Stochastic Curtailment Methods for Single-Arm Clinical Trials with a Time-to-Event Endpoint using Weibull Distribution

Muhammad Waleed

M.A. Mathematics, University of Maine, 2016

B.Sc. (Hons.) Mathematics, Lahore University of Management Sciences, 2013

Submitted to the graduate degree program in Biostatistics and the Graduate Faculty of the University of Kansas in partial fulfillment of the requirements for the degree of Doctor of Philosophy.

Co-Chair: Milind A. Phadnis, Ph.D.

Co-Chair: Jianghua He, Ph.D.

Byron J. Gajewski, Ph.D.

Matthew S. Mayo, Ph.D., MBA

Scott Weir, Pharm.D., Ph.D.

Date defended: April 20, 2021

The Dissertation Committee for Muhammad Waleed certifies
that this is the approved version of the following dissertation:

Stochastic Curtailment Methods for Single-Arm Clinical Trials with a Time-to-Event Endpoint using Weibull Distribution

Co-Chair: Milind A. Phadnis, Ph.D.

Co-Chair: Jianghua He, Ph.D.

Graduate Director: Jo A. Wick, Ph.D.

Date approved: April 26, 2021

Abstract

This dissertation is an outcome of three research projects which attempt to fill some existing gaps in the statistical literature related to the design and analysis of single-arm clinical trials with time-to-event endpoints following a Weibull distribution.

In the first project, we proposed a parametric maximum likelihood estimate based method for designing single-arm clinical trials with a time-to-event endpoint that follows a Weibull distribution with known shape parameter. The proposed method is quite flexible in the sense that it permits investigators to incorporate various design features, such as expected loss to follow-up rate, different accrual patterns, and administrative censoring. In the same context, three stochastic curtailment methods (conditional power, predictive power, Bayesian predictive probability) are presented which can be employed to obtain early evidence of efficacy or futility of an experimental treatment. Finally, we have also discussed the implementation of group sequential designs using the repeated significance approach.

The second project primarily focuses on the calculation of the Bayesian predictive probability when a reasonably accurate estimate of the shape parameter of the Weibull distribution for the underlying survival times is not available from historical studies. To suffice our purpose, two approaches based on the posterior mode and the entire posterior distribution of the shape parameter are presented. In addition to calculating the Bayesian predictive probability, we also explored the utility of the internal pilot study approach for reestimating the study sample size based on data accumulated at an interim stage.

In the third project, an R package is developed for designing single-arm clinical trials with a time-to-event endpoint following the Weibull distribution, and to implement stochastic curtailment methods discussed in the first two projects. The package will be made available to the scientific community on the Comprehensive R Archive Network (CRAN).

Acknowledgements

First and foremost, I would like to express deepest gratitude to my dissertation co-chairs, Dr. Jianghua He and Dr. Milind Phadnis, for their dedicated direction throughout the process of completing this dissertation. I am especially thankful to them for providing thoroughly constructive feedback on my progress, and for their patience and understanding whenever I stumbled across any obstacles in my work. They have always been incredibly gracious with their time taking keen interest in my academic and professional growth, and went above and beyond to encourage and motivate me to overcome my doubts and fears. I could not have wished for better mentors over the course of my graduate studies at KUMC.

I am profoundly grateful to Dr. Byron Gajewski, Dr. Matthew Mayo, and Dr. Scott Weir for agreeing to serve on my dissertation committee despite their demanding commitments, and for providing insightful comments and suggestions which greatly enhanced the presentation of my research. I would like to thank Dr. Gajewski for his unhesitating willingness to answer my numerous questions during the initial stages of my work in the Bayesian paradigm. I am also indebted to Dr. Mayo for admitting me into the doctoral program at KUMC, and for sponsoring my attendance in various short courses and conferences over the past few years. I extend him my absolute appreciation for his continued support and encouragement throughout this endeavor.

I am thankful to the entire graduate faculty of the Department of Biostatistics and Data Science for their exceptional instruction in all graduate-level courses. Under their tutelage, I have learnt a great deal about various domains within the discipline of biostatistics. I sincerely thank our graduate director Dr. Jo Wick for her unwavering support and guidance in addressing any issues I have had, and for ensuring a smooth and fruitful educational experience. I would also like to take this opportunity to thank Ms. Rosemary Morrow, Ms. Mandy Rametta, and Ms. Lori McElgunn for their tireless assistance in resolving administrative issues through my years at KUMC.

I would also like to acknowledge Dr. Devin Koestler and Dr. Dong Pei for providing and facilitating access to the high-performance computing (HPC) platform which saved an exorbitant amount of time to complete all computationally intensive simulations presented in Chapter 3 of this dissertation. These HPC facilities are supported by the K-INBRE Bioinformatics Core that is supported in part by the National Institute of General Medical Science award (P20 GM103418), the Biostatistics and Informatics Shared Resource, supported by the National Cancer Institute Cancer Center Support Grant (P30 CA168524), and the Kansas Institute for Precision Medicine COBRE, supported by the National Institute of General Medical Science award (P20 GM130423).

