UC Irvine

UC Irvine Electronic Theses and Dissertations

Title

Extensions of the Case-Cohort Design for Increasing Precision in Biomarker Discovery

Permalink

https://escholarship.org/uc/item/9b47b870

ISBN

9798173738394

Author

Xu, Yiren

Publication Date

2026-11-04

Copyright Information

This work is made available under the terms of a Creative Commons Attribution License, available at https://creativecommons.org/licenses/by/4.0/

Peer reviewed|Thesis/dissertation

eScholarship.org Powered by the California Digital Library
University of California

UNIVERSITY OF CALIFORNIA,
IRVINE

Extensions of the Case-Cohort Design for Increasing Precision in Biomarker Discovery

DISSERTATION

submitted in partial satisfaction of the requirements
for the degree of

DOCTOR OF PHILOSOPHY

in Statistics

by

Yiren Xu

Dissertation Committee:
Chancellor's Professor Daniel L. Gillen, Chair
Assistant Professor Ana Maria Kenney
Chancellor's Professor Bin Nan

2026

© 2026 Yiren Xu

DEDICATION

To my parents, for their unconditional love and support that made this journey possible.

TABLE OF CONTENTS

Page

LIST OF FIGURES **v**

LIST OF TABLES **vi**

ACKNOWLEDGMENTS **viii**

VITA **ix**

ABSTRACT OF THE DISSERTATION **xi**

1 Introduction **1**

2 Background **5**
2.1 The Cox Proportional Hazards Model 5
2.1.1 Andersen and Gill Counting Process Formulation 7
2.1.2 Alternative Sampling Designs 8
2.2 Univariate Case-Cohort Designs 10
2.2.1 Variance Estimation in Case-Cohort Designs 11
2.2.2 Efficiency in Case-Cohort Designs 15
2.3 Multivariate Case-Cohort Designs 16
2.3.1 Marginal Models for Multivariate Analysis 17
2.3.2 Recurrent Events Analysis 21
2.3.3 Case-cohort Designs for Recurrent Events Analysis 22
2.3.4 Sampling Considerations 24
2.4 Power and Sample Size Calculations in Survival Analysis 25
2.4.1 Overview of Power and Sample Size Calculations 25
2.4.2 Univariate Full Cohort Power and Sample Size Calculations 27
2.4.3 Univariate Case-Cohort Design Power and Sample Size Calculations . 29
2.4.4 Recurrent Events Full Cohort Power and Sample Size Calculations . . 31

3 Case-cohort designs to improve efficiency for detecting effect modification **37**
3.1 Introduction 37
3.2 Statistical Methods 41
3.2.1 Pseudo-Likelihood Estimators 41

3.2.2 Asymptotic Properties . . . 43
3.3 Simulation Study . . . 44
3.3.1 Bias and Variance . . . 45
3.3.2 Power and Coverage Probability . . . 47
3.4 Application to ADNI Data . . . 49
3.5 Discussion . . . 56
3.6 Acknowledgments . . . 58

4 Case-cohort designs to improve efficiency for recurrent events analysis 61
4.1 Introduction . . . 61
4.2 Methods: Event-Agnostic Model . . . 64
4.2.1 Proposed Model . . . 64
4.2.2 Impact of Sampling . . . 71
4.3 Simulation Study . . . 78
4.3.1 Performance of Proposed Estimators . . . 78
4.3.2 Sampling Comparisons and Impact on Efficiency . . . 80
4.4 Application . . . 82
4.5 Discussion . . . 88
4.6 Acknowledgments . . . 90

5 Sample size and power for recurrent events case-cohort designs 92
5.1 Introduction . . . 92
5.2 Methods . . . 94
5.2.1 Weighted Logrank Tests Using the Full Cohort . . . 94
5.2.2 Logrank-Type Tests for Recurrent Events in the Case Cohort Design 96
5.2.3 Power and Sample Size for Recurrent Events Data in the Case Cohort Design . . . 100
5.2.4 Using Subcohort Sampling Designs . . . 104
5.3 Simulation Study . . . 105
5.4 Application . . . 110
5.5 Discussion . . . 111

6 Conclusion and Future Directions 114

Bibliography 119

Appendix A Appendices for Chapter 2 130

Appendix B Appendices for Chapter 3 133

Appendix C Appendices for Chapter 4 138

LIST OF FIGURES

Page

3.1 Contour plots of the difference between power assessing β_3 for the stratified versus the classic case-cohort designs, in the interaction model. 48
3.2 Directed acyclic graph of response variable (time to AD progression), predictor of interest (pTau levels), and potential confounders and associated variables. 52

4.1 An illustration of one example of pooled sampling. 72
4.2 An illustration of one example of balanced sampling. 73
4.3 An illustration of one example of events-proportional sampling. 77
4.4 Kaplan-Meier plots for Event 1 (left) and Event 2 (right) based on full cohort data in the application of proposed method to A4 and LEARN data. 86

5.1 For $\lambda_1 = 0.05$, the comparison of estimated power obtained from power formula and empirical power obtained from simulations under exponential failure times and uniform censoring times, with no, mild, moderate, and substantial within-subject correlations. 109

6.1 Plots comparing estimated and empirical power when the baseline hazard function increase from Event 1 to Event 2, β_3, are $\log(1.1), \log(1.3)$, under mild correlations and the same settings as the simulation study in Chapter 5. 118

LIST OF TABLES

Page

3.1 Simulation results showing relative bias, robust, and empirical variance for the marginal model with $\beta_3 = \log(1.2)$ and 5% prevalence of $Z_2 = 1$. 46
3.2 Simulation results showing relative bias, robust, and empirical variance for the interaction model with $\beta_3 = \log(1.2)$ and 5% prevalence of $Z_2 = 1$. 47
3.3 Simulation results comparing power between different scenarios in assessing β_3 in the interaction model. 49
3.4 Simulation results comparing power between different scenarios in assessing γ_2 in the marginal model. 50
3.5 Simulation results of the coverage probability when estimating β_3 for the stratified case-cohort design in the interaction model. 51
3.6 Summary statistics of the analysis sets (including the subcohort and supplemented cases) for the classic case-cohort and stratified case-cohort designs. . 53
3.7 Estimates of hazard ratios (95% CI) and p-values of the model considering race and ethnicity as a potential effect modifier, with adjustment covariates. 59
3.8 Estimates of hazard ratios (95% CI) and p-values of the model considering APOE status as a potential effect modifier, with adjustment covariates. . . . 60

4.1 Notation for event-specific sampling schemes when $J = 2$. 74
4.2 Subcohort quantities for Step 1 of event-specific, balanced sampling, when we solve for the number of Event 1 control, denoted as x. 74
4.3 Subcohort quantities for Step 1 of event-specific, events-proportional sampling. 76
4.4 Simulation results showing relative bias and variance for the event-agnostic model. 78
4.5 Sample sizes of simulation results comparing case-cohort sampling schemes. . 80
4.6 Simulation results showing relative bias and variance for event-agnostic models with different sampling schemes. 80
4.7 Simulation results showing relative bias and variance for fully stratified models. 81
4.8 Table of summary statistics of A4 and LEARN participants at baseline. . . . 84
4.9 Event-agnostic model results for estimating the effect of cholesterol levels in association to PACC decline, adjusted for covariates. 85
4.10 Fully stratified model results for estimating the effect of cholesterol levels in association with PACC decline, adjusted for covariates. 85
4.11 Full event-agnostic model results using the events-proportional CC design. . 87

5.1 Table of estimated values of $\hat{\sigma}_w^2$ through simulations with 1,000,000 subjects in the full cohort. 105
5.2 Estimated power obtained from power formula and empirical power obtained from simulations under exponential failure times and uniform censoring times, with no ($\sigma_w^2 = 0$) within-subject correlations. 106
5.3 Estimated power obtained from power formula and empirical power obtained from simulations under exponential failure times and uniform censoring times, with moderate ($\sigma_w^2 = 0.09$) within-subject correlations. 107
5.4 Estimated power obtained from power formula and empirical power obtained from simulations under exponential failure times and uniform censoring times, using the events-proportional sampling scheme, with different levels of within-subject correlations. 108
5.5 For the A4 and LEARN studies data, assuming that failure times follow exponential distributions, under assumed hazard ratios, censoring, and correlations (σ_w^2), the minimum required subcohort size ($\tilde{n}$) and average subcohort sampling fraction (q^*) to achieve 70% or 80% power, based on the proposed sample size formula. 112

ACKNOWLEDGMENTS

First, I would like to thank my advisor Dr. Daniel Gillen. I am incredibly grateful to learn to become a statistician from someone who cares so deeply about his work and creating positive impact in the world. Thank you for tirelessly working with me on my papers and talks, patiently explaining statistical concepts, and kindly offering me advice in career decisions. You have helped me grow to be a better academic, mentor, and person. You have always believed in me, even when I did not myself, and for that I am truly grateful.

Thank you to Dr. Joshua Grill, who has been a great mentor throughout my time here. Thank you for your trust in me as a collaborator. I learned so much about Alzheimer's disease research and recruitment and the aspects we must consider for responsible and insightful research. Our work showed me how I can make real impact in the world.

Thank you to members of my dissertation committee, Dr. Ana Maria Kenney and Dr. Bin Nan. I deeply appreciate your time dedicated to and your interest in my work. Your expertise has been instrumental to improving this dissertation. I am also grateful for your advice and guidance in navigating my academics and career.

Thank you to the Grillen lab and SMART Meeting members. It has been an honor to work with you and I am grateful for your invaluable support. Thank you to my cohort mates and everyone in the UCI Statistics department for your help on this journey and our memories. Thank you to my professors at UCLA, for lighting the spark of my interest in statistics.

I would like to thank my friends for believing in me and being there for me through tough times. I would also like to thank my family members, who have always cheered me on.

Thank you to my parents, who are the best parents in the world, who never showed how tough it was to support me through school, and who always encouraged me to pursue greater things. I am happy to make you proud and I dedicate this dissertation to you.

Lastly, thank you to my husband Nico. I would not have been able to complete this journey without your unwavering support. In times of stress and self-doubt, your wise words and love have always helped me stay strong and continue on.

The text of Chapter 3 of this dissertation is an adaptation of the material as it appears in Xu et al. (2026), used with permission from *Springer Nature* [CC BY 4.0]. The coauthors listed in this publication are Joshua D. Grill, Daniel L. Gillen, for the Alzheimer's Disease Neuroimaging Initiative.

The research in this dissertation were supported by the National Institutes of Health (NIH) / National Institute on Aging (NIA) grant R01AG075107, the University of California, Irvine (UCI) CREATE Fellowship, the UCI Public Impact Fellowship, and the UCI Bob & Barbara Kleist Endowed Graduate Fellowship. Any opinions, findings, conclusions, and recommendations expressed in this material are those of the author(s) and do not necessarily reflect the views of the University of California or the National Institutes of Health.

VITA

Yiren Xu

EDUCATION

Doctor of Philosophy in Statistics **2026**
University of California, Irvine *Irvine, CA*

Master of Science in Statistics **2023**
University of California, Irvine *Irvine, CA*

Bachelor of Science in Statistics, Minor in English **2021**
University of California, Los Angeles *Los Angeles, CA*

RESEARCH EXPERIENCE

Graduate Student Researcher **2022–2026**
University of California, Irvine *Irvine, California*

TEACHING EXPERIENCE

Teaching Assistant **2022**
University of California, Irvine *Irvine, California*

Reader **2021**
University of California, Irvine *Irvine, California*

REFEREED JOURNAL PUBLICATIONS

Case-cohort designs to improve efficiency for detecting effect modification **2026**
Statistics in Biosciences

Research Attitudes Questionnaire scores and retention in a recruitment registry **2025**
Journal of Alzheimer's Disease

AWARDS AND HONORS

Bob & Barbara Kleist Endowed Graduate Fellowship **2026**
University of California, Irvine

Public Impact Fellowship **2025-2026**
University of California, Irvine

CREATE Fellowship **2023**
University of California, Irvine

ABSTRACT OF THE DISSERTATION

Extensions of the Case-Cohort Design for Increasing Precision in Biomarker Discovery

By

Yiren Xu

Doctor of Philosophy in Statistics

University of California, Irvine, 2026

Chancellor's Professor Daniel L. Gillen, Chair

The analysis of time-to-event outcomes, or survival analysis, is common in many areas of biomedical research. In a large cohort study where the event of interest may have lower rates and the predictor of interest is expensive to measure, the case-cohort design can be applied to save the valuable covariate while maintaining valid inference, by performing analysis on a random sample from the full cohort and the unsampled cases. In Alzheimer's disease (AD) biomarker research, where many current and future biomarkers can be discovered from cerebrospinal fluid samples, the case-cohort design can be very useful. We propose several methodological developments within univariate and multivariate case-cohort designs, aimed to assist in solving challenges in AD biomarker discovery. We address some of the case-cohort design's shortcomings in its potential lack of power when performing analysis on a subset of the full cohort data. First, when there is at most one event per subject, we demonstrate the potential lack of power when assessing differential biomarker effects across rare subpopulations. We propose a stratified case-cohort estimator for assessing the interaction of the covariate of interest and an effect modifier, which is shown to be more efficient for comparing subgroup differences while maintaining valid marginal inference. Turning to the analysis of recurrent events, which is motivated by the need to perform analysis on multiple disease progressions and multiple covariate measurements, we develop a sampling and analysis framework and propose three types of case-cohort sampling to obtain an analysis set

for efficiency gains. Lastly, to assist in planning for a recurrent events case-cohort design, we develop a logrank test statistic and derive the required sample size for a desired level of power.

Chapter 1

Introduction

Survival analysis, or time-to-event analysis, is widely used in epidemiological studies. In these studies, we are interested in understanding the time to some event of interest, such as the time to disease onset or disease progression. The cases, or individuals who experience events, generally contribute the most information to the analysis. Therefore, in large cohort studies when event rates are low and when covariate information is expensive to obtain, it would be more efficient and cost-effective to analyze a subset of the full cohort. To this end, researchers have considered alternative sampling designs, including the case-cohort design, which requires analysis on only a subsample of the full cohort (Prentice, 1986). Such designs conserve expensive covariates, which are often valuable biological samples, and reduce the time and cost of covariate processing. After some modifications to the estimation procedure, the estimators remain consistent and efficiency losses can be minimal.

The work in this dissertation is largely motivated by goals and challenges in Alzheimer's disease (AD) biomarker research. AD is a neurodegenerative disease and it is currently the sixth-leading cause of death in the United States, projected to affect 13.8 million individuals aged 65 and older by 2060 (A&D, 2025). The gold standard for AD diagnosis is through

pathological analysis of brain tissue, which is not feasible for mass screening due to its invasive nature (Gunes et al., 2022). It is therefore crucial to find more feasible and accessible ways to identify early indications of disease progression, as early diagnosis can inform patients and clinicians about important life choices and treatment options. One key goal in AD research is biomarker discovery, in which researchers aim to identify measurable biological markers associated with disease risk or prognosis to guide clinical diagnosis (Blennow et al., 2010). These biomarkers can also be potential targets for development of new interventions.

Biomarker measurements, however, can be of high burden and cost to obtain. One way is through positron emission tomography (PET) scans, which are scans of the brain that involve the injection of radioactive tracers to measure biomarkers including amyloid beta peptide and phosphorylated tau. While PET scans are generally highly accurate, they may be costly and not always accessible. In addition, the requirement of a specific tracer makes it unfeasible as a means for biomarker discovery. On the other hand, we may measure biomarker levels from the cerebrospinal fluid (CSF). The CSF is in contact with the extracellular space of the brain and can therefore reflect biochemical changes in the brain, making it an optimal source to discover biomarkers (Blennow et al., 2010). To acquire CSF samples, however, patients must undergo lumbar punctures and their apprehension to consent to this procedure can be a common barrier to research participation (Witbracht et al., 2020). Once a limited amount of CSF sample has been obtained and stored for each study participant, it is important to conserve as much of each sample as possible for use in multiple analyses. The case-cohort design can be very useful in this setting, where we may save time, cost, and valuable specimens in biomarker discovery.

In this dissertation, we develop novel statistical methods that build on the case-cohort design to adequately address challenges in AD biomarker research. In the following chapters, we keep a common goal in mind: to perform population-level inference with validity while increasing power and efficiency for subgroups or substrata, in the univariate and multivariate

survival settings. We also develop a framework in recurrent events case-cohort designs that span the design, sampling, and analysis stages.

Chapter 2 provides background and review for the following chapters. The review covers topics in survival analysis: first, an overview of inference using the Cox proportional hazards (PH) model, a widely used semiparametric model for survival analysis with full cohort data. We then summarize the literature on univariate case-cohort designs. This is followed by a discussion of approaches in multivariate survival analysis and recurrent events analysis, including two types of models, marginal and frailty models. Finally, we cover power and sample size calculations for time-to-event data, including in the univariate case-cohort design and the recurrent events setting.

In Chapter 3, we develop methods in the univariate survival setting motivated by the need to compare biomarker effects among subgroups. Some genetic variants are commonly found to be associated with an increased risk of AD. The Apolipoprotein E (APOE) gene encodes a protein involved in lipid transport; dysregulation of this process may contribute to the development of AD. Specifically, having two copies of the APOE $\varepsilon 4$ allele is associated with a higher risk of AD (A&D, 2025). In line with the goals of biomarker discovery, we seek to first establish the effect of biomarkers across the full patient population while ensuring power to assess subgroup differences in secondary analyses. To investigate whether the effect of biomarkers can be modified by factors including APOE status, we test interaction effects between the two in secondary analyses. As recruitment challenges persist in scientific research, we propose a stratified case-cohort design (Borgan et al., 2000) and demonstrate that subgroup effects can be assessed with higher power and greater efficiency while maintaining valid inference in the marginal model. Furthermore, we establish that these gains are relatively greater with rarer subpopulations and smaller subset sizes.

In Chapter 4, we extend case-cohort methodology to address issues in recurrent event analyses, with a focus on examining efficiency gains when using proposed sampling methods.

While there have been works to address case-cohort analysis of multiple diseases (Kang and Cai, 2009) and of recurrent events (Maitra et al., 2020), these methods do not allow for event-specific covariates and covariate effects under recurrent event settings. In biomarker discovery or similar studies where expensive covariate measurements are taken several times, and when covariate effects may vary from one event to the next, it is important to accommodate these changes in our models. We then propose the following sampling designs: "pooled" sampling, or sampling from all available records with a common proportion, and "event-specific" sampling, or flexibly sampling records using occurrence-specific proportions. We demonstrate that by using event-specific sampling, we may achieve greater efficiency in later events while maintaining consistency and a level of efficiency in the marginal model.

In Chapter 5, we shift our focus to the design phase of a recurrent events case-cohort study, and we derive power and sample size calculations for the various sampling schemes proposed in Chapter 4. In a large cohort study or database, such as the Alzheimer's Disease Neuroimaging Initiative (ADNI) (Petersen et al., 2010), there are generally biological samples stored, including CSF samples. These reserves are valuable resources that researchers can leverage for new biomarker discovery studies. A case-cohort design is advantageous in this setting, as only a subset of the stored samples need to be processed. In this chapter, we derive a logrank-type statistic and power and sample size formula that performs well in estimating the required subcohort size under various assumptions about within-subject correlations.

In Chapter 6, we summarize our contributions and discuss future directions of this work.

The objective of this dissertation is to guide and assist a broad range of researchers in planning a case-cohort study, which has the advantage of conserving expensive covariates, but also bring attention to the importance of adequately powering the case-cohort study. While the motivation of our work stems from challenges in our collaborative work in AD research, these methods may be applied to many other fields that would benefit from the employment of efficient sampling designs.

Chapter 2

Background

2.1 The Cox Proportional Hazards Model

In epidemiological studies, many scientific questions involve the analysis of censored time-to-event data. While nonparametric methods such as the Kaplan-Meier estimator (Kaplan and Meier, 1958) and fully parametric methods exist, the semiparametric Cox proportional hazards (PH) model (Cox, 1972) has been one of the most widely-used choices for modeling. Here we review inference for the Cox model to then introduce theory and highlight similarities in inference with the case-cohort design under multiplicative hazard models.

Let the full cohort contain individuals $i = 1, 2, \ldots, n$ and let $\boldsymbol{Z}_i$ be a p-dimensional covariate vector. Consider the hazard function $\lambda(t)$:

$$\lambda_i(t|\boldsymbol{Z}_i(t)) = \lambda_0(t)\exp(\boldsymbol{\beta}'\boldsymbol{Z}_i(t)), \tag{2.1}$$

where $\lambda_0(t)$ is the baseline hazard function when covariate values $\boldsymbol{Z}_i(t)$ are zero and the

relative risk of failure comparing subjects with $Z_k + 1$ values to those with Z_k is $\exp(\beta_k)$, $k = 1, 2, \ldots, p$. $\boldsymbol{Z}_i(t)$ may contain time-dependent covariates that are external, as defined in Kalbfleisch and Prentice (2002), or those whose future path is not affected by whether the event occurs. For simplicity, in the following sections we write $\boldsymbol{Z}_i$.

Assume that the data $(X_i, \boldsymbol{Z}_i, \delta_i)$ are independent and identically distributed, where X_i is the observed time, δ_i is the failure indicator, and failure times are t_j with $j = 1, \ldots, D$ (in the following sections, "event" and "failure" will be used interchangeably). To obtain an estimator of the parameter, we write the partial likelihood function as

$$L(\beta) = \prod_{j=1}^{D} \frac{\exp(\boldsymbol{\beta}' \boldsymbol{Z}_j)}{\sum_{i \in R_j} \exp(\boldsymbol{\beta}' \boldsymbol{Z}_i)}, \tag{2.2}$$

and the log partial likelihood function as

$$l(\boldsymbol{\beta}) = \sum_{j=1}^{D} \log\Big(\frac{\exp(\boldsymbol{\beta}' \boldsymbol{Z}_j)}{\sum_{i \in R_j} \exp(\boldsymbol{\beta}' \boldsymbol{Z}_i)} \Big), \tag{2.3}$$

where $\boldsymbol{Z}_j$ is the covariate vector of the individual who failed at time t_j and R_j is the set of subjects who are still at risk at time t_j. The score function, or the estimating function, is then

$$U(\boldsymbol{\beta}) = \sum_{j=1}^{D} \left\{ \boldsymbol{Z}_j - \frac{\sum_{i \in R_j} \boldsymbol{Z}_i \exp(\boldsymbol{\beta}' \boldsymbol{Z}_i)}{\sum_{l \in R_j} \exp(\boldsymbol{\beta}' \boldsymbol{Z}_l)} \right\}. \tag{2.4}$$

The score contributions are fairly interpretable: at each event time, we compare the covariate value of the individual who experienced the event to a weighted average of the covariates of all individuals in the risk set, $\frac{\sum_{i \in R_j} \boldsymbol{Z}_i \exp(\boldsymbol{\beta}' \boldsymbol{Z}_i)}{\sum_{l \in R_j} \exp(\boldsymbol{\beta}' \boldsymbol{Z}_l)}$, which represents the risk-set weighted mean of the covariates. To obtain the estimator for the true $\boldsymbol{\beta}_0$, $\hat{\boldsymbol{\beta}}$, we can solve the estimating equation $U(\boldsymbol{\beta}) \equiv 0$ using the Newton-Raphson method. Estimation using the Cox PH model does not require the correct specification of the baseline hazard function or the censoring distribution (under non-informative censoring) to perform valid inference.

2.1.1 Andersen and Gill Counting Process Formulation

The counting process formulation for the Cox PH model may utilize martingale theory for estimation and to prove useful properties of the estimators (Andersen and Gill, 1982). Consider again, for individual $i = 1, \ldots, n$, we have data including their failure time T_i, censoring time C_i, observed time $X_i = T_i \wedge C_i$, and covariate vector $\boldsymbol{Z}_i(t)$. Let Δ_i denote the failure indicator, while the at-risk process is $Y_i(t) = I(X_i \geq t)$, and the counting process is $N_i(t) = I(X_i \leq t, \Delta_i = 1)$. Observed data $(X_i, \Delta_i, \boldsymbol{Z}_i)$ are assumed to be independent and identically distributed. In counting process notation, the score function is

$$U(\boldsymbol{\beta}) = \sum_{i=1}^{n} \int_0^{\tau} \Big\{ \boldsymbol{Z}_i - \frac{\sum_{j=1}^{n} Y_j(t) \boldsymbol{Z}_j \exp(\boldsymbol{\beta}' \boldsymbol{Z}_j)}{\sum_{j=1}^{n} Y_j(t) \exp(\boldsymbol{\beta}' \boldsymbol{Z}_j)} \Big\} dN_i(t), \tag{2.5}$$

where study time $t \in [0, \tau]$. We solve the estimating equation $U(\boldsymbol{\beta}) \equiv 0$ to obtain $\hat{\boldsymbol{\beta}}$, the estimator for $\boldsymbol{\beta}_0$. The asymptotic properties of $\hat{\boldsymbol{\beta}}$ are as follows. Under regularity conditions (detailed in Appendix A), we have that, as $n \to \infty$,

$$\sqrt{n}(\hat{\boldsymbol{\beta}} - \boldsymbol{\beta}_0) \xrightarrow{D} N(0, I(\boldsymbol{\beta}_0)^{-1}), \tag{2.6}$$

where $I(\boldsymbol{\beta}_0)$ is the Fisher information matrix at the true coefficient value $\boldsymbol{\beta}_0$.

We note here that the score function of a Cox PH model when there is a single binary covariate under the null hypothesis is equivalent to the logrank test statistic. We revisit this in Section 2.4 on deriving power formulas in survival settings.

2.1.2 Alternative Sampling Designs

In epidemiological studies, researchers sometimes encounter the analytical challenge of low event rates. Furthermore, in biomedical studies aiming to understand the association between biological measurements and time to an event, samples may be valuable and limited, and processing costs may be high. Since most information is contributed by cases, or individuals who experience the event of interest, processing expensive covariates for all controls is inefficient. Instead, it is advantageous to perform analysis on a subset of the full cohort. To this end, alternative sampling designs have been proposed, including the nested case-control design and the case-cohort design, both relying on the key result that cases provide more information than controls. Here, we briefly describe the nested case-control design and some differences between the two designs, but a more comprehensive summary can be found in Nuño and Gillen (2017).

Under the nested case-control design, we randomly sample a prespecified number of M controls at each event time from individuals who are still at risk, and therefore match M controls to cases at their event times. The score function is then

$$U_{NCC}(\boldsymbol{\beta}) = \sum_{j=1}^{D} \Big\{ \boldsymbol{Z}_j - \frac{\boldsymbol{Z}_j \exp(\boldsymbol{\beta}' \boldsymbol{Z}_j) + \sum_{l=1}^{M} \boldsymbol{Z}_l \exp(\boldsymbol{\beta}' \boldsymbol{Z}_l)}{\exp(\boldsymbol{\beta}' \boldsymbol{Z}_j) + \sum_{l=1}^{M} \exp(\boldsymbol{\beta}' \boldsymbol{Z}_l)} \Big\}. \tag{2.7}$$

In contrast to the full cohort, the risk set at each event time contains only the case and the

M matched controls (which are typically one to four). Individuals who later become cases may still serve as controls at earlier event times.

Under the case-cohort design, at the beginning of the analysis, we perform random sampling to obtain a subset of the full cohort, called the subcohort, then supplement the subcohort with unsampled cases. In both designs, the idea is to reduce the number of controls needed. Both designs can save researchers time, cost, and valuable biological specimens from controls, while maintaining valid inference at a small efficiency cost.

Comparatively, Kim (2015) found the nested case-control design to be more efficient under random censoring when using inverse probability weights (IPW) proposed by Samuelsen (1997), which utilizes information on controls at other event times. As new methods are developed, however, relative performance can depend on the correlation between failure times and covariates, and IPW methods are not always feasible (Kim, 2015).

The case-cohort design, by contrast, offers the advantage of reusing the same subcohort to study multiple disease outcomes or secondary endpoints, since subcohort sampling is not outcome-specific. Due to the nature of matching controls to cases of a unique disease outcome, this is more challenging in the nested case-control design. It is worth noting that case-cohort designs require identifying only new cases as the study progresses, whereas nested case-control designs require identifying both new cases and new controls (O'Brien et al., 2022).

In the following section, we review current literature on the case-cohort design under the univariate setting, which is when individuals experience at most one event.

2.2 Univariate Case-Cohort Designs

The case-cohort design was originally proposed by Prentice (1986). To conduct case-cohort sampling, we select a random sample called the subcohort from the full dataset then supplement it with the rest of the unsampled cases. The inclusion of non-subcohort cases creates a form of "biased" sampling, and one must account for their contributions to the weighted covariate average when performing the analysis. The pseudolikelihood contribution of one subject and the score function are $L_{j,Prentice} = \frac{\exp(\boldsymbol{\beta}'\boldsymbol{Z}_j)}{\exp(\boldsymbol{\beta}'\boldsymbol{Z}_j)+\sum_{l\in R_{j,sc}}\exp(\boldsymbol{\beta}'\boldsymbol{Z}_l)}$ and

$$U_{Prentice}(\boldsymbol{\beta}) = \sum_{j=1}^{D}\left\{\boldsymbol{Z}_j - \frac{Z_j\exp(\boldsymbol{\beta}'Z_j) + \sum_{l\in R_{j,sc}}\boldsymbol{Z}_l\exp(\boldsymbol{\beta}'\boldsymbol{Z}_l)}{\exp(\boldsymbol{\beta}'Z_j) + \sum_{l\in R_{j,sc}}\exp(\boldsymbol{\beta}'\boldsymbol{Z}_l)}\right\}, \tag{2.8}$$

where $t_j, j = 1, \ldots, D$ denote the event times and $R_{j,sc}$ denotes the set of subcohort members at risk at t_j. Non-subcohort cases receive a weight of 1 at their own failure times but a weight of 0 otherwise in the weighted average.

Self and Prentice (1988) developed a design that is asymptotically equivalent, but which only includes subcohort members still at risk in the risk set. This means that the non-subcohort cases receive weights of 0 both before and at their own failure times. The partial likelihood contribution and score are $L_{j,SelfPrentice} = \frac{\exp(\boldsymbol{\beta}'\boldsymbol{Z}_j)}{\sum_{l\in R_{j,sc}}\exp(\boldsymbol{\beta}'\boldsymbol{Z}_l)}$ and

$$U_{SelfPrentice}(\boldsymbol{\beta}) = \sum_{j=1}^{D}\left\{\boldsymbol{Z}_j - \frac{\sum_{l\in R_{j,sc}}\boldsymbol{Z}_l\exp(\boldsymbol{\beta}'\boldsymbol{Z}_l)}{\sum_{l\in R_{j,sc}}\exp(\boldsymbol{\beta}'\boldsymbol{Z}_l)}\right\}. \tag{2.9}$$

By not including supplemented cases at their own event times, Self and Prentice (1988) attempts to better manage variance estimation.

Furthermore, Barlow (1994) proposed to allow for contributions from non-subcohort cases at all event times by applying an inverse probability of sampling weight. The score function is then

$$U_{Barlow}(\boldsymbol{\beta}) = \sum_{j=1}^{D} \left\{ \boldsymbol{Z}_j - \frac{w_j(t_j) Z_j \exp(\boldsymbol{\beta}' Z_j) + \sum_{l \in R_{j,sc}} w_l(t_j) \boldsymbol{Z}_l \exp(\boldsymbol{\beta}' \boldsymbol{Z}_l)}{w_j(t_j) \exp(\boldsymbol{\beta}' Z_j) + \sum_{l \in R_{j,sc}} w_l(t_j) \exp(\boldsymbol{\beta}' \boldsymbol{Z}_l)} \right\}, \qquad (2.10)$$

where $w(\cdot)$ is a weight function that varies depending on at-risk and subcohort status. For cases, $w_i = 1$ at their own event time. Define α as the subcohort sampling proportion. Prior to their own event time, subcohort cases receive a weight of $w_i = 1/\alpha$, while non-subcohort cases receive $w_i = 0$. Subcohort controls have a weight of $1/\alpha$ before their censoring times.

Further methodological developments have since been made to expand and improve estimation in case-cohort designs. For one, Kalbfleisch and Lawless (1988) discussed parametric models with maximum likelihood estimation for case-cohort designs and other retrospective schemes. Kulich and Lin (2000) proposed an additive hazards model in the case-cohort design, an analog of the proportional hazards model for the risk difference regression. Aside from modeling hazard functions, Nan et al. (2006) utilized a linear regression model, with modified linear ranks tests and pseudolikelihood-type estimating equations. Sitlani et al. (2020) discussed robust Cox model estimation which handle outliers and violations of the PH assumption.

2.2.1 Variance Estimation in Case-Cohort Designs

Different case-cohort methods also differ in their approach to variance estimation. We may broadly categorize the approaches as model-based and robust variance estimation. Other works use bootstrapping to calculate variance, for example in Wacholder et al. (1989), but bootstrapping may be more time-consuming and computationally burdensome.

For model-based variance estimation, estimates generally include two components: variance of the full cohort analysis and additional correlation induced from case-cohort sampling. When performing analysis on full cohort data, covariance between score contributions at different event times are zero when conditioning on the filtration, which simplifies variance estimation. This is no longer true under the case-cohort design due to the inclusion of the non-subcohort cases as part of the biased sample.

Prentice (1986) accounts for additional correlation by analytically deriving the covariance between score components. We have that the variance of the score consists of variances of each risk set's score contributions plus covariances between score contributions from different risk sets:

$$\text{var}\{U(\boldsymbol{\beta})\} = \sum_{j=1}^{D} \left(\text{var}\{U_j(\boldsymbol{\beta})\} + 2 \sum_{\{k|t_k<t_j\}} \text{cov}\{U_k(\boldsymbol{\beta}), U_j(\boldsymbol{\beta})\} \right). \tag{2.11}$$

Define $\delta_i = 1$ if $N_i(t_i) \neq N_i(t_i^-)$ and 0 otherwise, $\Delta(t) = 1$ if failure occurs outside of the subcohort at time t, and $v_{jj} = \text{var}\{U_j(\boldsymbol{\beta})|\mathscr{F}\left(t_j\right), N(t_j) \neq N(t_j^-)\}$. $\text{cov}\{U_k(\boldsymbol{\beta}), U_j(\boldsymbol{\beta})\}$ is present only when $\delta_k = \delta_j = 1$. If the failure occurred in the subcohort, $E\{U_k(\boldsymbol{\beta})U_j(\boldsymbol{\beta})|\mathscr{F}(t_j)$, $N(t_j) \neq N(t_j^-)\} = 0$. If the failure occurred outside the subcohort, then define $v_{kj} = E\{U_k(\boldsymbol{\beta})U_j(\boldsymbol{\beta})|\mathscr{F}(t_j), N(t_j) \neq N(t_j^-)\}$. For both sampling schemes, we have that

$$\tilde{V}(\hat{\boldsymbol{\beta}}) = \sum_{j=1}^{n} \delta_j \left\{ v_{jj} + 2\Delta\left(t_j\right) \sum_{(k|t_k<t_j)} \delta_k v_{kj} \right\},$$

and the maximum pseudolikelihood estimator, $\hat{\boldsymbol{\beta}}$, is asymptotically normal with mean $\boldsymbol{\beta}_0$, with the estimator of variance as

$$\text{var}(\hat{\boldsymbol{\beta}}) = I(\hat{\boldsymbol{\beta}})^{-1}\tilde{V}(\hat{\boldsymbol{\beta}})I(\hat{\boldsymbol{\beta}})^{-1}. \tag{2.12}$$

In Appendix A, we provide forms of the v_{jj}, v_{kj} terms.

Self and Prentice (1988) derived through the Taylor series expansion that $n^{-1/2}(\hat{\boldsymbol{\beta}} - \boldsymbol{\beta}_0)$ converges to a Normal distribution with mean 0 and variance

$$\text{var}(\hat{\boldsymbol{\beta}}) = \Sigma(\boldsymbol{\beta}_0)^{-1} + \Sigma(\boldsymbol{\beta}_0)^{-1}\Delta(\boldsymbol{\beta}_0)\Sigma(\boldsymbol{\beta}_0)^{-1}, \tag{2.13}$$

where $\boldsymbol{\beta}_0$ is the true coefficient, $\Sigma(\boldsymbol{\beta}_0)$ is the information matrix, and $\Delta(\boldsymbol{\beta}_0)$ accounts for covariance contributions induced by case-cohort sampling. $\Sigma(\boldsymbol{\beta}_0)$ and $\Delta(\boldsymbol{\beta}_0)$ may each be consistently estimated.

Alternatively, a class of robust variance estimators were proposed by Barlow (1994) and Lin and Ying (1993), among others. In Barlow (1994), the author developed a sampling scheme and a robust variance estimator that is the jackknife estimate of the variance of the individual influence function. The jackknife is a resampling method that leaves out each observation in turn, calculates the parameter estimate on the remaining observations, and aggregates these calculations. By estimating the influence of each individual on $\hat{\boldsymbol{\beta}}$ and on the overall score, this estimator is computationally efficient and allows for more complicated sampling mechanisms.

