Methods for the prediction of endpoint-occurrence times in clinical trials

Megan Therese Smith

A dissertation
submitted in partial fulfillment of the
requirements for the degree of

Doctor of Philosophy

University of Washington

2014

Reading Committee:

Ying Qing Chen, Chair

Elizabeth Brown

Peter Gilbert

Program Authorized to Offer Degree:
Biostatistics

University of Washington

Abstract

Methods for the prediction of endpoint-occurrence times
in clinical trials

Megan Therese Smith

Chair of the Supervisory Committee:
Ying Qing Chen
Program in Biostatistics and Biomethematics at Fred Hutchinson Cancer Research Center,
and Department of Biostatistics at University of Washington

This research is motivated by the need to anticipate the calendar time of landmark numbers of events in a clinical trial with a time-to-event monitoring endpoint. In such trials, the observed number of events usually determines the statistical information. Thus the timing of accumulating events in the trial factors strongly into the schedule of interim analyses and Data and Safety Monitoring Board reviews, as well as the overall length of the study. Once a trial is underway, patient accrual, event, and dropout rates may deviate from those anticipated during the design of the study. Then current data from the trial itself may provide the most relevant indicators of the timing of future event occurrences.

Our goal is to develop methods to predict the future calendar time of the N^{th} event in an ongoing clinical trial using data collected between the onset of the study and an interim data cutoff time. We describe two methods for this type of conditional event-time prediction. The first is a fully parametric method, based on maximum likelihood estimation and inference. The second is a semiparametric approach that allows for more flexible modeling of baseline hazards of patient events and loss-to-follow-up, as well as the inclusion of baseline covariate information. For each method, we discuss event-time point prediction and the construction of prediction intervals. Simulation studies are used to demonstrate the performance of the prediction methods in a variety of possible scenarios. This research was motivated by monitoring considerations in the HIV Prevention Trials

Network 052 Study of the prevention of HIV-1 infection with early antiretroviral therapy, and data from the trial is used to demonstrate the proposed methods. The dissertation concludes with a discussion of possible extensions, refinements, and further applications of this research.

TABLE OF CONTENTS

Page

LIST OF FIGURES

LIST OF TABLES

ACKNOWLEDGMENTS

I would like to express gratitude to my dissertation Chair, Ying Qing Chen, for contributing support, time, ideas, and encouragement throughout the course of my research, and for providing guidance and opportunities for me to develop as an independent researcher. I would also like to acknowledge and thank the members of my Supervisory Committee, Peter Gilbert and Elizabeth Brown, for their feedback and advice on the dissertation, and my Graduate School Representatives, Stephen Hawes and Ann Duerr, for their oversight of and assistance with my general and final exams. Many members of the University of Washington Biostatistics Department provided valuable and patient guidance during my course of study, including Graduate Program Assistants Gitana Garofalo and Renee Albert, and Graduate Program Directors David Yanez, Pat Wahl, and Scott Emerson. I am also grateful for the financial support I received from the National Institutes of Health through the Cardiovascular Biostatistics Training Grant.

DEDICATION

This dissertation is dedicated to my parents, Dick and Jennifer, for their unwavering support and belief in me throughout many pursuits and dreams in academics and beyond, to my siblings, Jessica, Andrea, and Nick, for their encouraging words and actions, and to my husband, Mike whose love and company make every challenge surmountable and every day a gift.

Chapter 1

INTRODUCTION AND MOTIVATION

1.1 Objectives

Clinical trials of new drugs or medical treatment regimens are often designed to investigate the length of time from the initiation of patient treatment until one or more clinical outcomes is observed, or "time-to-event." In such trials, patients are usually enrolled over a period of time, the length of observation varies among participants, and patients may or may not experience a primary disease endpoint before the study closes. Furthermore, some patients may drop out of the study or become lost to follow-up before an event of interest may be observed, resulting in censored observations. To accommodate for censoring and variable observation times, survival analysis methods are used with time-to-event data to evaluate and compare the event rates among groups.

In the survival setting many elements of clinical trial design, such as sample size, length of the study, and data analysis and monitoring plans, are based on the timing and total number of events expected to be observed in the study population. Due to the presence of censoring, and particularly when the event rate is relatively low, most of the statistical information in the data is in the number of observed events. For example, estimates of the survival curve for a treatment group, as well as statistical tests comparing survival outcomes for different treatment groups, depend on the number of events observed in each group within specific time intervals. The power of a survival analysis is governed by the number of events, as are the design and implementation of interim and final analyses [26].

This research is motivated by the need for reliable and convenient statistical methods to predict the time from the start of a clinical trial until a landmark number of events is observed, such as the 50th event or 100th event. Tools to accurately predict event times help facilitate optimal adherence to the preset trial protocol. Throughout the study, real-time predictions of the timing of future events may also be used to determine whether any

adjustments to the study design or implementation are warranted, for example modifications to recruitment procedures or the length of planned follow-up for each patient. Then timely adjustments may help preserve the statistical integrity of the trial.

Beyond the statistical considerations, there are also logistical, ethical, and economic reasons it is beneficial to be able to predict the timing of landmark numbers of events in an ongoing trial. Pivotal late-phase clinical trials, that is trials whose objective is to conclusively and reliably evaluate a treatment regimen, particularly addressing fatal or irreversible diseases or conditions, require an independent Data and Safety Monitoring Board (DSMB). The DSMB should include members with expertise in biostatistics, ethics, epidemiology, and the medical disciplines pertinent to the trial [24]. In addition to the DSMB, suggested clinical trial models used by the National Institutes of Health and industry may include a steering committee, sponsor, regulatory agencies, statistical analysis center, data coordinating center, labs, and clinical centers [26]. Reliable estimates of the progress of on ongoing trial help facilitate coordination among this diverse set of individuals and organizations.

The primary responsibility of a DSMB is to protect the interests of current and future trial participants. Thus efficient scheduling of DSMB reviews is vital to avoid unnecessary delays in the disclosure of proven efficacy, unfavorable results, or safety concerns [21]. Convening the DSMB may be logistically complicated or costly, and trial data must be collected, processed, and analyzed in advance of each DSMB meeting. Accurate predictions of landmark event times in the accumulating data can reduce the chances that the DSMB will be convened too early, before sufficient information is available to make important decisions about trial continuation, or unacceptably late, long after important results about benefit or harm were available. To this end, a trial's DSMB commonly requests that the statistical team present periodic predictions of the timing of endpoint occurrences in order to anticipate future reviews of interim trial results.

Furthermore, models for event accumulation can help inform resource management and budgetary considerations, such as determining if at any point the trial should be stopped for futility: that is, if the trial is unlikely to achieve its objectives with the available time and resources. Futility analyses are an important component of timely interim trial reviews. Late-phase trials often involve hundreds or even thousands of patients. If the underperfor-

mance or ineffectiveness of an intervention can be detected early, investigators may avoid subjecting patients to unnecessary therapy or adverse side effects. Futility methods may also be used to demonstrate when a trial may close with insufficient power to detect clinically relevant differences in treatment outcomes, perhaps due to slow recruitment or insufficient recruitment of high-risk patients. Then stopping the trial early for futility preserves money and staff for other, more promising research.

When a clinical trial is designed, extensive research goes into estimating plausible rates of patient participation, disease outcomes, and patient dropout in the study populations. However it is often the case that unforeseen trends develop once a trial is underway. Thus, it is sensible to use observed data from the ongoing trial itself to refine design-phase event-time projections. In this dissertation, we develop a collection of statistical tools to predict future event-occurrence times during an ongoing clinical trial utilizing the current observed trial data. We first introduce a general framework for event-time prediction that can jointly accommodate a variety of commonly used parametric models for patient accrual and time-to-event data. Then we extend the framework to employ more flexible, semi-parametric models; these models can account for dependencies in the data, such as similar clinical outcomes among patients treated at the same clinic, and permit the inclusion of patient characteristics, such as biological or behavioral risk factors for a disease. The resulting methods for event-time prediction are versatile and may be tailored to the scientific context and study design of a specific trial.

When researchers have the opportunity to conduct large and important clinical trials, it is imperative that they also have practical and reliable ways to predict landmark event times in order to assess, review, and disseminate useful information as efficiently as possible. Current event-time prediction methods in the literature rely on extensive data simulation and many prior assumptions about how trial data will accumulate. The new methods developed in this research are simpler to implement: they are computationally inexpensive, do not require the simulation of hypothetical data, and rely on less stringent assumptions about future trial data. Our methods also permit the inclusion of more patient-specific information, which may improve the accuracy and precision of event-time predictions. We believe these statistical methods will be a valuable addition to the clinical trialist's toolbox.

1.2 Motivating Example

The HIV Prevention Trials Network (HPTN) 052 Study is a Phase III, controlled, randomized clinical trial to assess the effectiveness of immediate versus delayed antiretroviral therapy (ART) strategies on sexual transmission of Human Immunodeficiency Virus Type-1 (HIV-1) [11]. It was designed as a long-term partner study. Infected patients who were in a stable, active sexual relationship with an uninfected partner, and who had cluster of differentiation 4 (CD4) counts between 350 and 550 cells/mm^3, were randomly assigned to receive either immediate ART or delayed therapy. Patients on the delayed arm received ART after two consecutive measurements of their CD4 cell counts fell below 250 cells/mm^3, or when they developed AIDS-defining illness, as per existing World Health Organization (WHO) guidelines. Participants in both arms received HIV-1 primary care and couples prevention counseling throughout the study. The HPTN 052 Study was motivated by earlier research indicating that sexual transmission of HIV-1 was strongly linked to levels of the virus in blood plasma and in the genital tract [1][43]. Since combination antiretroviral therapy had been shown to reduce concentrations of HIV-1 in genital secretions of an infected person, investigators hypothesized that ART could also curb the spread of the virus through sexual contact [12][28][49]. The primary objectives of the study were, first, to compare the rates of HIV-1 infection between the two treatment arms, and second, to assess the long-term effectiveness of early antiretroviral treatment as prevention for sexual HIV-1 transmission [11].

After screening, a total of 1,763 HIV-1-serodiscordant same- or opposite-sex couples were enrolled. Enrollment began in April 2005 for a pilot phase, and ran from June 2007 through May 2010 for the full study. Patients were recruited from four continents, nine countries, and thirteen clinical sites (earliest to latest: Boston, Massachusetts, USA; Pune and Chennai, India; Chiang Mai, Thailand; Harare, Zimbabwe; Porto Alegre and Rio de Janeiro, Brazil; Lilongwe and Blantyre, Malawi; Johannesburg, South Africa; Gaborone, Botswana; Soweto, South Africa; and Kisumu, Kenya). The last four sites were added in 2009 to offset the slower-than-expected rate of enrollment among the original nine sites. Couples were randomized in a 1:1 ratio, via permuted-block randomization stratified by

site, resulting in 886 couples on the early treatment arm and 877 on delayed therapy. Two primary endpoints were defined, both of which were time-to-event: genetically linked HIV-1 incident infection in the HIV-negative partner (prevention endpoint), and in the HIV-positive partner, the earliest of death, a WHO stage 4 diagnosis, a severe bacterial infection, or pulmonary tuberculosis (clinical endpoint). Observation of all couples was planned to continue until the last enrolled couple was followed for five years.

For the purpose of demonstrating the methods in this dissertation, we will focus on a composite mortality/morbidity (M/M) time-to-event endpoint defined for the oversight of efficacy and safety throughout the trial. An independent Data and Safety Monitoring Board (DSMB) of the National Institutes of Health and National Institute of AIDS and Infectious Disease (NIH/NIAID) Division of AIDS (DAIDS) monitored the ethics of trial continuation. The DSMB's interim reviews included analysis of the M/M endpoint comprised of the incidence of the earliest of a genetically linked HIV transmission to the HIV-negative partner or the occurrence in the HIV-positive partner of pulmonary tuberculosis, severe bacterial infection, a WHO stage 4 event, or death. The total number of observed composite endpoints among all enrolled couples governed the planned schedule of interim analyses for trial continuation [10]. The interim data were evaluated using a Lan-DeMets implementation of an O'Brien-Fleming monitoring boundary [17][41]. In practice, this approach is often implemented using equally spaced information time intervals (e.g. analyses after observing 25%, 50%, 75%, and 100% of the expected endpoints). For the HPTN 052 Study specifically, three interim analyses and one final analysis were planned when the total number of composite endpoints reached 85, 170, 255, and 340, respectively. The study's DSMB requested the statistical team to present predictions of the timing of endpoint occurrences, so that its review of the interim analyses could be scheduled accordingly.

On April 28, 2011, the DSMB reviewed analyses of data collected through February 21, 2011, by which time the first landmark number of M/M endpoints had been observed (86 composite endpoints between the pooled treatment arms). The early results were compelling: immediate ART reduced genetically linked HIV transmissions by 96% and clinical endpoints by 41% compared with the delayed strategy. The DSMB recommended immediate disclosure of the results to trial participants and the general public [11]. This discovery

is considered a "game-changer" in the effort to control the AIDS epidemic. These astounding early trial results, and the subsequent implications for patient care and policymaking, illustrate the importance of timely, efficient analysis and review of interim clinical trial data. The HPTN 052 Study is currently ongoing and is expected to be complete in 2015. Due to changes in protocol after the April 2011 review, participants on the delayed ART arm are being provided immediate ART. The continuing trial aims to address questions about the magnitude and durability of the Treatment as Prevention strategy [13].

The HPTN 052 Study provides a good example of the difficulty in obtaining accurate pretrial estimates of how the data will accumulate, even for widely researched diseases and treatments. In the best-case scenario, information from previously published studies, pilot studies, and population demographics may be available to inform the design of a new clinical trial. Nonetheless in reality, as was the case for the HPTN 052 Study, things do not always happen according to expectations. When the full study became active worldwide, many of its design parameters, such as patient enrollment and retention rates and baseline incidence rates of HIV transmission and other clinical outcomes, deviated from the anticipated values.

For example, the study team planned an 18-month period for recruitment of 1,750 serodiscordant couples; actual recruitment lasted more than 34 months, mainly due to stringent CD4 enrollment criteria for the HIV-infected partner. The expected annual loss to follow-up rate was 5% per year, and the expected annual event rates (both arms pooled) were approximately 4% for transmissions and 5% for clinical events. However, during the first 3.7 years of the study (through February 2011), the actual average loss rate was approximately 2% per year, the average transmission event rate was approximately 2% per year, and the average clinical event rate was approximately 7% per year [11]. Thus estimation of composite event occurrence times made based on the pretrial estimates was inaccurate. To maintain good coordination among the NIH, the six pharmaceutical companies providing treatment, and the thirteen clinical sites involved once the trial was underway, periodic recalibrations of the anticipated timing of future landmark events in the accumulating data were vital. We will use the HPTN 052 data accumulated in the first 3.7 years of the full study to demonstrate how our estimation method could be applied periodically throughout data collection to anticipate the calendar time of future landmark events, based on the

trends observed in the trial itself.

Unforeseen trends in accrual, health outcomes, or patient dropout may affect the reliability of endpoint occurrence time predictions in clinical trials. Thus it is important to identify trends in the trial data as early as possible. To properly guide the DSMB, adaptive methods are needed that can provide efficient estimates of future endpoint occurrence times, fully utilize the available study data, and accommodate unexpected discrepancies in the design parameters. Preliminary versions of the methods described in this dissertation were applied in the HPTN 052 Study and proved to be pivotal for the last successful DSMB review.

1.3 Specific Aims of the Research

When researchers have the opportunity to conduct large and important clinical trials, it is imperative that they also have practical and reliable ways to predict landmark event times in order to assess, review, and disseminate useful information as efficiently as possible. Investigators should have access to a variety of tools for the prediction of endpoint occurrence times. The determination of which tools are appropriate for use at a particular time in a given study depends on many factors: reliability of preliminary information about patient availability, local disease incidence rates near trial sites, and likelihood of patient retention, in addition to the amount of accumulated trial data and what assumptions an investigator is comfortable making about the future course of the trial.

When a clinical trial is designed, extensive research goes into estimating plausible patient participation rates, rates of clinical outcomes, and patient dropout rates in the study populations. Then, it is sensible to use accumulating data from the trial itself to refine design-phase event-time projections. Methods for event-time prediction currently described in the statistical literature either include only point predictions without uncertainty estimates and/or rely on Bayesian bootstrap or resampling procedures, simulation of hypothetical trial data, and the specification of prior distributions for several model parameters. These are discussed in detail in Chapter 2. In many cases, reliable event-time predictions may be obtained without extensive data simulation and computation. Sometimes it may in fact be advantageous to consider predictions made using few assumptions about the

trial data, as in the case of the HPTN 052 Study, where design-phase rate estimates were inaccurate.

Thus, we will propose a basic prediction framework that will be used to develop convenient frequentist versions of the existing parametric approaches. This framework will then be extended to incorporate semiparametric modeling of survival outcomes utilizing baseline covariate information. With this research we seek to contribute to the statistical literature a collection of flexible, computationally efficient methods for the real-time prediction of endpoint occurrences in clinical trials using available trial data effectively while staying within the bounds of trial confidentiality and credibility.

The proposed methods provide straightforward approaches to use current trial data to obtain point predictions and prediction intervals for future event times. The practicality of these approaches lies in the versatility and simplicity of their implementation. A variety of parametric models may be accommodated, and although no parametric model perfectly fits real data, many provide useful approximations if chosen thoughtfully by the statistician. When parametric modeling assumptions are not appropriate, and/or when information about prognostic risk factors is available for the patients, the semiparametric method provides a flexible and more sophisticated approach. These data-driven, frequentist approaches require neither prior assumptions about model parameter distributions nor the simulation of additional data.

1.3.1 Basic prediction framework and parametric event occurrence time models

First we establish a basic framework for estimating event occurrence times. We specify parametric models for patient accrual, event, and loss-to-follow-up times while assuming mutual independence among a patient's enrollment, event, and dropout times. Point predictions and prediction intervals are obtained for landmark event times using maximum likelihood theory and a modified delta method. The methods allow all study participants to be pooled into a single arm or stratified by a categorical baseline variable, such as treatment group or clinical study site. The validity of point and interval predictions is demonstrated through simulations. The methods are demonstrated and compared to existing Bayesian methods

using real data from the HPTN 052 Study and data collected by the International Chronic Granulomatous Disease (CGD) Cooperative Study Group.

1.3.2 Semiparametric event occurrence time model

The general framework described above is then adapted to accommodate more flexible modeling of time-to-event and time-to-dropout in the clinical trial. The fully parametric model of the event-time distribution is replaced with a more flexible, semiparametric approach employing the Cox model and the Breslow estimate of cumulative hazards.. Under the assumption of proportional hazards, this allows for the inclusion of baseline patient covariate data in the event-time predictions, while making no strict assumptions about the functional form of the baseline hazard of an event. An analogous nonparametric estimate of the cumulative hazards that does not include covariates, the Nelson-Aalen estimator, is also used. Each of these models appears extensively in survival data analysis, and it is sensible to also consider their use in event-time prediction for clinical trials. Point predictions and bootstrap prediction intervals are computed for this approach. The performance of this model is demonstarted for multiple types of baseline covariates through simulation studies.

1.4 Organization

This dissertation is organized as follows. Chapter 2 is a discussion of statistical literature related to event-time prediction in clinical trials. We also include brief discussions of the commonly used parametric, semiparametric, and nonparametric methods for survival data analysis that will be utilized in our prediction methods. In Chapter 3 we describe a likelihood-based parametric approach for predicting landmark event times. Several example scenarios illustrated through simulation studies follow the methodology. The methods are demonstrated on trial data from the HPTN 052 study, including diagnostics for model selection and a comparison of two modeling options. We also compare our prediction method with a previously published Bayesian approach. Chapter 4 describes the semiparametric prediction method, including a point prediction and a bootstrap prediction interval. The formulation of the estimating equation and point predictor are detailed, along with simulations characterizing the behavior of component estimates and the point prediction itself.

Chapter 5 summarizes possible directions for future work, including extensions and refinements of the proposed event-time prediction methods.

Chapter 2

BACKGROUND AND LITERATURE REVIEW

2.1 Prediction of subject event times in clinical trials

Patient enrollment times, event times, and loss-to-follow-up times are all stochastic in nature and must be considered collectively to accurately project the calendar times of future event occurrences in clinical trials. Accrual may be modeled, for example, using Brownian motion, a homogeneous or nonhomogeneous Poisson process, or a truncated parametric distribution. Event and dropout times may be modeled parametrically or nonparametrically, depending on the strength of evidence for, or the confidence in, *a priori* assumptions about hazard rates [55].

Early methods to estimate the total length of a clinical trial with a primary survival endpoint were derived from power and sample size formulas based on the exponential maximum likelihood estimation (MLE) test or the (Mantel-Haenszel) logrank test for the equality of survival curves. These methods require preliminary estimates of constant accrual, event, and loss rates. Rubenstein, et al. proposed a method for projecting the length of a trial consisting of an accrual period followed by a period of continued observation, assuming Poisson (uniform) enrollment and exponential event and loss times [45]. Lachin and Foulkes extended existing formulas relating power, sample size, and trial length to allow for nonuniform enrollment, noncompliance, and life-table analyses stratified on one or more prognostic covariates [32].

Bagiella and Heitjan adapted the model of Rubenstein, et al. to obtain point predictions for landmark event times corresponding to interim and final trial analyses [3]. These predictions utilize accumulating data from the trial and assume Poisson enrollment and exponential event and loss times. Point predictions for specific event times are calculated by plugging maximum likelihood estimates of the accrual, loss, and event rates for each treatment arm into a parametric model. Bagiella and Heitjan also proposed a prediction

interval calculation employing a Bayesian simulation scheme to obtain quantiles of the posterior distribution of a specific event time order statistic, conditional on the observed trial data. Conjugate prior distributions are specified for the Poisson accrual rate and for the exponential event and loss rates on each arm, and completions of the trial data are simulated according to the collection of current posterior distributions.

This approach has been modified to accommodate alternative parametric event and loss time distributions. Ying and Heitjan proposed a fully Bayesian method that assumes Weibull survival and loss times and obtains both the point prediction and prediction interval of a future landmark event time using simulated completions of the trial data [51]. Though more flexible and robust than its predecessor, the algorithm for Weibull prediction is computationally more intensive since the two-parameter Weibull has no conjugate prior and the posterior must be approximated numerically. The above approaches both assume access to unmasked data in which treatment groups are distinguished. Further extensions using parametric mixture models demonstrate the use of the method in the case where treatment arms are masked from the project team [38], and where treatment is masked but information about randomization blocks is available [18]. Later, a Bayesian method utilizing a parametric model that incorporates event-reporting lag was developed by Wang, et al and described for a single (or combined) treatment arm [50].

To address concerns that parametric prediction methods may be biased and inefficient if their underlying distributional assumptions are misspecified, Ying, Heitjan, and Chen proposed a hybrid nonparametric-parametric approach for obtaining point and interval predictions of future event times during an ongoing trial [53]. A Kaplan-Meier curve derived from current accumulated trial data estimates the overall survival curve for event or loss, and a Weibull tail extends each curve from the largest observed event or loss time. Parameters for the Weibull tail are estimated as Bayesian posterior modes, and prediction intervals are constructed using a Bayesian bootstrap. Ying and Heitjan later compared predictions obtained from the Bayesian exponential, Weibull, and nonparametric approaches by applying all three methods to data from the Randomized Evaluation of Mechanical Assistance for the Treatment of Congestive Heart Failure (REMATCH) Trial [52]. All three methods yielded similar point predictions, and the exponential and Weibull methods produced simi-

lar intervals. The nonparametric method, being less efficient, yielded much wider prediction intervals that were also sensitive to short periods of high or low event rates in the data.

Fang and Su modified Ying, Heitjan, and Chen's nonparametric point prediction for a future event time by first detecting change points in the observed survival curve, then retaining the nonparametric estimate before the last change point, and estimating a parametric tail based only on event and loss data reported after the last change point [22]. No method for obtaining prediction intervals is provided for this approach. Recently, Zhang and Long combined a nonhomogeneous Poisson process accrual model with the existing Bayesian parametric and nonparametric event and loss-to-follow-up models for increased flexibility in modeling enrollment when the assumption of a constant accrual rate is not met [54].

2.2 Parametric modeling of time-to-event data

Several families of parametric distributions have been used widely to model survival data, particularly when analyzing univariate failure time data for homogeneous populations. These models may be used in the case of complete or censored data. Detailed treatments and examples are provided in several texts [30][34][35]. Although a single parametric model or family of models can rarely account mathematically for the many complex factors influencing an event time, these tools have been found useful in a broad range of biological scenarios.

The exponential distribution is the simplest parametric survival distribution. It is characterized by a single parameter, which is interpreted as the constant hazard rate for failure. The constant hazard assumption may be unrealistically restrictive over long periods of time, as it ignores factors such as aging, disease progression, or the gradual onset of drug resistance. However the exponential model has proven to be reasonable for some short- or medium-length studies, particularly when the event of interest is rare, and is often employed in the planning phase of a clinical trial.

The Weibull family of distributions is more general and contains the exponential model as a special case. Typically characterized by two parameters, shape and scale, the Weibull model allows for a more flexible power dependence of the hazard rate on time. Populations

with increasing, decreasing, or constant risk may all be modeled by Weibull distributions, making this the most broadly applicable parametric survival model [34]. In nonhomogeneous populations that necessitate covariate adjustment, the Weibull hazard may be used as the baseline hazard in a proportional hazards model, yielding event hazard ratios, or in an accelerated failure time model, yielding acceleration factors on survival time [8].

The methods in this dissertation are developed and demonstrated using exponential and Weibull time-to-event models. However they could be modified to accommodate other frequently applied distributions, such as lognormal, gamma, generalized gamma, or log-logistic distributions. The lognormal is appropriate if the hazard of an event initially increases to a maximum, and then decreases. This may occur in biological scenarios where the occurrence or onset of a disease is right-skewed, such as Alzheimer's disease, Hodgkin's disease, chronic leukemia, or cancer. Hazard rates for the gamma distribution either increase or decrease monotonically to a particular constant value, which may be appropriate when a failure or death occurs after n subfailures or stages have happened. The log-logistic model can describe a monotonically decreasing hazard or one that first increases and then decreases. Thus it encompasses scenarios similar to Weibull or lognormal distributions, but handles censored data better than the lognormal model [30][35].

Selection of a particular parametric model or family should be a careful process, based on compelling empirical or scientific evidence. Poorly specified models may result in biased, inefficient, or inaccurate estimates. A variety of graphical and analytical methods have been developed to compare parametric distributions and assess goodness-of-fit [34][35]. When model diagnostics can be used to justify the choice, parametric models are mathematically convenient and offer clear interpretations.

2.3 Semiparametric and nonparametric modeling of time-to-event data

The most common method for estimating the cumulative hazard function in the context of censored data, making no *a priori* assumptions about the shape of the distribution, is the nonparametric Nelson-Aalen estimate, $\widehat{\Lambda}_{np}(t)$, a right-continuous step function with jumps at each observed failure time equal to the empirical hazard estimates [39]. The Nelson-Aalen estimator is simple to calculate and interpret and has well-developed finite-sample

and asymptotic theory [25][30][47]. An estimate of the survival function, suggested by Breslow, is simply a transformation of the Nelson-Aalen estimator.

$$\widehat{S}_{np}(t) = \exp -\widehat{\Lambda}_{np}(t).$$

The Breslow estimate is very closely related, mathematically, to the Kaplan-Meier estimate, another standard tool for describing the survival distribution of a sample population [31]. Because they are empirical estimates, the Nelson-Aalen and Breslow estimators are only defined over the range of observed survival times in the data.

When both the survival distribution and the relationship of covariates to survival are of interest, the semiparametric Cox model is the most widely used analytic tool. The Cox model is a proportional hazards model that leaves the baseline hazards an unspecified function of time and assumes that covariates have a multiplicative effect on the baseline hazard [14]. That is, the hazard of an event for patient i is

$$\lambda_i(t|z_i) = \lambda_0(t)e^{\beta z_i},$$

where z_i is the covariate vector for patient i, and β is a vector of coefficients. A partial likelihood equation is used to estimate the log hazard ratio(s) β and yields a consistent and asymptotically normal solution, $\widehat{\beta}$ [30][47]. The same partial likelihood estimate $\widehat{\beta}$ may be incorporated in the Breslow estimator to obtain a covariate-adjusted version of the cumulative hazard function.

The fundamental assumption of proportional hazards in the Cox model states that the ratio of hazards for two subjects or groups, each defined by a fixed set of covariates, is constant over time. The validity of any data analysis based on the Cox model (or covariate-adjusted Breslow estimate) depends on the appropriateness of the proportional hazards assumption and the goodness-of-fit of the model. Several basic graphical, analytical, and residual based methods to assess these modeling assumptions have been described [35][47]. Bagdonavicus also suggests a test of the goodness-of-fit for the Cox model based on a modified score statistic and a general family of survival models that includes the Cox model as a special case [2].

Chapter 3

PARAMETRIC PREDICTION OF EVENT TIMES IN CLINICAL TRIALS

3.1 Introduction

We begin our development of methods for event-time prediction in clinical trials with a fully parametric framework. Although parametric models are currently less commonly used than their semiparametric and nonparametric counterparts for data analysis and inference in the survival setting, several attributes make them convenient tools for prediction. Many different distributional families with varying shapes and levels of flexibility are available and provide sensible choices for modeling biological and medical processes, as mentioned in section 2.2. Estimation procedures for assessing the goodness-of-fit for a particular family, or for comparing the fits of two different families are available. Different methods for parameter estimation have been well documented. The basic shape of a parametric distribution, along with the associated survival, cumulative hazard, and hazard functions, is also clearly defined over its entire range, making the prediction and monitoring with these models accessible and tractable.

We begin by defining basic notation describing the clinical trial data and setting in which predictions will be implemented. Our estimand and the estimating procedures to predict it are clearly defined. First we demonstrate briefly how a marginal version of the model may be used as a convenient and informative tool in the planning stage of a trial. Then a conditional model for prediction is defined and implemented under several modeling assumptions. We demonstrate the process of model selection and landmark event-time prediction using data from the HIV Prevention Trials Network 052 Study. Finally, we compare our prediction method to an existing Bayesian approach using publicly available data from a study of Gamma Interferon in Chronic Granulomatous Disease.

3.2 Methods

3.2.1 Notation

Consider a clinical trial, consisting of a patient enrollment period followed by a period of continued observation. Denote the desired total sample size by n. In the following, random variables are represented by capital letters, and their lower case counterparts denote observed data values. Assume enrollment begins at time zero, and let W_i denote the enrollment time of the ith patient, measured from time zero, with the indices $i = 1, \ldots, n$ corresponding to the order in which patients accrue. Then the enrollment period ends at time W_n, with the accrual of the final nth patient. Investigators follow each patient from enrollment until the study closes or until the patient experiences the event of interest or is lost to follow-up. We measure individual event or loss times beginning from the patient's enrollment time. The observation time for patient i, denoted X_i, is the minimum of the event time, loss time, or time from enrollment to a data cutoff time, C. Three binary indicator variables designate whether the ith patient is observed to experience the study event, Δ_i^E, is lost-to-follow-up, Δ_i^L, or undergoes administrative censoring at an interim data cutoff or end-of-study time, Δ_i^A. The observed values, $\{w_i, x_i, \delta_i^E, \delta_i^L, \delta_i^A\}$ may vary for a patient depending on when in the course of the trial the data is viewed. To clarify data observed in a particular study time interval $[0, C]$, we could indicate a patient's data by

$$\{w_{C,i}, x_{C,i}, \delta_{C,i}^E, \delta_{C,i}^L, \delta_{C,i}^A\}.$$

However, since we will calculate each set of predictions from data accrued before a single cutoff time, the subscript C is hence suppressed for brevity.

Suppose N is a landmark number of events, such as the optimal number of events for an interim or final analysis. Let R_N denote the time from the beginning of enrollment to the Nth observed event. We want to estimate the expected value of the time to the Nth observed event, $E[R_N]$. Define $Y_i \equiv W_i + X_i$, the elapsed time from the start of enrollment until observation of the ith patient terminates. These absolute observation times and corresponding event indicators for all patients, ordered by magnitude, are the

pairs

$$(Y_{(1)}, \delta^E_{(1)}), (Y_{(2)}, \delta^E_{(2)}), \ldots, (Y_{(n)}, \delta^E_{(n)}).$$

Thus our goal is to estimate the expected value of the Nth order statistic from the collection of observed absolute event times,

$$E[R_N] = E\Big[Y_{(k)} : \sum_{j=1}^{k} \delta^E_{(j)} = N; \delta^E_{(k)} = 1\Big]. \tag{3.1}$$

Estimation of event times in clinical trials is a difficult task and depends on many factors. There is random variation in patient enrollment times, event times, and loss-to-follow-up times. These random variables must all be included in an adequate prediction model. A variety of parametric distributions may be used to model patient enrollment, event, and loss times. One simple approach to modeling patient accrual is to assume enrollment times are independent and identically distributed (iid) with density $f_W(t)$. For example, an exponential density could be used when enrollment is expected to be rapid at the study start and taper throughout the enrollment period. Alternatively, we could model the distribution of enrollment times for all n patients jointly, either using a homogeneous Poisson process for approximate constant-rate enrollment, or a non-homogeneous Poisson Process for enrollment with a time-varying rate, or by designing a model that captures specific conditions of recruitment for a particular trial. Assume the hazard and survival functions for event and loss are given, respectively, by $\lambda_E(t)$, $S_E(t)$, $\lambda_L(t)$, and $S_L(t)$. For example, an exponential survival time distribution is an appropriate choice when the hazard of an event (or dropout) is approximately constant; a Weibull distribution has a more flexible power dependence of the hazard on time, allowing for increasing, decreasing, or constant hazards over time. In the occurrence time estimation models proposed here, enrollment, event, and loss time distributions are assumed to be mutually independent.

3.2.2 General event-occurrence-time model for study planning

In the planning phase of a trial, estimates of the approximate timing of landmark events are essential. The general method described here may be used to produce estimates for landmark event occurrence times. The estimates will reflect the projected enrollment, event, and loss

rates and corresponding probability distributions deemed appropriate by the investigators for the design and scientific context of the trial.