I also wish to express gratitude to my previous mentors and friends I have come to know in this journey. To my brilliant mentors during the summer research fellowship at the U.S. FDA, Dr. Cesar Torres and Dr. Paul Schuette, thank you for providing me this rewarding opportunity, both intellectually and professionally, which also helped me address some aspects of my dissertation. To my inspiring mentors during the masters degree program at UMaine, Dr. Ramesh Gupta and Dr. Pushpa Gupta, thank you for your sincere support and guidance during my initial years in the US. To my amazing mentor at LUMS, Dr. Mudassar Imran, thank you for noticing potential in me even after academically tumultuous years during my undergraduate studies, and for encouraging me to pursue graduate studies. To Syed Aamir Ali Shah and Guanlin Zhang, thank you for your friendship, and unabated support in my personal and professional matters alike. To the Khan family in Maine, thank you for your sheer kindness and generosity, and for always making me feel at home.

Last but not the least, I wish to express my heartfelt thanks to my family: my mother (ammi), Shagufta, and my siblings, Amir, Munazza, and Junaid, for their continuous encouragement, prayers, love and support which kept me grounded during this long and arduous, yet rewarding endeavor. In particular, no words can do justice to express how appreciative I am to my brother, Junaid, for being my greatest support system through thick and thin, and for sacrificing various career opportunities and choosing to stay with ammi so that I could pursue my ambitions with complete peace of mind. May Allah SWT always bless you all, Ameen!

Dedication

To my beloved mother, Shagufta Perveen, for her countless sacrifices, and
for investing every ounce of her blood, sweat, and tear in pursuit of our better future.
Without her unconditional love, sincere prayers, and constant support and encouragement,
I would not have made this far.

Contents

List of Figures

List of Tables

Chapter 1

Introduction

1.1 Background

Clinical trials are research studies involving human subjects to find novel ways of diagnosis, prevention or treatment of a condition, and to manage symptoms and side effects from an existing or a new treatment [1, 2]. There are four phases of clinical trials [2]. Phase I trials are small sample studies, with about 15–30 subjects, to determine the maximum tolerated dose (MTD) or the recommended phase II dose (RP2D) of an experimental treatment which is further evaluated in a phase II trial with small to moderate sample sizes, usually less than 100, to obtain preliminary evidence of its therapeutic effect and safety profile [2, 3]. In the subsequent large sample phase III trials, an experimental treatment is compared with a standard treatment to obtain evidence of efficacy and to monitor any adverse events, and finally, phave IV trials are post-marketing large sample studies conducted to find evidence of safety and efficacy of an approved treatment [2].

The primary endpoints in different phases of oncology clinical trials include, but are not limited to, objective response rate (ORR) or tumor response rate (TRR), progression-free survival (PFS), disease-free survival (DFS), and overall survival (OS) [4]. Although ORR or TRR have popularly been used as an endpoint in the early phase single-arm oncology trials, Rubinstein noted that the time-to-event (TTE) endpoints PFS and OS are also being considered in the recent times due to some medical and practical considerations [5]. Since the entire drug development process is extremely expensive, both in terms of time as well as resources, it is critical to design early phase single-arm trials with sound statistical properties that enable researchers to detect ineffective treatments as soon as possible. This dissertation focuses on the design and other planning aspects for single-arm phase II clinical trials with a TTE primary endpoint that permit researchers to evaluate

the potential of an experimental treatment at an interim stage.

1.2 Aims and Scope

In this dissertation, we aim to fill gaps in the statistical literature related to various design aspects of single-arm phase-II clinical trials when the Weibull distribution is appropriate for modeling survival data derived from these studies. Our specific goals and objectives listed as below:

1. When a reasonably accurate estimate of the shape parameter of the Weibull distribution is available from historical studies, we propose a parametric maximum likelihood estimate (MLE) based method for designing single-arm clinical trials while incorporating flexible features. In the same context, we describe three stochastic curtailment methods (conditional power, predictive power, Bayesian predictive probability) for efficacy or futility testing based on data accumulated at an interim stage. Furthermore, we discuss the implementation of repeated significance testing approach for evaluating experimental treatments over the course of an ongoing clinical trial. A detailed introduction to the concepts of stochastic curtailment and repeated significance testing shall be presented in the later chapters.

2. We explore the effect of misspecification of the shape parameter of the Weibull distribution on the desired operating characteristics (Type-I error and power). When no reliable estimates of the nuisance shape parameter is available, we also discuss adaptation to the study sample size based on data collected at an interim stage, and present two approaches for calculating the Bayesian predictive probability for efficacy or futility testing.

3. We develop a computer package implementing the methods corresponding to the above-mentioned objectives that will be made freely available to the scientific community.

1.3 Organization of Research

We followed a three manuscript format for this dissertation and it is organized in the following order. The first manuscript is presented in Chapter 2 and it deals with the objectives listed in the

bullet point 1 in Section 1.2. The second manuscript, presented in Chapter 3, addresses the bullet point 2 in Section 1.2. In Chapter 4, the third manuscript introducing the usage and functionalities of our computer package is presented. Finally, we present some concluding remarks and future research direction in Chapter 5.