Specifically, the conditional probability that individual i fails at failure time t_j given that someone failed at time t_j is

$$p_i(t_j) = \frac{Y_i(t_j)w_i(t_j)\exp\{\boldsymbol{\beta}'\boldsymbol{Z}_i(t_j)\}}{\sum_{l=1}^{n} Y_l(t_j)w_l(t_j)\exp\{\boldsymbol{\beta}'\boldsymbol{Z}_l(t_j)\}}, \tag{2.14}$$

and the conditional expectation of the covariate is

$$E_Z(t_j) = \sum_{k=1}^{n} p_k(t_j)Z_k(t_j). \tag{2.15}$$

The influence of an individual observation on the overall score for observation i at time t_0 is $c_i(t_0) = \int_0^{t_0} Y_i(t)[dN_i(t) - \lambda_i(t)][\boldsymbol{Z}_i(t) - E_Z(t)]d\bar{N}(t)$, which can be estimated by

$$\hat{c}_i(t_0) = \int_0^{t_0} Y_i(t)[dN_i(t) - \hat{p}_i(t)][\boldsymbol{Z}_i(t) - \hat{E}_Z(t)]d\bar{N}(t). \tag{2.16}$$

The estimate of the effect of deletion of observation i on $\hat{\boldsymbol{\beta}}$, denoted by $\hat{\boldsymbol{\beta}}_{-i}$, is $\hat{e}_i = \hat{\boldsymbol{\beta}} - \hat{\boldsymbol{\beta}}_{-i} = I^{-1}(\hat{\boldsymbol{\beta}})\hat{c}_i(t_0)$, where $I(\hat{\boldsymbol{\beta}}) = \sum_t \sum_i \hat{p}_i(t))[\boldsymbol{Z}_i(t) - \hat{E}_Z(t)][\boldsymbol{Z}_i(t) - \hat{E}_Z(t)]'$. Since the mean of $\hat{c}_i(t_0)$ is zero, the infinitesimal jackknife estimator of the variance of the influence function $c(t_0)$ can be estimated by $\hat{V}(\hat{\boldsymbol{\beta}}) = \frac{1}{n}\sum_i \hat{c}_i(t_0)\hat{c}_i(t_0)'$, so the robust variance estimator is

$$\text{var}(\hat{\boldsymbol{\beta}}) = I(\hat{\boldsymbol{\beta}})^{-1}\hat{V}(\hat{\boldsymbol{\beta}})I(\hat{\boldsymbol{\beta}})^{-1} = \frac{1}{n}\sum_i \hat{e}_i\hat{e}_i'.$$

Lin and Ying (1993), on the other hand, developed the variance estimator to deal with missing data in covariate measurements using inverse probability weighted estimation of the

cohort total of the efficient influence function contributions.

These robust variance estimators are useful for their robustness and ability to handle modeling mis-specification as we continue to incorporate stratified case-cohort sampling or when we later consider multiple outcomes.

2.2.2 Efficiency in Case-Cohort Designs

There has been considerable focus on improving estimation in case-cohort studies, and many methods have improved efficiency for assessing a main effect of interest. Breslow and Wellner (2007) considered Horvitz–Thompson (Horvitz and Thompson, 1952) or inverse probability weighted estimators, and described a general Horvitz–Thompson approach to semiparametric inference for data from complex probability samples. Nan and Wellner (2013) derived asymptotic properties for case-cohort studies using a general Z-estimation theory and empirical process theory. Breslow et al. (2009a) reduced the variance component induced by case-cohort sampling by adjusting sampling weights through calibration, to either known or estimated auxiliary variables available for the full cohort. Breslow et al. (2009b) also discussed using survey sampling methods and information from the full cohort to improve efficiency in case-cohort estimation. Building on the ideas of generalized regression estimation using auxiliary variables, Lumley et al. (2011) related these estimators to augmented inverse probability weighted estimators and constructed calibration estimators. Nan (2004) proposed a one-step efficient estimator for case-cohort designs based on efficient score functions and proved its asymptotic properties. Choi and Kang (2025) performed a balanced sampling of controls in case-cohort designs, and used delta-betas as auxiliary variables to improve efficiency.

Stratified sampling in case-cohort designs refers to basing subcohort sampling proportions on covariate strata defined in the full cohort. Borgan et al. (2000) considered covariate-stratified sampling, which enables us to sample into the subcohort a desired covariate distribution,

and developed three estimators. Nan et al. (2006) proposed a method for linear regression in stratified case-cohort designs, while Noma and Tanaka (2017) considered logistic regression in stratified case-cohort designs when the outcome is binary. Hu et al. (2014) provided sample size and power calculations for a stratified case–cohort design.

2.3 Multivariate Case-Cohort Designs

In analyzing time-to-event outcomes, researchers may further be interested in the time to multiple events or endpoints. For one, we may want to analyze clustered or correlated survival times for the same disease among family members or those in the same treatment center. There also exist recurrent events of the same type, for example, the recurrence of strokes on the same patient. In the context of AD research, we may be interested in the progression of disease, which can be categorized as the progression from normal to mild cognitive impairment (MCI) and from MCI to dementia. We could classify this as recurrent events of different types. There are mainly two types of models to handle such problems: frailty ("random effects") models and marginal ("population-average") models.

In frailty models, covariate effects are interpreted at the cluster level (e.g., individuals or treatment centers). Most commonly, researchers will utilize a shared frailty model with random effects that are assumed to follow a gamma distribution with mean one and unknown variance (Amorim and Cai, 2015). This is a common choice because the gamma distribution can be a conjugate prior and the parameter can be interpreted as a time-independent cross-ratio function, which is a natural extension of the odds ratio (Hsu et al., 2007). Interpretation of regression coefficients in frailty models, however, is on the individual level, and there are debates on how many events and subjects are needed for stable estimates (Amorim and Cai, 2015). One Bayesian model for incorporating random effects and a cure fraction was developed by Chen et al. (2002). Additional frailty models include Gorfine and Hsu (2011), Hsu et al. (2007), and more.

Within random effects models, we also have a class of proportional odds models, which can be thought of as proportional hazards models with unobservable heterogeneity (Murphy et al., 1997). The treatment effect decreases gradually and the homogeneity of the population increases over time. Such works include Bennett (1983a), Bennett (1983b), Prentice et al. (1981), Cheng et al. (1995), etc.

For analysis of bivariate failure times, scholars have developed copula models (c.f., Shih and Louis (1995), Othus and Li (2010)). A review can be found in Georges et al. (2001).

2.3.1 Marginal Models for Multivariate Analysis

In marginal models, covariate effects are interpreted at the population level. Marginal models that assume multiplicative hazards can be a direct extension of the Cox PH model. For one, the counting process formulation described in Andersen and Gill (1982) can be extended to the multivariate case. This AG model uses a scale of total time, which means the time to an event since entry into the study, and assumes a common baseline hazard function and global covariate effects across events (Amorim and Cai, 2015).

Prentice et al. (1981) formulated two marginal models for recurrent events, one under total time and the other under gap time, which means the time from the immediately preceding failure. Contrasting from Andersen and Gill (1982), the Prentice et al. (1981) models accommodated varying baseline hazard functions and varying covariate effects for each event occurrence. In the gap time model, subjects were only permitted to contribute once to each event occurrence. We summarize these models here.

Let $Z(t) = \{z(u) : u \leq t\}$ be the covariate process up to time t and $N(t) = \{n(u) : u \leq t\}$, where $n(u)$ is the number of events for one subject prior to time u. The hazard or intensity function at time t, or the instantaneous rate of failure at time t given the covariate and counting processes at time t is

$$\lambda\{t|N(t), Z(t)\} = \lim_{\Delta t \to 0} Pr\{t \leq T_{n(t)+1} < t + \Delta t|N(t), Z(t)\}/\Delta t. \tag{2.17}$$

In the absence of covariates, $\lambda\{t|N(t), Z(t)\} = \lambda_0(t)$ for some $\lambda_0(\cdot) \geq 0$ can be the intensity function for a nonhomogeneous Poisson process. We revisit using a Poisson process to model recurrent events in Section 2.4.

The two models proposed are

1. $\lambda\{t|N(t), Z(t)\} = \lambda_{0s}(t) \exp\{z(t)\beta_s\}$,

2. $\lambda\{t|N(t), Z(t)\} = \lambda_{0s}(t - t_{n(t)}) \exp\{z(t)\beta_s\}$,

corresponding to the total and gap time models, respectively, and where $\lambda_{0s}(\cdot)$ are arbitrary baseline intensity functions and $s = s\{N(t), Z(t), t\}$ is a stratification variable that may change as a function of time for a given subject, β_s is a vector of stratum-specific regression coefficients. Here, stratum refers to event occurrence. The partial likelihood functions are then

$$L_{total}(\beta) = \prod_{s \geq 1} \prod_{i=1}^{d_s} \frac{\exp\{z_{si}(t_{si})\beta_s\}}{\sum_{l \in R(t_{si}, s)} \exp\{z_l(t_{si})\beta_s\}}, \tag{2.18}$$

$$L_{gap}(\beta) = \prod_{s \geq 1} \prod_{i=1}^{k_s} \frac{\exp\{z_{si}(t_{si})\beta_s\}}{\sum_{l \in \tilde{R}(u_{si}, s)} \exp\{z_l(t_l + u_{si})\beta_s\}}, \tag{2.19}$$

respectively, where $R(t, s)$ denotes the risk set in stratum s just prior to time t, $\tilde{R}(u, s)$ denotes the risk set in stratum s at gap time $u - 0$, and t_l is the last event time on subject i prior to entry into stratum s. We may apply asymptotic likelihood methods (Cox, 1975) or

an approximate likelihood (Breslow, 1974) to the above functions.

Wei et al. (1989) modeled each marginal distribution of failure time with a multiplicative hazards model and imposed no particular structure on the dependence. For failure type $k = 1, \ldots, K$ and subject $i = 1, \ldots, n$, $\tilde{X}_{ki}$ is the failure time, C_{ki} is the censoring time, $X_{ki} = min(\tilde{X}_{ki}, C_{ki})$ is the observed time, and Δ_{ki} is the failure indicator. Conditional on covariates $Z_{ki}(t)$, failure time vector $(\tilde{X}_{1i}, \ldots, \tilde{X}_{Ki})$ and censoring time vector $(C_{1i}, \ldots, C_{Ki})$ are independent. The hazard function for the kth type of failure of the ith subject is

$$\lambda_{ki}(t) = \lambda_{k0}(t) \exp\{\beta_k Z_{ki}(t)\}, t \geq 0, \tag{2.20}$$

where $\lambda_{k0}(t)$ is an unspecified baseline hazard function and regression coefficients β_k are specific to the kth failure type. Then the kth failure-specific partial likelihood function is

$$L_{WLW,k} = \prod_{i=1}^{n} \frac{\exp\{\beta' Z_{ki}(X_{ki})\}}{\sum_{l \in R_k(X_{ki})} \exp\{\beta' Z_{kl}(X_{ki})\}}, \tag{2.21}$$

where $R_k(t)$ is the risk set just prior to time t with respect to the kth failure type. We can solve for the maximum partial likelihood estimator $\hat{\beta}_k$. Wei et al. (1989) showed that as $n \to \infty$, the vector $(\hat{\beta}_1', \ldots, \hat{\beta}_K')$ follows a normal distribution with mean vector $(\beta_1', \ldots, \beta_K')$ and covariance matrix estimates as follows.

Let $N_{ki}(t) = I\{X_{ki} \leq t, \Delta_{ki} = 1\}$, $Y_{ki}(t) = I\{X_{ki} \geq t\}$, $M_{ki}(t) = N_{ki}(t) = \int_0^t Y_{ki}(u)\lambda_{ki}(u)du$. Then by the Taylor series expansion of the score $U_k(\beta_k; \infty)$ around β_k, we have that $n^{-1/2}U_k(\beta_k; \infty) = \hat{A}_k(\beta_k^*)n^{1/2}(\hat{\beta}_k - \beta_k)$, and

$$\hat{A}_k(\beta_k) = n^{-1}\sum_{j=1}^{n}\Delta_{kj}\left[\frac{\sum_{i=1}^{n}Y_{ki}(X_{kj})Z_{ki}(X_{kj})^{\otimes 2}\exp\{\beta' Z_{ki}(X_{kj})\}}{\sum_{i=1}^{n}Y_{ki}(X_{kj})\exp\{\beta' Z_{ki}(X_{kj})\}}\right. \tag{2.22}$$

$$\left. -\left(\frac{\sum_{i=1}^{n}Y_{ki}(X_{kj})Z_{ki}(X_{kj})\exp\{\beta' Z_{ki}(X_{kj})\}}{\sum_{i=1}^{n}Y_{ki}(X_{kj})\exp\{\beta' Z_{ki}(X_{kj})\}}\right)^{\otimes 2}\right], \tag{2.23}$$

where $a^{\otimes 2}$ denotes aa' and β_k^* is on the line segment between $\hat{\beta}_k$ and β_k. By Andersen and Gill (1982), $\hat{A}_k(\beta_k^*)$ can be consistently estimated by $\hat{A}_k(\beta_k)$.

The asymptotic covariance matrix can be consistently estimated by

$$\hat{A}_k^{-1}(\hat{\beta}_k)\hat{B}_k(\hat{\beta}_k,\hat{\beta}_l)\hat{A}_l^{-1}(\hat{\beta}_l), \tag{2.24}$$

where $\hat{B}_k(\hat{\beta}_k,\hat{\beta}_l) = n^{-1}\sum_{j=1}^{n}W_{kj}(\hat{\beta}_k)W_{lj}(\hat{\beta}_l)'$, and

$$W_{kj}(\beta_k) = \Delta_{kj}\left\{Z_{kj}(X_{kj}) - \frac{S_k^{(1)}(\beta_k;X_{kj})}{S_k^{(0)}(\beta_k;X_{kj})}\right\} \tag{2.25}$$

$$-\sum_{m=1}^{n}\frac{\Delta_{km}Y_{kj}(X_{km})\exp\{\beta_k' Z_{kj}(X_{km})\}}{nS_k^{(0)}(\beta_k;X_{kj})}\times\left\{Z_{kj}(X_{km}) - \frac{S_k^{(1)}(\beta_k;X_{km})}{S_k^{(0)}(\beta_k;X_{km})}\right\}, \tag{2.26}$$

where $S_k^{(d)}(\beta;t) = n^{-1}\sum_{i=1}^{n}Y_{ki}(t)Z_{ki}(t)^d\exp\{\beta' Z_{ki}(t)\}$, $d = 0, 1$. Hence the covariance between failure times of the kth and lth type of event are accounted for through a robust variance estimator.

Lee et al. (1992) proposed estimation and variance-covariance estimators for marginal models

with the independence assumption. Then, Spiekerman and Lin (1998) extended the Wei et al. (1989) and Lee et al. (1992) models to include a covariate vector $\boldsymbol{Z}_{kl}$ so that the marginal hazard function can depend on the lth component of the kth type of failure. The strata are correlated and failure times within each stratum are clustered.

Other discussions of multivariate survival methods and marginal models include Lin (1994), Liang et al. (1993), and Prentice and Hsu (1997). Cai and Prentice (1995) and Cai and Prentice (1997) developed weighted estimating equations with varying and common baseline hazards, respectively, with the goal of improving efficiency. Peng et al. (2007) extended the marginal model to include a cure fraction, terming it a mixture cure model. As alternatives to the Cox PH modeling approach, there are also developments in the accelerated failure time (AFT) model with the GEE approach, for example in Chiou et al. (2015) and Liya et al. (2021).

We note that these models make different assumptions about the baseline hazard and how covariate effects apply. We may consider common or varying baseline hazards for multiple outcomes and consider common or varying covariate effects across the outcomes. For example, the formulation in Wei et al. (1989) is helpful when we do not want to place assumptions on the correlation structure, and instead utilize robust sandwich-type variance estimators to deal with the induced variance.

2.3.2 Recurrent Events Analysis

While there exist the aforementioned methods on analyzing time-to-event for multiple diseases concurrently or for clustered time-to-events for different subjects, we focus on the recurrent events setting, which is most relevant to our motivating area. We use the term "recurrent events" to encompass both events of the same type (e.g., stroke recurrences in the same subject) and events of different but related, possibly sequential types (e.g., stages of disease progression in AD).

Many multivariate survival analysis methods can be applied in recurrent events. We shall, however, exercise care in defining the risk set and other quantities. For example, in a univariate survival setting, subjects are not considered to be at risk after experiencing an event. In a recurrent events setting, subjects may continue to be at risk after their first event and may only be at risk for the next event after they experienced the previous event.

Again, the methods may be broadly categorized into marginal models and frailty models. Additionally, there are what researchers term "multi-state" models, which can allow subjects to transition between states (alive and deceased, for example), where Andersen and Gill (1982) and Prentice et al. (1981) models may be applied to model transition intensities (Amorim and Cai, 2015).

2.3.3 Case-cohort Designs for Recurrent Events Analysis

With the efficiency and cost-saving advantages of the case-cohort design, it is natural to consider an extension into employing it in the recurrent events setting. Maitra et al. (2020) developed a proportional rates model to handle estimation in case-cohort designs for recurrent events.

Consider n subjects in the cohort. For individual i, let $N_i^*(t)$ be the number of recurrent events on $[0, t)$, then let $Y_i(t)$ be the at-risk process and $N_i(t) = N_i^*(t \wedge C_i)$. Let T_{ij}^* denote the j-th recurrent event time for individual i and C_i denote the censoring time. In this formulation, all events for the same individual are in the same counting process. Observed time is then $T_{ij} = (T_{ij}^* \wedge C_i), j = 1, 2, \ldots, n_i + 1$, where n_i is the number of events observed for individual i. We also note that Δ_i is the indicator that individual i experienced *at least* one event. $\boldsymbol{Z}_i(\cdot)$ is the p-dimensional covariate of interest.

The rate function is defined as $E(dN^*(t)|\boldsymbol{Z}(t)) = d\mu_z(t)$, and the proportional rates model outlined is

$$\mu_Z(t;\beta_0) = \int_0^t \exp\{\beta_0' \boldsymbol{Z}(u)\} d\mu_0(u), \tag{2.27}$$

where $\mu_0(\cdot)$ is an unspecified continuous baseline mean function. Let ζ_i denote the subcohort membership indicator and $\tilde{n}/n$ is the subcohort proportion. The weighted score function is

$$U(\beta) = \sum_{i=1}^{n} \int_0^\tau \Big\{ \boldsymbol{Z}_i(t) - \frac{\tilde{S}^{(1)}(\beta,t)}{\tilde{S}^{(0)}(\beta,t)} \Big\} dN_i(t), \tag{2.28}$$

where $\tilde{S}^{(d)}(\beta,t) = n^{-1} \sum_{i=1}^n w_i(t) Y_i(t) \boldsymbol{Z}_i(t)^{\otimes d} e^{\beta' \boldsymbol{Z}_i(t)}$, $d = 0, 1$. The weight function is $\Delta_i + (1-\Delta_i)\frac{\zeta_i}{\hat{\alpha}(t)}$, where $\hat{\alpha}(t) = \frac{\sum_i (1-\Delta_i)\zeta_i Y_i(t)}{\sum_i (1-\Delta_i) Y_i(t)}$ estimates the sampling proportion for controls.

The authors proposed that subcohort sampling be done, either randomly or with covariate-stratified sampling, on the subject level. This indicates that if a subject is sampled into the subcohort or experienced one or more events, all of their records are included in the case-cohort analysis.

We consider alternative ways to perform subcohort sampling on the record level. This is especially motivated by differential baseline hazards in disease progressions from different stages, the potential of new measurements taken at new event times, and the decrease of risk set size for later disease progressions, which can be common in AD biomarker discovery studies. We investigate the inclusion of event-specific covariate measurements and assessing event-specific effects through a framework of sampling and analysis in Chapter 4. We therefore review some motivating literature on sampling designs in the next section.

2.3.4 Sampling Considerations

Zhang et al. (2011) developed a sampling method for performing inference across clusters of patients being treated in various treatment centers. The authors proposed methods with three types of Bernoulli sampling. Consider a full cohort of independent clusters, where each cluster has m_i correlated subjects. Assume that subjects within the same cluster are exchangeable. To obtain the subcohort, there are three ways to perform sampling:

1. Randomly sample individuals from each cluster with Bernoulli sampling.

2. Randomly sample clusters from the full cohort with Bernoulli sampling.

3. Randomly sample clusters from the full cohort with Bernoulli sampling, then randomly sample subjects with Bernoulli sampling from the selected clusters.

Zhang et al. (2011) assumes a marginal proportional hazards model with a common baseline hazard and common regression coefficient across clusters. If subcohort sizes are approximately equal, the first design results in more efficient estimators.

Under the univariate setting, Hu et al. (2014) developed a stratified logrank-type statistic for performing analysis in case-cohort designs. Moreover, the authors propose designs to perform sampling within various strata, including

1. *Proportional design.* The size of subjects in the subcohort at each stratum is sampled to be proportional to that in the full cohort.

   With n total subjects, n_l subjects in each stratum, p_{lj} the proportion of subjects in group j in stratum l, p_{Dl} the failure proportion in stratum l, $v_l = n_l/n$, to detect a log hazard ratio of γ with power β and significance level α, the minimal required subcohort size is

$$\tilde{n} = \left[\frac{n \sum_{l=1}^{L} (p_{l1} p_{l2} p_{Dl}^2 v_l / (1 - p_{Dl}/2))}{B_2^2 - \sum_{l=1}^{L} (p_{l1} p_{l2} p_{Dl} v_l (1 - p_{Dl}/(1 - p_{Dl}/2)))} \right], \tag{2.29}$$

where $[x]$ is the smallest integer bigger than x, $B_2 = \frac{n^{1/2} \gamma \sum_{n=1}^{L} p_{l1} p_{l2} p_{Dl} v_l}{Z_{1-\alpha/2} + Z_\beta}$. Then for each stratum we sample $\tilde{n}_l = \tilde{n} v_l$.

2. *Balanced design.* The number of subjects in the subcohort is the same across each stratum. Here, the minimal required subcohort size is

$$\tilde{n} = \left[\frac{Ln \sum_{l=1}^{L} (p_{l1} p_{l2} p_{Dl}^2 v_l / (1 - p_{Dl}/2))}{B_2^2 - \sum_{l=1}^{L} (p_{l1} p_{l2} p_{Dl} v_l (1 - p_{Dl}/(1 - p_{Dl}/2)))} \right]. \tag{2.30}$$

For each stratum we sample $\tilde{n}_l = \tilde{n}/L$.

These designs show improved flexibility and efficiency compared with simple random sampling, so it would be advantageous to investigate possible efficiency gains under the recurrent events setting.

2.4 Power and Sample Size Calculations in Survival Analysis

2.4.1 Overview of Power and Sample Size Calculations

Before beginning any study involving human participants, it is important to carefully plan and design the study to ensure its safety, ethical conduct, and efficiency. A critical consideration in the planning stage is the determination of sample size, or, in the case of human

research, the required number of individuals to enroll. Adequate sample size estimation helps ensure sufficient statistical power for inference while conserving time and resources. Combined with information about accrual and event rate, sample size estimation also gives researchers an idea of study length.

While the case-cohort design can be considered "retrospective" as all event/censoring information needs to be observed on the full cohort before we begin the analysis, there is still a critical need to plan and determine the required sample size, which is often the size of the subcohort needed. It can be difficult to estimate sample size and calculations largely rely on assumptions, which becomes even more challenging in case-cohort studies for recurrent events analysis, where the final analysis set relies on many factors including event rate, subcohort size, and correlations between observations for the same individual.

In a hypothesis test, a type I error is defined as an error of incorrectly rejecting the null hypothesis and a type II error is defined as failing to reject the null hypothesis when it is false. A power function is the probability that the observed sample belongs to the rejection region of a test for a given parameter γ. For a fixed sample size, the general approach to finding a good test is to fix type I error probability at a small level and find a test with the smallest type II error probability. The power of a test, β^P, is the probability of correctly rejecting the null hypothesis when the null hypothesis is false, and is equal to one minus the type II error probability. In developing the power and sample size formula for an analysis, one goal is to determine the sample size needed for the test to have adequate power to detect an effect of a specified magnitude at a prespecified significance level given some assumptions placed on other study parameters (StataCorp, 2025).

In time-to-event data settings, it is common to determine power and sample size based on the logrank test. The logrank test is a nonparametric test of whether the survival distributions of two or more independent groups differ significantly (Mantel, 1966). This section provides background for methods in Chapter 5, which develop power and sample size (PSS) formula

for recurrent events case-cohort designs based on the logrank test statistic.

2.4.2 Univariate Full Cohort Power and Sample Size Calculations

The logrank test compares the survival functions or equivalently, the cumulative hazard functions, of two independent groups, to test if there is a significant difference. While the final study analysis may use a Cox PH model with covariate adjustment, it is standard practice to use the logrank test and its asymptotic properties for sample size determination prior to the study. For a continuous biomarker of interest, we may dichotomize it into two groups, for example, low versus high levels of the biomarker based on the median or some predetermined threshold. Other possible groups of interest may be treatment versus control groups. Under these somewhat simplified assumptions, it is possible to determine sample size more straightforwardly. The power of the logrank test is related to the number of events observed in the study, and in the presence of censoring, the probability of censoring needs to be estimated to obtain the final estimate of sample size (StataCorp, 2025).

For the univariate full cohort analysis, both Freedman (1982) and Schoenfeld (1981) derive formulas based on the asymptotic distribution of the logrank test statistic. Freedman (1982) expresses the asymptotic mean and variance of the logrank test statistic as a function of the true hazard ratio, while Schoenfeld (1981) writes them as a function of the true log hazard ratio. When the hazard ratio is close to one, the two methods perform similarly. In general, Freedman (1982) tends to slightly overestimate the sample size (StataCorp, 2025).

Suppose there are two groups, indexed $j = 1, 2$, with n_j individuals in each group. Then the full cohort size is $n = n_1 + n_2$. Let T_{ij} and C_{ij} be the failure and censoring time for individual i in group j, respectively. Then, $X_{ij} = T_{ij} \wedge C_{ij}$ is the observed time and $\Delta_{ij} = I(T_{ij} \leq C_{ij})$ is the failure indicator for individual i in group j. Let the cumulative hazard functions be $\Lambda_j(t)$ for group j. Under the proportional hazards assumption, we want to test $H_0 : \Lambda_1(t) = \Lambda_2(t)$, $t \in [0, \tau]$.

Let $\bar{Y}_j(t)$ be the number of subjects at risk at time t in group j, $w(t)$ be a weight function, $N_{ij}(t)$ be the counting process, and $\bar{N}_j(t) = \sum_{i=1}^{n_j} N_{ij}(t)$. Then the logrank test statistic is

$$\sum_{i=1}^{n_1} \frac{\Delta_{i1} w(X_{i1}) \bar{Y}_2(X_{i1})}{\bar{Y}_1(X_{i1}) + \bar{Y}_2(X_{i1})} - \sum_{i=1}^{n_2} \frac{\Delta_{i2} w(X_{i2}) \bar{Y}_1(X_{i2})}{\bar{Y}_1(X_{i2}) + \bar{Y}_2(X_{i2})}, \tag{2.31}$$

or, equivalently in the counting process notation,

$$\int_0^t w(s) \frac{\bar{Y}_1(s)\bar{Y}_2(s)}{\bar{Y}_1(s) + \bar{Y}_2(s)} \times \left\{ \frac{d\bar{N}_1(s)}{\bar{Y}_1(s)} - \frac{d\bar{N}_2(s)}{\bar{Y}_2(s)} \right\}, \tag{2.32}$$

which is asymptotically normal. By Schoenfeld (1981), for a given power β^P, two-sided significance level α, proportions of groups $j = 1, 2$ in the cohort of p_j, failure rate p_D, and to detect a log hazard ratio of γ, the required sample size is

$$n = \frac{(z_{1-\alpha/2} + z_{\beta^P})^2}{\gamma^2 p_1 p_2 p_D}, \tag{2.33}$$

and the power function for a size n cohort is

$$\Phi(z_{\alpha/2} + \gamma\sqrt{n p_1 p_2 p_D}), \tag{2.34}$$

where $\Phi(\cdot)$ is the standard Normal cumulative distribution function.

2.4.3 Univariate Case-Cohort Design Power and Sample Size Calculations

For a case-cohort design under the univariate setting, Cai and Zeng (2004) developed a logrank-type test and approximations to obtain power and sample size calculations.

Assume that the full cohort has size $n = n_1 + n_2$, where there is a dichotomized expensive exposure labeled as group $j = 1, 2$ with n_j subjects in each, for example, high or low biomarker levels based on some threshold. Let T_{ij} and C_{ij} denote the failure time and censoring time for subject i in group j, $T_{ij} \perp C_{ij}$. Let $X_{ij} = T_{ij} \wedge C_{ij}$ be the observed time, $\Delta_{ij} = I(T_{ij} \leq C_{ij})$ be the failure indicator. We take a random sample as the subcohort, and let ζ_{ij} be the subcohort indicator for subject i in group j and subcohort size be $\tilde{n}$. We test the null hypothesis $H_0 : \Lambda_1(t) = \Lambda_2(t), t \in [0, \tau]$, where $\Lambda_j(t)$ is the cumulative hazard function of group j.

When using a case-cohort design, $\tilde{Y}_j(t)$, the number of subjects at risk in group j in the subcohort, can be used to estimate the full cohort quantity $\bar{Y}_j(t)$. If the sampling fraction is q, we may approximate the full cohort quantity $\bar{Y}_j(t)$ by $\tilde{Y}_j(t)/q$. Then in the univariate case-cohort design, the logrank-type test statistic is

$$SP_n = \sum_{i=1}^{n_1} \frac{\Delta_{i1} w(X_{i1}) \tilde{Y}_2(X_{i1})}{\tilde{Y}_1(X_{i1}) + \tilde{Y}_2(X_{i1})} - \sum_{i=1}^{n_2} \frac{\Delta_{i2} w(X_{i2}) \tilde{Y}_1(X_{i2})}{\tilde{Y}_1(X_{i2}) + \tilde{Y}_2(X_{i2})}. \tag{2.35}$$

When n_j/n approximates some $p_j \in (0, 1)$ and $\tilde{n}/n$ converges to q, under H_0, $n^{-1/2} SP_n$ is asymptotically normal with asymptotic variance $\hat{\sigma}^2_{SP} = \hat{\sigma}^2_{UFC} + \hat{\sigma}^2_{UCC}$ where

$$\hat{\sigma}^2_{UCC} = \frac{2(1-\hat{q})}{n} \times \sum_{j=1}^{2}\sum_{i=1}^{n_j}\Bigg\{ \frac{\Delta_{ij} w(X_{ij})\tilde{Y}_1(X_{ij})\tilde{Y}_2(X_{ij})}{(\tilde{Y}_1(X_{ij})+\tilde{Y}_2(X_{ij}))^2} \tag{2.36}$$

$$\times \left[\sum_{j'=1}^{2}\sum_{i'=1}^{n_j} \frac{\Delta_{i'j'} w(X_{i'j'}) I(X_{i'j'} \le X_{ij})}{\tilde{Y}_1(X_{i'j'})+\tilde{Y}_2(X_{i'j'})}\right]\Bigg\} \tag{2.37}$$

$$- \frac{(1-\hat{q})}{n}\sum_{j=1}^{2}\sum_{i=1}^{n_j} \frac{\Delta_{ij} w(X_{ij})\tilde{Y}_1(X_{ij})\tilde{Y}_2(X_{ij})}{(\tilde{Y}_1(X_{ij})+\tilde{Y}_2(X_{ij}))^3}, \tag{2.38}$$

by Cai and Zeng (2004) and Self and Prentice (1988).

Then, to obtain a two-sided level α test of H_0 versus $H_A : \Lambda_1(t) < \Lambda_2(t), t \in [0, \tau]$, we consider $w(t) = 1$ and make the following assumptions:

1. Censoring distributions are the same in the two groups.

2. The number of failures is very small in the full cohort but much larger than 1.

3. No ties of failures are observed.

So $H_A : \Lambda_2(t) = e^{\gamma}\Lambda_1(t)$ where $\gamma = O(\tilde{n}^{-1/2})$, the log-hazard ratio of the two groups.

To simplify and obtain power and sample size formula, under the previously stated assumptions, first note that the asymptotic expectation of $n^{-1/2}SP_n$ under H_A is the same as that of the full cohort logrank test statistic. This asymptotic expectation can be approximated by $n^{-1/2}\int \frac{\bar{Y}_1(t)\bar{Y}_2(t)}{\bar{Y}_1(t)+\bar{Y}_2(t)}(d\Lambda_1(t) - d\Lambda_2(t)) \approx \tilde{n}^{-1/2}q^{1/2}\gamma p_2 D_1$, where D_j is the number of failures in group j.

To simplify $\hat{\sigma}^2_{UCC}$, order failure times with indexes $c = 1, \ldots,$ and denote the risk set size of group j in the full cohort corresponding to each failure time as $m^j_1, m^j_2, \ldots,$ $j = 1, 2$. Then $\hat{\sigma}^2_{UCC} \approx \frac{2}{q}p_1 p_2 \sum_c \frac{d_c - 1/2}{m^1_c + m^2_c}$, where d_c is the number of failure times greater than or equal to the cth failure time. With the assumption that there are few number of failures in the full

cohort, $m_c^1 + m_c^2 \approx n$, so $\hat{\sigma}^2_{UCC} \approx (n^2 q)^{-1}(1-q)p_1 p_2 (D_1 + D_2)^2$. So $\hat{\sigma}^2_{SP}$ can be approximated by $n^{-1}(p_2^2 D_1 + p_1^2 D_2) + (n^2 q)^{-1}(1-q)p_1 p_2 (D_1 + D_2)^2$.

Hence the power function can be approximated by

$$\Phi\left(z_{\alpha/2} + \gamma\sqrt{\frac{\tilde{n}p_1 p_2 p_D}{q + (1-q)p_D}}\right), \tag{2.39}$$

where p_D is the proportion of failures, or event rate, in the full cohort. Conversely, to approximate the subcohort size needed for level α, power β^P, and subcohort fraction q, we have

$$\tilde{n} = \frac{(z_{1-\alpha/2} + z_{\beta^P})^2 (q + (1-q)p_D)}{\gamma^2 p_1 p_2 p_D}. \tag{2.40}$$

As the authors pointed out, $q \leq q + (1-q)p_D \leq 1$, so the power above is between $\Phi(z_{\alpha/2} + \gamma\sqrt{\tilde{n}p_1 p_2 p_D})$ and $\Phi(z_{\alpha/2} + \gamma\sqrt{n p_1 p_2 p_D})$, which are the power estimated only using the subcohort and the power estimated using the full cohort, respectively (Cai and Zeng, 2004). This is important to note, as the case-cohort design performs analysis not only on the subcohort but also on the supplemented cases, which provide information and hence increase the power of the analysis.

2.4.4 Recurrent Events Full Cohort Power and Sample Size Calculations

We now review the logrank-type test for performing recurrent events analysis in the full cohort, as well as the power and sample size formulas based on this test statistic (Song

et al., 2008). We do not consider adjustment for additional covariates in Chapter 5, as Song et al. (2008) does, in order to use minimal assumptions in the planning stage.

Let there be an expensive exposure that can be dichotomized into groups $j = 1, 2$. For each subject i, $i = 1, 2, \ldots, n_j$, the failure times are $T_{ij} = (T_{ij1}, T_{ij2}, \ldots, T_{ijK})$, with $k = 1, \ldots, K$ events for each subject and K the maximum number of observed events for each subject. Let C_{ij} be the censoring time for each subject in group j.

Let $N_{ijk} = I\{T_{ijk} \leq t, C_{ij} \geq t\}$ and $N_{ij}(t) = \sum_{k=1}^{\infty} N_{ijk}(t) = \sup_k\{k : T_{ijk} \leq t, C_{ij} \geq t\}$. Note here that the counting process $N_{ij}(t)$ may have more than one jump. The at-risk process is $Y_{ij} = I\{C_{ij} \geq t\}$ and the probability of subjects being under study in group j at time t is $\pi_j^n = n_j^{-1} \sum_{i=1}^{n_j} EI\{C_{ij} \geq t\}$. $N_{ijk}^0(t)$ and $N_{ij}^0(t)$ are the underlying event process versions of $N_{ijk}(t)$ and $N_{ij}(t)$, where $N_{ij}(t) = \int_0^t Y_{ij}(s) dN_{ij}^0(s)$. Define $\Lambda_{ij}(t) = E\{N_{ij}^0(t)\}$ to be the expected number of events for subject i in group j by time t.

The following assumptions are needed:

1. T_{ij} and C_{ij} are independent for $i = 1, \ldots, n_j, j = 1, 2$.

2. $\lim_{n\to\infty} n_j/n = p_j \in (0, 1)$ and $\sup_{t\in(0,\tau)} |\pi_j^n(t) - \pi_j(t)| \to 0$, where τ is the maximum event time.

3. The cumulative mean functions of events for subjects in the same group are identical, $\Lambda_{ij}(t) = \Lambda_j^n(t)$, for $i = 1, 2, \ldots, n_j$, $t \in (0, \tau)$. The rate function is $\lambda_j^n(t) = \partial\Lambda_j^n(t)/\partial t$, $j = 1, 2$. Here, the superscript n permits contiguous alternatives.

4. $\sup_{t\in(0,\tau)} |\sqrt{n}\{d\Lambda_1^n(t)/d\Lambda_2^n(t) - 1\} - \psi(t)\{1 + \eta(t)\}| \to 0$, where ψ is either cadlag or caglad with bounded total variation and η is bounded and zero except at event times.

We test the hypothesis $H_0 : \Lambda_1(t) = \Lambda_2(t), t \in [0, \tau]$. The logrank test statistic is

$$L_n(t) = n^{-1/2} \int_0^t \hat{W}_n(s) \frac{\bar{Y}_1(s)\bar{Y}_2(s)}{\bar{Y}_1(s) + \bar{Y}_2(s)} \times \left\{ \frac{d\bar{N}_1(s)}{\bar{Y}_1(s)} - \frac{d\bar{N}_2(s)}{\bar{Y}_2(s)} \right\} \tag{2.41}$$

where $\bar{N}_j(t) = \sum_{i=1}^{n_j} N_{ij}(t)$ and $\bar{Y}_j(t) = \sum_{i=1}^{n_j} Y_{ij}(t)$. $\hat{W}_n$ is caglad or cadlag with total bounded variation, and nonnegative so that L_n is sensitive to ordered alternatives. Assume that $\sup_{t\in(0,\tau)}|\hat{W}_n(t) - W(t)| \to 0$ in probability for some uniformly bounded integrable function $W(t)$.

