containing BPs (Ganesan et al., 2020). Nitrogen-containing BPs include alendronate, risedronate, and zoledronic acid; and non-nitrogen containing bisphosphonates include etidronate (Ganesan et al., 2020). A common method to compare drug utilization patterns across different regions or countries is to consider the World Health Organization (WHO) Defined Daily Dose (DDD). Appendix Table A.2 provides a summary of WHO DDD in milligrams of different types of BP medications prescribed in this study.

However, WHO DDD fails to consider the condition under treatment and includes only a single DDD for a medication, regardless of indication or strength. For example, etidronate is prescribed as a daily dose of 200 mg to treat Pagets disease, yet 400mg/day for 14 days covers therapy for 90 days, and thus effective daily treatment dose is $(400 \times 14)/90 = 62$mg/day for osteoporosis (Burden et al., 2013; Merlotti et al., 2009). This is considered equivalent therapy to alendronate 10mg daily, or risedronate 5mg daily (Burden et al., 2013; WHO, 2019b,e). We thus used clinical guidelines for therapy to treat osteoporosis and prevent fractures to define alendronate equivalent daily dose of bisphosphonates to enable modeling bisphosphonate therapy as a single exposure. Although some data suggest that risedronate reduces fracture risk earlier than alendronate, while alendronate benefits persist in bone longer after discontinuation than risedronate, and that the nitrogen-containing bisphosphonates are more effective in reducing fracture risk than the non-nitrogen-containing bisphosphonate etidronate, the evidence in the real world is unclear (Cadarette et al., 2008, 2013), particularly in the context of GC-induced osteoporosis (Amiche et al., 2018). Therefore for the purposes of this study, an

alendronate-equivalent dose was calculated consider exposure to all types of bisphosphonates.

Table 4.5: 10mg daily alendronate equivalent bisphosphonates (Burden et al., 2013)

Drug Names	Dose	Daily Alendronate Equivalent	Quantity	Days Supply Imputation
Alendronate	5mg/day	5mg	Q	Q
Alendronate	10mg/day 70mg/week	10mg	Q	Q Q × 7
Risedronate	5mg/day 35mg/week 150mg/month	10mg	Q	Q Q × 7 Q × 30
Etidronate	400mg/day for 14 days (then 500mg calcium for 76 days)	10mg	Q*	Q* × 90
Zoledronic Acid	1 infusion of 5mg/100ml in a year	10mg	Q*	Q* × 365

* Q=1 for etidronate and zoledronic acid.

Similarly, we need to harmonize the difference between WHO suggested dosage and the actual dosage we have. All our documented prescriptions for zoledronic acid are 5mg/100ml per infusion and lasts 365 days, so we adopted this value, instead of using the WHO suggested 4mg/100ml that includes common cancer dosage (Burden et al., 2013). Due to the common days-supply entry error in Ontario osteoporosis administrative data, we imputed the days-supply value as indicated in Table 4.5 and made BP prescriptions strictly to 10mg alendronate equivalent (Burden et al., 2013). After cleaning alendronate equivalent BPs with the above derived guidelines, we further cleaned the overlapping exposure windows of consecutive dispensations of a patient, with a small variation of the strategy we used when cleaning GC exposure overlaps due to different pharmacokinetics between GC and BP. If two dispensations are made on the same date, we deleted the lower dose dispensation. To make sure there was no overlap between dispensations, if a patient received the next dispensation of the same molecule, then we shifted all subsequent dispensations until the patient finished the current days-supply value; if a patient changes treatment molecule (e.g. alendronate to risedronate, or alendronate to denosumab), then we truncate the current days-supply to the new dispensation date (Pazzagli et al., 2018).

From our primary summary statistics, female has more prevalent TD confounders compare to male - 31.3% of female in our GC new-users cohort had taken BP therapy, compare to only 12.0% in male; the prevalence of LTC initiation is roughly the same for female (2.6%) and male (2.1%).

4.3 Models for GC usage patterns

A baseline covariate is considered a confounder if it impacts both the dosage assignment and the outcome. Variables that are highly associated with the exposure but do not affect the time-to-event outcome should be removed from the analysis. Therefore, univariate pre-screening is conducted for each baseline variable with the initial dose assignment using a chi-squared test and the time-to-event outcome using a log-rank test. Analyses are stratified by sex, because we hypothesize that different sex have different hormone treatment. Among our selected baseline covariates, anti-sex hormone therapies, bronchodilators, and previous denosumab use are not significantly ($p > 0.05$) correlated with the outcome for both sex and thus are excluded from the analysis. For males specifically, we further excluded the insignificant variable of diabetes and also the previous raloxifene or calcitonin use due to zero or small counts that would lead to positivity assumption violation when fitting the model.

Afterwards, the selected baseline covariates are integrated into our long-format exposure data set which was discretized with five-day intervals. For any continuous variable in our exposure and outcome data sets, we center it by subtracting the mean before fitting the model. The conditional GC exposure models are adapted from the models for D_{i0}, V_{ik}, and D_{ik} introduced in Section 2.3, with the modification that $\eta_1 Z_i = \sum_{\iota=1}^{\iota=q} \eta_{1\iota} Z_{i\iota}$ includes the q selected baseline variables and $\eta_2 X_{ik} = \sum_{r=1}^{r=2} \eta_{2r} X_{ir,k}$ stands for long term care (LTC) initiation, where we lag one interval for LTC confounder labeling in discrete-time setting, and calculate BP cumulative days up to time $k-1$. Instead of fitting the visiting model with a linear time effect k, we replace $\eta_5 k$ and $\phi_5 k$ with a quadratic spline basis transformation $\sum_p \eta_{5p} b_{pk}$ and $\sum_p \phi_{5p} b_{pk}$, where $p = 1, 2$ is the spline degree indicator.

The marginal versions of the three exposure models all contain the baseline covariates, as we will adjust the baseline covariates through the outcome regression model, and we only adjust time-dependent confounders with IPTW. Therefore, the marginal D_{i0} model is the same as its conditional model, while the marginal version of our multiple visit MPP V_{ik} model (Eq. 2.27) and D_{ik} model (Eq. 2.30) are specified excluding the time-dependent covariates $\sum_{r=1}^{r=2} \eta_{2r} X_{ir,k}$. Our fitted conditional initial dose D_{i0} model has the form:

$$\text{logit}[P(D_{i0} \leq d_0 \mid \Omega; \phi)] = \log\left[\frac{P(D_{i0} \leq d_0 \mid \Omega; \phi)}{1 - P(D_{i0} \leq d_0 \mid \Omega; \phi)}\right] = \xi_{0d_0} + \sum_{\iota=1}^{\iota=q} \xi_{1\iota} Z_{i\iota}, \qquad (4.1)$$

where the coefficient estimates $\hat{\xi} = \{\hat{\xi}_{0d_0}, \hat{\xi}_{1\iota}\}$ are listed in Table 4.6.

Table 4.6: Parameter estimates for the initial dosage assignment D_0 model - Male

Covariate	Notation	Coef.Estimate	Std.Error	OR	95% CI	p
Age	Z_{i1}	-0.03	0.00	0.97	(0.97, 0.97)	<0.001 *
Fall related drugs use	Z_{i2}	0.12	0.02	1.13	(1.08, 1.18)	<0.001 *
Sex hormone	Z_{i3}	-0.11	0.08	0.89	(0.76, 1.05)	0.165
Prev. BP duration 0-1 year (vs. 0 days)	Z_{i4}	-0.20	0.04	0.82	(0.76, 0.88)	<0.001 *
Prev. BP duration 1-3 years (vs. 0 days)	Z_{i5}	-0.05	0.06	0.95	(0.85, 1.07)	0.383
Prev. BP duration 3-5 years (vs. 0 days)	Z_{i6}	-0.06	0.09	0.95	(0.79, 1.13)	0.537
Prev. BP duration $\geq$ 5 year (vs. 0 days)	Z_{i7}	0.07	0.09	1.07	(0.90, 1.28)	0.432
Inhale GC	Z_{i8}	0.16	0.02	1.18	(1.14, 1.22)	<0.001 *
Thiazide	Z_{i9}	-0.19	0.02	0.82	(0.79, 0.86)	<0.001 *
Arthritis	Z_{i10}	-0.72	0.03	0.49	(0.46, 0.52)	<0.001 *
Fall related conditions	Z_{i11}	-0.21	0.02	0.81	(0.78, 0.84)	<0.001 *
Emergency or hospitalization	Z_{i12}	0.30	0.02	1.35	(1.29, 1.40)	<0.001 *
Prev. fracture	Z_{i13}	-0.18	0.08	0.83	(0.71, 0.97)	0.023 *
Intercept $d_0 \leq 1 \mid d_0 \geq 2$	ξ_{01}	-2.84	0.03			<0.001 *
Intercept $d_0 \leq 2 \mid d_0 \geq 3$	ξ_{02}	-1.55	0.03			<0.001 *
Intercept $d_0 \leq 3 \mid d_0 \geq 4$	ξ_{03}	0.17	0.02			<0.001 *
Intercept $d_0 \leq 4 \mid d_0 \geq 5$	ξ_{04}	1.01	0.03			<0.001 *

The columns corresponds to covariate, notation, coefficient estimates, standard error of coefficient estimates, odds ratio (OR) estimates, 95% confidence interval of OR, and p-value.

Alternatively, a forest plot can provide an efficient visualization to check whether the confidence interval (CI) bands cross the threshold value of 1 for odds ratios or hazard ratios. Several baseline covariates are predictive for initial dose prescription for males. Some factors increase the odds of lower initial dose prescription, such as fall-related drug use, inhale GC use, or visit an emergency unit or hospitalization. In comparison, other factors decrease the odds of lower initial dose prescriptions, such as less than one year of previous BP treatment, thiazide, arthritis, fall-related conditions, and a previous fracture. As an interpretation example, patients being previously hospitalized or have visited an emergency unit at baseline increase the odds of receiving a low initial dose by 1.35 folds. On the other hand, having arthritis at baseline would likely decrease the odds of having a low initial dose compare to those who do not have arthritis by 51% (i.e., patients with baseline arthritis condition tend to have higher initial GC dose). The interpretations for these predictive variables stay the same no matter which initial dose caegories are being compared, e.g. $d \leq 1$ vs. $d \geq 1$ or $d \leq 2$ vs. $d \geq 2$, etc., due to the proportional odds assumption.

Also, the predictive factors for female patients' initial dose assignment do not differ drastically from those for males. According to Appendix Figure B.1, the effect of previous long term ($\geq$ 5 years) BP treatment and baseline sex hormone are additionally predictive for female. Female patients with previous long term BP increase the odds of getting a low initial GC dose by 1.20 folds, while those treated with sex hormone decrease the odds of getting a low dose

by 13%.

The D_{i0} data set used for fitting the initial dose model has one row per patient, which has a size of $40,620$ for males and $46,222$ for females. On the other hand, the exposure data sets for subsequent multiple visit MPP models are in long-format, so we need to use robust sandwich standard error due to data replications and unobserved heterogeneity. We apply the R function `infjack.glm` for our visiting model, and we use the R package `sandwich` for the two multiordinal dosage models to derive the corresponding robust standard error.

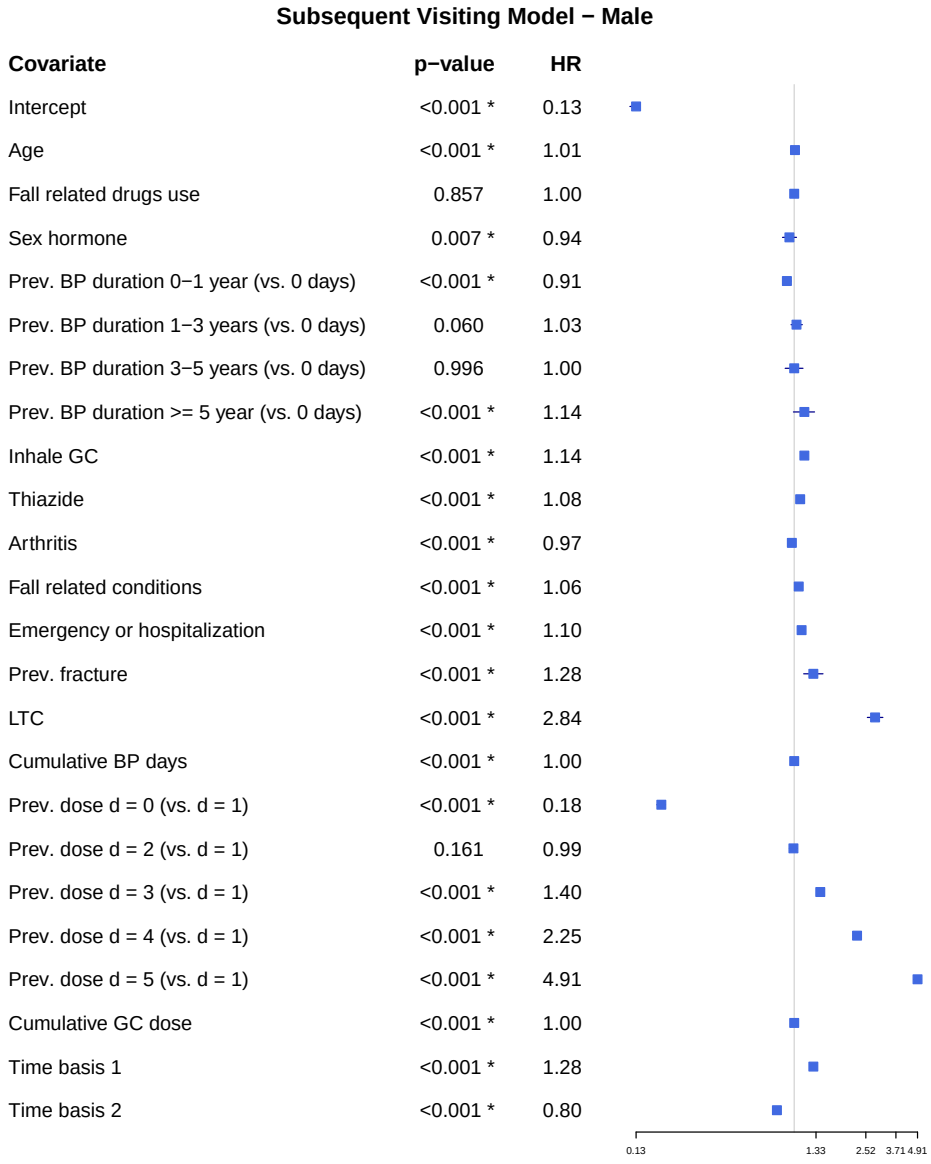


Figure 4.7: The forest plot of HR for the subsequent visit model V_{ik} for male.

The forest plot for conditional subsequent visiting models for male Figure 4.7 shows that

being in LTC at the previous five-day time point $k-1$ increases the odds of getting a GC dispensation at the current time-point k by 2.84 folds, and the effect of time-dependent BP treatment appears insignificant at predicting subsequent dispensation initiation. Furthermore, patients previously unexposed tend to stay unexposed, while patients previously on high dose group ($d \geq 3$) tend to have a new dispensation at the current time point. Besides, baseline covariates such as age, previous long term BP treatment, inhale GC, thiazide, falls-related conditions, emergency or hospitalization, and the previous fracture shows an increase in odds to get a subsequent visit/dispensation, while sex hormone, previous short term BP therapy and arthritis decrease the odds of getting a visit. For female patients, Appendix Figure B.2 shows similar results, except that having three to five years of previous BP therapy has slightly more significance in predicting a subsequent visit than having more than five years of previous BP therapy. Having diabetes at baseline also weakly increases the odds of receiving a subsequent dispensation.

We then present the subsequent D_{ik} models for our MPP. Among male patients who are unexposed at the previous time-point Figure 4.8, patients with baseline fall-related drug use, inhale GC use, emergency or hospitalization, and more than one year previous BP therapy increase the odds of receiving low dose GC treatment the current time k when there is a visit. All other time-fixed or time-dependent covariates increase the odds of receiving a higher dose, among which the first-degree time effect spline basis, arthritis, and LTC are the most significant.

Secondly, for those who are previous exposed with some dose $d \neq 0$ Figure 4.9, baseline inhale GC use, time effect spline bases, and on a high dose ($d \geq 3$) at time $k-1$ increase the odds of getting a high dose at k, while all other factors, except insignificant arthritis and cumulative BP days, increase the odds of receiving lower dose at k. We also found the higher the previous dose, the less likely a patient transfers to a low dose category at k, but those who are on maintenance dose tend to stay in low dose.

Combining the results from the three subsequent exposure models, we notice an interesting phenomenon for the time-dependent LTC status: patients in LTC tend to get GC dispensations; if they were not taking GC at $k-1$, then they tend to receive a higher dose GC dispensation at k; however, if they are already taking GC at $k-1$, then they have higher odds of receiving a lower dose.

Similar patterns can be observed from female D_{ik} models in Appendix Figure B.3 and Appendix Figure B.4.

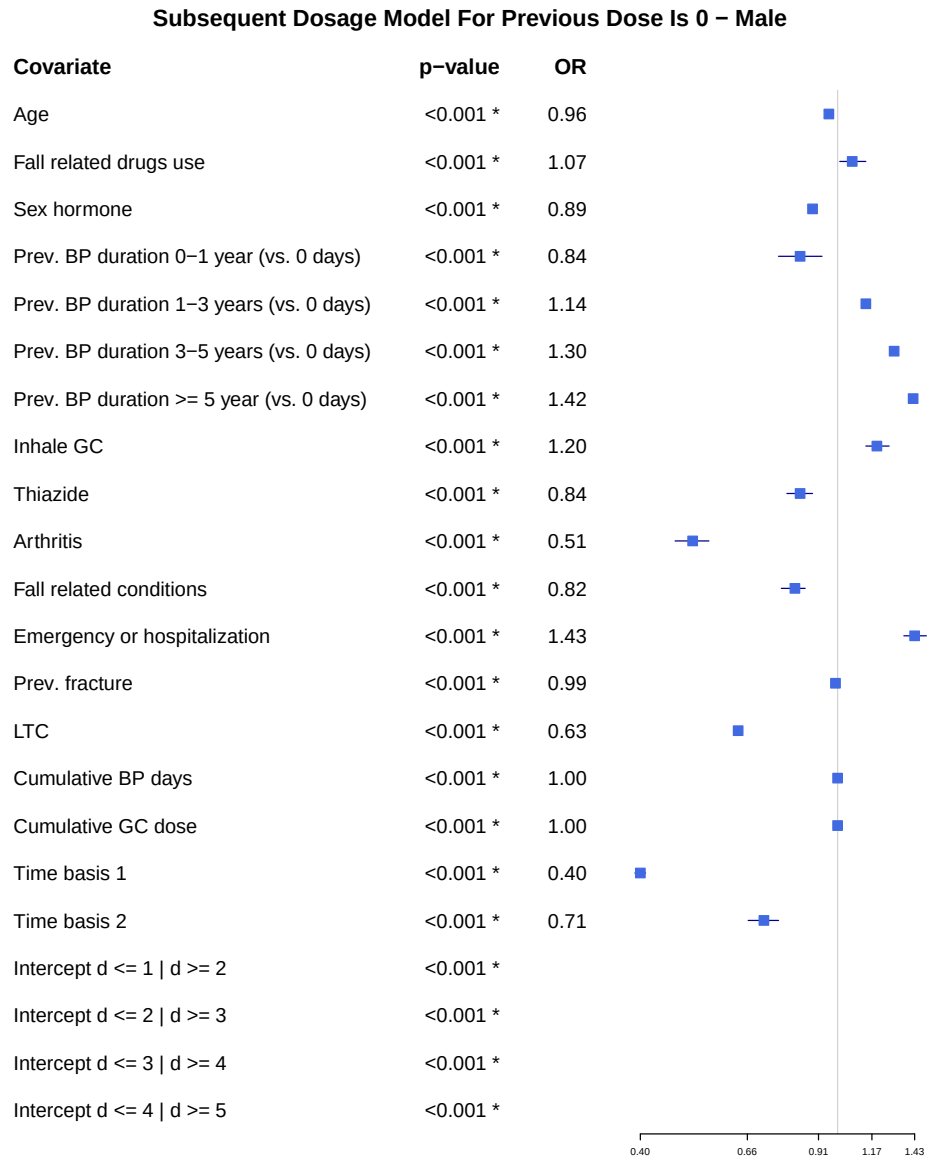


Figure 4.8: The forest plot of OR for the subsequent dose model D_{ik}, where at previous dose $D_{i,k-1}$, the dose level is unexposed for male.