Chapter 2

Some Design Considerations Incorporating Early Futility for Single-Arm Clinical Trials with Time-to-Event Primary Endpoints using Weibull Distribution

This chapter has previously been published and is reprinted here with permission with minor modifications. Waleed M, He J, Phadnis MA. Some design considerations incorporating early futility for single-arm clinical trials with time-to-event primary endpoints using Weibull distribution. *Pharm Stat.* 2021; 20(3): 610–644. https://doi.org/10.1002/pst.2097

Abstract

Sample size calculation is an essential component of the planning phase of a clinical trial. In the context of single-arm clinical trials with time-to-event (TTE) endpoints, only a few options with limited design features are available. Motivated from ethical or practical considerations, two-stage designs are implemented for single-arm studies to obtain early evidence of futility. A major drawback of such designs is that early stopping may only occur at the conclusion of the first stage, even if lack of efficacy becomes apparent at any other time point over the course of the clinical trial. In this manuscript, we attempt to fill some existing gaps in the literature related to single-arm clinical trials with TTE endpoints. We propose a parametric maximum likelihood estimate-based test whose variance component accounts for the expected proportion of loss to follow-up and different accrual patterns (early, late, or uniform accrual). For the proposed method, we present three stochastic curtailment methods (conditional power, predictive power, Bayesian predictive probability) which can be employed for efficacy or futility testing purposes. Finally, we discuss the implementation of

group sequential designs for obtaining an early evidence of efficacy or futility at pre-planned timings of interim analyses. Through extensive simulations, it is shown that our proposed method performs well for designing these studies with moderate to large sample sizes. Some examples are presented to demonstrate various aspects of the stochastic curtailment and repeated significance testing methods presented in this manuscript.

2.1 Introduction

Single-arm trials are generally implemented to evaluate novel treatments during the early phases of clinical research when it is unethical to conduct placebo-controlled trials, or investigators are unable to implement randomized controlled trials due to practical considerations [3]. For instance, single-arm trials are often conducted when the target population is small in size (as in the case of rare diseases), or "window-of-opportunity" trials where subjects with a particular condition are available for a limited period of time prior to receiving any available standard treatment [6]. After determining the maximum tolerated dose (MTD) or the recommended phase II dose (RP2D) of an experimental treatment in a phase I trial, researchers conduct phase II studies with small to moderate sample sizes to evaluate whether it has sufficient efficacy to warrant further investigation in a subsequent phase III trial with a larger sample size [3]. In the absence of a widely used standard treatment, single-arm phase II studies are often conducted in oncology to obtain preliminary evidence of therapeutic effect of new cancer treatments, and to obtain additional safety data [3, 5]. In the context of single-arm phase II oncology trials, tumor response rate is popularly used as the primary endpoint [5]. In the recent times, however, progression-free survival (PFS) and overall survival (OS) have also been used as the primary endpoints of interest (see Rubinstein [5] for details). Given the time-consuming nature and skyrocketing costs associated with developing new treatments, it is imperative to pay considerable attention towards designing such early phase studies in a manner that can expedite evaluation of their efficacy or lack thereof, while preserving desired operating characteristics. The subject matter of this manuscript deals with sample size calculation and other planning aspects concerning single-arm phase II studies with time-to-event

(TTE) endpoints.

Upon literature review, it appears that only a limited number of options, primarily based on the log-rank test and its weighted versions, are available at our disposal for calculating the sample size needed to design single-arm clinical trials with TTE endpoints. Some of the existing methods include the ones proposed by Finkelstein et al. [7], Kwak and Jung [8], Sun et al. [9], Wu [10], and Phadnis [6]. Among these approaches, Wu's method [10] offers an improvement to the earlier methods by incorporating the exact variance of the log-rank test statistic into sample size calculations, and his method has been implemented in commercial software `PASS` [11] and `nQuery` [12]. Very recently, Phadnis [6] has extended the exact parametric approach by Narula and Li [13] assuming Weibull distributed survival times, and proposed sample size calculation procedure that adjusts for administrative censoring along with an ad-hoc inflation for random loss to follow-up. Parametric maximum likelihood estimate (MLE) test based on the exponential model has also been studied, but it is cautioned that this method may not be reliable under certain scenarios [14]. For instance, Owzar and Jung [14] demonstrated that the MLE-based test may fail to maintain the Type-I error when the underlying survival distribution does not follow the exponential distribution, and a uniform censoring mechanism is considered. To our knowledge, the utility of the MLE-based parametric test under the Weibull distribution has not been explored when the censoring mechanism also follows the Weibull distribution, and subject accrual is not necessarily uniform during accrual phase, such as observed in the studies with very early or late accrual patterns.

Two-stage designs (such as Simon's optimal and minimax designs) are popularly used in single-arm phase II oncology trials to obtain early evidence of futility of an experimental treatment which warrants early stopping of the clinical trial [15, 16]. Although a single futility analysis is the most likely scenario in a phase II trial, Kunz and Kieser [16] noted that a potential disadvantage of these two-stage designs is that "early stopping is only allowed at the end of the first stage, even if it becomes evident during the trial that a significant result is unlikely." To address this limitation, stochastic curtailment (SC) methods can be alternatively used to decide whether there is sufficient evidence in favor of the null hypothesis to 'curtail' sampling beyond an interim analysis conducted

over the course of a clinical trial [16, 17, 18]. Popular SC methods include: conditional power (Lan, Simon and Halprin [19], Andersen [20]), predictive power (Spiegelhalter et al. [21]), and Bayesian predictive probability (Herson [22], Choi and Pepple [23], Geisser [24], Dmitrienko and Wang [25]). For studies with normal- and binary- endpoints, most of these methods have been implemented in packages available in the statistical software `R` [26] and `SAS` [27]. For designing two-arm studies with TTE endpoints, the software `PASS` [11] can be used to carryout conditional power and predictive power calculations that are based on the log-rank test. From our review, it appears that the SC methods have not been well-studied when the underlying survival data follows the Weibull distribution, and various censoring mechanisms specific to survival analysis are under consideration.