Assume that $\lim_{n\to\infty} \sigma_n^2 = \sigma^2 < \infty$. Under assumptions, we have that L_n converges in distribution to a normally distributed random variable with mean μ and variance σ, where

$$\mu = \int_0^\tau W(s)\psi(s)\{1 + \eta(s)\} \frac{p_1\pi_1(s)p_2\pi_2(s)}{p_1\pi_1(s) + p_2\pi_2(s)} d\Lambda_0(s), \tag{2.42}$$

$$\sigma_n^2 = \frac{1}{n} \sum_{j=1}^{2} \sum_{i=1}^{n_j} \left\{ \int_0^\tau W_n(s) \frac{p_{j'}\pi_{j'}(s)}{p_1\pi_1(s) + p_2\pi_2(s)} dM_{ij}(s) \right\}^2, \tag{2.43}$$

where $M_{ij}(t) = N_{ij}(t) - \int_0^t Y_{ij}(s)d\Lambda_j(s)$ and $\sigma_n^2 \to \sigma^2$ as $n \to \infty$, for some $0 < \sigma^2 < \infty$. We also have $\tilde{\sigma}_n^2 \to \sigma^2$ in probability as $n \to \infty$. For a robust variance estimator, we have

$$\hat{\sigma}_n^2 = \frac{1}{n} \sum_{j=1}^{2} \sum_{i=1}^{n_j} \left\{ \int_0^\tau \hat{W}_n(s) \frac{\bar{Y}_{j'}(s)}{\bar{Y}_1(s) + \bar{Y}_2(s)} d\hat{M}_{ij}(s) \right\}^2, \tag{2.44}$$

where $\hat{M}_{ij}(t) = N_{ij}(t) - \int_0^t \bar{Y}_{ij}(s)d\bar{N}_j(s)/\bar{Y}_j(s)$, and $i = 1, \ldots, n_j, j = 1, 2, j' = 3 - j$.

The proportional means local alternative is $H_A : \Lambda_j(t) = \Lambda_0(t)\exp\{(-1)^{j-1}\psi(t)/2\}, j = 1, 2,$ $\psi(t) \neq 0, \psi(t) = o(1), t \in (0, \tau)$. Under H_A, the asymptotic mean is

$$\mu_1 = \int_0^\tau \psi(s)\{1 + \eta(s)\}\frac{p_1\pi_1(s)p_2\pi_2(s)}{p_1\pi_1(s) + p_2\pi_2(s)}d\Lambda_0(s) \tag{2.45}$$

For variance, $\tilde{\sigma}_n^2 \to \sigma^2$ in probability as $n \to \infty$ still holds.

Now we assume that

1. $W(s) = 1$.
2. The marginal censoring distributions of the times C_{ij} are the same.
3. The number of failures is very small in the full cohort for each event occurrence, but are all much larger than 1.
4. There are no ties.

With assumption 2, we have that $\pi_1 = \pi_2 \equiv \pi_0$. Baseline cumulative mean functions are continuous and local alternatives satisfy $\psi = \gamma, \gamma \in \mathbb{R}$. $\mu_1 = \gamma p_1 p_2 D_{1g} + o(\gamma)$, where $D_{1g} = \int_0^\tau \pi_0(s)d\Lambda_0(s)$. To estimate D_{1g}, we take the geometric mean:

$$\hat{D}_{1g} = \sqrt{\frac{\sum_{i=1}^{n_1}\hat{\Lambda}_1(C_{i1})}{n_1} \times \frac{\sum_{i=1}^{n_2}\hat{\Lambda}_2(C_{i2})}{n_2}}, \tag{2.46}$$

where $\hat{\Lambda}_j(C_{ij}) = \int_0^{C_{ij}} \hat{Y}_{ij}(t)d\bar{N}_j(t)/\hat{Y}_j(t)$.

To account for the correlations within each subject, assume that N_{0ij} is a nonstationary

Poisson process with cumulative intensity function $w_{ij}\Lambda_j(t)$, where w_{ij} is a positive, latent real random variable with mean 1 and variance σ_w^2. Assume that w_{ij} is independent of the censoring process.

After some simplifications, the sample size required for the alternative $\psi = \gamma$ for a two-sided test of size α and power β^P is

$$n = \frac{(z_{1-\alpha/2} + z_{1-\beta^P})^2(\hat{D}_{1a} + \hat{\sigma}_w^2\hat{D}_2)}{\gamma^2 p_1 p_2 \hat{D}_{1g}^2}, \tag{2.47}$$

where the failure rate is now obtained through estimators of the quantities

$$D_{1a} = \sum_{j=1}^{2} \int_0^\tau p_{j'}\pi_0(s)d\Lambda_j(s), \tag{2.48}$$

$$D_2 = \sum_{j=1}^{2} p_{j'} E[\{\int_0^\tau Y_{1j}(s)d\Lambda_j(s)\}^2], j' = 3 - j, \tag{2.49}$$

which are

$$\hat{D}_{1a} = n^{-1} \sum_{j=1}^{2} \sum_{i=1}^{n_j} \hat{\Lambda}_j(C_{ij}), \tag{2.50}$$

$$\hat{D}_2 = n^{-1} \sum_{j=1}^{2} \sum_{i=1}^{n_j} \hat{\Lambda}_j^2(C_{ij}), \tag{2.51}$$

$$\hat{\sigma}_w^2 = (n^{-1} \sum_{j=1}^{2} \sum_{i=1}^{n_j} \frac{N_{ij}(C_{ij})\{N_{ij}(C_{ij}) - 1\}}{\hat{D}_2} - 1)^+, \quad x^+ = max\{x, 0\}. \tag{2.52}$$

When $p_1 = p_2 = 0.5$, D_{1a} and D_2 can be interpreted as the average number and the average squared number of observed events among the two groups. When the hazard ratio size is small, we may have the approximation that $D_{1g} \approx D_{1a}$.

The proposed test accounts for within-subject correlation among multiple events by assuming that N_{ij} follows a Poisson process, which provides a useful framework for characterizing the level of correlation in terms of overdispersion.

Chapter 3

Case-cohort designs to improve efficiency for detecting effect modification

3.1 Introduction

In large cohort analyses with time-to-event outcomes, a low event rate can motivate performing analysis on a smaller subset of the full cohort. Additionally, using efficient designs that only analyzes on a subset can save processing time and cost, as well as expensive covariates. The case-cohort design is one such design, developed by Prentice (1986). In this design, under the Cox proportional hazards model, a random sample called the subcohort is selected from the full dataset, and it is supplemented with all other individuals from the full study cohort who experience an event, termed the "cases". The estimating equation is then modified to return valid estimates. Hence, this analysis requires covariate histories only on subcohort members and supplemented cases, which will save the amount of valuable samples that must be processed. Self and Prentice (1988) developed an asymptotically

equivalent sampling scheme for the case-cohort design with efforts to obtain better theoretical properties. Later, Lin and Ying (1993) and Barlow (1994) developed methods that allow for contributions from non-subcohort cases at all event times as well as robust variance estimation.

The case-cohort design can be useful for dealing with challenges in Alzheimer's disease (AD) biomarker discovery. AD is a serious neurodegenerative disease projecting to affect 13.8 million Americans by 2060 (A&D, 2025), and identifying early indications of disease progression is important for informing patients and clinicians on early treatment options. We may estimate the association of disease biomarkers of pathologies with the time to AD symptom onset and progression, to improve early diagnosis. Additionally, biomarkers can be treatment targets and valuable outcomes for clinical trials. Some biomarkers currently thought to be related to AD pathology include amyloid beta peptide (Aβ) and phosphorylated tau (pTau). Biomarkers, however, can be of high burden and costly to obtain. For example, the cerebrospinal fluid (CSF) contains valuable biomarker measurements including measures of Aβ and pTau, and potentially new biomarkers. However, to obtain CSF samples, patients must undergo lumbar punctures and their apprehension to do so has been identified as a barrier to research participation (Witbracht et al., 2020). Furthermore, while the instances of AD are growing, AD prevention and preclinical studies can still suffer from low event rates. For the above reasons, case-cohort designs can be well utilized in AD biomarker discovery.

In this context, we seek to first assess the effect of biomarkers across the full patient population while ensuring the ability to assess potential subgroup differences in secondary analyses. To motivate this goal, a previous study has shown potentially higher AD prevalence in Black and Hispanic subpopulations (Matthews et al., 2019), yet there may be higher odds of amyloid biomarker criteria ineligibility among Black, Hispanic, and Asian participants compared to non-Hispanic (NH) White participants in four AD trials (Grill et al., 2024). This suggests the need to further investigate potential effect modification of biomarker effects by

race and ethnicity. Additionally, the Apolipoprotein E (APOE) gene $\epsilon 4$ allele is the most well-replicated genetic risk factor for AD. Specifically, having one or two copies of the APOE $\epsilon 4$ allele is associated with a higher risk of AD (A&D, 2025). To quantify whether biomarker effects can be modified by these factors, we would like to test interaction effects in secondary analyses.

One challenge in comparing subgroup biomarker effects is having limited available data, especially for rare subgroups. Across examined AD trials, a study found that the median percentage of Black, Hispanic, and Asian participants was a combined 11.2% (Franzen et al., 2022). Considering effect modification by APOE status, we encounter the analytical challenge that only around 5% of the population carry two copies of APOE $\epsilon 4$ (A&D, 2025). Further, APOE testing is not widely implemented in the general public and one recent study found that 29% of first-degree relatives of AD patients were apprehensive about testing due to negative implications of sharing genetic information (Kopits et al., 2011). Under these circumstances, especially when using the case-cohort design, we may further deplete the sample of rare subgroups. This is exacerbated in later event times as the size of the risk set continues to decrease, and as a result, we have limited power to detect interaction effects. While there are efforts to alleviate this issue through recruitment, we aim to utilize sampling and weighting methods to increase power for detecting effect modification.

Stratified sampling, or two-phase sampling, has been incorporated into case-cohort designs to improve precision in main effect estimation. To do so, we may perform biased sampling to obtain a subcohort with a desired distribution of the covariate and subsequently incorporate weights in estimation to remove bias. Borgan et al. (2000) performed sampling based on stratification of a correlate of the predictor of interest. Resulting main treatment effect estimates were shown to be more efficient compared to a randomly sampled subcohort. Breslow et al. (2009b) showed that we may further improve the precision of hazard ratio estimates by adjusting subcohort sampling weights through calibration or estimation. Choi

and Kang (2025) also made improvements in efficiency by performing balanced sampling of controls using delta-betas as auxiliary variables. On a real-world dataset, Cologne et al. (2012) considered a gene-environment interaction effect in an application of the stratified case-cohort design, showing increasing efficiency with decreasing subcohort fractions. As a guide for epidemiologists, O'Brien et al. (2022) proposed stratified case-cohort designs for assessing interaction effects for enhanced efficiency.

Our work aims to more systematically propose a framework consistent with goals in AD biomarker research. That is, useful biomarker discovery should first establish the marginal biomarker effect on AD progression, pooling as much information as possible to first identify a signal across the entire study population, while then retaining the ability to assess potential effect modification across pre-defined subgroups in secondary analyses. There are no works highlighting the impact of stratified sampling on the power for detecting a potential effect modifier while maintaining inferential validity in the marginal effect, particularly through simulation results and in varying scenarios of interaction effect sizes, subcohort sizes, and subgroup rarities. This chapter will, from a statistical standpoint, emphasize the role of such sampling methods, which are especially useful in the context of AD research but has utility across application areas.

In the following sections, we will first summarize developments in the case-cohort design. We will detail a framework to design and perform case-cohort analysis using a stratified case-cohort estimator with inverse-probability-based sampling weights, along with its theoretical properties, for increased power in subgroup analyses. Next, we perform a simulation study to illustrate the consistency of the stratified case-cohort estimates and emphasize power gains in detecting effect modification, particularly when rare subgroups are present. Finally, we investigate the heterogeneity of phosphorylated tau as a biomarker for AD progression across ethnoracial groups and APOE $\epsilon 4$ status in the Alzheimer's Disease Neuroimaging Initiative (ADNI) dataset. In addition to validating existing biomarkers, this method can

be implemented for new biomarker discovery efforts.

3.2 Statistical Methods

3.2.1 Pseudo-Likelihood Estimators

We will set up the class of pseudo-likelihood estimators for case-cohort designs and describe the estimator in case-cohort designs utilizing stratified sampling. Consider the Cox (1972) proportional hazards model. We have the hazard function $\lambda_i(t;\mathbf{Z}_i) = \lambda_0(t)\exp\{\boldsymbol{\beta}'\mathbf{Z}_i(t)\}$ for individual i, $i = 1, 2, \ldots, n$. $\lambda_0(t)$ is the baseline hazard function, $\boldsymbol{\beta}' = (\beta_1, \ldots, \beta_q)$ is a vector of regression coefficients, and $\mathbf{Z}_i'(t) = (Z_{i1}(t), \ldots, Z_{iq}(t))$ is a vector of covariates. We assume untied failure times t_j, where $t_1 < t_2 < \cdots < t_D$. We also assume that failure and censoring times are independent, conditional on the covariates.

Prentice (1986) first proposed the case-cohort design, where we randomly sample a subcohort from the full cohort of individuals. The subcohort is supplemented with all other individuals who experience an event (the "cases"). The risk set at each failure time consists of subcohort members at risk plus the non-subcohort case at its own failure time. Self and Prentice (1988) discussed a design that is asymptotically equivalent, but which includes only subcohort members at risk in the risk set. We will call these "classic case-cohort designs" to differentiate from subsequent designs with stratified sampling.

We consider a predictor of interest denoted as $Z_{i1}(t)$. At the same time, it may be of interest to consider a potential effect modifier denoted as $Z_{i2}(t)$, i.e. we model an interaction between $Z_{i1}(t)$ and $Z_{i2}(t)$. Such covariates may be categorical such as race and ethnicity, or continuous and categorized into levels. In this case, a randomly sampled subcohort may produce low power due to rarer subgroups of $Z_{i2}(t)$ and fewer individuals at risk in later event times. Instead, we may perform stratified sampling at each level of $Z_{i2}(t)$ to achieve a uniform sample distribution of $Z_{i2}(t)$ in the subcohort, which maximizes the information

on rarer subgroups.

We first partition the cohort into L levels based on $Z_{i2}(t)$, and sample approximately m_l subcohort members from n_l cohort members in each level l through binomial sampling with probability m_l/n_l. The subcohort $\mathcal{C}$ consists of approximately $m = \sum_l m_l$ individuals while the cohort $\mathcal{S}$ consists of $n = \sum_l n_l$ individuals. For our purposes, we take an approximately equal number of individuals, m_l, at each level l so that the distribution of the variable in the subcohort is approximately uniform. When there is a low number of a certain subpopulation, we may adjust how we perform stratified sampling to include all the members of the rare subpopulation. As with the classic case-cohort design, covariate history is required only on members of the subcohort and non-subcohort individuals that experience an event.

Consider the pseudo-likelihood function $\mathcal{L}_v(\beta) = \prod_{t_j} \frac{\exp\{\beta' Z_{(j)}(t_j)\} v_{(j)}(t_j)}{\sum_k Y_k(t_j) \exp\{\beta' Z_k(t_j)\} v_k(t_j)}$. $Y_k(t_j)$ is the at-risk indicator of individual k at t_j, $v_{(j)}(t_j)$ is a weight for the individual with the jth event time t_j, and $v_k(t_j)$ is a weight for individual k at t_j. The weight incorporates both the Self and Prentice (1988) sampling scheme and inverse-probability-of-sampling weights to correct for bias introduced by stratified sampling. We have

$$v_k(t_j) = \begin{cases} 0, & \text{if } k \notin \mathcal{C} \text{ at } t_j, \\ \frac{n_{l(k)}}{m_{l(k)}}, & \text{if } k \in \mathcal{C} \text{ at } t_j, \end{cases} \tag{3.1}$$

where $l(k)$ is the stratum that individual k belongs to in $Z_{i2}(t)$, so $n_{l(k)}$ is the number of individuals in this stratum in the full cohort, and $m_{l(k)}$ is the number of individuals in this stratum sampled into the subcohort. This means that if individual k is not a subcohort member, they are not counted in the risk set. If individual k is a subcohort member, they are weighted by the inverse-probability-of-sampling weight. This estimator is considered in Borgan et al. (2000) as "Estimator I".

As previously noted, a first step of biomarker discovery and validation is to quantify the association between a biomarker of interest and the event marginally across all subpopulations. That is, we consider a "marginal model" where we only model the main effects. We denote the model as $\lambda_i(t;\mathbf{Z}_i) = \lambda_0(t)\exp\{\gamma_1 Z_{i1}(t) + \gamma_2 Z_{i2}(t)\}$. Without modelling interactions, we may still obtain consistent estimates for the true marginal effect across subpopulations by incorporating sampling weights as defined above. By plugging in $\gamma_1 Z_{i1}(t) + \gamma_2 Z_{i2}(t)$, we obtain the pseudo-likelihood function for this model. We solve for estimator $\hat{\boldsymbol{\gamma}} = (\gamma_1, \gamma_2)$ by finding the solution to $\frac{\partial \log \mathcal{L}_v(\boldsymbol{\gamma})}{\partial \boldsymbol{\gamma}} = 0$. After establishing a marginal effect, we then consider an "interaction model" for secondary analysis, where interest lies in estimating potentially differential subgroup effects. In this case, we model the interaction $Z_{i1}(t) \times Z_{i2}(t)$. We write this model as $\lambda_i(t;\mathbf{Z}_i) = \lambda_0(t)\exp\{\beta_1 Z_{i1}(t) + \beta_2 Z_{i2}(t) + \beta_3 Z_{i1}(t) \times Z_{i2}(t)\}$. We solve for estimator $\hat{\boldsymbol{\beta}} = (\hat{\beta}_1, \hat{\beta}_2, \hat{\beta}_3)$ by solving $\frac{\partial \log \mathcal{L}_v(\boldsymbol{\beta})}{\partial \boldsymbol{\beta}} = 0$.

3.2.2 Asymptotic Properties

For $\zeta = 0, 1$, if we perform stratified sampling within each level of the proposed effect modifier, then for each stratum l in the subcohort $\mathcal{C}$, we denote $S_{\mathcal{C}_l}^{(\zeta)}(\beta, t) = \sum_{k \in \mathcal{C}_l} Y_k(t) Z_k(t)^\zeta \exp\{\beta' Z_k(t)\} \frac{n_l}{m_l}$ and $E_{\mathcal{C}_l}(\beta, t) = S_{\mathcal{C}_l}^{(1)}(\beta, t) / S_{\mathcal{C}_l}^{(0)}(\beta, t)$.

For a general model $\lambda_i(t;\mathbf{Z}_i) = \lambda_0(t)\exp\{\boldsymbol{\beta}'\mathbf{Z}_i(t)\}$, denoting the true parameter as $\boldsymbol{\beta}_0$, we have that the asymptotic distribution of estimator $\hat{\boldsymbol{\beta}}$, as $n \to \infty$, is

$$\sqrt{n}(\hat{\boldsymbol{\beta}} - \boldsymbol{\beta}_0) \to_d N(0, \Sigma^{-1} + \Sigma^{-1}\Delta\Sigma^{-1}). \tag{3.2}$$

The covariance matrix can be estimated consistently by $n(\hat{\Sigma}^{-1} + \hat{\Sigma}^{-1}\hat{\Delta}\hat{\Sigma}^{-1})$, where $\hat{\Sigma}^{-1} = \hat{\mathcal{I}}_v(\hat{\boldsymbol{\beta}})$ is the observed information matrix at $\hat{\boldsymbol{\beta}}$ after incorporating weights, and $\hat{\Sigma}^{-1} + \hat{\Sigma}^{-1}\hat{\Delta}\hat{\Sigma}^{-1}$ is

$$\hat{\mathcal{I}}_v^{-1}(\hat{\boldsymbol{\beta}}) + \sum_{l=1}^{L} m_l \frac{1-\pi_{l,n}}{\pi_{l,n}^2} \hat{\mathcal{I}}_v^{-1}(\hat{\boldsymbol{\beta}}) \hat{\Delta}_l \hat{\mathcal{I}}_v^{-1}(\hat{\boldsymbol{\beta}}) \tag{3.3}$$

where $\pi_{l,n} = m_l/n_l$, $\hat{\Delta}_l = m_l^{-1} \sum_{k \in \mathcal{C}_l} (X_k - \bar{X}_l)(X_k - \bar{X}_l)'$, $X_k = \int_0^\tau \{Z_k(t) - E_{\mathcal{C}_{l(k)}}(\hat{\boldsymbol{\beta}}, t)\} Y_k(t)$ $\exp\{\hat{\boldsymbol{\beta}}' Z_k(t)\} S^{(0)}_{\mathcal{C}_{l(k)}}(\hat{\boldsymbol{\beta}}, t)^{-1} d\bar{N}(t)$ for subcohort members in each stratum l, $\bar{N}(t) = \sum_{k \in \mathcal{C}_{l(k)}} N_k(t)$. These properties hold for both the marginal and interaction models. For details on these results, see the Appendix B.

3.3 Simulation Study

To investigate the finite sample performance and properties of the estimators, we conducted a simulation study. In the marginal model, we investigated the validity of the estimator for assessing marginal effects under oversampling. In the interaction model, we assessed whether the estimator demonstrated increased efficiency and power, compared to a classic case-cohort estimator.

We compared results from the full cohort analysis, the classic case-cohort design, the stratified case-cohort design without incorporating needed weights, and the stratified case-cohort design correcting with weights. True survival times were simulated from an exponential distribution with hazard rate $0.5\exp\{\beta_1 Z_1 + \beta_2 Z_2 + \beta_3 Z_1 \times Z_2\}$, with a continuous predictor of interest $Z_1 \sim N(0.1 \times Z_2, 1)$. Z_2, the proposed effect modifier, was a binary variable taking values 0 or 1. Other forms of Z_2 can be easily accommodated. We designated the $Z_2 = 1$ subgroup to be the rare and/or under-represented subpopulation, and we varied the prevalence of the $Z_2 = 1$ subgroup to investigate the performance of the designs when a rarer subgroup is present.

True parameter values were set to be $\beta_1 = \log(1.3) \approx 0.262$, $\beta_2 = \log(1.25) \approx 0.223$. We

varied the true values of β_3 to investigate the performance of the method under different relative sizes of the interaction effect to the main effect, i.e. β_3/β_1. There were approximately 20% events, obtained using a Uniform(0.1, 0.7) distribution for censoring. 1,000 simulations with 5,000 subjects each were performed in each scenario, where we also compared results from taking subcohort sizes of 10% or 20% of the full cohort. Scenarios performed with the classic case-cohort designs were in the Prentice (1986) sampling and weighting scheme. The classic case-cohort designs drew random samples to be the subcohort, while stratified sampling of the subcohort took Z_2 to be from a discrete uniform distribution. We note that when we took 10% of the cohort to be subcohorts and the prevalence of the $Z_2 = 1$ group was 5%, all members of the rarer subgroup were sampled into the subcohort, and the classic case-cohort analysis sets would be larger in these scenarios due to the limited size of the $Z_2 = 1$ group. Variances were calculated using the robust (sandwich) variance from functions coxph() and cch() in the `survival` package (Therneau and Lumley, 2015). We also reported the relative bias and empirical variance of the estimator compared to the true values. "True" values for the marginal model coefficients were estimated through 100 simulations of 100,000 subjects. Analyses throughout are performed using `survival` (version 3.8-3) in R (version 4.5.5-arm64).

3.3.1 Bias and Variance

Coefficient estimates in the stratified sampling designs, after correcting for the sampling scheme, were consistent in both the marginal and interaction models. Stratified sampling case-cohort estimates were found to have comparable relative bias to the classic case-cohort design, slightly higher than the full cohort estimates. This was found to hold for both the marginal and interaction models in various scenarios. As an example, we show in detail the estimates for the scenario where $\beta_3 = \log(1.2)$ and the prevalence of $Z_2 = 1$ was 5% in

	$\gamma_1 \approx 0.273$			$\gamma_2 \approx 0.291$		
	%Bias	Rob. Var.	Emp. Var.	%Bias	Rob. Var.	Emp. Var.
10% Subcohort						
Full Cohort	0.00	0.0011	0.0011	-2.02	0.0180	0.0178
Classic CC	0.46	0.0037	0.0037	8.59	0.0719	0.0788
Strat. CC: Uncor.	2.29	0.0058	0.0060	-1×10^3	0.0285	0.0248
Strat. CC: Cor.	2.29	0.0060	0.0060	-5.13	0.0199	0.0195
20% Subcohort						
Full Cohort	-0.27	0.0011	0.0011	0.21	0.0180	0.0175
Classic CC	0.49	0.0023	0.0024	2.41	0.0415	0.0434
Strat. CC: Uncor.	1.50	0.0033	0.0035	-7.8×10^3	0.0253	0.0204
Strat. CC: Cor.	1.50	0.0034	0.0035	-1.65	0.0189	0.0184

Table 3.1: Simulation results showing relative bias, robust, and empirical variance for the marginal model with $\beta_3 = \log(1.2)$ and 5% prevalence of $Z_2 = 1$. Classic CC: classic case-cohort design taking a random subcohort. Strat. CC: Uncor.: stratified case-cohort design without correcting for biased sampling. Strat. CC: Cor.: stratified case-cohort design after correcting for biased sampling. %Bias: relative bias. Rob. Var: average robust variance estimate across simulations. Emp. Var.: Empirical variance across simulations.

tables 3.1 and 3.2. Table 3.1 shows results in the estimation of γ_2 in the marginal model, comparing across the full cohort analysis, the classic case-cohort analysis (where we take a random subcohort), the stratified case-cohort analysis without correcting for biased sampling, and the stratified case-cohort analysis after correcting for biased sampling. Table 3.2 shows results in the estimation of β_3 in the interaction model.

Comparing across the designs and estimators, we find that in stratified sampling settings, the robust variance estimator performed well in estimating the variance. The variances of the proposed effect-modifying variable coefficients (β_3 and γ_2) were also lower than those in the classic case-cohort designs. This demonstrated the desired precision gains when performing stratified sampling.

We also note that the classic case-cohort design could produce higher bias and variance, possibly due to the high variability of random sampling. For example, in Table 3.2, there was around 46% relative bias in estimating $\hat{\beta}_3$, coupled by higher variance, when prevalence

	$\beta_1 \approx 0.262$			$\beta_2 \approx 0.223$			$\beta_3 \approx 0.182$		
	%Bias	Var	Emp	%Bias	Var	Emp	%Bias	Var	Emp
10% Subcohort									
Full Cohort	-0.22	0.0012	0.0012	-6.37	0.0226	0.0231	4.07	0.0192	0.0196
Classic CC	-0.04	0.0039	0.0038	6.35	0.0808	0.0941	46.6	0.1043	0.1398
Strat.: Uncor.	2.27	0.0065	0.0067	-1×10^3	0.0323	0.0296	1.19	0.0311	0.0251
Strat.: Cor.	2.27	0.0067	0.0067	-8.73	0.0237	0.0241	1.19	0.0250	0.0251
20% Subcohort									
Full Cohort	-0.57	0.0012	0.0012	-3.37	0.0224	0.0229	4.32	0.0191	0.0194
Classic CC	-0.10	0.0024	0.0026	-2.01	0.0465	0.0499	27.2	0.0521	0.0657
Strat.: Uncor.	1.35	0.0037	0.0039	-1×10^3	0.0293	0.0251	1.65	0.0276	0.0226
Strat.: Cor.	1.35	0.0038	0.0039	-4.55	0.0229	0.0233	1.65	0.0217	0.0226

Table 3.2: Simulation results showing relative bias, robust, and empirical variance for the interaction model with $\beta_3 = \log(1.2)$ and 5% prevalence of $Z_2 = 1$. Classic CC: classic case-cohort design taking a random subcohort. Strat. CC: Uncor.: stratified case-cohort design without correcting for biased sampling. Strat. CC: Cor.: stratified case-cohort design after correcting for biased sampling. %Bias: Relative bias. Var: Average robust variance estimate across simulations. Emp: Empirical variance across simulations.

of the rare subgroup was low and when we took a 10% subcohort. The stratified sampling design proved to be useful in combating this variability by ensuring that more, even all, individuals in the rarer subgroup were sampled into the analysis.

3.3.2 Power and Coverage Probability

The contour plot of the difference of power in assessing the interaction effect (Figure 3.1) showed that as the difference between β_3 and β_1 increased, the power gains from using the stratified design increased as well. The power gains also grew greater as the subgroup $Z_2 = 1$ became rarer in the cohort. Especially when the prevalence of $Z_2 = 1$ was as low as 5%,

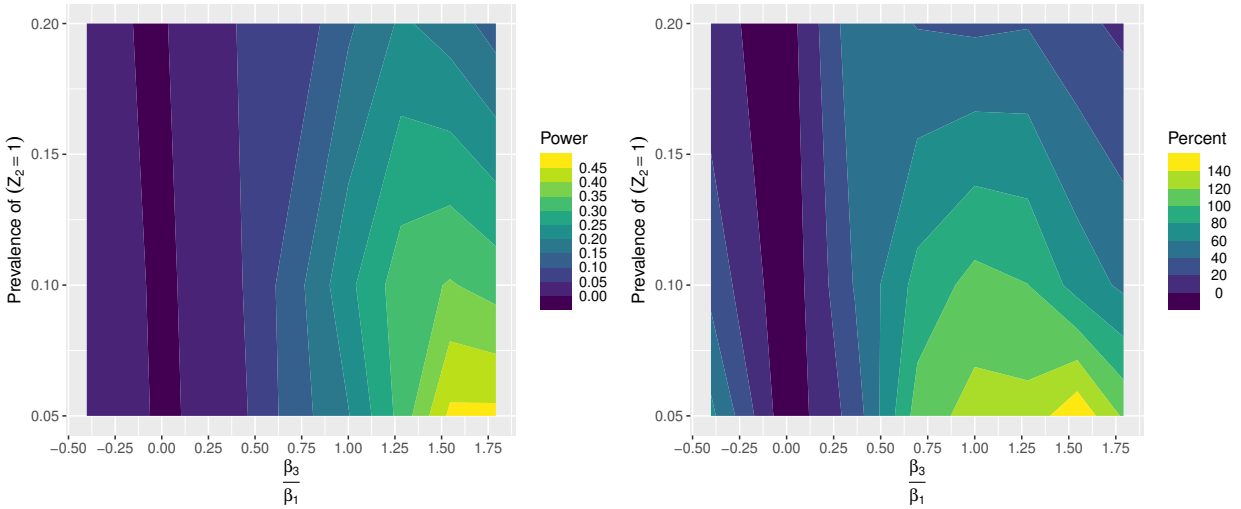


Figure 3.1: Contour plots of the difference between power assessing β_3 for the stratified versus the classic case-cohort designs, in the interaction model. The size of the subcohort is 10% of the full cohort. Left: absolute difference in power. Right: relative difference in power.

we might gain as much as over 140% higher power to assess the effect modification of Z_2 through the stratified design.

Tables 3.3 and 3.4 further illustrate the desired increase in power when we are interested in investigating subgroup effects. The type I error rates at $\beta_3 = 0$ were controlled at around 5% for the stratified design across different scenarios. Compared to the full cohort analysis, we retained the advantages of the case-cohort design by requiring less covariate histories but still maintained power to detect subgroup differences. While power was limited as the size of the analysis set (including the subcohort and the unsampled cases) was small, the stratified design was able to retain much more information on the rare subgroup. In the marginal model, the power to assess the coefficient corresponding to the binary variable Z_2, γ_2, was shown to be higher. We note that this, however, may be due to a larger effect in the $Z_2 = 1$ subgroup and may not always hold under other data-generating models.

Additionally, the case-cohort design with stratified sampling showed good coverage probability for β_3 in the interaction model (Table 3.5). We note that the simulations occasionally

β_3	Prevalence of $Z_2 = 1$	10% Subcohort		20% Subcohort	
		Classic CC	Strat. CC	Classic CC	Strat. CC
$\log(1) = 0$	5%	0.058	0.050	0.057	0.052
	10%	0.062	0.053	0.056	0.052
	20%	0.049	0.044	0.055	0.052
$\log(1.3)$	5%	0.154	0.351	0.243	0.461
	10%	0.222	0.459	0.371	0.661
	20%	0.389	0.530	0.611	0.749
$\log(1.4)$	5%	0.249	0.566	0.320	0.649
	10%	0.326	0.653	0.574	0.866
	20%	0.537	0.745	0.866	0.912
$\log(1.5)$	5%	0.296	0.757	0.392	0.808
	10%	0.492	0.846	0.709	0.948
	20%	0.701	0.878	0.924	0.979
$\log(1.6)$	5%	0.396	0.859	0.577	0.908
	10%	0.590	0.920	0.842	0.986
	20%	0.826	0.953	0.975	0.998

Table 3.3: Simulation results comparing power between different scenarios in assessing β_3 in the interaction model. $\beta_1 = \log(1.3), \beta_2 = \log(1.25)$, and β_3 takes values from $\log(1)$ to $\log(1.6)$. In all scenarios, the sample size is 5,000 and 1,000 iterations were run. (For $\beta_3 = \log(1)$, 5,000 iterations were run to assess type I error.) Strat.: Stratified.

had errors in the coxph() function due to the convergence of some coefficient estimates being faster than others, and we removed results from the one error iteration in these scenarios.

3.4 Application to ADNI Data

We demonstrated the utility of the method in biomarker discovery by application to Alzheimer's Disease Neuroimaging Initiative (ADNI) data obtained from the ADNI database,

β_3	Prevalence of $Z_2 = 1$	10% Subcohort		20% Subcohort	
		Classic CC	Strat. CC	Classic CC	Strat. CC
	5%	0.133	0.331	0.199	0.361
log(1) = 0	10%	0.199	0.479	0.334	0.549
	20%	0.334	0.652	0.515	0.748
	5%	0.233	0.587	0.356	0.643
log(1.3)	10%	0.389	0.797	0.609	0.875
	20%	0.612	0.922	0.833	0.977
	5%	0.268	0.724	0.437	0.737
log(1.4)	10%	0.463	0.877	0.715	0.933
	20%	0.704	0.966	0.906	0.996
	5%	0.349	0.803	0.532	0.801
log(1.5)	10%	0.577	0.936	0.781	0.959
	20%	0.772	0.987	0.962	0.998
	5%	0.368	0.854	0.563	0.870
log(1.6)	10%	0.628	0.979	0.878	0.987
	20%	0.852	0.992	0.978	0.998

Table 3.4: Simulation results comparing power between different scenarios in assessing γ_2 in the marginal model. $\beta_1 = \log(1.3), \beta_2 = \log(1.25)$, and β_3 takes values from $\log(1)$ to $\log(1.6)$. In all scenarios, the sample size is 5,000 and 1,000 iterations were run. (For $\beta_3 = \log(1)$, 5,000 iterations were run.)

to assess subgroup differences in biomarker effects. The ADNI was launched in 2003 to instruct clinical trials to discover treatments to slow or stop the progression of AD. The dataset used included participants enrolled in ADNI1, ADNIGO, ADNI2, and ADNI3 phases. To note, if researchers employed a case-cohort design in a new study, they would not process expensive covariates for the full cohort and they would not have full cohort biomarker data.

In our analysis, we defined the event as disease progression of AD, which was progression from cognitively normal (CN) to mild cognitive impairment (MCI), or from MCI to dementia. We did not consider recurrent events in this analysis (i.e., two progressions for the same individual) or the reversion from dementia to MCI or MCI to CN. Diagnoses were as provided in the ADNI dataset, and to further collapse the diagnostic categories, we categorized SMC (subjective memory complaint) as CN, and EMCI and LMCI (early and late MCI, respectively) as MCI.

β_3	Prevalence of $Z_2 = 1$	10% Subcohort	20% Subcohort
	5%	0.950	0.948
$\log(1) = 0$	10%	0.947	0.948
	20%	0.956	0.948
	5%	0.955	0.946
$\log(1.3)$	10%	0.931	0.942
	20%	0.947	0.958
	5%	0.934	0.952
$\log(1.4)$	10%	0.935	0.960
	20%	0.955	0.951
	5%	0.951	0.942
$\log(1.5)$	10%	0.952	0.939
	20%	0.948	0.952
	5%	0.953	0.942
$\log(1.6)$	10%	0.957	0.939
	20%	0.951	0.952

Table 3.5: Simulation results of the coverage probability when estimating β_3 for the stratified case-cohort design in the interaction model.

The predictor of interest was chosen to be CSF phosphorylated tau (p-Tau181) levels. Pathological modification of tau leads to hyperphosphorylated tau, causing axonal dysfunction, synaptic loss and degeneration of neurons (Rawat et al., 2022). The level of pTau is a key biomarker in AD research, with changes that can be observed prior to a dementia diagnosis. pTau levels of < 8 pg/mL were truncated to be 8, > 120 pg/mL to be 120, and those without any measurements of pTau were removed from the dataset. For the analysis, we standardized the measurements by its standard deviation (SD). We considered, for each individual, the time from their first disease-free measurement of pTau to the time of progression. Censoring time was set as the date of the last record for the individual. There were, in total, 999 individuals with complete data for the variables used in the analysis. We used subcohort sizes that were 20% of the full cohort.