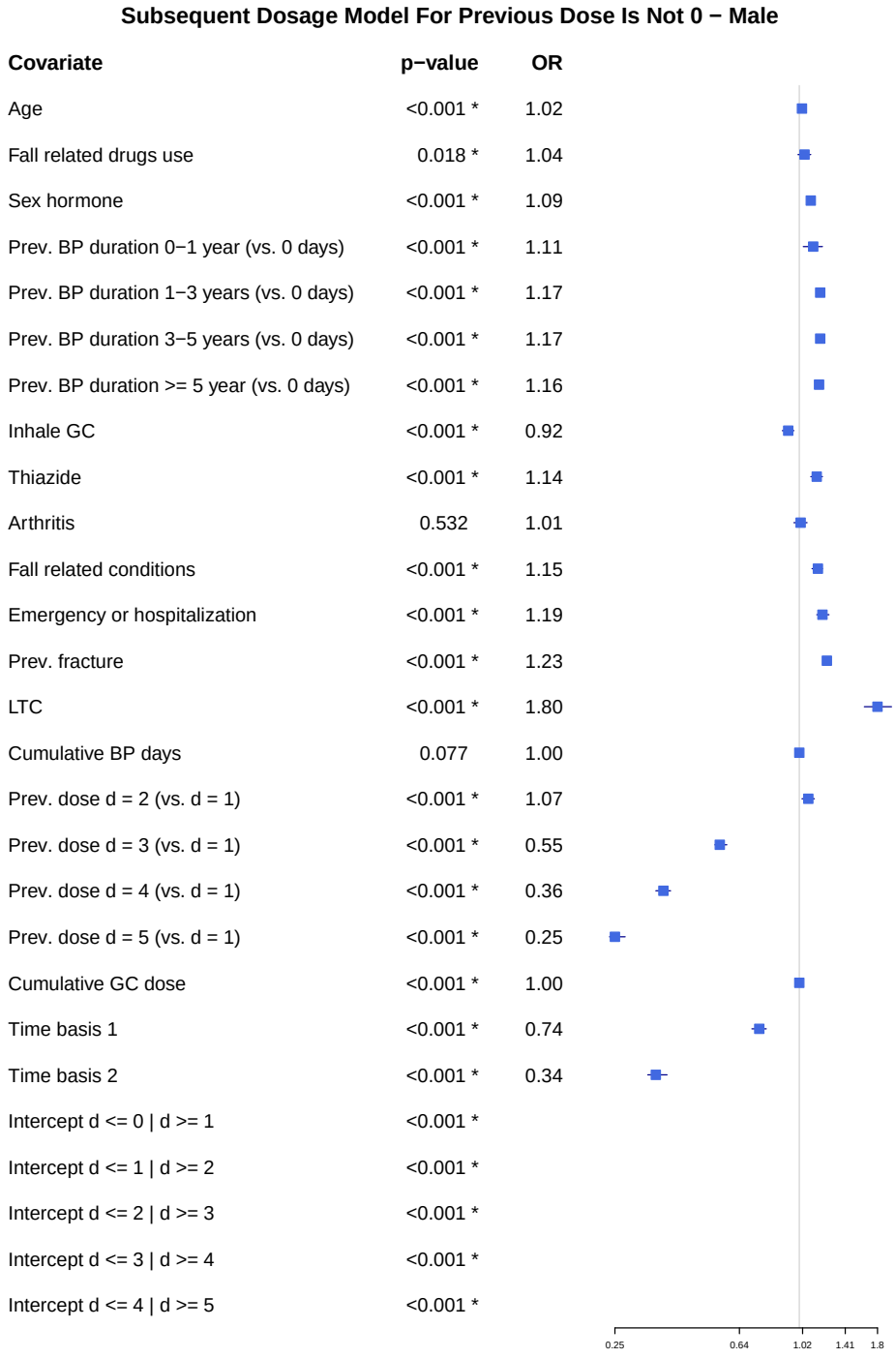


Figure 4.9: The forest plot of OR for the subsequent dose model D_{ik}, where at previous dose $D_{i,k-1}$, the dose level is unexposed for male.

4.4 Combined IPTWs and checking covariate balance

After deriving the IPTWs by predicting the marginal and conditional propensity scores on the discrete-time exposure data, the weights were accumulated up to the interval previous to the time point being considered in the outcome model to ensure correct ordering of time-varying exposure and covariate status. We then obtain the cumulative products of visiting weights over time. The stabilized weighting function is the ratio between each pair of the marginal and the conditional models as specified in (Eq. 2.32).

Similar to risk set sampling, some patients with short person-time contribution are not selected by case-base sampling, so the final number of selected individuals are 38,065 (out of 40,620) and 45,360 (out of 46,222) for male and female, respectively. The outcome data set is constructed using case-base sampling of size 804 $\times$ 200 + 804 = 161,604 rows for male and 1,835 $\times$ 200 + 1835 = 368,835 rows for female. For any time point in the outcome data set, we calculate the corresponding five-day time interval that it falls into.

Then, we match the combined stabilized weights from the discrete-time exposure data set over to the sampled time points in the case-base long format continuous-time outcome data set. We can plot the density distribution of weights for exposure A_{ik} value, and we also plot the matched weights from each of the exposure models, D_{i0}, V_{ik} and D_{ik}, for diagnostic purpose. As shown in Figure 4.10, the top left graph is the overall distribution of all combined stabilized weights in the outcome data set for male, and it is highly skewed to the right because of extreme time-dependent V_{ik} weights (bottom left panel). We decide to truncate at 0.5% and 99.5% on both ends, and the resulting weights are within the range of approximately $(0.6, 2.1)$ as shown in the top right graph. The weights from D_{ik} are well centered at 1 with a slight positive skew (bottom right panel). Because the marginal and conditional models are the same for the initial dose model, the stabilized weights for D_{i0} are exactly all at 1. We see similar pattern for female Appendix Figure B.5, except the outliers in time-dependent visiting weights are much more extreme than those for male, which eventually result in a wider range of weights $(0.6, 3.0)$ after truncation.

We further investigate the trajectory of the truncated combined stabilized weights over time Figure 4.11. For both sex, the median of the weights lie on 1 over time, but the variability of the weights gradually increase, which can be seen from the 90% interval lines. Because of truncation, the minimum and maximum weights stays the same at later time-points. The wider range of weights in female might due to higher proportion of female patients ever had a BP therapy during follow-up as a time-dependent confounder.

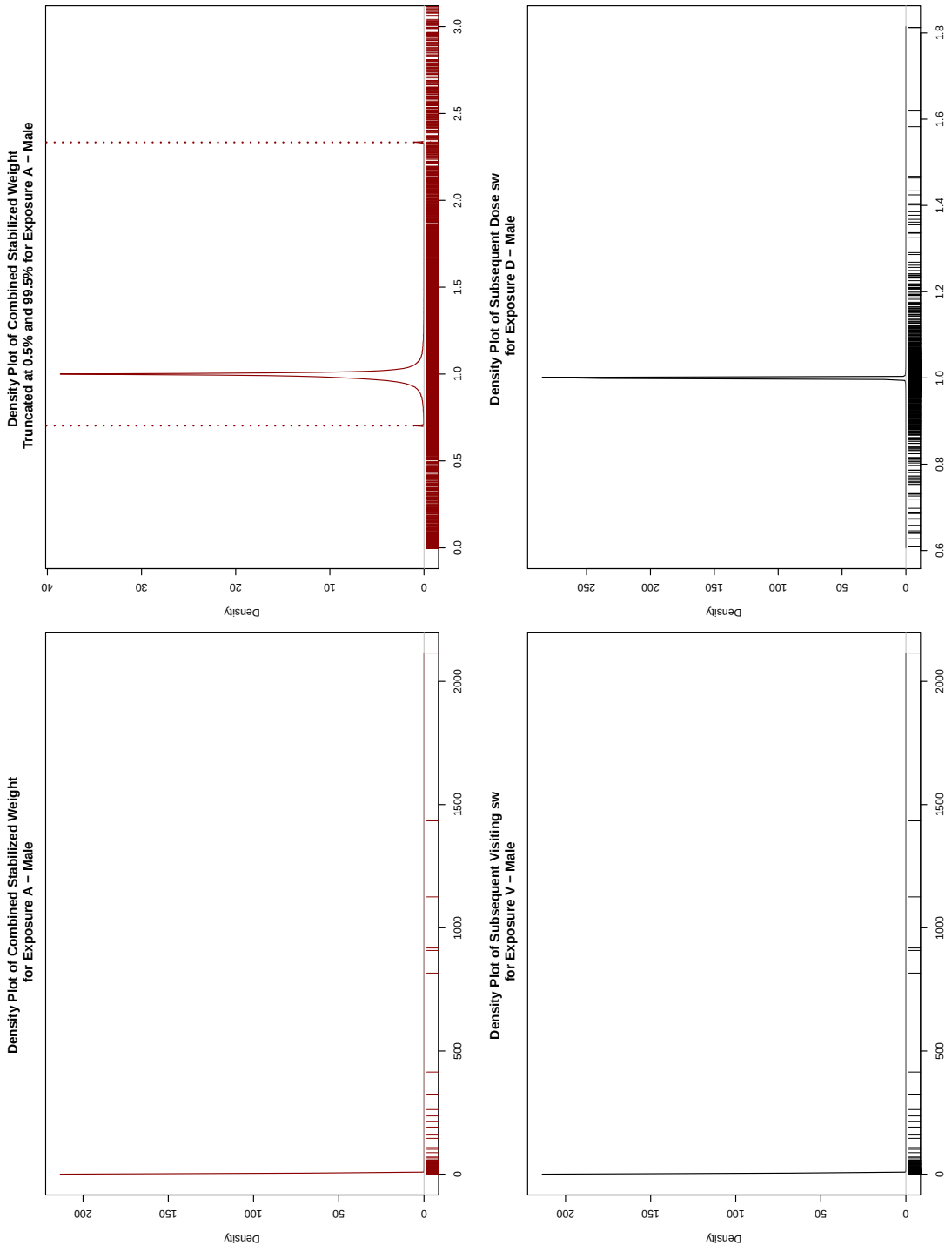


Figure 4.10: Density Plot of IPTW for Multiple Visit MPP - Male

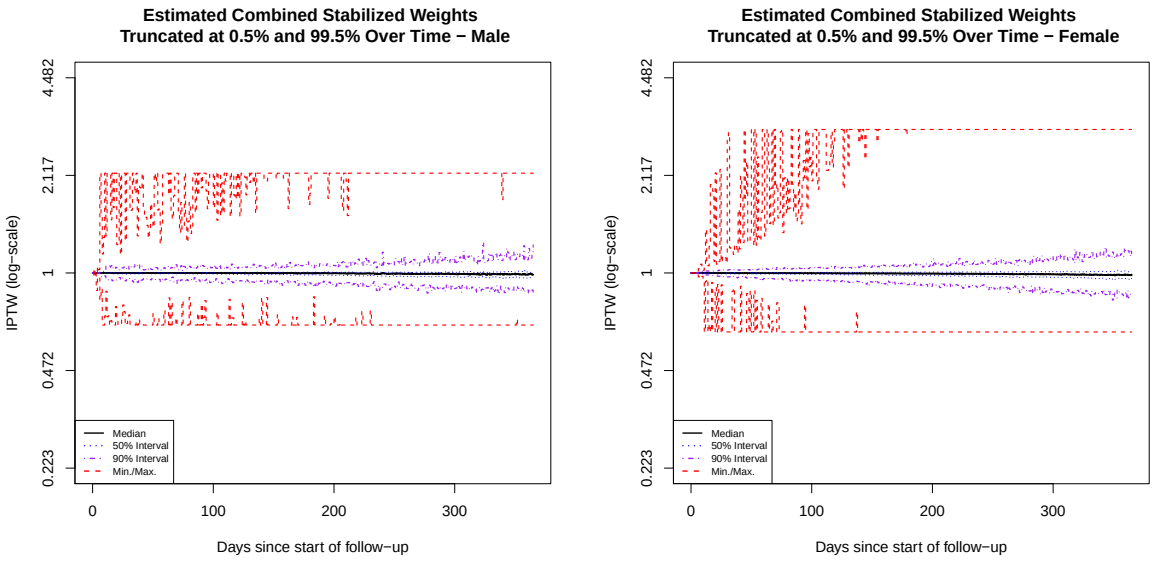


Figure 4.11: Truncated stabilized weights as a function of time for both genders

We then assess the covariate balance with and without weightings between different dosage groups for each sex. Whether the distributions of a confounder are balanced in different dosage groups can be assessed by using a t-test for continuous variables or a Chi-squared test for categorical variables. However, these methods are reliant on sample size, and they may return falsely significant results in large sample size studies (Lanza et al., 2013). Therefore, we consider using standardized mean difference (SMD, or Cohen's d) as a metric for assessing balance (Cohen, 2013). For a binary covariate, SMD is the pair-wise difference in the treatment group means divided by the pooled standard deviation:

$$\text{SMD} = \frac{\bar{X}_a - \bar{X}_b}{\sqrt{\frac{s_a^2 + s_b^2}{2}}} \tag{4.2}$$

For a categorical variable with more than two levels, the Mahalanobis distance is returned as a generalization of SMD (Yang and Dalton, 2012). A covariate is considered balanced in different treatment groups if the SMD is less than an effect size value of 0.2 (Cohen, 2013).

For diagnostic purpose only, the marginal model for the initial dose model D'_{i0} is fitted differently than that of our D_{i0} exposure model introduced in the previous section, because now we are interested in adjusting baseline covariates through weighting. Therefore, we keep only the intercept for our marginal model, while the conditional model remains the same as (Eq. 2.31). The implementation of SMD plots before and after weighting are adapted from Tang et al. (2020), using the packages `survey` and `TableOne` in R.

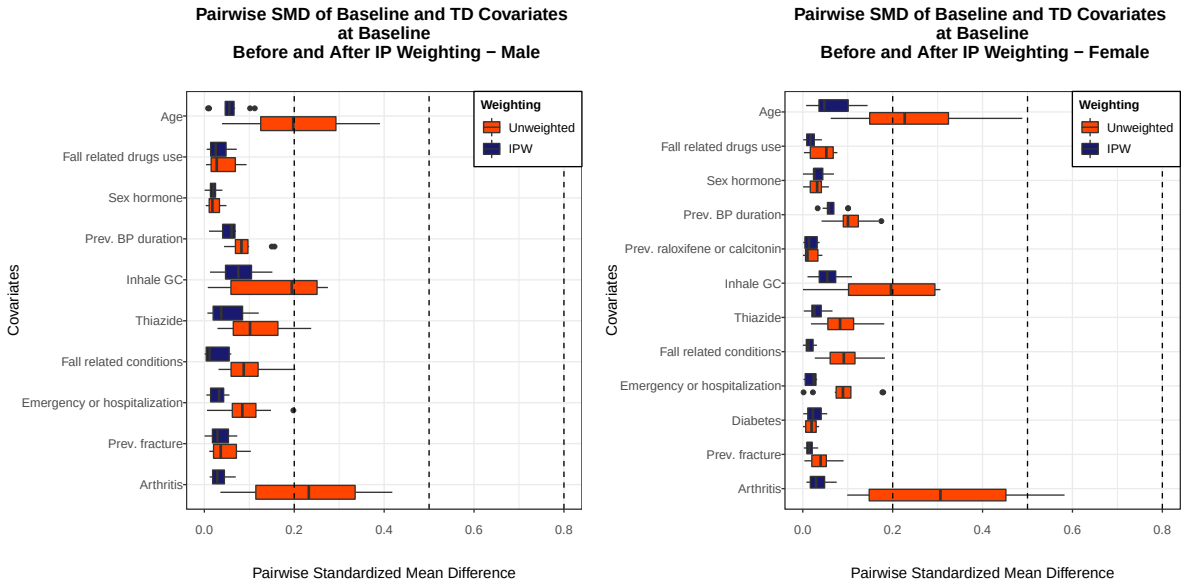


Figure 4.12: Standardized Mean Difference (SMD) plot at baseline for both genders

In Figure 4.12, most of our unweighted covariates are already well-balanced with a SMD below the threshold 0.2. After the weighting, we see notable improvement for covariates age, inhale GC and arthritis in both genders, and thus balance is achieved for all weighted covariates at the baseline.

We also attempt checking effect of covariate balance using the truncated combined stabilized time-varying weights in Appendix Figure B.6. Previously from Figure 4.3, the first three months of follow-up appears to be crucial for fracture incidence. Therefore, we split our case-base outcome model into three sections based on if the sampled time-point falls in: i) (0,100] days, ii) (100, 183] days, or iii) (183, 365] days. However, the time-varying weights across these three time points appear to dilute the effect of baseline weighting. In addition, our time-dependent confounders, LTC and cumulative BP duration, are weak confounders due to infrequent onset or prescriptions, so the improvements from weighting are not obvious. Nevertheless, the balance assessment for time-dependent confounders remains a methodological challenge, and the technical details require more consideration.