Contrary to the SC approach, if the total number and the corresponding timings of interim analyses are pre-specified, group sequential design (GSD) plans can be developed using the repeated significance testing (RST) approach wherein 'stopping boundaries' for each interim analyses are constructed which dictate whether or not to continue the study beyond an interim analysis [17, 18, 28]. More specifically, if the observed test statistic at an interim time point (often referred as the 'look time') crosses either efficacy or futility boundary, researchers may terminate the trial in consultation with the Data Safety Monitoring Board (DSMB) overseeing the clinical trial [18, 29]. Some classical procedures for the implementation of RST approach include the works by Pocock [30], O'Brien and Fleming [31], and Wang and Tsiatis [32]. Under the umbrella of RST designs, error spending approaches (such as proposed by Lan and DeMets [33], Hwang, Shih and DeCani [34], and Jennison and Turnbull [35]) are extremely flexible in the sense that prior specification of the number and timing of interim analyses is not required.

The RST approach has been well-studied for normal- and binary- endpoints, such as described in Jennison and Turnbull [17], and the references therein. The RST approach for such endpoints has been implemented in various software such as `PASS` [11], `nQuery` [12], `R` [26], and `SAS` [27]. In comparison to studies with continuous or binary data where RST approach is implemented after a 'group' of subjects complete the study, the distinguishing characteristic of TTE data is

that all subjects, regardless of their survival status, contribute to the test statistic during interim analyses [18, 29]. In the context of two-arm phase III studies with TTE endpoints, methods based on the log-rank and its weighted versions, which assume exponentially distributed survival times or proportional hazards (PH), are usually employed to construct GSD plans. Simulation-based approach to implement such design with a variety of design options is available in the commercial software PASS [11, 29]. When the assumption of proportional hazards or exponentially distributed survival times is inappropriate, one may consider simulation-based approach recently proposed by Phadnis and Mayo [29] which utilizes the concept of proportional time (PT) to construct GSD plans for two-arm phase III studies. The method proposed by Phadnis and Mayo [29] for constructing GSD plans for the special case of single-arm phase II clinical trials is yet to be addressed.

Based on our discussion above, it appears that various design aspects of single-arm phase-II studies still need to be investigated when the parametric Weibull model is appropriate for modeling survival data derived from such studies. To fill these gaps in the statistical literature, we aim to achieve the following objectives in this manuscript. For the parametric Weibull model, we first propose a parametric MLE-based test whose variance component can account for the expected proportion of loss to follow-up and different subject accrual patterns. We discuss power and sample size calculations for single-arm single-stage phase-II studies using our method. From a practical perspective, the SC methods are extremely attractive, especially for futility monitoring, in such single-arm phase II studies. To cater such needs, we present detailed mathematical development and derive general results for performing interim analyses using three SC methods (conditional power, predictive power, Bayesian predictive probability) for the Weibull model for single-arm phase II studies with TTE endpoints. Finally, we briefly discuss the implementation of GSD plans for designing single-arm studies with Weibull survival data using the RST approach.

This manuscript is organized in the following order. In Section 2.2, we describe a real-life example which motivated the work presented in this manuscript. We describe our proposed method for calculating the required sample size for single-arm phase II studies with TTE endpoints specific to Weibull distributed survival times, derive results corresponding to the SC methods, and discuss

implementation of the GSD plans in Section 2.3. We present some simulation studies and examples to demonstrate these methods in Section 2.4. Finally, Section 2.5 entails a brief discussion on the methods presented in this manuscript.

2.2 Motivating Example

In this section, we briefly introduce a real-life example of a single-arm phase II clinical trial in oncology which motivated the methods presented in this manuscript. Recently, Phadnis [6] contributed in the design of a phase II clinical trial to investigate whether the use of new combination therapies improves the PFS among patients suffering from chemotherapy refractory advanced metastatic biliary cholangiocarcinoma, a "rare" but aggressive neoplasm. Such patients have metastatic disease and undergo an initial treatment followed by a second-line treatment which has a PFS rate of 5%–10% by 1 year. From historical control studies reported in the literature, it is understood that such patients have a median PFS of 2.5 months with an interquartile range of around 2–5 months. Due to dismal survival rates, researchers believe that the newly proposed combination therapies hold sufficient promise to be evaluated in future large sample studies if they yield a statistically significant improvement in the 25th, 50th and 75th percentile of PFS by a factor of 1.5. Therefore, for design purposes, researchers hypothesized a consistent improvement in PFS for all quantiles of the survival curve of the historical controls by a factor of 1.5, and the Weibull distribution was deemed to be an appropriate choice for performing the sample size calculations. The shape parameter for the Weibull distribution was estimated from the historical controls to be 1.25 (increasing hazard). Due to practical considerations, researchers envisioned to conduct a fixed sample study with an accrual time period of 2 years and a follow-up period of 3 years. In addition, they anticipated the random loss to follow-up rate to be around 15%–20%. Using the above design parameters, the method proposed by Phadnis [6] yielded a required sample size of 28 subjects when the nominal Type-I error and power were assumed to be 5% and 80%, respectively.

Due to ethical and administrative reasons, it would have been reasonable to design the above-mentioned study in such a manner that permits study investigators to conduct interim analysis to

obtain early evidence of efficacy or futility of the newly proposed combination therapies. To do so, one may employ popular SC approaches such as conditional power, predictive power, or Bayesian predictive probability. Since these methods have not been well-studied when the underlying survival data follows the Weibull distribution, it is worthwhile to fill these gaps in the literature as we anticipate to encounter similar studies with different design features (such as accrual patterns and random loss to follow-up) in the future. For the sake of exposition, we shall use simulated data sets to illustrate the methods presented in this manuscript.