One proposed effect modifier was race and ethnicity, where categories were first combined to be mutually exclusive. Then, due to a lack of individuals who were not NH White in the dataset, we collapsed the categories to form a binary variable with levels "NH White" and

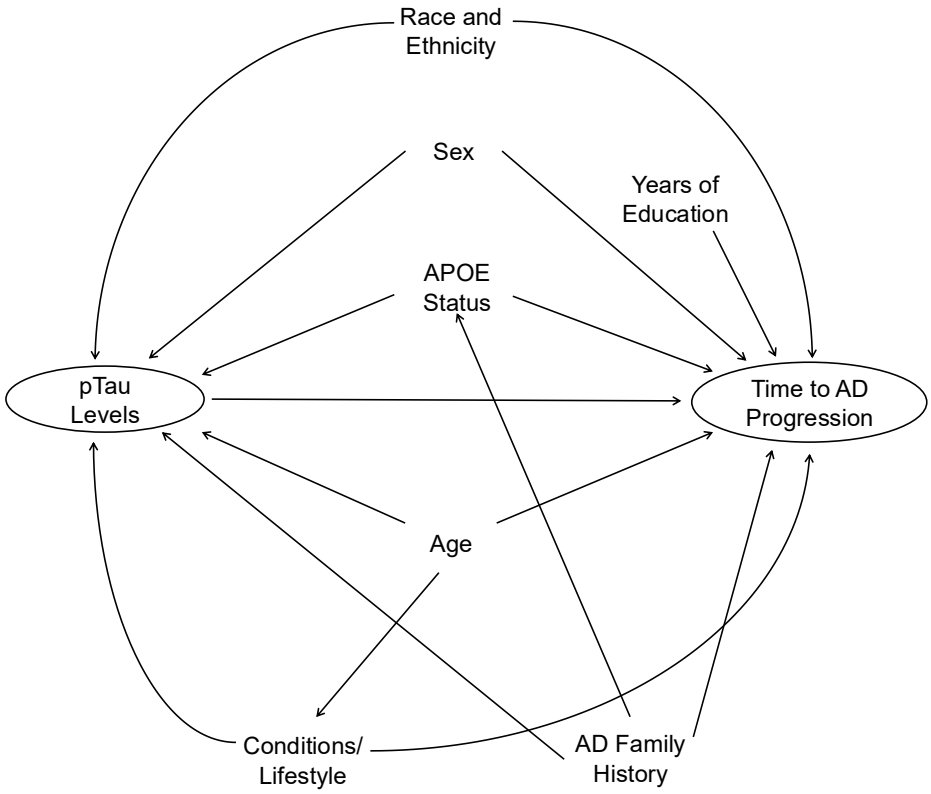


Figure 3.2: Directed acyclic graph of response variable (time to AD progression), predictor of interest (pTau levels), and potential confounders and associated variables.

"all other racial and ethnic groups".

We considered adjustment variables thought to be related to the response and the predictor of interest. In a directed acyclic graph (Figure 3.2), we show potential confounders and variables that may increase precision. Demographic variables including sex and age are adjusted for due to potential confounding and for being generally informative for understanding biomarker effects. Years of education can potentially be related to cognitive reserve, which is a protective factor associated with risk of AD (Silva et al., 2019), while AD and APOE status are strongly linked to AD family history. Additionally, we considered a comorbidities

Covariate	Classic Case-Cohort (N = 473)	Stratified Case-Cohort (N = 471)
AD Progression		
No Progression	131 (27.7%)	129 (27.4%)
Progression	342 (72.3%)	342 (72.6%)
pTau (pg/mL)	29.4 (14.9)	28.5 (14.6)
APOE4 (copies)		
0	248 (52.4%)	240 (51.0%)
1	172 (36.4%)	183 (38.9%)
2	53 (11.2%)	48 (10.2%)
Race and Ethnicity		
Hispanic/Latino	6 (1.3%)	30 (6.4%)
Multiracial	5 (1.2%)	10 (2.1%)
NH Am Indian/Alaskan	1 (0.2%)	2 (0.4%)
NH Asian	5 (1.2%)	14 (3.0%)
NH Black	18 (3.8%)	36 (7.6%)
NH Hawaiian/Other PI	1 (0.2%)	2 (0.4%)
NH White	437 (92.4%)	377 (80.0%)
Age (yrs)	73.4 (7.1)	73.4 (7.1)
Sex		
Male	266 (56.2%)	265 (56.3%)
Female	207 (43.8%)	206 (43.7%)
Education (yrs)	16.1 (2.7)	16.1 (2.7)
No. of Comorbidities		
Low (0-6)	246 (52.0%)	261 (55.4%)
High (7-15)	227 (48.0%)	210 (44.6%)
Family History of Dementia/AD		
No	207 (43.8%)	213 (45.2%)
Yes	266 (56.2%)	258 (54.8%)

Table 3.6: Summary statistics of the analysis sets (including the subcohort and supplemented cases) for the classic case-cohort and stratified case-cohort designs. For categorical variables, we reported N(%) and for numerical variables, we reported mean (SD).

variable that sums the number of conditions or lifestyle factors for individuals, which include: 1) psychiatric, 2) neurologic (other than cognitive disorder), 3) head, eyes, ears, nose, throat, 4) cardiovascular, 5) respiratory, 6) hepatic, 7) dermatologic-connective tissue, 8) musculoskeletal, 9) endocrine-metabolic, 10) gastrointestinal, 11) hematopoietic-lymphatic, 12) renal-genitourinary, 13) allergies or drug sensitivities, 14) smoking, alcohol use, and/or drug use, 15) malignancy. These conditions can be linked to possible increased risk of AD

(Silva et al., 2019; Zhang et al., 2011). We binarized the variable into low (0-6 conditions) and high (7-15 conditions). As a limitation, certain potential confounding factors are not measured. Some conditions and lifestyle factors including level of social interaction, obesity, and diet are not included due to limited data. Additionally, we highlighted the potentially strongest associations in the figure but we acknowledge that other associations may exist between the variables. We do not consider the interaction of the potential effect modifiers with adjustment covariates in this analysis.

For illustration, we compared the analysis set when using the classic versus the stratified case-cohort designs in one iteration of sampling, and present a table of summary statistics (Table 3.6). In general, the age of the individuals in the study were older, with higher years of education on average. In comparison, by using the stratified design, we may maximally retain the information we have on non-NH White individuals by sampling all of the individuals into the subcohort. On the other hand, the classic case-cohort design only included 7.6% individuals who were not NH White in the analysis set.

In this analysis (Table 3.7), in the stratified sampling design we estimated that in NH White individuals, one standardized unit increase in pTau was associated with a 54% higher hazard for progression of dementia (HR: 1.54; 95% CI: (1.31, 1.80), p-value: < 0.001), after adjusting for other variables in the model. For individuals in all other racial and ethnic groups, one standardized unit increase in pTau was associated with a 6% higher hazard for progression of AD (HR: 1.06; 95% CI: (0.57, 1.95), p-value: 0.86). The interaction of pTau with race and ethnicity was not found to be statistically significant (full cohort, classic case-cohort, stratified case-cohort analyses p-values were 0.574, 0.905, 0.443, respectively). The 95% CI was 63% narrower than in the classic design (95% CI: (0.40, 4.13)), indicating a much higher precision when performing stratified sampling.

There will exist some sampling variability when taking one random or stratified sample leading to some differences in point estimates, hence we also provided estimated quantities

from averaging across 100 iterations of sampling in Table B.1 in Appendix B. The point estimates of the case-cohort designs stabilized and became closer to the full cohort analysis when averaging across iterations. Compared to the full cohort analysis, while we lose some efficiency, we only require processing of high burden and expensive covariates on less than half of the cohort. Additionally, the stratified design showed comparable precision to the full cohort when assessing the effect of pTau in the rare subgroup.

With the proposed effect modifier of APOE status, where individuals may have 0, 1, or 2 copies of the $\epsilon 4$ allele (Table 3.8), we performed stratified sampling with regard to APOE status. We were also able to retain more information on the rare group of individuals with 2 copies of the $\epsilon 4$ allele. In the stratified model, we found that one standardized unit increase in pTau was associated with a 3.2-fold higher hazard for progression of AD in those with 2 copies of the $\epsilon 4$ allele (HR: 3.21; 95% CI: (2.07, 4.98), p-value: < 0.001), a 2.2-fold higher hazard in those with 1 copy (HR: 2.23; 95% CI: (1.52, 3.27), p-value: < 0.001), and a 2.3-fold higher hazard in those with 0 copies (HR: 2.34; 95% CI: (1.55, 3.53), p-value: < 0.001), after adjusting for other variables in the model. The interactions of pTau with APOE status of 2 copies (the rare subgroup) in the full cohort, classic case-cohort, and stratified case-cohort analyses returned p-values of < 0.001, 0.555, and 0.002, respectively. The stratified design demonstrated a statistically significant interaction effect that would otherwise be overlooked in the classic design. Furthermore, there was a 41% reduction in the width of the CI for assessing the effect of pTau in individuals with 2 copies when using the stratified design. This again indicated the potential to discover more subgroup differences in biomarker effects with sampling and weighting methods alone. Similarly, we provided estimated quantities from averaging across 100 iterations of sampling in Table B.2 in Appendix B.

3.5 Discussion

The first step of biomarker discovery typically focuses on valid inference of the biomarker effect across the full study population, in what we termed the marginal model. Additionally, it is often crucial to further investigate any possible subgroup differences to better ensure valid inference among various subpopulations. Using the case-cohort design can largely reduce the burden and cost of processing samples for studies with lower proportions of events. While the case-cohort design is effective in reducing the required time, cost, and valuable samples, it may face the issue of decreased power when analyzing subgroup effects, especially in later event times. We have demonstrated through simulations that, when interested in subgroup-specific results in case-cohort analyses, we may perform stratified sampling to achieve greater power. Although we may lose some power in the marginal model, we reiterate the importance of gaining much greater power in subgroup analyses, a trade-off that can lead to important results and discoveries. Our analysis of ADNI data showed a great increase in efficiency when assessing effect modification of race and ethnicity as well as APOE status with pTau, which is one example of discovering subgroup heterogeneity when more information is retained in the subcohort. We also see more stable estimates of the interaction effect compared to taking random subcohorts.

While we have examined the effect modification of APOE status and biomarker effects in AD, one could also consider the effect modification of APOE status in AD trials for adverse events such as amyloid-related imaging abnormalities (ARIA). For example, in trials assessing lecanemab, an antibody that targets amyloid beta aggregates for the treatment of AD, individuals with one or two copies of APOE $\epsilon 4$ were found to be at higher risk for ARIA-E and -H, or ARIA due to edema and hemorrhage, respectively (Kane, 2024). In assessments of treatments, it is important to investigate subgroup differences especially as it could point to higher risks for certain subgroups.

The estimator discussed is approximately score-unbiased as shown in Borgan et al. (2000).

For a score-unbiased stratified estimator, Borgan et al. (2000) proposed "Estimator III" where, if the case occurred outside the subcohort, then a subcohort member from the same stratum swaps place with the case. Although this estimator performs slightly better, the results are very similar unless the subcohort is very small (Cologne et al., 2012). Generally, researchers will consider subcohort sizes that still ensure a pre-specified level of power (Hu et al., 2014; Cai and Zeng, 2004). Hence, for ease of implementation, especially for a wider audience of researchers, we still propose utilizing the discussed "Estimator I". Furthermore, we only considered time-fixed weights, i.e., weights calculated using the number of individuals in the full cohort and subcohort in each stratum at the beginning of the study. Kulich and Lin (2004) proved that incorporating whole cohort information using time-dependent weights will further increase efficiency of the estimation. We plan to investigate power gains in implementing time-dependent weights for interaction effects as a future direction. Moreover, the progression of AD can be regarded as recurrent events of different types. Although there is existing work on handling recurrent events in case-cohort designs (Maitra et al., 2020), we plan to further investigate methods to increase efficiency through sampling considerations.

While the use of the proportional hazards model is ubiquitous in clinical research, we acknowledge that the assumption is strong and cannot hold under both the marginal and interaction models. Despite this, the hazard ratio continues to yield a meaningful summary measure of association in many cases, representing a weighted average of the hazard ratio over time, and multiple authors have written about the interpretation of a Cox proportional hazards estimator in non-proportional hazards settings (Struthers and Kalbfleisch, 1986; Xu, 2000; Boyd et al., 2012). Further, we find that in most scientific settings where there is not strong mechanistic rationale to *a priori* hypothesize effect modification, it is useful to first pool as much information in establishing an overall association then to further investigate differences in effects across subpopulations. This is the approach that we have advocated for here.

In the application to ADNI data, the low proportions of individuals from the non-NH White group was not ideal, but it may be reflective of many datasets in AD research. For example, across six Alzheimer's Disease Cooperative Study trials, only 5% and 6% of participants were Hispanic and African American, respectively (Grill et al., 2013). We believe it remains important to investigate such subgroup differences and to maximize all possible information, to better understand how well biomarkers perform in diagnosis or interventions for diverse subpopulations.

In datasets that have severely under-represented subpopulations, the analysis may still suffer from low power even when we perform stratified sampling and retain all subjects in rare subgroups. In such scenarios, to answer questions relating to subgroup effects, researchers would need to address the issue with recruitment strategies and study designs that can alleviate under-representation from a data collection standpoint. Such efforts may include targeted interventions to increase attitudes toward research, especially for under-represented subpopulations (Witbracht et al., 2025).

3.6 Acknowledgments

Data used in preparation of Chapter 3 were obtained from the Alzheimer's Disease Neuroimaging Initiative (ADNI) database (adni.loni.usc.edu). As such, the investigators within the ADNI contributed to the design and implementation of ADNI and/or provided data but did not participate in analysis or writing of this report. A complete listing of ADNI investigators can be found at: https://adni.loni.usc.edu.

	Full Cohort				Classic Case-Cohort				Stratified Case-Cohort			
	N	HR	95% CI	p-value	N	HR	95% CI	p-value	N	HR	95% CI	p-value
pTau												
NH White	905	1.50	(1.36, 1.65)	< 0.001	437	1.24	(0.99, 1.56)	0.06	377	1.54	(1.31, 1.80)	< 0.001
All other groups	94	1.25	(0.71, 2.21)	0.44	36	1.29	(0.40, 4.13)	0.67	94	1.06	(0.57, 1.95)	0.86
Age	999	1.04	(1.02, 1.05)	< 0.001	473	1.04	(1.01, 1.07)	0.01	471	1.06	(1.04, 1.08)	< 0.001
Female vs. Male	458	0.74	(0.59, 0.93)	0.01	207	0.56	(0.36, 0.86)	0.01	206	0.66	(0.44, 0.98)	0.04
Education	999	0.97	(0.93, 1.01)	0.16	473	0.96	(0.89, 1.04)	0.28	471	1.05	(1.00, 1.09)	0.04
APOE4												
0 copies	588	Ref.	-	-	248	Ref.	-	-	240	Ref.	-	-
1 copy	336	1.82	(1.43, 2.32)	< 0.001	172	4.19	(2.63, 6.68)	< 0.001	183	2.06	(1.36, 3.12)	< 0.001
2 copies	75	2.45	(1.68, 3.57)	< 0.001	53	2.38	(1.02, 5.52)	0.04	48	1.52	(0.78, 2.98)	0.22
No. Comorbidities: High	472	0.75	(0.61, 0.94)	0.01	227	0.77	(0.51, 1.18)	0.23	210	0.65	(0.45, 0.95)	0.03
Family History: Yes	549	1.02	(0.81, 1.27)	0.89	266	0.73	(0.47, 1.11)	0.14	258	0.82	(0.64, 1.05)	0.12

Table 3.7: Estimates of hazard ratios (95% CI) and p-values of the model considering race and ethnicity as a potential effect modifier, with adjustment covariates. "All other groups" in variable race and ethnicity includes Hispanic/Latino and non-Hispanic American Indian, Alaskan, Asian, Black, Hawaiian or other PI, and multi-racial individuals. "Ref." stands for referent level.

	Full Cohort				Classic Case-Cohort				Stratified Case-Cohort			
	N	HR	95% CI	p-value	N	HR	95% CI	p-value	N	HR	95% CI	p-value
pTau												
0 copies	588	1.77	(1.50, 2.08)	<0.001	222	1.81	(1.27, 2.58)	<0.001	206	2.34	(1.55, 3.53)	<0.001
1 copy	336	2.78	(2.19, 3.53)	<0.001	182	2.67	(1.68, 4.25)	<0.001	184	2.23	(1.52, 3.27)	<0.001
2 copies	75	3.75	(2.64, 5.31)	<0.001	53	5.12	(2.15, 12.19)	<0.001	72	3.21	(2.07, 4.98)	<0.001
Age	999	1.03	(1.02, 1.05)	<0.001	457	1.01	(0.97, 1.04)	0.70	462	1.03	(0.99, 1.06)	0.13
Female vs. Male	458	0.78	(0.62, 0.98)	0.03	193	0.79	(0.51, 1.24)	0.30	199	0.82	(0.52, 1.29)	0.39
Race and Eth.												
NH White	905	Ref.	-	-	422	Ref.	-	-	426	Ref.	-	-
All other groups	94	0.94	(0.61, 1.43)	0.76	35	0.64	(0.31, 1.29)	0.21	36	0.58	(0.26, 1.28)	0.18
Education	999	0.96	(0.93, 1.00)	0.07	457	1.03	(0.95, 1.11)	0.46	462	0.99	(0.92, 1.07)	0.84
No. Comorbidities: High	472	0.78	(0.63, 0.97)	0.02	222	0.75	(0.49, 1.14)	0.18	215	0.63	(0.41, 0.97)	0.04
Family History: Yes	549	0.99	(0.79, 1.24)	0.92	264	0.77	(0.50, 1.18)	0.23	262	0.70	(0.45, 1.09)	0.12

Table 3.8: Estimates of hazard ratios (95% CI) and p-values of the model considering APOE status as a potential effect modifier, with adjustment covariates. "All other groups" in variable race and ethnicity includes Hispanic/Latino and non-Hispanic American Indian, Alaskan, Asian, Black, Hawaiian or other PI, and multi-racial individuals. "Ref." stands for referent level.

Chapter 4

Case-cohort designs to improve efficiency for recurrent events analysis

4.1 Introduction

In the previous chapter, we modified the case-cohort design under the univariate setting, where individuals experience at most one event, to be more efficient for comparing subgroup differences in secondary analyses. We considered a multiplicative hazards model where the hazard function is modeled as a baseline hazard function multiplied by a function of covariates such as an exponentiated linear predictor. Under the univariate survival setting, this model is known as the Cox proportional hazards model (Cox, 1972). Prentice (1986) introduced the case-cohort design for the univariate survival setting.

Using a similar sampling strategy, it is possible to apply the case-cohort design in studies where each individual may experience more than one event, or where there are clustered units or individuals. Analysis of the former, known as recurrent events analysis, is particularly relevant in the context of Alzheimer's disease (AD) research. In AD pathology, the disease

progresses through clinical stages including cognitively normal (where some patients may be characterized as having preclinical AD, due to elevated levels of biomarkers such as amyloid beta), to mild cognitive impairment (MCI), then to AD dementia (Tahami Monfared et al., 2022). As patients progress through these stages, they may experience up to two disease progression events. Hence, the usual univariate analysis may not be sufficient.

Most models developed for multivariate analysis fall into two general categories: frailty models, where covariate effects apply at the individual level; and marginal models, where covariate effects apply at the population level. In this paper, we focus on marginal models. An extension of Andersen and Gill (1982) to the multivariate setting assumes that the correlation between event times is captured by appropriate specification of time-dependent covariates, with a common baseline hazard function across events and global covariate effects. The marginal model discussed in Prentice et al. (1981) accommodated varying baseline hazard functions and varying covariate effects. Wei et al. (1989) developed a model without a specific structure of dependence among the failure times of each subject. The model considered in Spiekerman and Lin (1998) included covariates $\boldsymbol{Z}_{kl}$ so that the marginal hazard can depend on the lth component of the kth type of failure.

Considering marginal models under the case-cohort design, Maitra et al. (2020) proposed a model for analyzing recurrent event data. The authors utilized a weighted estimating equation stemming from a pseudo-likelihood function and proved asymptotic properties for this estimator. Within the estimating equation, an individual's contribution is identical across event occurrences because covariate values are not updated at each event. However, when the analysis involves expensive covariates such as biomarker measurements that are updated at each event, event-specific covariate histories must be incorporated. Kang and Cai (2009) developed a marginal hazards model for case-cohort studies with multiple disease outcomes, a setting in which observations on different disease types within the same individual are correlated. Both papers utilized case-cohort weights similar to those proposed in (Kalbfleisch

and Lawless, 1988; Borgan et al., 2000; Kulich and Lin, 2004), which were developed in the univariate and stratified settings. As an extension to both papers, we propose a model under the recurrent setting that can account for event-specific covariate values.

Furthermore, there are no works that discuss subcohort sampling for case-cohort designs in the recurrent setting when each individual has multiple records stratified by event occurrence. If we treat each record as independent and account for correlations through variance estimation as in Wei et al. (1989), subcohort sampling is not individual-level but instead record-level. An issue that may arise is that risk sets are smaller in later event occurrences. If we take a simple random sample of the entire pool of available records regardless of event occurrence, records from later event occurrences will be underrepresented in the subcohort, limiting statistical power to perform inference. Kulich and Lin (2004) noted that, for the proposed class of estimators, asymptotic theory requires the number of subcohort controls to be large in every stratum (i.e., every event occurrence in our setting) throughout the entire study period, and violations of this assumption in practice may adversely affect performance.

There are no current works that consider sampling strategies in recurrent event case-cohort designs with updated covariate measurements at each event time, and that discuss the impact of sampling on efficiency and power. This paper develops a sampling and estimation framework to address this gap and demonstrates efficiency gains, particularly for later event occurrences, under the appropriate sampling approach.

In this chapter we make three primary contributions. First, we extend the recurrent events case-cohort framework to accommodate event-specific covariate histories that are updated at each event occurrence. Second, we identify that the choice of subcohort sampling strategy – pooled versus event-specific – has a substantive impact on the efficiency of inference for later event occurrences, a consideration not addressed in prior work. Third, we propose two event-specific sampling schemes, balanced and events-proportional sampling, and characterize the operating properties of each through simulation and application.

In the following sections, we first describe an estimator for the event-agnostic model that may incorporate two types of weights in the estimating equation. We derive asymptotic properties of the estimator. We then detail three sampling schemes and the impact of each on estimation and precision through a simulation study, with recommendations for researchers in practice. Finally, we provide an illustrative example of the proposed framework using the A4 Study and the LEARN Study datasets, which enrolled participants with brain amyloid pathology and without, respectively, to investigate the association of cholesterol levels and cognitive decline (ACTC; Sperling et al., 2020).

4.2 Methods: Event-Agnostic Model

4.2.1 Proposed Model

In the following sections, we call models that perform analysis on all records of individuals regardless of event occurrence "event-agnostic models". To detail the proposed event-agnostic model, we follow counting process notation (Andersen and Gill, 1982). Suppose there are n individuals in the full study cohort and within each event occurrence j, there are n_j observations or records, $j = 1, 2, \ldots, J$. For each individual i, where $i = 1, 2, \ldots, n$, let T_{ij} denote the jth event time. Denote the censoring time for each individual as C_i, and failure time as $\boldsymbol{T}_i = (T_{i1}, \ldots, T_{iJ})'$. $X_{ij} = (T_{ij} \wedge C_i)$ is the observed time and $\Delta_{ij} = I(T_{ij} \leq C_i)$ is the event indicator for individual i at occurrence j. Let $Y_{ij}(t) = I(X_{ij} \geq t)$ be the at risk process, $N_{ij}(t) = I(X_{ij} \leq t, \Delta_{ij} = 1)$ be the counting process of the jth event for individual i on $[0, t)$. $\mathbf{Z}_{ij}(\cdot)$ is a p-dimensional covariate of interest for individual i, which may be event-specific for event occurrence j. This is common in studies where covariate measurements are updated regularly at each visit or when the individual experiences an event of interest. We assume that C_i is independent of $\boldsymbol{T}_i$, conditional on the covariates. We assume that time-dependent covariates are external, as defined in Kalbfleisch and Prentice (2002). Specifically, external time-dependent covariates are those whose future path are not affected by whether the event

occurs (as opposed to internal covariates, such as a physiological measurement that change at the moment of an event).

For this event-agnostic model, we consider a multiplicative hazards model with hazard function $\lambda(t)$ and different (event-specific) baseline hazard functions, $\lambda_{0j}(t)$, for each event occurrence j:

$$\lambda_i\{t|\mathbf{Z}_{ij}(t)\} = \lambda_{0j}(t)\exp\{\beta_0'\mathbf{Z}_{ij}(t)\}, \tag{4.1}$$

in which we consider common covariate effect vector β_0 across event occurrences. Although this model can be extended to accommodate event-specific covariate effects, $\lambda_i\{t|\mathbf{Z}_{ij}(t)\} = \lambda_{0j}(t)\exp\{\beta_{0j}'\mathbf{Z}_{ij}(t)\}$, when no information is shared across event occurrences, fitting a collection of separate univariate models is typically more appropriate. We term these fully stratified models, which we discuss in more detail in the next section.

To perform case-cohort sampling on recurrent event data, we may obtain a random sample from the full cohort through the following strategies, which we term:

1. Pooled sampling. With a pre-specified subcohort sampling fraction q, we take a random sample of size $\tilde{n}$ from all records of all individuals in the full cohort to be the subcohort.

2. Event-specific sampling. Within records of each event occurrence j, we obtain random samples of size $\tilde{n}_j$ with subcohort sampling fraction q_j. We describe two types of proposed event-specific sampling designs in a later section.

Let ζ_{ij} be the subcohort sampling indicator for the record of individual i in event j. We also define the quantity

$$q^* = \begin{cases} q, & \text{if performing pooled sampling,} \\ \frac{\sum_{j=1}^{J} n_j q_j}{\sum_{j=1}^{J} n_j}, & \text{if performing event-specific sampling,} \end{cases} \tag{4.2}$$

which we will call the average subcohort sampling fraction. We assume that covariate histories $\mathbf{Z}_{ij}(t), t \in [0, \tau]$, are only available for records sampled into the subcohort and for records in which the individual experienced an event.

With full cohort data, the estimating equation for obtaining an estimator for β_0 through a partial likelihood equation would be

$$U(\beta) = \sum_{j=1}^{J} \sum_{i=1}^{n} \int_0^\tau Y_{ij}(t) \Big\{ \boldsymbol{Z}_{ij}(t) - \frac{S_j^{(1)}(\beta, t)}{S_j^{(0)}(\beta, t)} \Big\} dN_{ij}(t) \equiv 0, \tag{4.3}$$

where $S_j^{(d)}(\beta, t) = n_j^{-1} \sum_{i=1}^{n} Y_{ij}(t) \boldsymbol{Z}_{ij}(t)^{\otimes d} e^{\beta' \boldsymbol{Z}_{ij}(t)}, d = 0, 1, 2$, where for vector a, $a^{\otimes 2} = aa', a^{\otimes 1} = a, a^{\otimes 0} = 1$.

Replacing the full cohort score function by a weighted version that upweights the contribution of sampled subcohort members to account for their selection probability, the weighted estimating equation under the case-cohort design is

$$\tilde{U}(\beta) = \sum_{j=1}^{J} \sum_{i=1}^{n} \int_0^\tau Y_{ij}(t) \Big\{ \boldsymbol{Z}_{ij}(t) - \frac{\tilde{S}_j^{(1)}(\beta, t)}{\tilde{S}_j^{(0)}(\beta, t)} \Big\} dN_{ij}(t) \equiv 0, \tag{4.4}$$

where $\tilde{S}_j^{(d)}(\beta, t) = n_j^{-1} \sum_{i=1}^{n} w_{ij}(t) Y_{ij}(t) \boldsymbol{Z}_{ij}(t)^{\otimes d} e^{\beta' \boldsymbol{Z}_{ij}(t)}, d = 0, 1, 2.$ $w_{ij}(t)$ is a weight to

account for the case-cohort sampling scheme. It can take the form of inverse-probability-of-sampling weights

$$w_{ij}^{KC}(t) = \Delta_{ij} + (1 - \Delta_{ij})\frac{\zeta_{ij}}{\tilde{\alpha}_j(t)}, \tag{4.5}$$

where $\tilde{\alpha}_j(t) = \frac{\sum_{i=1}^{n}(1-\Delta_{ij})\zeta_{ij}Y_{ij}(t)}{\sum_{i=1}^{n}(1-\Delta_{ij})Y_{ij}(t)}$. A similar weight was proposed by Kang and Cai (2009) for modeling multiple disease outcomes, building on the Estimator II weights of Kulich and Lin (2004) and the framework of Borgan et al. (2000), developed in the univariate setting with covariate strata. Here, $\tilde{\alpha}_j(t)$ is an estimator of the true subcohort non-case sampling probability α_j, which is common across all event occurrences j in pooled sampling and may be specific to each j in event-specific sampling. In Kang and Cai (2009), the weight is $\tilde{\alpha}(t)$, common across all event occurrences. The proposed weight may be time-varying, calculating the number of at-risk and non-case records in the subcohort divided by the number of at-risk and non-case records in the full cohort at time t. It may also be time-invariant, calculating the total number of subcohort non-case records divided by total number of cohort non-case records at the beginning of the study.

Analogous to the Prentice (1986) weighting scheme in the univariate setting, we also propose using weights of the form

$$w_{ij}^{P}(t) = \begin{cases} 1, & t = T_{ij} \\ \zeta_{ij}, & t < T_{ij} \end{cases} \tag{4.6}$$

where a records sampled into the subcohort have weight 1 while at risk ($t < T_{ij}$), and any

case has weight 1 at its own failure time. Using terminology in Kulich and Lin (2004), $w_{ij}^{KC}(t)$ are weights for a class of D-estimators while estimators with $w_{ij}^{P}(t)$ belong to a class of N-estimators. We may solve the estimating equation to obtain an estimator of β_0, $\tilde{\beta}$.

An estimator for the baseline cumulative hazard function for event j, following the form of the Breslow-Aalen estimator, is

$$\tilde{\Lambda}_{0j}(\beta, t) = \int_0^t \frac{\sum_{i=1}^n w_{ij}(u) dN_{ij}(u)}{n_j \tilde{S}_j^{(0)}(\beta, u)}. \tag{4.7}$$

Properties of the Estimators

To detail the asymptotic properties of the estimators, we follow conventional notations

$$s_j^{(d)}(\beta, t) = E\{\tilde{S}_j^{(d)}(\beta, t)\}, d = 0, 1, 2, \quad e_j(\beta, t) = \frac{s_j^{(1)}(\beta, t)}{s_j^{(0)}(\beta, t)},$$
$$v_j(\beta, t) = \frac{s_j^{(2)}(\beta, t) s_j^{(0)}(\beta, t) - s_j^{(1)}(\beta, t)^{\otimes 2}}{s_j^{(0)}(\beta, t)^2}, \quad A_j(\beta) = \int_0^\tau v_j(\beta, t) s_j^{(0)}(\beta, t) \lambda_{0j}(t) dt,$$
$$M_{ij}(t) = N_{ij}(t) - \int_0^t Y_{ij}(u) e^{\beta_0' \mathbf{Z}_{ij}(u)} \lambda_{0j}(u) du,$$
$$M_{\tilde{\mathbf{Z}},ij}(\beta) = \int_0^\tau \{\mathbf{Z}_{ij}(t) - e_j(\beta, t)\} dM_{ij}(t).$$

Proposition 4.1. *Under the above regularity conditions stated in Appendix C, $\tilde{\beta}$, which is the solution to estimating equation $\tilde{U}(\beta) = 0$, is a consistent estimator for β_0.*

When we treat each record as an independent observation in its contribution to the estimating equation, the regularity conditions hold and consistency and asymptotic normality proofs follow a similar outline as in Kang and Cai (2009) and Maitra et al. (2020). While the model

takes a form similar to that in Kang and Cai (2009), some differences arise in how we must treat α_j, which may be different across event occurrences j when performing event-specific sampling. We summarize the proof of Proposition 4.1 and point out its differences from previous work in Appendix C.

Proposition 4.2. *Under regularity conditions,* $n^{1/2}\{\tilde{\beta} - \beta_0\}$ *converges to a normal distribution with mean zero and variance* Σ*. The form of* Σ *is*

$$A(\beta_0)^{-1}Q(\beta_0)A(\beta_0)^{-1}, \tag{4.8}$$

where

$$A(\beta) = \sum_{j=1}^{J} A_j(\beta), \quad Q(\beta) = E\{\sum_{j=1}^{J} M_{\tilde{\mathbf{Z}},ij}(\beta)\}^{\otimes 2}. \tag{4.9}$$

To account for correlations between the observations, we propose a robust variance estimator that incorporates record-level weights for estimation of $\Sigma(\beta_0)$, where each component can be estimated by

$$\tilde{\Sigma}(\tilde{\beta}) = \tilde{A}(\tilde{\beta})^{-1}\tilde{Q}(\tilde{\beta})\tilde{A}(\tilde{\beta})^{-1}, \tag{4.10}$$

$$\tilde{Q}(\tilde{\beta}) = n^{-1}\sum_{i=1}^{n}\Big\{\sum_{j=1}^{J} w_{ij}(t)\tilde{M}_{\tilde{\mathbf{Z}},ij}(\tilde{\beta})\Big\}^{\otimes 2}, \quad \tilde{A}(\tilde{\beta}) = \sum_{j=1}^{J}\tilde{A}_j(\tilde{\beta}), \tag{4.11}$$

$$\tilde{M}_{\tilde{\mathbf{Z}},ij}(\tilde{\beta}) = \Delta_{ij}\left\{\mathbf{Z}_{ij}(X_{ij}) - \frac{\tilde{S}_j^{(1)}(\tilde{\beta}, X_{ij})}{\tilde{S}_j^{(0)}(\tilde{\beta}, X_{ij})}\right\} \tag{4.12}$$

$$- \sum_{k=1}^{n}\sum_{l=1}^{J}\left\{\frac{\Delta_{kl}Y_{ij}(X_{kl})\exp(\tilde{\beta}'\mathbf{Z}_{ij}(X_{kl}))}{n_j\tilde{S}_j^{(0)}(\tilde{\beta}, X_{kl})}\right\}\times\left\{\mathbf{Z}_{ij}(X_{kl}) - \frac{\tilde{S}_j^{(1)}(\tilde{\beta}, X_{kl})}{\tilde{S}_j^{(0)}(\tilde{\beta}, X_{kl})}\right\}, \tag{4.13}$$

$$\tilde{A}_j(\tilde{\beta}) = n_j^{-1}\sum_{i=1}^{n}\Delta_{ij}\left\{\frac{S_j^{(2)}(\tilde{\beta}, X_{ij})}{S_j^{(0)}(\tilde{\beta}, X_{ij})} - \frac{S_j^{(1)}(\tilde{\beta}, X_{ij})^{\otimes 2}}{S_j^{(0)}(\tilde{\beta}, X_{ij})}\right\}. \tag{4.14}$$

Compared to the variance estimator in Maitra et al. (2020), we apply a weight of w_{ij} on the contribution of the record of the ith individual and jth event, rather than an averaged weight across events applied at the individual-level, $\frac{\zeta_i}{\tilde{\alpha}}$, in $\tilde{Q}(\tilde{\beta})$.

The proof of asymptotic normality follows from the proof of consistency and the Taylor series expansion of $\tilde{U}(\beta)$.

Based on Kang and Cai (2009), under regularity conditions, the estimator of the baseline cumulative hazard functions $\tilde{\Lambda}_{0j}(\tilde{\beta}, t)$ converges in probability to $\Lambda_{0j}(t)$ uniformly in $t \in [0, \tau]$, $j = 1, \ldots, J$. $n^{1/2}[\{\tilde{\Lambda}_{01}(\tilde{\beta}, t) - \Lambda_{01}(t)\}, \ldots, \{\tilde{\Lambda}_{0J}(\tilde{\beta}, t) - \Lambda_{0J}(t)\}]'$ converges weakly to a mean zero Gaussian process $\mathcal{W}(t)$ in $D[0, \tau]^J$, where $\mathcal{W}(t) = \{\mathcal{W}_1(t), \ldots, \mathcal{W}_J(t)\}^T$. The covariance function between $\mathcal{W}_j(t_1)$ and $\mathcal{W}_k(t_2)$ is $\phi_{jk}(t_1, t_2)(\beta_0) = E\{v_{ij}(\beta_0, t_1)v_{ik}(\beta_0, t_2)\}$, where

$$r_j(\beta, t) = -\int_0^t e_j(\beta, t) s_j^{(0)}(\beta, u) d\mu_0(u), \tag{4.15}$$

$$v_{ij}(\beta, t) = r_j(\beta, t)' A(\beta)^{-1} \sum_{m=1}^{J} M_{\tilde{\mathbf{Z}},im}(\beta) + \int_0^t \{s_j^{(0)}(\beta, u)\}^{-1} dM_{ij}(u). \tag{4.16}$$

4.2.2 Impact of Sampling

In this subsection, we propose three feasible sampling schemes to obtain the subcohort and describe how to perform each sampling scheme. Then in the next section, we compare the effect each scheme will have on estimation through a simulation study and provide practical guidance on choosing a sampling scheme in practice.

Pooled Sampling

The most straightforward approach is to sample the subcohort by drawing a random sample of all available records, regardless of which event occurrence the record is from. We term this sampling scheme "pooled sampling" as we pool together all records before obtaining the subcohort. After performing subcohort sampling, we supplement with the rest of the cases from each event occurrence.