Suppose a patient is enrolled at time W. The probability of that patient experiencing an observed event in the time interval $[W, W + \Delta]$, $\Delta > 0$, is given by the function $Q(\Delta)$, where

$$Q(\Delta) = \int_0^{\Delta} \lambda_E(u) \cdot S_E(u) \cdot S_L(u) du$$

Let $a(t)$ be a function for the anticipated instantaneous rate of change in enrollment, derived from the prespecified enrollment distribution, i.e. the anticipated number of new patients joining the study at time t. For a study time interval of $[0, R]$ the expected number of observed events may be computed as

$$\int_0^{U} a(v) \cdot Q(R - v) dv,$$

where $U = \min(E[W_n], R)$. The upper limit of integration is R if patient accrual is still anticipated to be underway at R, or it is the expected end of the enrollment period, $E[W_n]$, if accrual is to be complete before R.

Define the function $G(\cdot)$ as

$$G(R; E[W_n], N) = \int_0^{U} a(v) \cdot Q(R - v) dv - N. \tag{3.2}$$

Given a particular sample size, n, the projected length of the accrual period, $E[W_n]$, a desired number of observed events, N, and a set of event, dropout, and enrollment distributions, we may calculate the expected Nth event time, $E[R_N]$, as the solution to the equation $G(R; E[W_n], N) = 0$. That is to say, $E[R_N]$ is implicitly defined by n, $E[W_n]$, and N in the function $G(\cdot)$. Extensions of the model to accommodate multiple treatment arms or clinical sites, when necessary, are straightforward to implement. Example scenarios are illustrated in section 3.3.2.

3.2.3 Methods for conditional event-occurrence-time model for study monitoring

When a study is underway, investigators may be interested primarily in a method to predict the timing of future events, based on trends in the accumulating trial data, and conditional

on the exact number and timing of events already observed. That is, the random quantity of interest is the time from the beginning of the study to the Nth observed event, given the current trial data. If we denote the trial data accumulated in the interval $[0, C]$ as $\mathscr{F}_C$, then our goal is to estimate

$$E[R_N|\mathscr{F}_C] = E\Big[Y_{(k)} : \sum_{j=1}^{k} \delta_{(j)}^E = N; \delta_{(k)}^E = 1|\mathscr{F}_C\Big]. \tag{3.3}$$

To this end we modify (3.2) to calculate event-time predictions from the time of the most recently observed event. At an interim cutoff time, $C > 0$, measured from the start of enrollment, we will use the data collected in the interval $[0, C]$ to predict the timing of one or more future event occurrences. The cutoff time C may occur while enrollment is underway or during the observation period.

Let n_C represent the total number of patients enrolled in the study by time C, and let N_C be the number of events observed by time C. Conditional on $\mathscr{F}_C$, we estimate the number of observed events expected to occur by a future study time $R > C$ as a sum of three terms: the events already observed by time C, the events expected to occur in the interval $[C, R]$ among patients enrolled and still at risk at time C, and the events expected to occur by time R among patients to be enrolled after time C. Recall the observed data as of time C (with subscript C suppressed) for the ith patient enrolled is

$$\{w_i, x_i, \delta_i^E, \delta_i^L, \delta_i^A\}.$$

Similar to the general model, given a particular sample size, n, a desired number of observed events, N, and a set of distributions selected for enrollment, event, and dropout times, we calculate $E[R_N]$ as the solution to the equation $G^c(R; \widetilde{W}_n, N|\mathscr{F}_C) = 0$, with $G^c(\cdot)$ defined as follows. Let $\widetilde{W}_n$ be the observed value of W_n if the full sample has been accrued or $E[W_n]$ if accrual is still underway. Then let

$$\begin{aligned} &G^c(R; \widetilde{W}_n, N|\mathscr{F}_C) = \\ &N_C + \sum_{i=1}^{n_C} \left\{ \delta_i^A \cdot (1 - \delta_i^E) \cdot (1 - \delta_i^L) \cdot \int_C^R \frac{\lambda_E(u - w_i) \cdot S_E(u - w_i) \cdot S_L(u - w_i)}{S_E(C - w_i) \cdot S_L(C - w_i)} du \right\} \\ &+ I(n_C < n) \cdot \int_C^U a(v) \int_0^{R-v} \lambda_E(u) \cdot S_E(u) \cdot S_L(u) du dv - N, \end{aligned} \tag{3.4}$$

where $U = \min(\widetilde{W}_n, R)$. The number of events already observed by time C is given by N_C. The second term calculates the expected probability of an observed event in the interval $[C, R]$ among enrolled patients, conditional on their being event- and dropout-free at C. If the full sample is not yet accrued at time C, the third term is nonzero and represents the expected number of events to be observed by time R among patients to be enrolled after C.

Appropriate parametric distributions for enrollment, event, and loss times may be selected based on the scientific framework of the trial, original or updated design features, and empirical estimates from the observed data. Let θ denote the parameters governing these distributions. Then we may alternatively write equation (3.4) as $G^c(R; \widetilde{W}_n, N, \theta | \mathscr{F}_C)$. Note that the value of the study time R is implicitly defined by $\widetilde{W}_n$, N, θ, and the observed data $\mathscr{F}_C$ as the solution to the equation $G^c(R; \widetilde{W}_n, N, \theta | \mathscr{F}_C) = 0$.

Using the data observed in $[0, C]$, we obtain standard ML estimates for each parameter, yielding $\widehat{\theta}_{nC}$. If enrollment concluded before C, the value W_n is observed in the available data. If enrollment is still underway at C, an estimate, $\widehat{E[W_n]}$, of $E[W_n]$ may be calculated using the enrollment parameters in $\widehat{\theta}_{n_C}$. Plugging the ML estimates into G^c yields an estimating equation for $\widehat{E[R_N]}$,

$$\widehat{G^c}(R) = G^c(R; N, \widetilde{W}_n, \widehat{\theta}_{nC} | \mathscr{F}_C).$$

The point estimator $\widehat{E[R_N]}$ is calculated as the solution to

$$\widehat{G^c}(R) = 0.$$

Maximum likelihood estimators are used widely because of their asymptotic optimality. Under fairly general conditions, ML estimates are consistent and asymptotically normal. For example, assuming the accrual, event, and dropout distributions are differentiable, as is the case with the parametric distributions discussed in this chapter, the ML estimate $\widehat{\theta}_{n_C}$ converges to the true θ as the available sample size n_C tends to infinity [36]. Under additional smoothness conditions described in detail in [36] and [23], $\widehat{\theta}_{n_C}$ also satisfies the convergence in distribution

$$\sqrt{n}(\widehat{\theta}_{n_C} - \theta) \to \text{Normal}\Big(0, I(\theta)^{-1}\Big),$$

with Fisher information $I(\theta)$. The smoothness conditions are satisfied, for example, when a third derivative of the distribution exists. The value $E[R_N]$ may be viewed as a function of θ. Let $G^*(\cdot)$ be defined as the root of $G^c(R; N, \widetilde{W_n}, \cdot|\mathscr{F}_C) = 0$. Then by the invariance property of ML estimates, the ML estimate of $G^*(\theta)$ is $G^*(\widehat{\theta})$ [9]. Thus $\widehat{E[R_N]}$ is a consistent estimator of $E[R_N]$.

We also construct a prediction interval for the point estimator $\widehat{E[R_N]}$ using the observed information bounds for each ML parameter estimate, in conjunction with a modified delta method for estimating the variance of an implicitly defined random variable [5]. Each of the components of $\widehat{\theta}_{nC}$ is a ML estimate, and thus $\sqrt{n_C}\widehat{\theta}_{n_C}$ is asymptotically normal with covariance matrix, Σ, equal to the inverse of the Fisher information matrix for the distribution parameters. Since the random variable $\widehat{E[R_N]}$ is implicitly defined as a function of $\widehat{\theta}_{n_C}$,

$$\sqrt{n_C}\Big(\widehat{E[R_N]} - E[R_N]\Big)$$

is asymptotically normal with estimated variance

$$\widehat{V} = \Big(\frac{\partial G^c}{\partial R}\Big)^{-1} H\Sigma H^T \Big\{\Big(\frac{\partial G^c}{\partial R}\Big)^{-1}\Big\}^T,$$

where H is the $(1 \times k)$ vector with elements $\partial G^c/\partial\theta_j$, $j = 1, \ldots, k$.

An interval is constructed using the asymptotic normality of $\widehat{E[R_N]}$ and the standard formula for a frequentist prediction interval, or "predictive confidence interval," as detailed in [27]. Let $\mathscr{T}^m$ be Student's t-distribution with m degrees of freedom. Then

$$\frac{E[R_N] - \widehat{E[R_N]}}{\sqrt{\Big(\frac{\widehat{V}}{n_C}\Big)\Big(1 + \frac{1}{n_C}\Big)}} \sim \mathscr{T}^{n_C - 1},$$

Solving for $E[R_N]$ yields the predictive distribution

$$E[R_N] \sim \widehat{E[R_N]} + \sqrt{\Big(\frac{\widehat{V}}{n_C}\Big)\Big(1 + \frac{1}{n_C}\Big)}\,\mathscr{T}^{n_C - 1}.$$

The endpoints of a symmetric $100(1 - a)\%$ prediction interval are

$$\widehat{E[R_N]} \pm t_{(1-a/2, n_C-1)}\sqrt{\Big(\frac{\widehat{V}}{n_C}\Big)\Big(1 + \frac{1}{n_C}\Big)},$$

where $t_{(1-a/2,n_C-1)}$ is the critical value for the Student's t-distribution with significance level a and $n_C - 1$ degrees of freedom.

The interpretation of the frequentist prediction interval is analogous to that of the confidence interval. For our conditional model (3.4), we interpret the prediction interval assuming all previously observed data values are fixed. Given the observed data, $\mathscr{F}_C$, and assuming that the distribution of data observed in $[0, C]$ is representative of that to be collected in $(C, R]$, then in repeated experiments we would expect the prediction interval to cover the true value $E[R_N]$ $(1-\alpha)\%$ of the time. Also note that in practice we may truncate the left end of the prediction interval at $R = C$, since the method will always be implemented to predict the time of an event known to be not yet observed at C.

To calculate a point prediction using this parametric method, we require at least one observed event in $\mathscr{F}_C$; to obtain the estimated variance and prediction interval for $\widehat{E[R_N]}$, $\mathscr{F}_C$ must include at least one observed event and one observed dropout. In studies where patient dropout is not an issue, the point prediction and prediction interval formulas may easily be modified by setting the survival function for loss-to-follow-up equal to one and $\delta_i^L = 0$ $\forall i$ in (3.4). In this case no distribution for dropout is specified, and no dropout parameters are estimated.

3.3 Example scenarios and simulation studies

3.3.1 Parametric model selection in monitoring clinical trials

The general and conditional models described above may be used with any parametric distributions for which we can obtain consistent ML parameter estimates that have asymptotically normal distributions. Some parametric models are not well defined if the exact finite range of the value of the random variable is not known in advance, such as uniform or beta distributions. Maximum likelihood parameter estimation in those models is not possible when the available data is right truncated, as is the case in study monitoring. Consequently these models are appropriate for use as enrollment distributions in the general model for planning, but not in the conditional model for study monitoring.

The following examples and subsequent data analyses use either exponential or Weibull

models for time-to-event or time-to-dropout distributions. We note that other distributions, such as the lognormal or gamma, may alternatively be used when hazards are not assumed to be monotonic. However, the more complex the hazard function is in the assumed time-to-event distribution family, the more unlikely it is that reliable predictions may be obtained with limited amounts of data, e.g. early in the clinical trial.

The Weibull distribution is the most widely used parametric lifetime model due to its applicability in a variety of biological, medical, and many other types of analyses [34][44]. Its appeal stems from the flexibility of modeling hazards as constant, increasing, or decreasing with various shapes. Although there are certainly many disease processes with hazards that follow U-shaped or other, more complicated trajectories over a lifetime, most clinical trials only capture a relatively short interval, or snapshot, of the entire progression. The hazard of an event during a truncated interval may often be well approximated with a monotonic function. The two-parameter Weibull model is sufficient for the analyses presented in this dissertation. However the Weibull distribution may also be generalized to a three-parameter version. The three-parameter model can, for example, be used to incorporate a "guarantee time" at the beginning of observation, during which no deaths may occur [35]. A host of graphical and analytical methods have been described in the literature to assess the goodness of fit of the Weibull model to a set of observed data [34][35][44]. We demonstrate particular large- and small-sample diagnostic methods in section 3.4.

3.3.2 Example scenarios of event-time estimation for study planning

For a clinical trialist planning a study, the general model can be useful for considering a variety of plausible scenarios for enrollment, event, and dropout. In this section we demonstrate how to use the general model to obtain estimates of the expected time from the start of the study until the observation of landmark event times. The first scenario illustrates distributional choices similar to those assumed in the planning of the HIV Prevention Trials 052 Study, pertinent to the safety M/M monitoring endpoint, our motivating example (Poisson enrollment, and exponential event and loss times). The second example demonstrates the use of alternative modeling assumptions (beta accrual, Weibull event times, and

exponential loss times) and describes circumstances which might justify these more flexible choices. The HPTN 052 Study is described in detail in section 1.2.

Example 1: Poisson enrollment and exponential event and dropout times

The study design, methods, and early trial results for HPTN 052 were described by Cohen, et al. [13]. Ying et al. further discussed statistical considerations of the HPTN 052 Study that informed decisions about the design and analysis for this trial [10]. Investigators designed the trial with two treatment arms (early ART vs. delayed ART), randomly assigned in a 1:1 ratio. A sample size of 1,750 serodiscordant couples was selected to provide 92% power to demonstrate that early ART would yield a 20% reduction in risk of serious clinical events in the HIV positive partner (clinical endpoint), and 87% power to detect a 39% reduction in incidence of HIV transmission to the HIV negative partner (prevention endpoint). A combination of these two endpoints comprised the time-to-event M/M safety endpoint.

We assume accrual was planned to last 1.5 years, with expected enrollment rates of 75 patients per month for the first 6 months and 110 patients per month thereafter. (Note: the actual accrual estimates were slightly lower for the first 6 months, due to patient accrual during a run-in phase of the trial preceding the start of the full study [10]. The run-in period is excluded here to clarify the example.) Investigators anticipated cumulative incidence rates of approximately 10% for transmissions and approximately 9% for clinical events in the early treatment arm, and cumulative incidence rates of approximately 13% for transmissions and 18% for clinical events in the delayed treatment arm, resulting in a total of 340 M/M events [13]. Anticipating that some couples could experience both clinical and transmission events, assume an annual M/M composite event rate of 4% in the early treatment arm and 6% per year in the delayed arm. Investigators planned to allow for an overall 5% annual loss-to-follow-up rate.

We will use the general model to produce estimates of the 85th event, or 25% of the total 340 M/M events, i.e. the target number of events for the first interim safety analysis. Enrollment is modeled as a Poisson process with rate $\alpha_{P,\text{rate1}} = 900$ couples per year for the first six months and $\alpha_{P,\text{rate2}} = 1,320$ couples per year for the next year. The times to event

(early and delayed arms) and dropout may be modeled using exponential distributions with rate parameters $\gamma_{E1} = 0.04$, $\gamma_{E2} = 0.06$, and $\gamma_L = 0.05$, respectively. Expanded ranges of plausible event and dropout rates around the target rates are calculated to illustrate a possible range for the 85th event time in the actual trial.

Letting $j = 1, 2$ denote treatment arm, distribution parameters

$$\theta = \{\alpha_{P,\text{rate1}}, \alpha_{P,\text{rate2}}, \gamma_{E1}, \gamma_{E2}, \gamma_L\},$$

and the instantaneous accrual rate function defined

$$a(v) = 1(v \leq 0.5)\alpha_{P,\text{rate1}} + 1(0.5 < v)\alpha_{P,\text{rate2}},$$

the general event occurrence time model, following equation (3.2), for this scenario is

$$\begin{aligned} G(R; E[W_n], N, \theta) &= \sum_{j=1}^{2} \left\{ \int_0^U \frac{a(v)}{2} \cdot \int_0^{R-v} \lambda_{Ej}(u) S_{Ej}(u) S_L(u) du dv \right\} - N \qquad (3.5) \\ &= \sum_{j=1}^{2} \left\{ \int_0^U \frac{a(v)}{2} \cdot \int_0^{R-v} \gamma_{Ej} e^{-\gamma_{Ej} u} e^{-\gamma_L u}) du dv \right\} - N \\ &= \sum_{j=1}^{2} \left\{ \left[\frac{\alpha_{P1}\gamma_{Ej}}{2(\gamma_{Ej} + \gamma_L)^2} \int_0^U e^{-(\gamma_{Ej}+\gamma_L)(R-v)} - 1 dv \right] \right. \\ &\quad \left. + \left[\frac{\alpha_{P2}\gamma_{Ej}}{2(\gamma_{Ej} + \gamma_L)^2} \int_{0.5}^U e^{-(\gamma_{Ej}+\gamma_L)(R-v)} - 1 dv \right] \right\} - N \end{aligned}$$

The estimated timing of a landmark event, measured from the beginning of enrollment, can easily be calculated by plugging the prescribed values into equation (3.5) and solving for R in

$$G(R; E[W_n], N, \theta) = 0.$$

The estimated time of the $N = 85$th event for hypothesized ranges of event and dropout rates in this example are shown in Table 3.1.

Example 2: beta accrual, Weibull event times, and exponential loss times

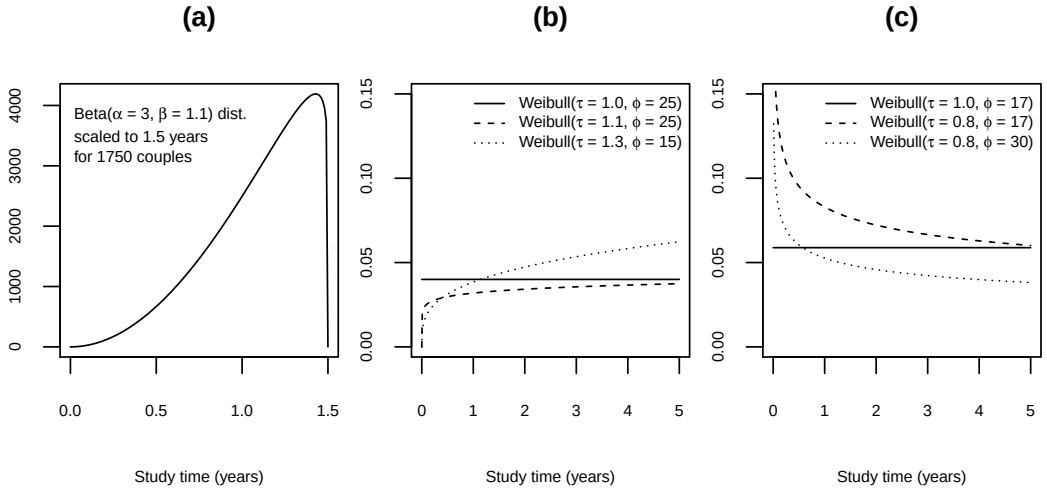
The Poisson enrollment time and exponential event and dropout time distributions of the previous example assume a piecewise constant entry rate and constant failure rates. In some

Table 3.1: General model predictions of the time of the 85th observed event, measured in years after the study start, assuming piecewise Poisson accrual and exponential event and loss times (Poisson rate parameters: $\alpha_{P1} = 900$, $\alpha_{P2} = 1320$).

Event rate arm 1	Event rate arm 2	Loss to follow-up rate		
		Predicted event time (yrs)		
γ_{E1}	γ_{E2}	3%	5%	7%
1%	2%	4.32	4.45	4.61
	6%	2.28	2.31	2.33
	10%	1.75	1.76	1.77
4%	2%	2.51	2.55	2.58
	6%	1.82	1.83	1.85
	10%	1.53	1.54	1.55
7%	2%	1.94	1.96	1.98
	6%	1.59	1.59	1.60
	10%	1.40	1.41	1.41

settings, more flexible modeling may be warranted to adequately represent characteristics of the study design and implementation. For example, HPTN 052 ran at thirteen study sites in nine different countries. The site openings were staggered due to varying regulatory processes in each region [10]. Thus it would be reasonable to assume enrollment would be slow at first, as study sites implement procedures and train staff, and then more rapid later. Trends like this and a variety of others may be produced using a beta distribution, scaled to the anticipated length of the enrollment period. The trend described here could be produced, for example, by choosing a beta distribution with shape parameters $\alpha = 3$ and $\beta = 1.1$. See Figure 3.1(a).

Figure 3.1: Distribution and parameter selections for Study Planning Example 2. (a) Instantaneous accrual rate function derived from scaled Beta distribution. (b) Hazard functions for range of Weibull distributions for event times on treatment arm 1 (early ART). (c) Hazard functions for range of Weibull distributions for event times on treatment arm 2 (delayed ART).



Furthermore, the long-term goal of assessing durability of the ART treatment effect complicated the estimation of event rates over the course of the years-long trial. On the early treatment arm, investigators anticipated a possible reduction over time in efficacy due to developing drug resistance, lack of adherence, and ineffective therapy (increasing hazard rate for an event). Conversely as HIV-positive patients on the delayed therapy

arm gradually initiated treatment, a possible improvement in clinical and transmission outcomes was anticipated (decreasing hazard rate for an event). Event rates that increase or decrease over time may be modeled using Weibull distributions. In this scenario, one could calculate landmark event occurrence times using a range of Weibull parameters to determine the magnitude of influence on landmark event times resulting from different event rates, as in Figure 3.1(b) and 3.1(c). Here we use the general model to calculate landmark event times assuming a scaled Beta(α_B, β_B) distribution for enrollment times, Weibull(shape $= \tau_{E1}$, scale $= \phi_{E1}$) and Weibull(shape $= \tau_{E2}$, scale $= \phi_{E2}$) distributions for event times on each of the $j = 1, 2$ treatment arms, and exponential(rate $= \gamma_L$) distributed dropout times.

Letting

$$\theta = (\alpha, \beta, \tau_{E1}, \phi_{E1}, \tau_{E2}, \phi_{E2}, \gamma_L),$$

the general event occurrence time model, following equation (3.2), for this scenario is

$$\begin{aligned}
&G(R; E[W_n], N, \theta) \qquad (3.6)\\
&= \sum_{j=1}^{2} \left\{ \int_0^U \frac{a(v)}{2} \int_0^{R-v} \lambda_{Ej}(u) S_{Ej}(u) \lambda_L(u) du dv \right\} - N \\
&= \sum_{j=1}^{2} \left\{ \int_0^U \frac{n}{2} \left(\frac{v}{E[W_n]} \right)^{\alpha_B - 1} \left(1 - \frac{v}{E[W_n]} \right)^{\beta_B - 1} \right. \\
&\cdot \left[\int_0^{E[W_n]} \left(\frac{w}{E[W_n]} \right)^{\alpha_B - 1} \left(1 - \frac{w}{E[W_n]} \right)^{\beta_B - 1} dw \right]^{-1} \\
&\left. \cdot \int_0^{R-v} \left[\frac{\tau_{Ej}}{\phi_{Ej}} \left(\frac{u}{\phi_{Ej}} \right)^{\tau_{Ej} - 1} \right] \exp\left[- \left(\frac{u}{\phi_{Ej}} \right)^{\tau_{Ej}} - \gamma_L u \right] du dv \right\} - N
\end{aligned}$$

The timing of a landmark event, measured from the beginning of enrollment, is calculated by plugging the prescribed values into equation (3.6) and solving $G(R; E[W_n], N, \theta) = 0$. The predicted times of the $N = 85$th event for the hypothesized ranges of event and dropout rates in this example are shown in Table 3.2

Table 3.2: General model predictions of the time of the 85th observed event, measured in years after the study start, assuming Beta accrual, Weibull event times, and exponential loss times (Beta parameters: $\alpha = 1$, $\beta = 3$).

Event distribution arm 1	Event distribution arm 2	Loss to follow-up rate		
		Predicted event time (yrs)		
Weibull(τ_{E1}, ϕ_{E1})	Weibull(τ_{E2}, ϕ_{E2})	3%	5%	7%
($\tau_{E1} = 1, \phi_{E1} = 25$)	($\tau_{E2} = 1, \phi_{E2} = 17$)	2.13	2.14	2.15
	($\tau_{E2} = 0.8, \phi_{E2} = 17$)	1.77	1.78	1.79
	($\tau_{E2} = 0.8, \phi_{E2} = 30$)	2.06	2.07	2.08
($\tau_{E1} = 1.1, \phi_{E1} = 25$)	($\tau_{E2} = 1, \phi_{E2} = 17$)	2.25	2.27	2.28
	($\tau_{E2} = 0.8, \phi_{E2} = 17$)	2.84	1.85	1.86
	($\tau_{E2} = 0.8, \phi_{E2} = 30$)	2.19	2.20	2.21
($\tau_{E1} = 1.3, \phi_{E1} = 15$)	($\tau_{E2} = 1, \phi_{E2} = 17$)	2.23	2.24	2.26
	($\tau_{E2} = 0.8, \phi_{E2} = 17$)	1.85	1.85	1.86
	($\tau_{E2} = 0.8, \phi_{E2} = 30$)	2.17	2.18	2.19

3.3.3 Simulation studies of event-time prediction for study monitoring

When a clinical trial is underway, the conditional model may be used at an interim data cutoff time to obtain point predictions and prediction intervals for future event occurrence times. The performance of this method under various distributional assumptions is demonstrated here through simulation studies. Parameters for the simulated clinical trial scenarios were chosen to be similar in size and scope to observed data in the HPTN 052 Study.

Poisson accrual and exponential event and loss

Suppose enrollment follows a homogeneous Poisson process with constant rate α_P, and that event and loss times are exponentially distributed with rate parameters γ_E and γ_L, respectively. In this scenario the hazard rates for the event and for dropout are both constant, and the conditional model for event-occurrence times, equation (3.4), simplifies to

$$
\begin{aligned}
&G^c(R;\widetilde{W}_n, N, \theta|\mathscr{F}_C) = \qquad (3.7)\\
&\quad N_C + \left(\sum_{i=1}^{n_C} \delta_i^A(1-\delta_i^E)(1-\delta_i^L)\right) \cdot \frac{\gamma_E}{\gamma_E+\gamma_L} \cdot \Big[1 - \exp\big\{-(\gamma_E+\gamma_L)(R-C)\big\}\Big]\\
&\quad + I(n_C < n) \cdot \frac{\alpha_P \gamma_E}{(\gamma_E+\gamma_L)^2}\\
&\quad \cdot \Big[(U-C)(\gamma_E+\gamma_L) + \exp\big\{-(\gamma_E+\gamma_L)(R-C)\big\} - \exp\big\{-(\gamma_E+\gamma_L)(R-U)\big\}\Big] - N,
\end{aligned}
$$

where $\theta = \{\alpha_P, \gamma_E, \gamma_L\}$, and $U = \min(\widetilde{W}_n, R)$.

The inter-accrual times, $W_i^* = W_i - W_{i-1}$, $W_0 \equiv 0$, for $i = 1, 2, \ldots, n_C$, are iid exponential random variables with rate α_P. Thus the likelihood for the full data simplifies to

$$
L_n(\theta) = \prod_{i=1}^{n_C} \Big[\alpha_P \exp(-\alpha_P w_i^*) \cdot \gamma_E^{\delta_i^E} \cdot \gamma_L^{\delta_i^L} \cdot \exp(-\gamma_E x_i) \cdot \exp(-\gamma_L x_i)\Big].
$$

The ML estimates $\widehat{\theta} = (\widehat{\alpha}_P, \widehat{\gamma}_E, \widehat{\gamma}_L)$, are calculated from the likelihood and substituted for α, γ_E, and γ_L in equation (3.7). If enrollment is still underway at the data cutoff, C, $\widetilde{W}_n = E[W_n]$ is estimated by n/α_P. Otherwise, the value of $\widetilde{W}_n$ is the observed end of the accrual period, w_n.

We simulated data for a clinical trial comprised of 1,750 patients enrolled over approximately three years and followed throughout an observation period. Patients were enrolled according to a Poisson process with rate $\alpha_P = 600$ (approximately 600 patients per year). Each patient was susceptible to experiencing an observed event or being lost to follow-up once enrolled. Event times and loss times were generated for each patient as exponential random variables with rate $\gamma_E = 0.06$ (6% per year), and $\gamma_L = 0.03$ (3% per year), respectively. A patient's observation terminated at the earliest of their event time, dropout time, or the data cutoff time, C. The total desired number of observed monitoring events was $N = 340$. Thus the landmark 85th , 170th , 255th , and 340th events marked the information fractions 0.25, 0.50, 0.75, and 1.00, respectively. We simulated 5,000 data sets, and at cutoff times, C, corresponding to information fractions of 0.10 (34 events), 0.25 (85 events), 0.50 (170 events), and 0.75 (255 events), we obtained point predictions, $\widehat{E[R_N]}$, and standard error estimates for the time of each of the landmark numbers of events. Simulation results for this scenario are shown in Table 3.3.

The prediction method is shown to be both accurate and precise when the model is appropriately specified. The point predictions obtained for event occurrence times have little or no bias when compared to the true expected Nth order statistic, R_N. The accuracy of point predictions improves with increasing accumulated trial data. We also found the predictions of $E[R_N]$ to be consistent for the true value using increasing study sample sizes n (tables not included). The (model-based) mean estimated standard error for each $\widehat{E[R_N]}$ is based on asymptotics, and is slightly lower than the observed, finite-sample, empirical standard error of the point predictions. As a result, the method yields Normal approximation prediction interval coverage probabilities that are slightly anticonservative but very close to the nominal 95%. Because the prediction interval formula relies on asymptotics, the coverage probability approaches 95% as the trial sample size increases.

Table 3.3: Simulation results (in years) for event time prediction assuming Poisson process accrual and exponential event and loss times. (True parameter values: $\alpha_P = 600$, $\gamma_E = 0.06$, and $\gamma_L = 0.03$.)

	Information fraction cutoff (average years from start of study)															
	0.10 (1.11)				0.25 (1.75)				0.50 (2.53)				0.75 (3.15)			
Parameter	Bias	SSE	MSE	CP	Bias	SSE	MSE	CP	Bias	SSE	MSE	CP	Bias	SSE	MSE	CP
$\widehat{\alpha}_P$	1.51	23.55	20.82	0.95	0.96	18.97	18.54	0.95	0.73	15.62	15.43	0.95	0.42	14.44	14.35	0.95
$\widehat{\gamma}_E$	0.00	0.02	0.02	0.95	0.00	0.01	0.01	0.94	0.00	0.01	0.01	0.95	0.00	0.01	0.01	0.95
$\widehat{\gamma}_L$	0.00	0.01	0.01	0.93	0.00	0.01	0.01	0.94	0.00	0.01	0.01	0.95	0.00	0.00	0.00	0.95
$\widehat{E[R_{85}]}$	-0.01	0.15	0.09	0.94	—	—	—	—	—	—	—	—	—	—	—	—
$\widehat{E[R_{170}]}$	-0.01	0.22	0.17	0.94	0.00	0.14	0.07	0.94	—	—	—	—	—	—	—	—
$\widehat{E[R_{255}]}$	0.00	0.30	0.26	0.94	0.00	0.19	0.13	0.94	0.00	0.14	0.05	0.94	—	—	—	—
$\widehat{E[R_{340}]}$	0.00	0.42	0.38	0.93	0.00	0.27	0.21	0.94	0.00	0.19	0.10	0.94	0.00	0.16	0.04	0.94

NOTE: Bias, sample mean of the differences between predictions and their true values; SSE, sample standard error of the estimates; MSE, mean of estimated standard errors; CP, nominal 95% coverage probabilities of prediction intervals; $\widehat{E[R_{85}]}$, $\widehat{E[R_{170}]}$, $\widehat{E[R_{255}]}$, $\widehat{E[R_{340}]}$, expected times of the 85th, 170th, 255th, and 340th events, corresponding to information fractions 0.25, 0.50, 0.75 and 1.00, respectively.

Exponential enrollment and Weibull event and loss times

Suppose we model enrollment as simultaneous iid exponential random variables with rate parameter α_e. Here the rate of patient accrual is assumed to be rapid at the start of the study and decreasing over time, as in Figure 3.2(a). Assume also that event times follow a Weibull distribution with shape parameter τ_E and scale parameter ϕ_E, and loss times follow a Weibull distribution with shape parameter τ_L and scale parameter ϕ_L. The conditional model for event occurrence times, following equation (3.4), is

$$
\begin{aligned}
&G^c(R;\widetilde{W}_n, N, \theta|\mathscr{F}_C) = \qquad (3.8)\\
&N_C + \sum_{i=1}^{n_C}\Bigg(\delta_i^A\cdot(1-\delta_i^E)\cdot(1-\delta_i^L)\cdot\int_C^R \frac{\tau_E}{\phi_E^{\tau_E}}\cdot(u-w_i)^{\tau_E-1}\\
&\cdot\exp\left[\frac{1}{\phi_E^{\tau_E}}\cdot\left\{(c-w_i)^{\tau_E}-(u-w_i)^{\tau_E}\right\}+\frac{1}{\phi_L^{\tau_L}}\cdot\left\{(c-w_i)^{\tau_L}-(u-w_i)^{\tau_L}\right\}\right]du\Bigg)\\
&+I(n_C<n)\cdot\int_C^U \alpha_e n\exp(-\alpha_e v)\int_0^{R-v}\frac{\tau_E}{\phi_E^{\tau_E}}\cdot u^{\tau_E-1}\\
&\cdot\exp\left\{-\left(\frac{u}{\phi_E}\right)^{\tau_E}-\left(\frac{u}{\phi_L}\right)^{\tau_L}\right\}dudv - N.
\end{aligned}
$$

where $\theta=\{\alpha_e,\tau_E,\phi_E,\tau_L,\phi_L\}$, and $U=\min(\widetilde{W}_n,R)$.