2.3 Methods

2.3.1 Notation and Preliminaries

Suppose that a total of n subjects enroll during accrual phase of a single-arm phase II trial with a TTE primary endpoint. Due to practical constraints, it may be infeasible to wait until all subjects accrued into the study either experience an event or are lost to follow-up. Therefore, investigators may decide to incorporate administrative censoring at a pre-defined calendar time τ, when all remaining subjects in the study are censored and the resulting data are analyzed. For the ith subject, let E_i denote the calendar time of his/her accrual into the clinical trial; Y_i denote the amount of time from E_i to the time of the event corresponding to the primary outcome; C_i denote the amount of time from E_i to the time of loss to follow-up; $A_i := \max(0, \tau - E_i)$ denote the amount of time from E_i to the calendar time of administrative censoring at time τ, and $Z_i := \min(A_i, C_i)$ represents the amount of time from E_i to the time of administrative censoring, or loss to follow-up. We assume that the failure time is independent of the censoring time, and $\{Y_i, Z_i, \forall i = 1, \ldots, n\}$ are independent and identically distributed. In summary, we have two pieces of information available for the ith subject: (1) $X_i := \min(Y_i, Z_i)$, the amount of time in the study without experiencing the event, being censored administratively, or being lost to follow-up, and (2) $\delta_i := \mathbb{1}_{(Y_i < Z_i)}$, an indicator variable representing whether the ith subject was observed to experience an event. Thus, we have n pairs of data $\{(X_i, \delta_i), i = 1, \ldots, n\}$ for subjects enrolled in the study.

Distributional Assumptions

Throughout this manuscript, we assume the following event and censoring time distributions for the ith subject ($i = 1, \ldots, n$) enrolled into the study:

1. Event or failure time Y_i follows the Weibull distribution, with shape parameter κ and scale parameter θ, having the probability density function (pdf) expressed as below:

$$f_{Y_i}(y) = \frac{\kappa}{\theta^{\kappa}} y^{\kappa-1} \exp\left\{-\left(\frac{y}{\theta}\right)^{\kappa}\right\}, \qquad \text{where: } y > 0, \kappa > 0, \theta > 0. \tag{2.1}$$

   The shape parameter κ of the Weibull(κ, θ) distribution determines the shape of the underlying hazard function. More specifically, $\kappa > 1$ $(\kappa < 1)$ corresponds to an increasing (decreasing) hazard, and $\kappa = 1$ yields the special case of the exponential distribution having a constant hazard rate [36].

   We assume that a reasonably accurate estimate of the shape parameter κ is known in advance, and therefore the scale parameter θ is the only unknown quantity. This assumption is deemed justifiable based on the recent work by Phadnis et al. [37] in which they demonstrated that a point estimate of the shape parameter is reasonably accurate when estimated from some historical studies with at least 50 subjects, and censoring rate close to 20%.

2. Although maximum efforts should be put in place to avoid any non-negligible random loss to follow-up in the early phase clinical trials, there are real-life situations, such as described in Section 2.2, where researchers may feel appropriate to accommodate any expected proportion of loss to follow-up during the design stage of a study. Assume that the random loss to follow-up time C_i follows the Weibull distribution with the shape parameter κ_c and the scale parameter η. We feel important to point out that the loss to follow-up is assumed to be unrelated to the event of interest, that is, non-informative of the survival process. When the shape parameters of event and censoring time distributions are different, that is $\kappa \neq \kappa_c$, the appropriate scale parameter $\eta \equiv \eta(\kappa, \kappa_c, \theta, \upsilon)$ of the censoring time distribution which ensures the expected loss to follow-up rate υ can be obtained by following the method

outlined by Wan [38] and it satisfies the equation:

$$\upsilon = P(\delta_i = 0|\kappa, \kappa_c, \theta, \eta) = P(C_i \leq Y_i \leq \infty, 0 \leq C_i \leq \infty)$$
$$= \int_0^{\infty} \frac{\kappa_c}{\eta^{\kappa_c}} c_i^{\kappa_c - 1} \exp\left\{-\left(\frac{c_i}{\eta}\right)^{\kappa_c}\right\} \exp\left\{-\left(\frac{c_i}{\theta}\right)^{\kappa}\right\} dc_i$$

Since there is no closed form solution to the above equation, one must rely on numerical calculations to obtain the corresponding estimate of the scale parameter η. It is also worth noting that we would require reliable estimates of the two shape parameters (for the distributions of C_i and Y_i) from the historical studies whenever we aim to accommodate the general case $\kappa \neq \kappa_c$. It may not be feasible to incorporate the general case in the early phase studies related to rare diseases such as the one described in Section 2.2. Since our main objective is to accommodate the anticipated loss to follow-up rate υ within our sample size calculations at the design stage instead of doing an ad-hoc inflation of the sample size (as described in Phadnis [6]), we assume $\kappa_c = \kappa$ throughout this manuscript. It can be conveniently verified that $\eta = \theta \cdot \left(\frac{1-\upsilon}{\upsilon}\right)^{1/\kappa}$ ensures the expected loss to follow-up rate υ.