We first focus on scenarios where there is a maximum of $J = 2$ event occurrences. Let $\mathcal{C}_1$ and $\mathcal{C}_2$ denote the set of all records within the first and second event occurrences, respectively. We denote the full cohort to contain n individuals. In $\mathcal{C}_1$ and $\mathcal{C}_2$, let there be n_1 and n_2 records, and p_{D1} and p_{D2} event rates, respectively. The pre-specified subcohort proportion is q and total subcohort size is $\tilde{n} = n_1 q + n_2 q$. We note that while we may assume that those who experienced the first event are at risk for the second event, i.e. $n_2 = n_1 p_{D1}$, in practice some individuals may experience the first event and be lost to follow-up right after their

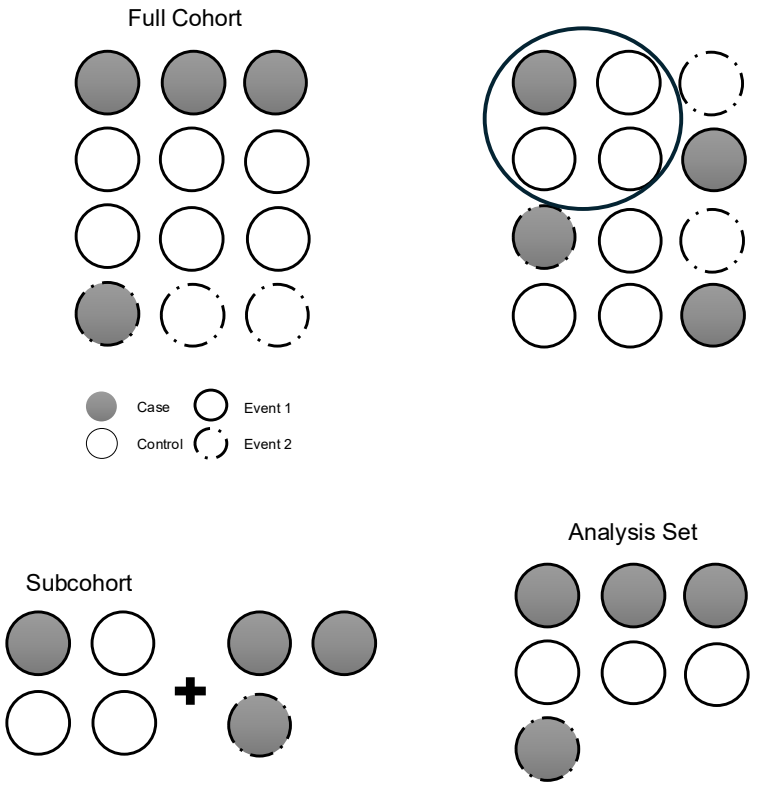


Figure 4.1: An illustration of one example of pooled sampling. Records which are cases are shaded and controls are not shaded. Records in Event 1 have solid borders and records in Event 2 have dotted borders. Random subcohort sampling is illustrated by the circle in the top-right graphic.

event time. If such dropout occurs frequently, for example if participants who experienced the first event withdrew because of worsening disease symptoms, it would be important to assess the reasonableness of a missing completely at random (MCAR) or missing at random (MAR) assumption.

This sampling scheme is easy to implement. However, as shown in Figure 4.1, when there are more records in the first event occurrence than in the second event occurrence ($n_1 > n_2$), then $n_1 q > n_2 q$, and controls from the second event occurrence may not be well-represented in the analysis set.

In Maitra et al. (2020), sampling is done at the individual-level. If an individual is sampled into the subcohort or experiences at least one event, the case-cohort analysis set will include all records of this individual. This may also be considered a type of pooled sampling. As with our proposed pooled sampling, it may lead to over-represented controls for the event occurrence with a larger set of records.

Event-Specific, Balanced Sampling

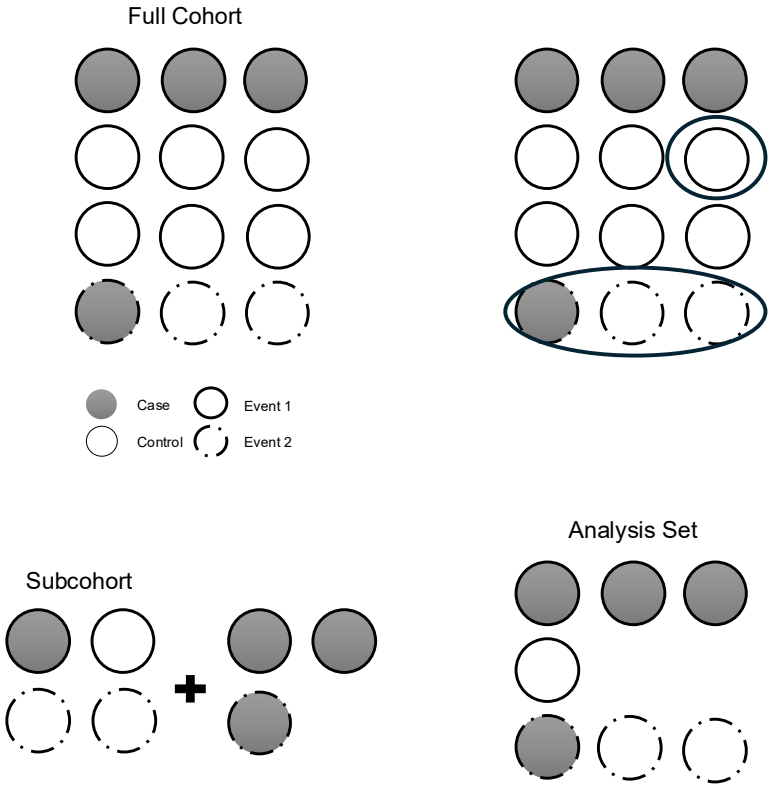


Figure 4.2: An illustration of one example of balanced sampling. Records which are cases are shaded and controls are not shaded. Records in Event 1 have solid borders and records in Event 2 have dotted borders. Random subcohort sampling is illustrated by the circle in the top-right graphic.

To exercise greater control over how many records to sample into the subcohort within each

event occurrence, we propose utilizing "event-specific sampling", where we may designate the proportion of the subcohort we plan to sample from each event occurrence.

We term one way to perform event-specific sampling to be "balanced sampling", which attempts to balance representation in the subcohort by increasing or maximizing the proportion of records coming from later event occurrences. An illustration of balanced sampling is shown in Figure 4.2.

For a cohort with a maximum of two event occurrences, or $J = 2$, notations for event-specific sampling schemes are shown in Table 4.1.

	Total	Event Rates	Full Cohort Cases	Full Cohort Controls	Sampling Proportion	Subcohort Size
Event 1	n_1	p_{D1}	$n_1 p_{D1}$	$n_1(1-p_{D1})$	q_1	$\tilde{n}_1$
Event 2	n_2	p_{D2}	$n_2 p_{D2}$	$n_2(1-p_{D2})$	q_2	$\tilde{n}_2$

Table 4.1: Notation for event-specific sampling schemes when $J = 2$.

	Subcohort Cases	Subcohort Controls
Event 1	$\frac{p_{D1}}{(1-p_{D1})}x$	x
Event 2	$\frac{p_{D2}}{(1-p_{D2})}mn_2p_{D2}$	mn_2p_{D2}
Total		$\{n_1(1-p_{D1}) + n_2(1-p_{D2})\}q^*$

Table 4.2: Subcohort quantities for Step 1 of event-specific, balanced sampling, when we solve for the number of Event 1 control, denoted as x.

To perform sampling of the subcohort, we follow the steps below:

1. To maximize representation of Event 2 controls but still leave "room" to sample Event 1 controls, first let the case-to-control ratio in Event 2 be 1:m. m can represent a practical upper bound for case-cohort efficiency, similar to those in case-control studies, where having more than 4 controls per case does not substantially increase efficiency. In subsequent empirical studies we set $m \leq 3,\ m \in \mathbb{N}$.

   As shown in Table 4.2, we may fill in each cell by first letting x be the number of controls we sample into the subcohort from Event 1. As the case-to-control ratio in

Event 1 is $p_{D1} : (1 - p_{D1})$, the number of cases in Event 1 sampled into the subcohort is $\frac{p_{D1}}{(1-p_{D1})}x$.

To enable a fair comparison with pooled sampling, we match the total controls in each event-specific scheme to that in the pooled design. Using the defined $q*$ that we set to be common across pooled and event-specific designs, we can solve for x by solving $x + mn_2p_{D2} = \{n_1(1 - p_{D1}) + n_2(1 - p_{D2})\}q^*$, which equates the total number of subcohort controls to the number expected under pooled sampling.

2. Next, evaluate x:

   (a) If $x > 0$, then $\tilde{n}_1 = \frac{p_{D1}}{(1-p_{D1})}x + x = \frac{1}{1-p_{D1}}x$.

   (b) If $x \leq 0$, the desired Event 2 case-to-control ratio is not achievable at the target total subcohort size. Reduce m by 1 and solve again for x. Repeat until $x > 0$. Then, follow step 2(a).

3. Desired subcohort sampling proportions are $q_1 = \tilde{n}_1/n_1$ and $q_2 = \tilde{n}_2/n_2$.

Event-Specific, Events-Proportional Sampling

In the balanced sampling scheme described above, the reverse problem may arise: later event occurrences can become over-represented in the subcohort. As an alternative event-specific approach, we propose basing the proportion of controls sampled in each event occurrence on the corresponding proportion of cases in that occurrence. We term this "events-proportional sampling". With this sampling scheme, we achieve similar case-to-control ratios in each event occurrence such that each stratum (and its controls) is adequately represented in the subcohort. Figure 4.3 shows an illustration of events-proportional sampling.

To perform sampling, we follow the steps below:

1. First, we set a 1:1 case-to-control ratio in each event occurrence, hence there will

	Subcohort Cases	Subcohort Controls
Event 1	$\frac{p_{D1}}{(1-p_{D1})} n_1 p_{D1}$	$n_1 p_{D1}$
Event 2	$\frac{p_{D2}}{(1-p_{D2})} n_2 p_{D2}$	$n_2 p_{D2}$
Total		$\{n_1(1-p_{D1}) + n_2(1-p_{D2})\}q^*$

Table 4.3: Subcohort quantities for Step 1 of event-specific, events-proportional sampling.

be $n_1 p_{D1}$ cases and $n_1 p_{D1}$ controls in Event 1 in the analysis set. To account for the cases randomly sampled into the subcohort (not those supplemented in after subcohort sampling), based on the case-to-control ratio in Event 1, we calculate that there are approximately $\frac{p_{D1}}{(1-p_{D1})} n_1 p_{D1}$ cases sampled into the subcohort, as shown in Table 4.3.

2. By constructing Table 4.3, we may solve for $a = \{n_1(1-p_{D1}) + n_2(1-p_{D2})\}q^* - (n_1 p_{D1} + n_2 p_{D2})$, which can be interpreted as the "room" still available if we equate the number of controls to sample into the subcohort (and therefore in the analysis set) to that in the pooled design.

   (a) If $a \leq 0$, the case-cohort design may not be useful, as event rates may be high in at least one event occurrence.

   (b) If $a > 0$, increase sampling probability. Let x be the number of controls to add to the subcohort allocated to Event 1, and y be the number of controls to add to the subcohort allocated to Event 2. Then, solve the following equations for x and y:

$$\begin{cases} x + y = a \\ \frac{n_1 p_{D1}}{n_1 p_{D1} + x} = \frac{n_2 p_{D2}}{n_2 p_{D2} + y} \end{cases}$$

   Hence, $\tilde{n}_1 = \frac{n_1 p_{D1} + x}{(1-p_{D1})}$, $\tilde{n}_2 = \frac{n_2 p_{D2} + y}{(1-p_{D2})}$.

3. Desired subcohort sampling proportions are $q_1 = \tilde{n}_1/n_1$ and $q_2 = \tilde{n}_2/n_2$.

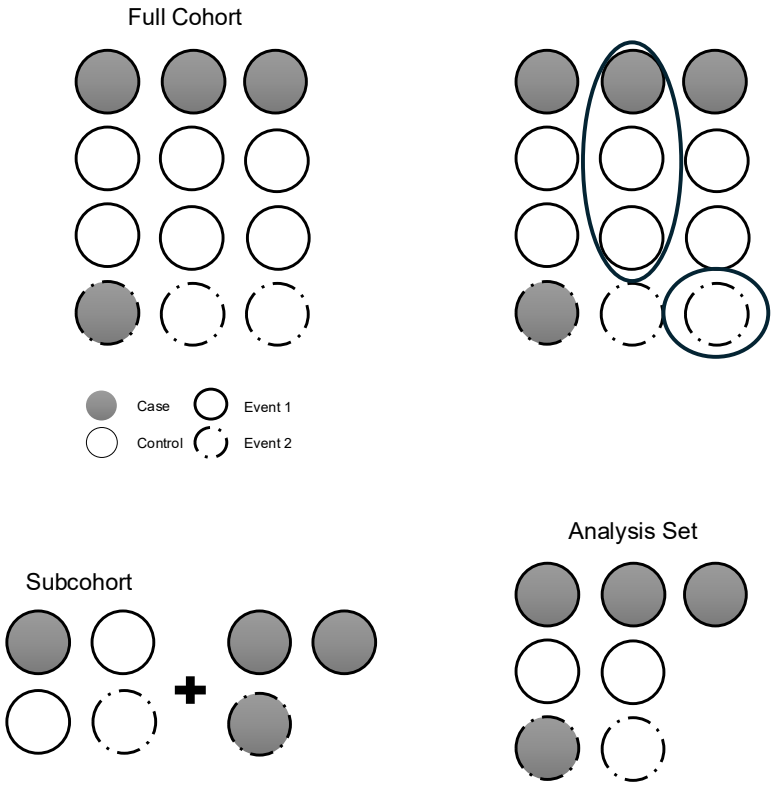


Figure 4.3: An illustration of one example of events-proportional sampling. Records which are cases are shaded and controls are not shaded. Records in Event 1 have solid borders and records in Event 2 have dotted borders. Random subcohort sampling is illustrated by the circle in the top-right graphic.

Sampling for More Than Two Events

For $J > 2$, pooled sampling remains the same, where we take a random sample of all records pooled together, with a total subcohort size of $\sum_{j=1}^{J} n_j q$.

In event-specific sampling, we may follow the same principles described in the previous sections but researchers will need to make decisions based on the objective of the study. For

example, researchers may choose to maximize the inclusion of all events $j > 2$, allocating the rest of the subcohort to be split among events $j = 1, 2$, and calculate q_j accordingly. Or, the researcher may choose to maintain the case-to-control ratio in all events, which will require more unknown quantities in the systems of linear equations. For instance:

$$\begin{cases} x + y + z = a \\ \frac{n_1 p_{D1}}{n_1 p_{D1} + x} = \frac{n_2 p_{D2}}{n_2 p_{D2} + y} = \frac{n_3 p_{D3}}{n_3 p_{D3} + z} \end{cases}$$

where z is the number of controls to add in Event 3, for $J = 3$. This would require other constraints set by the researcher to solve for a unique solution.

4.3 Simulation Study

4.3.1 Performance of Proposed Estimators

	$\beta_1 \approx 0.2607$			$\beta_2 \approx 0.4063$		
	Est. (% Bias)	Robust Var.	Empirical Var.	Est. (% Bias)	Robust Var.	Empirical Var.
Full Cohort	0.2628 (0.81%)	0.0006	0.0006	0.4044 (-0.47%)	0.0031	0.0030
CC (KC)	0.2639 (1.24%)	0.0012	0.0011	0.4102 (0.97%)	0.0061	0.0061
CC (KC time-varying)	0.2640 (1.25%)	0.0012	0.0011	0.4103 (0.98%)	0.0061	0.0061
CC (Prentice)	0.2646 (1.50%)	0.0016	0.0020	0.4057 (-0.15%)	0.0106	0.0106

Table 4.4: Simulation results showing relative bias and variance for the event-agnostic model. All case-cohort (CC) models used pooled sampling. CC (KC) used time-invariant weights w_{ij}^{KC}, CC (KC time-varying) used time-varying weights $w_{ij}^{KC}(t)$, and CC (Prentice) used weights $w_{ij}^{P}(t)$. For the CC designs, Robust Var. is the estimated variance using the proposed variance estimator. "True" values of coefficients are obtained from 10 simulations of 100,000 subjects.

We conducted simulations to demonstrate the performance of the event-agnostic model coefficient estimator and the proposed robust variance estimator. We performed 1,000 simulations with 10,000 individuals. For individual i and the jth event, true failure times T_{ij} were generated with hazard rate $0.1 \times \exp\{\beta_1 Z_{1,ij}(t) + \beta_2 Z_{2,i}(t) + \beta_3 I_{ij}(j = 2) + b_{0i}\}$. The

maximum number of events, J, was set to 2, but more event occurrences can be designated. The coefficient values were $\beta_1 = \log(1.3),\ \beta_2 = \log(1.5),\ \beta_3 = \log(1.1)$. Z_2 was a binary covariate and $Z_{2,i} \sim \text{Bern}(0.2)$. We generated $\mathbf{Z}_{1,i} = (Z_{1,i1}, Z_{1,i2}) \sim MVN(\mu_i, \Sigma)$, where $\mu_i = 0.1 \times \mathbf{Z}_{2,i} = 0.1(Z_{2,i}, Z_{2,i})$ and $\Sigma = \left(\begin{smallmatrix} 1 & 0.5 \\ 0.5 & 1 \end{smallmatrix}\right)$. Additionally, $b_{0i} \sim \text{Normal}(0,\ SD = 0.1)$ is a within-subject frailty term to induce correlation within each individual. For each event, we generated a censoring time $C_{i1} \sim \text{Uniform}(0,\ 2.5)$ and $C_{i2} \sim \text{Uniform}(0,\ 4)$, which can reflect studies where there are differential censoring times for each event and to obtain desired differential event rates. The total censoring time for each individual is $C_i = C_{i1} + C_{i2}$. The event/failure rates were 13% and 23% for the first and second events, respectively, which were chosen since case-cohort designs are most effective when event rates are low. We take 20% of the full cohort records as the subcohort. To the generated data, we fit models marginalizing over the frailty term and assuming that the baseline hazards are different for event j. Because the marginal hazard ratio is noncollapsible (it differs from the conditional hazard ratio in the presence of a frailty term) (Struthers and Kalbfleisch, 1986; Greenland et al., 1999), the true marginal values of β_1 and β_2 that the estimators target were approximated by fitting the marginal model to 10 simulations of 100,000 individuals each.

In these simulations, we compared results across analyses on the full cohort, the case-cohort design with Kang and Cai-type weights, and the case-cohort design with Prentice-type weights. All models assumed a multiplicative hazards model with different baseline hazard functions across event occurrences: $\lambda\{t|\mathbf{Z}_{ij}(t)\} = Y_{ij}(t)\lambda_{0j}(t)\exp\{\beta'\mathbf{Z}_{ij}(t)\}$.

We found that case-cohort designs, regardless of weighting scheme, consistently estimated the coefficients and are comparable to full cohort analysis. The robust variance estimator performed well under each CC design. Designs using KC weights showed higher efficiency than those using Prentice weights. However, utilizing time-varying weights may cause estimates to become large in later event times, as $\tilde{\alpha}_j$ may approach zero. One solution is to perform weight-trimming by truncating weights that exceed a specified percentile of the

empirical weight distribution.

4.3.2 Sampling Comparisons and Impact on Efficiency

	Event 1 (13% Event Rate)			Event 2 (23% Event Rate)		
	Cases	Controls	Ratio	Cases	Controls	Ratio
Pooled CC	1298	1740	1:1.3	292	202	1.4:1
Balanced CC	1298	1065	1.2:1	292	877	1:3
Events-prop. CC	1298	1586	1:1.2	292	357	1:1.2

Table 4.5: Sample sizes of simulation results comparing case-cohort sampling schemes.

	$\beta_1 \approx 0.2607$			$\beta_2 \approx 0.4063$		
	Est. (% Bias)	Robust Var.	Emp. Var.	Est. (% Bias)	Robust Var.	Emp. Var.
Full Cohort	0.2627 (0.76%)	0.0006	0.0006	0.4066 (0.08%)	0.0031	0.0032
Pooled CC	0.2621 (0.52%)	0.0012	0.0012	0.4083 (0.50%)	0.0065	0.0064
Balanced CC	0.2641 (1.30%)	0.0014	0.0015	0.4094 (0.76%)	0.0074	0.0076
Events-prop. CC	0.2622 (0.59%)	0.0012	0.0012	0.4084 (0.52%)	0.0062	0.0062

Table 4.6: Simulation results showing relative bias and variance for event-agnostic models with different sampling schemes. All case-cohort (CC) models use $w_{ij}^P(t)$ weights. "True" values of coefficients are obtained from 10 simulations of 100,000 subjects.

To empirically evaluate the effect of sampling on estimation, we first examine the bias and variance under simulations in the event-agnostic model, i.e. combining all records to perform recurrent events analysis with one model. We then examine these same quantities for the fully stratified models to assess event-specific effects. We performed 1,000 simulations with 10,000 individuals under the same settings as the previous section, using Prentice-type weights. We take 20% of the full cohort records as the subcohort for pooled sampling and match the sizes of the event-specific design subcohorts to be approximately equal in size.

In Table 4.5, we show the average sample sizes, including the number of cases and controls, across simulations. We find that the pooled CC design over-represents Event 1 controls while under-representing Event 2 controls. The balanced design aims to combat this issue with higher representation of Event 2 controls and achieved a case-to-control ratio of 1:3 in

Event 2, on average. The events-proportional design, on the other hand, achieved a similar case-to-control ratio across event occurrences of approximately 1:1.2.

In Table 4.6, we see that when compared to the full cohort event-agnostic model, all case-cohort event-agnostic models show comparable bias in estimation. Compared to the pooled sampling design, the balanced design shows an estimated 14% higher variance while the events-proportional design shows an estimated 5% lower variance in assessing β_2.

Event 1	$\beta_{event1,1} \approx 0.2606$			$\beta_{event1,2} \approx 0.4000$		
	Est. (% Bias)	Robust Var.	Emp. Var.	Est. (% Bias)	Robust Var.	Emp. Var.
Full Cohort	0.2634 (1.08%)	0.0008	0.0008	0.4082 (2.05%)	0.0039	0.0039
Pooled CC	0.2626 (0.78%)	0.0011	0.0014	0.4097 (2.41%)	0.0055	0.0075
Balanced CC	0.2653 (1.81%)	0.0014	0.0021	0.4120 (3.00%)	0.0077	0.0108
Events-prop. CC	0.2633 (1.02%)	0.0011	0.0015	0.4097 (2.43%)	0.0058	0.0078
Event 2	$\beta_{event2,1} \approx 0.2589$			$\beta_{event2,2} \approx 0.3967$		
	Est. (% Bias)	Robust Var.	Emp. Var.	Est. (% Bias)	Robust Var.	Emp. Var.
Full Cohort	0.2600 (0.43%)	0.0035	0.0034	0.4002 (0.87%)	0.0154	0.0171
Pooled CC	0.2635 (1.78%)	0.0067	0.0090	0.4084 (2.96%)	0.0300	0.0419
Balanced CC	0.2599 (0.38%)	0.0036	0.0036	0.4006 (0.98%)	0.0156	0.0185
Events-prop. CC	0.2602 (0.51%)	0.0046	0.0060	0.4068 (2.55%)	0.0204	0.0274

Table 4.7: Simulation results showing relative bias and variance for fully stratified models. All case-cohort models use Prentice weights. True coefficient values are obtained from 100 simulations of 10,000 subjects.

In Table 4.7, we apply the same analysis sets derived for the event-agnostic models to the fully stratified model. For Event 1, variance estimates of $\tilde{\beta}_{event1,1}$ are comparable across all CC designs. In the estimation of $\tilde{\beta}_{event1,2}$, compared to the pooled design, the balanced design shows approximately 40% increase in variance while the events-proportional design shows similar variance to the pooled design.

For Event 2, however, where analyses generally suffer from smaller risk sets, we observe efficiency increases when using event-specific designs. Compared to the pooled design, in assessing $\beta_{event2,1}$ and $\beta_{event2,2}$, the variance estimates are approximately 46% and 48% lower when using the balanced design; the estimates are approximately 31% and 32% lower when using the events-proportional design. This illustrates that with event-agnostic model esti-

mates remaining consistent and efficiency remaining similar, event-specific CC models may be highly efficient in assessing effects in later events compared to the pooled design.

We note that the robust variance estimates in the fully stratified models fall below the empirical variance estimates, particularly for the pooled design in Event 2. This underestimation likely reflects the sparse Event 2 controls under pooled sampling, which can cause the robust estimator to understate the true sampling variability. This is an important practical limitation of the pooled design.

4.4 Application

We applied the proposed method and sampling designs to data from the Anti-Amyloid Treatment in Asymptomatic Alzheimer's (A4) Study and the Longitudinal Evaluation of Amyloid Risk and Neurodegeneration (LEARN) Study (ACTC; Sperling et al., 2020). The A4 Study was a Phase III study designed to test the drug solanezumab in individuals with preclinical AD. Preclinical AD was defined as having evidence of brain amyloid pathology on Positron Emission Tomography (PET) imaging, so these participants were clinically normal but at high risk for cognitive decline. The A4 Study enrolled 1,169 randomized participants. The LEARN Study enrolled 540 individuals who screen-failed for the A4 Study for not showing elevated amyloid and were followed longitudinally with all of the same assessments as in A4.

The recurrent events in our analysis were defined based on the Preclinical Alzheimer Cognitive Composite (PACC), a composite of 4 measures established as sensitive to decline specific to the amyloid-positive and cognitively normal population, as opposed to decline that is associated with aging (Donohue et al., 2014). The PACC was regularly measured in the studies at screening, Week 0, then at regular 6-week intervals after Week 0. We limited our study period to the placebo-controlled period, not extending into the open-label extension period.

We defined the first event as a 1.5-standard deviation (SD) decrease in standardized PACC compared to baseline (measured at Week 0) and the second event as a 3-SD decrease in standardized PACC compared to baseline.

We aimed to investigate the association of cholesterol with cognitive decline within preclinical AD participants. Past studies have shown somewhat conflicting evidence on this association. While some studies found modest positive association between cholesterol and dementia (i.e. higher levels associated with higher risk), others found no significant association or even negative association (Iwagami et al., 2021; Mielke et al., 2005; Shepardson et al., 2011). Some reasons for these discrepancies may be the difference between studies that measure total cholesterol versus high-density lipoprotein (HDL) and low-density lipoprotein (LDL) levels, and whether the study was conducted in participants' later life when substantial neuropathological AD type might be present (Shepardson et al., 2011). Hence, we are interested in further investigating this association in the preclinical AD study population of the A4 and LEARN studies. In these studies, cholesterol measurements were taken at 12-week intervals until Week 60, after which they were taken at Weeks 108, 168, and 240. At each event/censoring time, the most recent measurement of cholesterol was used. We standardized cholesterol by mean and SD.

In addition, we considered adjustment by variables that were hypothesized to be confounders and/or associated with the outcome. We adjusted for substudy (A4 vs. LEARN), as the amyloid eligibility criteria differed between the two studies. We also adjusted for demographic variables including age, sex, and race and ethnicity (collapsing all groups of individuals who were not non-Hispanic White (NH White) due to small cell sizes). Additionally, education level has been found to be potentially related to cognitive reserve, a protective factor associated with AD risk (Silva et al., 2019), while genetic status of being an APOE e4 (APOE4) carrier is strongly linked to AD risk (A&D, 2025). Having a family history of dementia, defined here as any parent with dementia or significant memory impairment, is a known risk

factor for AD. Therefore, we adjusted for these variables.

In the following, we provide full cohort analysis results as reference. In applications of case-cohort designs, we would not know expensive covariate histories on non-case records outside of the subcohort.

Covariate	Total (N=1632)	Low Baseline Cholesterol (N=850)	High Baseline Cholesterol (N=782)
Substudy			
A4	1112 (68.1%)	612 (72.0%)	500 (63.9%)
LEARN	520 (31.9%)	238 (28.0%)	282 (36.1%)
Age (yrs)	71.5 (4.7)	72.0 (4.8)	70.9 (4.5)
Sex			
Male	659 (40.4%)	473 (55.6%)	186 (23.8%)
Female	973 (59.6%)	377 (44.4%)	596 (76.2%)
Race and Ethnicity			
NH White	1490 (91.3%)	781 (91.9%)	709 (90.7%)
All other groups	142 (8.7%)	69 (8.1%)	73 (9.3%)
Education (yrs)	16.6 (2.7)	16.7 (2.8)	16.6 (2.6)
APOE4 Status			
Non-carrier	852 (52.2%)	462 (54.4%)	390 (49.9%)
Carrier	778 (47.7%)	386 (45.4%)	392 (50.1%)
AD Family History			
No	540 (33.1%)	295 (34.7%)	245 (31.3%)
Yes	1085 (66.5%)	554 (65.2%)	531 (67.9%)

Table 4.8: Table of summary statistics of A4 and LEARN participants at baseline. Low and high cholesterol groups are categorized by lower than and higher than cohort mean cholesterol levels at baseline. "All other groups" in race and ethnicity includes Hispanic or Latino, NH Asian, NH Black, or other race.

In Table 4.8, we find that when stratifying the sample (examining only the first event records) by the mean of cholesterol levels, there is a higher proportion of female participants among the high cholesterol group. Other covariates show similar distributions across the cholesterol groups. Study participants were generally older, non-Hispanic White, and had a family history of dementia or significant memory impairment.

Based on the full cohort data, the Kaplan-Meier estimates of the event-free survival functions

for Event 1 and Event 2, displayed in Figure 4.4, show no large separation between the low and high cholesterol groups in either event occurrence. However, Event 2 shows slightly steeper curves and a marginally larger separation between cholesterol groups, suggesting a higher baseline hazard and a potentially stronger association at the second progression.

	N (Records)	Est. (Adjusted 95% CI)	SE	p-value
Full Cohort	1884	1.060 (0.957, 1.173)	0.052	0.263
Pooled CC	702	1.123 (0.938, 1.343)	0.092	0.207
Balanced CC	695	1.014 (0.839, 1.226)	0.097	0.885
Events-prop. CC	703	1.013 (0.866, 1.184)	0.080	0.874

Table 4.9: Event-agnostic model results for estimating the effect of cholesterol levels in association to PACC decline, adjusted for covariates. All case-cohort models use Prentice weights.

Event 1	N (Records)	Est. (Adjusted 95% CI)	SE	p-value
Full Cohort	1623	1.057 (0.937, 1.192)	0.061	0.371
Pooled CC	568	1.232 (1.030, 1.472)	0.091	0.022
Balanced CC	434	0.967 (0.759, 1.232)	0.124	0.786
Events-prop. CC	540	0.939 (0.792, 1.114)	0.087	0.471
Event 2	N (Records)	Est. (Adjusted 95% CI)	SE	p-value
Full Cohort	261	1.007 (0.828, 1.225)	0.100	0.944
Pooled CC	134	0.855 (0.536, 1.364)	0.238	0.511
Balanced CC*	261	1.007 (0.828, 1.225)	0.100	0.944
Events-prop. CC	163	1.127 (0.878, 1.446)	0.127	0.348

Table 4.10: Fully stratified model results for estimating the effect of cholesterol levels in association with PACC decline, adjusted for covariates. All case-cohort models use Prentice weights. *In Event 2, balanced CC sampled all available records, so full cohort analysis was performed.

We performed analysis of the association between cholesterol levels (standardized) and time to PACC decline using the full cohort, pooled sampling case-cohort (CC), balanced sampling CC, and events-proportional CC designs, for illustration. We set q^* in the pooled design to be 0.2 and match other schemes to its sample size. Comparing the results of the predictor

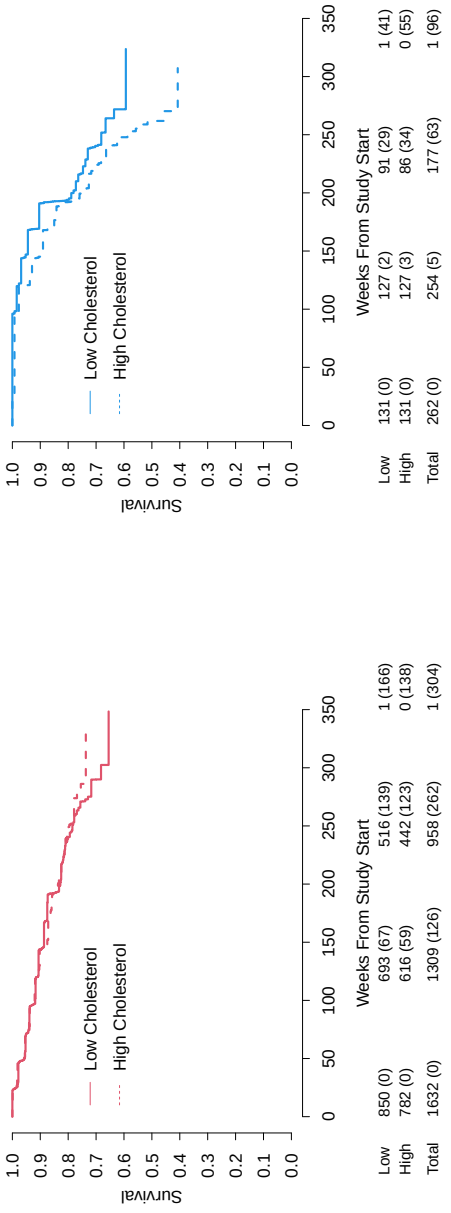


Figure 4.4: Kaplan-Meier plots for Event 1 (left) and Event 2 (right) based on full cohort data in the application of proposed method to A4 and LEARN data. Low and high cholesterol levels are binarized based on median cholesterol levels of records at risk for Events 1 and 2, respectively.

			Unadjusted			Adjusted		
	N	Ne	Est. HR	95% CI	p-value	Est. HR	95% CI	p-value
Cholesterol (standardized unit)	703	399	1.01	(0.88, 1.15)	0.934	1.01	(0.87, 1.18)	0.874
Substudy								
A4	534	338	Referent	-	-	Referent	-	-
LEARN	169	61	0.35	(0.25, 0.48)	< 0.001	0.53	(0.35, 0.78)	0.002
Age (10-year)	703	399	2.19	(1.66, 2.89)	< 0.001	2.19	(1.59, 3.02)	< 0.001
Sex								
Male	287	166	Referent	-	-	Referent	-	-
Female	416	233	0.91	(0.69, 1.19)	0.49	0.95	(0.69, 1.30)	0.731
Race and Ethnicity								
NH White	639	363	Referent	-	-	Referent	-	-
All Other Groups	64	36	0.88	(0.55, 1.41)	0.597	0.86	(0.51, 1.44)	0.563
Education (year)	703	399	1.01	(0.96, 1.06)	0.8	0.99	(0.94, 1.04)	0.680
APOE4 Status								
Non-Carrier	302	136	Referent	-	-	Referent	-	-
Carrier	401	263	2.08	(1.57, 2.75)	< 0.001	1.81	(1.29, 2.53)	0.001
Family History								
No	235	134	Referent	-	-	Referent	-	-
Yes	468	265	1.02	(0.77, 1.35)	0.913	0.95	(0.69, 1.31)	0.752

Table 4.11: Full event-agnostic model results using the events-proportional CC design. N: number of records. Ne: number of events.

of interest for these designs in the event-agnostic model (Table 4.9), we find that using approximately 37% of the full cohort records, all CC designs yield estimates comparable to the full cohort. As shown in Table 4.9, there is approximately 13% and 12% lower standard error (SE) in the estimate of the effect of interest when using the events-proportional CC design compared to the pooled and balanced CC designs, respectively. We do not find a significant effect of cholesterol marginally across progressions.

In the fully stratified models (Table 4.10), we do not find significant associations of cholesterol with PACC decline except in the pooled CC design for Event 1. The pooled CC estimates for Event 1 differ substantially from the full cohort results, likely reflecting chance variation arising from the inadequate number of Event 2 controls under this design. While the SE in the balanced CC design shows a 58% decrease compared to the pooled CC design in Event

2, the SE in Event 1 increases by 36%. On the other hand, while there is a similar SE in Event 1 when using the events-proportional CC design, compared to the pooled CC design, there is a 47% decrease in SE in Event 2. This presents a desirable trade-off for increasing efficiency in later events, where we may suffer from lower power and efficiency.

Table 4.11 shows results when performing analysis using the events-proportional CC design. We do not find a significant association of cholesterol with PACC decline, (HR (95% CI): 1.01 (0.87, 1.18); p-value: 0.874). We find that those in the LEARN study have an estimated 47% lower hazard of PACC decline (95% CI: (0.35, 0.78); p-value: 0.002), while those 10 years older are estimated to have a 2.2-fold greater hazard of PACC decline (95% CI: (1.59, 3.02); p-value: < 0.001), and those who are APOE4 carriers (compared to non-carriers) are estimated to have a 1.8-fold greater hazard of PACC decline (95% CI: (1.29, 2.53); p-value: 0.001), when adjusting for other variables, respectively.

4.5 Discussion

In this work, we first proposed a marginal model for performing case-cohort analysis of recurrent events data, which extends previous work to include event-specific covariate histories. Our empirical results support consistency of coefficient and variance estimators. We then detailed a framework for performing subcohort sampling under recurrent event settings. For a similar subcohort size, we compared efficiency and power results when using the pooled and event-specific sampling schemes through a simulation study and application to preclinical AD trial data.

Our recommendations for choosing a sampling scheme for a study are as follows. We suggest using the events-proportional sampling scheme when both marginal and event-specific inference is desired. The events-proportional scheme can generally show efficiency gains for later event occurrences while marginal efficiency results remain comparable to pooled sampling.

We recommend balanced sampling when the primary goal is maximizing efficiency for later event occurrences, at some cost to Event 1 efficiency. And finally, pooled sampling can be utilized when desiring simplicity and when inference for later events is not a primary focus.

As a limitation, as with all case-cohort designs, there is loss of efficiency and variation in point estimates due to sampling. In general, the savings in costly covariate processing relative to the loss in efficiency are favorable, though researchers should consider carefully whether a case-cohort design is appropriate for their study. Furthermore, we note that if one performs event-specific sampling yet decides to consider an event-agnostic model assuming common baseline hazards across events, there may be bias in estimation as we have differentially sampled records. In addition, the Kang and Cai-type weights, both time-varying and time-invariant, were shown to be more efficient but may be more computationally intensive and could require weight trimming. As an alternative, researchers may use the proposed Prentice-type weights and simplified variance estimator.