When we assume that accrual times are iid as $f_W(t)$, e.g. iid exponential, the enrollment times observed in $[0,C]$ come from a right-truncated version of the accrual distribution,

$$Pr(W=w_i|0\le W\le C)=\frac{f_W(w_i)}{F_W(C)},$$

with truncation time C. Thus the likelihood for the full data in the scenario described above is

$$
\begin{aligned}
L_n(\theta)=\prod_{i=1}^{n_C}\Bigg[&\frac{\alpha_e\exp(-\alpha_e w_i)}{1-\exp(-\alpha_e C)}\cdot\left\{\frac{\tau_E}{\phi_E}\left(\frac{t}{\phi_E}\right)^{(\tau_E-1)}\right\}^{\delta_i^E}\cdot\left\{\frac{\tau_L}{\phi_L}\left(\frac{t}{\phi_L}\right)^{(\tau_L-1)}\right\}^{\delta_i^L}\\
&\cdot\exp\left(-\frac{t}{\phi_E}\right)^{\tau_E}\cdot\exp\left(-\frac{t}{\phi_L}\right)^{\tau_L}\Bigg].
\end{aligned}
$$

The ML estimates, $\widehat{\theta}=(\widehat{\alpha}_e,\widehat{\tau}_E,\widehat{\phi}_E,\widehat{\tau}_L,\widehat{\phi}_L)$, for the enrollment, event, and loss time distribution parameters are substituted for the values of θ in equation (3.8). The ML estimate of the rate parameter, $\widehat{\alpha}_e$, for the truncated exponential distribution is nonzero if and only if

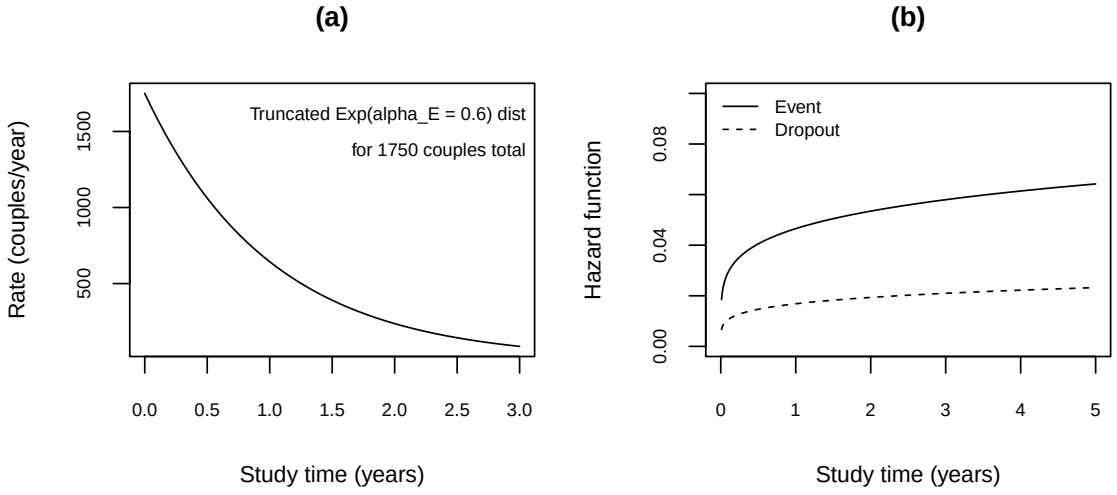
the mean of the observed enrollment times is less than half of the truncation time [16], i.e.

$$\frac{1}{n_C}\sum_{i=1}^{n_C} w_i < \frac{C}{2}. \tag{3.9}$$

When condition (3.9) does not hold, the ML estimate of α_e is zero, which is clearly inappropriate as an accrual rate. Thus if (3.9) does not hold, the investigator should rethink the choice of the truncated exponential for modeling the enrollment distribution.

If enrollment is complete by the data cutoff, C, $\widetilde{W}_n$ is the observed end of the accrual period, w_n. Otherwise, use $\widetilde{W}_n = E[W_n]$ estimated by $1/\widehat{\alpha}_e \sum_{i=1}^{n} 1/i$, the expected value of the maximum order statistic of n iid exponential random variables.

Figure 3.2: Distribution and parameter selections for Study Monitoring Example 2. (a) Instantaneous accrual rate function derived from simultaneous exponential distributions. (b) Hazard functions for selected Weibull distributions for event times and dropout times (both arms pooled).



Simulation study results for predictions in this scenario are shown in Table 3.4. Here we generated data for $n = 1,750$ patients with exponential enrollment times (rate $\alpha_e = 1$), Weibull event times (shape $\tau_E = 1.2$, scale $\phi_E = 15$), and Weibull loss times (shape $\tau_L = 1.2$, scale $\phi_L = 35$). In this scenario, the hazards for both event and dropout gradually increase throughout the course of follow-up, shown in Figure 3.2(b). As in the previous example, each patient's observation time was the earliest of their event time, dropout time, or the data

cutoff time, C, and the landmark 85th, 170th, 255th, and 340th observed events marked each quarter of the desired statistical information. We simulated 5,000 data sets, and at cutoff times corresponding to information fractions of 0.10, 0.25, 0.50, and 0.75, we obtained point predictions, $\widehat{E[R_N]}$, and prediction intervals for the time of each of the landmark numbers of events. Results are shown in Table 3.4.

Once again, for the properly specified model, the bias of point predictions $\widehat{E[R_N]}$ is small and becomes negligible throughout the study as the available data increases. Because more parameters are estimated here than in the previous example, the individual estimates are more variable and have greater finite sample bias. Weibull scale parameters, in particular, require a large amount of data to be accurately estimated. Nevertheless, point predictions from the full conditional prediction model have little sensitivity to bias in individual parameter estimates. Prediction interval coverage approaches 95% with increasing available data for a fixed sample size, as well as with increasing overall sample size (results not shown). Lower than nominal coverage at early cutoff times results from model-based asymptotic standard error estimates that are, on average, underestimates of the true finite-sample standard error of the point predictions.

Table 3.4: Simulation results (in years) for event time prediction assuming simultaneous exponential accrual and Weibull event and loss times. (True parameter values: $\alpha_e = 1$, $\tau_E = 1.2$, $\phi_E = 15$, $\tau_L 1.2$, and $\phi_L = 35$.)

	Information fraction cutoff (average years from start of study)															
	0.10 (1.26)				0.25 (2.08)				0.50 (3.22)				0.75 (4.29)			
Parameter	Bias	SSE	MSE	CP	Bias	SSE	MSE	CP	Bias	SSE	MSE	CP	Bias	SSE	MSE	CP
$\widehat{\alpha}_e$	0.00	0.08	0.08	0.95	0.00	0.05	0.05	0.96	0.00	0.03	0.03	0.96	0.00	0.03	0.03	0.96
$\widehat{\tau}_E$	0.05	0.20	0.19	0.95	0.02	0.13	0.12	0.95	0.01	0.09	0.09	0.95	0.01	0.07	0.07	0.95
$\widehat{\phi}_E$	0.82	17.93	7.72	0.86	0.10	15.58	3.79	0.91	-0.03	15.12	2.18	0.93	-0.04	15.04	1.54	0.94
$\widehat{\tau}_L$	0.09	1.35	0.34	0.95	0.03	1.25	0.21	0.95	0.01	1.22	0.15	0.95	0.01	1.21	0.12	0.95
$\widehat{\phi}_L$	$2.42x10^2$	$13.11x10^3$	$17.12x10^2$	0.81	7.95	56.68	28.09	0.87	2.78	40.64	13.64	0.91	1.56	37.84	9.42	0.93
$\widehat{E[R_{85}]}$	-0.02	0.24	0.17	0.92	—	—	—	—	—	—	—	—	—	—	—	—
$\widehat{E[R_{170}]}$	-0.02	0.53	0.49	0.91	-0.01	0.25	0.17	0.94	—	—	—	—	—	—	—	—
$\widehat{E[R_{255}]}$	0.02	0.96	0.94	0.90	-0.01	0.42	0.36	0.93	-0.01	0.24	0.11	0.94	—	—	—	—
$\widehat{E[R_{340}]}$	0.09	1.55	1.68	0.89	-0.01	0.65	0.59	0.93	-0.01	0.35	0.24	0.94	0.00	0.25	0.09	0.94

NOTE: Bias, sample mean of the differences between predictions and their true values; SSE, sample standard error of the estimates; MSE, mean of estimated standard errors; CP, nominal 95% coverage probabilities of prediction intervals; $\widehat{E[R_{85}]}$, $\widehat{E[R_{170}]}$, $\widehat{E[R_{255}]}$, $\widehat{E[R_{340}]}$, expected times of the 85th, 170th, 255th, and 340th events, corresponding to information fractions 0.25, 0.50, 0.75 and 1.00, respectively.

3.4 A Real data example: The HIV Prevention Trials Network 052 Study

3.4.1 Description of the data and monitoring plan

As introduced earlier, the HIV Prevention Trials Network 052 (HPTN 052) Study is a Phase III, multi-continent, randomized clinical trial comparing the effectiveness of immediate versus delayed antiretroviral (ART) therapy strategies on sexual transmission of HIV-1 [10][11]. When this study rolled out in eight countries on three continents, many of its design parameters, such as patient enrollment and retention rates and baseline incidence rates of HIV transmission and other clinical outcomes, deviated from the anticipated values. Nevertheless, timely Data and Safety Monitoring Board (DSMB) reviews of trial data were vital and ultimately led to an important discovery in the effort to control the AIDS epidemic: immediate ART reduced genetically linked HIV transmissions by 96% compared with a delayed strategy. The DSMB recommended immediate disclosure of the results to trial participants and the general public [11]. The study is described in more detail in section 1.2.

The HPTN 052 trial enrolled serodiscordant couples, in which the HIV-1-infected partner had CD4 (cluster of differentiation 4) cell counts between 350 and 550 cells/mm^3. Couples were randomly assigned to either the immediate ART therapy or the delayed ART therapy arm. Patients on the delayed arm received ART after two consecutive measurements of their CD4 cell counts fell below 250 cells/mm^3, or when they developed AIDS-defining illness, as defined in World Health Organization (WHO) guidelines.

The DSMB's monitoring of trial continuation included analysis of a composite morbidity and mortality (M/M) endpoint comprised of the earliest of either the incidence of genetically linked HIV transmission to the HIV-negative partner (prevention endpoint) or the occurrence in the HIV-positive partner of pulmonary tuberculosis, severe bacterial infection, a WHO stage 4 event, or death (clinical endpoint). The planned schedule of interim analyses for trial continuation was governed by the total number of observed composite M/M endpoints among all enrolled couples. The interim safety monitoring analysis plan was based on a Lan-DeMets implementation of an O'Brien-Fleming monitoring boundary [17][41] in which three interim analyses and one final analysis were planned when the total number of composite endpoints reached 85, 170, 255, and 340.

This study demonstrates the difficulty in obtaining accurate pretrial estimates of data accumulation patterns and landmark event times for interim analyses, despite extensive existing literature on the disease progression and treatment options for HIV-1. Differences in expected versus observed enrollment rates and rates of transmissions, clinical events, and patient dropout are detailed in section 1.2. At the time these methods were developed, data were available through the observation of the 86th composite M/M event, the information cutoff for the first interim safety monitoring analysis (0.25). This was approximately 3.7 years after the start of the full study. We will use the HPTN 052 data accumulated up to this 0.25 information fraction cutoff to demonstrate how our estimation method could be applied periodically throughout data collection to anticipate the calendar time of future landmark events, based on the trends observed in the trial itself.

After screening, a total of 1,763 HIV-1-serodiscordant couples were enrolled. For the purpose of demonstrating our method, we excluded 82 couples that were part of a pilot phase that began in April 2005. Enrollment for the full study took place between June 2007 and May 2010 at sites in eight countries (Pune and Chennai, India; Chiang Mai, Thailand; Harare, Zimbabwe; Porto Alegre and Rio de Janeiro, Brazil; Lilongwe and Blantyre, Malawi; Johannesburg, South Africa; Gaborone, Botswana; Soweto, South Africa; and Kisumu, Kenya). The time to composite event or loss-to-follow-up was missing for seven of the couples. Our analyses are based on data collected for the remaining 1,674 couples between June 2007 and the data cutoff in February 2011. The event, dropout, and administrative censoring times for couples at each of the sites are shown in Figure 3.3. Data is only shown for twelve of the thirteen study sites because the site in Boston was only operational during the pilot phase, and not during the full study. Couples' observation times were censored as of their last clinical visit before the February 2011 cutoff, which accounts for the variability among administrative censoring times in the figure.

To test the utility of our likelihood-based approach we compared our model predictions with observed composite event times in the HPTN 052 data. Each time-to-event model is based on data accrued from the start of enrollment to a data cutoff time dictated by an information fraction benchmark. Eighty-six composite events were observed before the February 2011 data cutoff (3.69 years into the study). A total of 340 composite M/M end-

Figure 3.3: Observed morbidity and mortality (M/M) composite event times and loss-to-follow-up times in the HPTN 052 Study

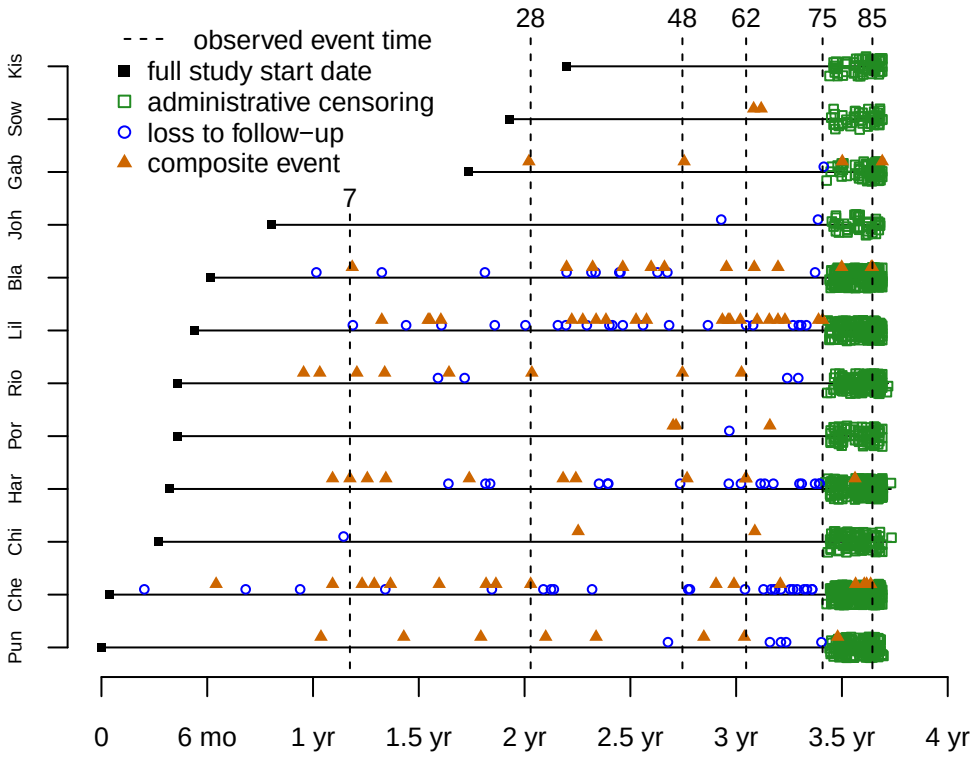


Table 3.5: Summary of HPTN 052 Study data available at each cutoff time

	Information fraction cutoff (years from study start)				
	0.02 (1.17)	0.08 (2.03)	0.14 (2.75)	0.18 (3.04)	0.22 (3.41)
Couples enrolled (% of total sample)	466 (28%)	1029 (61%)	1611 (96%)	1674 (100%)	1674(100%)
Total observed person-years	199	784	1681	2153	2709
Observed composite endpoints	7	28	48	62	75
Couples lost to follow-up	5	19	43	51	85

points were anticipated, and information fractions of 0.14 (48 events), 0.18 (62 events), 0.22 (75 events), and 0.25 (85 events) occurred at 2.75, 3.04, 3.41, 3.64 years, respectively, after the start of enrollment. We treated these as the landmark events whose timing we wished to predict. Event occurrence time predictions were calculated at information fractions of 0.02 (7 events), 0.08 (28 events), 0.14, 0.18, and 0.22, corresponding to cutoff times of $C =$ 1.17, 2.03, 2.75, 3.04, and 3.41 years. Table 3.5 summarizes the data available for prediction calculations at each cutoff time. The data accumulated prior to each time C was used to calculate ML estimates of the distribution parameters and point predictions and 95% prediction intervals for the event times R_{48}, R_{62}, R_{75}, and R_{85}.

3.4.2 Parametric model selection

At the time of each prediction, we can use empirical estimates from the observed data and knowledge of study design features to select appropriate distributions for accrual, event, and loss times. The trial was initially planned to run at nine study sites. By 1.17 years into the study, our first information fraction cutoff time, all nine sites had opened and been operational for at least four months, and approximately 28% of the study population had been enrolled. At $C = 1.17$ and at the subsequent information fraction benchmark, $C = 2.03$, we modeled continuing enrollment as a Poisson process with constant rate equal to the sum of the estimated enrollment rates at each of the nine sites.

Due to slower than expected enrollment, three additional study sites were opened approximately two years into the trial. By 2.75 years into the study, information fraction cutoff 0.14, all twelve sites were operational, and approximately 72% of the study population had been enrolled. For this cutoff, we modeled continuing enrollment as a Poisson process with constant rate equal to the sum of the estimated enrollment rates at each of the twelve sites. At the 0.18 and 0.22 information cutoffs, $C = 3.04$ and 3.41 years into the study, enrollment had concluded; thus no accrual model was necessary for further event-time predictions.

A variety of graphical methods, goodness-of-fit tests, and methods for comparing survival distributions are available to assist in the selection of parametric time-to-event models, as referenced in section 3.3.1. Biological plausibility should be considered in combination with statistical tools. First, it is useful to determine whether a chosen parametric family of time-to-event distributions is appropriate, versus a nonparametric alternative. Hazard plotting is a simple preliminary method to visually assess the strength of evidence that a set of censored survival data follows a particular model [35]. For the Weibull distribution, we note that the cumulative hazard function has the form

$$\Lambda(t) = \left(\frac{t}{\phi}\right)^{\tau}. \tag{3.10}$$

Solving for t and taking the logarithm yields the equation

$$\log t = \log \phi + \frac{1}{\tau} \log \Lambda(t). \tag{3.11}$$

Thus the log of the survival time is a linear function of the log of the cumulative hazard. If the time-to-event data is truly Weibull, a plot of the log observed survival times against the log observed cumulative hazards at each event time will be linear. Hazard plots for the observed time-to-event at each information fraction cutoff time are shown in Figure 3.4. A lowess smoother is fit to each scatterplot to help determine the deviation from linearity. Also, a line indicating the log cumulative hazards according to the ML estimates of the Weibull shape and scale parameters is shown for comparison. With the exception of the earliest cutoff time, at which there are very few observed events (or dropouts) to analyze, the hazard plots appear approximately linear. The lowess curve and the model-based Weibull curve nearly coincide, which also indicates the distribution is well approximated by a Weibull

model. The corresponding plots for time-to dropout yield similar results and are shown in Figure 3.5. In both the event- and dropout-time hazard plots, the strength of the linear relationship increases after the earliest few observed failures.

Analytic methods may also be used to inform the decision to model event and dropout distributions as Weibull. Bagdonavicius, et al., describe modified chi-square goodness-of-fit tests which may be used to determine whether a parametric family of time-to-event distributions is appropriate, versus a nonparametric distribution, in the context of censored data [2]. Consider the null hypothesis,

$$H_0 : F(x) \in \mathscr{F}_0 = \{F_0(x; \theta = (\tau, \phi), \theta \in \Theta = \mathbb{R}^+ \mathrm{x} \mathbb{R}^+\},$$

that the true event-time distribution is a member of the Weibull family. First, we divide the interval of observation times $[0, x_{(n_C)}]$ into k subintervals, $I_i = (a_{j-1}, a_j]$, $a_0 = 0$, $a_k = x_{(n_C)}$. The choice of k is left to the investigator. General guidelines state that all intervals should have a nonzero expected event count, and at least 80% of intervals should have expected counts of five or higher under H_0 [4]. An algorithm for choosing the interval endpoints, a_j, meant to produce intervals in which the expected number of events is approximately equal, is given by the authors [2]. A test statistic X^2 is derived based on the vector of differences between the numbers of observed and expected events in the intervals $I_1, \ldots, I_k$. We reject H_0 with an approximate significance level α if $X^2 \geq \chi^2_\alpha(k-1)$.

The decisions to accept or reject H_0 based on this test at each data cutoff time C are shown in Figures 3.4 and 3.5. A caveat of the modified chi square test is that it is based on the asymptotic distribution of the test statistic, and therefore requires a large number of observed events for accuracy. The number of observed events influences both k and the bias and precision of the ML Weibull parameter estimates, on which the expected interval event counts depend.

The above graphical and analytic diagnostic methods support for the choice of Weibull event- and dropout-time distributions, with stronger evidence of Weibull distributions later in data collection. We could further refine our prediction model by performing a parametric test to determine the strength of evidence that the time-to-event data are exponentially distributed (Weibull shape parameter $\tau = 1$). Figure 3.6 and Figure 3.7 compare the cu-

Figure 3.4: Hazard plots and results of the nonparametric modified chi square test of the hypothesis that the distribution is Weibull for the observed event-time distribution at each information fraction cutoff time.

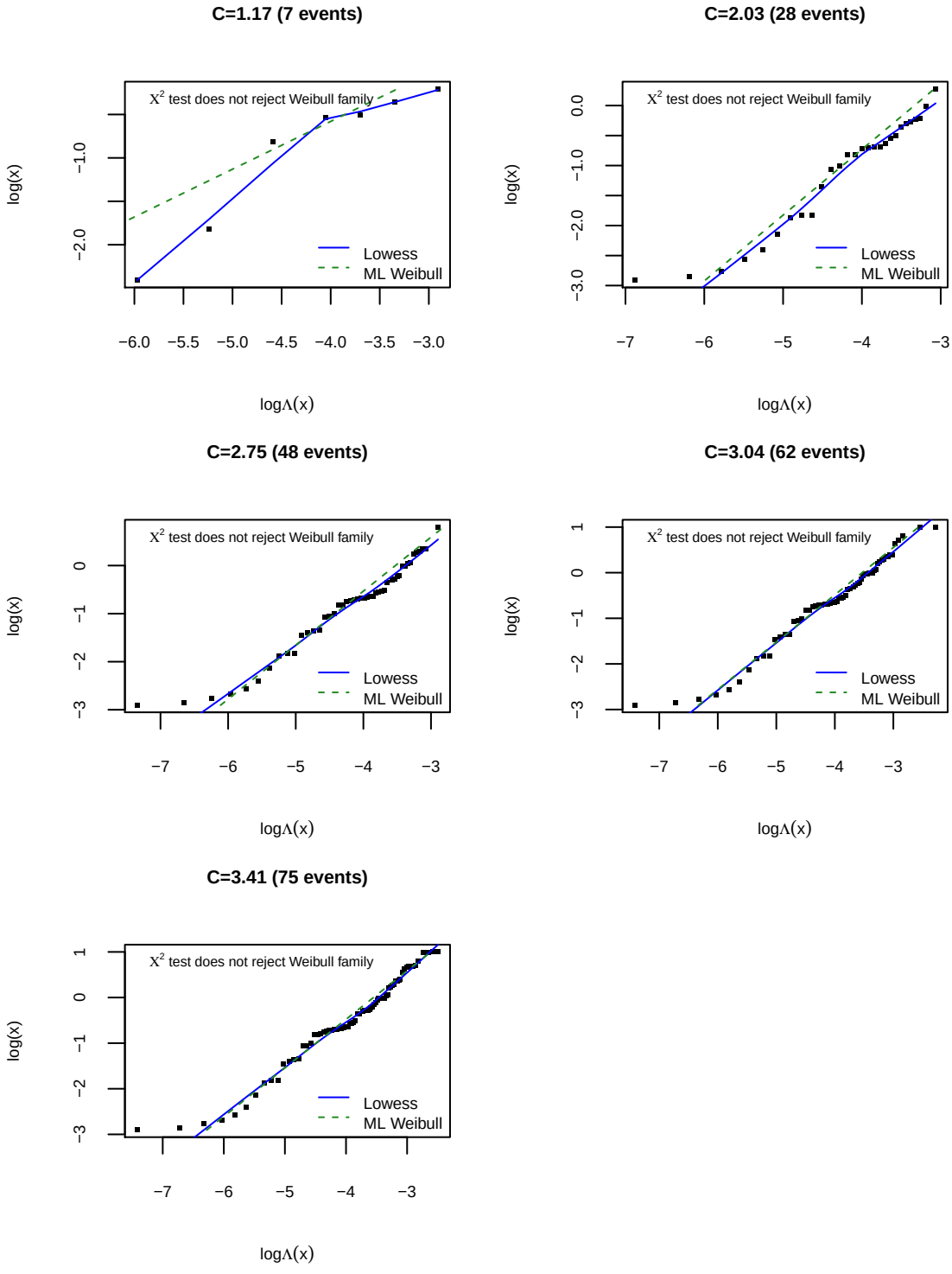
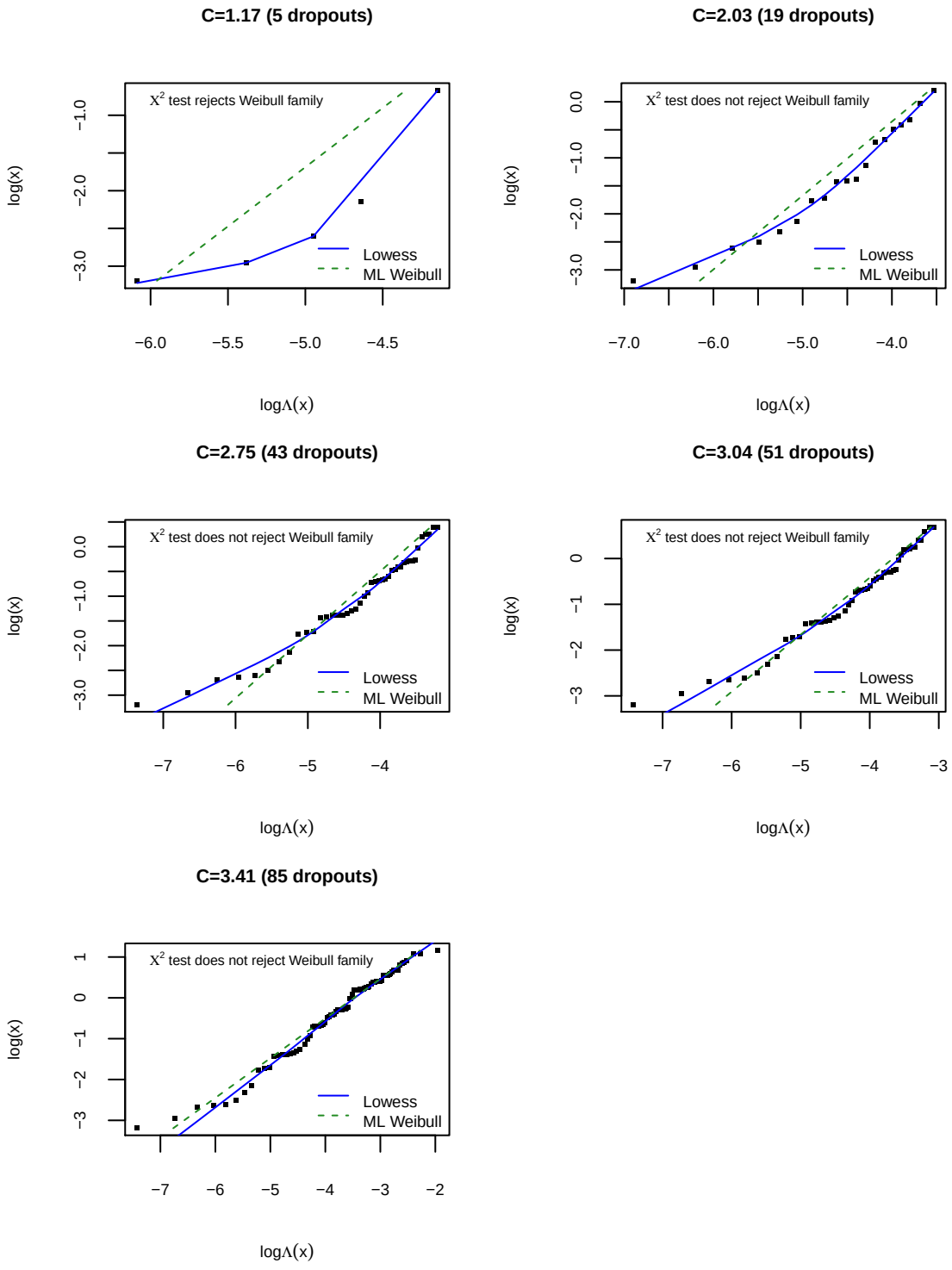


Figure 3.5: Hazard plots and results of the nonparametric modified chi square test of the hypothesis that the distribution is Weibull for the observed dropout-time distribution at each information fraction cutoff time..



mulative hazard curves for events and dropout, based on ML estimates of the exponential and Weibull parameters, with nonparametric Nelson-Aalen estimates of the observed cumulative hazards for events and dropouts, respectively. The corresponding (extrapolated) hazard functions are also shown. In nearly every case, a likelihood ratio test of nested models fails to reject the null hypothesis that the Weibull shape parameter $\tau = 1$, i.e. the distribution is in the exponential family. This is not surprising given the similarity of the estimated exponential and Weibull hazard functions.

Here we note that in general, when there is sufficient data to estimate parameters of a more flexible model (e.g. Weibull versus exponential), it is beneficial to do so. Although models with more parameters will always yield wider prediction intervals than simpler models, a potential decrease in bias may be worth sacrificing some efficiency. However if predictions are made early in the trial when data is limited, the less flexible model may be a more prudent choice because it helps guard against overfitting the data. We have included predictions at a very small information fraction (0.02) to illustrate this issue (see Figures 3.8 and 3.9, and the resulting prediction intervals in Figures 3.10 and 3.11). The estimated Weibull parameters at 0.02 (1.17 years) are strongly influenced by a cluster of events occurring approximately one year into the study for couples with similar observation times. The extrapolated hazard function is much larger than the true hazards, yielding too-early predictions. The exponential model is less sensitive to tightly clustered events because it is simply based on an overall average hazard rate. Nevertheless, the prediction methods described here are not meant to be used, and cannot be reliable if implemented, too early in data collection. Adequate data accumulation is required to assess the appropriateness of distributional families or to accurately estimate parameters of a chosen distribution.

For the purpose of demonstration we implement predictions using Weibull event and dropout distributions, as well as predictions using exponential distributions. The exponential predictions are based on the more restrictive assumption of constant hazards. The resulting point predictions and prediction intervals for the 48th, 62nd, 75th, and 85th composite events are shown in Tables 3.6 and 3.7. The predictions under each set of modeling assumptions are compared graphically for each landmark event in Figure 3.10, and for each information fraction prediction time in Figure 3.11. Predicted event times tend to be early

Figure 3.6: Observed and ML fitted cumulative hazard functions and corresponding hazard functions for events at data cutoffs of $C = 1.43, 2.24, 2.85$, and 3.16 years. (Λ = LR test statistic of H_0: underlying distribution is exponential, or Weibull shape parameter =1.)

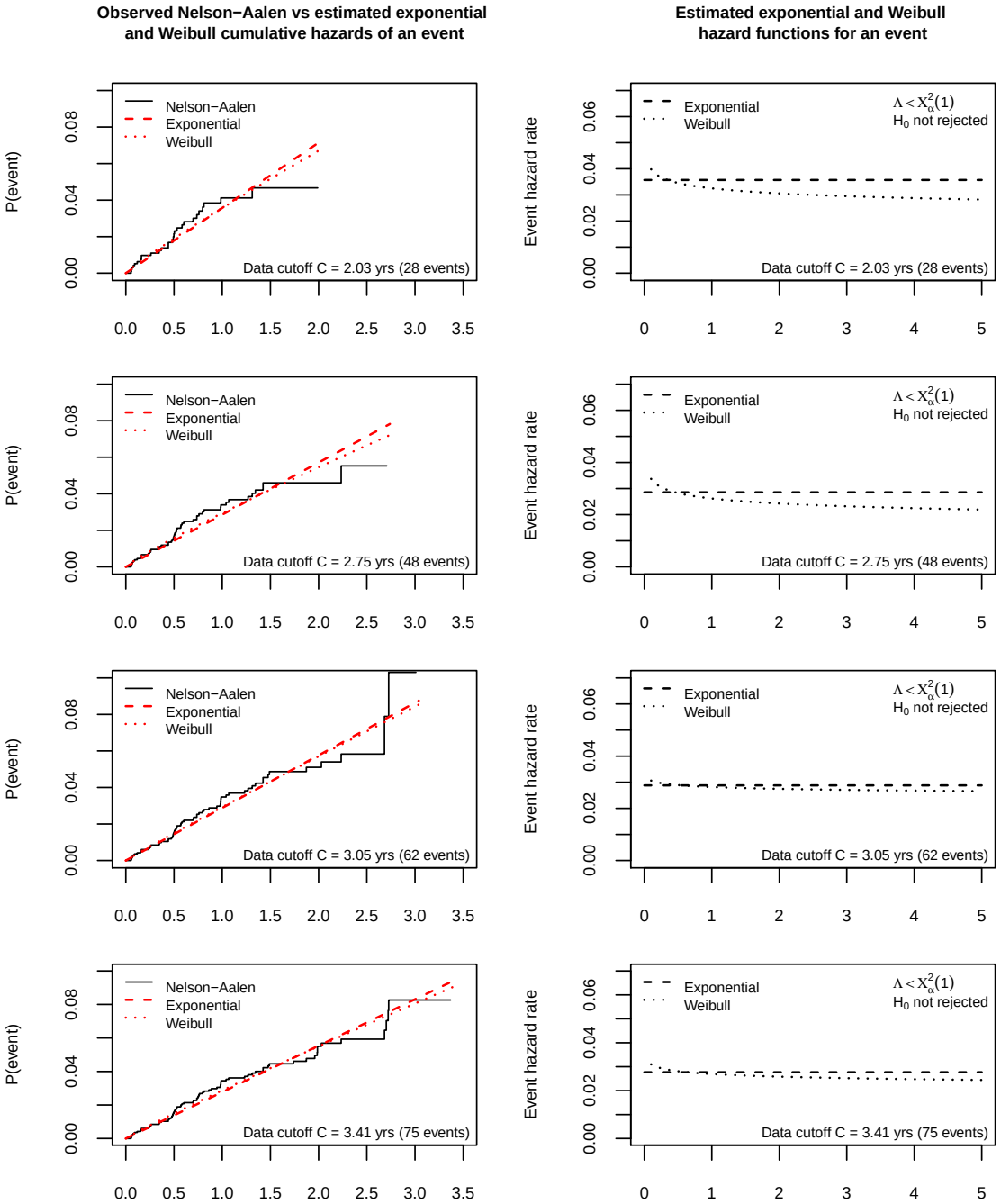


Figure 3.7: Observed and ML fitted cumulative hazard functions and corresponding hazard functions for loss-to-follow-up at data cutoffs of $C = 1.43, 2.24, 2.85$, and 3.16 years.