3. Suppose that ω represents the maximum calendar time of accrual into the study. Instead of using a uniform$(0, \omega)$ distribution for accrual time E_i, as used in the earlier works by Phadnis [6] and Wu [10], we consider a rather general form of the continuous uniform distribution, with an additional power parameter φ, having the following pdf (at a realized value e of E_i):

$$f_{E_i}(e) = \frac{\varphi e^{\varphi - 1}}{\omega^{\varphi}}, \qquad \text{where: } e \in [0, \omega], \varphi > 0. \tag{2.2}$$

Since there is no standardized terminology in the literature, we shall refer to the above pdf as Gen-Uniform$(0, \omega, \varphi)$ distribution in this manuscript. By letting $\varphi = 1$, we note that the uniform$(0, \omega)$ distribution is a special of the Gen-Uniform$(0, \omega, \varphi)$ distribution. This choice of accrual distribution is flexible in the sense that it enables us to incorporate accrual patterns that may tend to occur early on or very late in the accrual phase. More specifically, subjects

tend to get accrued early on in the accrual phase when $\varphi \to 0$, and they tend to get accrued very late into the accrual period when φ gets larger in magnitude.

4. Using the results presented in Casella and Berger [39], it can be verified that the pdf of $Z_i = \min(A_i, C_i)$ having parameters $\kappa > 0, \eta > 0, \omega > 0$, and $\varphi > 0$ is given as:

$$f_{Z_i}(z) = \begin{cases} \dfrac{\kappa}{\eta^\kappa} z^{\kappa-1} \exp\left\{-\left(\dfrac{z}{\eta}\right)^\kappa\right\} & \text{if } z \in [0, \tau-\omega) \\ \left(\dfrac{\varphi}{\tau-z} + \dfrac{\kappa}{\eta^\kappa} z^{\kappa-1}\right)\left(\dfrac{\tau-z}{\omega}\right)^\varphi \exp\left\{-\left(\dfrac{z}{\eta}\right)^\kappa\right\} & \text{if } z \in [\tau-\omega, \tau] \\ 0 & \text{otherwise} \end{cases} \tag{2.3}$$

The pdfs for the special cases of Z_i are presented in Section A.1 of Appendix A.

2.3.2 Fixed Sample Design

In this section, we introduce a fixed sample design that can be used to obtain preliminary evidence of efficacy of an experimental treatment in a single-arm phase II clinical trial with a TTE primary endpoint. To suffice our purpose, we may test the null hypothesis $H_0 : M \leq M_0$ against the alternative $H_1 : M > M_0$, where $M = \theta(\ln 2)^{1/\kappa}$ is the median of the Weibull(κ, θ) distribution specified in Eq. (2.1). Since a reasonable estimate of the shape parameter κ is assumed to be known from historical studies, we may alternatively define our hypotheses as: $H_0 : \theta \leq \theta_0$ versus $H_1 : \theta > \theta_0$. In the context of parametric models in survival analysis, covariates are commonly introduced through the scale parameter as $\theta = \exp\{\boldsymbol{\gamma}^T \boldsymbol{x}\}$, where: $\boldsymbol{x} = (1, x_1, \ldots, x_k)^T$ is the vector of $k+1$ covariates, and $\boldsymbol{\gamma}^T = (\gamma_0, \gamma_1, \ldots, \gamma_k)$ is the corresponding vector of parameters to be estimated [36]. When no covariates other than the experimental treatment administered to the subjects are introduced into the model, the scale parameter can be expressed as $\theta = \exp\{\gamma\}$. Thus, our hypotheses can be equivalently expressed as:

$$\begin{aligned} H_0 &: \gamma \leq \gamma_0 \\ H_1 &: \gamma > \gamma_0 \end{aligned} \tag{2.4}$$

From Klein and Moeschberger [36], it is straightforward to verify that the maximum likelihood estimator (MLE) $\widehat{\gamma}$ of γ is given as:

$$\widehat{\gamma} = \log\left(\frac{\sum_{i=1}^{n} X_i^{\kappa}}{\sum_{i=1}^{n} \delta_i}\right)^{1/\kappa} = \frac{1}{\kappa}\log\left(\frac{\overline{X^{\kappa}}}{\overline{\delta}}\right), \tag{2.5}$$

where $\overline{X^{\kappa}} = \frac{1}{n}\sum_{i=1}^{n} X_i^{\kappa}$, $\overline{\delta} = \frac{1}{n}\sum_{i=1}^{n} \delta_i$, and the definitions of these random variables and distributional assumptions have been presented in Section 2.3.1.

It appears analytically intractable to obtain the exact distribution of $\widehat{\gamma}$ due to the underlying correlation between a subject's survival time and the corresponding survival status. Therefore, we rely on asymptotic calculations to construct a parametric MLE-based statistic for testing the hypotheses in Eq. (2.4). For this purpose, we first obtain the joint asymptotic distribution of $\overline{X^{\kappa}}$ and $\overline{\delta}$ using the multivariate central limit theorem, and subsequently employ the multivariate delta method to obtain the asymptotic distribution of $\widehat{\gamma}$. Without loss of generality, it can be shown that

$$\lim_{n\to\infty} \sqrt{n}\,(\widehat{\gamma} - \gamma) \xrightarrow{d} \text{Normal}\left(0, \sigma^2 = \frac{1}{\kappa^2 \mu_{\bar{\delta}}}\right), \tag{2.6}$$

where $\mu_{\bar{\delta}} = 1 - E_{Z_1}\left(\exp\left\{-\left(\frac{Z_1}{\exp\{\gamma\}}\right)^{\kappa}\right\}\right)$, and $Z_1 = \min(A_1, C_1)$. We skip a detailed derivation of this result, and present the derivation of a similar but more general result in a later section.