As previous works have pointed out, in gap time models, the event time of later event occurrences (after the first event) may have dependence with censoring (Wang, 1999; Huang, 1999). Since the second event becomes observable only if the first event has occurred, the length of the first duration affects the probability of the second duration being censored (Wang and Wells, 1998). Methods, including those nonparametric and semiparametric, have been developed to deal with this issue (Wang and Wells, 1998; Chen, 2004). While in practice we do not know the true censoring mechanism, it is helpful to guard against this potential dependence. One approach is to apply inverse probability of censoring weights (Willems et al., 2018). Another approach is to condition on, or adjust for, the observed time of Event 1. We include results from this approach in the Appendix C. For the simulation study, we now generate one censoring time $C_i \sim \text{Uniform}(0, 3.5)$ for subject i and compare the results obtained through the model unadjusted and adjusted for the time of Event 1 for later events. For the application of the proposed method to A4 and LEARN data, we perform

an analysis adjusted for Event 1 time as well. In these analyses, we did not find substantial impact of censoring on the point estimates or significance of the predictors. We note that the framework we have proposed shows utility in identifying this dependence when performing both the event-agnostic and fully stratified models.

In practice, it may also be desirable to maintain a certain level of power in the analysis of the first event. We may follow the steps outlined for performing event-specific sampling to determine the subcohort proportion for Event 1, q_1, but add a step to calculate $q_{1,min}$ for desired minimal power in Event 1, then take a maximum of q_1 and $q_{1,min}$ and adjust $q_j, j = 2, \ldots$ accordingly.

Kang and Cai (2009) and Maitra et al. (2020) also consider what is termed "generalized case-cohort designs", where, in addition to subcohort sampling, we also sample a proportion of the cases when event rates are high. We only focus on the setting when event rates are low, which is generally when case-cohort designs are most useful and which is common in AD cohort studies.

As a future direction, we consider developing sample size and power calculation formulas for these sampling schemes. Such formulas have been considered by Song et al. (2008) and Cai and Zeng (2004), for full cohort recurrent event settings and case-cohort univariate settings, respectively. To date, no such formulas have been developed for the recurrent events case-cohort setting, which represents an important gap for planning future cohort studies.

4.6 Acknowledgments

Data used in Chapters 4 and 5 were obtained from the A4 and LEARN Studies. The A4 Study was a secondary prevention trial in preclinical Alzheimer's disease, aiming to slow cognitive decline associated with brain amyloid accumulation in clinically normal older individuals. The A4 Study was funded by a public-private-philanthropic partnership, including

funding from the National Institutes of Health-National Institute on Aging, Eli Lilly and Company, Alzheimer's Association, Accelerating Medicines Partnership, GHR Foundation, an anonymous foundation, and additional private donors, with in-kind support from Avid Radiopharmaceuticals, Cogstate, Albert Einstein College of Medicine and the Foundation for Neurologic Diseases. The companion observational Longitudinal Evaluation of Amyloid Risk and Neurodegeneration (LEARN) Study was funded by the Alzheimer's Association and GHR Foundation. The A4 and LEARN Studies were led by Dr. Reisa Sperling at Brigham and Women's Hospital, Harvard Medical School, and Dr. Paul Aisen at the Alzheimer's Therapeutic Research Institute (ATRI) at the University of Southern California. The A4 and LEARN Studies were coordinated by ATRI at the University of Southern California, and the data are made available under the auspices of Alzheimer's Clinical Trial Consortium through the Global Research & Imaging Platform (GRIP). The complete A4 Study Team list is available on: https://www.actcinfo.org/a4-study-team-lists/. We would like to acknowledge the dedication of the study participants and their study partners who made the A4 and LEARN Studies possible.

Chapter 5

Sample size and power for recurrent events case-cohort designs

5.1 Introduction

In Alzheimer's disease (AD) research, many large cohort studies, such as those conducted by the Alzheimer's Disease Research Center, may collect and store cerebrospinal fluid (CSF) samples on enrolled participants. When specific biomarkers may be of interest, researchers will then thaw participants' samples and process a fraction of the sample to obtain measurements. As CSF samples are high burden to obtain and limited in quantity due to the procedure of lumbar punctures (Blazel et al., 2020; Marcantonio et al., 2008; Witbracht et al., 2020), it is important to utilize statistical methods to reduce the amount required to be processed, so that we may plan for more future studies and save processing time and cost. Alternative sampling designs that require processing only a subsample of the full cohort can be very useful, especially in the presence of low event rates, which is common in AD studies.

In the previous chapter, we performed recurrent events analysis to investigate the time to

AD progression from normal to mild cognitive impairment then to dementia. Under these settings, case-cohort designs remain important for preserving biomarker samples and for cost-efficiency.

At the design stage of any rigorous scientific investigation that relies upon empirical evidence, statisticians are crucial to study design decisions, including identification the required sample size to discriminate between scientifically meaningful hypotheses. For a case-cohort study, determining the appropriate subcohort size is an essential component of sample size calculations, as it ensures adequate statistical power while avoiding unnecessary expenditure of resources.

Several works have developed power and sample size (PSS) formulas for survival data. In a full cohort analysis for univariate data, the logrank test was developed to compare the survival distribution between two (or more) groups and determine PSS (Schoenfeld, 1983). Cai and Zeng (2004) developed a logrank-type test and derived PSS formula for the case-cohort design, while Hu et al. (2014) extended this work for stratified case-cohort designs. On the other hand, Gangnon and Kosorok (2004) developed a logrank-type test and PSS formula for clustered survival data and Song et al. (2008) did so for recurrent events data. However, there are no works on study planning and deriving a logrank-type test and PSS formula for a case-cohort design performed under the recurrent events setting.

In this chapter, we first propose a logrank-type test statistic and derive its asymptotic properties. Then, to best assist in the study-design phase, we provide a simplified PSS formula with simplified quantities from the asymptotic distribution. This is derived for what we term the pooled and event-specific subcohort sampling schemes, which can allow for simplicity or flexibility in subcohort sampling. We conduct a simulation study to compare the performance of the PSS formula to empirical results from simulated data. Finally, we apply the method to data from the Anti-Amyloid Treatment in Asymptomatic Alzheimer's (A4) Study and the Longitudinal Evaluation of Amyloid Risk and Neurodegeneration (LEARN)

Study (ACTC; Sperling et al., 2020), to determine the required subcohort size for a recurrent events case-cohort design assessing a new biomarker among preclinical AD participants.

5.2 Methods

5.2.1 Weighted Logrank Tests Using the Full Cohort

We first define the following quantities for the full cohort data, which contains n independent subjects. Let there be an expensive exposure of interest that can be dichotomized into groups $j = 1, 2$. For example, in AD biomarker research, groups may be high versus low levels of biomarker based on some threshold. Before the study, researchers may specify a meaningful threshold or dichotomize by the median level, for example. For each subject i, $i = 1, 2, \ldots, n_j$, let T^*_{ijk} denote the calendar time of the kth event for subject i in group j, $k = 1, \ldots, K$, with the convention $T^*_{ij0} = 0$ at study entry. Let C_{ij} denote the subject-level calendar censoring time.

The counting process recording all observed events up to calendar time $t \in [0, \tau_0]$ for subject i in group j is $N_{ij}(t) = \sum_{k=1}^{K} I(T^*_{ijk} \leq 1, C_{ij} \geq t)$ and the corresponding at-risk indicator in calendar time is $Y_{ij}(t) = I(C_{ij} \geq t)$. $N^0_{ij}(t)$ is the underlying event process version of $N_{ij}(t)$, where $N_{ij}(t) = \int_0^t Y_{ij}(s) dN^0_{ij}(s)$. Then the aggregate full cohort quantities are given by $\bar{N}_j(t) = \sum_{i=1}^{n_j} N_{ij}(t)$ and $\bar{Y}_j(t) = \sum_{i=1}^{n_j} Y_{ij}(t)$. Let the probability of subjects being under study in group j at time t be $\pi^n_j = n_j^{-1} \sum_{i=1}^{n_j} EI\{C_{ij} \geq t\}$ and $\Lambda_{ij}(t) = E\{N^0_{ij}(t)\}$ to be the expected number of events for subject i in group j by time t.

We make the following assumptions:

1. T_{ij} and C_{ij} are independent for $i = 1, \ldots, n_j, j = 1, 2$.

2. $\lim_{n\to\infty} n_j/n = p_j \in (0, 1)$ and $\sup_{t\in(0,\tau_0)} |\pi^n_j(t) - \pi_j(t)| \to 0$, where τ_0 is the maximum event time.

3. The cumulative mean functions of events for subjects in the same group are identical, $\Lambda_{ij}(t) = \Lambda_j^n(t)$, for $i = 1, 2, \ldots, n_j$, $t \in (0, \tau_0)$. The rate function is $\lambda_j^n(t) = \partial\Lambda_j^n(t)/\partial t$, $j = 1, 2$. Here, the superscript n permits contiguous alternatives.

4. $\sup_{t\in(0,\tau_0)}|\sqrt{n}\{d\Lambda_1^n(t)/d\Lambda_2^n(t) - 1\} - \psi(t)\{1 + \eta(t)\}| \to 0$, where ψ is either cadlag or caglad with bounded total variation and η is bounded and zero except at event times.

5. The cumulative mean functions of each event occurrence for subjects in group j are equal, $\Lambda_{j1}(t) = \Lambda_{j2}(t) = \cdots = \Lambda_{jK}(t) = \Lambda_j(t), j = 1, 2$.

To test the hypothesis $H_0 : \Lambda_1(t) = \Lambda_2(t), t \in [0, \tau_0]$, if all data from the full cohort were available, then the logrank test statistic is

$$L_n = \int_0^{\tau_0} w(s)\frac{\bar{Y}_1(s)\bar{Y}_2(s)}{\bar{Y}_1(s) + \bar{Y}_2(s)} \times \left\{ \frac{d\bar{N}_1(s)}{\bar{Y}_1(s)} - \frac{d\bar{N}_2(s)}{\bar{Y}_2(s)} \right\} \tag{5.1}$$

where $\bar{N}_j(t) = \sum_{i=1}^{n_j} N_{ij}(t)$ and $\bar{Y}_j(t) = \sum_{i=1}^{n_j} Y_{ij}(t)$. $w(\cdot)$ is a weight function, which may be 1 for the usual logrank test or other standard weight functions such as the $G^{\rho,\gamma}$ family (Fleming and Harrington, 1981). Let $\sigma_{FC,n}^2$ denote the variance of L_n. Assuming $\lim_{n\to\infty}\sigma_{FC,n}^2 = \sigma^2 < \infty$, it follows from Song et al. (2008), $n^{-1/2}L_n$ converges in distribution to a normally distributed random variable with mean μ and variance σ^2, where

$$\begin{aligned}
\mu &= \int_0^{\tau_0} w(s)\psi(s)\{1 + \eta(s)\}\frac{p_1\pi_1(s)p_2\pi_2(s)}{p_1\pi_1(s) + p_2\pi_2(s)}d\Lambda_0(s), \\
\sigma_{FC,n}^2 &= \frac{1}{n}\sum_{j=1}^{2}\sum_{i=1}^{n_j}\left\{ \int_0^{\tau_0} w(s)\frac{p_{j'}\pi_{j'}^n(s)}{p_1\pi_1^n(s) + p_2\pi_2^n(s)}dM_{ij}(s) \right\}^2,
\end{aligned}$$

and $M_{ij}(t) = N_{ij}(t) - \int_0^t Y_{ij}d\Lambda_j(s)$.

5.2.2 Logrank-Type Tests for Recurrent Events in the Case Cohort Design

To implement the case-cohort design, we take a simple random sample to form the subcohort and then supplement the subcohort with unsampled cases. Such sampling is straightforward in the univariate setting when each subject has at most one event and therefore one record. To perform subcohort sampling in the recurrent events setting where subjects have more than one record, we may perform sampling at the record-level, which is especially useful when there are event-specific covariates, i.e., when expensive covariates are taken at each event occurrence. One approach is what we term "pooled sampling", where the simple random sample is taken from all records pooled together, regardless of event occurrence, with a sampling probability of q to form the subcohort. For simplicity, we first consider the development of power and sample size formulas for pooled sampling, and address "event-specific" sampling, where stratified sampling on event number is performed, later in this section.

For record-level case-cohort sampling, we define the gap time to event k, or the time between event $k-1$ and event k, for subject i in group j as $T_{ijk} = T^*_{ijk} - T^*_{ij,k-1}$. The corresponding residual follow-up available for event k is defined as $C_{ijk} = C_{ij} - T^*_{ij,k-1}$, defined only for subjects who have experienced events $1, \ldots, k-1$ and set to zero otherwise. Define the observed event-specific time and event indicator for record k of subject i in group j as $X_{ijk} = \min(T_{ijk}, C_{ijk}), \Delta_{ijk} = I(T_{ijk} \leq C_{ijk})$.

Let u denote time on this event-specific scale, $u \in [0, \tau_k]$. Define an eligibility indicator $e_{ijk} = I(\Delta_{ij,k-1} = 1)$ for $k \geq 2$ and $e_{ij1} = 1$, so that subject i in group j is eligible for event k only after experiencing event $k-1$. The event-specific at-risk indicator is then $Y_{ijk}(t) = e_{ijk} \cdot I(X_{ijk} \geq u)$. Then we have that $\bar{Y}_{ij}(t) = \sum_{k=1}^{K} e_{ijk} \cdot I(T^*_{ij,k-1} < t \geq T^*_{ij,k-1} + X_{ijk})$, that is, the calendar-time at-risk process is recovered by summing the event-specific records over K event occurrences.

Now, let $\zeta_{ijk} \in \{0,1\}$ denote the subcohort sampling indicator. The subcohort at-risk indicator, defined in event-specific time u, is $\tilde{Y}_{ijk}(u) = \zeta_{ijk} \cdot Y_{ijk}(u) = \zeta_{ijk} \cdot e_{ijk} \cdot I(X_{ijk} \geq u)$. The aggregate subcohort at-risk count in group j at time u is $\tilde{Y}_j(u) = \sum_{i=1}^{n_j} \sum_{k=1}^{K} \tilde{Y}_{ijk}(u)$. Under Assumption 5, each sampled record into the subcohort represents approximately $1/q$ full cohort records, so we may approximate $\bar{Y}(t)$ with $\tilde{Y}_j(u)/q$.

Then to test the hypothesis $H_0 : \Lambda_1(t) = \Lambda_2(t), t \in [0, \tau_0]$ using the case-cohort design for recurrent events analysis, we consider a stratified logrank test statistic

$$L_{RCC} = \sum_{k=1}^{K} \int_0^{\tau_k} w(u) \frac{\tilde{Y}_{1k}(u)\tilde{Y}_{2k}(u)}{\tilde{Y}_{1k}(u) + \tilde{Y}_{2k}(u)} \times \left\{ \frac{d\bar{N}_{1k}(u)}{\tilde{Y}_{1k}(u)} - \frac{d\bar{N}_{2k}(u)}{\tilde{Y}_{2k}(u)} \right\}, \tag{5.2}$$

where $\tilde{Y}_{jk}(u) = \sum_{i=1}^{n_j} \tilde{Y}_{ijk}(u)$ and $\bar{N}_{jk}(u) = \sum_{i=1}^{n_j} N_{ijk}(u)$ with $N_{ijk}(u) = \Delta_{ijk} \cdot I(T_{ijk} \leq u)$.

To estimate the variance of L_{RCC}, defined as $\sigma^2_{RCC,n}$, we propose an estimator $\hat{\sigma}^2_{RCC}$ with two components: $\hat{\sigma}^2_{FC}$, the full cohort variance estimator of $\sigma^2_{FC,n}$, and $\hat{\sigma}^2_{CC}$, which estimates the extra variance component induced by case-cohort sampling. We propose estimators

$$\hat{\sigma}^2_{FC} = \frac{1}{n} \sum_{j=1}^{2} \sum_{i=1}^{n_j} \left\{ \int_0^{\tau_0} w(s) \frac{\tilde{Y}_{j'}(s)}{\tilde{Y}_1(s) + \tilde{Y}_2(s)} d\tilde{M}_{ij}(s) \right\}^2, \tag{5.3}$$

where $\tilde{M}_{ij}(t) = N_{ij}(t) - \int_0^t \tilde{Y}_{ij} d\bar{N}_j(s)/\tilde{Y}_j(s)$, and $i = 1, \ldots, n_j, j = 1, 2, j' = 3 - j$, and

$$\hat{\sigma}^2_{CC} = \frac{2(1-q)}{n} \times \sum_{k=1}^{K} \sum_{j=1}^{2} \sum_{i=1}^{n_j} \Bigg\{ \frac{\Delta_{ijk} w(X_{ijk}) \tilde{Y}_{1k}(X_{ijk}) \tilde{Y}_{2k}(X_{ijk})}{(\tilde{Y}_{1k}(X_{ijk}) + \tilde{Y}_{2k}(X_{ijk}))^2} \tag{5.4}$$

$$\times \Bigg[\sum_{j'=1}^{2} \sum_{i'=1}^{n_j} \frac{\Delta_{i'j'k} w(X_{i'j'k}) I(X_{i'j'k} \leq X_{ijk})}{\tilde{Y}_{1k}(X_{i'j'k}) + \tilde{Y}_{2k}(X_{i'j'k})} \Bigg] \Bigg\} \tag{5.5}$$

$$- \frac{(1-q)}{n} \sum_{k=1}^{K} \sum_{j=1}^{2} \sum_{i=1}^{n_j} \frac{\Delta_{ijk} w(X_{ijk}) \tilde{Y}_{1k}(X_{ijk}) \tilde{Y}_{2k}(X_{ijk})}{(\tilde{Y}_{1k}(X_{ijk}) + \tilde{Y}_{2k}(X_{ijk}))^3}. \tag{5.6}$$

Theorem 5.1. *Under Assumptions A1-A5, q converges to some $0 < \alpha < 1$, and $\sigma^2_{FC,n} \to \sigma^2_{FC}$ as $n \to \infty$ ($0 < \sigma^2_{FC} < \infty$), then $n^{-1/2} L_{RCC}$ has an asymptotically normal distribution with mean μ and variance σ^2_{RCC}, where*

$$\mu = \int_0^{\tau_0} w(s)\psi(s)\{1+\eta(s)\} \frac{p_1\pi_1(s)p_2\pi_2(s)}{p_1\pi_1(s) + p_2\pi_2(s)} d\Lambda_0(s). \tag{5.7}$$

Theorem 5.2. *Under assumptions outlined in Theorem 5.1, $\hat{\sigma}^2_{RCC} \to \sigma^2_{RCC}$ in probability as $n \to \infty$, where $0 < \sigma^2_{RCC} < \infty$.*

The proof is as follows. In L_{RCC}, the contribution from stratum k to $dN_{ij}(t)$ in the calendar time occurs at $t = T^*_{ijk} = T^*_{ij,k-1} + T_{ijk}$. This corresponds to $u = T_{ijk}$ in event-specific time. Under Assumption 5, summing $\bar{N}_{jk}(u)$ over strata k in event-specific time recovers $\bar{N}_j(t)$ in calendar time since $t = T^*_{ij,k-1} + u$. $\tilde{Y}_j(t)/q$ approximates $\bar{Y}_j(t)$ in the full cohort and all cases are included in the case-cohort design. Hence we have

$$L_{RCC} = \int_0^{\tau_0} w(s) \frac{\tilde{Y}_1(s)\tilde{Y}_2(s)}{\tilde{Y}_1(s) + \tilde{Y}_2(s)} \times \left\{ \frac{d\bar{N}_1(s)}{\tilde{Y}_1(s)} - \frac{d\bar{N}_2(s)}{\tilde{Y}_2(s)} \right\} \tag{5.8}$$

$$= \int_0^{\tau_0} w(s) \frac{\bar{Y}_1(s)\bar{Y}_2(s)}{\bar{Y}_1(s) + \bar{Y}_2(s)} \times \left\{ \frac{d\bar{N}_1(s)}{\bar{Y}_1(s)} - \frac{d\bar{N}_2(s)}{\bar{Y}_2(s)} \right\}, \tag{5.9}$$

and therefore $n^{-1/2}L_{RCC}$ is asymptotically normal and μ equals that in the full cohort analysis, which is derived by arguments in Song et al. (2008).

Further, $L_{RCC}(t)$ can be written similar to the univariate stratified logrank statistic, $\sum_{k=1}^{K}\sum_{i=1}^{n_1} \frac{\Delta_{i1k} w(X_{i1k})\tilde{Y}_{2k}(X_{i1k})}{\tilde{Y}_{1k}(X_{i1k})+\tilde{Y}_{2k}(X_{i1k})} - \sum_{k=1}^{K}\sum_{i=1}^{n_2} \frac{\Delta_{i2k} w(X_{i2k})\tilde{Y}_{1k}(X_{i2k})}{\tilde{Y}_{1k}(X_{i2k})+\tilde{Y}_{2k}(X_{i2k})}$. By Hu et al. (2014), $n^{-1/2}L_{RCC}$ has

asymptotic variance given by full cohort variance plus variance induced by case-cohort sampling.

By arguments in Song et al. (2008), $\hat{\sigma}^2_{FC}$ is a robust estimator of $\sigma^2_{FC,n}$, where

$$\sigma^2_{FC,n} = \frac{1}{n}\sum_{j=1}^{2}\sum_{i=1}^{n_j} \left\{ \int_0^{\tau_0} w(s) \frac{p_{j'}\pi^n_{j'}(s)}{p_1\pi^n_1(s) + p_2\pi^n_2(s)} dM_{ij}(s) \right\}^2. \tag{5.10}$$

By assumption, $\sigma^2_{FC,n} \to \sigma^2_{FC}$ for some $\sigma^2_{FC} < \infty$ as $n \to \infty$, so $\hat{\sigma}^2_{FC} \to \sigma^2_{FC}$ in probability as $n \to \infty$. Combine records across the groups $j = 1, 2$ and let b_k be the total number of records in event occurrence k in the full cohort. Then we have

$$\sigma^2_{CC} = \sum_{k=1}^{K} v_k \int\int \frac{w(t)}{p_1\pi_1(t) + p_2\pi_2(t)} \frac{w(s)}{p_1\pi_1(s) + p_2\pi_2(s)} Q_k(t,s) \frac{dS_k(s)}{S_k(s)} \frac{dS_k(t)}{S_k(t)}, \tag{5.11}$$

where $Q_k(t,s) = \frac{1-\alpha}{\alpha} p_1 p_2 S_k(t \vee s)[p_1\pi_1(t)\pi_1(s)\pi_2(t \vee s) + p_2\pi_2(t)\pi_2(s)\pi_1(t \vee s)]$, $s \in [0, \tau_0]$, $v_k = \frac{b_k}{\sum_k b_k}$ and $t \vee s = \max(t,s)$. With $\alpha = \lim_{n\to\infty} q$, the form of the variance estimator follows from Cai and Zeng (2004) and Hu et al. (2014).

By Slutsky's theorem, with $\hat{\sigma}^2_{RCC} = \hat{\sigma}^2_{FC} + \hat{\sigma}^2_{CC}$, we have $\hat{\sigma}^2_{RCC} \to \sigma^2_{RCC}$ where $\sigma^2_{RCC} = \sigma^2_{FC} + \sigma^2_{CC}$ and $\sigma^2_{RCC} < \infty$.

5.2.3 Power and Sample Size for Recurrent Events Data in the Case Cohort Design

In this section, we derive a power and sample size formula for the recurrent events case-cohort design under some additional assumptions. The proportional means local alternative is

$$H_A : \Lambda_j(t) = \Lambda_0(t)\exp\{(-1)^{j-1}\psi(t)/2\} \text{ for } j = 1, 2,$$

with $\psi(t) \neq 0$ and $\psi(t) = o(1), t \in (0, \tau_0)$. The logrank statistic is also equivalent to the score test statistic under a Cox proportional hazards model when there is a single binary predictor. From Song et al. (2008), we have that the asymptotic mean of $\hat{L}$ under H_A is

$$\mu_1 = \int_0^{\tau_0} \psi(s)\{1 + \eta(s)\}\frac{p_1\pi_1(s)p_2\pi_2(s)}{p_1\pi_1(s) + p_2\pi_2(s)} d\Lambda_0(s), \tag{5.12}$$

and $\hat{\sigma}^2_{RCC} \to \sigma^2_{RCC}$ in probability as $n \to \infty$ still holds under H_A.

We make the following additional assumptions:

1. The distribution of censoring times C_{ij} are the same in both groups, so $\pi_1 = \pi_2 \equiv \pi_0$.

2. Baseline cumulative mean functions are continuous and local alternatives satisfy $\psi = \gamma, \gamma \in \mathbb{R}$.

3. $1 << d\bar{N}_j(\tau_0) << n_1 + n_2$.

4. There are no ties.

We additionally set $w(s) = 1$ throughout this section. We now use quantities to characterize event rate and within-subject correlations following the approach in Song et al. (2008). Define $D_{1g} = \int_0^{\tau_0} \pi_0(s) d\Lambda_0(s)$. To estimate D_{1g}, we take the geometric mean:

$$\hat{D}_{1g} = \sqrt{\frac{\sum_{i=1}^{n_1} \hat{\Lambda}_1(C_{i1})}{n_1} \times \frac{\sum_{i=1}^{n_2} \hat{\Lambda}_2(C_{i2})}{n_2}}, \tag{5.13}$$

where $\hat{\Lambda}_j(C_{ij}) = \int_0^{C_{ij}} \hat{Y}_{ij}(t) d\bar{N}_j(t) / \hat{Y}_j(t)$.

Then, assume that N_{ij}^0 is, conditional on a positive, latent real random variable w_{ij}, a nonstationary Poisson process with cumulative intensity function $w_{ij}\Lambda_j(t)$, where w_{ij} has mean 1 and variance σ_w^2. Assume that w_{ij} is independent of the censoring process. By Song et al. (2008), $\sigma_{FC}^2 = p_1 p_2 (D_{1a} + \sigma_w^2 D_2) + o(1)$, where $D_{1a} = \sum_{j=1}^2 \int_0^{\tau_0} p_{j'} \pi_0(s) d\Lambda_j(s), D_2 = \sum_{j=1}^2 p_{j'}$
$E[\{\int_0^{\tau_0} Y_{1j}(s) d\Lambda_j(s)\}^2]$, and $j' = 3 - j$. To estimate these quantities, we have

$$\hat{D}_{1a} = n^{-1} \sum_{j=1}^{2} \sum_{i=1}^{n_j} \hat{\Lambda}_j(C_{ij}), \tag{5.14}$$

$$\hat{D}_2 = n^{-1} \sum_{j=1}^{2} \sum_{i=1}^{n_j} \hat{\Lambda}_j^2(C_{ij}), \tag{5.15}$$

$$\hat{\sigma}_w^2 = (n^{-1} \sum_{j=1}^{2} \sum_{i=1}^{n_j} \frac{N_{ij}(C_{ij})\{N_{ij}(C_{ij}) - 1\}}{\hat{D}_2} - 1)^+, \quad x^+ = max\{x, 0\}. \tag{5.16}$$

Under H_A, the asymptotic mean of L_{RCC} can be approximated by $\gamma p_1 p_2 D_{1g}$ and $\hat{\sigma}_{FC}^2$ can

be approximated by $p_1p_2(D_{1a} + \sigma_w^2 D_2)$ by arguments in Song et al. (2008).

To simplify $\hat{\sigma}^2_{CC}$, we follow a similar argument as in Cai and Zeng (2004). Order failure times by $h_k = 1, 2, \ldots,$ for each event occurrence k. Then, denote the risk sizes in the full cohort by $m^j_{h_k}$ for group j, time h_k. The approximation is

$$\hat{\sigma}^2_{CC} \approx \frac{(1-q)}{nq} \sum_k \left[2 \sum_{h_k} p_1 p_2 \sum_{h_{k'} < h_k} \frac{1}{m^1_{h_{k'}} + m^2_{h_{k'}}} - \sum_{h_k} p_1 p_2 \frac{1}{m^1_{h_k} + m^2_{h_k}} \right] \tag{5.17}$$

$$= \frac{(1-q)}{nq} p_1 p_2 \sum_k \left[2 \sum_{h_k} \frac{d_{h_k} - 1/2}{m^1_{h_k} + m^2_{h_k}} \right], \tag{5.18}$$

where d_{h_k} is the number of failure times greater than or equal to the h_kth failure time. Furthermore, by assumption, there are few failure times, so $\sum_k (m^1_{h_k} + m^2_{h_k}) \approx n$. Then $\hat{\sigma}^2_{CC}$ can be approximated by $\frac{(1-q)}{n^2 q} p_1 p_2 (D_1 + D_2)^2$, where we define D_j to be the total number of events in group j.

By definition, we have

$$\frac{1}{n_j} \sum_{i=1}^{n_j} \hat{\Lambda}_j(C_{ij}) = \frac{1}{n_j} \sum_{i=1}^{n_j} \sum_{t_{lj} \le C_{ij}} \frac{dN_j(t_{lj})}{Y_j(t_{lj})} \tag{5.19}$$

$$= \frac{1}{n_j} \sum_{l=1}^{n_j} \frac{dN_j(t_{lj})}{Y_j(t_{lj})} \sum_{i=1}^{n_j} I(C_{ij} \ge t_{lj}) \tag{5.20}$$

$$= \frac{\hat{D}_j}{n_j}, \tag{5.21}$$

so $\hat{D}_{1g} = \sqrt{\frac{\sum_{i=1}^{n_1} \hat{\Lambda}_1(C_{i1})}{n_1} \times \frac{\sum_{i=1}^{n_2} \hat{\Lambda}_2(C_{i2})}{n_2}} = \sqrt{\frac{\hat{D}_1}{n_1} \times \frac{\hat{D}_2}{n_2}}$. If we assume γ, or the log-hazard ratio, is close to 0, and groups $j = 1, 2$ are roughly balanced ($p_1 \approx p_2$), then $\hat{D}_{1g} \approx \hat{D}_{1a} = \frac{\hat{D}_1 + \hat{D}_2}{n}$, since

$\frac{2}{n}\sqrt{D_1 D_2} \leq \frac{D_1+D_2}{n}$ and equality holds when $\gamma = 0$ and $n_1 = n_2$. If the above assumptions are violated, $\hat{D}_{1g}$ may be a more conservative estimate.

Therefore, $\hat{\sigma}^2_{CC}$ can be approximated by $\frac{(1-q)}{q} p_1 p_2 D^2_{1g}$, and $\hat{\sigma}^2_{RCC}$ can be approximated by $p_1 p_2 (D_{1a} + \sigma^2_w D_2) + \frac{(1-q)}{q} p_1 p_2 D^2_{1g}$.

Under H_A, for the alternative $\psi = \gamma$ and a two-sided test of level α, the power is calculated by $Pr\left[\hat{L}_{RCC,s} > z_{1-\alpha/2} | L_{RCC}\right]$, where $\hat{L}_{RCC,s}$ is the statistic $\hat{L}_{RCC}$ standardized by μ_1 and σ^2_{RCC}. So for a size n full cohort and subcohort sampling fraction q, power can be calculated by

$$\Phi\left(z_{\alpha/2} + \gamma \sqrt{\frac{nq p_1 p_2 \kappa}{q + (1-q)\kappa}} \right), \tag{5.22}$$

where $\Phi(\cdot)$ is the standard normal CDF, $\kappa \equiv \frac{D^2_{1g}}{(D_{1a} + \sigma^2_w D_2)}$, α is the significance level, γ is the log-hazard ratio of group 2 to group 1, and p_j is the proportion of group j in the full cohort.

Conversely, with n subjects in the full cohort, the average subcohort sampling fraction required for the alternative $\psi = \gamma$ for a two-sided test of level α and power β is

$$q = \frac{(z_{1-\alpha/2} + z_\beta)^2 \kappa}{n\gamma^2 p_1 p_2 \kappa - (z_{1-\alpha/2} + z_\beta)^2 (1-\kappa)} \tag{5.23}$$

$$= \frac{B\kappa}{n - B(1-\kappa)}, \tag{5.24}$$

where $\kappa \equiv \frac{D^2_{1g}}{(D_{1a} + \sigma^2_w D_2)}$ and $B = \frac{(z_{1-\alpha/2} + z_\beta)^2}{\gamma^2 p_1 p_2 \kappa}$.

Connecting to previous work, when $\sigma^2_w = 0$ and $q = 1$, formula 5.22 is equal to the power

formula in the univariate full cohort setting, $\Phi(z_{\alpha/2} + \gamma\sqrt{np_1p_2p_D})$, where p_D is the event rate in the full cohort (Schoenfeld, 1983). When $\sigma_w^2 = 0$ and $q < 1$, formula 5.22 is equal to the power formula in the univariate case-cohort design (Cai and Zeng, 2004), $\Phi(z_{\alpha/2} + \gamma\sqrt{\frac{\tilde{n}p_1p_2p_D}{q+(1-q)p_D}})$. Intuitively, M in the recurrent events setting characterizes event rate. When $q = 1$, under various σ_w^2, formula 5.22 equals that developed for full cohort analysis of recurrent events, $\Phi(z_{\alpha/2} + \gamma\sqrt{np_1p_2\kappa})$ (Song et al., 2008).

5.2.4 Using Subcohort Sampling Designs

Alternatively, in Chapter 3 of unpublished dissertation work (Xu and Gillen, 2026), we demonstrated that event-specific sampling may flexibly include more records from later event occurrences, which can be more efficient for estimating the marginal effect as well as the event-specific effect. Event-specific sampling refers to taking random samples from records from each event occurrence k with sampling fractions q_k, $k = 1, 2, \ldots, K$, then combining the samples to form a subcohort. We may perform:

1. Balanced sampling: when we increase or maximize the proportion of subcohort controls from later event occurrences, which can usually be under-represented in the subcohort. For example, when $K = 2$, let the case-to-control ratio in Event 2 be 1:m. m can represent a practical upper bound for case-cohort efficiency and we set $m \leq 3,\ m \in \mathbb{N}$. m is decreased until there can be a certain amount of Event 1 controls in the target subcohort size.

2. Events-proportional (EP) sampling: when we base proportions of controls sampled in each occurrence on the proportions of cases in each occurrence. This way, we maintain the same case-to-control ratio in each event occurrence.

We define a quantity termed "average subcohort sampling fraction" that may be used for each of the three sampling schemes as

$$q^* = \frac{\sum_{k=1}^{K} b_k q_k}{\sum_{k=1}^{K} b_k}, \tag{5.25}$$

which can be interpreted as the weighted average of records sampled into the subcohort. We may use this quantity in place of q as it is a measure common to each of the sampling schemes.

5.3 Simulation Study

Within-Subject Correlation	τ of Frailty Term	$\hat{\sigma}_w^2$
None	0	0
Mild	0.1, 0.2, 0.3	0
Moderate	0.5	0.09
Substantial	0.7	0.25

Table 5.1: Table of estimated values of $\hat{\sigma}_w^2$ through simulations with 1,000,000 subjects in the full cohort. Failure times are generated by $0.05 \times \exp\{\log(1.1)Z_{ij}(t) + b_{0i}\}$, $b_{01} \sim N(0, \tau)$, where τ is the standard deviation, and $C_{ij} \sim \text{Uniform}(0, 2.5)$.

In study planning stages, there are two general approaches to obtain an estimate for the required sample size. One approach is to perform a simulation study and calculate the minimum sample size needed for a desired level of empirical power obtained from the simulations. Another approach is to utilize a formula such as 5.22, which requires no simulations and is more straightforward. Both require some assumptions based on prior knowledge or expert opinions. In this section, we evaluate the performance of the power formula derived in Section 5.2.3 by comparing its estimated power to the empirical power obtained from simulations under similar assumptions.

In a hypothetical study assessing the effect of an expensive covariate which we binarize into

No Within-Subject Correlation ($\sigma_w^2 = 0$)				
$\exp(\gamma)$	$\exp(\gamma_{marginal})$	Design	Estimated Power	Empirical Power
1.00	1.00	Full Cohort		0.021
		Pooled	0.025	0.030
		Balanced	0.025	0.016
		EP	0.025	0.024
1.10	1.10	Full Cohort		0.219
		Pooled	0.189	0.186
		Balanced	0.189	0.181
		EP	0.189	0.177
1.15	1.15	Full Cohort		0.444
		Pooled	0.357	0.359
		Balanced	0.357	0.355
		EP	0.357	0.328
1.20	1.20	Full Cohort		0.674
		Pooled	0.554	0.539
		Balanced	0.554	0.545
		EP	0.554	0.561
1.25	1.25	Full Cohort		0.833
		Pooled	0.733	0.691
		Balanced	0.733	0.689
		EP	0.733	0.709

Table 5.2: Estimated power obtained from power formula and empirical power obtained from simulations under exponential failure times and uniform censoring times, with no ($\sigma_w^2 = 0$) within-subject correlations. $\exp(\gamma_{marginal})$ is obtained from 10 simulations of 500,000 subjects.

two groups ($j = 1, 2$), with maximum $K = 2$ events, we further assume:

1. Failure times in both groups follow exponential distributions. The hazard in group 1 is λ_1 and the hazard in group 2 is $\lambda_2 = \lambda_1 \exp(\gamma)$, i.e. assuming proportional hazards, the hazard ratio is $\exp(\gamma)$.