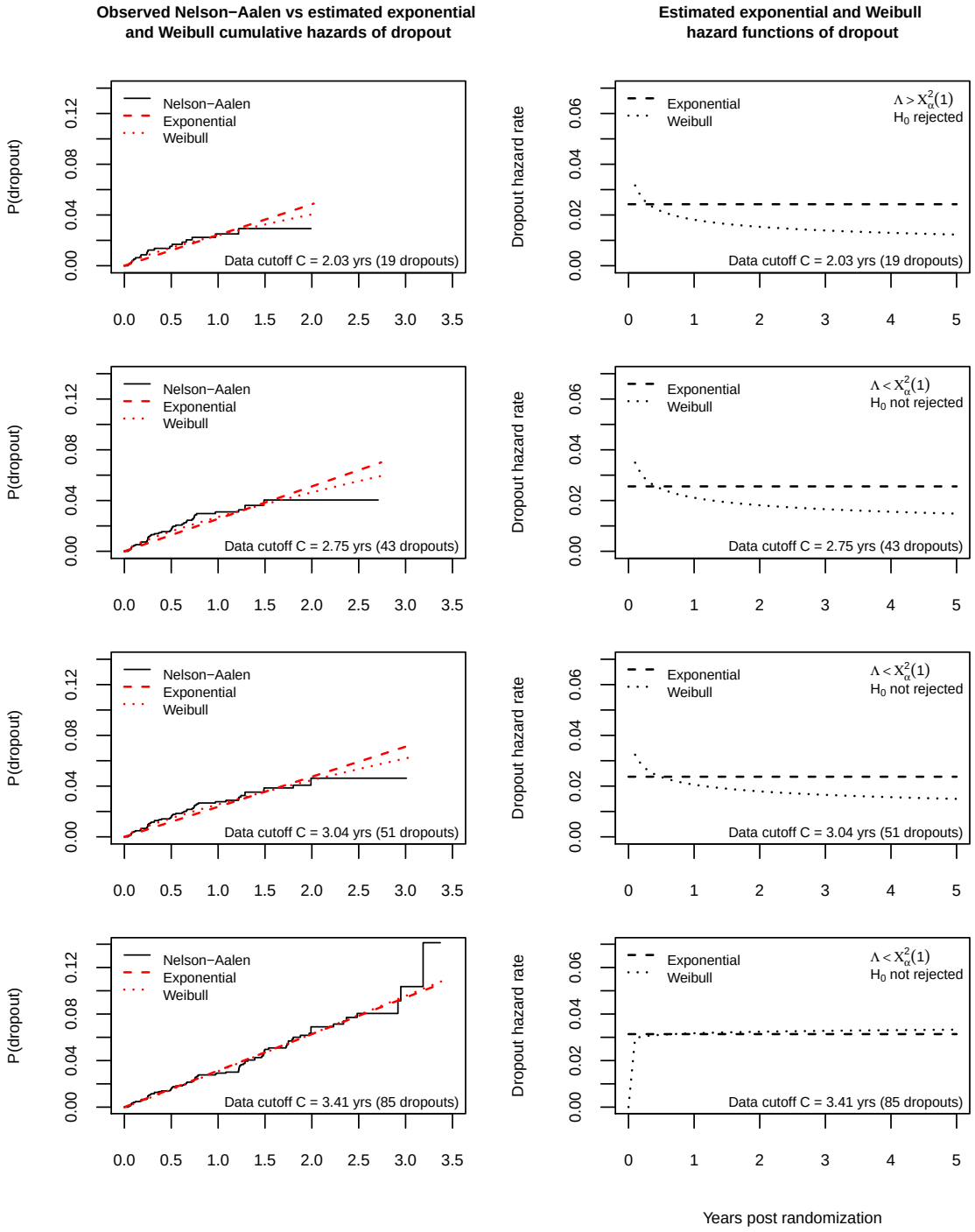


Figure 3.8: Observed and ML fitted cumulative hazard functions and corresponding hazard functions for events at data cutoff $C = 1.17$ years (7 observed events).

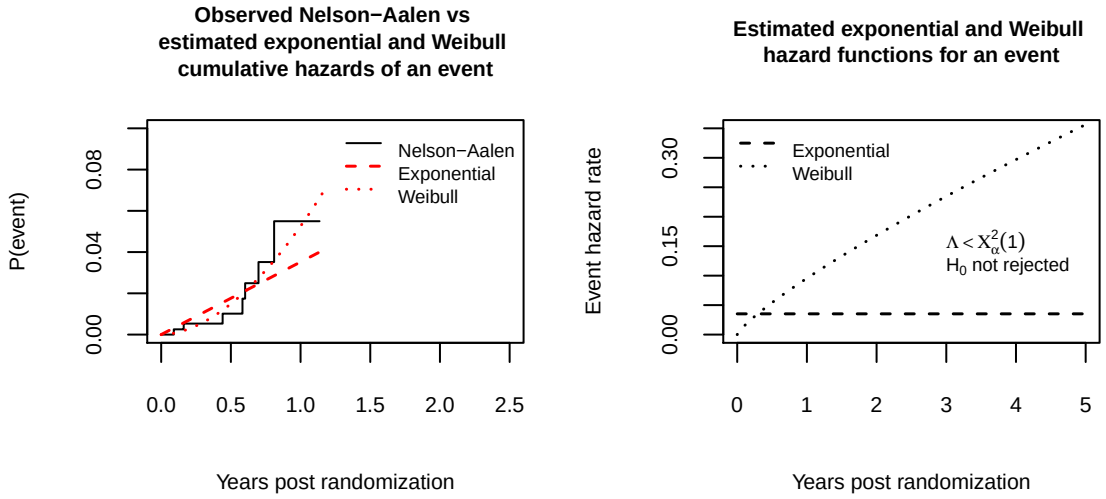


Figure 3.9: Observed and ML fitted cumulative hazard functions and corresponding hazard functions for dropout at data cutoff $C = 1.17$ years (5 observed dropouts).

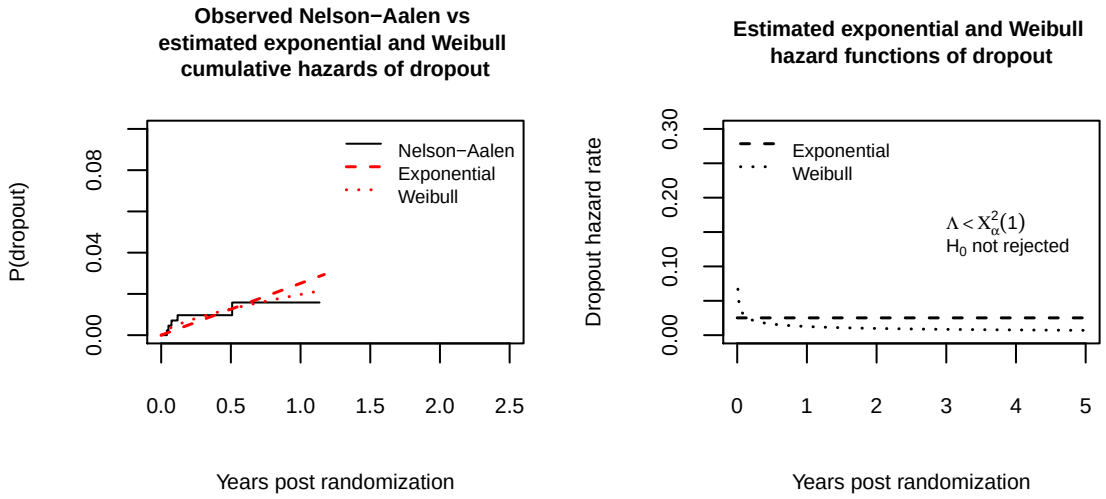


Table 3.6: Event-occurrence time predictions for the HPTN 052 Study measured in years from start of study, assuming Poisson enrollment times, exponential composite event times, and exponential loss-to-follow-up.

		Information fraction cutoff (years from start of study)				
		0.02 (1.17)	0.08 (2.03)	0.14 (2.75)	0.18 (3.04)	0.22 (3.41)
R_{48}	2.75					
$\widehat{E[R_{48}]}$		2.60	2.51	—	—	—
95% PI		(1.87, 3.32)	(2.40, 2.67)	—	—	—
R_{62}	3.04					
$\widehat{E[R_{62}]}$		2.92	2.80	3.06	—	—
95% PI		(2.06, 3.77)	(2.56, 3.03)	(2.97, 3.15)	—	—
R_{75}	3.41					
$\widehat{E[R_{75}]}$		3.18	3.03	3.35	3.34	—
95% PI		(2.22, 4.15)	(2.72, 3.34)	(3.18, 3.53)	(3.27, 3.41)	—
R_{85}	3.64					
$\widehat{E[R_{85}]}$		3.37	3.21	3.58	3.57	3.65
95% PI		(2.33, 4.42)	(2.83, 3.59)	(3.35, 3.82)	(3.44, 3.70)	(3.59, 3.70)

R_{48}, R_{62}, R_{75}, and R_{85} denote the observed times of the 48th , 62nd , 75th, and 85th M/M composite events, respectively. $\widehat{E[R_{48}]}$, $\widehat{E[R_{62}]}$, $\widehat{E[R_{75}]}$, and $\widehat{E[R_{85}]}$ denote the predicted times of these events.

throughout the trial and improve with more available data. By information fraction 0.14, prediction intervals consistently cover the true observed event time and are narrow enough to be useful for planning interim analyses, despite some degree of model misspecification. The Weibull model produces severely biased predictions at information fraction 0.02, but recovers quickly for subsequent predictions. Beginning at 2.03 years, the point predictions are essentially the same for both models, but the Weibull prediction intervals, being wider, are slightly more conservative and have better coverage of the true observed landmark event times.

Table 3.7: Event-occurrence time predictions for the HPTN 052 Study measured in years from start of study, assuming Poisson enrollment times, Weibull composite event times, and Weibull loss-to-follow-up.

		Information fraction cutoff (years from start of study)				
		0.02 (1.17)	0.08 (2.03)	0.14 (2.75)	0.18 (3.04)	0.22 (3.41)
R_{48}	2.75					
$\widehat{E[R_{48}]}$		2.08	2.54	—	—	—
95% PI		(1.51, 2.66)	(2.36, 2.72)	—	—	—
R_{62}	3.04					
$\widehat{E[R_{62}]}$		2.26	2.83	3.08	—	—
95% PI		(1.57, 2.94)	(2.55, 3.11)	(2.97, 3.20)	—	—
R_{75}	3.41					
$\widehat{E[R_{75}]}$		2.40	3.08	3.42	3.35	—
95% PI		(1.62, 3.18)	(2.69, 3.48)	(3.17, 3.66)	(3.25, 3.44)	—
R_{85}	3.64					
$\widehat{E[R_{85}]}$		2.50	3.29	3.68	3.58	3.66
95% PI		(1.66, 3.34)	(2.77, 3.80)	(3.32, 4.04)	(3.41, 3.76)	(3.59, 3.74)

R_{48}, R_{62}, R_{75}, and R_{85} denote the observed times of the 48th , 62nd , 75th, and 85th M/M composite events, respectively. $\widehat{E[R_{48}]}$, $\widehat{E[R_{62}]}$, $\widehat{E[R_{75}]}$, and $\widehat{E[R_{85}]}$ denote the predicted times of these events.

Figure 3.10: Comparison of prediction intervals assuming Poisson enrollment times and Weibull event and dropout distributions (PWW) versus assuming Poisson enrollmenttimes and exponential event and dropout distributions (PEE). Predictions are calculated at information fraction cutoff times of 0.02, 0.08, 0.14, 0.18, and 0.22.

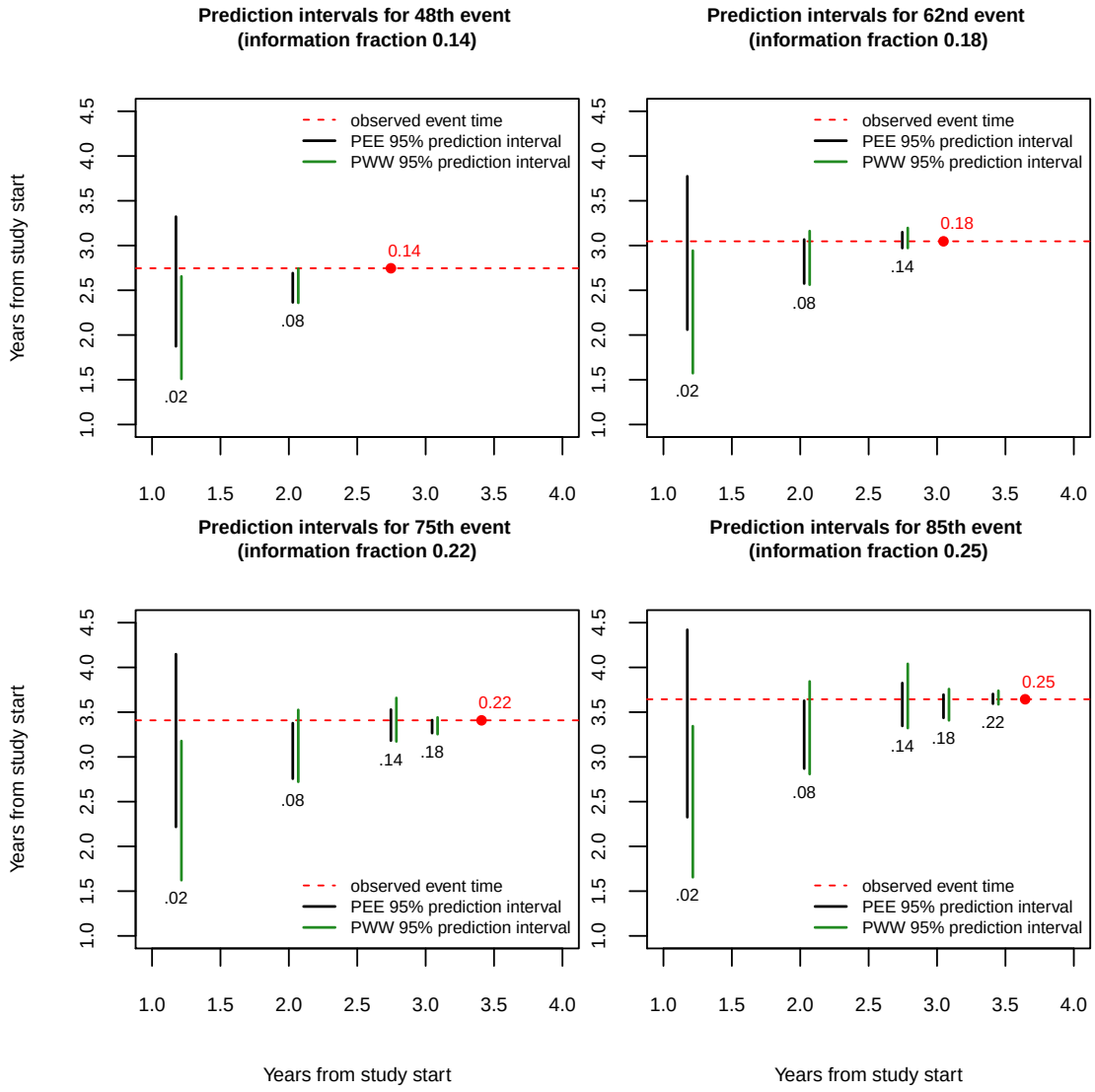
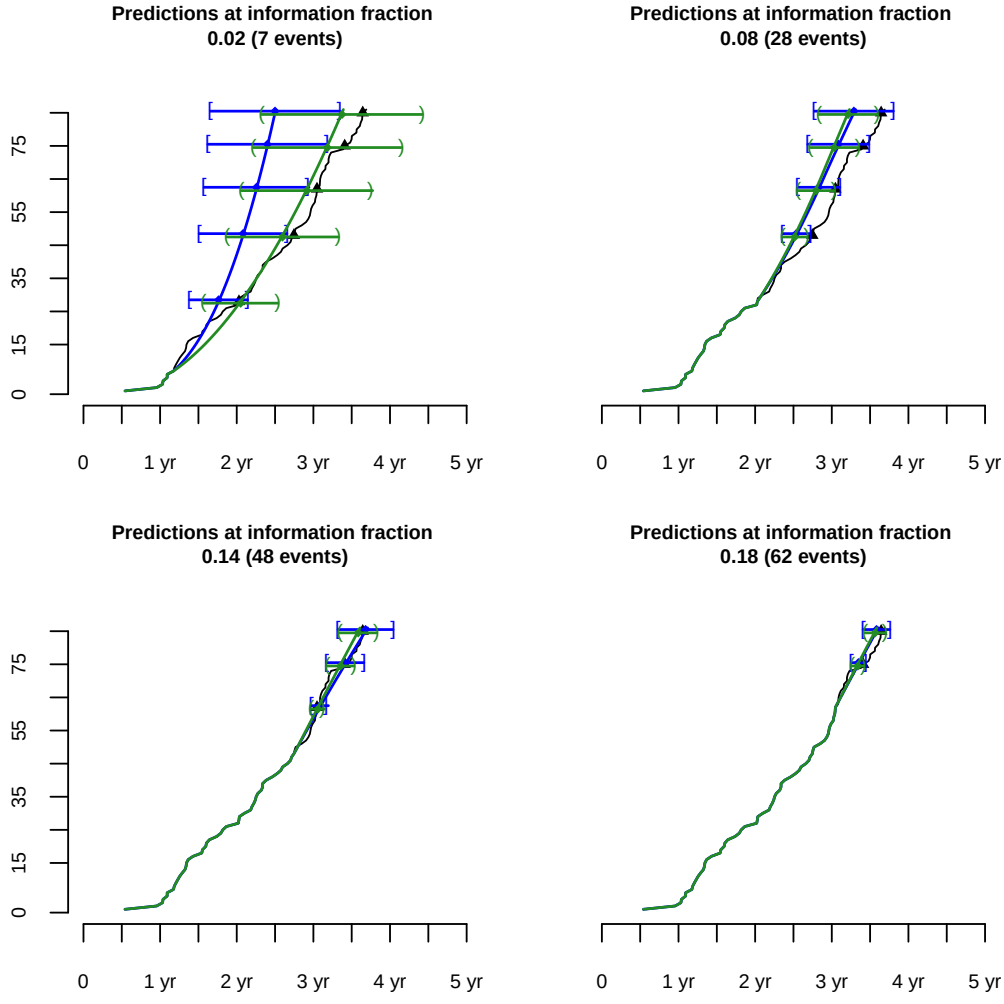


Figure 3.11: Comparison of prediction intervals assuming Poisson enrollment times and Weibull event and dropout distributions (PWW) versus assuming Poisson enrollment times and exponential event and dropout distributions (PEE). Predictions are calculated at information fraction cutoff times of 0.02, 0.08, 0.14, and 0.18.



3.5 Comparison of likelihood-based and Bayesian event-time prediction methods

Several of the previously published Bayesian methods for event-time prediction have been demonstrated on publicly available data from the Chronic Granulomatous Disease (CGD) Study [3][38][51][53]. Results of the original CGD trial may be found in *The New England Journal of Medicine*, and the data set is published [25][29]. The authors of some Bayesian prediction methods have made S-plus code for specific analyses available online [3][51][53]. Thus the CGD Study data provide a good basis for comparison of the performance of the parametric method described in this dissertation to other existing methods.

In this section we compare our likelihood-based predictions, assuming homogeneous Poisson process enrollment and exponential event and dropout times, to the corresponding Bayesian method described by Bagiella and Heitjan, which also assumes homogeneous Poisson enrollment and exponential events and loss-to-follow-up [3]. The results shown here may be directly compared to published results using the Bayesian Weibull method [51], the Bayesian nonparametric method [53], and a Bayesian method for data in which treatment arm is masked [38], all of which are applied to the CGD data. An additional comparison of the Bayesian exponential, Weibull, and nonparametric methods was published using data from the Randomized Evaluation of Mechanical Assistance for the Treatment of Congestive Heart Failure (REMATCH) Trial [52]. At the writing of this dissertation the REMATCH data was not publicly available. Alternatively, we also compare predictions using the likelihood-based method to those of the Bayesian exponential method for the HIV Prevention Trials Network 052 Study data.

3.5.1 Prediction of landmark event times in the Chronic Granulomatous Disease Study

The CGD Study was a randomized, placebo-controlled, double blind trial of gamma interferon in the treatment of Chronic Granulomatous Disease, a group of rare, inherited disorders of immune function. Patients with CGD have an increased susceptibility to severe infections. Gamma interferon was hypothesized to reduce the recurrence rate of serious infections. Over a period of approximately seven months, 128 patients were enrolled and

randomized to gamma interferon or placebo.

For the purpose of demonstrating event-time predictions, we consider time until first infection as the primary time-to-event endpoint to be monitored. Bagiella and Heitjan proposed an interim monitoring plan with analyses after the 18th and 35th observed events [3]. Predictions of these two landmark event times were made at 30-day intervals from the beginning of enrollment. The 18th event occurred at day 251 of the study, and the 35th event was observed at day 353.

As in our likelihood-based method, the point prediction for the Bayesian exponential approach is calculated by conditioning on the number of events already observed in the data and finding the expected time of the landmark event using maximum likelihood estimates of the accrual, event, and dropout rate parameters in the respective distribution functions. The fundamental difference between the likelihood-based and Bayesian methods is in quantifying the variability of the point prediction. Bagiella and Heitjan propose to obtain prediction intervals by assigning a gamma (conjugate) prior to the accrual rate and event and dropout rates on each treatment arm, finding the posterior parameter densities based on the observed trial data, and then simulating completions of the trial data set based on random draws from the posterior distributions of the parameters. The "landmark" rank statistics from the resulting simulated data sets form a predictive distribution for the landmark event time, from which quantiles are calculated to form the prediction interval.

Bagiella and Heitjan assumed the Poisson enrollment rate, and the exponential event and dropout rates to be mutually independent and put gamma prior densities on each rate parameter. The gamma prior distribution has two convenient properties in this setting. First, it is a conjugate prior in each case. Second, it has a very accessible conceptualization in terms of historical, pilot, or early-phase trial data: the gamma shape parameter, A, may be viewed as a prior number of events, and the gamma rate parameter, B, (where rate = 1/scale) may be viewed as a prior total length of observation time. Bagiella and Heitjan selected the following "original" priors for their prediction model. The homogeneous Poisson process, with inter-accrual times having exponential(rate=α) distribution, had a prior distribution on α assuming accrual of 30 patients in 15 days (2 per day), i.e. $\Gamma(A_P = 30, B_P = 15)$. Treatment arms were separate, not pooled, with each arm having its own

exponential event and dropout rates. The event rate priors were $\Gamma(A_{E0} = 1, B_{E0} = 730)$ (1 event in 730 days) for the placebo arm and $\Gamma(A_{E1} = 1, B_{E1} = 2,190)$ (1 event in 2,190 days) for the gamma interferon arm. The dropout rate priors for both arms were $\Gamma(A_L = 1, B_L = 3,650)$ (1 loss-to-follow-up in 10 years).

One of the potential complications of the Bayesian approach is the requirement to specify a prior distribution on each of the design parameters of the study (accrual, event, and dropout rate). Since clinical trial data may deviate from the patterns anticipated during the design of the study, one may wish to take an agnostic approach to prediction, relying only on the observed data. To accomplish this using the Bayesian exponential approach, we can specify a "noninformative" prior distribution from the gamma family. One possible choice is $\Gamma(A = \epsilon, B = \epsilon)$, where $\epsilon \approx 0$. Using the notation of section 3.2.1, the posterior density at cutoff time C for the event-time exponential rate parameter with prior density $\Gamma(A, B)$ is

$$\Gamma\left(A + \sum_{i=1}^{n_C} \delta_{Ci}^E, B + \sum_{i=1}^{n_C} x_{Ci}\right)$$

[33]. We will use $\Gamma(A = \epsilon, B = \epsilon)$ priors with $\epsilon = 10^{-4}$. Note that the mean of the posterior distribution in this case is approximately equal to the ML estimate for the exponential rate parameter,

$$\text{posterior mean } = \frac{\epsilon + \sum_{i=1}^{n_C} \delta_{Ci}^E}{\epsilon + \sum_{i=1}^{n_C} x_{Ci}} \approx \frac{\sum_{i=1}^{n_C} \delta_{Ci}^E}{\sum_{i=1}^{n_C} x_{Ci}},$$

and the posterior variance is approximately equal to the estimated variance of the rate ML estimate, or the inverse information bound,

$$\text{posterior variance } = \frac{\epsilon + \sum_{i=1}^{n_C} \delta_{Ci}^E}{(\epsilon + \sum_{i=1}^{n_C} x_{Ci})^2} \approx \frac{\sum_{i=1}^{n_C} \delta_{Ci}^E}{(\sum_{i=1}^{n_C} x_{Ci})}.$$

Similarly, assuming a prior $\Gamma(A = \epsilon, B = \epsilon)$ density on the exponential rate parameter, α, which dictates inter-accrual times in homogeneous Poisson process enrollment, we obtain the posterior density

$$\Gamma\left(\epsilon + n_C, \epsilon + \sum_{i=1}^{n_C} w_{Ci}^*\right)$$

(recall w^* notation from section 3.3.3). The posterior mean and variance in this case also

Table 3.8: Summary of CGD Study data available at each cutoff time

		Gamma interferon			Placebo			Pooled arms		
Study day	enrolled	events	dropouts	person-days	events	dropouts	person-days	events	dropouts	person-days
30	4	0	0	70	1	0	8	1	0	78
60	18	0	0	275	2	0	142	2	0	417
90	36	0	0	664	3	0	457	3	0	1,121
120	67	0	0	1,539	4	0	1,158	4	0	2,697
150	86	0	0	2,722	4	0	2,086	4	0	4,808
180	107	1	0	4,195	10	0	3,288	11	0	7,483
210	128	2	0	5,954	11	0	4,665	13	0	10,619
240	128	3	1	7,758	13	0	6,237	16	1	13,995
270	128	5	1	9,494	16	0	7,749	21	1	17,243
300	128	6	1	11,186	18	2	9,155	24	3	20,341
330	128	6	21	12,866	21	2	10,437	27	3	23,339

correspond to the ML estimate for α and the inverse information bound for α, respectively.

$$\text{posterior mean} = \frac{\epsilon + n_C}{\epsilon + \sum_{i=1}^{n_C} x_{Ci}} \approx \frac{n_C}{\sum_{i=1}^{n_C} x_{Ci},}$$
$$\text{posterior variance} = \frac{\epsilon + n_C}{(\epsilon + \sum_{i=1}^{n_C} x_{Ci})^2} \approx \frac{n_C}{(\sum_{i=1}^{n_C} x_{Ci})^2.}$$

Thus we would expect this choice of prior to yield prediction intervals as close as possible to the truly agnostic (frequentist) likelihood-based method under the constraints of the Bayesian method.

We applied the method of Bagiella and Heitjan to the CGD study data under the original priors and the noninformative prior, using the S-plus codes available online (implemented in R). The available algorithm separates treatment arms and estimates distinct event and loss rates for each arm. We also applied the likelihood-based method described in section 3.3.3 in two ways: with pooled treatment arms, and stratified by treatment arm. Since our method requires the occurrence of at least one event in the pooled version, and at least one event on each arm in the stratified version, predictions using the stratified model are only shown starting at 180 days. Table 3.8 summarizes the available data used to calculate predictions at each 30-day cutoff time. If no dropout has been observed in the pooled

version, predictions from the likelihood-based method assume the future dropout rate will remain zero. This is also true for an individual treatment arm in the stratified version. The resulting prediction intervals for the 18th event under all four setups are shown in Figure 3.12, and prediction intervals for the 35th event are shown in Figure 3.13.

The likelihood-based (frequentist) prediction interval and the Bayesian prediction interval differ in their interpretations. The interpretations of both intervals assume all previously observed data values, $\mathscr{F}_C$, are fixed, and that the distributional assumptions for accrual, event, and dropout times are correct. In repeated experiments, we would expect the level α frequentist prediction interval to cover the true value, $E[R_N]$, $100 \cdot (1-\alpha)\%$ of the time; i.e. the random interval has probability 0.95 of covering the true future target event time when $\alpha = 0.05$. The level $\alpha = 0.05$ Bayesian interval may be interpreted as the middle 95% of the predictive distribution of the landmark event time $E[R_N]$, given the selected priors for each distribution parameter. Thus the probability of the true target event time being within the Bayesian prediction interval is 0.95, conditional on the choice of priors.

At each cutoff time, the Bayesian prediction intervals are far wider than the likelihood-based intervals. Predictions from all models shorten significantly at 180 days, when the total number of observed events jumps from four to eleven, and the first event is observed on the gamma interferon arm. Before that time there is too little data for either method to produce predictions with enough precision to be practically useful. The likelihood-based prediction intervals at day 180 (pooled and separate arms) narrowly miss the observed time of the 18th event. Otherwise, all four prediction intervals at all data cutoff times cover the true 18th event time. In the 71 days preceding the 18th event, the Bayesian prediction intervals remain quite wide, at approximately 100, 80, and 70 days. By comparison, the pooled-arm likelihood prediction intervals are more precise close to the landmark event time, having lengths of approximately 50, 40, and 15 days. The separate-arm likelihood prediction intervals are even shorter.

All prediction intervals for the 35th event time cover the true time except for the likelihood-based intervals at days 300 and 330, and the Bayesian intervals at day 330. These overestimates of time to the 35th event are due to a series of seven events that occur in quick succession between day 330 and day 353. The increase in event frequency is too close to

the landmark event for the prediction methods to detect. In the six months preceding the occurrence of the 35th event, the likelihood-based prediction intervals are approximately half to two-thirds of the length of the Bayesian intervals.

The CGD data set is quite small, which is not optimal for either prediction approach. In general for this trial data, the coverage of the Bayesian prediction intervals is slightly better than that of the likelihood-based intervals. However, the improved coverage is due to the longer length of the prediction intervals, rather than the specific location of the intervals, and the expanded reach of the Bayesian intervals often limits their practicality. When applied to the CGD data, the Bayesian Weibull and nonparametric prediction methods (not shown here) produce even wider prediction intervals than does the exponential method [51][53].

3.5.2 Prediction of landmark event times in the HIV Prevention Trials Network 052 Study

We also compared Bayesian and likelihood-based prediction intervals for landmark events in the HPTN 052 Study data. The HPTN 052 data set has a far greater sample size than the CGD data, includes about twice as many monitoring events (although a much lower overall event rate), and far more patient dropout (in terms of both overall loss-to-follow-up rate and absolute dropout counts). In section 3.4.1 we determined that a model assuming homogeneous Poisson process enrollment and Weibull event and loss time distributions would be the most prudent choice for mid-trial predictions. In order to make a more direct comparison, here we focus on prediction intervals obtained using the likelihood-based method with exponential event and loss times and the corresponding Bayesian exponential method.

Assuming that a total of 340 M/M monitoring events represents 100% of the statistical information in the interim monitoring plan, we once again selected as landmark events the 48th (0.14 information), 62nd (0.18 information), 75th (0.22 information), and 85th (0.25 information) observed M/M events. Predictions were made at the observation times of the 28th (0.08 information), 48th, 62nd, and 75th events. Likelihood-based predictions were calculated using a model that pooled treatment arms as well as one stratified by treatment arm.

Figure 3.12: Comparison of Bayesian and likelihood method prediction intervals for the date of the 18th event in the CGD trial. All models assume Poisson enrollment times and exponential event and dropout distributions. Predictions are calculated at 30-day intervals from the start of enrollment.