4.5 Models for GC usage-fracture outcome relationship

Consider the simplest models without interaction with time, we maximize the weighted case-base partial likelihood by fitting one marginal hazard model for dosage level, with $a = 1$ as the reference ($A_i(t)$ takes the same value as D_{ik} mapped from discrete-time exposure data, for $t \in (5(k-1), 5k]$) (Eq. 4.3) and another marginal hazard model for log-linear effect of

cumulative dose (Eq. 4.4). We can also consider modelling the effect of cumulative dose through a flexible quadratic spline with $p = 1, 2$. (Eq. 4.5).

$$\log \psi_i(t;\theta) = \sum_p \theta_{0p}(t) + \sum_{\iota=1}^{\iota=q} \theta_{1\iota} Z_{i\iota} + \sum_{a\in\{0,2,...,5\}} \theta_{2a} \mathbf{1}\{A_i(t) = a\} \tag{4.3}$$

$$\log \psi_i(t;\theta) = \sum_p \theta_{0p}(t) + \sum_{\iota=1}^{\iota=q} \theta_{1\iota} Z_{i\iota} + \theta_2 \text{cum}(\bar{a}) \tag{4.4}$$

$$\log \psi_i(t;\theta) = \sum_p \theta_{0p}(t) + \sum_{\iota=1}^{\iota=q} \theta_{1\iota} Z_{i\iota} + \sum_p \theta_{2p} \text{Basis}(\text{cum}(\bar{a}))_p \tag{4.5}$$

Like the stabilized weights, cumulative dose are first calculated in the discrete-time exposure data set and then matched over to the sampled time points in the continuous-time outcome data set. In the exposure data set, we take the median daily dose value of all dispensations within each dose category, and the resulting daily dose values are 0mg, 5mg, 10mg, 20mg, 34mg and 50mg for $d \in \{0, 1, ..., 5\}$, respectively. The cumulative dose is calculated by accumulating number of days exposed (a constant of 5) $\times$ the derived median daily dose value at each interval over a patient's follow-up period.

Table 4.7: Parameter estimates for the unweighted dose level outcome model - Male

Covariate	Notation	Coef.Estimate	Std.Error	HR	95% CI	p
Intercept	η_0	-10.13	0.20	0.00	(0.00, 0.00)	<0.001 *
Baseline hazard basis 1 (cont.)	θ_{01}	-0.14	0.24	0.87	(0.54, 1.41)	0.578
Baseline hazard basis 2 (cont.)	θ_{02}	-0.09	0.13	0.92	(0.72, 1.17)	0.491
Age (cont.)	Z_{i1}	0.05	0.01	1.05	(1.04, 1.06)	<0.001 *
Fall related drugs use (yes vs. no)	Z_{i2}	-0.09	0.09	0.91	(0.76, 1.09)	0.314
Sex hormone (yes vs. no)	Z_{i3}	0.69	0.23	1.99	(1.26, 3.15)	0.003 *
Prev. BP duration 0-1 year (vs. 0 days)	Z_{i4}	0.05	0.14	1.05	(0.80, 1.40)	0.711
Prev. BP duration 1-3 years (vs. 0 days)	Z_{i5}	0.38	0.20	1.46	(0.99, 2.15)	0.055
Prev. BP duration 3-5 years (vs. 0 days)	Z_{i6}	0.62	0.25	1.86	(1.14, 3.03)	0.013 *
Prev. BP duration $\geq$ 5 year (vs. 0 days)	Z_{i7}	0.05	0.31	1.05	(0.57, 1.91)	0.882
Inhale GC (yes vs. no)	Z_{i8}	-0.13	0.07	0.88	(0.76, 1.01)	0.079
Thiazide (yes vs. no)	Z_{i9}	0.19	0.08	1.21	(1.03, 1.43)	0.022 *
Arthritis (yes vs. no)	Z_{i10}	0.09	0.11	1.10	(0.89, 1.36)	0.384
Fall related conditions (yes vs. no)	Z_{i11}	0.31	0.08	1.36	(1.17, 1.57)	<0.001 *
Emergency or hospitalization (yes vs. no)	Z_{i12}	0.38	0.09	1.46	(1.22, 1.76)	<0.001 *
Prev. fracture (yes vs. no)	Z_{i13}	1.58	0.14	4.84	(3.66, 6.39)	<0.001 *
Treatment $A = 0$ (vs. $A = 1$)	$\mathbf{1}\{A_i(t) = 0\}$	-0.02	0.15	0.98	(0.73, 1.33)	0.908
Treatment $A = 2$ (vs. $A = 1$)	$\mathbf{1}\{A_i(t) = 2\}$	-0.04	0.20	0.96	(0.65, 1.42)	0.832
Treatment $A = 3$ (vs. $A = 1$)	$\mathbf{1}\{A_i(t) = 3\}$	0.28	0.18	1.32	(0.92, 1.89)	0.126
Treatment $A = 4$ (vs. $A = 1$)	$\mathbf{1}\{A_i(t) = 4\}$	0.26	0.25	1.30	(0.80, 2.11)	0.284
Treatment $A = 5$ (vs. $A = 1$)	$\mathbf{1}\{A_i(t) = 5\}$	-0.02	0.29	0.98	(0.56, 1.72)	0.944

The marginal hazard models are fitted with and without the weights, in both settings, adjusting for the baseline covariates. Because our IPTWs only adjust for the weak time-dependent confounders, the effect of weighting is not apparent. The weighted Table 4.8 and unweighted Table 4.7 model outputs has minimal difference in terms of the coefficient estimates, which suggest BP and LTC are weak time-dependent confounders. Comparing across different models, all three models returns similar results, and we demonstrate the similarities by presenting Table 4.8 with Appendix Table B.1 and Appendix Table B.2. We also noticed that the cumulative GC dose spline is only significant at the first degree, so a linear spline basis would be equivalent to directly modeling cumulative dose as a continuous variable and would be enough to capture the flexibility.

Table 4.8: Parameter estimates for the weighted dose level outcome model - Male

Covariate	Notation	Coef.Estimate	Std.Error	HR	95% CI	p
Intercept	η_0	-10.15	0.20	0.00	(0.00, 0.00)	<0.001 *
Baseline hazard basis 1 (cont.)	θ_{01}	-0.16	0.24	0.85	(0.53, 1.37)	0.498
Baseline hazard basis 2 (cont.)	θ_{02}	-0.08	0.12	0.93	(0.73, 1.18)	0.547
Age (cont.)	Z_{i1}	0.05	0.01	1.05	(1.04, 1.06)	<0.001 *
Fall related drugs use (yes vs. no)	Z_{i2}	-0.10	0.09	0.91	(0.76, 1.09)	0.290
Sex hormone (yes vs. no)	Z_{i3}	0.75	0.23	2.12	(1.36, 3.31)	0.001 *
Prev. BP duration 0-1 year (vs. 0 days)	Z_{i4}	0.07	0.14	1.07	(0.81, 1.42)	0.632
Prev. BP duration 1-3 years (vs. 0 days)	Z_{i5}	0.36	0.20	1.43	(0.97, 2.11)	0.069
Prev. BP duration 3-5 years (vs. 0 days)	Z_{i6}	0.70	0.24	2.01	(1.25, 3.21)	0.004 *
Prev. BP duration $\geq$ 5 year (vs. 0 days)	Z_{i7}	-0.03	0.31	0.97	(0.52, 1.78)	0.911
Inhale GC (yes vs. no)	Z_{i8}	-0.12	0.07	0.89	(0.77, 1.02)	0.090
Thiazide (yes vs. no)	Z_{i9}	0.17	0.08	1.19	(1.01, 1.40)	0.038 *
Arthritis (yes vs. no)	Z_{i10}	0.11	0.11	1.12	(0.91, 1.38)	0.299
Fall related conditions (yes vs. no)	Z_{i11}	0.30	0.07	1.36	(1.17, 1.57)	<0.001 *
Emergency or hospitalization (yes vs. no)	Z_{i12}	0.38	0.09	1.47	(1.22, 1.76)	<0.001 *
Prev. fracture (yes vs. no)	Z_{i13}	1.57	0.14	4.79	(3.63, 6.31)	<0.001 *
Treatment $A = 0$ (vs. $A = 1$)	$\mathbf{1}\{A_i(t) = 0\}$	0.01	0.15	1.01	(0.75, 1.37)	0.929
Treatment $A = 2$ (vs. $A = 1$)	$\mathbf{1}\{A_i(t) = 2\}$	0.00	0.20	1.00	(0.68, 1.47)	0.995
Treatment $A = 3$ (vs. $A = 1$)	$\mathbf{1}\{A_i(t) = 3\}$	0.28	0.18	1.33	(0.93, 1.90)	0.117
Treatment $A = 4$ (vs. $A = 1$)	$\mathbf{1}\{A_i(t) = 4\}$	0.29	0.25	1.34	(0.83, 2.17)	0.232
Treatment $A = 5$ (vs. $A = 1$)	$\mathbf{1}\{A_i(t) = 5\}$	0.01	0.29	1.01	(0.57, 1.77)	0.985

The columns corresponds to covariate, notation, hazard ratio (HR) estimates, 95% confidence interval, and p-value. 'cont.' means continuous variable.

From the model output for male patients Table 4.8, the daily dosage levels ($A = 3$) and ($A = 4$) show high but insignificant impact on fracture for male patients, compared to the lowest daily dose level ($A = 1$).

Several differences in estimated cause-specific fracture HRs are observed for female patients Table 4.9. A daily dosage level of $A = 3$ or $A = 4$ significantly increase the cause-specific fracture hazard for female, in which $A = 4$: 30 - 50mg has the highest increase of 1.65 folds.

Table 4.9: Parameter estimates for the weighted dose level outcome model - Female

Covariate	Notation	Coef.Estimate	Std.Error	HR	95% CI	p
Intercept	η_0	-9.52	0.13	0.00	(0.00, 0.00)	<0.001 *
Baseline hazard basis 1 (cont.)	θ_{01}	-0.02	0.16	0.98	(0.71, 1.34)	0.892
Baseline hazard basis 2 (cont.)	θ_{02}	-0.11	0.08	0.90	(0.76, 1.06)	0.197
Age (cont.)	Z_{i1}	0.06	0.00	1.06	(1.05, 1.06)	<0.001 *
Fall related drugs use (yes vs. no)	Z_{i2}	0.08	0.07	1.08	(0.95, 1.23)	0.231
Sex hormone (yes vs. no)	Z_{i3}	-0.19	0.09	0.83	(0.69, 0.99)	0.039 *
Prev. BP duration 0-1 year (vs. 0 days)	Z_{i4}	0.18	0.07	1.20	(1.05, 1.37)	0.007 *
Prev. BP duration 1-3 years (vs. 0 days)	Z_{i5}	0.30	0.07	1.36	(1.17, 1.57)	<0.001 *
Prev. BP duration 3-5 years (vs. 0 days)	Z_{i6}	0.21	0.09	1.23	(1.03, 1.48)	0.025 *
Prev. BP duration $\geq$ 5 year (vs. 0 days)	Z_{i7}	0.09	0.08	1.10	(0.94, 1.29)	0.251
Inhale GC (yes vs. no)	Z_{i8}	0.04	0.05	1.04	(0.95, 1.14)	0.435
Thiazide (yes vs. no)	Z_{i9}	0.07	0.05	1.07	(0.97, 1.19)	0.168
Arthritis (yes vs. no)	Z_{i10}	0.03	0.07	1.03	(0.91, 1.17)	0.620
Fall related conditions (yes vs. no)	Z_{i11}	0.20	0.05	1.23	(1.11, 1.35)	<0.001 *
Emergency or hospitalization (yes vs. no)	Z_{i12}	0.25	0.06	1.28	(1.15, 1.44)	<0.001 *
Prev. fracture (yes vs. no)	Z_{i13}	0.95	0.09	2.57	(2.17, 3.05)	<0.001 *
Prev. raloxifene or calcitonin (yes vs. no)	Z_{i14}	0.13	0.26	1.14	(0.69, 1.88)	0.620
Diabetes (yes vs. no)	Z_{i15}	-0.20	0.06	0.82	(0.73, 0.92)	0.001 *
Treatment $A = 0$ (vs. $A = 1$)	$\mathbf{1}\{A_i(t) = 0\}$	-0.05	0.09	0.95	(0.80, 1.13)	0.559
Treatment $A = 2$ (vs. $A = 1$)	$\mathbf{1}\{A_i(t) = 2\}$	-0.02	0.12	0.98	(0.77, 1.23)	0.848
Treatment $A = 3$ (vs. $A = 1$)	$\mathbf{1}\{A_i(t) = 3\}$	0.26	0.11	1.30	(1.05, 1.62)	0.016 *
Treatment $A = 4$ (vs. $A = 1$)	$\mathbf{1}\{A_i(t) = 4\}$	0.50	0.15	1.65	(1.24, 2.20)	0.001 *
Treatment $A = 5$ (vs. $A = 1$)	$\mathbf{1}\{A_i(t) = 5\}$	0.19	0.19	1.21	(0.84, 1.74)	0.315

We perform a likelihood ratio test for each of the weighted outcome models, comparing to their corresponding weighted null model, in which the treatment term is excluded. With the addition of any one of the three treatment terms, the full model fits the data set significantly better than their null model Appendix Table B.6. Similar result are shown in the likelihood ratio test table for females Appendix Table B.7.

We are also interested in the fracture hazard over time, both descriptively in the current data set and in terms of potential outcomes under always treated schemes at different doses. For male patients, Figure 4.13 top left panel shows the fracture hazard over time. The impact of GC manifests since the beginning of follow-up and has an overall trend of monotonically decreasing over time. The upper bound of the 90% interval curve shows slightly exaggerated increase beyond the 300^{th} day of follow-up. Due to our exceptionally rare event setting (2% and 4% cause-specific fracture incidence for male and female, respectively), the hazard appears minimal on the y-axis.

To answer the causal question of what could we potentially observe in fracture hazard had the patients been always treated with each dosage level over time, we create the potential outcome data sets always treated with each of the five dosage levels and predict the potential hazards with our fitted marginal hazard models.

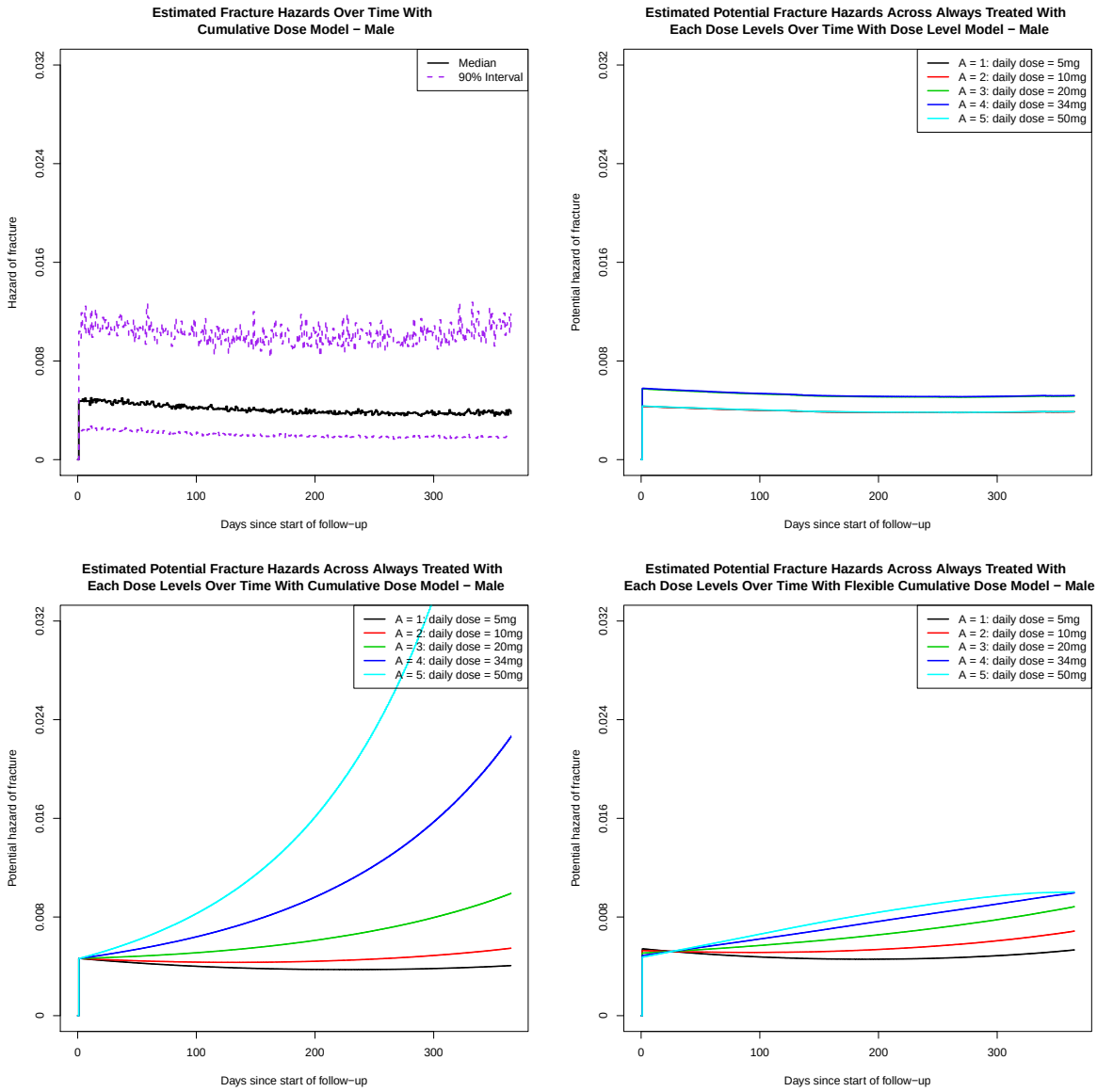


Figure 4.13: Descriptive and potential hazard under always treated with each dose

Figure 4.13 top right panel shows the potential hazard curves if the male patients in our study population are always treated with each dose level. Contrary to our belief that the higher dose would leads to higher hazard, the second highest dose level ($A = 4$) results in the highest cause-specific fracture hazard, graphically overlapped with the median dose level ($A = 3$) curve. The cause-specific fracture hazard curve of the highest dose group ($A = 5$) is overlapped with the curves of the low dose groups (daily dose $\leq$ 10mg). With the two cumulative dose models in Figure 4.13 bottom panels, we see the expected result that the higher the dose level, the higher the cumulative dose, and the higher the hazard. The cause-specific hazard grows exponentially since day 1 if the patients are always on high dose, so

this explains why GC treatment tapering (transfer from a high initial dose to a maintenance dose) should be recommended in future CTS protocols Table 4.4. In practice, it is unlikely for a patient to be always on high dose, so the log-linear effect assumption near the end of the follow-up is extrapolated. The flexible cumulative dose model provides a regularized representation of the cumulative dose model at later time points, and we still observe the same functional form of hazard over time.

Similar trends are observed for female patients Appendix Figure B.7, while female has slightly higher hazards compare to male, which might be due to a slightly higher event incidence rate in female. The cause-specific fracture hazard curves from the dose level model are more separated for female patients, and with flexible cumulative dose model, a heavier penalization is applied to the high dose groups at later time points.