2. Uniform censoring in both groups, $C_{ij} \sim Uniform(0, b), j = 1, 2$.

Under these assumptions, we may further obtain the following simplifications:

Moderate Within-Subject Correlation ($\sigma_w^2 = 0.09$)				
$\exp(\gamma)$	$\exp(\gamma_{marginal})$	Design	Estimated Power	Empirical Power
1.00	1.00	Full Cohort		0.029
		Pooled	0.025	0.029
		Balanced	0.025	0.032
		EP	0.025	0.028
1.10	1.10	Full Cohort		0.254
		Pooled	0.188	0.203
		Balanced	0.188	0.179
		EP	0.188	0.182
1.15	1.14	Full Cohort		0.449
		Pooled	0.355	0.358
		Balanced	0.355	0.354
		EP	0.355	0.357
1.20	1.19	Full Cohort		0.720
		Pooled	0.551	0.591
		Balanced	0.551	0.542
		EP	0.551	0.578
1.25	1.24	Full Cohort		0.873
		Pooled	0.731	0.735
		Balanced	0.731	0.731
		EP	0.731	0.739

Table 5.3: Estimated power obtained from power formula and empirical power obtained from simulations under exponential failure times and uniform censoring times, with moderate ($\sigma_w^2 = 0.09$) within-subject correlations. $\exp(\gamma_{marginal})$ is obtained from 10 simulations of 500,000 subjects.

$$\hat{D}_{1a} \approx \big(\sum_{j=1}^{2} p_j(1 - \frac{1}{b\lambda_j}(1 - e^{-b\lambda_j}))\big)\big(\sum_{j=1}^{2} p_j(1 - \frac{1}{b\lambda_j}(1 - e^{-b\lambda_j})) + 1\big), \tag{5.26}$$

$$\hat{D}_{1g} \approx b\sqrt{p_1 p_2 \lambda_1 \lambda_2}, \tag{5.27}$$

$$\hat{D}_2 \approx \frac{b^2}{3}(p_1\lambda_1^2 + p_2\lambda_2^2). \tag{5.28}$$

To simulate data, we performed 1,000 simulations with 10,000 subjects each and generated an expensive exposure Z_{ij} from $Bernoulli(0.5)$, where 0 represents membership in group 1 and 1 represents membership in group 2. True failure times T_{ijk} were generated with hazard rate

$\exp(\gamma)$	$\exp(\gamma_{marginal})$	τ	σ_w^2	Estimated Power	Empirical Power
1.00	1.00			0.025	0.024
1.10	1.10			0.189	0.177
1.15	1.15	0	0	0.357	0.328
1.20	1.20			0.554	0.561
1.25	1.25			0.733	0.709
1.00	1.00			0.025	0.029
1.10	1.10			0.189	0.186
1.15	1.15	0.3	0	0.357	0.356
1.20	1.20			0.554	0.578
1.25	1.25			0.733	0.731
1.00	1.00			0.025	0.028
1.10	1.10			0.188	0.182
1.15	1.14	0.5	0.09	0.355	0.357
1.20	1.19			0.552	0.578
1.25	1.24			0.731	0.739
1.00	1.00			0.025	0.033
1.10	1.09			0.186	0.189
1.15	1.14	0.7	0.25	0.352	0.371
1.20	1.19			0.547	0.552
1.25	1.24			0.726	0.761

Table 5.4: Estimated power obtained from power formula and empirical power obtained from simulations under exponential failure times and uniform censoring times, using the events-proportional sampling scheme, with different levels of within-subject correlations. $\exp(\gamma_{marginal})$ is obtained from 10 simulations of 500,000 subjects.

$\lambda_1 \times \exp\{\beta_1 Z_{ij}(t) + b_{0i}\}$, where $\beta_1 = \gamma$. $b_{0i} \sim \text{Normal}(0,\ \tau^2)$ is a within-subject frailty term, where τ is a standard deviation varied to induce different levels of correlation. Censoring times are $C_{ij} \sim \text{Uniform}(0,\ b)$ for both groups, and $b = 2.5$. With equal proportions of subjects in the two groups in the full cohort at the beginning of the study, $p_1 = p_2 = 0.5$, we have $\hat{D}_{1a} \approx Pr[\text{Event 1}]^2 + Pr[\text{Event 1}] = 0.25\{\sum_j (1 - \frac{1}{b\lambda_j}(1 - e^{-b\lambda_j}))\}^2 + 0.5\{\sum_j (1 - \frac{1}{b\lambda_j}(1 - e^{-b\lambda_j}))\}, \hat{D}_{1g} \approx \frac{b}{2}\sqrt{\lambda_1\lambda_2}, \hat{D}_2 \approx \frac{b^2}{6}(\lambda_1^2 + \lambda_2^2)$.

To obtain estimated power from 5.22, we also need to make assumptions about the level of correlations within each subject, or the degree of over-dispersion of N_{ij}, which is reflected in

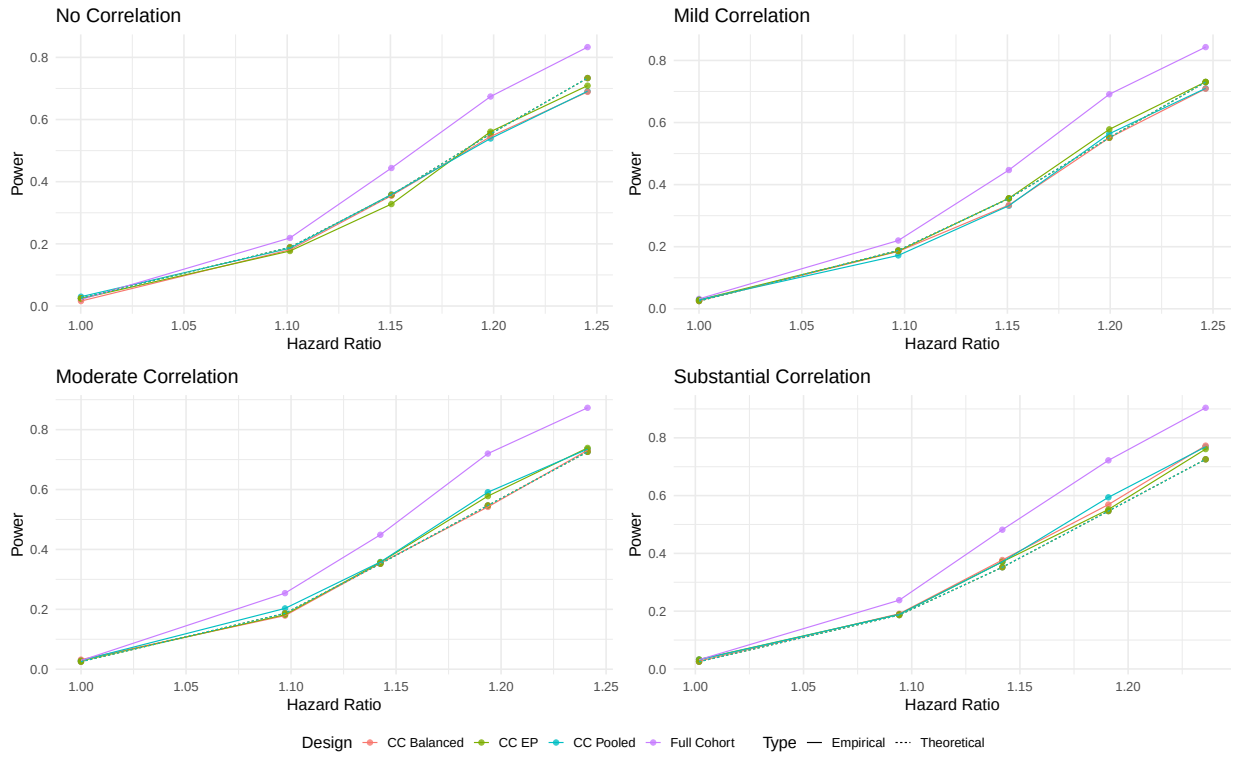


Figure 5.1: For $\lambda_1 = 0.05$, the comparison of estimated power obtained from power formula and empirical power obtained from simulations under exponential failure times and uniform censoring times, with no, mild, moderate, and substantial within-subject correlations. Dotted lines represent empirical power and solid lines represent estimated power. Full cohort and case-cohort (CC) designed with pooled, balanced, and events-proportional (EP) sampling schemes are in purple, blue, red, and green, respectively.

the size of σ_w^2. To do so, we performed simulations with 1,000,000 subjects in the full cohort and obtained values that $\hat{\sigma}_w^2$ is consistent for, under each size of τ of the frailty term b_{0i} when $\beta_1 = \log(1.1)$, shown in Table 5.1. We note that the estimated size of $\hat{\sigma}_w^2$ is reflective of the relative size of the within-subject correlations compared to other parameter values including λ_1, β_1, and b. These are general recommended values for these levels of correlations, but for specific values more tailored to the study at hand, researchers may obtain estimates through simulation.

We set λ_1 to be 0.05 and vary hazard ratio values to be 1, 1.1, 1.15, 1.2, and 1.25. Under these settings, the event rate is approximately 5% across both events and we take subcohorts

with $q^* = 0.2$ in each sampling scheme.

In Figure 5.1, we compare the performance of estimated power from the power formula in the case-cohort design with pooled, balanced, and events-proportional sampling schemes, as well as the empirical power achieved in the full cohort analysis for reference. Table 5.2 and 5.3 shows these results and the comparisons for mild and moderate within-subject correlations ($\sigma_w^2 = 0, 0.09$). Under the current settings, the power formula performs well in these sampling schemes and shows similar estimates compared to empirical results. Table 5.4 shows results across mild to substantial levels of within-subject correlations for the events-proportional sampling design. The power formula approximates the empirical power well across correlation levels, although it tends to be more conservative for higher correlations and larger hazard ratios. This conservatism arises because certain approximations used in the derivation are most accurate near the null, which is a well-known characteristic of logrank-based power formulas. We also note that in real-world datasets, correlations will likely be mild or moderate levels.

5.4 Application

We applied the proposed method to data from the Anti-Amyloid Treatment in Asymptomatic Alzheimer's (A4) Study and the Longitudinal Evaluation of Amyloid Risk and Neurodegeneration (LEARN) Study, to determine the minimal sample size needed to conduct possible future analyses (ACTC; Sperling et al., 2020). The A4 Study was a Phase III trial that tested the experimental drug solanezumab in subjects with preclinical AD, or those who had evidence of brain amyloid pathology on Positron Emission Tomography (PET) imaging at screening. The LEARN Study enrolled subjects who screen-failed for the A4 Study for not showing elevated amyloid levels. The two studies had a combined 1709 subjects enrolled.

We defined recurrent events based on the Preclinical Alzheimer Cognitive Composite (PACC),

which is a composite of 4 measures established to be sensitive to decline specific to the amyloid-positive and cognitively normal population, as opposed to decline that is associated with aging. The PACC was regularly measured in the studies at screening, Week 0, then at regular 6-week intervals after Week 0. We defined the first event as a 1.5-standard deviation (SD) decrease in standardized PACC compared to baseline (Week 0) and the second event as a 3-SD decrease compared to baseline. The longest follow-up observed within the trial period, not counting the open-label period, was 87 months.

Using the proposed sample size formula (5.24) and approximations as shown in Table 5.1, we derived the required subcohort size for a test size of 0.05 (two-sided) and power of 70% and 80%, assuming censoring time is Uniform$(0, 87)$ to detect hazard ratio $\exp(\gamma) = 1.15, 1.2, 1.3$ with various correlation levels, shown in Table 5.5. For example, to reach 70% power to detect a 1.20 hazard ratio, and assuming there are mild correlations, we plan to sample 667 subjects into the subcohort. By setting q^* to be 0.390, we may employ any of the proposed sampling schemes to obtain the subcohort. We note that when q^* is larger than 1, then it is not possible to reach the desired level of power even with full cohort data, for example in the scenario of desiring 80% power to detect a 1.15 hazard ratio and assuming $\sigma_w^2 = 0.25$. If covariates are instead event-specific, i.e. measurements are taken at each new event time, then we may also calculate the required subcohort size $\tilde{n}$ multiplied by the expected number of measurements per subject to obtain the number of samples needed to be processed.

5.5 Discussion

In this work, we developed a logrank-type test statistic and derived power and sample size formulas for planning recurrent events analysis that utilize case-cohort designs. When expensive covariates such as a reserve of biological samples exist, it is important to maximize the number of studies we may conduct by preserving available samples, and efficient sampling designs such as the case-cohort design can achieve this goal. To plan for a case-cohort

Power	Hazard Ratio	σ_w^2	$\tilde{n}$	q^*
		0	1056	0.618
	1.15	0.09	1142	0.668
		0.25	1334	0.780
		0	667	0.390
70%	1.20	0.09	700	0.410
		0.25	768	0.449
		0	461	0.270
	1.25	0.09	477	0.280
		0.25	508	0.297
		0	1286	0.752
	1.15	0.09	1415	0.828
		0.25	1722†	1.007†
		0	824	0.482
80%	1.20	0.09	876	0.513
		0.25	985	0.576
		0	575	0.336
	1.25	0.09	600	0.351
		0.25	650	0.380

Table 5.5: For the A4 and LEARN studies data, assuming that failure times follow exponential distributions, under assumed hazard ratios, censoring, and correlations (σ_w^2), the minimum required subcohort size ($\tilde{n}$) and average subcohort sampling fraction (q^*) to achieve 70% or 80% power, based on the proposed sample size formula. †When q^* is larger than 1, the desired power under these assumptions is not reachable even with full cohort data.

design, the proposed logrank-type test statistic and derived formula can be used to calculate the minimum required subcohort size for a desired level of power. We demonstrated the formula's finite sample performance through a simulation study and derived the subcohort size required for a hypothetical study based on the A4 and LEARN studies.

A limitation of the proposed method is that we place several assumptions on failure and censoring time, including distributional assumptions, to obtain a sample size calculation. It is, however, common in the design stage to place assumptions and there generally exists a trade-off between ease of implementation and assumptions. By using the proposed formula, we would not need to perform a simulation study, but if circumstances are very specific or complex, it may be helpful to instead obtain empirical estimates from a simulation. Moreover, the term for incorporating within-subject correlations, σ_w^2, may be more challenging to

approximate. Generally, we suggest assuming milder levels of correlation, which is common in real datasets.

In Song et al. (2008), the authors consider adjusting for covariates $V_{ij}(t)$ and incorporating a function $h(V_{ij}(t);\theta)$, which builds on work in covariate-adjusted logrank tests and may be more efficient when the relationship is correctly specified. In the design phase, it may be difficult to correctly identify relevant covariates beyond the predictor of interest and it may be unfeasible to propose sample size estimates based on extra covariate information in a study proposal, especially when proposing a case-cohort study, so we do not incorporate covariate-adjustment in this work.

To perform subcohort sampling, another approach is to sample on the subject level, and when covariates do not vary between event occurrences, we may include all events $k = 1, \ldots, K$ for individuals sampled into the subcohort to increase efficiency, needing no extra processing. If we take $\tilde{q}$ as the subcohort sampling proportion, then we obtain a subcohort with $b_1\tilde{q}$ individuals. However, the later controls that are supplemented in will need to be weighted by the inverse probability weight of sampling, as they are no longer a simple random sample. We performed an empirical assessment of the power in this case and the current formula still performs relatively well to estimate the empirical power. A future direction is to formally develop this estimator.

As another future direction, we plan to develop approximations for other failure time distributions including Weibull or lognormal distributions, for ease of implementation for a wider audience. Furthermore, we plan to extend the current method to use approximations for accommodating different baseline hazard functions for different event occurrences. Power and sample size calculations under differential baseline hazards can be obtained if we can reasonably estimate D_{1a}, D_{1g}, D_2, but they may be more difficult to approximate beforehand when event rates are assumed to differ across event occurrences.

Chapter 6

Conclusion and Future Directions

In this dissertation, we made methodological developments in the univariate and multivariate case-cohort designs with three main contributions. First, we utilized a case-cohort design that performed stratified sampling on a proposed effect modifier of the predictor of interest, which is shown to be more efficient for assessing the interaction effect while still maintaining valid inference across the full study population. Power and efficiency gains were found to be greatest in assessing larger effect sizes, smaller subcohort sizes, and rarer subpopulations. Extending to recurrent events analyses, we provided an estimator under recurrent event settings that allowed for event-specific covariate values and event-specific effects estimation, and demonstrated its consistency and asymptotic normality. Furthermore, we developed three sampling schemes that can flexibly include controls into the subcohort for various goals: simplicity, maximizing efficiency for assessing effects in later events, or a balance of the two. Lastly, we proposed a logrank-type statistic for recurrent events case-cohort designs and derived power and sample size formula to assist in study planning.

The case-cohort design can be greatly advantageous for biomarker discovery in AD research due to its ability to save time, cost, and valuable biological samples. It is, however, important

to further develop sampling that can ensure statistical power to assess, broadly, subgroup differences. This may manifest in ensuring power to detect differential effects within covariate subpopulations or event occurrence strata, and in design phase considerations. The methods in this dissertation aim to do the above, to alleviate challenges in study design and recruitment, and to better understand and minimize health disparities. The goals of this dissertation are aligned with goals in useful biomarker discovery in AD research, but can also be employed in any study and field that would benefit from resource-saving and efficient study designs.

For extensions of Chapter 3, we consider further increasing model efficiency with time-varying weights rather than time-invariant weights, which are calculated using the number of individuals in the full cohort and subcohort in each stratum at each event time rather than at the beginning of the study. Kulich and Lin (2004) demonstrated that incorporating cohort information using time-dependent weights will further increase efficiency of estimation. We plan to investigate potential power gains in implementing time-dependent weights when assessing interaction effects as a future direction.

To expand on the work in Chapters 4 and 5, when performing subcohort sampling for recurrent events analyses on the record level, it would be most ideal to sample the lowest number of records and hence the least amount of samples needing to be processed. Under the univariate setting, Hu et al. (2014) derived sample size formulas for case-cohort designs with an "optimal design", meaning if restricted by resources to only sample $\tilde{n}$ subjects, this would be an optimal design strategy to provide the highest power. Let there be $l = 1, \ldots, L$ strata. With n total subjects, n_l subjects in each stratum, p_{lj} the proportion of subjects in group j in stratum l, p_{Dl} the failure proportion in stratum l, $v_l = n_l/n$, then the subcohort size to provide the highest power is

$$\tilde{n} = \left[\frac{n(\sum_{l=1}^{L} \sqrt{p_{l1}p_{l2}/(1-p_{Dl}/2)}p_{Dl}v_l)^2}{B_2^2 - \sum_{l=1}^{L}(p_{l1}p_{l2}p_{Dl}v_l(1-p_{Dl}/(1-p_{Dl}/2)))} \right], \tag{6.1}$$

where $B_2 = \frac{n^{1/2}\theta \sum_{l=1}^{L}(p_{l1}p_{l2}p_{Dl}v_l)}{z_{1-\alpha/2}+z_\beta}$, and this optimal power can be calculated through

$$\Phi\left(Z_{\alpha/2} + \frac{\tilde{n}^{1/2}|\theta| \sum_{l=1}^{L}(p_{l1}p_{l2}p_{Dl}v_l)}{\sqrt{\tilde{n}/nP + Q^2}} \right), \tag{6.2}$$

where $P = \sum_{l=1}^{L}(p_{l1}p_{l2}p_{Dl}v_l(1-p_{Dl}/(1-p_{Dl}/2))), Q = (\sum_{l=1}^{L} \sqrt{p_{l1}p_{l2}/(1-p_{Dl}/2)} \times p_{Dl}v_l))$. The optimal sampling proportion in each stratum can also be obtained. Under the recurrent events setting, we may consider each event occurrence as a stratum. We plan to derive a similar design with optimal sampling proportions, subcohort size, and power formula as an extension to Chapter 5. This sampling scheme may then be applied to our work in Chapter 4, to assess optimal efficiency gains when performing inference for later event occurrences.

In Chapter 5, we would like to extend the method to accommodate event-specific covariates. We assume in this work that N_{ij}^0 is a nonstationary Poisson process with cumulative intensity function $w_{ij}\Lambda_j(t)$, where w_{ij} has mean 1 and variance σ_w^2. With group membership j indicating different biomarker levels, if a subject's covariate value changes across different event occurrences, the group membership of their counting process will change. This may require characterization of the correlations between N_{i1}^0 and N_{i2}^0 which is not currently in the model.

Moreover, while the current approximations in Chapter 5 that operate under exponential distribution assumptions do not allow for baseline hazard functions to be different across event occurrences, the approximations may still work well to estimate results under varying

baseline hazards. In Figure 6.1, we simulated data under the same settings as in the simulation study of Chapter 5, but where true failure times T_{ijk} were generated with hazard rate $\lambda_1 \times \exp\{\beta_1 Z_{ij}(t) + I(k = 2)\beta_2 + b_{0i}\}$, where the added baseline hazard for Event 2 were $\beta_2 = \log(1.1), \log(1.3)$, respectively. These preliminary empirical results indicate that the power formula may still provide a reasonable estimate. Our next steps will be to formally derive a power formula under these settings.

Finally, this work focuses on the case-cohort design, which may be interpreted as a retrospective study design that is conducted after a cohort study is concluded, when events have been observed and expensive samples have been collected. The idea of saving valuable samples and the time and cost of enrolling study participants is also a common goal in prospective study designs, including clinical trial designs. We consider developing group sequential methods in clinical trials to derive gains in power and efficiency when switching randomization strategies at interim analyses. Biomarker stratified designs are designs that randomize all patients to one of the treatment therapies based on a biomarker. Enriched biomarker stratified designs keep all biomarker positives and only a proportion of biomarker negative and therefore retain more information on the treatment effect (Wang et al., 2018). Clinical trials may start out with an emphasis on oversample subgroups to ensure power in assessing subgroup effects. If a pre-specified test of futility at the interim analysis yields that there is no significant effect modification, trialists may consider a randomly sampled scheme to overcome recruitment costs and challenges. We aim to provide a framework for such procedures and derive the implications in statistical power and efficiency.

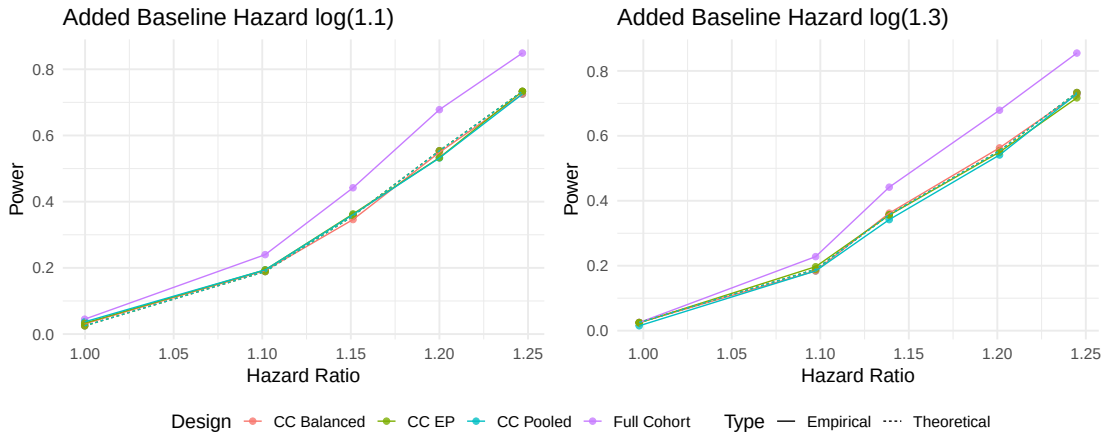


Figure 6.1: Plots comparing estimated and empirical power when the baseline hazard function increase from Event 1 to Event 2, β_3, are $\log(1.1), \log(1.3)$, under mild correlations and the same settings as the simulation study in Chapter 5.

Bibliography

ACTC. The A4 Study and LEARN Study. URL a4studydata.org.

A&D. 2025 Alzheimer's disease facts and figures. *Alzheimer's & Dementia*, 21(4): e70235, Apr. 2025. ISSN 1552-5260, 1552-5279. doi: 10.1002/alz.70235. URL https://alz-journals.onlinelibrary.wiley.com/doi/10.1002/alz.70235.

L. D. Amorim and J. Cai. Modelling recurrent events: a tutorial for analysis in epidemiology. *International journal of epidemiology*, 44(1):324–333, 2015.

P. K. Andersen and R. D. Gill. Cox's Regression Model for Counting Processes: A Large Sample Study. *The Annals of Statistics*, 10(4), Dec. 1982. ISSN 0090-5364. doi: 10.1214/aos/1176345976.

W. E. Barlow. Robust Variance Estimation for the Case-Cohort Design. *Biometrics*, 50(4): 1064, Dec. 1994. ISSN 0006-341X. doi: 10.2307/2533444. URL https://www.jstor.org/stable/2533444?origin=crossref.

S. Bennett. Analysis of survival data by the proportional odds model. *Statistics in medicine*, 2(2):273–277, 1983a.

S. Bennett. Log-logistic regression models for survival data. *Journal of the Royal Statistical Society Series C: Applied Statistics*, 32(2):165–171, 1983b.

M. M. Blazel, K. K. Lazar, C. A. Van Hulle, Y. Ma, A. Cole, A. Spalitta, N. Davenport-Sis, B. B. Bendlin, M. Wahoske, C. Illingworth, C. E. Gleason, D. F. Edwards, H. Blazel, S. Asthana, S. C. Johnson, and C. M. Carlsson. Factors Associated with Lumbar Puncture Participation in Alzheimer's Disease Research. *Journal of Alzheimer's Disease*, 77(4): 1559–1567, Oct. 2020. ISSN 13872877, 18758908. doi: 10.3233/JAD-200394. URL https://journals.sagepub.com/doi/full/10.3233/JAD-200394.

K. Blennow, H. Hampel, M. Weiner, and H. Zetterberg. Cerebrospinal fluid and plasma biomarkers in Alzheimer disease. *Nature Reviews Neurology*, 6(3):131–144, Mar. 2010. ISSN 1759-4758, 1759-4766. doi: 10.1038/nrneurol.2010.4. URL https://www.nature.com/articles/nrneurol.2010.4.

O. Borgan, B. Langholz, S. O. Samuelsen, L. Goldstein, and J. Pogoda. Exposure Stratified Case-Cohort Designs. *Lifetime Data Analysis*, 6(1):39–58, Mar. 2000. ISSN 1380-7870,

1572-9249. doi: 10.1023/A:1009661900674. URL https://link.springer.com/10.1023/A:1009661900674.

A. P. Boyd, J. M. Kittelson, and D. L. Gillen. Estimation of treatment effect under non-proportional hazards and conditionally independent censoring. *Statistics in Medicine*, 31 (28):3504–3515, Dec. 2012. ISSN 0277-6715, 1097-0258. doi: 10.1002/sim.5440. URL https://onlinelibrary.wiley.com/doi/10.1002/sim.5440.

N. Breslow. Covariance Analysis of Censored Survival Data. *Biometrics*, 30(1):89, Mar. 1974. ISSN 0006341X. doi: 10.2307/2529620. URL https://www.jstor.org/stable/2529620?origin=crossref.

N. E. Breslow and J. A. Wellner. Weighted Likelihood for Semiparametric Models and Two-phase Stratified Samples, with Application to Cox Regression. *Scandinavian Journal of Statistics*, 34(1):86–102, Mar. 2007. ISSN 0303-6898, 1467-9469. doi: 10.1111/j.1467-9469.2006.00523.x. URL https://onlinelibrary.wiley.com/doi/10.1111/j.1467-9469.2006.00523.x.

N. E. Breslow, T. Lumley, C. M. Ballantyne, L. E. Chambless, and M. Kulich. Improved Horvitz–Thompson estimation of model parameters from two-phase stratified samples: applications in epidemiology. *Statistics in biosciences*, 1:32–49, 2009a.

N. E. Breslow, T. Lumley, C. M. Ballantyne, L. E. Chambless, and M. Kulich. Using the whole cohort in the analysis of case-cohort data. *American journal of epidemiology*, 169 (11):1398–1405, 2009b.

J. Cai and R. L. Prentice. Estimating equations for hazard ratio parameters based on correlated failure time data. *Biometrika*, 82(1):151–164, 1995.

J. Cai and R. L. Prentice. Regression estimation using multivariate failure time data and a common baseline hazard function model. *Lifetime Data Analysis*, 3:197–213, 1997.

J. Cai and D. Zeng. Sample Size/Power Calculation for Case–Cohort Studies. *Biometrics*, 60(4):1015–1024, Dec. 2004. ISSN 0006-341X, 1541-0420. doi: 10.1111/j.0006-341x.2004.00257.x. URL https://academic.oup.com/biometrics/article/60/4/1015-1024/7286732.

M.-H. Chen, J. G. Ibrahim, and D. Sinha. Bayesian inference for multivariate survival data with a cure fraction. *Journal of Multivariate Analysis*, 80(1):101–126, 2002.

Y. Q. Chen. Semiparametric regression analysis on longitudinal pattern of recurrent gap times. *Biostatistics*, 5(2):277–290, Apr. 2004. ISSN 1465-4644, 1468-4357. doi: 10.1093/biostatistics/5.2.277. URL https://academic.oup.com/biostatistics/article-lookup/doi/10.1093/biostatistics/5.2.277.

S. Cheng, L. J. Wei, and Z. Ying. Analysis of transformation models with censored data. *Biometrika*, 82(4):835–845, 1995.

S. H. Chiou, S. Kang, and J. Yan. Semiparametric accelerated failure time modeling for clustered failure times from stratified sampling. *Journal of the American Statistical Association*, 110(510):621–629, 2015.

K. Choi and S. Kang. Efficient case-cohort design using balanced sampling. *Journal of Statistical Computation and Simulation*, 95(8):1682–1708, May 2025. ISSN 0094-9655, 1563-5163. doi: 10.1080/00949655.2025.2469907. URL https://www.tandfonline.com/doi/full/10.1080/00949655.2025.2469907.

J. Cologne, D. L. Preston, K. Imai, M. Misumi, K. Yoshida, T. Hayashi, and K. Nakachi. Conventional case–cohort design and analysis for studies of interaction. *International journal of epidemiology*, 41(4):1174–1186, 2012.

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

D. R. Cox. Partial likelihood. *Biometrika*, 62(2):269–276, Aug. 1975. ISSN 0006-3444, 1464-3510. doi: 10.1093/biomet/62.2.269. URL https://academic.oup.com/biomet/article-lookup/doi/10.1093/biomet/62.2.269.

M. C. Donohue, R. A. Sperling, D. P. Salmon, D. M. Rentz, R. Raman, R. G. Thomas, M. Weiner, and P. S. Aisen. The Preclinical Alzheimer Cognitive Composite: Measuring Amyloid-Related Decline. *JAMA Neurology*, 71(8):961, Aug. 2014. ISSN 2168-6149. doi: 10.1001/jamaneurol.2014.803. URL http://archneur.jamanetwork.com/article.aspx?doi=10.1001/jamaneurol.2014.803.

T. R. Fleming and D. P. Harrington. A class of hypothesis tests for one and two sample censored survival data. *Communications in Statistics - Theory and Methods*, 10(8):763–794, Jan. 1981. ISSN 0361-0926, 1532-415X. doi: 10.1080/03610928108828073. URL https://www.tandfonline.com/doi/full/10.1080/03610928108828073.

S. Franzen, J. E. Smith, E. Van Den Berg, M. Rivera Mindt, R. L. Van Bruchem-Visser, E. L. Abner, L. S. Schneider, N. D. Prins, G. M. Babulal, and J. M. Papma. Diversity in Alzheimer's disease drug trials: The importance of eligibility criteria. *Alzheimer's & Dementia*, 18(4):810–823, Apr. 2022. ISSN 1552-5260, 1552-5279. doi: 10.1002/alz.12433. URL https://alz-journals.onlinelibrary.wiley.com/doi/10.1002/alz.12433.

L. S. Freedman. Tables of the number of patients required in clinical trials using the logrank test. *Statistics in Medicine*, 1(2):121–129, Jan. 1982. ISSN 0277-6715, 1097-0258. doi: 10.1002/sim.4780010204. URL https://onlinelibrary.wiley.com/doi/10.1002/sim.4780010204.

R. E. Gangnon and M. R. Kosorok. Sample-size formula for clustered survival data using weighted log-rank statistics. *Biometrika*, 91(2):263–275, June 2004. ISSN 0006-3444, 1464-3510. doi: 10.1093/biomet/91.2.263. URL https://academic.oup.com/biomet/article-lookup/doi/10.1093/biomet/91.2.263.

P. Georges, A.-G. Lamy, E. Nicolas, G. Quibel, and T. Roncalli. Multivariate survival modelling: a unified approach with copulas. *Available at SSRN 1032559*, 2001.

M. Gorfine and L. Hsu. Frailty-based competing risks model for multivariate survival data. *Biometrics*, 67(2):415–426, 2011.

S. Greenland, J. Pearl, and J. M. Robins. Confounding and Collapsibility in Causal Inference. *Statistical Science*, 14(1), Feb. 1999. ISSN 0883-4237. doi: 10.1214/ss/1009211805. URL https://projecteuclid.org/journals/statistical-science/volume-14/issue-1/Confounding-and-Collapsibility-in-Causal-Inference/10.1214/ss/1009211805.full.

J. D. Grill, R. Raman, K. Ernstrom, P. Aisen, and J. Karlawish. Effect of study partner on the conduct of Alzheimer disease clinical trials. *Neurology*, 80(3):282–288, Jan. 2013. ISSN 0028-3878, 1526-632X. doi: 10.1212/WNL.0b013e31827debfe. URL https://www.neurology.org/doi/10.1212/WNL.0b013e31827debfe.

J. D. Grill, C. Flournoy, S. Dhadda, K. Ernstrom, R. Sperling, D. Molina-Henry, K. Tranotti, R. Harris, M. Kanekiyo, M. Gee, M. Irizarry, L. Kramer, P. Aisen, and R. Raman. Eligibility Rates among Racially and Ethnically Diverse US Participants in Phase 2 and Phase 3 Placebo-Controlled, Double-Blind, Randomized Trials of Lecanemab and Elenbecestat in Early Alzheimer Disease. *Annals of Neurology*, 95(2):288–298, Feb. 2024. ISSN 0364-5134, 1531-8249. doi: 10.1002/ana.26819. URL https://onlinelibrary.wiley.com/doi/10.1002/ana.26819.

S. Gunes, Y. Aizawa, T. Sugashi, M. Sugimoto, and P. P. Rodrigues. Biomarkers for Alzheimer's disease in the current state: a narrative review. *International journal of molecular sciences*, 23(9):4962, 2022.

D. G. Horvitz and D. J. Thompson. A generalization of sampling without replacement from a finite universe. *Journal of the American statistical Association*, 47(260):663–685, 1952.

L. Hsu, M. Gorfine, and K. Malone. On robustness of marginal regression coefficient estimates and hazard functions in multivariate survival analysis of family data when the frailty distribution is mis-specified. *Statistics in Medicine*, 26(25):4657–4678, 2007.

W. Hu, J. Cai, and D. Zeng. Sample size/power calculation for stratified case-cohort design. *Statistics in Medicine*, 33(23):3973–3985, Oct. 2014. ISSN 0277-6715. doi: 10.1002/sim.6215. URL https://onlinelibrary.wiley.com/doi/10.1002/sim.6215.

Y. Huang. The Two-Sample Problem with Induced Dependent Censorship. *Biometrics*, 55 (4):1108–1113, Dec. 1999. ISSN 0006-341X, 1541-0420. doi: 10.1111/j.0006-341X.1999.01108.x. URL https://academic.oup.com/biometrics/article/55/4/1108-1113/7263306.

M. Iwagami, N. Qizilbash, J. Gregson, I. Douglas, M. Johnson, N. Pearce, S. Evans, and S. Pocock. Blood cholesterol and risk of dementia in more than 1.8 million people over

two decades: a retrospective cohort study. *The Lancet Healthy Longevity*, 2(8):e498–e506, Aug. 2021. ISSN 26667568. doi: 10.1016/S2666-7568(21)00150-1. URL https://linkinghub.elsevier.com/retrieve/pii/S2666756821001501.

J. Kalbfleisch and J. Lawless. Likelihood analysis of multi-state models for disease incidence and mortality. *Statistics in medicine*, 7(1-2):149–160, 1988.

J. D. Kalbfleisch and R. L. Prentice. *The statistical analysis of failure time data*. John Wiley & Sons, 2002.

M. Kane. Lecanemab Therapy and APOE Genotype. *Medical Genetics Summaries*, 2024. URL https://www.ncbi.nlm.nih.gov/books/NBK605938/.

S. Kang and J. Cai. Marginal hazards model for case-cohort studies with multiple disease outcomes. *Biometrika*, 96(4):887–901, Dec. 2009. ISSN 0006-3444, 1464-3510. doi: 10.1093/biomet/asp059. URL https://academic.oup.com/biomet/article-lookup/doi/10.1093/biomet/asp059.

E. L. Kaplan and P. Meier. Nonparametric estimation from incomplete observations. *Journal of the American statistical association*, 53(282):457–481, 1958.

R. S. Kim. A new comparison of nested case–control and case–cohort designs and methods. *European Journal of Epidemiology*, 30(3):197–207, Mar. 2015. ISSN 0393-2990, 1573-7284. doi: 10.1007/s10654-014-9974-4. URL http://link.springer.com/10.1007/s10654-014-9974-4.

I. M. Kopits, C. Chen, J. S. Roberts, W. Uhlmann, and R. C. Green. Willingness to pay for genetic testing for Alzheimer's disease: a measure of personal utility. *Genetic testing and molecular biomarkers*, 15(12):871–875, 2011.

M. Kulich and D. Lin. Additive hazards regression for case-cohort studies. *Biometrika*, 87 (1):73–87, 2000.

M. Kulich and D. Y. Lin. Improving the Efficiency of Relative-Risk Estimation in Case-Cohort Studies. *Journal of the American Statistical Association*, 99(467):832–844, Sept. 2004. ISSN 0162-1459, 1537-274X. doi: 10.1198/016214504000000584. URL http://www.tandfonline.com/doi/abs/10.1198/016214504000000584.

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

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

D. Lin. On fitting Cox's proportional hazards models to survey data. *Biometrika*, 87(1): 37–47, Mar. 2000. ISSN 0006-3444, 1464-3510. doi: 10.1093/biomet/87.1.37. URL https://academic.oup.com/biomet/article-lookup/doi/10.1093/biomet/87.1.37.

D. Y. Lin. Cox regression analysis of multivariate failure time data: The marginal approach. *Statistics in Medicine*, 13(21):2233–2247, Nov. 1994. ISSN 0277-6715, 1097-0258. doi: 10.1002/sim.4780132105. URL https://onlinelibrary.wiley.com/doi/10.1002/sim.4780132105.