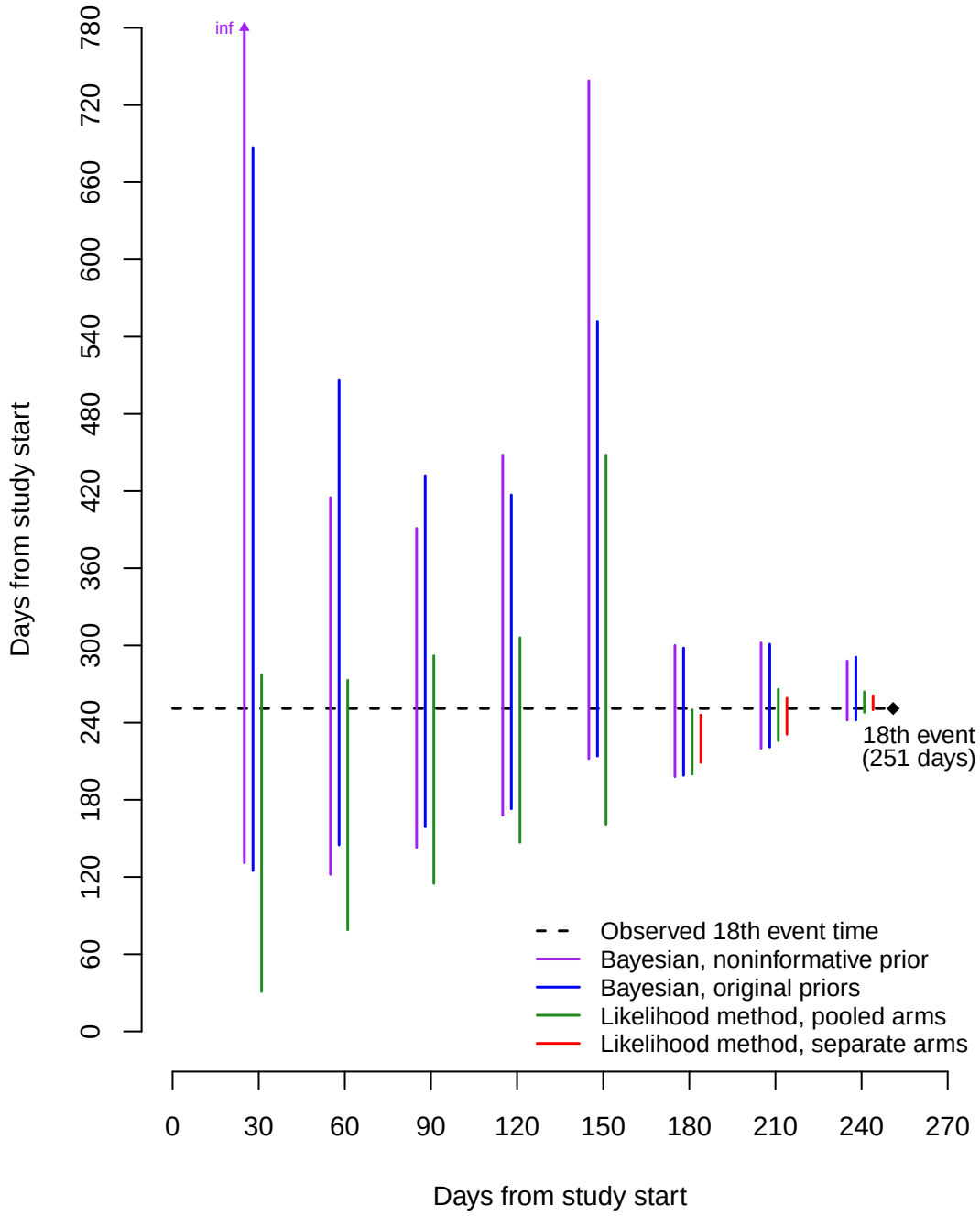


Figure 3.13: Comparison of Bayesian and likelihood method prediction intervals for the date of the 35th event in the CGD trial assuming Poisson enrollment times and exponential event and dropout distributions. Predictions are calculated at 30-day intervals from the start of enrollment.

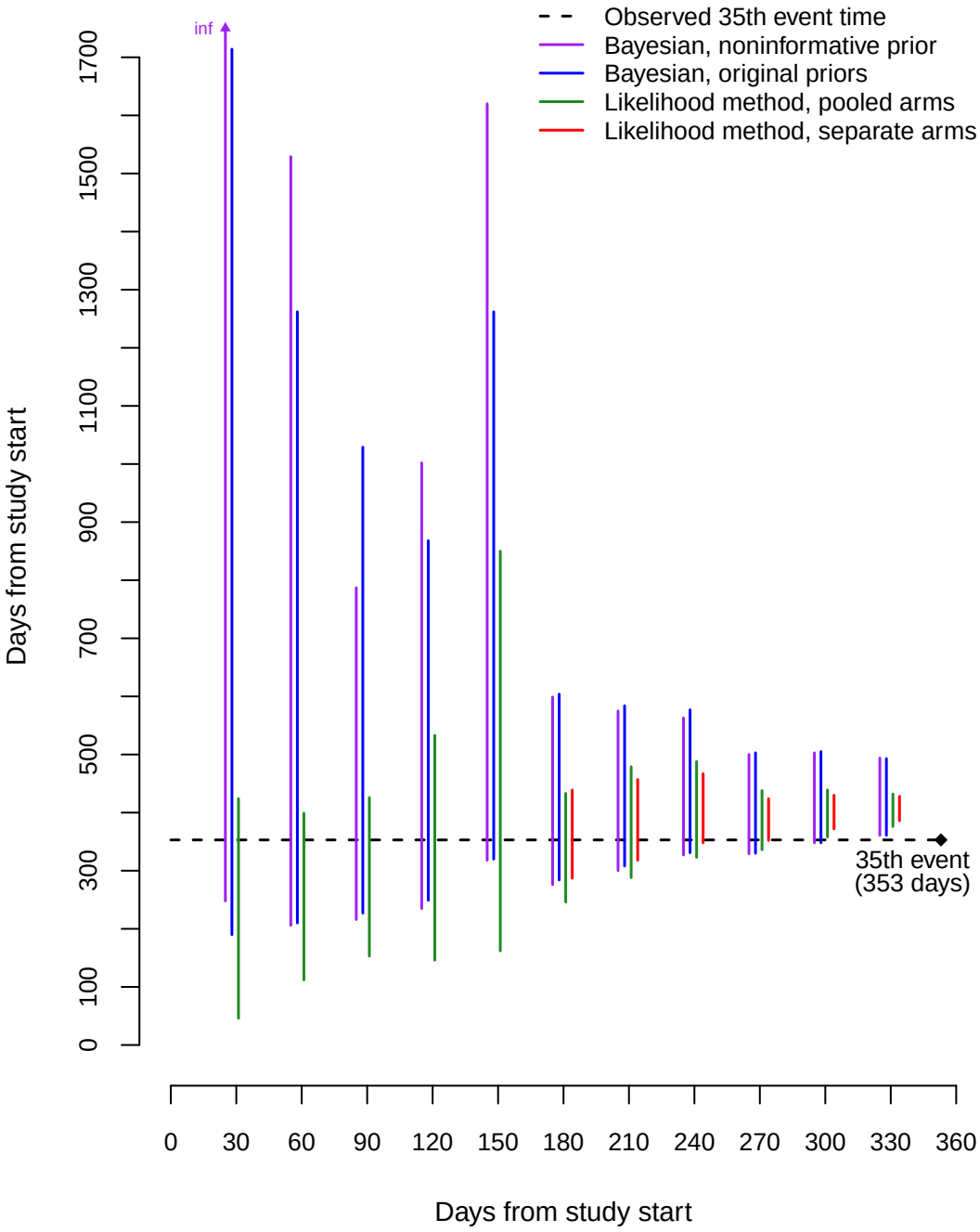


Table 3.9: Prior distributions for the exponential inter-accrual rate and the exponential event and dropout rates for Bayesian event-time prediction in the HPTN 052 Study.

	Set of prior distributions		
Exponential rate parameter	Noninformative	Design-phase (strong)	Design-phase (weak)
inter-accrual	$\Gamma(10^{-4}, 10^{-4})$	$\Gamma(1674, 1.5 * 365)$	$\Gamma(167.4, 1.5 * 36.5)$
event, early ART arm	$\Gamma(10^{-4}, 10^{-4})$	$\Gamma(6, 365 * 100)$	$\Gamma(0.6, 365 * 10)$
event, delayed ART arm	$\Gamma(10^{-4}, 10^{-4})$	$\Gamma(4, 365 * 100)$	$\Gamma(0.4, 365 * 10)$
dropout, early ART arm	$\Gamma(10^{-4}, 10^{-4})$	$\Gamma(5, 365 * 100)$	$\Gamma(0.5, 365 * 10)$
dropout, delayed ART arm	$\Gamma(10^{-4}, 10^{-4})$	$\Gamma(5, 365 * 100)$	$\Gamma(0.5, 365 * 10)$

We demonstrate Bayesian prediction intervals calculated under three different sets of prior distributions on the accrual, event, and dropout rate parameters. All prior distributions are $\Gamma(\text{shape} = A, \text{rate} = B)$ conjugate priors. First, we used noninformative priors on all rate parameters; second, we used prior distributions based on the design-phase estimates of accrual, events, and dropout in HPTN 052 ("strong" design-phase priors); and third, we used weakened versions of the design-phase priors, for which the shape and rate parameters were each decreased by a factor of ten ("weak" design-phase priors). The weak priors have the same mean as the strong priors, but a variance that is larger by a factor of ten. For a description of the design-phase estimates, see section 3.3.2, Example 1. The specific prior distribution parameters are shown in Table 3.9.

Bayesian and likelihood-based method prediction intervals are shown in Figure 3.14. Recall the differences in the interpretation of each type of interval, as described in section 3.5.1. The noninformative and weak design-phase priors produce very similar results: wide prediction intervals that cover the observed landmark event time at each information fraction prediction time. Prediction intervals obtained using the strong design-phase priors at information fractions 0.08 and 0.14 tend be narrower and cover a slightly earlier range of study times than those using noninformative and weak design-phase priors. All three Bayesian prediction intervals are similar at the 0.18 and 0.22 information fraction cutoffs.

The Bayesian intervals range in length from approximately 4 to 10 months.

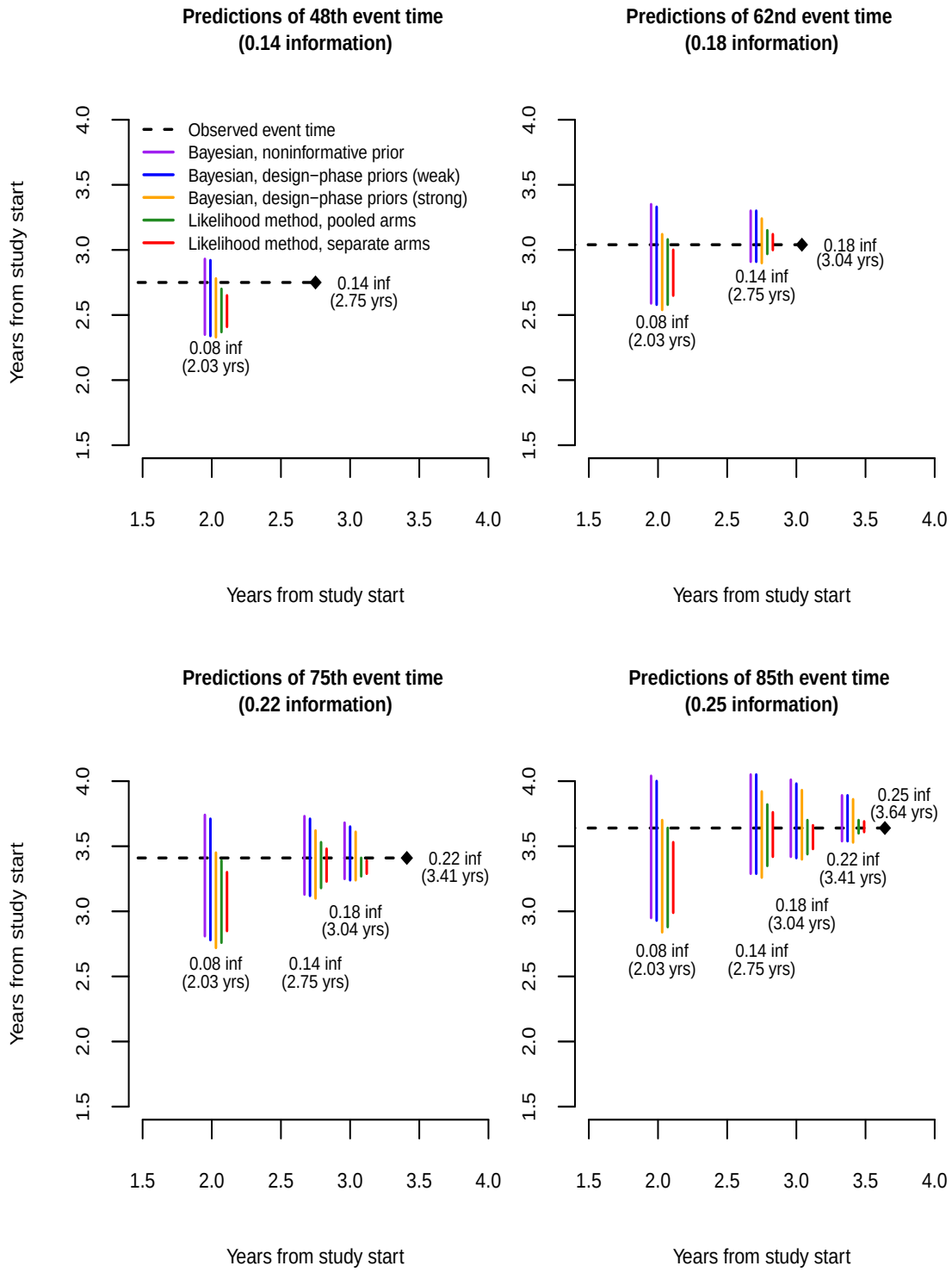
The likelihood-based prediction intervals, both for pooled and separate treatment arms, are narrower than the Bayesian intervals at each prediction time, ranging in length from 1 to 6.5 months. Most of the likelihood prediction intervals calculated at information fraction 0.08 fall earlier than the true landmark event time. The remaining likelihood prediction intervals cover the true event time, with the exception of the stratified treatment arm model prediction of information fraction 0.22 at information fraction 0.18, which ends 0.02 years short of the actual 75th event.

3.6 Summary

Our parametric prediction method provides a straightforward approach to use accumulating trial data to obtain point predictions and prediction intervals for future event times. Both the general model for study planning and the conditional model for study monitoring allow investigators to select the enrollment, event, and loss-time distributions most appropriate to their study based on the scientific context or on empirical estimates. The practicality of the approach lies in its versatility and simplicity of implementation: a variety of parametric models may be accommodated, and although no parametric model perfectly fits real data, many provide useful approximations if chosen thoughtfully by the statistician. This data-driven, frequentist approach requires neither prior assumptions about model parameter distributions nor the simulation of additional data. Furthermore, the resulting prediction intervals do not depend on prior distributional assumptions for individual accrual, event, and dropout time distributions.

The reliability of point predictions and inference obtained using our approach was demonstrated through simulation studies. Point predictions were consistent for the true expected times of landmark events, and prediction intervals were well calibrated even at early data cutoff times. The event time predictions were shown to be robust even when ML methods yielded inaccurate estimates of individual distribution parameter values and their standard errors (as occurred most notably for the Weibull shape parameters). We also demonstrated that the method produced good point predictions and reasonably narrow prediction intervals when applied to the early HPTN 052 Study data. The HPTN

Figure 3.14: Comparison of Bayesian and likelihood method prediction intervals for the observation dates of information fractions 0.14, 0.18, 0.22, and 0.25 for the M/M monitoring endpoint in the HPTN 052 trial. All methods assume Poisson enrollment times and exponential event and dropout distributions. Predictions are calculated at observed information fractions of 0.08, 0.14, 0.18, and 0.22.



052 example illustrates the bias/variance tradeoff present in many prediction problems: the two-parameter Weibull event-time model over-fits a cluster of closely-timed events at the first data cutoff time, and if used at this early stage, would yield too-early predictions of future event times. Thus the simpler exponential model, which relies only on the average event rate, is a more appropriate choice for estimation early in data collection; i.e. flexibility and parsimony should be carefully weighed when predictions are based on little data.

Although parametric modeling is convenient, it comes with several limitations. Misspecification of enrollment, event, or loss-time models may result in biased or inefficient estimators. The approach presented here also requires strong assumptions of independence both among patients, and for a single patient's accrual time and event or dropout time. In practice, many clinical trials are conducted in high-risk catchment areas, in which case accrual and event rates are likely to be correlated. Furthermore, the methods cannot account for any type of informative censoring.

Nevertheless, this approach is a practical and useful tool, and it provides a statistical framework that could potentially be extended in a number of ways. Using survival models that allow for more dependent structures, we could relax the stringent distributional and independence assumptions. An ideal occurrence time model would also permit the inclusion of baseline covariates, such as study site or risk indicators, to account for dependence in the data. We explore this idea in Chapter 4.

Chapter 4

SEMIPARAMETRIC PREDICTION OF EVENT TIMES IN CLINICAL TRIALS

4.1 Introduction

Many late-phase clinical trials are large and employ study participants carefully selected on the basis of specific inclusion and exclusion criteria. Homogeneity in the sample of patients allows investigators to administer experimental treatment to the appropriate candidates, accurately analyze and interpret trial data, and properly generalize the results in the broader population. Nevertheless the study population may still include patients with important differences in susceptibility to a particular outcome, such as age, sex, ethnicity, geographical location, or progression of disease. Baseline assessments of patient characteristics and exposures, including risk factors, prognostic covariates, or even genetic determinants for the response of interest, are made before the start of the intervention, often as part of the randomization scheme [26]. Baseline covariates known or theorized to be strongly related to a monitored endpoint may improve the efficiency and accuracy of endpoint-time predictions when included in the working prediction model. With a rich data set that includes covariate information in addition to event indicators and accrual and observation times, we may ask the question: given trial data accumulated through an interim cutoff time, along with baseline characteristics of the patients remaining at risk, what is the expected future study time of the Nth observed event?

The fully parametric prediction method described in Chapter 3 allows for limited use of participant characteristics. Specifically, the prediction models can be stratified according to one or more categorical variables, such as sex, treatment group, or study site. Stratification into multiple categories or by multiple binary variables will quickly increase the number of parameters to be estimated and decrease the available data to estimate each one. Without a very large study population, this could greatly reduce the efficiency of event-time

predictions.

In this Chapter we introduce a semiparametric event-time prediction method with expanded capabilities compared to the parametric approach. The Cox proportional hazards model is the most commonly used method for modeling the relationship of covariates to a censored survival outcome. In the Cox model, the hazard function $\lambda(t|z)$ is an unspecified function of time, and covariates have a multiplicative effect on the hazards [14].

$$\lambda_i(t|z) = \lambda_0(t)e^{z_i\beta}$$

The nonparametric hazard function is more flexible than its parametric counterparts. Assuming the proportional hazards assumption holds, model misspecification is not a concern, and consistent estimates of the log hazard ratios, $\widehat{\beta}$, may be obtained using a partial likelihood function [25]. A semiparametric estimate of the baseline cumulative hazard function, corresponding to the Cox model, is given by the Breslow estimator [7], for which asymptotic theory is well-developed [6][25][30]. For prediction, we require a model for the distribution of loss-to-follow-up times as well for the event times. The counterpart of the Breslow estimator for a single homogeneous population (no covariates) is the nonparametric Nelson-Aalen estimate of cumulative hazards [40]. Nonparametric modeling is inherently less efficient than a parametric model when a parametric model is a good fit for the sampled data. However, when it is difficult or impossible to select an appropriate parametric fit for the data, and when there is sufficient data for nonparametric modeling, a nonparametric model may provide a more accurate representation of the data.

Although these nonparametric and semiparametric methods allow for unrestricted modeling of the hazard function and the inclusion of covariates, the resulting estimated cumulative hazard and survival functions are only defined over the range of observation times in the underlying data. To make event-occurrence-time predictions beyond the observed range of the data, we must extrapolate tails of the event-time and dropout-time distributions. We use parametric Weibull extensions for both the proportional hazards event distribution and the distribution for time-to-dropout.

Among parametric survival models, the Weibull has been shown to have several useful properties, even when the data do not follow an exact Weibull distribution. If proportion-

ality holds, the estimated hazard ratios from the Weibull and Cox models are very similar, as both estimate the average risk of death on one treatment arm relative to the average risk of death on the other. This is often true even when the data are known not to come from a Weibull distribution, as has been demonstrated for lognormal, gamma and piecewise exponential data [8]. Also, the standard error of the log hazard ratio estimate is asymptotically identical for the Cox model and Weibull model, so the models are equally efficient, asymptotically.

In the following, we first define the basic notation of the clinical trial data and our setting. In many cases, the notation is the same or similar to that used in Chapter 3 but is expanded to include more patient-specific variables and the larger set of finite- and infinite-dimensional parameters that defines the semiparametric prediction method. The estimating equation used to calculate our point prediction is quite similar to that described in Chapter 3. However here we derive the estimating equation in a manner meant to motivate the asymptotic consistency and normality on which we rely for inference. Estimation of each component of the estimating equation is described in detail, including pseudo-maximum likelihood parameter estimates we develop specifically for this prediction model. Behavior of the resulting point predictions is characterized through simulation studies.

4.2 Methods

4.2.1 Notation

Consider a clinical trial with a time-to-event endpoint we wish to monitor for efficacy or safety considerations throughout the study. A sample of n patients is required and may be entered into the study either simultaneously or in a staggered fashion over an interval of time. When patient i, $i = 1, \ldots, n$, is accrued, baseline characteristics, or covariates, are recorded. Let the finite-dimensional vector Z denote a subset of the measured baseline covariates assumed to be predictive of the time to the monitoring endpoint. After enrollment, each patient is followed until they experience the event of interest or are lost to follow-up. At any data cutoff time throughout the study, a patient who has not yet experienced the event or been lost to follow-up is administratively censored. The total number of monitoring

endpoints required for a final analysis is considered 100% of the statistical information for the monitoring analysis plan. At certain information fractions, each corresponding to a specific number of observed monitoring events, interim analyses will be performed to evaluate the benefit-to-risk profile of the trial.

At the close of the study, the data for an individual patient consists of an enrollment time W measured from the start of the study, the baseline covariate vector Z, and an observation time X, which is the minimum of the patient's time-to-event, time-to-dropout, or time-to-administrative censoring, measured from his or her enrollment. Additionally, we include three indicator variables denoting the occurrence of an observed event, Δ^E, the patient being lost-to-follow-up, Δ^L, and the patient being administratively censored, Δ^A. The complete observed data for patient i is

$$\mathbb{X}_i = \{w_i, z_i, x_i, \delta_i^E, \delta_i^L, \delta_i^A\}.$$

A general time during the study is denoted R, and the start of the study (the beginning of enrollment) is at time $R = 0$. Suppose we require a certain number of observed events, N, for an interim or final data analysis. At a time C during the study, we wish to use the data accumulated in the time interval $[0, C]$ to predict the future study time R_N, at which the Nth event will be observed among the sample of size n. We will assume that all n patients have been accrued before the data cutoff time C, so all enrollment times w_i and all baseline covariate values z_i are known. At time C, some patients may have already been observed to experience the event or drop out of the study. For clarity, a patient's data at a specific data cutoff time C could be denoted

$$\mathbb{X}_{C,i} = \{w_{C,i}, z_{C,i}, x_{Ci}, \delta_{Ci}^E, \delta_{Ci}^L, \delta_{Ci}^A\},$$

to indicate that the values of the variables are specific to time C. Note that whenever $\delta_{C,i}^E = 0$ and $\delta_{C,i}^L = 0$, then we consider the patient to be administratively censored at the cutoff time, that is $\delta_{C,i}^A = 1$. At any time throughout the study, it is always the case that $\delta_{C,i}^E + \delta_{C,i}^L + \delta_{C,i}^A = 1$. Also, since baseline covariates and accrual time are fixed once a patient enrolls, $z_{Ci} = z_i$, and $w_{Ci} = w_i$ for any cutoff time C after the end of enrollment. The collection of all observed data in $[0, C]$ is represented by $\mathscr{F}_C$.

As in Chapter 3, our goal is to predict the expected time of the Nth observed event conditional on the current observed trial data at C, $E[R_N|\mathscr{F}_C]$. The elapsed time from $R = 0$ until the end of observation for the ith patient, or the absolute observation time for patient i, is defined as $Y_i \equiv W_i + X_i$. The absolute observation times for all patients, ordered by magnitude, are

$$Y_{(1)} \leq Y_{(2)} \leq \cdots \leq Y_{(n)}.$$

Thus we wish to find the expected value of an order statistic,

$$E[R_N|\mathscr{F}_C] = E\left[Y_{(k)}|\mathscr{F}_C : \sum_{j=1}^{k} \delta_{(j)}^R = N; \delta_{(k)}^E = 1\right]. \tag{4.1}$$

Assume the true time from enrollment to the monitoring endpoint is iid for each patient in the study population and follows distribution F_{E^*}. This true event-time distribution may depend on a variety of baseline patient characteristics, exposures, or fluctuations in treatment implementation, say Z^*. Suppose the finite-dimensional covarariate vector Z^* has distribution F_{Z^*}. Then $F_{E^*} = F_{E^*|Z^*}F_{Z^*}$. The time from enrollment to dropout is iid for each patient and follows distribution F_L. We assume censoring is noninformative, and that $F_{E^*|Z^*}$ and F_L are independent. The corresponding survival function for dropout and survival and cumulative hazard functions for an event are given by S_L, S_{E^*}, and Λ_{E^*}, respectively. For now, all patients are assumed to be accrued before the data cutoff C. So our prediction model need not allow for random variation in enrollment times.

4.2.2 A Conditional event-occurrence time model

For simplicity, initially suppose enrollment is simultaneous for a sample of n patients, so that $w_i = 0$ for $i = 1, \ldots, n$. Let $E[R_N]$ be the expected time of the Nth observed event in a sample of size n. Define the probability

$$p_N \equiv Pr(0 \leq x_i \leq E[R_N], \delta_i^E = 1). \tag{4.2}$$

Then $p_N = \frac{N}{n}$. Recall that Δ_R^E is the event indicator for a patient observed through study time R. Then $\Delta_{E[R_N]}^E \sim$ Bernoulli(p_N), and in the observed data,

$$\widehat{p_N} \equiv \frac{1}{n}\sum_{i=1}^{n} \delta_{E[R_N]i}^E$$

is well known to be a consistent estimator of p_N [9]. Thus

$$E\left[\frac{1}{n}\sum_{i=1}^{n}\delta^E_{E[R_N]i} - p_N\right] = E\left[\frac{1}{n}\sum_{i=1}^{n}\delta^E_{E[R_N]i} - \frac{N}{n}\right] = 0. \tag{4.3}$$

Now suppose data through study time $E[R_N]$ is not available, and instead we have data truncated at a cutoff time $C \in (0, E[R_N])$. We wish to consider a version of the estimator $\widehat{p}_N$, conditional on the truncated, observed data. Denote the data observed in study time $[0, C]$ as

$$\mathscr{F}_C \equiv \left\{w_i = 0, z_i, x_{Ci}, \delta^E_{Ci}, \delta^L_{Ci}, \delta^A_{Ci} : i = 1, \ldots, n\right\}.$$

At the data cutoff, C, some patients may have been observed to have the event of interest ($\delta^E_{C,i} = 1$), some may have been lost to follow-up ($\delta^L_{C,i} = 1$), and some may remain under observation with no event or dropout ($\delta^A_{C,i} = 1$). The set of administratively censored patients at time C is exactly the set of patients still at risk of having an event or dropping out of the study. Let $\mathscr{R}_C$ be the risk set at time C. If a patient has experienced an observed event or has been lost-to-follow-up before C, then the value $\delta^E_{E[R_N]i}$ is known; that is, $\delta^E_{Ci} = \delta^E_{E[R_N]i}$, and $\delta^E_{E[R_N]i} = 1$ for an observed event or $\delta^E_{E[R_N]i} = 0$ for an observed dropout. The probability of an individual experiencing an observed event by study time R, given that they are in the risk set at time C is

$$Pr(C < X + W \leq R, \Delta^E = 1 | X + W > C).$$

Here, W, X and Δ^E represent a patient's eventual true enrollment time, observation time and event indicator, not the truncated values recorded at time C.

Let $N_C \equiv \sum_{i=1}^{n}\delta^E_{Ci}$. Then, using the accumulated data $\mathscr{F}_C$, equation (4.3) may be

rewritten

$$E\left[E\left[\left\{\frac{1}{n}\sum_{i=1}^{n}\delta^E_{E[R_N]i}-\frac{N}{n}\right\}\Big|\mathscr{F}_C\right]\right]=0 \tag{4.4}$$

$$E\left[E\left[\left\{\frac{1}{n}\sum_{i\notin\mathscr{R}_{\mathscr{C}}}\delta^E_{E[R_N]i}+\frac{1}{n}\sum_{i\in\mathscr{R}_{\mathscr{C}}}\delta^E_{E[R_N]i}-\frac{N}{n}\right\}\Big|\mathscr{F}_C\right]\right]=0$$

$$E\left[\frac{N_C}{n}-\frac{N}{n}+E\left[\frac{1}{n}\sum_{i\in\mathscr{R}_C}\delta^E_{E[R_N]i}\Big|\mathscr{F}_C\right]\right]=0$$

$$\frac{N_C}{n}-\frac{N}{n}+\frac{1}{n}E\left[E\left[\sum_{i\in\mathscr{R}_C}\delta^E_{E[R_N]i}\Big|\mathscr{F}_C\right]\right]=0$$

$$\frac{N_C}{n}-\frac{N}{n}+\frac{1}{n}E\left[\sum_{i=1}^{n}\delta^A_{Ci}\Big\{Pr(C<x_i\le E[R_N],\delta^E_i=1|x_i>C,\mathscr{F}_C)\Big\}\right]=0$$

$$\frac{N_C}{n}-\frac{N}{n}+\frac{1}{n}\sum_{i=1}^{n}\delta^A_{Ci}\left\{Pr(C<x_i\le E[R_N],\delta^E_i=1|x_i>C,\mathscr{F}_C)\right\}=0 \tag{4.5}$$

Multiplying both sides of equation (4.5) by n, yields

$$N_C-N+\sum_{i=1}^{n}\delta^A_{Ci}\left\{Pr(C<x_i\le E[R_N],\delta^E_i=1|x_i>C,\mathscr{F}_C)\right\}=0. \tag{4.6}$$

When the probability

$$Pr(C<x_i\le R,\delta^E_i=1|x_i>C,\mathscr{F}_C)$$

is known for a general study time $R>C$, (4.6) is a deterministic equation that implicitly defines the expected time, $E[R_N|\mathscr{F}_C]$, by which N events will be observed, conditional on the current observed trial data at C. Define the function G as

$$G(R;N|\mathscr{F}_C)=N_C-N+\sum_{i\in\mathscr{R}_C}\delta^A_{Ci}\big\{Pr(C<x_i+w_i\le R,\delta^E_i=1|x_i+w_i>C,\mathscr{F}_C)\big\} \tag{4.7}$$

For a chosen number of events N, the root of this equation (if it exists) is the expected future study time of the Nth observed event, $E[R_N]$.

If we assume that $F_{E^*|Z^*}$ is differentiable with derivative $dF_{E^*|Z^*}$, then the function G

may be written as follows.

$$G(R;N|\mathscr{F}_C) = N_C - N + \sum_{i\in\mathscr{R}_C}\left\{\frac{Pr(C<x_i+w_i\le R,\delta_i^E=1|\mathscr{F}_C)}{Pr(x_i+w_i>C|\mathscr{F}_C)}\right\}$$
$$= N_C - N + \sum_{i=1}^{n}\left\{\delta_C i^A\int_{Z^*}\int_{C-w_i}^{R-w_i}\frac{S_L(u)dF_{E^*|Z^*}(u|z)dF_Z^*(z)}{S_L(C-w_i)S_{E^*|Z^*}(C-w_i)}\right\} \quad (4.8)$$
$$= N_C - N + \sum_{i=1}^{n}\left\{\delta_{Ci}^A\int_{Z^*}\int_{C-w_i}^{R-w_i}\frac{S_L(u)S_{E^*|Z^*}(u|z)d\Lambda_{E^*|Z^*}(u|z)dF_{Z^*}(z)}{S_L(C-w_i)S_{E^*|Z^*}(C-w_i|z_i)}\right\}. \quad (4.9)$$

The function G provides the framework for an estimating procedure that may be used to predict of the time to the Nth observed event, given $\mathscr{F}_C$. When both the true probability

$$Pr(C<x_i\le R,\delta_i^E=1|x_i>C,\mathscr{F}_C)$$

and the true time $E[R_N]$ are unknown, as is usually the case during a clinical trial, we propose to use a working approximation of

$$Pr(C<x_i\le R,\delta_i^E=1|x_i>C,\mathscr{F}_C)$$

to obtain an estimating function $\widehat{G}_{Cn}(R;N|\mathscr{F}_C)$, the root of which is an estimate $\widehat{E[R_N]}$ of the expected Nth observed event time. Trial data accumulated in the interval $[0,C]$ is used to obtain estimates $\widehat{S}_L$, $\widehat{S}_{E^*|Z^*}$, and $\widehat{\Lambda}_{E^*|Z^*}$, of the functions comprising the integrand of equation (4.9). Then an estimating function $\widehat{G}_{Cn}$ is obtained by plugging in the estimated functions and observed data values from $\mathscr{F}_C$ into equation (4.9).

$$\widehat{G}_{Cn}(R;N|\mathscr{F}_C) = N_C - N + \sum_{i=1}^{n}\left\{\delta_{Ci}^A\int_{C-W_i}^{R-w_i}\frac{\widehat{S}_L(u)\widehat{S}_{E^*|Z^*}(u|z_i)d\widehat{\Lambda}_{E^*|Z^*}(u|z_i)}{\widehat{S}_{E^*|Z^*}(C-w_i|z_i)\widehat{S}_L(C-w_i)}\right\} \quad (4.10)$$
$$= N_C - N + \sum_{i=1}^{n}\left\{\frac{\delta_{Ci}^A}{\widehat{S}_{E^*|Z^*}(C-w_i|z_i)\widehat{S}_L(C-w_i)}\int_{C-W_i}^{R-w_i}\widehat{S}_L(u)\widehat{S}_{E^*|Z^*}(u|z_i)d\widehat{\Lambda}_{E^*|Z^*}(u|z_i)\right\}$$

The solution to the equation $\widehat{G}_{Cn}(R)=0$ is our point estimator, $\widehat{E[R_N|\mathscr{F}_C]}$. The consistency of this estimator for the true value of $E[R_N]$ is discussed in section 4.2.7.