Chapter 5

Discussion

5.1 Strengths

This thesis has extended the existing statistical methods on continuous-time marginal structural models by making both the exposure and the outcome models fully-continuous, combined with considering pharmacoepidemiological dispensations as a marked point process. Instead of the discrete-time causal assumptions on consistency, conditional exchangeability, and positivity, we used the continuous-time causal assumptions: stability, which implies no unmeasured confounder, and absolute continuity.

Our research context is administrative data, where various exposure, confounder, and death databases are linked through a unique patient ID. Therefore, we generated these processes independently in our simulation study, and our confounder process is naturally observed independently from the exposure process in our data application. To the best of our knowledge, we have performed comprehensive baseline data cleaning and clustering with proper clinical guidance to identify possible confounders.

Regarding absolute continuity, in our simulation study, we have experimented with initial parameter values to guarantee that there is at least one event for each combination (e.g. confounder process happens before visit process, confounder process happens after visit process, etc.) in each data set generated under 1,000 seeds. With the administrative data type, pharmacological dispensations continuously occur without incurring the missing data problem at discrete clinical visits. In our data application, we had frequent dispensations to ensure no positivity violation in the study population. Additionally, we have assigned treatment gaps as an unexposed state, with its own treatment probability. In general, positivity is a testable assumption; we verified this by checking that all the parameters in the exposure models can be estimated without identifiability issues, as well as checking the distribution of the resulting weights. These properties guarantee that the treatment intensities under observed $\mathcal{O}$ and

experimental $\mathcal{E}$ settings are always positive, which gives us absolute continuity.

The different combinations of IPTW estimators and the marginal hazard models are validated through simulation studies. Under several possible study designs that are leveraging treatment initiation or termination, we proposed the data generating mechanism through multistate models. As we expected, the discrete and continuous-time exposure models return similar parameter estimates when discretization is reasonably chosen, and the weighted case-base sampling partial likelihood can serve as a computationally efficient alternative to the well-established weighted Cox partial likelihood, while also being able to estimate absolute hazards.

For our real data analysis, we uncovered GC prescription patterns over time descriptively. The lack of clear guidelines on GC duration cleaning motivates us to propose the novel data-driven criteria for dealing with treatment overlaps ($\leq$30 days push the overlaps forward, $>$ 30 days truncate the current dispensation and advance the next dispensation), and we have accounted for the treatment gaps by adding an unexposed level. To date, we are the first to construct GC episode on a non-overlapping individual level without any aggregation, which could lead to a better approximation to the real usage. On the other hand, previous literature has suggested the dispensation days-supply values are often inaccurate for BP in the Ontario Drug Benefits database (Burden et al., 2013). We have cleaned the days-supply values for BP with peer-reviewed guidance and taken into account the quantity dispensed. Once again, we emphasize that days-supply should not be relied on solely when cleaning the exposure data, as it tends to give an underestimated treatment duration. Besides, we have proposed a novel conversion metrics for alendronate equivalent BP treatment duration.

We used case-base partial likelihood for fitting our outcome model. Apart from smaller data size, another advantage of case-base sampling is its handling of tied event times in the outcome model. Many fracture event times are tied across individuals, but ties should not have any practical implications for the case-base sampling partial likelihood. Unlike Cox's partial likelihood, which conditions on the risksets, we can shift the times in case-base by infinitesimal amount without changing the likelihood. Therefore, case-base sampling can handle tied event times during a relatively short follow-up period. Besides, case-base sampling can also be used to fit absolute hazards, and thus can be used to calculate predictions and prediction intervals. These properties can further be used to calculate cumulative incidences as probabilities, and thus are easier to interpret as causal quantities. However, marginal hazards as causal quantities need to be interpreted with caution due to them involving both causal effects and selection.

Finally, we applied our exposure model to the administrative data to generate combined stabilized weights, identified significant risk factors for cause-specific fracture hazard for male and female separately after removing time-dependent confounders, and provided potential

fracture hazard over time under always treated regime with each GC daily dose level. Here, we discuss some limitations to our methodology and provide some suggestions for future direction.

5.2 Limitations

In our simulation study, we only considered counting to the first exposure event, while in reality there could be recurrent exposure episodes for a patient. We simplified the scenario in simulation studies so that the true parameter value could be calculated. However, if the true parameter value can be obtained in high dimensional settings, simulation studies could be conducted for recurrent exposures to check the MPP weighting performance. We also did not investigate the effect of discretization on continuous time scale on parameter estimation, although a recent study has validated our hypothesis that coarse discretization would result in bias (Ferreira Guerra et al., 2020).

In our real data analysis, different definitions of GC exposure might result in a divergent estimation of drug effects. The threshold criteria on 30-day (monthly) overlap proposed by us should be investigated further by clinical scientists regarding GC's pharmacokinetics. In terms of causal assumptions, inevitably, there might be underlying unmeasured confounders that we fail to access in some other electronic healthcare databases, and the assumption of no unmeasured confounders is not testable. In addition, in both our simulation study and the real data analysis, we assumed there is only administrative censoring.

The data generation mechanism we adopt in the simulation study guarantees there will be no misspecification. However, the exposure and outcome models in the data analysis require more consideration. We did not use a continuous-time model, such as Cox or Poisson, for our visiting process V_{ik}, due to our study's large sample size (86,842 patients in total). For example, using the Cox model would have to generate an enormous data set (number of rows $\approx$ 86,842 $\times$ 365 minus a few non-administratively censored person-times). Even with our current five-day discretization, we ended up with 5,721,411 rows. Besides, the five-day interval captures most of the days-supply values, but there are still around 6% short (daily to four-day) prescriptions being smoothed out in the approximation.

Furthermore, a surprising amount ($\sim$ 13%) of our study population experienced death as a competing risk. In the current analysis, we have only presented the cause-specific fracture hazard. Patients in both death and fracture subgroups took more high dose treatment early on during the follow-up (graphs are not shown due to small cells), and the peaks of the event time density curves are overlapped at around 30^{th} day for death and fracture subgroups, and high mortality can affect the estimation of dose-effect.

Moreover, due to our study's large sample size and high dispensation rates, the p-values tend to be significant by chance for the exposure models. Large sample studies usually give small standard errors for coefficient estimates, so the 95% confidence interval is less likely to cross 1 for estimates close to 1.

BP's effect in preventing fracture is often delayed because it persists in bone long after the therapy has stopped. So far, it has been difficult to quantify how much shifting we should address when modeling. On the other hand, GC's effect manifests rapidly within three months after the chronic users' cohort entry. Therefore, we suspect the therapeutic effect of BP prescribed during our one-year follow-up was not exhibited before the GC-induced fracture took place. Thus, BP therapy appears to be a weak time-dependent confounder.

In general, it is possible to have model misspecified in real data analysis. Our current results suggest the highest dose level does not result in the highest cause-specific fracture hazard, because the model did not enforce the monotonicity of the dose effect. Our current outcome model also assumed proportionality of the treatment effects, and such proportionality assumption can be relaxed by adjusting for an interaction term between time on treatment and treatment dosage. Methods discussed in this thesis are only singly robust, and doubly robust methods for longitudinal data have begun to receive more attention in the literature (Tran et al., 2019).

5.3 Outlook

Given how popular GC is used in both clinical trials for the emerging infectious diseases like SARS and COVID-19 (Zhang et al., 2020) and real-world setting as a steroid therapy for respiratory conditions and inflammatory arthritis, we strongly encourage the development of guidelines on how to adjust GC treatment duration overlaps in pharmacoepidemiological data, possibly separated for acute therapy and maintenance therapy.

We have assumed administrative censoring. If there were actual random censoring other than administrative censoring in the data, Inverse Probability Censoring (IPC) weights could be introduced to correct some biases, due to censoring explained by observed covariates. Modeling subdistribution fracture hazard would be an alternative to modeling the cause-specific hazard. Inspired by the recent study by Ryalen et al. (2020), we could further consider investigating the subdistribution hazards by constructing a different riskset (Austin et al., 2016). After that, cumulative incidence functions can be plotted to compare the two subgroups of death and fracture. However, modeling subdistribution hazard in the causal inference framework with time-dependent exposures may require further investigation.

One previous study has proposed using case-base sampling to derive continuous-time

weights for a continuous Cox outcome MSM (Saarela and Liu, 2016). With our data, using case-base sampling to construct the exposure data set would only result in 2,639 $\times$ 200 = 530,439 rows. We could replace our current discrete and conventional continuous-time exposure models with case-base sampling in both simulation and the data analysis as a future direction.

Several time-dependent confounders could be further adjusted, such as other OP drugs, visit to emergency department, and hospitalization. Only 0.3% of our study population ever had an other OP dispensation on either denosumab, raloxifene, or calcitonin, so we expect the confounding effect from these to be weak due to the infrequency. These drugs have different mechanisms for treating GC-induced osteoporosis and do not persist in bones after treatment ends, so we should not convert them into a single category and model each of them as on-or-off treatment separately. We list the days-supply imputation for each drug in (Appendix. Table. A.3). The days-supply values are determined, and since we do not convert them into the same type, we should shift all the subsequent exposure windows until the current dispense is finished within each drug. Furthermore, deriving a clinically meaningful definition for a visit to emergency and hospitalization during follow-up could be challenging for our statistical analysis because most patients are discharged on the same date as admission. Therefore, a time-dependent confounder exposure duration of zero could occur.

Other than unmeasured confounders, failing to adjust for the dosage level and time interaction term could also lead to outcome model misspecification. We considered the simplest outcome model to start with, but one future direction is to model the cumulative exposure through a time interaction term. We could also extend WCE to our outcome model, that is, weight the cumulative exposure effect by dispensation recency (Xiao et al., 2014).

To handle the possible outcome model misspecification, instead of using a flexible outcome model, we could consider constrain the outcome model to be monotonically decreasing, based on our current fracture hazard over time pattern. Doubly robust estimators combine IPTW and g-formula together, which allows us to get an unbiased estimator with either one of the exposure or the outcome model correctly specified (Hernan and Robins, 2020).

Even when modeling in continuous-time, we suggest horizontally shift the functional form of the BP effect forward. One future direction in study design is that we model the BP dispensations in the previous year, rather than during the current follow-up period. More generally, we can introduce latency periods to the modeling, both for the confounders and the exposure of interest.

Another methodological challenge is how to assess the time-varying covariate balance. We have attempted to split our outcome data set based on sampled time points, but a more continuous and systematic approach is needed.

Bibliography

Aalen, O., Borgan, O., and Gjessing, H. (2008). *Survival and event history analysis: a process point of view*. Springer Science & Business Media.

Aalen, O. O., Cook, R. J., and Roysland, K. (2015). Does cox analysis of a randomized survival study yield a causal treatment effect? *Lifetime data analysis*, 21(4):579–593.

Abadie, A., Diamond, A., and Hainmueller, J. (2010). Synthetic control methods for comparative case studies: Estimating the effect of californias tobacco control program. *Journal of the American statistical Association*, 105(490):493–505.

Album, J. (2014). Glucocorticoid-induced osteoporosis management among seniors. Master's thesis, University of Toronto.

Ali, M. S., Groenwold, R. H., Belitser, S. V., Souverein, P. C., Martin, E., Gatto, N. M., Huerta, C., Gardarsdottir, H., Roes, K. C., Hoes, A. W., et al. (2016). Methodological comparison of marginal structural model, time-varying cox regression, and propensity score methods: the example of antidepressant use and the risk of hip fracture. *pharmacoepidemiology and drug safety*, 25:114–121.

Amiche, M., Albaum, J., Tadrous, M., Pechlivanoglou, P., Lévesque, L., Adachi, J., and Cadarette, S. (2016). Fracture risk in oral glucocorticoid users: a bayesian meta-regression leveraging control arms of osteoporosis clinical trials. *Osteoporosis International*, 27(5):1709–1718.

Amiche, M. A. (2018). *Oral glucocorticoid: fracture risk and benefit of osteoporosis pharmacotherapy*. PhD thesis, University of Toronto.

Amiche, M. A., Lévesque, L. E., Gomes, T., Adachi, J. D., and Cadarette, S. M. (2018). Effectiveness of oral bisphosphonates in reducing fracture risk among oral glucocorticoid users: three matched cohort analyses. *Journal of Bone and Mineral Research*, 33(3):419–429.

Arjas, E., Keiding, N., Borgan, Ø., Andersen, P. K., and Natvig, B. (1989). Survival models and martingale dynamics [with discussion and reply]. *Scandinavian Journal of Statistics*, pages 177–225.

Austin, P. C., Lee, D. S., and Fine, J. P. (2016). Introduction to the analysis of survival data in the presence of competing risks. *Circulation*, 133(6):601–609.

Balasubramanian, A., Wade, S. W., Adler, R. A., Saag, K., Pannacciulli, N., and Curtis, J. R. (2018). Glucocorticoid exposure and fracture risk in a cohort of us patients with selected conditions. *Journal of Bone and Mineral Research*, 33(10):1881–1888.

Bartlett, J. (2019). Hazard ratios from randomised trials are causal. Lifetime Data Science: Foundations and Frontiers.

Beaulieu-Jones, B. K., Finlayson, S. G., Yuan, W., Altman, R. B., Kohane, I. S., Prasad, V., and Yu, K.-H. (2020). Examining the use of real-world evidence in the regulatory process. *Clinical Pharmacology & Therapeutics*, 107(4):843–852.

Becher, H. (2014). *Analysis of Continuous Covariates and Dose-Effect Analysis*, pages 1057–1086. Springer New York, New York, NY.

Bodnar, L. M., Davidian, M., Siega-Riz, A. M., and Tsiatis, A. A. (2004). Marginal structural models for analyzing causal effects of time-dependent treatments: an application in perinatal epidemiology. *American Journal of Epidemiology*, 159(10):926–934.

Boslaugh, S. (2007). *Encyclopedia of epidemiology*. Sage Publications.

Burden, A., Paterson, J., Solomon, D. H., Mamdani, M., Juurlink, D., and Cadarette, S. (2012). Bisphosphonate prescribing, persistence and cumulative exposure in ontario, canada. *Osteoporosis international*, 23(3):1075–1082.

Burden, A. M. (2014). *Lost in Translation: Exposure Misclassification when Relying on Days Supply in Pharmacy Claims Data*. PhD thesis, University of Toronto.

Burden, A. M., Huang, A., Tadrous, M., and Cadarette, S. M. (2013). Variation in the days supply field for osteoporosis medications in ontario. *Archives of osteoporosis*, 8(1-2):128.

Cadarette, S. M., Katz, J. N., Brookhart, M. A., Stürmer, T., Stedman, M. R., and Solomon, D. H. (2008). Relative effectiveness of osteoporosis drugs for preventing nonvertebral fracture. *Annals of internal medicine*, 148(9):637–646.

Cadarette, S. M., Levesque, L., Mamdani, M., Perreault, S., Juurlink, D. N., Paterson, J. M., Carney, G., Gunraj, N., Hawker, G. A., Tadrous, M., et al. (2013). Comparison of orally administered bisphosphonate drugs in reducing the risk of hip fracture in older adults: a population-based cohort study. *CMAJ open*, 1(3):E97.

Chalitsios, C. V., Shaw, D. E., and McKeever, T. M. (2020). A retrospective database study of oral corticosteroid and bisphosphonate prescribing patterns in england. *NPJ primary care respiratory medicine*, 30(1):1–8.

Cohen, J. (2013). *Statistical power analysis for the behavioral sciences*. Academic press.

Cole, S. R. and Hernan, M. A. (2008). Constructing inverse probability weights for marginal structural models. *American journal of epidemiology*, 168(6):656–664.

Commenges, D. (2019). Causality without potential outcomes and the dynamic approach. *arXiv preprint arXiv:1905.01195*.

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. (1975). Partial likelihood. *Biometrika*, 62(2):269–276.

Cranney, A., Guyatt, G., Griffith, L., Wells, G., Tugwell, P., Rosen, C., Group, O. M., and Group, O. R. A. (2002). Ix: Summary of meta-analyses of therapies for postmenopausal osteoporosis. *Endocrine reviews*, 23(4):570–578.

David, G. and Kleinbaum, K. (2016). *Survival analysis: a self-learning text*. Springer-Verlag New York.

Davies, J., Martinec, M., Delmar, P., Coudert, M., Bordogna, W., Golding, S., Martina, R., and Crane, G. (2018). Comparative effectiveness from a single-arm trial and real-world data: alectinib versus ceritinib. *Journal of comparative effectiveness research*, 7(09):855–865.

Dawid, A. P., Didelez, V., et al. (2010). Identifying the consequences of dynamic treatment strategies: A decision-theoretic overview. *Statistics Surveys*, 4:184–231.

De Vries, F., Bracke, M., Leufkens, H. G., Lammers, J.-W. J., Cooper, C., and Van Staa, T. P. (2007). Fracture risk with intermittent high-dose oral glucocorticoid therapy. *Arthritis & Rheumatism*, 56(1):208–214.

Dowd, R., Recker, R. R., and Heaney, R. (2000). Study subjects and ordinary patients. *Osteoporosis International*, 11(6):533–536.

Edsbacker, S. and Andersson, T. (2004). Pharmacokinetics of budesonide (entocort ec) capsules for crohns disease. *Clinical pharmacokinetics*, 43(12):803–821.

Fay, M. P. and Graubard, B. I. (2001). Small-sample adjustments for wald-type tests using sandwich estimators. *Biometrics*, 57(4):1198–1206.

Ferreira Guerra, S., Schnitzer, M. E., Forget, A., and Blais, L. (2020). Impact of discretization of the timeline for longitudinal causal inference methods. *Statistics in medicine*.

Folkestad, L., Hald, J. D., Ersboll, A. K., Gram, J., Hermann, A. P., Langdahl, B., Abrahamsen, B., and Brixen, K. (2017). Fracture rates and fracture sites in patients with osteogenesis imperfecta: a nationwide register-based cohort study. *Journal of Bone and Mineral Research*, 32(1):125–134.

Fosen, J., Ferkingstad, E., Borgan, O., and Aalen, O. O. (2006). Dynamic path analysisa new approach to analyzing time-dependent covariates. *Lifetime data analysis*, 12(2):143–167.

Ganesan, K., Bansal, P., Goyal, A., and Roane, D. (2020). Bisphosphonate.

Garbe, E. and Suissa, S. (2014). *Pharmacoepidemiology*, pages 1875–1925. Springer New York, New York, NY.

Ghassemi, M., Naumann, T., Schulam, P., Beam, A. L., Chen, I. Y., and Ranganath, R. (2020). A review of challenges and opportunities in machine learning for health. *AMIA Summits on Translational Science Proceedings*, 2020:191.

Greenland, S. (1995). Dose-response and trend analysis in epidemiology: alternatives to categorical analysis. *Epidemiology*, pages 356–365.