Let $\widehat{\sigma}^2 \equiv \sigma^2(\widehat{\gamma})$ denote the MLE plugged-in estimate of σ^2, then, under the null hypothesis, the Wald's test statistic is given as

$$Z_{stat} = \frac{\widehat{\gamma} - \gamma_0}{\widehat{\sigma}/\sqrt{n}} \;\dot{\sim}\; \text{Normal}(0,1). \tag{2.7}$$

For a given level of significance α, we reject the null hypothesis in favor of the alternative hypothesis when the observed test statistic $\widehat{Z}_{stat} > Z_{1-\alpha}$, where $Z_{1-\alpha} = \Phi^{-1}(1-\alpha)$ represents the upper α-quantile of the standard normal distribution.

For sample size calculations, researchers specify a clinically meaningful difference $\varepsilon > 0$ that they are interested in detecting under the alternative hypothesis $H_0 : \gamma > \gamma_1\ (= \gamma_0 + \varepsilon)$. Let $\sigma_1^2 \equiv$

$\sigma^2(\gamma_1)$ denote the plug-in estimator of σ^2 under H_1. The desired power $1-\beta$ of the Wald's test statistic should satisfy:

$$1-\beta = \text{Prob}(Z_{stat} > Z_{1-\alpha}|H_1) \simeq \Phi\left(\frac{\sqrt{n}\varepsilon}{\sigma_1} - Z_{1-\alpha}\right),$$

where $\Phi(\cdot)$ denotes the cumulative distribution function (cdf) of the standard normal distribution.

The required sample size to detect the difference ε using the Wald's test statistic in Eq. (2.7) with a Type I error rate α and power $1-\beta$ satisfies:

$$n = \sigma_1^2 \cdot \left(\frac{Z_{1-\beta} + Z_{1-\alpha}}{\varepsilon}\right)^2.$$

To compute the required sample size, we can use numerical integration to calculate σ_1^2.

When we are interested in testing the two-sided alternatives $H_0 : \gamma = \gamma_0$ vs. $H_1 : \gamma \neq \gamma_0$, we reject the null hypothesis H_0 at an α level of significance if the magnitude of the observed test statistic $|\widehat{Z}_{stat}| > Z_{1-\frac{\alpha}{2}}$. To design such a study with some pre-specified Type I error rate α and power $1-\beta$, the desired sample size n satisfies:

$$n = \sigma_1^2 \left(\frac{Z_{1-\beta} + Z_{1-\frac{\alpha}{2}}}{\varepsilon}\right)^2.$$

2.3.3 Stochastic Curtailment Methods

Due to ethical reasons, it is critical to detect lack of therapeutic effect of ineffective treatments being evaluated in the early phase studies as soon as possible in an effort to minimize risk to the subjects, and to direct resources to research on more promising treatments [17, 18]. When the primary motive behind interim analyses is to find early evidence of futility, SC methods are generally employed. These methods enable researcher to evaluate the likelihood of positive or negative trial outcome if the study were to continue to its planned end, given the current data [18]. In the context of parametric Weibull model considered in this manuscript, we discuss three SC methods: conditional power (purely frequentist approach), predictive power (mixed Bayesian-

frequentist approach), and Bayesian predictive probability (purely Bayesian approach).

2.3.3.1 Conditional Power

Introduced by Lan, Simon and Halprin [19], the conditional power relies on "predicting the distribution of the final outcome given the data already observed in the study [18]." Let Z_k denote the interim test statistic at the kth look time, and Z_K represents the test statistic at the trial end. For testing the one-sided hypotheses in Eq. (2.4) at an α level of significance, we may consider a decision rule based on the conditional power [17, 18, 19] at an interim stage k defined as follows:

$$P_k(\gamma) = \text{Prob}_\gamma(Z_K \text{ will reject } H_0 | Z_k) \tag{2.8}$$

Lan, Simon and Halprin [19] also suggested rules to aid decision making regarding early termination of a clinical trial. A high value of $P_k(\gamma_0)$ suggest that the test is unlikely to accept H_0 given the current data at the interim stage k, even if H_0 is true. Therefore, reject H_0 at the kth look if $P_k(\gamma_0) \geq \zeta$, where ζ can range from 0.5 to 1, but it is recommended to be 0.8 or 0.9 [17, 18]. On the other hand, a low value of $P_k(\gamma_1)$ suggests that the test is unlikely to reject H_0 given the current data at interim stage k, even if H_1 is true. Therefore, it fails to reject H_0 at the kth interim look if $1 - P_k(\gamma_1) \geq \zeta'$, where ζ' can range from 0.5 to 1. In the literature, the quantity $1 - P_k(\gamma_1)$ is called as *futility index* [17, 18].

Since SC methods can be implemented in a post-hoc manner without explicitly adjusting for the repeated testing approach, we now present the conditional power function for doing interim testing related to a fixed sample design in the context of the parametric Weibull model considered in this manuscript [18]. Since the sequence of test statistics $Z_1, \ldots, Z_K$ asymptotically follows a joint multivariate normal distribution (described in Section 2.3.4), it can be verified [17] that the conditional distribution of the final statistic Z_K given the interim statistic Z_k is:

$$Z_K | Z_k \sim \text{Normal}\left(\frac{Z_k \sqrt{\mathscr{I}_k} + (\gamma - \gamma_0)(\mathscr{I}_K - \mathscr{I}_k)}{\sqrt{\mathscr{I}_K}}, 1 - \frac{\mathscr{I}_k}{\mathscr{I}_K} \right),$$

where $\mathscr{I}_k$ and $\mathscr{I}_K$ represent the statistical information available at the interim look k, and the final look K, respectively.