It is rare or impossible in practice for investigators to identify and measure at baseline all the covariates Z^* necessary to precisely characterize the true event-time distribution $F_{E^*|Z^*}$. To apply the prediction method in practice we must first select a subset of baseline

covariates Z on which to base predictions. The collection of covariates Z includes patient and study site characteristics for which we may obtain baseline measurements, and which are known or believed to be associated with the time to the monitoring endpoint of interest. If the monitoring endpoint is a composite endpoint defined as the first occurrence of several possible events, then each element of Z is believed to be associated with one or more of the constituent events. When implementing the prediction method, we approximate $F_{E^*|Z^*}$ by the "working model" $F_E = F_{E|Z}F_Z$ for time to the monitoring event. Furthermore, we assume that proportional hazards hold for the covariates Z. That is, the covariates have a multiplicative relationship with the baseline hazard function, and the hazard ratios associated with elements of Z do not depend on time. The accuracy and utility of predictions will be heavily dependent on the appropriateness of the chosen working model. The estimating function for $\widehat{E[R_N]}$ corresponding to the working event-time model is

$$\widehat{G}^w_{Cn}(R; N|\mathscr{F}_C) = N_C - N + \sum_{i=1}^{n} \left\{ \delta^A_{Ci} \int_{C-W_i}^{R-w_i} \frac{\widehat{S}_L(u)\widehat{S}_{E|Z}(u|z_i)d\widehat{\Lambda}_{E|Z}(u|z_i)}{\widehat{S}_{E|Z}(C-w_i|z_i)\widehat{S}_L(C-w_i)} \right\} \tag{4.11}$$

$$= N_C - N + \sum_{i=1}^{n} \left\{ \frac{\delta^A_{Ci}}{\widehat{S}_{E|Z}(C-w_i|z_i)\widehat{S}_L(C-w_i)} \int_{C-W_i}^{R-w_i} \widehat{S}_L(u)\widehat{S}_{E|Z}(u|z_i)d\widehat{\Lambda}_{E|Z}(u|z_i) \right\}.$$

4.2.3 Piecewise estimation of the conditional event-time and the dropout-time distributions

Predictions are made after the full sample is enrolled, and baseline covariate information for the working event-time model is known. The observation time for each individual is $C - w_i$, the time between the patient's enrollment and the data cutoff time C. Let A_E denote the longest observation time corresponding to an observed event, and let A_L denote the longest observation time corresponding to a dropout. That is,

$$A_E = \max_{1\le i\le n} \{\delta^E_{Ci} \cdot x_{Ci}\}$$

$$A_L = \max_{1\le i\le n} \{\delta^L_{Ci} \cdot x_{Ci}\}$$

The value A_E partitions the range of possible patient observation times through time R into an interval over which functions for conditional event time ($S_{E|Z}$, $\Lambda_{E|Z}$, and $\lambda_{E|Z}$) are estimated semiparametrically, $[0, A_E]$, and an interval over which they are estimated parametrically, $(A_E, R]$. Similarly, we estimate S_L nonparametrically for possible observation

times in $[0, A_L]$ and parametrically for observation times in $(A_L, R]$.

4.2.4 Semiparametric and nonparametric estimation of the event- and dropout-time distributions

For estimation of $S_{E|Z}$, $\Lambda_{E|Z}$, and $\lambda_{E|Z}$, observation times terminating in dropout or administrative censoring are treated as censored. Over the interval $[0, A_E]$, we estimate the conditional event time functions semiparametrically using the Breslow estimate of cumulative hazards, conditional on baseline covariates Z.

To do this, we first fit a Cox proportional hazards model to the observed data $\mathscr{F}_C$ and, via partial likelihood, obtain an estimate $\widehat{\beta}_C$ of the log hazard ratio for event corresponding to Z [25]. The Breslow estimator, denoted $\widehat{\Lambda}_{spE|Z}$, is a semiparametric estimate of the conditional cumulative hazard function for an event that yields stratum-specific cumulative hazards for groups defined by the covariates Z. The estimated conditional survival function for event, $\widehat{S}_{spE|Z}$, is obtained as a transformation of $\widehat{\Lambda}_{spE|Z}$. For observation times $x \in [0, A_E]$ and covariate value z,

$$\widehat{\Lambda}_{spE|Z}(x|z) = \frac{\sum_{j=1}^{n} \delta_{Cj}^{E} \cdot 1(x_{Cj} \leq x) \cdot \exp(\widehat{\beta}_C z)}{\sum_{j=1}^{n} 1(x_{Cj} \geq x) \cdot \exp(\widehat{\beta}_C z_j)}$$

$$d\widehat{\Lambda}_{spE|Z}(x|z) = \frac{\sum_{j=1}^{n} \delta_{Cj}^{E} \cdot 1(x_{Cj} = x) \cdot \exp(\widehat{\beta}_C z)}{\sum_{j=1}^{n} 1(x_{Cj} \geq x) \cdot \exp(\widehat{\beta}_C z_j)},$$

and

$$\widehat{S}_{spE|Z}(x|z) = \exp\big\{ - \widehat{\Lambda}_{spE|Z}(x|z)\big\}.$$

For observation times $x > A_E$, we assume the event time functions may be reasonably approximated using a parametric Weibull distribution. The Weibull shape and scale parameters are estimated via a pseudo-maximum likelihood approach, described below. Both the semiparametric and the Weibull estimates of the event-time distribution assume proportional hazards conditional on the modeled baseline covariates.

The survival function for loss to follow-up, S_L, is similarly estimated piecewise over the range of possible observation times. Here patient dropout is the modeled endpoint, and observation times terminating in the monitoring event of interest or administrative censoring

are treated as censored. For study times in the interval $[0, A_L]$, the survival function for dropout, $\widehat{S}_{npL}$, is derived from a nonparametric Nelson-Aalen estimate of cumulative hazards for dropout, $\widehat{\Lambda}_{npL}$ [25]. Thus, for $x \in [0, A_L]$,

$$\widehat{\Lambda}_{npL}(x) = \frac{\sum_{i=1}^{n} \delta_{Ci}^{L} \cdot 1(x_{Ci} \leq x)}{\sum_{i=1}^{n} 1(x_{Ci} \geq x)}$$

and

$$\widehat{S}_{npL}(x) = \exp\big\{ - \widehat{\Lambda}_{npL}(x)\big\}.$$

For observation times $x > A_L$, the survival function for dropout is estimated parametrically using a Weibull distribution whose shape and scale parameters are obtained using the pseudo-maximum likelihood approach described below.

4.2.5 Parametric Weibull estimation of the event- and dropout-time tail distributions

The semiparametric and nonparametric estimates of the survival functions for event and dropout are only defined over the observed range of observation times. In order to make event-time predictions for longer observation, we append Weibull tails at the terminal times of each estimated survival function. A proportional hazards Weibull model is assumed for the distribution of event times, conditional on the modeled baseline covariates, Z.

Standard ML estimation of the Weibull shape and scale parameters using the available data $\mathscr{F}_C$ may result in a survival function tail beginning at a higher or lower probability than the terminal survival probability of the nonparametric estimate. A ML Weibull tail estimate that begins too high has the unfavorable result of yielding a working survival function that is not monotonically decreasing. A Weibull tail estimate that begins too low may create an unrealistic jump in survival probability at an arbitrary observation time. Furthermore, full maximum likelihood estimation of the hazard ratios in the proportional hazards Weibull model, along with the shape and scale parameters, could produce different hazard ratios than the partial likelihood estimates, $\exp \widehat{\beta}_C$ in section 4.2.4. We address these issues by estimating the Weibull shape and scale parameters using a pseudo-ML approach that restricts the estimated Weibull survival function tail to begin exactly at the terminal point of the nonparametric estimate. To maintain consistent hazard ratio estimates across

the entire range of observation times through R, the pseudo-ML estimates for event shape and scale parameters are obtained using the previously estimated partial likelihood log hazard ratios. In section 4.3.1 we use simulations to show that the pseudo-ML estimates of Weibull shape and scale are the same as the standard ML estimates when the underlying event- or dropout-time distribution is (proportional hazards) Weibull.

Pseudo-maximum likelihood estimation of the tail of the event-time distribution

Assume the tail of the event-time distribution, conditional on the baseline covariates in the working model, is well approximated by a proportional hazards Weibull model. That is, the baseline conditional hazard function should be monotonic (increasing, decreasing, or constant) with an approximate power-dependence on time. Then the tail of the conditional event hazard function has the form

$$\lambda_{E|Z}(x|z) = \frac{\tau_E}{\phi_E}\left(\frac{x}{\phi_E}\right)^{(\tau_E-1)} \cdot \exp(\beta z),$$

for $x \in (A_E, R]$. If Z has p elements, the score equations for the shape and scale parameters and the log hazard ratios β_p are

$$\begin{aligned}
\dot{l}_{\tau_E} &= \sum_{i=1}^{n} \Big\{ \frac{\delta_i^E}{\tau_E} + \delta_i^E \log(x_i) - \delta_i^E \log(\phi_E) - \Big(\frac{x_i}{\phi_E}\Big)^{\tau_E} \exp(\beta z_i) \log(x_i) \\
&\qquad + \Big(\frac{x_i}{\phi_E}\Big)^{\tau_E} \exp(\beta z_i) \log(\phi_E) \Big\} \\
\dot{l}_{\phi_E} &= \sum_{i=1}^{n} \Big\{ -\delta_i^E \frac{\tau_E}{\phi_E} + \frac{\tau_E}{\phi_E^{(\tau_E+1)}} x_i^{\tau_E} \exp(\beta z_i) \Big\} \\
\dot{l}_{\beta_j} &= \sum_{i=1}^{n} \Big\{ \delta_i^E z_{i,j} - \Big(\frac{x_i}{\phi_E}\Big)^{\tau_E} \exp(\beta_j z_{i,j}) z_{i,j} \Big\}, \text{ for j=1, \ldots, p.}
\end{aligned}$$

The standard ML estimates of τ_E, ϕ_E, and β are the simultaneous solutions to the system of equations ($\dot{l}_{\tau_E} = 0; \dot{l}_{\phi_E} = 0; \dot{l}_{\beta_j} = 0, \text{ for } j = 1, \ldots, p$).

Now, setting $\dot{l}_{\phi_E} = 0$, we obtain the following two equations relating ϕ_E and τ_E.

$$\phi_E^{\tau_E} = \frac{\sum_{i=1}^{n} x_i^{\tau_E} \exp(\beta z_i)}{\sum_{i=1}^{n} \delta_i^E}$$

$$\log(\phi_E) = \frac{1}{\tau_E} \log\Big(\frac{\sum_{i=1}^{n} x_i^{\tau_E} \exp(\beta z_i)}{\sum_{i=1}^{n} \delta_i^E}\Big)$$

Substituting these expressions into the equation $\dot{l}_{\tau_E}$ yields a version of the score equation that does not contain the variable ϕ_E.

$$\dot{l}_{\tau_E} = \frac{1}{\tau_E}\sum_{i=1}^{n}\delta_i^E + \sum_{i=1}^{n}\delta_i^E \log(x_i) - \frac{\Big(\sum_{i=1}^{n}\delta_i^E\Big)\Big(\sum_{i=1}^{n} x_i^{\tau_E}\exp(\beta z_i)\log(x_i)\Big)}{\sum_{i=1}^{n} x_i^{\tau_E}\exp(\beta z_i)} \tag{4.12}$$

Our pseudo-ML estimate of the shape parameter τ_E, denoted $\widetilde{\tau}_E$, is obtained by plugging the partial likelihood estimate $\widehat{\beta}_C$ in for β in (4.12) and solving for τ_E in $\dot{l}_{\tau_E} = 0$.

The Breslow estimate of $S_{E|Z}(x|z)$ is defined over the interval $[0, A_E]$, and the terminal point of the semiparametric baseline survival curve is $(A_E, \widehat{S}_{E|Znp}(A_E|z=0))$. Next, we propose a pseudo-ML estimate $\widetilde{\phi}_E$ of the Weibull scale parameter ϕ_E that constrains the estimated Weibull tail of the baseline survival curve to connect with the terminal point of $\widehat{S}_{E|Znp}(x|z=0)$.

The baseline survival function for a covariate-adjusted Weibull-distributed variable has the form

$$S(x|z=0) = \exp\Big[-\Big(\frac{x}{\phi_E}\Big)^{\tau_E}\Big].$$

Substituting in the terminal point of the Breslow estimate $\widehat{S}_{E|Znp}$ of the baseline survival function yields

$$\widehat{S}_{spE|Z}(A_E|z=0) = \exp\Big[-\Big(\frac{A_E}{\phi_E}\Big)^{\tau_E}\Big].$$

Plugging in the pseudo-ML estimate $\widetilde{\tau}_E$ for τ_E and solving for ϕ_E yields an equation for our pseudo maximum likelihood estimate of ϕ_E.

$$\widetilde{\phi}_E = A_E\big\{-\log\big[\widehat{S}_{spE|Z}(A_E|z=0)\big]\big\}^{-1/\widetilde{\tau}_E}.$$

The estimates $\widetilde{\tau}_E$ and $\widetilde{\phi}_E$ define the Weibull tail of the estimated baseline survival function for an event on the range of observation times $x > A_E$. The Weibull estimates of the tails of the survival and hazard functions for an event have the following forms.

$$\widehat{S}_{pE|Z}(x|z) = \exp\Big[-\Big(\frac{x}{\widetilde{\phi}_E}\Big)^{\widehat{\tau}_E}\Big]^{\exp(\widehat{\beta}_C z)}$$

$$\widehat{\lambda}_{pE|Z}(x|z) = \frac{\widehat{\tau}_E}{\widetilde{\phi}_E}\Big(\frac{x}{\widetilde{\phi}_E}\Big)^{\widehat{\tau}_E - 1}\exp(\widehat{\beta}_C z)$$

Pseudo-maximum likelihood estimation of the tail of the dropout-time distribution

The estimated parametric tail for the dropout-time distribution is constructed analogously. We assume the tail of the hazard function for dropout is monotonic with an approximate power-dependence on time, and can therefore be approximated by a Weibull hazard function,

$$\lambda_L(x) = \frac{\tau_L}{\phi_L}\Big(\frac{x}{\phi_L}\Big)^{(\tau_L - 1)},$$

for $x \in (A_L, R]$. The score equations for the Weibull shape and scale parameters, (τ_L, ϕ_L) are

$$\dot{l}_{\tau_E} = \sum_{i=1}^{n} \Big\{ \frac{\delta_i^L}{\tau_L} + \delta_i^L \log(x_i) - \delta_i^L \log(\phi_L) - \big(\frac{x_i}{\phi_L}\big)^{\tau_L} \log(x_i) + \big(\frac{x_i}{\phi_L}\big)^{\tau_L} \log(\phi_L) \Big\}$$

$$\dot{l}_{\phi_L} = \sum_{i=1}^{n} \Big\{ -\delta_i^L \frac{\tau_L}{\phi_L} + \frac{\tau_L}{\phi_L^{(\tau_L+1)}} x_i^{\tau_L} \Big\}$$

Setting $\dot{l}_{\phi_L} = 0$, we obtain the following two equations relating ϕ_L and τ_L.

$$\phi_L^{\tau_L} = \frac{\sum_{i=1}^n x_i^{\tau_L}}{\sum_{i=1}^n \delta_i^L}$$

$$\log(\phi_L) = \frac{1}{\tau_L} \log\Big(\frac{\sum_{i=1}^n x_i^{\tau_L}}{\sum_{i=1}^n \delta_i^L}\Big)$$

Substituting these expressions into the score equation $\dot{l}_{\tau_L}$ yields a version of the score equation that only contains the variable τ_L and observed data in $\mathscr{F}_C$.

$$\dot{l}_{\tau_L} = \frac{1}{\tau_L} \sum_{i=1}^{n} \delta_i^L + \sum_{i=1}^{n} \delta_i^L \log(x_i) - \frac{\Big(\sum_{i=1}^n \delta_i^L\Big)\Big(\sum_{i=1}^n x_i^{\tau_L} \log(x_i)\Big)}{\sum_{i=1}^n x_i^{\tau_L}}$$

The standard ML estimate for the shape parameter τ_L is the root of the score equation $\dot{l}_{\tau_L} = 0$. Let $\widetilde{\tau}_L$ denote this standard ML estimate. The standard ML estimate for the scale parameter, ϕ_L would then be obtained by plugging $\widetilde{\tau}_L$ in for τ_L in the score equation $\dot{l}_{\phi_L} = 0$ and solving for the root. We wish to constrain the estimate of the parametric tail of the survival function for dropout to connect with the Nelson-Aalen estimate of the survival curve at time $x = A_L$. The Nelson-Aalen estimate of the function $S_L(x)$ is defined over the interval $[0, A_L]$. The terminal point of this survival function is $(A_L, \widehat{S}_{npL}(A_L))$. We propose an alternative, pseudo-ML estimate, $\widetilde{\phi}_L$, of the ϕ_L that constrains the estimated Weibull tail of the survival curve to connect with the terminal point of $\widehat{S}_{npL}(x)$.

The survival function for a Weibull-distributed dropout time has the form

$$S(x) = \exp\Big[- \Big(\frac{x}{\phi_L}\Big)^{\tau_L}\Big].$$

Substituting in the terminal point of the Nelson-Aalen estimate $\widehat{S}_{npL}$ yields

$$\widehat{S}_{npL}(A_L) = \exp\Big[- \Big(\frac{A_L}{\phi_L}\Big)^{\tau_L}\Big].$$

Plugging in the ML estimate $\widetilde{\tau}_L$ for τ and solving for ϕ_L yields an equation for our pseudo maximum likelihood estimate of ϕ_L.

$$\widetilde{\phi}_L = A_L\big\{ - \log\big[\widehat{S}_{npL}(A_L)\big]\big\}^{-1/\widehat{\tau}_L}.$$

The estimates $(\widehat{\tau}_L, \widetilde{\phi}_L)$ define the Weibull tail of the estimated survival function for dropout on the range of observation times $x > A_L$. The Weibull estimate of the tails of the survival function for dropout has the form

$$\widehat{S}_{pL}(x) = \exp\Big[- \Big(\frac{x}{\widetilde{\phi}_L}\Big)^{\widehat{\tau}_L}\Big].$$

4.2.6 Estimation of an individual's probability of experiencing an observed event

In equation 4.10, the estimated version of our conditional event-occurrence time model, the integrand is the patient-specific probability of experiencing an observed event in the time interval $(C, R]$ for a patient in the risk set $\mathscr{R}_C$. The estimate of this probability depends not only on the patient's measured baseline covariate values, z_i, but also the length of his or her observation time, x_{Ci}, relative to that of other patients. Let $t_1^E < t_2^E < \cdots < t_{K_C}^E$ denote the K_C distinct event times observed by the data cutoff time C, and let $t_1^L < t_2^L < \cdots < t_{J_C}^L$ denote the J_C distinct dropout times observed by C. The estimated individual probability of an observed event in $(C, R]$ is

$$\begin{aligned}
&\big\{\widehat{Pr}(C < x_i + w_i \le R, \delta_i^E = 1 | x_i + w_i > C, \mathscr{F}_C)\big\} \\
=&\left\{ \delta_{Ci}^A \cdot \frac{1}{\widehat{S}_{E|Z}(C - w_i|z_i)\widehat{S}_L(C - w_i)} \int_{C-w_i}^{R-w_i} \widehat{S}_L(u)\widehat{S}_{E|Z}(u|z_i) d\widehat{\Lambda}_{E|Z}(u|z_i) \right\}
\end{aligned}$$

The mechanics of calculating the expression

$$\frac{1}{\widehat{S}_{E|Z}(C-w_i|z_i)\widehat{S}_L(C-w_i)}\int_{C-w_i}^{R-w_i}\widehat{S}_L(u)\widehat{S}_{E|Z}(u|z_i)d\widehat{\Lambda}_{E|Z}(u|z_i)$$

differ slightly for each patient i depending on the patient's possible length of observation relative to the data cutoff time and the maximum observed times to event or dropout. That is, the use of parametric or nonparametric estimated survival and hazard functions depend on the relative positions of the times A_L, A_E, $(C-w_i)$, and $(R-w_i)$. For example in the denominator of the leading coefficient, we select

$$\widehat{S}_{E|Z}(C-w_i|z_i)=\begin{cases}\widehat{S}_{spE|Z}(C-w_i|z_i), & C-w_i<A_E\\ \widehat{S}_{pE|Z}(C-w_i|z_i), & C-w_i\geq A_E\end{cases},$$

where the subscript sp (semiparametric) denotes the Breslow estimate of the survival curve for an event, and p (parametric) denotes the Weibull estimate of the survival curve for an event. Similarly, with np denoting the Nelson-Aalen estimate and p denoting the Weibull estimate of the survival curve for dropout,

$$\widehat{S}_L(C-w_i)=\begin{cases}\widehat{S}_{npL}(C-w_i), & C-w_i<A_L\\ \widehat{S}_{pL}(C-w_i), & C-w_i\geq A_L\end{cases}.$$

There are also four different versions of the term $\int_{C-w_i}^{R-w_i}\widehat{S}_L(u)\widehat{S}_{E|Z}(u|z_i)d\widehat{\Lambda}_{E|Z}(u|z_i)$ comprised of parametric and nonparametric components, depending on where the time $(R-w_i)$ is relative to A_E and A_L.

Case 1: $(R-w_i)<A_L$ and $(R-w_i)<A_E$.

This patient's observation through study time R falls entirely within the range we may estimate with semiparametric and nonparametric survival and cumulative hazard functions.

$$\begin{aligned}&Pr(C-w_i<x_i\leq(R-w_i),\delta_i^E=1|\mathscr{F}_C)\\&=\sum_{x_{Ci}<t_k^E\leq(R-w_i)}\widehat{S}_{E|Znp}(t_k^E|z_i)\cdot\widehat{S}_{Lnp}(t_k^E)\cdot d\widehat{\Lambda}_{E|Znp}(t_k^E|z_i).\end{aligned}$$

Case 2: $A_L \leq (R - w_i) < A_E$.

Here we use the nonparametric estimates of the survival and cumulative hazard functions for events, and we use the nonparametric estimate of the survival curve for dropout for $x \in [x_{Ci}, A_L]$ and the parametric estimate of the survival curve for dropout for $x \in (A_L, (R-w_i)]$.

$$
\begin{aligned}
&Pr(C - w_i < x_i \leq (R - w_i), \delta_i^E = 1 | \mathscr{F}_C) \\
&= \sum_{x_{Ci} < t_k^E \leq A_L} \widehat{S}_{E|Znp}(t_k^E|z_i) \cdot \widehat{S}_{Lnp}(t_k^E) \cdot d\widehat{\Lambda}_{E|Znp}(t_k^E|z_i) \\
&+ \sum_{A_L \leq t_k^E \leq (R-w_i)} \widehat{S}_{E|Znp}(t_k^E|z_i) \cdot \widehat{S}_{Lp}(t_k^E) \cdot d\widehat{\Lambda}_{E|Znp}(t_k^E|z_i).
\end{aligned}
$$

Case 3: $A_L < A_E \leq (R - w_i)$.

For $x \in (x_{Ci}, A_L]$ use the nonparametric estimates of both the event- and dropout-time functions. For $x \in (A_L, A_E]$ use the nonparametric estimates of the event-time functions and the parametric estimate of the survival function for dropout. For $x \in (A_E, (R - w_i)]$ use the parametric estimates of both the event- and dropout-time functions.

$$
\begin{aligned}
&Pr(C - w_i < x_i \leq (R - w_i), \delta_i^E = 1 | \mathscr{F}_C) \\
&= \sum_{x_{Ci} < t_k^E \leq A_L} \widehat{S}_{E|Znp}(t_k^E|z_i) \cdot \widehat{S}_{Lnp}(t_k^E) \cdot d\widehat{\Lambda}_{E|Znp}(t_k^E|z_i) \\
&+ \sum_{A_L \leq t_k^E \leq A_E} \widehat{S}_{E|Znp}(t_k^E|z_i) \cdot \widehat{S}_{Lp}(t_k^E) \cdot d\widehat{\Lambda}_{E|Znp}(t_k^E|z_i) \\
&+ \int_{A_E}^{(R-w_i)} \widehat{S}_{E|Zp}(u|z_i) \cdot \widehat{S}_{Lp}(u) \cdot \widehat{\lambda}_{E|Zp}(u|z_i) du.
\end{aligned}
$$

Case 4: $A_E \leq (R - w_i)$ and $A_E \leq A_L$.

For $x \in (x_{Ci}, A_E]$ use the nonparametric estimates of both the event- and dropout-time functions. For $x \in (A_E, min\{A_L, (R - w_i)\}]$ use the parametric estimate of the survival function for dropout and nonparametric estimates of the event-time functions. If $(R-w_i) > A_L$, then or $x \in (A_L, (R-w_i)]$ use the parametric estimates of both the event- and dropout-

time functions.

$$
\begin{aligned}
&Pr(C - w_i < x_i \leq (R - w_i), \delta_i^E = 1 | \mathscr{F}_C) \\
&= \sum_{x_{Ci} < t_k^E \leq A_E} \widehat{S}_{E|Znp}(t_k^E|z_i) \cdot \widehat{S}_{Lnp}(t_k^E) \cdot d\widehat{\Lambda}_{E|Znp}(t_k^E|z_i) \\
&+ \sum_{A_E < t_j^L \leq min\{A_L, (R-w_i)\}} \widehat{S}_{E|Znp}(t_{(j-1)}^L|z_i) \int_{max\{A_E, t_{(j-1)}^L\}}^{t_j^L} \widehat{S}_{E|Zp}(u|z_i) \cdot \widehat{\lambda}_{E|Zp}(u|z_i) du \\
&+ 1[(R - w_i) > A_L] \cdot \int_{A_L}^{(R-w_i)} \widehat{S}_{E|Zp}(u|z_i) \cdot \widehat{S}_{Lp}(u) \cdot \widehat{\lambda}_{E|Zp}(u|z_i) du.
\end{aligned}
$$

4.2.7 A point prediction for the expected Nth observed event time

The approximation $\widehat{G}_{C_n}(R; N|\mathscr{F}_C)$ of $G(R; N|\mathscr{F}_C)$ is obtained using the observed data $\mathscr{F}_C$, the estimated event- and dropout-time functions described in sections 4.2.4 and 4.2.5, and the expressions for patient-specific probabilities of experiencing observed events described in section 4.2.6. The point prediction, $\widehat{E[R_N]}$, of the expected Nth observed event time, $E[R_N]$, is the solution to the equation $\widehat{G}_{C_n}(R; N|\mathscr{F}_C) = 0$. This root is obtained numerically using an algorithm that requires the function to be continuous and to have a unique root. Our estimating function $\widehat{G}_{C_n}(R; N|\mathscr{F}_C)$ satisfies these constraints because it is monotone increasing and continuous in R for $R > 0$, as long as the true event time probability distribution function f_{E^*} is never identically zero on an open interval (i.e. F_{E^*} has no "flat spots").

Note that $\widehat{G}_{C_n}(R; N|\mathscr{F}_C)$ is a constant term plus the sum of iid random quantities

$$
\left\{ \delta_{Ci}^E + \frac{\delta_i^A}{\widehat{S}_{E^*|Z^*}(C - w_i|z_i)\widehat{S}_L(C - w_i)} \int_{C-W_i}^{R-w_i} \widehat{S}_L(u)\widehat{S}_{E^*|Z^*}(u|z_i) d\widehat{\Lambda}_{E^*|Z^*}(u|z_i) \right\} \tag{4.13}
$$

Each summand is a nonlinear function of five finite-dimensional parameters ($\widehat{\beta}_C$, $\widetilde{\tau}_L$, $\widetilde{\phi}_L$, $\widetilde{\tau}_E$, $\widetilde{\phi}_E$) and two infinite dimensional parameters ($\widehat{\Lambda}_{npL}$, $\widehat{\Lambda}_{spE|Z}$). Since $\widetilde{\phi}_L$ is a deterministic function of $\widetilde{\tau}_L$ and $\widehat{\Lambda}_{npL}(A_L)$, and $\widetilde{\phi}_E$ is a deterministic function of $\widehat{\beta}_C$, $\widetilde{\tau}_E$ and $\widehat{\Lambda}_{spE|Z}(A_E)$, we may view each summand (4.13) more simply as a nonlinear function of the parameters $\widehat{\beta}_C$, $\widetilde{\tau}_L$, $\widetilde{\tau}_E$, $\widehat{\Lambda}_{npL}$, and $\widehat{\Lambda}_{spE|Z}$.

Furthermore, our estimator $\widehat{E[R_N]}$ is the product of multiple, sequential estimation steps, rather than the solution to a single estimating equation. First, using the data, $\mathscr{F}_C$,

Table 4.1: Sequential estimation steps for $\widehat{E[R_N]}$, and interdependence of finite-and infinite-dimensional parameters

Estimation	Parameter	Required values	Nonzero covariance with
Step 1	$\widehat{\beta}_C$	$\mathscr{F}_C$	$\widehat{\Lambda}_{spE\|Z}$, $\widetilde{\tau}_E$
	$\widetilde{\tau}_L$	$\mathscr{F}_C$	$\widehat{\Lambda}_{npL}$
	$\widehat{\Lambda}_{npL}$	$\mathscr{F}_C$	$\widetilde{\tau}_L$
Step 2	$\widetilde{\tau}_E$	$\mathscr{F}_C$, $\widehat{\beta}_C$	$\widehat{\beta}_C$, $\widehat{\Lambda}_{spE\|Z}$
	$\widehat{\Lambda}_{spE\|Z}$	$\mathscr{F}_C$, $\widehat{\beta}_C$	$\widetilde{\tau}_E$, $\widehat{\beta}_C$
Step 3	Find $\widehat{E[R_N]}$, implicitly defined as nonlinear function of $\widehat{\beta}_C$, $\widetilde{\tau}_L$, $\widetilde{\tau}_E$, $\widehat{\Lambda}_{npL}$, and $\widehat{\Lambda}_{spE\|Z}$		

we estimate the partial likelihood log hazard ratio $\widehat{\beta}_C$, the ML Weibull shape parameter for dropout, $\widetilde{\tau}_L$, and the Nelson-Aalen cumulative hazards for dropout, $\widehat{\Lambda}_{npL}$. Next the fixed value $\widehat{\beta}_C$ is used with the observed data to estimate $\widehat{\Lambda}_{spE|Z}(A_E)$ and $\widetilde{\tau}_E$. Finally, we calculate $\widehat{E[R_N]}$ as an implicitly defined nonlinear function of $\widehat{\beta}_C$, $\widetilde{\tau}_L$, $\widetilde{\tau}_E$, $\widehat{\Lambda}_{npL}$, and $\widehat{\Lambda}_{spE|Z}$. See Table 4.1.

The complexity of the functional form of each summand (4.13) creates considerable challenges to making inference about $\widehat{E[R_N]}$ based on asymptotic distribution theory. Although Von Mises calculus and the functional delta method could be used to provide theory for the asymptotic consistency and normality of $\widehat{E[R_N]}$, including a formula for the variance of $\widehat{E[R_N]}$, these derivations would be prohibitively complex to calculate and implement [48]. Furthermore, the sequential estimation procedure prohibits the use of convenient methods for semiparametric inference. For example, the form of $G(R; N|\mathscr{F}_C)$ is somewhat similar to that of an M-estimator for a sample quantile, described in detail by Serfling [46]. However M-estimation requires a stochastic integral of a deterministic function, and due to the sequential estimation procedure described above, our integrand is comprised of random functions. M-estimation procedures for obtaining the asymptotic normal distribution of a statistic fail to capture the additional variability inherent in our random integrand. As described in section 4.2.2, $G(R; N|\mathscr{F}_C)$ may also be viewed as a sum of Bernoulli random variables (plus a constant term). Thus we can easily argue that the estimating function

$\widehat{G}_{C_n}(R; N|\mathscr{F}_C)$ has an asymptotic normal distribution, which might lend itself to inference based on a pivot resampling method [42]. Unfortunately, since $\widehat{E[R_N]}$ cannot be obtained immediately from the data using $\widehat{G}_{C_n}$ in a single step, and instead results from multiple sequential estimates, the true asymptotic variance of $\widehat{G}_{C_n}$ is very difficult to quantify. A naive approach using the normal approximation to the binomial variance,

$$\widehat{G}_{C_n}(R; N|\mathscr{F}_C) \dot{\sim} \text{Normal}\Big(0, \frac{p_N(1-p_N)}{n}\Big),$$

significantly underestimates the true asymptotic variance. Instead, we rely on the asymptotic unbiasedness of each of the component finite- and infinite-dimensional parameters ($\widehat{\beta}_C$, $\widetilde{\tau}_L$, $\widetilde{\phi}_L$, $\widetilde{\tau}_E$, $\widetilde{\phi}_E$), $\widehat{\Lambda}_{npL}$, and $\widehat{\Lambda}_{spE|Z}$) to justify the consistency of our prediction $\widehat{E[R_N]}$. We demonstrate the consistency of the point prediction with simulation studies, and describe a bootstrap resampling procedure to obtain a prediction interval for $\widehat{E[R_N]}$ on which prediction uncertainty may be gauged.

Heuristic argument for the consistency of $\widehat{G}_{C_n}(R; N|\mathscr{F}_C) = 0$ as an estimator of $E[R_N]$

Here we present a heuristic argument to justify the consistency of $\widehat{E[R_N]}$ under certain conditions. Consider a clinical trial in which the loss-to-follow-up times have distribution F_L, and event times have true distribution $F_{E^*} = F_{E^*|Z^*}F_{Z^*}$ for a set of baseline covariates Z^*. First, assume that the true dropout time distribution, F_L, and $F_{E^*|Z^*}$, the true event time distribution, conditional on baseline covariates Z^*, are independent and differentiable. Then

$$G(R; N|\mathscr{F}_C) = N_C - N + \sum_{i=1}^{n}\left\{\delta_i^A \int_{Z^*}\int_{C}^{R} \frac{S_L(u)S_{E^*|Z^*}(u|z)d\Lambda_{E^*|Z^*}(u|z)dF_{Z^*}(z)}{S_L(C)S_{E^*|Z^*}(C|z_i)}\right\}. \quad (4.14)$$

Replacing each of the component functions in the integrand of (4.6) with estimates based on $\mathscr{F}_C$ yields

$$\widehat{G}_{Cn}(R; N|\mathscr{F}_C) =$$
$$= N_C - N + \sum_{i=1}^{n}\left\{\frac{\delta_i^A}{\widehat{S}_{E^*|Z^*}(C-w_i|z_i)\widehat{S}_L(C-w_i)}\int_{C-W_i}^{R-w_i}\widehat{S}_L(u)\widehat{S}_{E^*|Z^*}(u|z_i)d\widehat{\Lambda}_{E^*|Z^*}(u|z_i)\right\} \quad (4.15)$$

Under certain assumptions, each of the estimated component functions is consistent for the true function it approximates. Note that each finite- and infinite-dimensional parameter is estimable at the $\sqrt{n}$ rate.