Greenland, S. (1996). Absence of confounding does not correspond to collapsibility of the rate ratio or rate difference. *Epidemiology*, pages 498–501.

Hanley, J. A. and Miettinen, O. S. (2009). Fitting smooth-in-time prognostic risk functions via logistic regression. *The International Journal of Biostatistics*, 5(1).

Hastie, T. J. and Tibshirani, R. J. (1990). *Generalized additive models*, volume 43. CRC press.

Havercroft, W. and Didelez, V. (2012). Simulating from marginal structural models with time-dependent confounding. *Statistics in medicine*, 31(30):4190–4206.

Hernan, M. A. (2010). The hazards of hazard ratios. *Epidemiology (Cambridge, Mass.)*, 21(1):13.

Hernán, M. Á., Brumback, B., and Robins, J. M. (2000). Marginal structural models to estimate the causal effect of zidovudine on the survival of hiv-positive men. *Epidemiology*, pages 561–570.

Hernan, M. A., Brumback, B. A., and Robins, J. M. (2002). Estimating the causal effect of zidovudine on cd4 count with a marginal structural model for repeated measures. *Statistics in medicine*, 21(12):1689–1709.

Hernan, M. A. and Hernandez-Diaz, S. (2012). Beyond the intention-to-treat in comparative effectiveness research. *Clinical trials*, 9(1):48–55.

Hernan, M. A., McAdams, M., McGrath, N., Lanoy, E., and Costagliola, D. (2009). Observation plans in longitudinal studies with time-varying treatments. *Statistical methods in medical research*, 18(1):27–52.

Hernan, M. A. and Robins, J. M. (2020). Causal inference: what if. *Boca Raton: Chapman & Hill/CRC*, 2020.

Holland, P. W. (1986). Statistics and causal inference. *Journal of the American statistical Association*, 81(396):945–960.

ICES (2017). Data dictionary: Ontario health insurance plan.

ICES (2020a). Data dictionary: Discharge abstract database.

ICES (2020b). Data dictionary: National ambulatory care reporting system, same day surgery.

Imai, K., Keele, L., and Tingley, D. (2010). A general approach to causal mediation analysis. *Psychological methods*, 15(4):309.

Kalbfleisch, J. D. and Prentice, R. L. (2011). *The statistical analysis of failure time data*, volume 360. John Wiley & Sons.

Lanza, S. T., Moore, J. E., and Butera, N. M. (2013). Drawing causal inferences using propensity scores: A practical guide for community psychologists. *American journal of community psychology*, 52(3-4):380–392.

Lau, B., Gange, S. J., and Moore, R. D. (2007). Interval and clinical cohort studies: epidemiological issues. *AIDS research and human retroviruses*, 23(6):769–776.

Lawless, J. F. (1987). Regression methods for poisson process data. *Journal of the American Statistical Association*, 82(399):808–815.

Lim, L. S., Fink, H. A., Kuskowski, M. A., Cauley, J. A., and Ensrud, K. E. (2005). Diuretic use and bone mineral density in older usa men: the osteoporotic fractures in men (mros) study. *Age and ageing*, 34(5):504–507.

Lin, H., Scharfstein, D. O., and Rosenheck, R. A. (2004). Analysis of longitudinal data with irregular, outcome-dependent follow-up. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, 66(3):791–813.

Lix, L. M., Azimaee, M., Osman, B. A., Caetano, P., Morin, S., Metge, C., Goltzman, D., Kreiger, N., Prior, J., and Leslie, W. D. (2012). Osteoporosis-related fracture case definitions for population-based administrative data. *BMC public health*, 12(1):301.

Lok, J. J. et al. (2008). Statistical modeling of causal effects in continuous time. *The Annals of Statistics*, 36(3):1464–1507.

Lougheed, M. D., Lemière, C., Dell, S. D., Ducharme, F. M., FitzGerald, J. M., Leigh, R., Licskai, C., Rowe, B. H., Bowie, D., Becker, A., et al. (2010). Canadian thoracic society asthma management continuum–2010 consensus summary for children six years of age and over, and adults. *Canadian Respiratory Journal*, 17.

Lougheed, M. D., Lemiere, C., Ducharme, F. M., Licskai, C., Dell, S. D., Rowe, B. H., FitzGerald, M., Leigh, R., Watson, W., Boulet, L.-P., et al. (2012). Canadian thoracic society 2012 guideline update: diagnosis and management of asthma in preschoolers, children and adults. *Canadian respiratory journal*, 19.

Lumley, T. and Heagerty, P. (1999). Weighted empirical adaptive variance estimators for correlated data regression. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, 61(2):459–477.

Mancini, L. and Paganoni, A. M. (2019). Marked point process models for the admissions of heart failure patients. *Statistical Analysis and Data Mining: The ASA Data Science Journal*, 12(2):125–135.

Martinussen, T. and Vansteelandt, S. (2013). On collapsibility and confounding bias in cox and aalen regression models. *Lifetime data analysis*, 19(3):279–296.

Merlotti, D., Gennari, L., Martini, G., and Nuti, R. (2009). Current options for the treatment of pagets disease of the bone. *Open Access Rheumatology: Research and Reviews*, 1:107.

Motulsky, H. and Christopoulos, A. (2004). *Fitting models to biological data using linear and nonlinear regression: a practical guide to curve fitting*. Oxford University Press.

Odonnell, D. E., Aaron, S., Bourbeau, J., Hernandez, P., Marciniuk, D. D., Balter, M., Ford, G., Gervais, A., Goldstein, R., Hodder, R., et al. (2007). Canadian thoracic society recommendations for management of chronic obstructive pulmonary disease–2007 update. *Canadian Respiratory Journal*, 14.

Odonnell, D. E., Hernandez, P., Kaplan, A., Aaron, S., Bourbeau, J., Marciniuk, D., Balter, M., Ford, G., Gervais, A., Lacasse, Y., et al. (2008). Canadian thoracic society recommendations for management of chronic obstructive pulmonary disease–2008 update–highlights for primary care. *Canadian Respiratory Journal*, 15.

ODonnell, S., Group, C. C. D. S. S. C. O. W., et al. (2013). Use of administrative data for national surveillance of osteoporosis and related fractures in canada: results from a feasibility study. *Archives of osteoporosis*, 8(1-2):143.

Panday, K., Gona, A., and Humphrey, M. B. (2014). Medication-induced osteoporosis: screening and treatment strategies. *Therapeutic advances in musculoskeletal disease*, 6(5):185–202.

Pazzagli, L., Linder, M., Zhang, M., Vago, E., Stang, P., Myers, D., Andersen, M., and Bahmanyar, S. (2018). Methods for time-varying exposure related problems in pharmacoepidemiology: An overview. *Pharmacoepidemiology and drug safety*, 27(2):148–160.

Pearl, J. (2009). *Causality*. Cambridge university press.

Petrone, J. (2018). Roche pays $1.9 billion for flatiron's army of electronic health record curators.

PHAC (2010). 2009 canadian community health surveyosteoporosis rapid response.

Pisani, M. A., Araujo, K. L., and Murphy, T. E. (2015). Association of cumulative dose of haloperidol with next day delirium in older medical intensive care unit patients. *Critical care medicine*, 43(5):996.

Pullenayegum, E. M. and Lim, L. S. (2016). Longitudinal data subject to irregular observation: A review of methods with a focus on visit processes, assumptions, and study design. *Statistical methods in medical research*, 25(6):2992–3014.

R Core Team (2013). *R: A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing, Vienna, Austria.

Ramsay, J. O. et al. (1988). Monotone regression splines in action. *Statistical science*, 3(4):425–441.

Ray, W. A. (2003). Evaluating medication effects outside of clinical trials: new-user designs. *American journal of epidemiology*, 158(9):915–920.

Robins, J. (1986). A new approach to causal inference in mortality studies with a sustained exposure periodapplication to control of the healthy worker survivor effect. *Mathematical modelling*, 7(9-12):1393–1512.

Rodriguez, G. (2007). Lecture notes on generalized linear models. *URL: http://data. princeton. edu/wws509/notes/c4. pdf*.

Rosenbaum, P. R. (2002). Overt bias in observational studies. In *Observational studies*, pages 71–104. Springer.

Ross, S. M. (2014). *Introduction to probability models*. Academic press.

Roysland, K. et al. (2011). A martingale approach to continuous-time marginal structural models. *Bernoulli*, 17(3):895–915.

Royston, P., Ambler, G., and Sauerbrei, W. (1999). The use of fractional polynomials to model continuous risk variables in epidemiology. *International journal of epidemiology*, 28(5):964–974.

Ryalen, P. C., Stensrud, M. J., Fosså, S., and Røysland, K. (2020). Causal inference in continuous time: an example on prostate cancer therapy. *Biostatistics*, 21(1):172–185.

Saarela, O. (2016). A case-base sampling method for estimating recurrent event intensities. *Lifetime data analysis*, 22(4):589–605.

Saarela, O. and Liu, Z. (2016). A flexible parametric approach for estimating continuous-time inverse probability of treatment and censoring weights. *Statistics in medicine*, 35(23):4238–4251.

Schmidli, H., Häring, D. A., Thomas, M., Cassidy, A., Weber, S., and Bretz, F. (2020). Beyond randomized clinical trials: Use of external controls. *Clinical Pharmacology & Therapeutics*, 107(4):806–816.

Steinbuch, M., Youket, T. E., and Cohen, S. (2004). Oral glucocorticoid use is associated with an increased risk of fracture. *Osteoporosis international*, 15(4):323–328.

Suissa, S., Moodie, E. E., and Dell'Aniello, S. (2017). Prevalent new-user cohort designs for comparative drug effect studies by time-conditional propensity scores. *Pharmacoepidemiology and drug safety*, 26(4):459–468.

Sylvestre, M.-P. and Abrahamowicz, M. (2009). Flexible modeling of the cumulative effects of time-dependent exposures on the hazard. *Statistics in medicine*, 28(27):3437–3453.

Tang, T.-S., Austin, P. C., Lawson, K. A., Finelli, A., and Saarela, O. (2020). Constructing inverse probability weights for institutional comparisons in healthcare. *Statistics in Medicine*.

Ten Have, T. R., Normand, S. L. T., Marcus, S. M., Brown, C. H., Lavori, P., and Duan, N. (2008). Intent-to-treat vs. non-intent-to-treat analyses under treatment non-adherence in mental health randomized trials. *Psychiatric annals*, 38(12).

Therneau, T. and Grambsch, P. M. (2000). *Modeling Survival Data: Extending the Cox Model*. Statistics for Biology and Health. Springer-Verlag New York, 1 edition.

Tran, L., Yiannoutsos, C., Wools-Kaloustian, K., Siika, A., Van Der Laan, M., and Petersen, M. (2019). Double robust efficient estimators of longitudinal treatment effects: comparative performance in simulations and a case study. *The international journal of biostatistics*, 15(2).

Van Norman, G. A. (2016). Drugs, devices, and the fda: part 1: an overview of approval processes for drugs. *JACC: Basic to Translational Science*, 1(3):170–179.

Van Staa, T., Leufkens, H., Abenhaim, L., Zhang, B., and Cooper, C. (2000). Use of oral corticosteroids and risk of fractures. *Journal of bone and mineral research*, 15(6):993–1000.

WHO (2019a). Collaborating centre for drug statistics methodology: Alendronate cholecalciferol.

WHO (2019b). Collaborating centre for drug statistics methodology: Alendronic acid.

WHO (2019c). Collaborating centre for drug statistics methodology: Calcitonin.

WHO (2019d). Collaborating centre for drug statistics methodology: Denosumab.

WHO (2019e). Collaborating centre for drug statistics methodology: Etidronic acid.

WHO (2019f). Collaborating centre for drug statistics methodology: Raloxifene.

WHO (2019g). Collaborating centre for drug statistics methodology: Risedronate.

WHO (2019h). Collaborating centre for drug statistics methodology: Zoledronic acid.

Williams, D. (1991). *Probability with martingales*. Cambridge university press.

Xiao, Y., Abrahamowicz, M., and Moodie, E. E. (2010). Accuracy of conventional and marginal structural cox model estimators: a simulation study. *The international journal of biostatistics*, 6(2).

Xiao, Y., Abrahamowicz, M., Moodie, E. E., Weber, R., and Young, J. (2014). Flexible marginal structural models for estimating the cumulative effect of a time-dependent treatment on the hazard: reassessing the cardiovascular risks of didanosine treatment in the swiss hiv cohort study. *Journal of the American Statistical Association*, 109(506):455–464.

Xiao, Y., Moodie, E. E., and Abrahamowicz, M. (2013). Comparison of approaches to weight truncation for marginal structural cox models. *Epidemiologic Methods*, 2(1):1–20.

Yang, D. and Dalton, J. E. (2012). A unified approach to measuring the effect size between two groups using sas®. In *SAS global forum*, volume 335, pages 1–6.

Yasir, M. and Sonthalia, S. (2019). Corticosteroid adverse effects.

Young, J. G., Hernan, M. A., Picciotto, S., and Robins, J. M. (2010). Relation between three classes of structural models for the effect of a time-varying exposure on survival. *Lifetime data analysis*, 16(1):71.

Young, J. G., Stensrud, M. J., Tchetgen Tchetgen, E. J., and Hernán, M. A. (2020). A causal framework for classical statistical estimands in failure-time settings with competing events. *Statistics in Medicine*, 39(8):1199–1236.

Young, J. G. and Tchetgen Tchetgen, E. J. (2014). Simulation from a known cox msm using standard parametric models for the g-formula. *Statistics in medicine*, 33(6):1001–1014.

Zeileis, A. (2006). Object-oriented computation of sandwich estimators. *Journal of Statistical Software*, 16(9):1–16.

Zeileis, A., Köll, S., and Graham, N. (2020). Various versatile variances: An object-oriented implementation of clustered covariances in R. *Journal of Statistical Software*, 95(1):1–36.

Zhang, M., Joffe, M. M., and Small, D. S. (2011). Causal inference for continuous-time processes when covariates are observed only at discrete times. *Annals of statistics*, 39(1).

Zhang, W., Zhao, Y., Zhang, F., Wang, Q., Li, T., Liu, Z., Wang, J., Qin, Y., Zhang, X., Yan, X., et al. (2020). The use of anti-inflammatory drugs in the treatment of people with severe coronavirus disease 2019 (covid-19): The experience of clinical immunologists from china. *Clinical Immunology*, page 108393.

Appendix A

Derivations & Data Cleaning

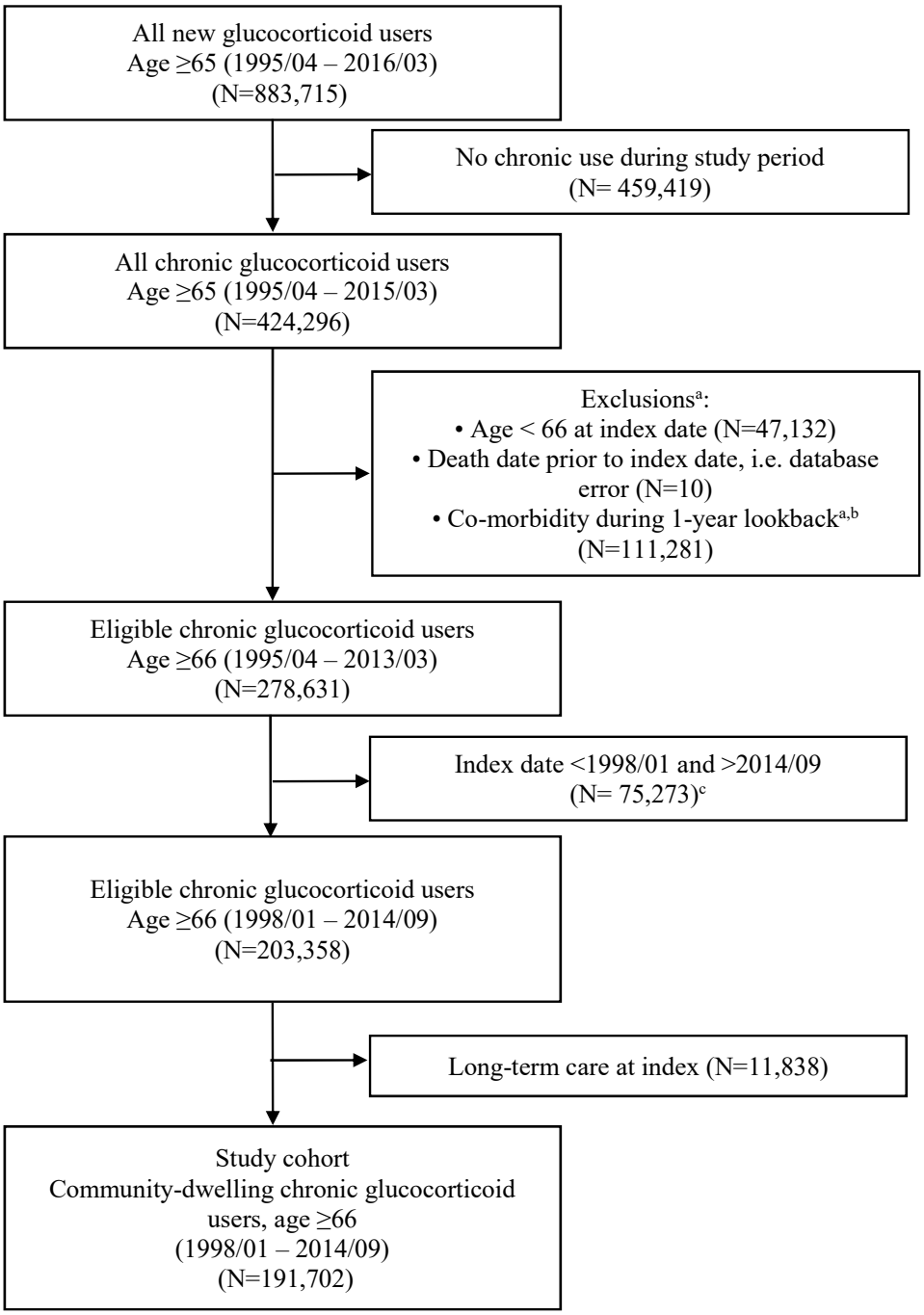
Cox partial likelihood derivation:

$$
\begin{aligned}
L(\eta_1, h_0(t)) &\overset{\eta}{=} \prod_i^n h_i(t;\eta)^{e_i} S_i(t;\eta) \\
&= \prod_i^n h_i(t;\eta)^{e_i} S_i(t;\eta) \cdot \frac{\left[\sum_{l\in G(u)} h_l(t;\eta)\right]^{e_i}}{\left[\sum_{l\in G(t)} h_l(t;\eta)\right]^{e_i}} \\
&= \prod_i^n \left[\frac{h_i(t;\eta)}{\sum_{l\in G(t)} h_l(t;\eta)}\right]^{e_i} \cdot S_i(t;\eta) \cdot \left[\sum_{l\in G(t)} h_l(t;\eta)\right]^{e_i} \\
&= \prod_i^n \left[\frac{h_0(t)\exp\{\eta_1 X_i\}}{\sum_{l\in G(t)} h_0(t)\exp\{\eta_1 X_l\}}\right]^{e_i} \cdot S_i(t;\eta) \cdot \left[\sum_{l\in G(t)} h_l(t;\eta)\right]^{e_i} \\
&= \prod_i^n \left[\frac{\exp\{\eta_1 X_i\}}{\sum_{l\in G(t)} \exp\{\eta_1 X_l\}}\right]^{e_i} \cdot S_i(t;\eta) \cdot \left[\sum_{l\in G(t)} h_l(t;\eta)\right]^{e_i} \\
L(\eta_1) &\overset{\eta_1}{\propto} \prod_i^n \left[\frac{\exp\{\eta_1 X_i\}}{\sum_{l\in G(t)} \exp\{\eta_1 X_l\}}\right]^{e_i}
\end{aligned}
\tag{A.1}
$$