D. Y. Lin and Z. Ying. Cox Regression with Incomplete Covariate Measurements. *Journal of the American Statistical Association*, 88(424):1341–1349, Dec. 1993. ISSN 0162-1459, 1537-274X. doi: 10.1080/01621459.1993.10476416. URL http://www.tandfonline.com/doi/abs/10.1080/01621459.1993.10476416.

F. Liya, Z. Yang, J. Zhang, A. Long, and Y. Zhou. Generalized estimating equations for analyzing multivariate survival data. *Communications in Statistics-Simulation and Computation*, 50(10):3060–3068, 2021.

T. Lumley, P. A. Shaw, and J. Y. Dai. Connections between survey calibration estimators and semiparametric models for incomplete data. *International Statistical Review*, 79(2): 200–220, 2011.

P. Maitra, L. D. A. F. Amorim, and J. Cai. Multiplicative rates model for recurrent events in case-cohort studies. *Lifetime Data Analysis*, 26(1):134–157, Jan. 2020. ISSN 1380-7870, 1572-9249. doi: 10.1007/s10985-019-09466-0. URL http://link.springer.com/10.1007/s10985-019-09466-0.

N. Mantel. Evaluation of survival data and two new rank order statistics arising in its consideration. *Cancer Chemother Rep*, 1966.

E. R. Marcantonio, J. Aneja, R. N. Jones, D. C. Alsop, T. G. Fong, G. J. Crosby, D. J. Culley, L. A. Cupples, and S. K. Inouye. Maximizing Clinical Research Participation in Vulnerable Older Persons: Identification of Barriers and Motivators. *Journal of the American Geriatrics Society*, 56(8):1522–1527, Aug. 2008. ISSN 0002-8614, 1532-5415. doi: 10.1111/j.1532-5415.2008.01829.x. URL https://agsjournals.onlinelibrary.wiley.com/doi/10.1111/j.1532-5415.2008.01829.x.

K. A. Matthews, W. Xu, A. H. Gaglioti, J. B. Holt, J. B. Croft, D. Mack, and L. C. McGuire. Racial and ethnic estimates of Alzheimer's disease and related dementias in the United States (2015–2060) in adults aged\geq 65 years. *Alzheimer's & Dementia*, 15(1):17–24, 2019.

M. M. Mielke, P. P. Zandi, M. Sjögren, D. Gustafson, S. Östling, B. Steen, and I. Skoog. High total cholesterol levels in late life associated with a reduced risk of dementia. *Neurology*, 64(10):1689–1695, May 2005. ISSN 0028-3878, 1526-632X. doi: 10.1212/01.WNL.0000161870.78572.A5. URL https://www.neurology.org/doi/10.1212/01.WNL.0000161870.78572.A5.

S. Murphy, A. Rossini, and A. W. van der Vaart. Maximum likelihood estimation in the proportional odds model. *Journal of the American Statistical Association*, 92(439):968–976, 1997.

B. Nan. Efficient estimation for case-cohort studies. *Canadian Journal of Statistics*, 32(4): 403–419, Dec. 2004. ISSN 0319-5724, 1708-945X. doi: 10.2307/3316024. URL https://onlinelibrary.wiley.com/doi/10.2307/3316024.

B. Nan and J. A. Wellner. A general semiparametric Z-estimation approach for case-cohort studies. *Stat Sin*, 2013. URL https://pmc.ncbi.nlm.nih.gov/articles/PMC3904394/.

B. Nan, M. Yu, and J. D. Kalbfleisch. Censored linear regression for case-cohort studies. *Biometrika*, 93(4):747–762, 2006.

H. Noma and S. Tanaka. Analysis of case-cohort designs with binary outcomes: improving efficiency using whole-cohort auxiliary information. *Statistical methods in medical research*, 26(2):691–706, 2017.

M. M. Nuño and D. L. Gillen. Alternative Sampling Designs for Time-to-Event Data With Applications to Biomarker Discovery in Alzheimer's Disease. In *Handbook of statistics*, volume 36, pages 105–166. Elsevier, 2017.

M. Othus and Y. Li. A gaussian copula model for multivariate survival data. *Statistics in biosciences*, 2:154–179, 2010.

K. M. O'Brien, K. G. Lawrence, and A. P. Keil. The Case for Case–Cohort: An Applied Epidemiologist's Guide to Reframing Case–Cohort Studies to Improve Usability and Flexibility. *Epidemiology*, 33(3):354–361, 2022.

Y. Peng, J. M. Taylor, and B. Yu. A marginal regression model for multivariate failure time data with a surviving fraction. *Lifetime data analysis*, 13:351–369, 2007.

R. C. Petersen, P. S. Aisen, L. A. Beckett, M. C. Donohue, A. C. Gamst, D. J. Harvey, C. R. Jack, W. J. Jagust, L. M. Shaw, A. W. Toga, J. Q. Trojanowski, and M. W. Weiner. Alzheimer's Disease Neuroimaging Initiative (ADNI): Clinical characterization. *Neurology*, 74(3):201–209, Jan. 2010. ISSN 0028-3878, 1526-632X. doi: 10.1212/WNL.0b013e3181cb3e25. URL https://www.neurology.org/doi/10.1212/WNL.0b013e3181cb3e25.

R. L. Prentice. A case-cohort design for epidemiologic cohort studies and disease prevention trials. *Biometrika*, 73(1):1–11, 1986.

R. L. Prentice and L. Hsu. Regression on hazard ratios and cross ratios in multivariate failure time analysis. *Biometrika*, 84(2):349–363, 1997.

R. L. Prentice, B. J. Williams, and A. V. Peterson. On the regression analysis of multivariate failure time data. *Biometrika*, 68(2):373–379, 1981. ISSN 0006-3444, 1464-3510. doi: 10.1093/biomet/68.2.373. URL https://academic.oup.com/biomet/article-lookup/doi/10.1093/biomet/68.2.373.

P. Rawat, U. Sehar, J. Bisht, A. Selman, J. Culberson, and P. H. Reddy. Phosphorylated Tau in Alzheimer's Disease and Other Tauopathies. *International Journal of Molecular Sciences*, 23(21):12841, Oct. 2022. ISSN 1422-0067. doi: 10.3390/ijms232112841. URL https://www.mdpi.com/1422-0067/23/21/12841.

S. Samuelsen. A pseudolikelihood approach to analysis of nested case-control studies. *Biometrika*, 84(2):379–394, June 1997. ISSN 0006-3444, 1464-3510. doi: 10.1093/biomet/84.2.379. URL https://academic.oup.com/biomet/article-lookup/doi/10.1093/biomet/84.2.379.

D. Schoenfeld. The asymptotic properties of nonparametric tests for comparing survival distributions. *Biometrika*, 68(1):316–319, 1981. ISSN 0006-3444, 1464-3510. doi: 10.1093/biomet/68.1.316. URL https://academic.oup.com/biomet/article-lookup/doi/10.1093/biomet/68.1.316.

D. A. Schoenfeld. Sample-Size Formula for the Proportional-Hazards Regression Model. *Biometrics*, 39(2):499, June 1983. ISSN 0006341X. doi: 10.2307/2531021. URL https://www.jstor.org/stable/2531021?origin=crossref.

S. G. Self and R. L. Prentice. Asymptotic Distribution Theory and Efficiency Results for Case-Cohort Studies. *The Annals of Statistics*, 16(1), Mar. 1988. ISSN 0090-5364. doi: 10.1214/aos/1176350691.

N. E. Shepardson, G. M. Shankar, and D. J. Selkoe. Cholesterol Level and Statin Use in Alzheimer Disease. *ARCH NEUROL*, 68(10), 2011.

J. H. Shih and T. A. Louis. Inferences on the association parameter in copula models for bivariate survival data. *Biometrics*, pages 1384–1399, 1995.

M. V. F. Silva, C. D. M. G. Loures, L. C. V. Alves, L. C. De Souza, K. B. G. Borges, and M. D. G. Carvalho. Alzheimer's disease: risk factors and potentially protective measures. *Journal of Biomedical Science*, 26(1):33, Dec. 2019. ISSN 1423-0127. doi: 10.1186/s12929-019-0524-y. URL https://jbiomedsci.biomedcentral.com/articles/10.1186/s12929-019-0524-y.

C. M. Sitlani, T. Lumley, B. McKnight, K. M. Rice, N. C. Olson, M. F. Doyle, S. A. Huber, R. P. Tracy, B. M. Psaty, and J. A. Delaney. Incorporating sampling weights into robust estimation of Cox proportional hazards regression model, with illustration in the Multi-Ethnic Study of Atherosclerosis. *BMC Medical Research Methodology*, 20:1–10, 2020.

R. Song, M. R. Kosorok, and J. Cai. Robust Covariate-Adjusted Log-Rank Statistics and Corresponding Sample Size Formula for Recurrent Events Data. *Biometrics*, 64(3):741–750, Sept. 2008. ISSN 0006-341X, 1541-0420. doi: 10.1111/j.1541-0420.2007.00948.x. URL https://academic.oup.com/biometrics/article/64/3/741-750/7331505.

R. A. Sperling, M. C. Donohue, R. Raman, C.-K. Sun, R. Yaari, K. Holdridge, E. Siemers, K. A. Johnson, P. S. Aisen, and for the A4 Study Team. Association of Factors With Elevated Amyloid Burden in Clinically Normal Older Individuals. *JAMA Neurology*, 77(6):735, June 2020. ISSN 2168-6149. doi: 10.1001/jamaneurol.2020.0387. URL https://jamanetwork.com/journals/jamaneurology/fullarticle/2763540.

C. F. Spiekerman and D. Y. Lin. Marginal Regression Models for Multivariate Failure Time Data. *Journal of the American Statistical Association*, 93(443):1164–1175, Sept.

1998. ISSN 0162-1459, 1537-274X. doi: 10.1080/01621459.1998.10473777. URL http://www.tandfonline.com/doi/full/10.1080/01621459.1998.10473777.

StataCorp. Stata 19 Power, Precision, and Sample-Size Reference Manual., 2025. URL https://www.stata.com/manuals/pss.pdf.

C. A. Struthers and J. D. Kalbfleisch. Misspecified proportional hazard models. *Biometrika*, 73(2):363–369, 1986. ISSN 0006-3444, 1464-3510. doi: 10.1093/biomet/73.2.363. URL https://academic.oup.com/biomet/article-lookup/doi/10.1093/biomet/73.2.363.

A. A. Tahami Monfared, M. J. Byrnes, L. A. White, and Q. Zhang. Alzheimer's Disease: Epidemiology and Clinical Progression. *Neurology and Therapy*, 11(2):553–569, June 2022. ISSN 2193-8253, 2193-6536. doi: 10.1007/s40120-022-00338-8. URL https://link.springer.com/10.1007/s40120-022-00338-8.

T. M. Therneau and T. Lumley. Package 'survival'. *R Top Doc*, 128(10):28–33, 2015.

S. Wacholder, M. H. Gail, D. Pee, and R. Brookmeyer. Alternative variance and efficiency calculations for the case-cohort design. *Biometrika*, 76(1):117–123, 1989.

M.-C. Wang. GAP TIME BIAS IN INCIDENT AND PREVALENT COHORTS. *Statistica Sinica*, 9(4):999–1010, 1999. ISSN 10170405, 19968507. URL http://www.jstor.org/stable/24306632.

W. Wang and M. T. Wells. Nonparametric Estimation of Successive Duration Times Under Dependent Censoring. *Biometrika*, 85(3):561–572, 1998. ISSN 00063444, 14643510. URL http://www.jstor.org/stable/2337386.

X. Wang, J. Zhou, T. Wang, and S. L. George. On Enrichment Strategies for Biomarker Stratified Clinical Trials. *Journal of Biopharmaceutical Statistics*, 28(2):292–308, Mar. 2018. ISSN 1054-3406, 1520-5711. doi: 10.1080/10543406.2017.1379532. URL https://www.tandfonline.com/doi/full/10.1080/10543406.2017.1379532.

L. J. Wei, D. Y. Lin, and L. Weissfeld. Regression Analysis of Multivariate Incomplete Failure Time Data by Modeling Marginal Distributions. *Journal of the American Statistical Association*, 84(408):1065–1073, Dec. 1989. ISSN 0162-1459, 1537-274X. doi: 10.1080/01621459.1989.10478873. URL http://www.tandfonline.com/doi/abs/10.1080/01621459.1989.10478873.

S. Willems, A. Schat, M. Van Noorden, and M. Fiocco. Correcting for dependent censoring in routine outcome monitoring data by applying the inverse probability censoring weighted estimator. *Statistical Methods in Medical Research*, 27(2):323–335, Feb. 2018. ISSN 0962-2802, 1477-0334. doi: 10.1177/0962280216628900. URL https://journals.sagepub.com/doi/10.1177/0962280216628900.

M. Witbracht, Y. Xu, O. B. Morgan, C. R. Salazar, D. Hoang, A. Kind, D. L. Gillen, and J. D. Grill. Research Attitudes Questionnaire scores and retention in a recruitment registry.

Journal of Alzheimer's Disease, 103(2):406–418, Jan. 2025. ISSN 1387-2877, 1875-8908. doi: 10.1177/13872877241302422. URL https://journals.sagepub.com/doi/10.1177/13872877241302422.

M. G. Witbracht, O. M. Bernstein, V. Lin, C. R. Salazar, S. A. Sajjadi, D. Hoang, C. G. Cox, D. L. Gillen, and J. D. Grill. Education and Message Framing Increase Willingness to Undergo Research Lumbar Puncture: A Randomized Controlled Trial. *Frontiers in Medicine*, 7, Sept. 2020. ISSN 2296-858X. doi: 10.3389/fmed.2020.00493. URL https://www.frontiersin.org/article/10.3389/fmed.2020.00493/full.

R. Xu. Estimating average regression effect under non-proportional hazards. *Biostatistics*, 1(4):423–439, Dec. 2000. ISSN 14654644, 14684357. doi: 10.1093/biostatistics/1.4.423. URL https://academic.oup.com/biostatistics/article-lookup/doi/10.1093/biostatistics/1.4.423.

Y. Xu and D. L. Gillen. Case-Cohort Designs to Improve Efficiency for Recurrent Events Analysis. *Unpublished doctoral dissertation*, 2026.

Y. Xu, J. D. Grill, D. L. Gillen, and for the Alzheimer's Disease Neuroimaging Initiative. Case-cohort designs to improve efficiency for detecting effect modification. *Statistics in Biosciences*, Aug. 2026. ISSN 1867-1764, 1867-1772. doi: 10.1007/s12561-026-09537-5. URL https://link.springer.com/10.1007/s12561-026-09537-5.

H. Zhang, D. E. Schaubel, and J. D. Kalbfleisch. Proportional hazards regression for the analysis of clustered survival data from case–cohort studies. *Biometrics*, 67(1):18–28, 2011.

Appendix A

Appendices for Chapter 2

A.1 Regularity Conditions for Andersen & Gill (1982) Results

1. $\{Y_i(\cdot), N_i(\cdot), \boldsymbol{Z}_i\}$ are i.i.d.

2. $\boldsymbol{Z}_i$ is bounded w.p. 1.

3. $\Lambda_0(\tau) < \infty$, where $\Lambda_0(t)$ is the cumulative hazard function.

4. $P\{Y_i(t) = 1\} > 0$ for all $t \in (0, \tau]$ where τ is the maximum event time.

5. $I(\boldsymbol{\beta}_0)$ is positive definite.

A.2 Prentice (1986) Variance Terms

Denote $b_{li} = Y_l(t_i)Z_l(t_i)e^{\beta Z_l(t_i)}$, $c_{li} = Y_l(t_i)Z_l(t_i)$, $r_{li} = Y_l(t_i)e^{\beta Z_l(t_i)}$, $R_j = \sum_{l \in R(t_j)} r_{li}$, $B_j = \sum_{l \in R(t_j)} b_{li}$. As developed in Prentice (1986), we have

$$v_{jj} = \text{var}\{U_j(\beta)|\mathscr{F}\left(t_j\right), N(t_j) \neq N(t_j^-)\} = \sum_{i \in \mathcal{R}(t_j)} c_{ij}' c_{ij} r_{ij} R_j^{-1} - B_j' B_j R_j^{-2}, \tag{A.1}$$

$$\begin{aligned} v_{kj} = R_k^{-1} R_j^{-1} \sum b_{ik}' b_{ij} + B_k' B_j R_k^{-2} R_j^{-2} \sum r_{ij} r_{ik} \\ - B_j' R_k^{-1} R_j^{-2} \sum b_{ik} r_{ij} - B_k' R_k^{-2} R_j^{-1} \sum b_{ij} r_{ik} \\ + R_j^{-1} R_k^{-1} \sum \left[\left(b_{ik} - b_{ij}\right)' c_{ij} + \left(b_{jk} - b_{ik}\right)' B_j R_j^{-1} \right] \\ \times r_{ij} \left(r_{ik} - r_{jk}\right) \left(R_k + r_{jk} - r_{ik}\right)^{-1}. \end{aligned} \tag{A.2}$$

Appendix B

Appendices for Chapter 3

B.1 Asymptotic Properties Details

Using the counting process formulation of Andersen and Gill (1982), on a time interval $t \in [0, \tau]$, let N_i, Y_i, Z_i be the counting, at risk, and covariate processes for the ith individual in the cohort. Assume that there is a non-decreasing family of σ-algebras $(\mathcal{F}_t)_{t\in(0,\tau)}$ such that N_i are $(\mathcal{F}_t)$-adapted, and Y_i and Z_i are predictable w.r.t. $(\mathcal{F}_t)$. $(\mathcal{F}_t)$ is termed the filtration. Assume censoring to be independent, and the intensity process is $\lambda_i(t) = Y_i(t)\lambda_0(t)\exp\{\boldsymbol{\beta}_0'\mathbf{Z}_i(t)\}$. Then, we assume that the limit of both n_l/n and m_l/n_l are positive numbers as $n \to \infty$. Assume also that within each level of $Z_{i2}(t)$ (each stratum), the regularity conditions of Self and Prentice (1988) hold.

For $\zeta = 0, 1$, we denote $S^{(\zeta)}(\beta, t) = \sum_{k\in\mathcal{S}} Y_k(t)Z_k(t)^\zeta \exp\{\beta' Z_k(t)\}$, and the conditional expectation of the covariate at time t as $E(\beta, t) = S^{(1)}(\beta, t)/S^{(0)}(\beta, t)$. For the corresponding quantities in a case-cohort design in the Self and Prentice (1988) scheme, we denote $S_{\mathcal{C}}^{(\zeta)}(\beta, t) = \sum_{k\in\mathcal{C}} Y_k(t)Z_k(t)^\zeta \exp\{\beta' Z_k(t)\}$ and $E_{\mathcal{C}}(\beta, t) = S_{\mathcal{C}}^{(1)}(\beta, t)/S_{\mathcal{C}}^{(0)}(\beta, t)$. Then, if we perform stratified sampling within each level of the proposed effect modifier, for each

stratum l in the cohort, $S_{(l)}^{(\zeta)}(\beta,t) = \sum_{k\in\mathcal{S}_l} Y_k(t)Z_k(t)^{\zeta}\exp\{\beta' Z_k(t)\}$. For the subcohort, $S_{\mathcal{C}_l}^{(\zeta)}(\beta,t) = \sum_{k\in\mathcal{C}_l} Y_k(t)Z_k(t)^{\zeta}\exp\{\beta' Z_k(t)\}\frac{n_l}{m_l}$. With the incorporation of the weight $\frac{n_l}{m_l}$, we can assume that $n_l^{-1}S_{\mathcal{C}_l}^{(\zeta)}(\beta,t)$ and $n_l^{-1}S_{(l)}^{(\zeta)}(\beta,t)$ both converge to the same limit as $n\to\infty$.

For the estimation of the covariance matrix, $\hat{\Delta}_l$ is the empirical covariance matrix of $X_k = \int_0^\tau \{Z_k(t) - E_{\mathcal{C}_{l(k)}}(\hat{\boldsymbol{\beta}},t)\}Y_k(t)\exp\{\hat{\boldsymbol{\beta}}' Z_k(t)\}S_{\mathcal{C}_{l(k)}}^{(0)}(\hat{\boldsymbol{\beta}},t)^{-1}d\bar{N}(t)$ for subcohort members in each stratum l. So, $\hat{\Delta}_l = m_l^{-1}\sum_{k\in\mathcal{C}_l}(X_k - \bar{X}_l)(X_k - \bar{X}_l)'$, where $\bar{X}_l$ is the mean of X_k in stratum l. Proofs are outlined in Borgan et al. (2000).

B.2 Results of Averaged Estimates in Application

Tables B.1 and B.2 display the results of averaging estimates of analysis on 100 independent random (for the classic case-cohort design) and stratified (for the stratified case-cohort design) samples when performing subcohort sampling. We notice much more stability on the estimates when considering the subgroup effect for the rare subgroups, i.e. the all other races and ethnicity group and the 2 copies of APOE $\epsilon 4$ group.

	Full Cohort				Classic Case-Cohort (Avg.)				Stratified Case-Cohort (Avg.)			
	N	HR	95% CI	p-value	N	HR	95% CI	p-value	N	HR	95% CI	p-value
pTau												
NH White	905	1.50	(1.36, 1.65)	< 0.001	428	1.56	(1.25, 1.95)	< 0.001	391	1.75	(1.37, 2.25)	0.01
All other groups	94	1.25	(0.71, 2.21)	0.44	35	3.56	(0.94, 19.5)	0.42	94	1.13	(0.60, 2.14)	0.64
Age	999	1.04	(1.02, 1.05)	< 0.001	463	1.04	(1.00, 1.07)	0.09	485	1.04	(1.01, 1.08)	0.11
Female vs. Male	458	0.74	(0.59, 0.93)	0.01	198	0.75	(0.49, 1.15)	0.26	212	0.74	(0.44, 1.23)	0.26
Education	999	0.97	(0.93, 1.01)	0.16	463	0.97	(0.90, 1.04)	0.43	485	0.96	(0.89, 1.05)	0.32
APOE4												
0 copies	588	Ref.	-	-	226	Ref.	-	-	243	Ref.	-	-
1 copy	336	1.82	(1.43, 2.32)	< 0.001	183	1.86	(1.18, 2.95)	0.06	191	2.04	(1.19, 3.52)	0.09
2 copies	75	2.45	(1.68, 3.57)	< 0.001	54	2.91	(1.32, 6.49)	0.07	51	4.72	(2.00, 12.4)	0.15
No. Comorbidities: High	472	0.75	(0.61, 0.94)	0.01	218	0.76	(0.51, 1.15)	0.29	217	0.80	(0.49, 1.32)	0.31
Family History: Yes	549	1.02	(0.81, 1.27)	0.89	256	1.03	(0.68, 1.56)	0.47	266	1.07	(0.69, 1.71)	0.41

Table B.1: Average estimates (across 100 iterations) of hazard ratios (95% CI) and p-values of the model considering race and ethnicity as a potential effect modifier, with adjustment covariates. "All other groups" in variable race and ethnicity includes Hispanic/Latino and non-Hispanic American Indian, Alaskan, Asian, Black, Hawaiian or other PI, and multi-racial individuals. "Ref." stands for referent level.

	Full Cohort				Classic Case-Cohort				Stratified Case-Cohort			
	N	HR	95% CI	p-value	N	HR	95% CI	p-value	N	HR	95% CI	p-value
pTau												
0 copies	588	1.77	(1.50, 2.08)	< 0.001	236	2.01	(1.40, 2.91)	0.01	198	2.29	(1.52, 3.52)	0.01
1 copy	336	2.78	(2.19, 3.53)	< 0.001	172	3.19	(1.94, 5.26)	< 0.001	176	2.88	(1.94, 4.30)	< 0.001
2 copies	75	3.75	(2.64, 5.31)	< 0.001	49	5.08	(2.26, 12.06)	< 0.001	74	3.43	(2.18, 5.40)	< 0.001
Age	999	1.03	(1.02, 1.05)	< 0.001	457	1.03	(1.00, 1.07)	0.15	448	1.03	(1.00, 1.06)	0.14
Female vs. Male	458	0.78	(0.62, 0.98)	0.03	201	0.78	(0.50, 1.22)	0.28	187	0.78	(0.51, 1.21)	0.32
Race and Eth.												
NH White	905	Ref.	-	-	419	Ref.	-	-	410	Ref.	-	-
All other groups	94	0.94	(0.61, 1.43)	0.76	38	1.16	(0.54, 2.54)	0.46	38	1.13	(0.55, 2.40)	0.43
Education	999	0.96	(0.93, 1.00)	0.07	457	0.96	(0.89, 1.04)	0.37	448	0.96	(0.90, 1.03)	0.32
No. Comorbidities: High	472	0.78	(0.63, 0.97)	0.02	216	0.82	(0.53, 1.28)	0.38	211	0.79	(0.53, 1.19)	0.31
Family History: Yes	549	0.99	(0.79, 1.24)	0.92	254	1.02	(0.66, 1.58)	0.48	253	1.05	(0.70, 1.58)	0.43

Table B.2: Average estimates (across 100 iterations) of hazard ratios (95% CI) and p-values of the model considering APOE status as a potential effect modifier, with adjustment covariates. "All other groups" in variable race and ethnicity includes Hispanic/Latino and non-Hispanic American Indian, Alaskan, Asian, Black, Hawaiian or other PI, and multi-racial individuals. "Ref." stands for referent level.

Appendix C

Appendices for Chapter 4

C.1 Regularity Conditions for Asymptotic Proofs

1. $(\boldsymbol{T}_i, C_i, \boldsymbol{Z}_i), i = 1, 2, \ldots, n$ are independent and identically distributed, where $\boldsymbol{Z}_i = (\boldsymbol{Z}_{i1}, \ldots, \boldsymbol{Z}_{iJ})$.

2. For $t \in [0, \tau]$ where $0 < \tau < \infty$, $P(Y(\tau) = 1) > 0$.

3. $|\mathbf{Z}_{ij}(0)| + \int_0^\tau |d\mathbf{Z}_{ij}(u)| < C_z < \infty$ almost surely for some constant C_z.

4. Matrix $A_j(\beta_0)$ is positive definite.

5. (Finite Interval) $\int_0^\tau \lambda_{0j}(t)dt < \infty$, for $j = 1, \ldots, J$

6. (Asymptotic Stability) There exists a neighborhood $\mathcal{B}$ of β_0 where

   (a) There exists functions $s_j^{(d)}(\beta, t), d = 0, 1, 2$ defined on $\mathcal{B} \times [0, \tau]$ such that

$$\sup_{t \in [0,\tau], \beta \in \mathcal{B}} \| S_j^{(d)}(\beta, t) - s_j^{(d)}(\beta, t) \| \xrightarrow{P} 0$$

(b) There exists a matrix $Q(\beta)$ such that

$$n^{-1}\sum_{i=1}^{n}\text{var}(\sum_{j=1}^{J}M_{\tilde{\mathbf{Z}},ij}(\beta_0)) \xrightarrow{P} Q(\beta_0)$$

7. (Asymptotic Regularity Conditions) For all $\beta \in \mathcal{B}, t \in [0,\tau]$, $s^{(d)}(\beta,t)$ are continuous functions of β uniformly in $t \in [0,\tau]$ and bounded on $\mathcal{B}\times[0,\tau]$ and $s^{(0)}(\beta,t)$ is bounded away from zero on $\mathcal{B}\times[0,\tau]$.

8. (Non-Trivial Strata) As $n\to\infty$, $n_j/n \to C_j > 0$ for all $j = 1,2,\ldots,J$.

9. (Non-Trivial Subcohort and Cases) As $n\to\infty$, $\tilde{\alpha}_j = \tilde{n}_j/n_j \xrightarrow{P} \alpha_j \in (0,1)$, $p_{Dj} \xrightarrow{P} p_j \in (0,1)$, where p_{Dj} is the proportion of individuals who experienced event j in the cohort.

10. (Asymptotic Normality of Subcohort Averages at β_0) For $\epsilon > 0$ and all j,

$$n_j^{-1}\sup_t\sum_{i=1}^{n}Y_{ij}(t)\{e^{\beta'\mathbf{Z}_{ij}(t)}\}^2 I(n_j^{-1/2}Y_{ij}(t)e^{\beta'\mathbf{Z}_{ij}(t)} > \epsilon) \xrightarrow{P} 0,$$

$$n_j^{-1}\sup_t\sum_{i=1}^{n}Y_{ij}(t)\{\parallel \mathbf{Z}_{ij}(t) \parallel\}^2\{e^{\beta'\mathbf{Z}_{ij}(t)}\}^2 I(n_j^{-1/2}Y_{ij}(t) \parallel \mathbf{Z}_{ij}(t) \parallel e^{\beta'\mathbf{Z}_{ij}(t)} > \epsilon) \xrightarrow{P} 0$$

11. (Asymptotic Normality of Samples) As $n\to\infty$

$$n_j^{-1}\sup_{i,t\in[0,\tau]}e^{2\beta'\mathbf{Z}_{ij}(t)} \xrightarrow{P} 0,$$

$$n_j^{-1} \sup_{i,t\in[0,\tau]} \{\| \mathbf{Z}_{ij}(t) \|\}^2 e^{2\beta'\mathbf{Z}_{ij}(t)} \xrightarrow{P} 0$$

C.2 Outline of the Proof of Proposition 4.1

Lemma C.1 is modified from Kang and Cai (2009):

Lemma C.1. *Let $\zeta_j = (\zeta_{1j}, \ldots, \zeta_{nj})$ be a random vector containing $\tilde{n}_j$ ones and $n_j - \tilde{n}_j$ zeroes, with each permutation equally likely. Let $B_i(t), i = 1, \ldots, n$, be independent and identically distributed real-valued random processes on $[0, \tau]$ with $E[B(t)] = \mu_B(t), var(B_i(0)) < \infty$ and $var(B_i(\tau)) < \infty$. Let $B(t) = \{B_1(t), \ldots, B_n(t)\}$ be independent. Suppose that almost all paths of $B_i(t)$ have finite variation. Then, $n^{1/2} \sum_{i=1}^n \zeta_{ij}\{B_i(t) - \mu_B(t)\}$ converges weakly in $\ell^\infty[0, \tau]$ to a zero-mean Gaussian process and therefore $n^{-1} \sum_{i=1}^n \zeta_{ij}\{B_i(t) - \mu_B(t)\}$ converges in probability to 0 uniformly in t.*

Lemma C.2 is from Maitra et al. (2020) and Lin (2000):

Lemma C.2. *Let $W_n(t)$ and $G_n(t)$ be two sequences of bounded processes. For some constant τ, assume that the following conditions hold: (i) $\sup_{0\le t\le\tau} \| W_n(t) - W(t) \| \xrightarrow{P} 0$ for some bounded process, $W(t)$, (ii) $W_n(t)$ is monotone on $[0, \tau]$, and (iii) $G_n(t)$ converges to a zero-mean process with continuous sample paths. Then*

$$\sup_{0\le s\le\tau} \| \int_0^s (W_n(t) - W(t)) dG_n(t) \| \xrightarrow{P} 0, \tag{C.1}$$

$$\sup_{0\le s\le\tau} \| \int_0^s (G_n(t) - G(t)) dW_n(t) \| \xrightarrow{P} 0. \tag{C.2}$$

The outline of proof of consistency is as follows. Define $U_n(\beta) = \frac{1}{n}\tilde{U}(\beta)$. Show

1. $\frac{\partial}{\partial \beta'} U_n(\beta)$ exists and is continuous in an open neighborhood of β_0 in $\mathcal{B}$

2. $\frac{\partial}{\partial \beta'} U_n(\beta)|_{\beta=\beta_0}$ is negative definite w.p. $\to 1$ as $n \to \infty$

3. $-\frac{\partial}{\partial \beta'} U_n(\beta) \xrightarrow{P} A(\beta_0)$ uniformly for β in a neighborhood of β_0

4. $U_n(\beta) \xrightarrow{P} 0$

Proof of the first three conditions follow the same arguments as in Maitra et al. (2020). Arguments using $\tilde{\alpha}_j(t)$ for non-case weights do not differ from those using $\tilde{\alpha}(t)$.

For Condition 4, we must show that $n^{-1/2}\tilde{U}(\beta)$ is asymptotically zero-mean normal with finite variance. To do so, decompose into:

$$n^{-1/2}\tilde{U}(\beta) = n^{-1/2} \sum_{i=1}^{n} \sum_{j=1}^{J} \int_0^{\tau} \{\mathbf{Z}_{ij}(u) - e_j(\beta, u)\} dM_{ij}(u) \tag{C.3}$$

$$+ n^{-1/2} \sum_{i=1}^{n} \sum_{j=1}^{J} \left(\sum_{k=1}^{K} \sum_{l=1}^{J} \left\{ e_j(\beta, X_{kl}) - \frac{\tilde{S}_j^{(1)}(\beta, X_{kl})}{\tilde{S}_j^{(0)}(\beta, X_{kl})} \right\} dM_{ij}(X_{kl}) \right) \tag{C.4}$$

Term (C.4) converges to 0 uniformly in t by Lemma C.2 and the uniform consistency of $\tilde{S}_j^{(d)}(\beta, t)$. Under regularity conditions, term (C.3) is asymptotically normal with mean zero and covariance matrix $Q(\beta_0)$ Lin (2000). Therefore, $n^{-1/2}\tilde{U}(\beta)$ is asymptotically normal with mean zero and finite variance $Q(\beta_0)$. Hence, $n^{-1}\tilde{U}(\beta)$ converges to zero in probability. Thus, Condition 4 is satisfied.

By Conditions 1, 2, 3, and 4, there is a unique sequence $\tilde{\beta}$ such that $n^{-1}\tilde{U}(\tilde{\beta}) = 0$ with probability converging to one as $n \to \infty$ and with $\tilde{\beta}$ converging in probability to β_0.

C.3 Simulation Study Adjusting for Time of Event 1

Simulation study settings remain the same except we generate censoring time for subject i as $C_i \sim \text{Uniform}(0, 3.5)$, with Event 1 and 2 rates of 17% and 14%, marginal event rate 17%.

Unadjusted						
	Estimate	Robust Var.	Emp. Var.	Estimate	Robust Var.	Emp. Var.
Full Cohort	0.2631	0.0005	0.0005	0.4053	0.0025	0.0024
CC (KC Time-invariant)	0.2642	0.0010	0.0010	0.4033	0.0053	0.0053
CC (Prentice)	0.2642	0.0011	0.0011	0.4043	0.0057	0.0058
Adjusted						
	Estimate	Robust Var.	Emp. Var.	Estimate	Robust Var.	Emp. Var.
Full Cohort	0.2631	0.0005	0.0005	0.4053	0.0025	0.0024
CC (KC Time-invariant)	0.2643	0.0010	0.0010	0.4035	0.0053	0.0054
CC (Prentice)	0.2643	0.0011	0.0011	0.4044	0.0057	0.0059

Table C.1: Simulation results showing estimates and variance for the event-agnostic model, unadjusted and adjusted for Event 1 time.

In this setting, we find that the case-cohort models have similar estimates compared to the full cohort analysis. Furthermore, we do not see much differences between the unadjusted and adjusted models as we accounted for the induced dependence through including the predictors of interest in the model. If there was an unmodeled predictor independent of the predictor of interest but associated with event time, we may see larger differences when not adjusting for Event 1 time. In such settings, it would be beneficial to adjust for Event 1 time to account for the dependence.

C.4 Analysis of A4 and LEARN Study Data Adjusting for Time of Event 1

After adjusting for Event 1 time, we find that in this analysis, point estimates and significance do not meaningfully differ from the models without Event 1 time adjustment.

			Unadjusted			Adjusted		
	N	Ne	Est. HR	95% CI	p-value	Est. HR	95% CI	p-value
Cholesterol	703	399	1.01	(0.88, 1.15)	0.934	0.99	(0.84, 1.16)	0.885
Substudy								
A4	534	338	Referent	-	-	Referent	-	-
LEARN	169	61	0.35	(0.25, 0.48)	< 0.001	0.56	(0.38, 0.83)	0.004
Age	703	399	2.19	(1.66, 2.89)	< 0.001	2.32	(1.68, 3.21)	< 0.001
Sex								
Male	287	166	Referent	-	-	Referent	-	-
Female	416	233	0.91	(0.69, 1.19)	0.49	0.93	(0.67, 1.29)	0.676
Race and Ethnicity								
NH White	639	363	Referent	-	-	Referent	-	-
All Other Groups	64	36	0.88	(0.55, 1.41)	0.597	0.84	(0.50, 1.42)	0.512
Education	703	399	1.01	(0.96, 1.06)	0.8	1.00	(0.95, 1.06)	0.977
APOE Status								
Non-Carrier	302	136	Referent	-	-	Referent	-	-
Carrier	401	263	2.08	(1.57, 2.75)	< 0.001	1.95	(1.39, 2.75)	< 0.001
Family History								
No	235	134	Referent	-	-	Referent	-	-
Yes	468	265	1.02	(0.77, 1.35)	0.913	1.01	(0.73, 1.40)	0.944

Table C.2: Full event-agnostic model results using the events-proportional CC design, adjusting for Event 1 time. N: number of records. Ne: number of events.

Event 2	N (Records)	Est. (Adjusted 95% CI)	SE	p-value
Full Cohort	261	0.959 (0.785, 1.173)	0.102	0.686
Pooled CC	134	0.766 (0.402, 1.459)	0.329	0.417
Balanced CC*	261	0.959 (0.785, 1.173)	0.102	0.686
Events-prop. CC	163	1.151 (0.855, 1.548)	0.151	0.354

Table C.3: Event 2 analysis in fully stratified model results for estimating the effect of cholesterol levels in association with PACC decline, adjusted for covariates and Event 1 time. All case-cohort models use Prentice weights. *In Event 2, balanced CC sampled all available records, so full cohort analysis was performed.