- Assuming the proportional hazards relative risk model is valid, and baseline covariate measurements Z^* are bounded, the partial likelihood estimate $\widehat{\beta}_C$ of the log hazard ratio β is consistent and asymptotically normal, $(\widehat{\beta}_p - \beta_p) \approx N(0, \widehat{I}^{pp})$, where $\widehat{I}^{pp}$ is the pth diagonal element of $I(\widehat{\beta})^{-1}$. Details of the minimum sufficient conditions for asymptotic unbiasedness and normality are described by Kalbfleisch and Prentice [30].

- Assume we have data collected over the finite interval $[0, C]$. If the number of patients at risk of dropout at any specific time $t \in [0, C]$ grows large as $n \to \infty$, then the Nelson Aalen estimator $\widehat{\Lambda}_{npL}$ can be shown to be uniformly consistent on $[0, C]$, with $n^{1/2}(\widehat{\Lambda}_{npL}(t) - \Lambda_L(t))$ weakly convergent to a mean zero Gaussian process. Due to the continuity of F_L, this also ensures the consistency of $\widehat{S}_{npL}$ [25][30].

- For right- and independently censored data from the proportional hazards relative risk model, the Breslow estimate of the baseline cumulative hazard function for an event converges weakly to a mean zero Gaussian process on $[0, C]$,

  $$n^{1/2}(\widehat{\Lambda}_{spE|Z}(t|z = 0) - \Lambda_{spE|Z}(t|z = 0)).$$

  The asymptotic unbiasedness of $\widehat{S}_{spE^*|Z^*}(t)$ and $d\widehat{\Lambda}_{spE^*|Z^*}(t)$ follow [25][30].

- Assume the time-to-dropout follows a Weibull distribution over the interval $[A_L, R]$. The ML estimate of the Weibull shape parameter for the tail of the dropout distribution, $\widetilde{\tau}_L$ is asymptotically unbiased and normal [44]. The Nelson-Aalen estimate of the terminal point of the survival curve for dropout, $\widehat{S}_{npL}(A_L)$, is also asymptotically unbiased and normal. Then, since the pseudo-ML estimate $\widetilde{\phi}_L$ is a differentiable function of $\widetilde{\tau}_L$ and $\widehat{S}_{npL}(A_L)$, $\widetilde{\phi}_L$ is asymptotically unbiased and normal by the delta method [9]. With consistent estimates of the Weibull shape and scale parameters, our estimate of the tail of the survival function for dropout, $\widehat{S}_{pL}(t)$, is consistent for the tail of $S_L(t)$.

- Suppose the event-time conditional distribution $F_{E^*|Z^*}$ follows a proportional hazards Weibull distribution over the interval $[A_E, R]$. The standard ML estimate of the Weibull shape parameter, $\tau_{E^*|Z^*}$, is asymptotically unbiased and normal. The pseudo-ML estimate $\widetilde{\tau}_{E^*|Z^*}$ is found using the log hazard ratio estimate $\widehat{\beta}_C$ instead of the standard ML estimate of β. The consistency of the estimate $\widehat{\beta}_C$ for β ensures that $\widetilde{\tau}_{E^*|Z^*}$ is also consistent for $\tau_{E^*|Z^*}$. Similarly, since $\widetilde{\phi}_{E^*|Z^*}$ is a differentiable function of the consistent estimates $\widetilde{\tau}_{E^*|Z^*}$ and $\widehat{S}_{spE^*|Z^*}(A_E|z=0)$, it is consistent for ϕ_E. Consequently our estimates of the tail of the baseline survival function for an event

$$\widehat{S}_{spE^*|Z^*}(t|z=0) = \exp\Big[-\Big(\frac{t}{\widetilde{\phi}_E}\Big)^{\widetilde{\tau}_E}\Big]^{\exp(\widehat{\beta}_C z)}$$

  and the tail of the hazard function for an event

$$\widehat{\lambda}_{pE^*|Z^*}(t|z) = \frac{\widetilde{\tau}_E}{\widetilde{\phi}_E}\Big(\frac{t}{\widetilde{\phi}_E}\Big)^{\widetilde{\tau}_E - 1}\exp(\widehat{\beta}_C z)$$

  are consistent for the tails of $S_{E^*|Z^*}(t)$ and $\lambda_{E^*|Z^*}(t)$.

When the above conditions on $F_{E^*|Z^*}$ and F_L for consistency of the component functions are satisfied, the estimating equation 4.6 is differentiable in R and yields a consistent point estimator $\widehat{E[R_N]}$ for $E[R_N]$.

The above arguments for consistency of our finite- and infinite-dimensional parameters, and thus the point estimator $\widehat{E[R_N]}$, depend on knowledge of the true event- and dropout-time models insofar as selection of the correct set of baseline covariates, the proportional hazards assumption across the entire study time interval, and properly specified tail distributions. Thus even with an asymptotically unbiased estimator and valid prediction interval for the expected time of the Nth observed event, the accuracy and utility of predictions depend strongly on the investigator's careful selection of baseline covariates in the working model $F_{E|Z}$ for event time. Failure to account for strongly prognostic baseline covariates may result in biased point and interval predictions.

4.2.8 A Bootstrap resampling method for prediction interval calculation

The bootstrap method may be used under very unrestrictive conditions to calculate approximate variance and prediction intervals in theoretically complex estimation problems [19].

When data is right-censored under a random censoring mechanism, rather than complete, the bootstrap method still provides valid asymptotic error measurements [20]. We calculate a bootstrap prediction interval for the expected time of the Nth observed event using the bootstrap distribution of this statistic and a percentile method construction.

Recall that our observed data

$$\mathscr{F}_C = \big\{w_i, z_i, x_{Ci}, \delta^E_{Ci}, \delta^L_{Ci}, \delta^A_{Ci}; i = 1, \ldots, n\big\}$$

are an iid sample from the conditionally independent distributions $F_{E^*|Z^*}$ and F_L. The point prediction $\widehat{E[R_N]}$ is the root of the equation $\widehat{G}_{C_n}(R; N|\mathscr{F}_C) = 0$. We construct the bootstrap distribution of $\widehat{E[R_N]}$ as follows. A bootstrap sample $\mathscr{F}^*_C$ is obtained by randomly selecting n times with replacement from the collection of observed data points

$$\mathbb{X}_{Ci} = \big\{w_i, z_i, x_{Ci}, \delta^E_{Ci}, \delta^L_{Ci}, \delta^A_{Ci}\big\}.$$

A bootstrap estimate $\widehat{E[R_N]}^*$ is calculated as the root of $\widehat{G}_{C_n}(R; N|\mathscr{F}^*_C) = 0$ for each bootstrap sample. This process is repeated $B = 1000$ times. The cumulative distribution function for $E[R_N]$ is estimated by the empirical cumulative distribution function of the observed bootstrap estimates $\widehat{E[R_N]}^*$,

$$\mathrm{CDF}_*(t) = \mathrm{Pr}_*\big\{\widehat{E[R_N]}^* \leq t\big\}.$$

For a given significance level α, we construct a $1-\alpha$ central prediction interval by selecting as the lower and upper endpoints, $\widehat{E[R_N]}^*_{LOW} = \mathrm{CDF}^{-1}_*(\alpha/2)$ and $\widehat{E[R_N]}^*_{UP} = \mathrm{CDF}^{-1}_*(1-\alpha/2)$, respectively.

The bootstrap prediction interval calculated in this way is interpreted as the 95% prediction interval of the expected time of the Nth observed event among n patients, $E[R_N]$, that would result from an infinitely large data set collected and analyzed in the same way as our actual data set $\mathscr{F}_C$ of a finite number of participants [20].

4.3 Simulation studies

4.3.1 Simulation studies of the pseudo-maximum likelihood estimates of Weibull distribution parameters

We performed simulation studies to demonstrate the performance of the pseudo-ML Weibull parameter estimates compared to the standard ML estimates. For Weibull-distributed time-to-event data, the pseudo-ML and standard ML estimates are nearly identical. Table 4.2 compares pseudo- and standard ML estimates of shape and scale for a range of Weibull distributions with no covariates. This scenario corresponds to modeling the tail of the dropout distribution, or the tail of the event distribution when no baseline covariates are included in our prediction method. Note that when the time-to-event working model includes no baseline covariates, we use the standard ML shape parameter estimate; only the scale parameter estimate is modified.

If one or more covariates influence time-to-event in proportional hazards Weibull data, exclusion of those variables from the working prediction model results in biased parameter estimates. Table 4.3 compares pseudo- and standard ML estimates of shape and scale for a Weibull(shape: $\tau = 0.8$, scale: $\phi = 30$) distribution in the presence of an unmodeled binary covariate. For example this could correspond to estimation for an event-time working model (with no covariate adjustment) in a two-arm randomized (1:1) clinical trial, in which treatment assignment is excluded from the prediction model to maintain confidentiality of the treatment effect at an interim study time. Estimation of the shape parameter τ is largely unaffected by the unmodeled covariate. However, ML and pseudo-ML estimates of the scale parameter ϕ may be substantially biased if covariates with large hazard ratios are not accounted for in the chosen working model. Once again, the bias and sample standard error are nearly identical for the ML and pseudo-ML estimates.

When baseline covariates are included in the proportional hazards Weibull event-time working model, none of the parameter estimates (log hazard ratios, shape, scale) are obtained with standard ML. Table 4.4 compares standard ML versus partial likelihood log hazard ratio parameter estimates, and standard ML versus pseudo-ML estimates for Weibull shape and scale parameters, when time-to-event is distributed as proportional hazards

Table 4.2: Comparison of standard ML Weibull shape and scale parameter estimates with pseudo-ML estimates, when event times are distributed Weibull(shape: τ, scale: ϕ).

		Estimated Weibull shape (τ) and scale (ϕ) parameters					
		τ: ML		ϕ: ML		ϕ: pseudo-ML	
True parameters	no. events	Bias	(SSE)	Bias	(SSE)	Bias	(SSE)
$\tau = 0.8, \phi = 30$	50	0.03	(0.12)	2.48	(21.58)	2.48	(21.57)
	100	0.02	(0.08)	0.43	(9.93)	0.42	(9.93)
	200	0.01	(0.06)	-0.02	(4.87)	-0.02	(4.87)
$\tau = 1.0, \phi = 60$	50	0.04	(0.15)	2.21	(31.35)	2.20	(31.34)
	100	0.02	(0.10)	0.21	(15.48)	0.20	(15.47)
	200	0.01	(0.07)	-0.15	(7.76)	-0.16	(7.76)
$\tau = 1.1, \phi = 80$	50	0.05	(0.17)	1.94	(36.91)	1.93	(36.91)
	100	0.02	(0.11)	0.04	(18.61)	0.03	(18.60)
	200	0.01	(0.08)	-0.24	(9.39)	-0.25	(9.39)
$\tau = 1.3, \phi = 20$	50	0.05	(0.20)	0.18	(7.51)	0.18	(7.51)
	100	0.03	(0.13)	-0.06	(3.89)	-0.06	(3.89)
	200	0.01	((0.09)	-0.07	(1.98)	-0.07	(1.98)

NOTE: Total sample size is $n = 1000$, and values are averages over 5000 simulated datasets. Bias: sample mean of the differences between estimates and their true values. SSE: sample standard error of the estimates.

Table 4.3: Comparison of standard ML Weibull shape and scale parameter estimates with pseudo-ML estimates, when event times are distributed Weibull(shape: $\tau = 0.8$, scale: $\phi = 30$), in the presence of an unmodeled binary covariate.

		Estimated Weibull shape (τ) and scale (ϕ) parameters					
		τ: ML		ϕ: ML		ϕ: pseudo-ML	
HR for unmodeled covariate	no. events	Bias	(SSE)	Bias	(SSE)	Bias	(SSE)
$e^{\beta} = 0.8$	50	0.03	(0.12)	6.50	(22.79)	6.49	(22.78)
	100	0.02	(0.08)	4.60	(11.03)	4.59	(11.02)
	200	0.01	(0.06)	4.45	(5.57)	4.45	(5.57)
$e^{\beta} = 1.0$	50	0.03	(0.12)	2.48	(21.58)	2.48	(21.57)
	100	0.02	(0.08)	0.43	(9.93)	0.42	(9.93)
	200	0.01	(0.06)	-0.02	(4.87)	-0.02	(4.87)
$e^{\beta} = 1.2$	50	0.03	(0.12)	-1.61	(17.90)	-1.62	(17.90)
	100	0.02	(0.08)	-3.03	(8.60)	-3.04	(8.59)
	200	0.01	(0.06)	-3.29	(4.38)	-3.29	(4.38)
$e^{\beta} = 1.5$	50	0.03	(0.12)	-5.79	(15.27)	-5.79	(15.27)
	100	0.01	(0.08)	-6.84	(7.39)	-6.85	(7.39)
	200	0.01	(0.06)	-7.07	(3.74)	-7.07	(3.74)
$e^{\beta} = 2.0$	50	0.03	(0.12)	-10.59	(12.34)	-10.59	(12.33)
	100	0.01	(0.08)	-11.36	(5.95)	-11.36	(5.95)
	200	0.00	(0.06)	-11.42	(3.07)	-11.42	(3.07)

NOTE: Total sample size is $n = 1000$, and values are averages over 5000 simulated datasets. Bias: sample mean of the differences between estimates and their true values. SSE: sample standard error of the estimates.

Weibull(shape:$\tau = 0.8$, scale:$\phi = 30$) with a binary covariate. The bias and sample standard error of ML and non-ML coincide almost exactly when the model is correctly specified. For data in which proportional hazards holds but the baseline hazard function deviates from the Weibull model, we would expect some differences. Then the estimate of β will generally be more accurate with partial likelihood than with ML. Consequently, the pseudo-ML shape and scale estimates, which depend on estimates of the partial likelihood β as well as the nonparametric baseline survival function, will yield a tail distribution more representative of the observed data than one obtained via standard ML estimation.

When the working model includes some, but not all of the patient covariates governing time-to-event, parameter estimates may be biased. Table 4.5 compares standard ML versus partial likelihood log hazard ratio parameter estimates, and standard ML versus pseudo-ML estimates for Weibull shape and scale parameters, when time-to-event is distributed as proportional hazards Weibull(shape:$\tau = 0.8$, scale:$\phi = 30$) with a binary covariate (hazard ratio: $\exp \beta = 1.2$), in the presence of an unmodeled binary covariate with varying levels of influence. When the data are truly proportional hazards Weibull-distributed, ML and non-ML estimates of the hazard ratio for the modeled covariate and the Weibull shape parameter coincide and are accurate. The scale parameter estimates have bias that increases in severity with larger effects of the unmodeled covariate. ML and pseudo-ML scale parameter estimates have the same accuracy and variability for the simulated Weibull data.

Table 4.4: Comparison of standard ML versus partial likelihood log hazard ratio parameter estimates, and standard ML versus pseudo-ML estimates for Weibull shape and scale parameters, when time-to-event is distributed as proportional hazards Weibull(shape:$\tau = 0.8$, scale:$\phi = 30$) with a binary covariate.

		Estimated coefficient (β) and Weibull shape (τ) and scale (ϕ) parameters											
		β: ML		β: Cox		τ: ML		τ: pseudo-ML		ϕ: ML		ϕ: pseudo-ML	
HR for binary covariate	no. events	Bias	(SSE)	Bias	(SSE)	Bias	(SSE)	Bias	(SSE)	Bias	(SSE)	Bias	(SSE)
$e^{\beta} = 0.8$	50	-0.01	(0.29)	-0.01	(0.29)	0.03	(0.12)	0.03	(0.12)	2.40	(20.97)	2.42	(20.98)
($\beta = -0.22$)	100	-0.01	(0.20)	-0.01	(0.20)	0.02	(0.08)	0.02	(0.08)	0.48	(10.18)	0.50	(10.18)
	200	0.00	(0.14)	0.00	(0.14)	0.01	(0.06)	0.01	(0.06)	0.23	(5.33)	0.25	(5.33)
$e^{\beta} = 1.0$	50	0.00	(0.28)	0.00	(0.29)	0.03	(0.12)	0.03	(0.12)	2.81	(22.32)	2.83	(22.33)
($\beta = 0$)	100	0.00	(0.20)	0.00	(0.20)	0.02	(0.08)	0.02	(0.08)	0.70	(10.72)	0.72	(10.73)
	200	0.00	(0.14)	0.00	(0.14)	0.01	(0.06)	0.01	(0.06)	0.26	(5.71)	0.28	(5.71)
$e^{\beta} = 1.2$	50	0.00	(0.29)	0.00	(0.29)	0.03	(0.12)	0.03	(0.12)	3.33	(23.94)	3.34	(23.95)
($\beta = 0.18$)	100	0.00	(0.20)	0.00	(0.20)	0.02	(0.08)	0.02	(0.08)	0.73	(11.18)	0.74	(11.19)
	200	0.00	(0.14)	0.00	(0.14)	0.01	(0.06)	0.01	(0.06)	0.26	(6.05)	0.27	(6.06)
$e^{\beta} = 1.5$	50	0.00	(0.29)	0.00	(0.29)	0.03	(0.12)	0.03	(0.12)	3.82	(25.76)	3.83	(25.77)
($\beta = 0.41$)	100	0.00	(0.21)	0.00	(0.20)	0.02	(0.08)	0.02	(0.08)	1.05	(12.11)	1.06	(12.12)
	200	0.00	(0.15)	0.00	(0.14)	0.01	(0.06)	0.01	(0.06)	0.27	(6.39)	0.29	(6.39)

NOTE: Total sample size is $n = 1000$, and values are averages over 5000 simulated datasets. Bias: sample mean of the differences between estimates and their true values. SSE: sample standard error of the estimates.

Table 4.5: Comparison of standard ML versus partial likelihood log hazard ratio parameter estimates, and standard ML versus pseudo-ML estimates for Weibull shape and scale parameters, when time-to-event is distributed as proportional hazards Weibull(shape:$\tau = 0.8$, scale:$\phi = 30$) with a binary covariate (hazard ratio: $\exp\beta = 1.2$), in the presence of an unmodeled binary covariate.)

		Estimated coefficient (β) and Weibull shape (τ) and scale (ϕ) parameters											
		β: ML		β: Cox		τ: ML		τ: pseudo-ML		ϕ: ML		ϕ: pseudo-ML	
HR for unmodeled covariate	no. events	Bias	(SSE)	Bias	(SSE)	Bias	(SSE)	Bias	(SSE)	Bias	(SSE)	Bias	(SSE)
$e^\beta = 0.8$	50	0.00	(0.29)	0.00	(0.29)	0.03	(0.12)	0.03	(0.12)	8.82	(28.34)	8.83	(28.36)
($\beta = -0.22$)	100	0.00	(0.21)	0.00	(0.20)	0.01	(0.08)	0.01	(0.08)	5.71	(13.33)	5.72	(13.33)
	200	0.00	(0.14)	0.00	(0.14)	0.01	(0.06)	0.01	(0.06)	4.86	(6.97)	4.88	(6.98)
$e^\beta = 1.0$	50	0.00	(0.30)	0.00	(0.29)	0.03	(0.12)	0.03	(0.12)	3.97	(24.47)	3.99	(24.48)
($\beta = 0$)	100	0.00	(0.21)	0.00	(0.20)	0.02	(0.08)	0.02	(0.08)	1.21	(11.55)	1.23	(11.56)
	200	0.00	(0.14)	0.00	(0.14)	0.01	(0.06)	0.01	(0.06)	0.44	(6.09)	0.45	(6.09)
$e^\beta = 1.2$	50	0.01	(0.30)	0.00	(0.29)	0.03	(0.12)	0.03	(0.12)	0.29	(21.83)	0.30	(21.85)
($\beta = 0.18$)	100	0.00	(0.21)	0.00	(0.20)	0.01	(0.08)	0.01	(0.08)	-2.30	(10.18)	-2.28	(10.18)
	200	0.00	(0.14)	0.00	0.14)	0.01	(0.06)	0.01	(0.06)	-3.00	(5.37)	-2.99	(5.37)
$e^\beta = 1.5$	50	0.00	(0.30)	0.00	(0.29)	0.03	(0.12)	0.03	(0.12)	-4.06	(18.55)	-4.05	(18.56)
($\beta = 0.41$)	100	0.00	(0.21)	0.00	(0.20)	0.01	(0.08)	0.01	(0.08)	-6.28	(8.73)	-6.27	(8.73)
	200	0.00	(0.14)	0.00	(0.14)	0.01	(0.060	0.01	(0.06)	-6.84	(4.60)	-6.83	(4.60)
$e^\beta = 2.0$	50	0.00	(0.30)	0.00	(0.29)	0.03	(0.12)	0.03	(0.12)	-9.10	(15.48)	-9.09	(15.49)
($\beta = 0.41$)	100	0.00	(0.21)	0.00	(0.20)	0.01	(0.08)	0.01	(0.08)	-10.93	(7.06)	-10.92	(7.07)
	200	0.00	(0.14)	0.00	(0.14)	0.00	(0.06)	0.00	(0.06)	-11.25	(3.75)	-11.24	(3.75)

NOTE: Total sample size is $n = 1000$, and values are averages over 5000 simulated datasets. Bias: sample mean of the differences between estimates and their true values. SSE: sample standard error of the estimates.

4.3.2 Simulation studies of the accuracy of the point estimator for the Nth observed event time

Simulation studies were used to investigate the behavior of our point estimator $\widehat{E[R_N]}$ of the expected time of the Nth observed event, conditional on the data observed from a chosen information fraction cutoff time. Table 4.6 shows event-time predictions for data generated from a proportional hazards Weibull distribution with a normal baseline covariate. (Tables 4.7 and 4.8 demonstrate uniform and binary covariates, respectively.) Observation times (in years) for participants were randomly censored at an approximate rate of 3% per year. The bias in our point predictions is averaged over 5000 data sets, each with a sample size of $n = 2000$. The hazard ratio for the normal covariate ranges from 0.8 to 2.0.

In general, we see that bias is quite small when the model is correctly specified. The timing of an event designating a particular information fraction may be well approximated by the data available at a much smaller information fraction, for example prediction of the time at which 20% of the information will be accrued when only 5% is available, or prediction of the time of the landmark event for 50% of the information when 30% has been observed. Bias is low for near predictions and increases for times of larger (more distant) numbers of events. When the available information is low (5%) the bias in predicting a large information fraction (50%) can be significant.

More precisely, the magnitude and direction of the bias in our point predictor are due to two finite-sample factors. First, the actual number of observed events available for the prediction model determines the accuracy of the individual parameter estimates, regardless of the information fraction to which that corresponds in a particular clinical trial. For example, the ML estimates of the Weibull parameters, in particular the scale parameter, tend to have a positive finite sample bias. A critical number events is necessary to achieve adequate precision.

Second, two patterns of bias visible throughout the simulations in this section are due to the estimation procedures for certain components in the estimating function $\widehat{G}_{C_n}(R; N|\mathscr{F}_C)$. Note that the bias for predicting numbers of events closer in time to the cutoff tends to be slightly negative, then becomes positive and increases in magnitude for farther away

Table 4.6: Bias and sample standard error (SSE) of semiparametric point predictions (in years), when event times are distributed Weibull(shape: $\tau = 0.8$, scale: $\phi = 30$) with modeled covariate $Z \sim \text{Normal}(0, 1)$ having varying levels of influence on the hazard of an event. Dropout times are distributed Weibull($\tau_L = 1$, $\phi_L = 30$), approximately 3% dopout per year.

		Proportion (number) of sample population with an observed event							
Target time		0.05 (100)		0.10 (200)		0.20 (400)		0.30 (600)	
(prop. obs. events)	Ave. true time	Bias	SSE	Bias	SSE	Bias	SSE	Bias	SSE
True hazard ratio for Z: $e^\beta = 0.8$									
R_{400} (0.20)	1.55	-0.02	0.38	-0.04	0.18	—	—	—	—
R_{600} (0.30)	2.88	0.09	1.03	-0.03	0.45	-0.05	0.22	—	—
R_{800} (0.40)	4.73	0.32	2.18	0.05	1.01	-0.06	0.45	-0.06	0.30
R_{1000} (0.50)	7.45	0.49	3.77	0.39	2.55	-0.03	0.95	-0.07	0.59
True hazard ratio for Z: $e^\beta = 1.2$									
R_{400} (0.20)	1.56	-0.02	0.38	-0.04	0.18	—	—	—	—
R_{600} (0.30)	2.90	0.08	0.99	-0.03	0.45	-0.04	0.22	—	—
R_{800} (0.40)	4.75	0.30	2.21	0.04	1.02	-0.05	0.45	-0.05	0.31
R_{1000} (0.50)	7.47	0.55	3.95	0.34	2.42	-0.01	0.95	-0.06	0.60
True hazard ratio for Z: $e^\beta = 1.5$									
R_{400} (0.20)	1.27	-0.02	0.35	-0.04	0.17	—	—	—	—
R_{600} (0.30)	2.46	0.07	0.95	-0.03	0.43	-0.04	0.22	—	—
R_{800} (0.40)	4.22	0.27	2.07	0.05	1.00	-0.04	0.45	-0.06	0.30
R_{1000} (0.50)	7.06	0.40	3.53	0.37	2.36	0.00	0.97	-0.07	0.61
True hazard ratio for Z: $e^\beta = 2.0$									
R_{400} (0.20)	1.27	-0.02	0.31	-0.04	0.15	—	—	—	—
R_{600} (0.30)	2.46	0.06	0.88	-0.04	0.40	-0.04	0.20	—	—
R_{800} (0.40)	4.22	0.22	1.89	0.04	0.97	-0.05	0.44	-0.06	0.30
R_{1000} (0.50)	7.06	0.39	3.78	0.39	2.54	0.01	1.03	-0.07	0.65

NOTE: Total sample size is $n = 2000$, and values are averages over 5000 simulated datasets. Bias: sample mean of the differences between predictions and their true values. SSE: sample standard error of the estimates.

Table 4.7: Bias and sample standard error (SSE) of semiparametric point predictions (in years), when event times are distributed Weibull(shape: $\tau = 0.8$, scale: $\phi = 30$) with modeled covariate $Z \sim \text{Uniform}(0, 1)$ having varying levels of influence on the hazard of an event. Dropout times are distributed Weibull($\tau_L = 1$, $\phi_L = 30$), approximately 3% dopout per year.

		Proportion (number) of sample population with an observed event							
Target time		0.05 (100)		0.10 (200)		0.20 (400)		0.30 (600)	
(prop. obs. events)	Ave. true time	Bias	SSE	Bias	SSE	Bias	SSE	Bias	SSE
True hazard ratio for Z: $e^\beta = 0.8$									
R_{400} (0.20)	1.83	-0.02	0.46	-0.04	0.21	—	—	—	—
R_{600} (0.30)	3.40	0.12	1.27	-0.02	0.52	-0.04	0.26	—	—
R_{800} (0.40)	5.61	0.37	2.64	0.08	1.19	-0.05	0.54	-0.06	0.36
R_{1000} (0.50)	8.91	0.56	4.78	0.48	3.00	0.01	1.17	-0.07	0.73
True hazard ratio for Z: $e^\beta = 1.2$									
R_{400} (0.20)	1.41	-0.02	0.35	-0.03	0.16	—	—	—	—
R_{600} (0.30)	2.60	0.05	0.87	-0.03	0.39	-0.04	0.19	—	—
R_{800} (0.40)	4.23	0.25	1.85	0.03	0.86	-0.04	0.39	-0.05	0.26
R_{1000} (0.50)	6.57	0.41	3.22	0.28	2.00	-0.01	0.79	-0.06	0.50
True hazard ratio for Z: $e^\beta = 1.5$									
R_{400} (0.20)	1.21	-0.02	0.29	-0.03	0.14	—	—	—	—
R_{600} (0.30)	2.24	0.04	0.74	-0.03	0.33	-0.04	0.17	—	—
R_{800} (0.40)	3.62	0.19	1.55	0.02	0.72	-0.04	0.33	-0.05	0.22
R_{1000} (0.50)	5.58	0.38	2.75	0.21	1.58	-0.03	0.65	-0.06	0.42
True hazard ratio for Z: $e^\beta = 2.0$									
R_{400} (0.20)	1.00	-0.02	0.24	-0.03	0.11	—	—	—	—
R_{600} (0.30)	1.84	0.03	0.61	-0.03	0.27	-0.04	0.14	—	—
R_{800} (0.40)	2.96	0.12	1.17	0.00	0.57	-0.04	0.27	-0.05	0.18
R_{1000} (0.50)	4.54	0.31	2.19	0.14	1.20	-0.04	0.51	-0.06	0.33

NOTE: Total sample size is $n = 2000$, and values are averages over 5000 simulated datasets. Bias: sample mean of the differences between predictions and their true values. SSE: sample standard error of the estimates.

Table 4.8: Bias and sample standard error (SSE) of semiparametric point predictions (in years), when event times are distributed Weibull(shape: $\tau = 0.8$, scale: $\phi = 30$) with modeled covariate $Z \sim$ Bernoulli(0.5) having varying levels of influence on the hazard of an event. Dropout times are distributed Weibull($\tau_L = 1$, $\phi_L = 30$), approximately 3% dopout per year.

		Proportion (number) of sample population with an observed event							
Target timet		0.05 (100)		0.10 (200)		0.20 (400)		0.30 (600)	
(prop. obs. events)	Ave. true time	Bias	SSE	Bias	SSE	Bias	SSE	Bias	SSE
True hazard ratio for Z: $e^\beta = 0.8$									
R_{400} (0.20)	1.82	-0.02	0.45	-0.04	0.21	—	—	—	—
R_{600} (0.30)	3.39	0.09	1.22	-0.02	0.53	-0.04	0.26	—	—
R_{800} (0.40)	5.59	0.32	2.54	0.08	1.20	-0.04	0.54	-0.06	0.36
R_{1000} (0.50)	8.89	0.51	4.62	0.51	3.15	0.01	1.16	-0.07	0.73
True hazard ratio for Z: $e^\beta = 1.2$									
R_{400} (0.20)	1.41	-0.02	0.34	-0.03	0.16	—	—	—	—
R_{600} (0.30)	2.59	0.05	0.87	-0.03	0.39	-0.04	0.19	—	—
R_{800} (0.40)	4.22	0.24	1.81	0.04	0.87	-0.04	0.39	-0.05	0.26
R_{1000} (0.50)	6.56	0.43	3.24	0.28	1.97	-0.02	0.79	-0.06	0.51
True hazard ratio for Z: $e^\beta = 1.5$									
R_{400} (0.20)	1.20	-0.03	0.29	-0.04	0.14	—	—	—	—
R_{600} (0.30)	2.21	0.03	0.72	-0.03	0.33	-0.04	0.17	—	—
R_{800} (0.40)	3.59	0.18	1.51	0.01	0.72	-0.04	0.33	-0.05	0.22
R_{1000} (0.50)	5.54	0.40	2.76	0.19	1.53	-0.03	0.65	-0.06	0.42
True hazard ratio for Z: $e^\beta = 2.0$									
R_{400} (0.20)	0.96	-0.03	0.23	-0.03	0.11	—	—	—	—
R_{600} (0.30)	1.78	0.02	0.57	-0.04	0.27	-0.04	0.13	—	—
R_{800} (0.40)	2.89	0.15	1.21	0.00	0.57	-0.04	0.26	-0.05	0.18
R_{1000} (0.50)	4.47	0.33	2.27	0.13	1.25	-0.04	0.52	-0.06	0.34

NOTE: Total sample size is $n = 2000$, and values are averages over 5000 simulated datasets. Bias: sample mean of the differences between predictions and their true values. SSE: sample standard error of the estimates.

predictions (or times of larger numbers of events). The culprit behind both trends is the (negative) finite sample bias of the nonparametric estimate of cumulative hazard (Breslow or Nelson-Aalen), which is well documented in the statistical literature and is summarized in seminal survival analysis texts [25][30][47]. The closer a target landmark event time is to the information cutoff, the more the estimate of an at-risk individual's probability of an observed event in $[C, R]$ depends on the semi-and nonparametric estimates. The bias of these infinite-dimensional parameters slightly inflates the integrand function in $\widehat{G}_{C_n}(R; N|\mathscr{F}_C)$, overestimating the hazard of an event, and resulting in an early prediction of $E[R_N]$.

Conversely, the same negative bias of the Breslow and Nelson-Aalen estimators yields underestimated hazards in the parametric tail of the event distributions. The estimated terminal point of the nonparametric survival curve tends to be higher than the truth. The pseudo-ML estimate of the Weibull scale parameter is consequently positively biased, resulting in a negatively biased estimate of the tail of the hazard function. Thus the number of events at a future study time will be underestimated, and the prediction $\widehat{E[R_N]}$ too large. Underestimation of the hazard function also affects the tail of the dropout-time distribution. However, since monitoring events typically occur more frequently than dropouts in a randomized clinical trial, the behavior of the tail of the event-time distribution dominates the point prediction.

Finally, we also note that the degree of bias in the prediction of a landmark event time at a given information fraction is similar for a range of hazard ratios associated with the modeled covariate. This suggests that baseline covariates with large hazard ratios may not necessarily be more important to include in the working prediction model than those with smaller hazard ratios. Rather, the certainty of the association may be the most important consideration. The simulations presented here characterize the "best case" scenario, when the prediction model is properly specified for the underlying data. For a misspecified model, or when the working model for prediction poorly approximates the truth, the patterns of bias will be less predictable and could be much more substantial.