Table A.1: ICD codes used for fracture outcome derivation

Fractures	**Hip Fracture**		
	ICD-9 (CIHI DAD)	820.xx	Fx* of neck of femur
	ICD-10 (CIHI DAD)	S72.0x	Fx of neck of femur
		S72.1x	Pertrochanteric fx
		S72.2	Subtrochanteric fx
	Radius or ulna fracture		
	OHIP	813	Radius or ulna
	ICD-9 (CIHI DAD)	813.x	Fx of radius and ulna
	ICD-10 (CIHI DAD)	S52.x	Fx of forearm
	Humerus fracture		
	OHIP	812	Fx of humerus
	ICD-9 (CIHI DAD)	812.x	Fx of humerus
	ICD-10 (CIHI DAD)	S42.2x	Fx of upper end of humerus
		S42.3x	Fx of shaft of humerus
		S42.4x	Fx of lower end of humerus
	Vertebral fracture		
	OHIP	805	Vertebral without spinal cord damage
	ICD-9 (CIHI DAD)	733.13	Path fx of vertebrae
		805.xx	Fx of vertebral column w/o spinal cord injury
	ICD-10 (CIHI DAD)	S22.0	Fx of thoracic vertebra
		S22.1	Multiple fx of thoracic spine
		S32.0	
		S32.7-S32.8	
		M48.4	Fatigue or stress fracture of vertebral

* Fx = fracture



Index date = First dispensing date in the 6-month window that meets the "chronic use" criteria (cumulative prednisone equivalent dose ≥450 mg and ≥2 prescriptions dispensed)
[a]Total may not add to 100% due to overlap between exclusion criteria
[b]Includes chronic renal disease (N=29,235), malignancy other than skin (N=88,456), organ transplant (N=1,552), or bone metabolic disorders (Paget's disease (N=595), osteomalacia (N=12))
[c]Before 01/1998 the days supply variable is not available. Patients identified after 2014/09 do not have a minimum of 6-months before the start of follow-up

Figure A.1: Existing cohort exclusion criteria flowchart (Amiche, 2018)

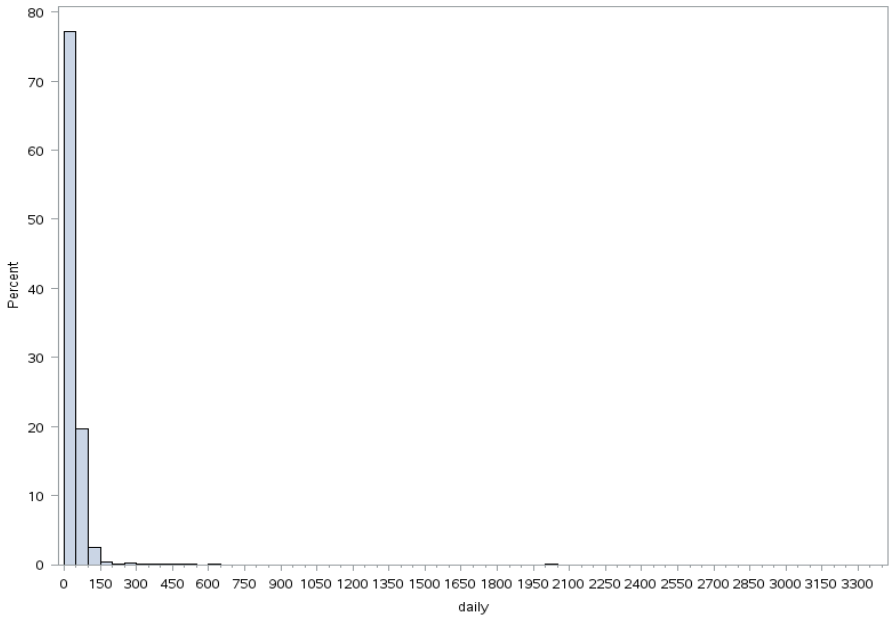


Figure A.2: The distribution of daily dose from all dispensations of all patients. 97.5% of the data lies within 100mg.

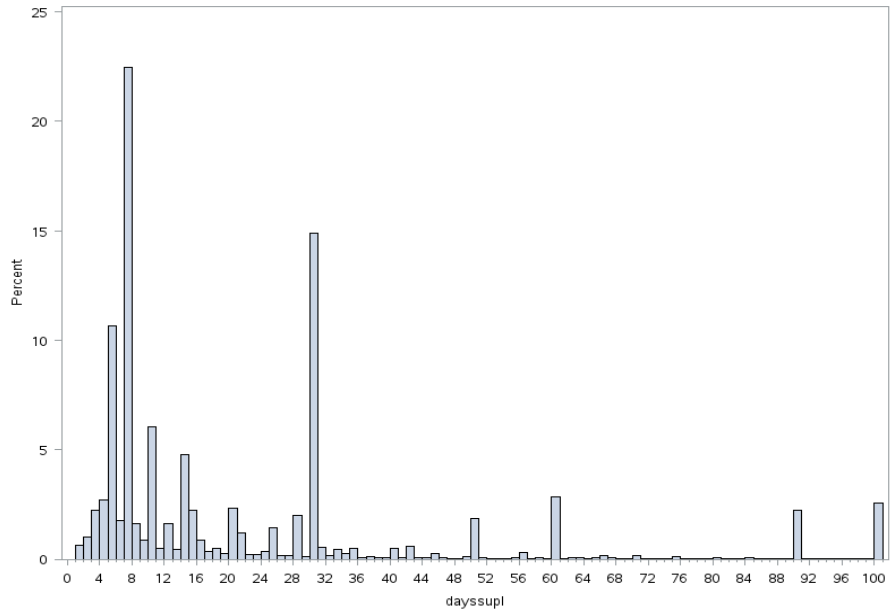


Figure A.3: The distribution of days-supply value from all dispensations of all patients. The most common days-supply is 7 days, followed by a second peak at 30 days, and a third peak at 5 days.

[1]For all histograms in the Appendix, bars with < 6 counts were hidden.

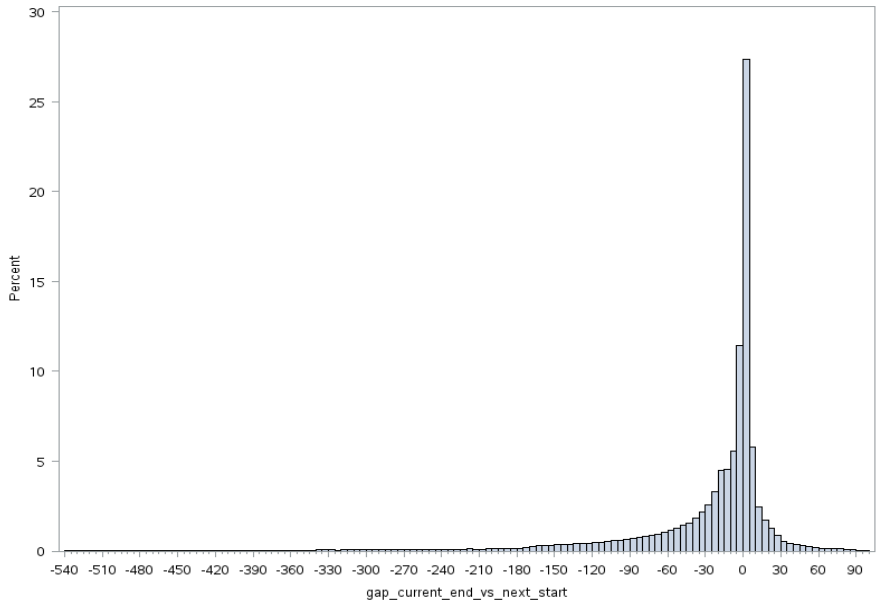


Figure A.4: The distribution of gap - current dispense date plus current days supply value minus the next dispensation date - from all dispensations of all patients. A positive value means there is an overlap in exposure, while a negative value (e.g.-90) means a patient waited for 90 days until receiving the next dispensation. The peak of the distribution is at 0 days, which means around 28% of two subsequent dispensations has no overlap in time. The range of the distribution goes from -540 days to 100 days. We determine the threshold value at gap $= 30$ days for either pushing forward all subsequent dispensation dates ($0 <$ gap $\leq$ 30 days) or truncating current days supply value to next dispensation date (gap > 30 days).

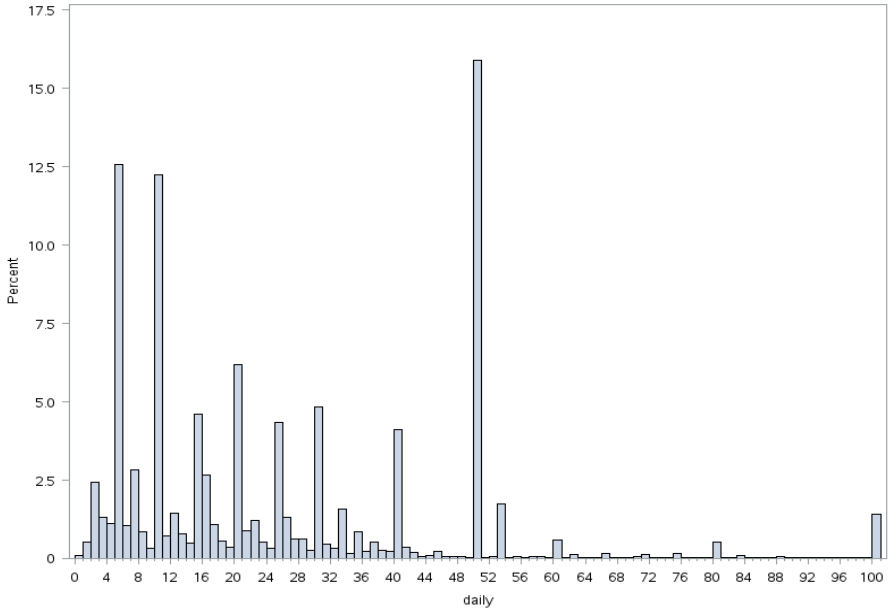


Figure A.5: The distribution of daily dose after excluding extreme values of $>$ 100mg

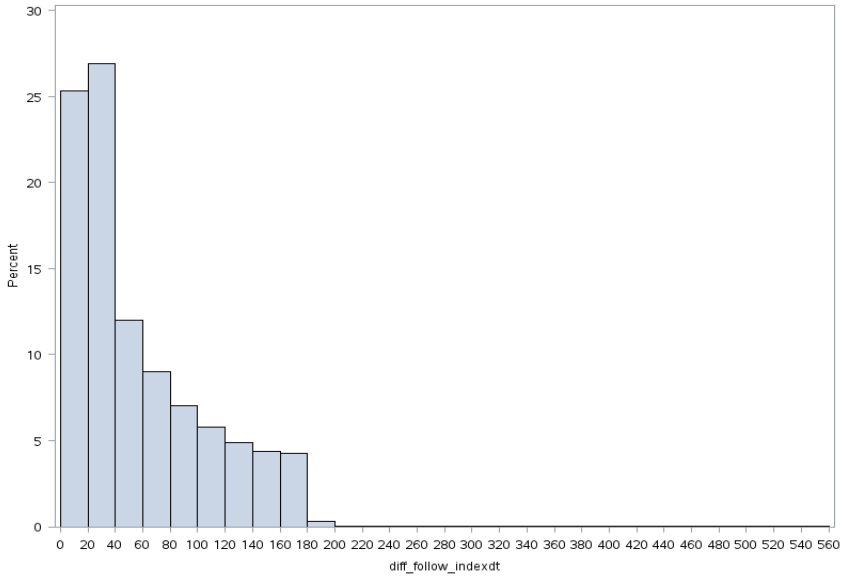


Figure A.6: The difference of days between t_1 and t_0 for patients who satisfied dispensation and cumulative dose requirements. We determine the cut-off value at 180 days, and exclude $< 1\%$ patients with $t_1 > t_0 + 180$ days.

Table A.2: World Health Organization defined daily dose for bisphosphonates (WHO, 2019a,b,c,d,e,f,g,h)

Osteoporosis Drug Types	Drug Names	WHO DDD (mg)
Nitrogen-containing BP	Alendronate	10
	Alendronate Sodium	10
	Alendronate Sodium Cholecalciferol	10
	Risedronate	5
	Zoledronic Acid	4
Non-nitrogen containing BP	Etidronate	400
Other OP Drugs	Denosumab	0.33
	Raloxifene	60
	Calcitonin	–

Table A.3: Other osteoporosis drugs days-supply imputation (Burden et al., 2013)

Other OP Drugs	Quantity	Days-supply Imputation
Denosumab	1	183‖
Raloxifene	Q	Q**
Calcitonin	Q	Q × 28

‖ Based on dispensation date

** Based on quantity dispensed

Table A.4: Different washout periods for fracture subtypes comparison. Hip fracture (HIP), radius or ulna (RAD) fractures, vertebral (VERT) fractures, and humerus (HUM) fractures were evaluated.

No washout	**HIP (N)**	**RAD (N)**	**VERT (N)**	**HUM (N)**
OHIP	NA	595	474	404
ICD diagnostic codes				
a. ICD DAD	773	138	456	138
b. ICD NACRS	587	333	333	232
c. ICD DAD or NACRS	**804**	390	688	283
OHIP or ICD diagnostic codes	804	669	980	470
OHIP or ICD diagnostic + PR codes*	683	655	980	451

90 day washout	**HIP (N)**	**RAD (N)**	**VERT (N)**	**HUM (N)**
OHIP	NA	562	466	372
ICD diagnostic codes				
a. ICD DAD	768	137	453	137
b. ICD NACRS	585	333	330	232
c. ICD DAD or NACRS	799	388	682	282
OHIP or ICD diagnostic codes	799	**634**	**963**	**437**
OHIP or ICD diagnostic + PR codes*	681	621	963	417

120 day washout	**HIP (N)**	**RAD (N)**	**VERT (N)**	**HUM (N)**
OHIP	NA	557	461	365
ICD diagnostic codes				
a. ICD DAD	767	137	452	137
b. ICD NACRS	585	333	330	232
c. ICD DAD or NACRS	798	388	681	282
OHIP or ICD diagnostic codes	798	629	957	431
OHIP or ICD diagnostic + PR codes*	681	616	957	411

* Procedure codes were listed in Appendix Table A.5.

Table A.5: Procedure codes used only for exploring the fracture definition versus washout period, unused in actual fracture date derivation

Procedure Codes			
	Hip Fracture		
	CCI	Fixation: 1VA73, 1VC73	Reduction: 1VA74, 1VC74, 1VA80, 1VC80
	CCP	Fixation: 9104, 9124	Reduction: 9054, 9114, 9134
	Radius or ulna fracture		
	CCI	Fixation: 1TV74 Repair: 1TV80, 1TV82	Reduction: 1TV73 Immobilization: 1TV03
	CCP	Fixation: 9111, 9131, 9052	Reduction: 9101, 9121, 9141
	Humerus fracture		
	CCI	Fixation: 1TK74 Repair: 1TK80, 1TK82	Reduction:1TK73 Immobilization: 1TK03
	CCP	Fixation: 9130, 9130, 9051	Reduction: 9100, 9120
	Vertebral fracture		
	NA		

Appendix B

Additional Analysis Outputs

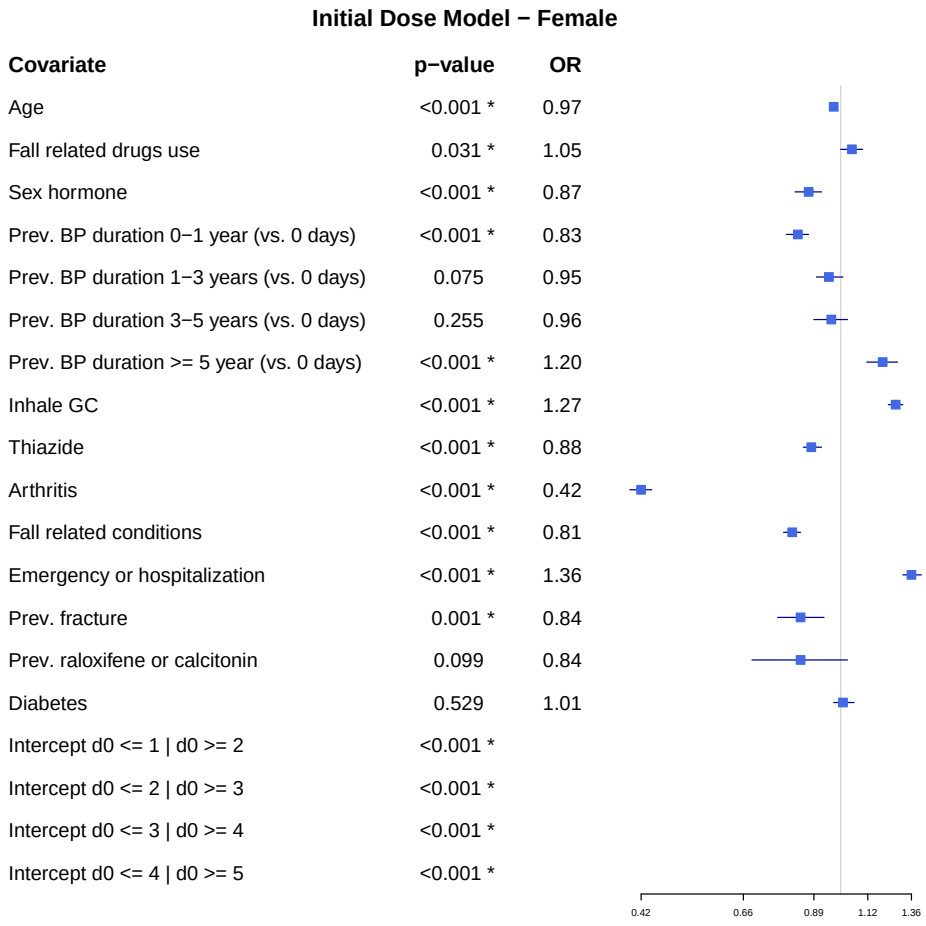


Figure B.1: The forest plot of OR for the initial dose model D_{i0} for female.

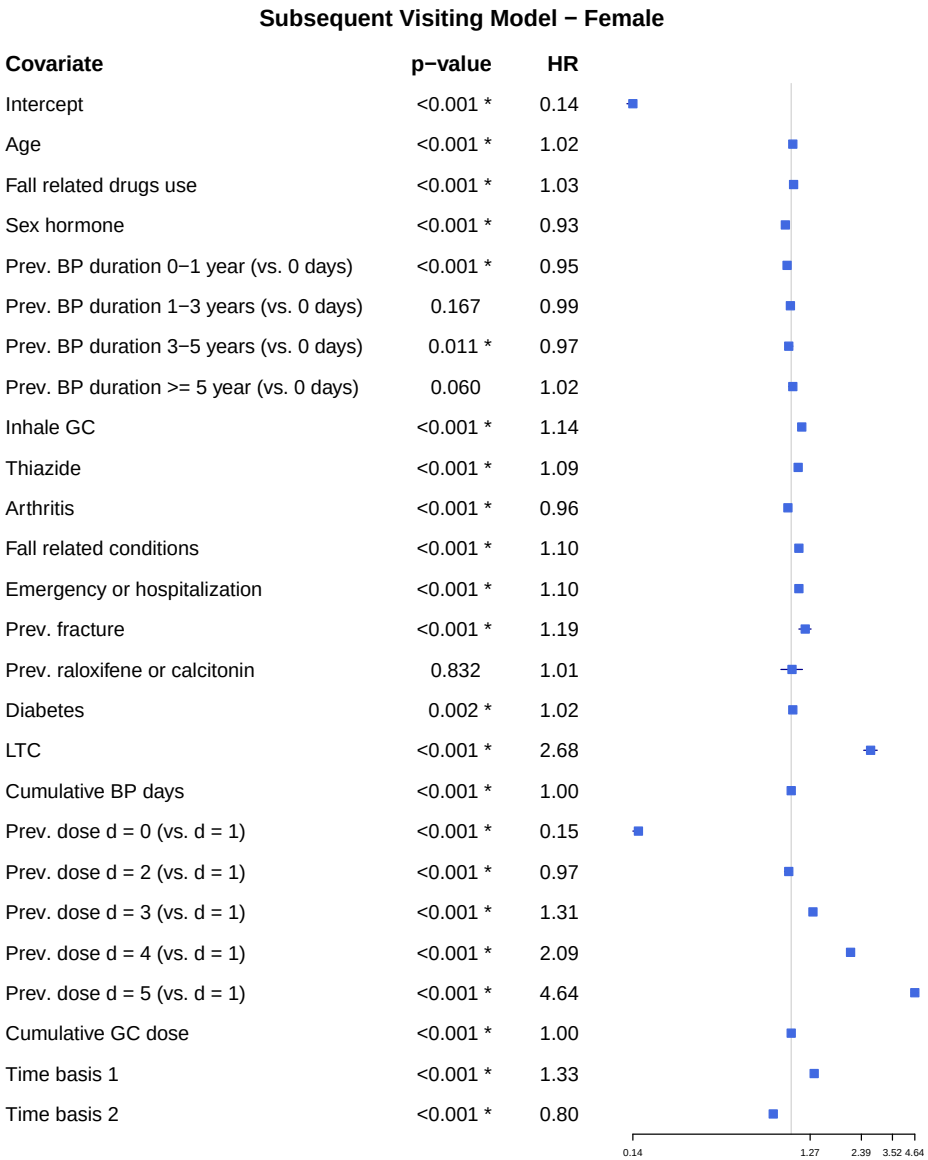


Figure B.2: The forest plot of HR for the subsequent visit model V_{ik} for female.

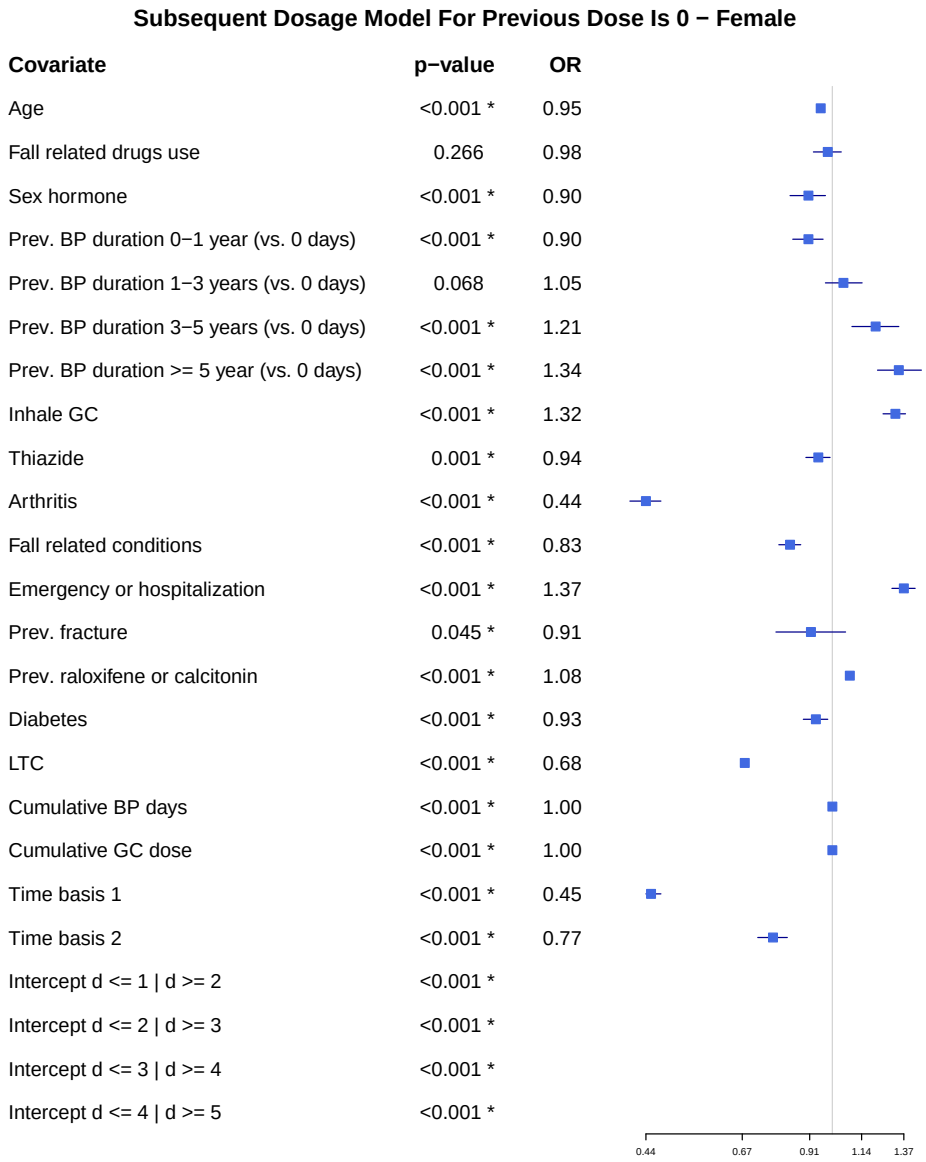


Figure B.3: The forest plot of OR for the subsequent dose model D_{ik}, where at previous dose $D_{i,k-1}$, the dose level is unexposed for female.

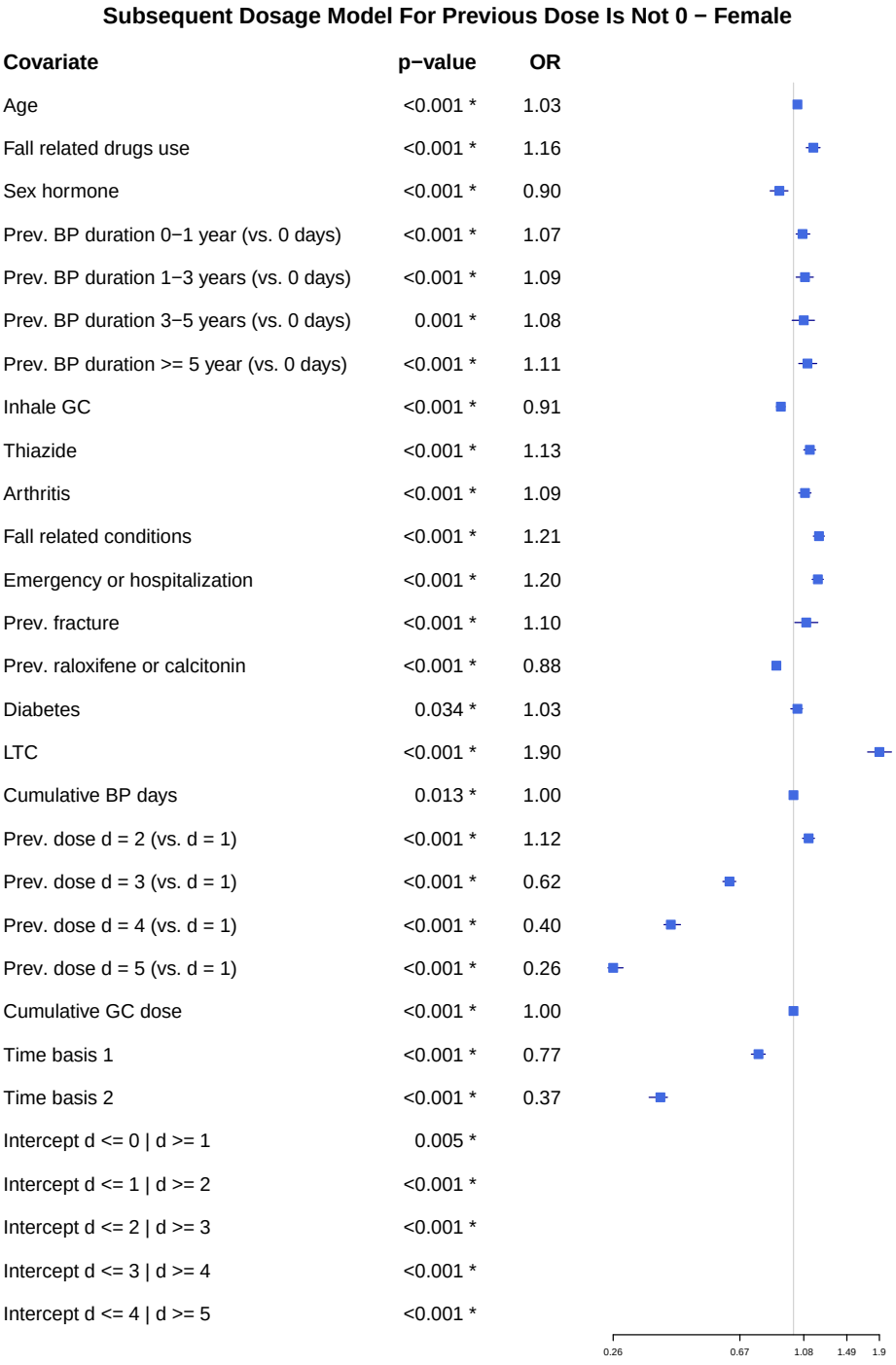


Figure B.4: The forest plot of OR for the subsequent dose model D_{ik}, where at previous dose $D_{i,k-1}$, the dose level is unexposed for female.

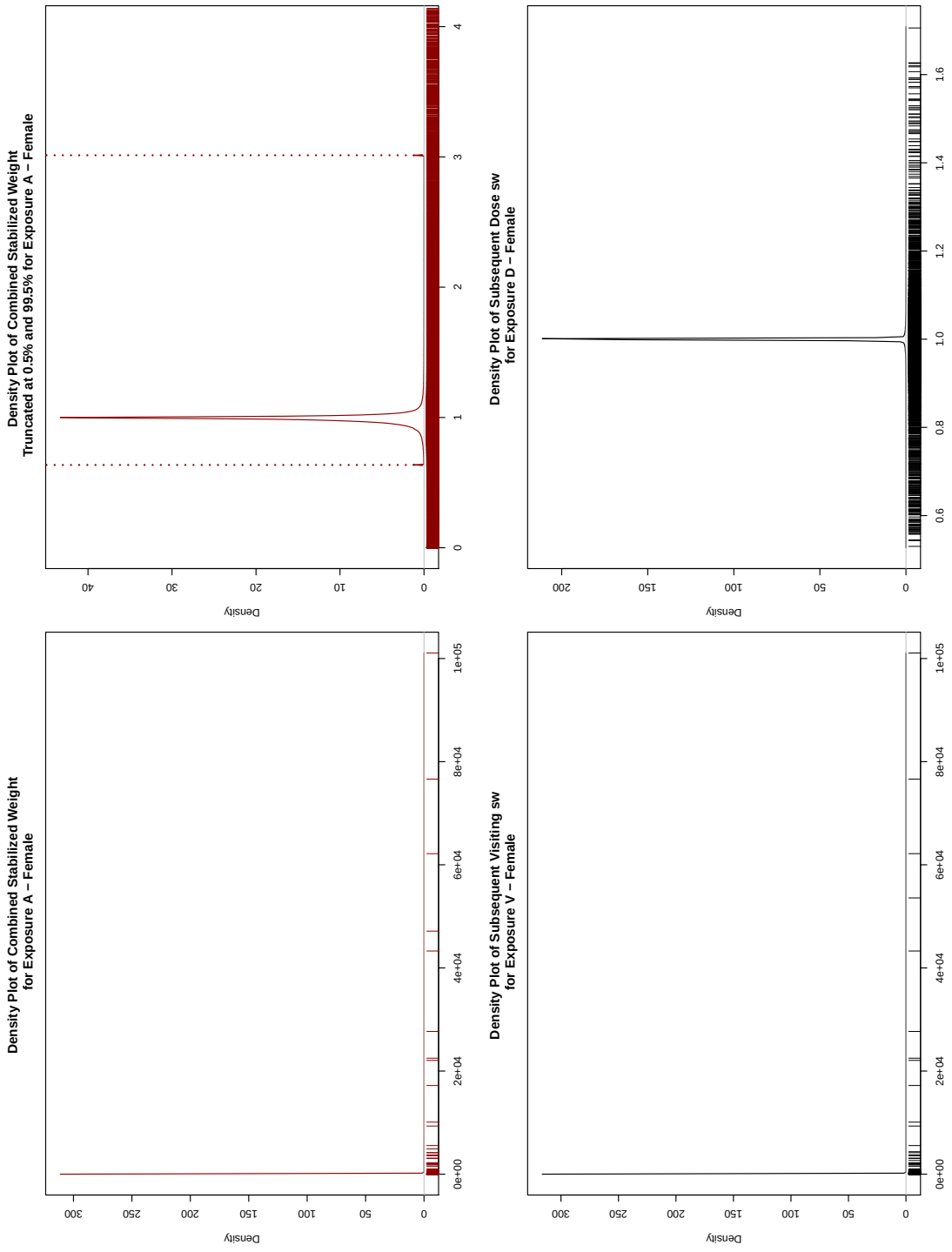


Figure B.5: Density Plot of IPTW for Multiple Visit MPP - Female

Table B.1: Parameter estimates for the weighted cumulative dose outcome model - Male

Covariate	Notation	Coef.Estimate	Std.Error	HR	95% CI	p
Intercept	η_0	-9.93	0.14	0.00	(0.00, 0.00)	<0.001 *
Baseline hazard basis 1 (cont.)	θ_{01}	-0.49	0.24	0.61	(0.38, 0.98)	<0.039 *
Baseline hazard basis 2 (cont.)	θ_{02}	-0.40	0.13	0.67	(0.52, 0.86)	0.002 *
Age	Z_{i1}	0.05	0.01	1.05	(1.04, 1.06)	<0.001 *
Fall related drugs use (yes vs. no)	Z_{i2}	-0.09	0.09	0.91	(0.76, 1.09)	0.304
Sex hormone (yes vs. no)	Z_{i3}	0.75	0.23	2.12	(1.36, 3.30)	0.001 *
Prev. BP duration 0-1 year (vs. 0 days)	Z_{i4}	-0.02	0.14	0.98	(0.74, 1.29)	0.863
Prev. BP duration 1-3 years (vs. 0 days)	Z_{i5}	0.36	0.20	1.43	(0.97, 2.11)	0.069
Prev. BP duration 3-5 years (vs. 0 days)	Z_{i6}	0.65	0.24	1.92	(1.20, 3.09)	0.007 *
Prev. BP duration $\geq$ 5 year (vs. 0 days)	Z_{i7}	-0.04	0.31	0.96	(0.52, 1.78)	0.907
Inhale GC (yes vs. no)	Z_{i8}	-0.06	0.07	0.94	(0.82, 1.09)	0.409
Thiazide (yes vs. no)	Z_{i9}	0.18	0.08	1.19	(1.01, 1.40)	0.033 *
Arthritis (yes vs. no)	Z_{i10}	0.08	0.11	1.09	(0.88, 1.34)	0.434
Fall related conditions (yes vs. no)	Z_{i11}	0.29	0.07	1.34	(1.16, 1.55)	<0.001 *
Emergency or hospitalization (yes vs. no)	Z_{i12}	0.37	0.09	1.45	(1.20, 1.74)	<0.001 *
Prev. fracture (yes vs. no)	Z_{i13}	1.57	0.14	4.81	(3.64, 6.35)	<0.001 *
Cumulative GC dose (cont.)	$\mathrm{cum}(\tilde{a})$	0.00	0.00	1.00	(1.00, 1.00)	<0.001 *

cont. means continuous variable.