Sometimes, influential baseline covariates may be excluded from the event-time prediction model either accidentally, as in the case of the "unknown unknown," or purposefully, such as when treatment arms are pooled to protect the confidentiality of the trial results at

Table 4.9: Bias and sample standard error (SSE) of semiparametric point predictions (in years), when event times are distributed Weibull(shape: $\tau = 0.8$, scale: $\phi = 30$) with modeled covariate $Z \sim \text{Normal}(0, 1)$ with true hazard ratio $\exp \beta = 1.2$, in the presence of an unmodeled binary covariate. Dropout times are distributed Weibull($\tau_L = 1$, $\phi_L = 30$), approximately 3% dopout per year.

		Proportion (number) of sample population with an observed event							
Target time		0.05 (100)		0.10 (200)		0.20 (400)		0.30 (600)	
(prop. obs. events)	Ave. true time	Bias	SSE	Bias	SSE	Bias	SSE	Bias	SSE
Hazard ratio for unmodeled binary covariate: $e^\beta = 0.8$									
R_{400} (0.20)	1.79	-0.02	0.43	-0.03	0.21	—	—	—	—
R_{600} (0.30)	3.34	0.08	1.15	-0.02	0.52	-0.04	0.25	—	—
R_{800} (0.40)	5.52	0.29	2.53	0.07	1.22	-0.05	0.53	-0.05	0.36
R_{1000} (0.50)	8.83	0.42	4.56	0.45	2.94	-0.02	1.15	-0.07	0.74
Hazard ratio for unmodeled binary covariate: $e^\beta = 1.1$									
R_{400} (0.20)	1.46	-0.02	0.35	-0.03	0.17	—	—	—	—
R_{600} (0.30)	2.71	0.06	0.91	-0.03	0.42	-0.03	0.20	—	—
R_{800} (0.40)	4.43	0.23	1.95	0.04	0.94	-0.04	0.41	-0.04	0.28
R_{1000} (0.50)	6.95	0.42	3.45	0.27	2.06	-0.01	0.85	-0.05	0.55
Hazard ratio for unmodeled binary covariate: $e^\beta = 1.2$									
R_{400} (0.20)	1.38	-0.02	0.33	-0.03	0.16	—	—	—	—
R_{600} (0.30)	2.55	0.05	0.85	-0.03	0.39	-0.04	0.19	—	—
R_{800} (0.40)	4.17	0.22	1.84	0.03	0.86	-0.04	0.39	-0.04	0.26
R_{1000} (0.50)	6.51	0.35	3.12	0.25	1.92	-0.03	0.79	-0.06	0.50
Hazard ratio for unmodeled binary covariate: $e^\beta = 1.5$									
R_{400} (0.20)	1.18	-0.02	0.28	-0.03	0.14	—	—	—	—
R_{600} (0.30)	2.17	0.02	0.69	-0.04	0.33	-0.04	0.16	—	—
R_{800} (0.40)	3.54	0.12	1.46	-0.01	0.71	-0.06	0.32	-0.05	0.22
R_{1000} (0.50)	5.51	0.21	2.69	0.10	1.53	-0.09	0.63	-0.09	0.41

NOTE: Total sample size is $n = 2000$, and values are averages over 5000 simulated datasets. Bias: sample mean of the differences between predictions and their true values. SSE: sample standard error of the estimates.

an interim stage. In Table 4.9 we demonstrate the behavior of point predictions when one prognostic covariate is accounted for, and another is omitted. Here, data are sampled from a proportional hazards Weibull model with one normal covariate (hazard ratio $e^{\beta} = 0.12$ and one (uncorrelated) binary covariate with a range of possible associated hazard ratios. The omission of the binary variable has a negligible effect on the overall bias of event-time predictions. We further demonstrate the consistency of the point estimator in the case of an unmodeled covariate in Table 4.10. Predictions made at fixed proportions of observed events are shown for sample sizes of $n = 500$, 2000, and 5000. Both negative and positive bias decrease toward zero as the sample size increases.

Most clinical trials recruit and enroll patients in a staggered fashion over several weeks, months, or even years. Point predictions for data generated with varying speeds of staggered enrollment are shown in Table 4.11 (accurately specified event-time model) and Table 4.12 (event-time model with prognostic binary covariate excluded). Recall that our prediction method requires knowledge of baseline covariate values for all patients, and is therefore only implemented after the full sample is enrolled. The longer the accrual period, the greater the semiparametric and nonparametric cumulative hazard and survival functions factor into individual patients' probabilities of an observed event in the interval $[C, R]$. This accounts for the increasing negative bias in point predictions visible across slowing enrollment rates.

Table 4.10: Comparison of bias and sample standard error (SSE) of semiparametric point predictions (in years) for various sample sizes. Event times are distributed Weibull(shape: $\tau = 0.8$, scale: $\phi = 30$) with modeled covariate $Z \sim \text{Normal}(0, 1)$ with true hazard ratio $\exp(\beta) = 1.2$, in the presence of an unmodeled binary covariate with hazard ratio $\exp(\beta) = 1.2$. Dropout times are distributed Weibull($\tau_L = 1$, $\phi_L = 30$), approximately 3% dopout per year.

		Proportion of sample population with an observed event							
Target proportion		0.05		0.10		0.20		0.30	
(no. obs. events)	Ave. true time	Bias	SSE	Bias	SSE	Bias	SSE	Bias	SSE
Sample size $n = 500$		25 events		50 events		100 events		150 events	
0.20 (R_{100})	1.38	NA	NA	-0.10	0.33	—	—	—	—
0.30 (R_{150})	2.55	NA	NA	-0.06	0.86	-0.13	0.39	—	—
0.40 (R_{200})	4.16	NA	NA	0.07	1.84	-0.15	0.80	-0.16	0.53
0.50 (R_{250})	6.51	NA	NA	0.22	3.58	-0.06	1.75	-0.20	1.04
Sample size $n = 2000$		100 events		200 events		400 events		600 events	
R_{400} (0.20)	1.38	-0.02	0.33	-0.03	0.16	—	—	—	—
R_{600} (0.30)	2.55	0.05	0.85	-0.03	0.39	-0.04	0.19	—	—
R_{800} (0.40)	4.17	0.22	1.84	0.03	0.86	-0.04	0.39	-0.04	0.26
R_{1000} (0.50)	6.51	0.35	3.12	0.25	1.92	-0.03	0.79	-0.06	0.50
Sample size $n = 5000$		250 events		500 events		1000 events		1500 events	
R_{1000} (0.20)	1.38	-0.01	0.20	-0.01	0.10	—	—	—	—
R_{1500} (0.30)	2.55	0.02	0.50	-0.01	0.24	-0.01	0.12	—	—
R_{2000} (0.40)	4.17	0.16	1.20	0.00	0.50	-0.02	0.24	-0.02	0.16
R_{2500} (0.50)	6.50	0.42	2.33	0.07	1.05	-0.02	0.49	-0.04	0.31

NOTE: Values are averaged over 5000 data sets. Results are NA when the information cutoff occurrs before any dropout is observed. Bias: sample mean of the differences between predictions and their true values. SSE: sample standard error of the estimates.

Table 4.11: Bias and sample standard error (SSE) of semiparametric point predictions (in years), when event times are distributed Weibull(shape: $\tau = 0.8$, scale: $\phi = 30$) with modeled covariate $Z \sim \text{Normal}(0, 1)$ with true hazard ratio $\exp(\beta) = 1.2$, with varying lengths of staggered patient enrollment. Dropout times are distributed Weibull($\tau_L = 1$, $\phi_L = 30$), approximately 3% dopout per year.

		Proportion (number) of sample population with an observed event							
Target time		0.05 (100)		0.10 (200)		0.20 (400)		0.30 (600)	
(prop. obs. events)	Ave. true time	Bias	SSE	Bias	SSE	Bias	SSE	Bias	SSE
Simultaneous enrollment of full sample									
R_{400} (0.20)	1.56	-0.02	0.38	-0.04	0.18	—	—	—	—
R_{600} (0.30)	2.90	0.08	0.99	-0.03	0.45	-0.04	0.22	—	—
R_{800} (0.40)	4.75	0.03	2.21	0.04	1.02	-0.05	0.45	-0.05	0.31
R_{1000} (0.50)	7.47	0.55	3.95	0.34	2.42	-0.01	0.95	-0.06	0.60
Poisson enrollment rate $\alpha = 4000$ (approx. 0.5 yrs to accrue full sample)									
R_{400} (0.20)	1.82	-0.15	0.32	-0.14	0.20	—	—	—	—
R_{600} (0.30)	3.15	-0.19	0.83	-0.24	0.47	-0.17	0.26	—	—
R_{800} (0.40)	5.00	-0.11	1.88	-0.32	1.02	-0.27	0.52	-0.21	0.35
R_{1000} (0.50)	7.73	-0.04	3.66	-0.28	2.22	-0.38	1.02	-0.32	0.68
Poisson enrollment rate $\alpha = 2000$ (approx. 1.0 yrs to accrue full sample)									
R_{400} (0.20)	2.08	NA	NA	-0.12	0.18	—	—	—	—
R_{600} (0.30)	3.41	NA	NA	-0.23	0.45	-0.18	0.26	—	—
R_{800} (0.40)	5.26	NA	NA	-0.31	0.93	-0.29	0.53	-0.23	0.36
R_{1000} (0.50)	7.98	NA	NA	-0.24	2.16	-0.41	1.03	-0.36	0.71
Poisson enrollment rate $\alpha = 1000$ (approx. 2.0 yrs to accrue full sample)									
R_{200} (0.20)	2.61	NA	NA	NA	NA	—	—	—	—
R_{400} (0.30)	3.93	NA	NA	NA	NA	-0.14	0.24	—	—
R_{600} (0.40)	5.78	NA	NA	NA	NA	-0.31	0.53	-0.22	0.34
R_{1000} (0.50)	8.50	NA	NA	NA	NA	-0.44	1.05	-0.39	0.72

NOTE: Sample size is n=2000, and results are averaged over 5000 replications. Values are NA where information cutoff occurrs on average before the full sample is enrolled. Bias: sample mean of the differences between predictions and their true values. SSE: sample standard error of the estimates.

Table 4.12: Bias and sample standard error (SSE) of semiparametric point predictions (in years), when event times are distributed Weibull(shape: $\tau = 0.8$, scale: $\phi = 30$) with modeled covariate $Z \sim \text{Normal}(0, 1)$ with true hazard ratio $\exp(\beta) = 1.2$ and an unmodeled binary covariate with true hazard ratio $\exp(\beta) = 1.2$. Patient accrual is staggered over various length enrollment periods. Dropout times are distributed Weibull($\tau_L = 1$, $\phi_L = 30$), approximately 3% dopout per year.

		Proportion (number) of sample population with an observed event							
Target time		0.05 (100)		0.10 (200)		0.20 (400)		0.30 (600)	
(prop. obs. events)	Ave. true time	Bias	SSE	Bias	SSE	Bias	SSE	Bias	SSE
Simultaneous enrollment of full sample									
R_{400} (0.20)	1.38	-0.02	0.33	-0.03	0.16	—	—	—	—
R_{600} (0.30)	2.55	0.05	0.85	-0.03	0.39	-0.04	0.19	—	—
R_{800} (0.40)	4.17	0.22	1.84	0.03	0.86	-0.04	0.39	-0.04	0.26
R_{1000} (0.50)	6.51	0.35	3.12	0.25	1.92	-0.03	0.79	-0.06	0.50
Poisson enrollment rate $\alpha = 4000$ (approx. 0.5 yrs to accrue full sample)									
R_{400} (0.20)	1.63	-0.11	0.28	-0.12	0.18	—	—	—	—
R_{600} (0.30)	2.80	-0.13	0.70	-0.21	0.41	-0.15	0.23	—	—
R_{800} (0.40)	4.42	-0.04	1.69	-0.28	0.85	-0.24	0.46	-0.18	0.30
R_{1000} (0.50)	6.75	0.01	2.86	-0.26	1.82	-0.35	0.88	-0.28	0.57
Poisson enrollment rate $\alpha = 2000$ (approx. 1.0 yrs to accrue full sample)									
R_{400} (0.20)	1.89	NA	NA	-0.10	0.15	—	—	—	—
R_{600} (0.30)	3.06	NA	NA	-0.21	0.38	-0.15	0.22	—	—
R_{800} (0.40)	4.67	NA	NA	-0.29	0.77	-0.26	0.45	-0.20	0.31
R_{1000} (0.50)	7.01	NA	NA	-0.30	1.63	-0.37	0.87	-0.32	0.59
Poisson enrollment rate $\alpha = 1000$ (approx. 2.0 yrs to accrue full sample)									
R_{400} (0.20)	2.43	NA	NA	NA	NA	—	—	—	—
R_{600} (0.30)	3.59	NA	NA	NA	NA	-0.11	0.20	—	—
R_{800} (0.40)	5.20	NA	NA	NA	NA	-0.27	0.44	-0.18	0.29
R_{1000} (0.50)	7.53	NA	NA	NA	NA	-0.40	0.85	-0.35	0.61

NOTE: Sample size is n=2000, and results are averaged over 5000 replications. Values are NA where information cutoff occurrs on average before the full sample is enrolled. Bias: sample mean of the differences between predictions and their true values. SSE: sample standard error of the estimates.

4.4 Summary

This semiparametric method is a promising extension of the previously described parametric framework for prediction of landmark event-times in randomized clinical trials. There are currently no event-time prediction methods in the statistical literature that can accommodate baseline covariate information, beyond stratifying the prediction model by a categorical covariate. This method provides a first step, which may be modified and further developed to address monitoring issues in clinical trials. We demonstrated through theoretical justification and simulation studies that, when modeling assumptions are met, our point predictions and prediction intervals are valid. The method can clearly provide useful guidance to investigators and the oversight teams in a clinical trial.

The semiparametric prediction model relies on key assumptions that cannot be ignored. Any baseline covariates included in the model should satisfy the proportional hazards assumption across the entire range of possible observation times for the study. It may be impossible to definitively verify this assumption outside of the range of observed data at an interim cutoff time. Nevertheless it should be checked as thoroughly as possible. For the projected tail probabilities of event and dropout to be accurate, we must assume the tails of the hazard functions are approximately Weibull-distributed, and therefore monotonic. In fact, in the examples provided above, we estimated parameters of the Weibull tail distributions using all of the available data collected before the cutoff time. This implicitly assumes that the entire event distribution is proportional hazards Weibull and the entire dropout distribution is Weibull. When an adequate amount of data is available, this stringent assumption could be relaxed by estimating the Weibull parameters from a subset of the available data, for example data collected in a shorter interval of time immediately preceding the data cutoff.

It is also important to emphasize the distinction between the true event-time distribution and the "working model" event distribution approximation ultimately implemented in predictions. The consistency and asymptotic normality of our point predictor hold when estimates of the true event distribution are used in the model, i.e. the event-time distribution is properly specified. In most cases it is impossible for investigators to know the true

event distribution, including all the relevant baseline covariates. If the working model is too severely misspecified, the semiparametric prediction method will not be able to compensate, and predictions will be inaccurate. Thus, investigators charged with the task of selecting the prediction model face two primary challenges of prediction: first, observed data may not be representative of future data, and second, the set of predetermined measured baseline covariates may exclude a variable unknown to be critical to the outcome.

We note, however, that the semiparametric Cox model and the parametric Weibull have been and continue to be employed extensively in survival data analysis not only for their convenience, but also because they are versatile and valid for a wide range of scientific questions. Even if assumptions of these models do not hold over long periods of time, they are often adequate approximations over the duration of study for a clinical trial.

Chapter 5

DISCUSSION

5.1 Summary

The question of how best to predict event-occurrence times in randomized clinical trials is fundamental to the successful implementation of the study from statistical, logistic, and economic perspectives. Challenges faced in scheduling the optimally efficient and ethical times for interim monitoring analyses in the HIV Prevention Trials Network 052 Study motivated our research on this topic. We proposed two methods for using accumulating current data to predict the time of the Nth event in an ongoing trial. The first is a fully parametric, likelihood-based approach, and the second is a more flexible semiparametric approach that allows for the inclusion of prognostic baseline covariate information in predictions. We demonstrated the performance of these methods in simulation studies and through application to real clinical trial data. Our frequentist parametric approach was also compared to an existing Bayesian prediction procedure, yielding generally comparable, and sometimes more efficient, results.

We believe these methods are worthwhile contributions to the statistical literature. The proposed parametric method is simple to understand and implement, and it yields useful and accurate point and interval predictions when the accrual, event, and dropout time distributions are carefully selected. By contrast, previously published methods, either parametric or nonparametric, are more complicated, computationally intensive, and require more preliminary assumptions, which may restrict their wider accessibility and implementation in clinical trials. Furthermore, none of the existing methods in the statistical literature can accommodate patient-specific, prognostic baseline covariate information beyond categorical stratification of the model. Our semiparametric method allows for the inclusion of one or more categorical or continuous covariates, giving it the potential to produce more accurate and efficient predictions.

The basic framework for event-occurrence time point and interval predictions provided by our proposed methods may easily be extended or modified to address additional scenarios that arise in clinical trials due to specifics in the design or implementation. Some possible modifications are discussed in section 5.2. There are several limitations to the prediction methods proposed here, as discussed in earlier chapters. Further research and refinement of the models and estimation procedures will help us to more clearly characterize the limitations and potentially eliminate some of them.

5.2 Future Work

5.2.1 Potential extensions of the parametric event-occurrence time method

Inclusion of historical data or pilot data

Early-phase and pilot-phase studies usually precede the type of large, confirmatory clinical trial for which prediction methods like those described in this dissertation would be implemented. In certain cases, there may be pilot data or historical data that could be included in event-occurrence time predictions to increase the efficiency of predictions. For example, in our analyses of the early HPTN 052 data, we excluded 82 couples that were part of a pilot phase that began two years before the full study. Pilot studies are often conducted, "...in order to evaluate the feasibility of recruitment, randomization, retention, assessment procedures, new methods, and/or implementation of the novel intervention."[37]. If the sample population and the enrollment and treatment procedures of preliminary studies are very similar to those used in the full study, it may be safe to assume the data are representative of the data in the ongoing trial. Then either the pilot enrollment data, the time-to-event data, or both, may be included in the ML estimation of the parameters for the parametric prediction model. The results may be used as a sensitivity analysis for the standard event-time predictions. It would be worthwhile to investigate the inclusion of pilot data in the HPTN 052 data.

Prediction using a fully parametric proportional hazards Weibull model

As presented, the parametric prediction method uses no baseline covariate information beyond the possible stratification by a categorical covariate, such as treatment arm. The proportional hazards Weibull model has been suggested as a reasonable alternative to the Cox model for covariate-adjusted analyses due to its flexibility and sensible application in many situations. In fact, when the baseline hazards are truly well approximated by a Weibull distribution, the estimated log hazard ratios are asymptotically unbiased and equal in efficiency to the partial likelihood estimates of the Cox model [8].

Because the log hazard ratios, in addition to the shape and scale parameters, may be estimated using standard maximum likelihood, our parametric prediction method could easily be extended to the Weibull model. When a Weibull working model for the event- and dropout-time distributions is justified through model diagnostics, a fully parametric approach may be more efficient than the semiparametric method. A fully parametric covariate-adjusted prediction model would also be less computationally intensive than the semiparametric model because prediction intervals could be calculated directly, rather than constructed through bootstrap resampling. Moreover the same Weibull model could be used over the range of observed observation times as well as for future observation times. Nevertheless, a covariate-adjusted Weibull approach would be subject to some of the same constraints as the semiparametric approach, in that the proportional hazards assumption must be valid for all modeled covariates throughout the range of observation times, and the reliability of the predictions would be heavily dependent on the appropriate selection of a working event-time model.

Publicly available code for the parametric method

General versions of the parametric prediction methods for a variety of distributional assumptions should be made available online to promote the further use and investigation of the proposed approaches.

5.2.2 Next steps and potential extensions of the semiparametric event-occurrence time method

Investigation of the effects of model misspecification

In sections 4.3.1 and 4.3.2, respectively, we demonstrated the behavior of the pseudo-ML estimates of Weibull shape and scale and the semiparametric point prediction under mild model misspecifition, i.e. unmodeled covariate effects. Since model misspecification is difficult or impossible to avoid in a prediction setting, it is important to investigate the degree to which our approach can compensate and yield usable predictions. The method should be applied in scenarios where the working model for an event deviates from the true distribution. In particular, predictions should be assessed when data come from distributions having a variety of (non-Weibull) baseline hazard functions, as well as in situations where the proportional hazards assumption does not hold.

Alternative estimation of tail probabilities

We believe the pseudo-ML Weibull estimation and projection of tail probabilities is a reasonable approach that will produce useful predictions in many situations when the proportional hazards assumption is valid for the modeled covariates. However, as illustrated in section 4.3.2, the pseudo-ML scale parameter estimate will always be influenced by the finite-sample bias of the Breslow and Nelson-Aalen estimates of cumulative hazards, resulting in positively biased long-term predictions of the expected time of the Nth event. It would be worthwhile to brainstorm and investigate the performance of alternative approaches to estimating the tail distributions of event and dropout.

Even within the framework of Weibull-extended survival curves, a number of alternative estimation approaches are possible. For example, the Weibull tails need not be estimated from the full set of accumulated trial data. Rather, it may be prudent to estimate Weibull parameters from a truncated subset of the data, excluding the earliest interval of observation, as later data may be more pertinent to the behavior of the tails of the survival curves. Alternatively, a weighting scheme could be applied to the full observed data set to emphasize the shape of the event- and dropout-time distributions closer to the tail.

Investigation of bootstrap prediction interval properties and alternative bootstrap resampling methods

We plan to conduct simulation studies to investigate the characteristics and performance of the bootstrap prediction intervals in the semiparametric prediction method. We will describe the width and coverage properties of prediction intervals for the Nth observed event time in scenarios consistent with the parameter estimate and point prediction simulations shown in Chapter 4.

The prediction interval in section 4.2.8 is obtained via a standard bootstrap resampling algorithm, as described in Efron (1981) [20]. Further study of the semiparametric prediction method may suggest that an alternative resampling method may be more appropriate. For example, Davison and Hinkley discuss a number of alternative bootstrap resampling algorithms for the context of right-censored survival data, both for a no-covariate model and a proportional hazards model [15]. These resampling schemes are intended to deal more carefully with two issues that arise when applying the bootstrap to survival data. First, the observed data are the result of a combination of two distinct distributions, the conditional event-time distribution $F_{E^*|Z^*}$ and the dropout-time distribution F_L. Even when censoring is assumed to be noninformative, it may be prudent to consider the two distinct distributions in the resampling algorithm. Second, when sampling with replacement from the observed data, we inevitably end up with instances of tied survival times.

In the case of no covariates, alternatives to the standard resampling of cases (the observation time and event-indicator pairs) suggested by Davison and Hinkley include a "conditional bootstrap" that is conditional on the observed pattern of censorship, or a "weird bootstrap" that treats the numbers of failures at each observed failure time as independent binomial variables, while conditioning on the number of subjects at risk at each time. For patient data that includes baseline covariates, two algorithms are presented. The first, "conditional resampling for censored survival data" may be used when the censoring distribution is (assumed to be) independent of the covariates, and the second, "resampling for censored survival data," assumes proportional hazards for the modeled covariates also applies to the censoring distribution.

Application of the semiparametric prediction method to real clinical trial data

The early HPTN 052 Study data was well suited to demonstrate the parametric prediction method; however it does not include baseline covariates strongly predictive of the composite mortality and morbidity monitoring event. Also, the range of observed data ends shortly after the last patient is enrolled, so it is impossible to compare the accuracy of post-enrollment predictions to actual observed event times, unless we discard a large portion of the data and restrict analyses to a subset of patients enrolled early in the study. We would like to find one or more alternative real clinical trial data sets with which to demonstrate the semiparametric approach.

Application of the semiparametric method with incomplete enrollment

The development of the semiparametric prediction approach assumes complete enrollment and baseline covariate information for all participants. It may be useful to consider ways to implement the semiparametric method while enrollment is still underway. To do this, we would need to determine appropriate ways to model continuing patient accrual, based on the observed data. Also, reasonable methods to impute baseline covariate values for the set of patients not-yet-enrolled would need to be identified.

Publicly available code for the parametric method

The semiparametric prediction procedure, in particular the calculation of patient-specific probability of an event, is not straightforward to implement in R. A general version should be refined and made available online to promote the further use and investigation of the method.

BIBLIOGRAPHY

[1] Jared M Baeten, Erin Kahle, Jairam R Lingappa, Robert W Coombs, Sinead Delany-Moretlwe, Edith Nakku-Joloba, Nelly R Mugo, Anna Wald, Lawrence Corey, Deborah Donnell, et al. Genital hiv-1 rna predicts risk of heterosexual hiv-1 transmission. Science translational medicine, 3(77):77ra29–77ra29, 2011.

[2] Vilijandas Bagdonavicus, Julius Kruopis, and Mikhail Nikulin. Nonparametric tests for censored data. John Wiley & Sons, 2013.

[3] Emilia Bagiella and Daniel F Heitjan. Predicting analysis times in randomized clinical trials. Statistics in medicine, 20(14):2055–2063, 2001.

[4] Brigitte Baldi and David S Moore. The practice of statistics in the life sciences. WH Freeman, 2014.

[5] Jacques Benichou and Mitchell H Gail. A delta method for implicitly defined random variables. The American Statistician, 43(1):41–44, 1989.

[6] Norman Breslow, John Crowley, et al. A large sample study of the life table and product limit estimates under random censorship. The Annals of Statistics, 2(3):437–453, 1974.

[7] Norman E Breslow. Contribution to the discussion of the paper by dr cox. Journal of the Royal Statistical Society, Series B, 34(2):216–217, 1972.

[8] Kevin J Carroll. On the use and utility of the weibull model in the analysis of survival data. Controlled clinical trials, 24(6):682–701, 2003.

[9] George Casella and Roger L Berger. Statistical inference, volume 2. Duxbury Pacific Grove, CA, 2002.

[10] Ying Qing Chen, Benoit Masse, Lei Wang, San-San Ou, Xin Li, Deborah Donnell, Marybeth McCauley, Theresa Gamble, Heather J Ribauldo, Myron S Cohen, et al. Statistical considerations for the hptn 052 study to evaluate the effectiveness of early versus delayed antiretroviral strategies to prevent the sexual transmission of hiv-1 in serodiscordant couples. Contemporary clinical trials, 33(6):1280–1286, 2012.

[11] Myron S Cohen, Ying Q Chen, Marybeth McCauley, Theresa Gamble, Mina C Hosseinipour, Nagalingeswaran Kumarasamy, James G Hakim, Johnstone Kumwenda, Beatriz Grinsztejn, Jose HS Pilotto, et al. Prevention of hiv-1 infection with early antiretroviral therapy. New England journal of medicine, 365(6):493–505, 2011.

[12] Myron S Cohen and Cynthia L Gay. Treatment to prevent transmission of hiv-1. Clinical Infectious Diseases, 50(Supplement 3):S85–S95, 2010.

[13] Myron S Cohen, Marybeth McCauley, and Theresa R Gamble. Hiv treatment as prevention and hptn 052. Current Opinion in HIV and AIDS, 7(2):99, 2012.

[14] Cox R David. Regression models and life tables (with discussion). Journal of the Royal Statistical Society, 34:187–220, 1972.

[15] Anthony C Davison and David V Hinkley. Bootstrap methods and their applications, cambridge series in statistical and probabilistic mathematics, 1997.

[16] Walter L Deemer Jr and David F Votaw Jr. Estimation of parameters of truncated or censored exponential distributions. The Annals of Mathematical Statistics, pages 498–504, 1955.

[17] David L Demets and KK Lan. Interim analysis: the alpha spending function approach. Statistics in medicine, 13(13-14):1341–1352, 1994.

[18] J Mark Donovan, Michael R Elliott, and Daniel F Heitjan. Predicting event times in clinical trials when randomization is masked and blocked. Clinical Trials, 4(5):481–490, 2007.

[19] Bradley Efron. Bootstrap methods: another look at the jackknife. The annals of Statistics, pages 1–26, 1979.

[20] Bradley Efron. Censored data and the bootstrap. Journal of the American Statistical Association, 76(374):312–319, 1981.

[21] Susan S Ellenberg, Thomas R Fleming, and David L DeMets. Data monitoring committees in clinical trials: a practical perspective. John Wiley & Sons, 2003.

[22] Liang Fang and Zheng Su. A hybrid approach to predicting events in clinical trials with time-to-event outcomes. Contemporary clinical trials, 32(5):755–759, 2011.

[23] Thomas Shelburne Ferguson. A course in large sample theory, volume 49. Chapman & Hall London, 1996.

[24] Thomas R Fleming and David L DeMets. Monitoring of clinical trials: issues and recommendations. Controlled Clinical Trials, 14(3):183–197, 1993.

[25] Thomas R Fleming and David P Harrington. Counting processes and survival analysis, volume 169. John Wiley & Sons, 2011.

[26] Lawrence M Friedman, Curt Furberg, and David L DeMets. Fundamentals of clinical trials, volume 4. Springer, 2010.

[27] Seymour Geisser. Predictive inference, volume 55. CRC Press, 1993.

[28] Susan M Graham, Sarah E Holte, Norbert M Peshu, Barbra A Richardson, Dana D Panteleeff, Walter G Jaoko, Jeckoniah O Ndinya-Achola, Kishorchandra N Mandaliya, Julie M Overbaugh, and R Scott McClelland. Initiation of antiretroviral therapy leads to a rapid decline in cervical and vaginal hiv-1 shedding. Aids, 21(4):501–507, 2007.

[29] International Chronic Granulomatous Disease Cooperative Study Group et al. A. phase iii study establishing efficacy of recombinant human interferon-gamma for infection prophylaxis in chronic granulomatous disease. N Engl J Med, 324:509–16, 1991.

[30] John D Kalbfleisch and Ross L Prentice. The statistical analysis of failure time data, volume 360. John Wiley & Sons, 2011.

[31] Edward L Kaplan and Paul Meier. Nonparametric estimation from incomplete observations. Journal of the American statistical association, 53(282):457–481, 1958.

[32] John M Lachin and Mary A Foulkes. Evaluation of sample size and power for analyses of survival with allowance for nonuniform patient entry, losses to follow-up, noncompliance, and stratification. Biometrics, pages 507–519, 1986.

[33] Nan M Laird and Thomas A Louis. Approximate posterior distributions for incomplete data problems. Journal of the Royal Statistical Society. Series B (Methodological), pages 190–200, 1982.

[34] Jerald F Lawless. Statistical models and methods for lifetime data, volume 362. John Wiley & Sons, 2011.

[35] Elisa T Lee and John Wenyu Wang. Statistical methods for survival data analysis. John Wiley & Sons, 2013.

[36] Erich Leo Lehmann and George Casella. Theory of point estimation, volume 31. Springer, 1998.

[37] Andrew C Leon, Lori L Davis, and Helena C Kraemer. The role and interpretation of pilot studies in clinical research. Journal of psychiatric research, 45(5):626–629, 2011.

[38] J Mark Donovan, Michael R Elliott, and Daniel F Heitjan. Predicting event times in clinical trials when treatment arm is masked. Journal of biopharmaceutical statistics, 16(3):343–356, 2006.

[39] Wayne Nelson. Hazard plotting for incomplete failure data. Journal of Quality Technology, 1(1), 1969.

[40] Wayne Nelson. Theory and applications of hazard plotting for censored failure data. Technometrics, 14(4):945–966, 1972.

[41] Peter C O'Brien and Thomas R Fleming. A multiple testing procedure for clinical trials. Biometrics, pages 549–556, 1979.

[42] MI Parzen, LJ Wei, and Z Ying. A resampling method based on pivotal estimating functions. Biometrika, 81(2):341–350, 1994.

[43] Thomas C Quinn, Maria J Wawer, Nelson Sewankambo, David Serwadda, Chuanjun Li, Fred Wabwire-Mangen, Mary O Meehan, Thomas Lutalo, and Ronald H Gray. Viral load and heterosexual transmission of human immunodeficiency virus type 1. New England journal of medicine, 342(13):921–929, 2000.

[44] Horst Rinne. The Weibull distribution: a handbook. CRC Press, 2010.

[45] Lawrence V Rubinstein, Mitchell H Gail, and Thomas J Santner. Planning the duration of a comparative clinical trial with loss to follow-up and a period of continued observation. Journal of chronic diseases, 34(9):469–479, 1981.

[46] Robert J Serfling. Approximation theorems of mathematical statistics, volume 162. John Wiley & Sons, 2009.

[47] Terry M Therneau. Modeling survival data: extending the Cox model. Springer, 2000.

[48] Aad W Van der Vaart. Asymptotic statistics, volume 3. Cambridge university press, 2000.

[49] Pietro L Vernazza, Luigi Troiani, Markus J Flepp, Richard W Cone, Jody Schock, Felix Roth, Katia Boggian, Myron S Cohen, Susan A Fiscus, Joseph J Eron, et al. Potent antiretroviral treatment of hiv-infection results in suppression of the seminal shedding of hiv. Aids, 14(2):117–121, 2000.

[50] Jianming Wang, Chunlei Ke, Qi Jiang, Charlie Zhang, and Steven Snapinn. Predicting analysis time in event-driven clinical trials with event-reporting lag. Statistics in medicine, 31(9):801–811, 2012.

[51] Gui-shuang Ying and Daniel F Heitjan. Weibull prediction of event times in clinical trials. Pharmaceutical statistics, 7(2):107–120, 2008.

[52] Gui-Shuang Ying and Daniel F Heitjan. Prediction of event times in the rematch trial. Clinical Trials, 10(2):197–206, 2013.

[53] Gui-shuang Ying, Daniel F Heitjan, and Tai-Tsang Chen. Nonparametric prediction of event times in randomized clinical trials. Clinical Trials, 1(4):352–361, 2004.

[54] Xiaoxi Zhang and Qi Long. Joint monitoring and prediction of accrual and event times in clinical trials. Biometrical Journal, 54(6):735–749, 2012.

[55] Xiaoxi Zhang and Qi Long. Modeling and prediction of subject accrual and event times in clinical trials: a systematic review. Clinical Trials, 9(6):681–688, 2012.