Table B.2: Parameter estimates for the weighted flexible cumulative dose outcome model - Male

Covariate	Notation	Coef.Estimate	Std.Error	HR	95% CI	p
Intercept	η_0	-10.22	0.15	0.00	(0.00, 0.00)	<0.001 *
Baseline hazard basis 1 (cont.)	θ_{01}	-0.57	0.24	0.57	(0.35, 0.91)	<0.018 *
Baseline hazard basis 2 (cont.)	θ_{02}	-0.43	0.13	0.65	(0.50, 0.84)	0.001 *
Age (cont.)	Z_{i1}	0.05	0.01	1.05	(1.04, 1.06)	<0.001 *
Fall related drugs use (yes vs. no)	Z_{i2}	-0.09	0.09	0.91	(0.77, 1.09)	0.329
Sex hormone (yes vs. no)	Z_{i3}	0.75	0.23	2.12	(1.36, 3.30)	0.001 *
Prev. BP duration 0-1 year (vs. 0 days)	Z_{i4}	-0.03	0.14	0.97	(0.74, 1.29)	0.859
Prev. BP duration 1-3 years (vs. 0 days)	Z_{i5}	0.36	0.20	1.43	(0.97, 2.11)	0.069
Prev. BP duration 3-5 years (vs. 0 days)	Z_{i6}	0.66	0.24	1.94	(1.21, 3.11)	0.006 *
Prev. BP duration $\geq$ 5 year (vs. 0 days)	Z_{i7}	-0.04	0.31	0.96	(0.52, 1.77)	0.902
Inhale GC (yes vs. no)	Z_{i8}	-0.05	0.07	0.95	(0.82, 1.09)	0.471
Thiazide (yes vs. no)	Z_{i9}	0.17	0.08	1.19	(1.01, 1.40)	0.034 *
Arthritis (yes vs. no)	Z_{i10}	0.07	0.11	1.07	(0.87, 1.32)	0.507
Fall related conditions (yes vs. no)	Z_{i11}	0.29	0.07	1.34	(1.15, 1.55)	<0.001 *
Emergency or hospitalization (yes vs. no)	Z_{i12}	0.37	0.09	1.45	(1.20, 1.74)	<0.001 *
Prev. fracture (yes vs. no)	Z_{i13}	1.57	0.14	4.83	(3.66, 6.37)	<0.001 *
Cumulative GC dose spline basis 1 (cont.)	$\mathrm{Basis}(\mathrm{cum}(\tilde{a}))_1$	2.35	0.52	10.45	(3.78, 28.92)	<0.001 *
Cumulative GC dose spline basis 2 (cont.)	$\mathrm{Basis}(\mathrm{cum}(\tilde{a}))_2$	0.67	1.47	1.95	(0.11, 34.60)	0.650

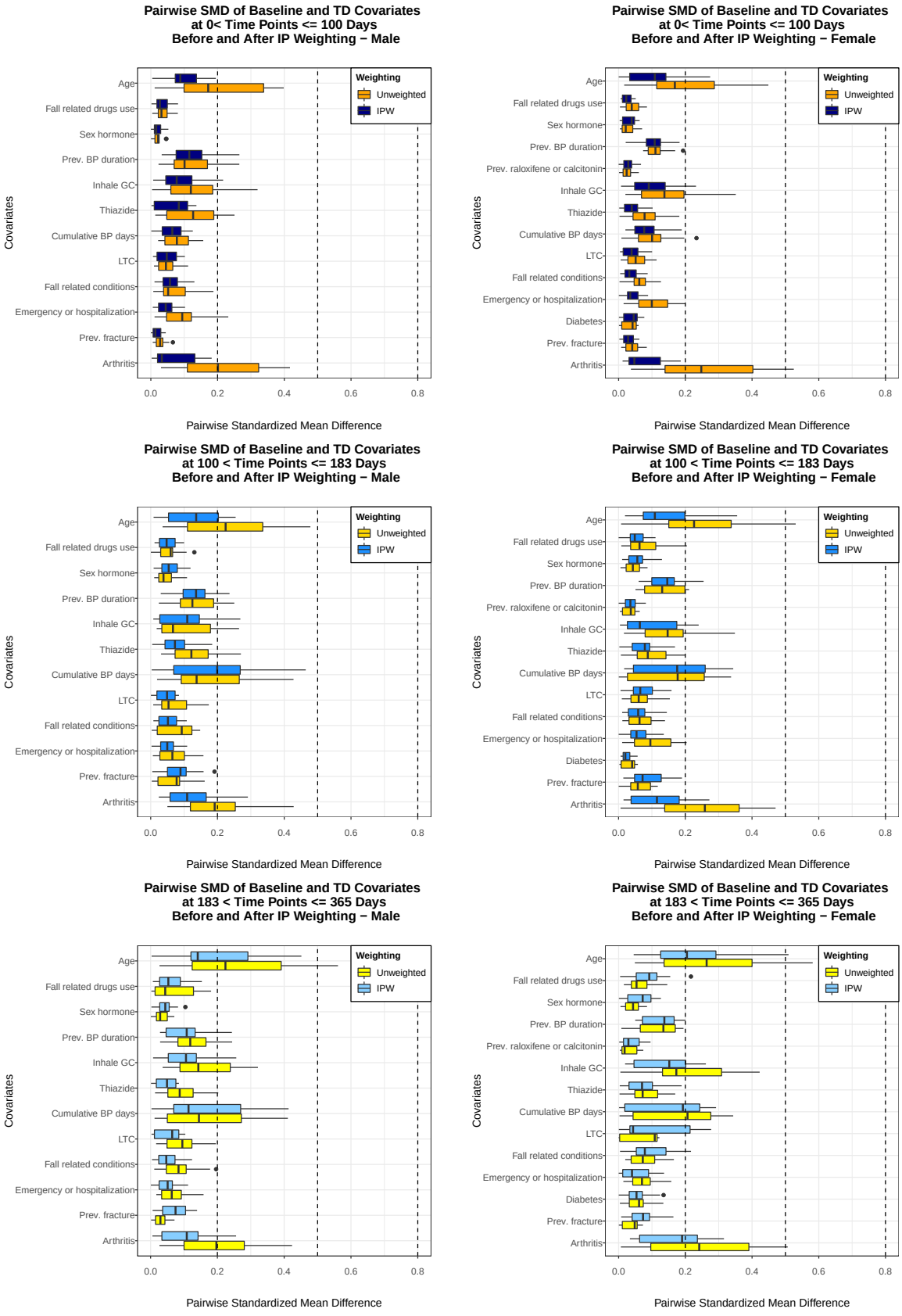


Figure B.6: Male and female pair-wise SMD over time

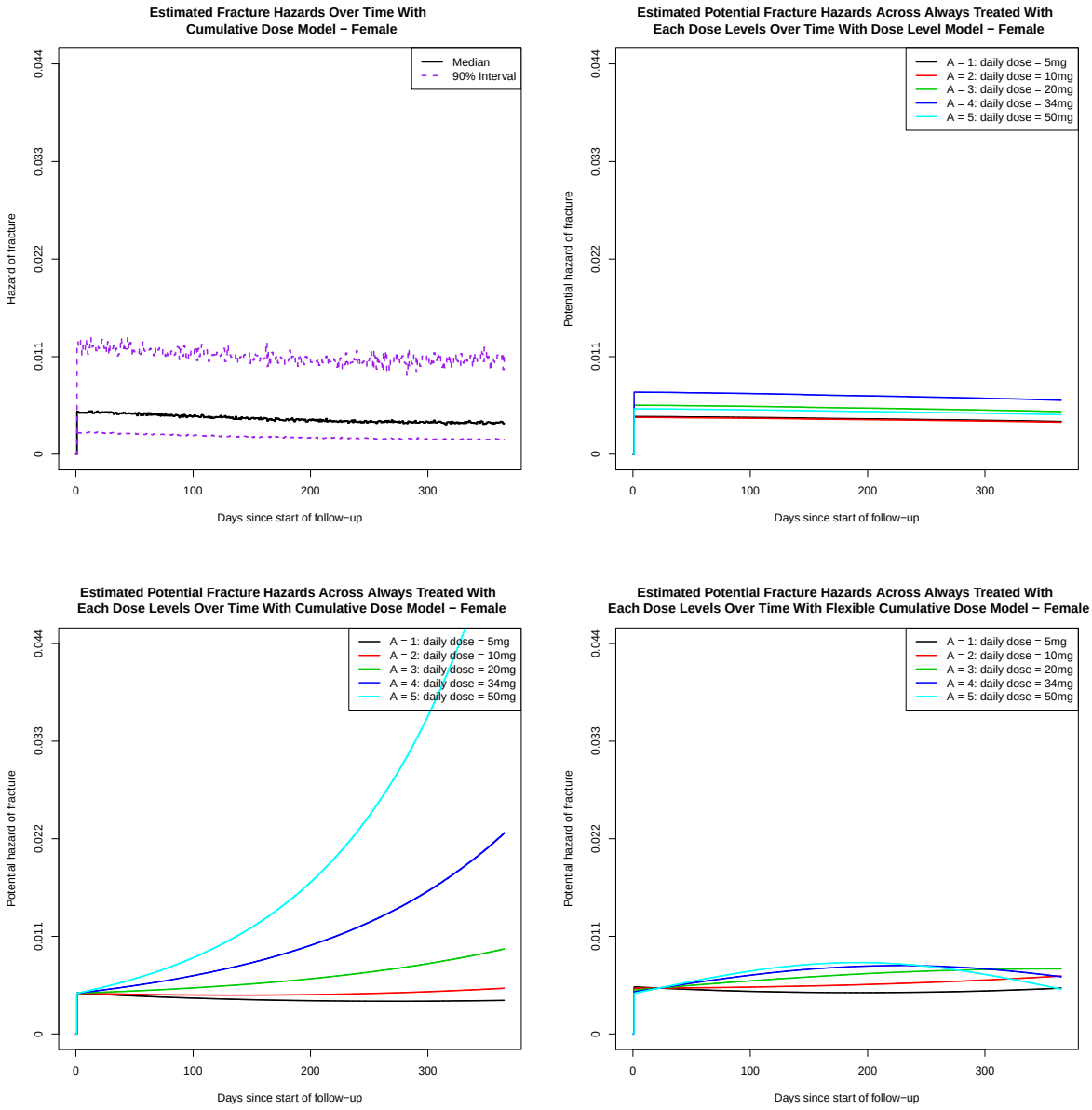


Figure B.7: Descriptive and potential hazard under always treated with each dose

Table B.3: Parameter estimates for the unweighted dose level outcome model - Female

Covariate	Notation	Coef.Estimate	Std.Error	HR	95% CI	p
Intercept	η_0	-9.55	0.13	0.00	(0.00, 0.00)	<0.001 *
Baseline hazard basis 1 (cont.)	θ_{01}	0.06	0.16	1.06	(0.77, 1.46)	0.715
Baseline hazard basis 2 (cont.)	θ_{02}	-0.17	0.09	0.84	(0.71, 1.00)	0.045 *
Age (cont.)	Z_{i1}	0.06	0.00	1.06	(1.05, 1.06)	<0.001 *
Fall related drugs use (yes vs. no)	Z_{i2}	0.08	0.07	1.09	(0.96, 1.24)	0.205
Sex hormone (yes vs. no)	Z_{i3}	-0.23	0.09	0.79	(0.66, 0.95)	0.014 *
Prev. BP duration 0-1 year (vs. 0 days)	Z_{i4}	0.17	0.07	1.18	(1.03, 1.35)	0.016 *
Prev. BP duration 1-3 years (vs. 0 days)	Z_{i5}	0.32	0.07	1.38	(1.19, 1.60)	<0.001 *
Prev. BP duration 3-5 years (vs. 0 days)	Z_{i6}	0.21	0.09	1.24	(1.03, 1.49)	0.025 *
Prev. BP duration $\geq$ 5 year (vs. 0 days)	Z_{i7}	0.10	0.08	1.11	(0.94, 1.30)	0.207
Inhale GC (yes vs. no)	Z_{i8}	0.04	0.05	1.04	(0.95, 1.15)	0.372
Thiazide (yes vs. no)	Z_{i9}	0.06	0.05	1.07	(0.96, 1.18)	0.222
Arthritis (yes vs. no)	Z_{i10}	0.04	0.07	1.04	(0.92, 1.19)	0.509
Fall related conditions (yes vs. no)	Z_{i11}	0.20	0.05	1.22	(1.11, 1.35)	<0.001 *
Emergency or hospitalization (yes vs. no)	Z_{i12}	0.24	0.06	1.28	(1.14, 1.43)	<0.001 *
Prev. fracture (yes vs. no)	Z_{i13}	0.95	0.09	2.59	(2.18, 3.09)	<0.001 *
Prev. raloxifene or calcitonin (yes vs. no)	Z_{i14}	0.17	0.25	1.18	(0.72, 1.94)	0.508
Diabetes (yes vs. no)	Z_{i15}	-0.19	0.06	0.83	(0.73, 0.93)	0.002 *
Treatment $A = 0$ (vs. $A = 1$)	$\mathbf{1}\{A_i(t) = 0\}$	-0.03	0.09	0.97	(0.81, 1.16)	0.758
Treatment $A = 2$ (vs. $A = 1$)	$\mathbf{1}\{A_i(t) = 2\}$	0.01	0.12	1.01	(0.79, 1.29)	0.940
Treatment $A = 3$ (vs. $A = 1$)	$\mathbf{1}\{A_i(t) = 3\}$	0.27	0.11	1.31	(1.05, 1.63)	0.019 *
Treatment $A = 4$ (vs. $A = 1$)	$\mathbf{1}\{A_i(t) = 4\}$	0.53	0.15	1.70	(1.27, 2.27)	<0.001 *
Treatment $A = 5$ (vs. $A = 1$)	$\mathbf{1}\{A_i(t) = 5\}$	0.24	0.19	1.27	(0.88, 1.83)	0.204

Table B.4: Parameter estimates for the weighted cumulative dose outcome model - Female

Covariate	Notation	Coef.Estimate	Std.Error	HR	95% CI	p
Intercept	η_0	-9.30	0.10	0.00	(0.00, 0.00)	<0.001 *
Baseline hazard basis 1 (cont.)	θ_{01}	-0.44	0.16	0.64	(0.47, 0.88)	0.005 *
Baseline hazard basis 2 (cont.)	θ_{02}	-0.47	0.09	0.62	(0.53, 0.74)	<0.001 *
Age (cont.)	Z_{i1}	0.06	0.00	1.06	(1.05, 1.06)	<0.001 *
Fall related drugs use (yes vs. no)	Z_{i2}	0.08	0.07	1.09	(0.96, 1.24)	0.197
Sex hormone (yes vs. no)	Z_{i3}	-0.20	0.09	0.82	(0.69, 0.98)	0.030 *
Prev. BP duration 0-1 year (vs. 0 days)	Z_{i4}	0.14	0.07	1.15	(1.01, 1.32)	0.037 *
Prev. BP duration 1-3 years (vs. 0 days)	Z_{i5}	0.30	0.07	1.35	(1.17, 1.56)	<0.001 *
Prev. BP duration 3-5 years (vs. 0 days)	Z_{i6}	0.20	0.09	1.22	(1.02, 1.47)	0.030 *
Prev. BP duration $\geq$ 5 year (vs. 0 days)	Z_{i7}	0.08	0.08	1.09	(0.93, 1.27)	0.298
Inhale GC (yes vs. no)	Z_{i8}	0.09	0.05	1.10	(1.00, 1.21)	0.052
Thiazide (yes vs. no)	Z_{i9}	0.07	0.05	1.07	(0.97, 1.19)	0.191
Arthritis (yes vs. no)	Z_{i10}	0.01	0.06	1.01	(0.89, 1.14)	0.916
Fall related conditions (yes vs. no)	Z_{i11}	0.20	0.05	1.22	(1.11, 1.35)	<0.001 *
Emergency or hospitalization (yes vs. no)	Z_{i12}	0.24	0.06	1.27	(1.13, 1.42)	<0.001 *
Prev. fracture (yes vs. no)	Z_{i13}	0.94	0.09	2.56	(2.16, 3.03)	<0.001 *
Prev. raloxifene or calcitonin (yes vs. no)	Z_{i14}	0.13	0.26	1.14	(0.69, 1.89)	0.606
Diabetes (yes vs. no)	Z_{i15}	-0.19	0.06	0.82	(0.73, 0.93)	0.001 *
Cumulative GC dose (cont.)	$\text{cum}(\bar{a})$	0.00	0.00	1.00	(1.00, 1.00)	<0.001 *

Table B.5: Parameter estimates for the weighted flexible cumulative dose outcome model - Female

Covariate	Notation	Coef.Estimate	Std.Error	HR	95% CI	p
Intercept	η_0	-9.61	0.10	0.00	(0.00, 0.00)	<0.001 *
Baseline hazard basis 1 (cont.)	θ_{01}	-0.55	0.16	0.58	(0.42, 0.79)	0.001 *
Baseline hazard basis 2 (cont.)	θ_{02}	-0.51	0.09	0.60	(0.51, 0.71)	<0.001 *
Age (cont.)	Z_{i1}	0.06	0.00	1.06	(1.05, 1.06)	<0.001 *
Fall related drugs use (yes vs. no)	Z_{i2}	0.09	0.07	1.09	(0.96, 1.24)	0.193
Sex hormone (yes vs. no)	Z_{i3}	-0.20	0.09	0.82	(0.68, 0.98)	0.029 *
Prev. BP duration 0-1 year (vs. 0 days)	Z_{i4}	0.14	0.07	1.15	(1.01, 1.31)	0.039 *
Prev. BP duration 1-3 years (vs. 0 days)	Z_{i5}	0.30	0.07	1.35	(1.16, 1.56)	<0.001 *
Prev. BP duration 3-5 years (vs. 0 days)	Z_{i6}	0.21	0.09	1.23	(1.02, 1.48)	0.027 *
Prev. BP duration $\geq$ 5 year (vs. 0 days)	Z_{i7}	0.09	0.08	1.09	(0.93, 1.28)	0.293
Inhale GC (yes vs. no)	Z_{i8}	0.10	0.05	1.11	(1.01, 1.22)	0.031 *
Thiazide (yes vs. no)	Z_{i9}	0.07	0.05	1.07	(0.97, 1.19)	0.198
Arthritis (yes vs. no)	Z_{i10}	-0.01	0.06	0.99	(0.87, 1.12)	0.858
Fall related conditions (yes vs. no)	Z_{i11}	0.20	0.05	1.22	(1.11, 1.34)	<0.001 *
Emergency or hospitalization (yes vs. no)	Z_{i12}	0.24	0.06	1.27	(1.13, 1.42)	<0.001 *
Prev. fracture (yes vs. no)	Z_{i13}	0.93	0.09	2.54	(2.15, 3.01)	<0.001 *
Prev. raloxifene or calcitonin (yes vs. no)	Z_{i14}	0.14	0.26	1.14	(0.69, 1.89)	0.599
Diabetes (yes vs. no)	Z_{i15}	-0.19	0.06	0.82	(0.73, 0.93)	0.001 *
Cumulative GC dose spline basis 1 (cont.)	$\text{Basis}(\text{cum}(\bar{a}))_1$	2.61	0.37	13.64	(6.63, 28.07)	<0.001 *
Cumulative GC dose spline basis 2 (cont.)	$\text{Basis}(\text{cum}(\bar{a}))_2$	-0.61	1.12	0.55	(0.06, 4.92)	0.589

Table B.6: Likelihood ratio test for all three outcome models - Male

	Outcome Models		
	(Eq. 4.3)	**(Eq. 4.4)**	**(Eq. 4.5)**
Degree of freedom	5	1	2
Pr(> chisq)	0.009 *	<0.001 *	<0.001 *

Table B.7: Likelihood ratio test for all three outcome models - Female

	Outcome Models		
	(Eq. 4.3)	**(Eq. 4.4)**	**(Eq. 4.5)**
Degree of freedom	5	1	2
Pr(> chisq)	<0.001 *	<0.001 *	<0.001 *