Following Jennison and Turnbull [17], the conditional power at the interim stage $k = 1, \ldots, K-1$ can be expressed as:

$$P_k(\gamma) = \Phi\left(\frac{Z_k\sqrt{\mathscr{I}_k} + (\gamma - \gamma_0)(\mathscr{I}_K - \mathscr{I}_k) - Z_{1-\alpha}\sqrt{\mathscr{I}_K}}{\sqrt{\mathscr{I}_K - \mathscr{I}_k}}\right), \tag{2.9}$$

where $Z_{1-\alpha} = \Phi^{-1}(1-\alpha)$ is the upper α-quantile of the standard normal distribution.

To do conditional power calculations, we can compute $\mathscr{I}_K$ using the information formula from the fixed sample design, and $\mathscr{I}_k = \mathscr{F}_k \times \mathscr{I}_K$ where $\mathscr{F}_k$ is the proportion of the total information $\mathscr{I}_K$ available at the kth interim look. When unscheduled interim analyses are performed using the conditional power calculations for fixed sample studies originally designed with Type-I error α and Type-II error rate β, it has been shown by Lan, Simon and Halprin [19] that the Type-I and Type-II error rates are bounded above by α/ζ and β/ζ', respectively. To ensure maximum Type-I and Type-II error rates to be α and β, respectively, Jennison and Turnbull [17] have recommended to design fixed sample studies with Type-I error rate $\alpha\zeta$ and Type-II error rate $\beta\zeta'$.

Corresponding to the decision rules for early termination using the conditional power in Eq. (2.9), it has been shown [17, 18] that the formal sequential stopping boundaries to reject H_0 are given by:

$$Z_k \geq Z_{1-\alpha}\sqrt{\frac{\mathscr{I}_K}{\mathscr{I}_k}} + Z_\zeta\sqrt{\frac{\mathscr{I}_K - \mathscr{I}_k}{\mathscr{I}_k}}. \tag{2.10}$$

The sequential stopping boundaries to accept H_0 at the kth interim look are given by:

$$Z_k \leq Z_{1-\alpha}\sqrt{\frac{\mathscr{I}_K}{\mathscr{I}_k}} - Z_{\zeta'}\sqrt{\frac{\mathscr{I}_K - \mathscr{I}_k}{\mathscr{I}_k}} - (\gamma - \gamma_0)\left(\frac{\mathscr{I}_K - \mathscr{I}_k}{\sqrt{\mathscr{I}_k}}\right), \tag{2.11}$$

where $Z_{1-\alpha} = \Phi^{-1}(1-\alpha)$, $Z_\zeta = \Phi^{-1}(\zeta)$, $Z_{\zeta'} = \Phi^{-1}(\zeta')$, and $\Phi^{-1}(\cdot)$ is the inverse cdf of the standard normal distribution.

2.3.3.2 Predictive Power

One of the major criticism of conditional power is that it is computed using the values of γ under the hypotheses which may not be supported by the current data [17, 18]. To address this issue, Spiegelhalter et al. [21] recommended the use of "predictive power" at the interim stage k, whereby conditional power function (frequentist component) is averaged over the posterior distribution (Bayesian component) of γ given its interim estimate $\widehat{\gamma}_k$. Mathematically, it is given as:

$$P_k = \int P_k(\gamma)\pi(\gamma|\widehat{\gamma}_k)d\gamma \tag{2.12}$$

where $P_k(\gamma)$ denotes the conditional power function expressed in Eq. (2.9), and $\pi(\gamma|\widehat{\gamma}_k)$ is the posterior distribution of γ given its estimate $\widehat{\gamma}_k$ at the interim stage k.

In the context of our problem, we have already expressed the asymptotic distribution of $\widehat{\gamma}|\gamma$ in Eq. (2.6). Since there exists a mean-variance relationship in the asymptotic distribution of $\widehat{\gamma}|\gamma$, it appears intractable to derive a nice closed form asymptotic posterior distribution $\pi(\gamma|\widehat{\gamma}_k)$. In this situation, we resort to the Metropolis-Hastings (MH) algorithm for generating samples from the asymptotic posterior distribution:

$$\pi(\gamma|\widehat{\gamma}_k) \propto \pi(\widehat{\gamma}_k|\gamma)\pi(\gamma),$$

where $\pi(\gamma)$ denotes the prior distribution for the parameter γ. Assuming a vague normal prior for γ, a step-by-step implementation of the MH algorithm is outlined in Section A.2 of Appendix A. Alternatively, the posterior distribution $\pi(\gamma|\widehat{\gamma}_k)$ can also be generated using `R2OPenBUGS` package available in the statistical software `R`.

After generating asymptotic posterior distribution of the parameter γ, we can numerically evaluate the predictive power at the kth interim stage using Eq. (2.12). We can use similar decision rules to stop for efficacy or futility as the ones in conditional power calculations [17].

2.3.3.3 Bayesian Predictive Probability

Since predictive power is a mixed Bayesian-frequentist approach, Jennison and Turnbull stated that "neither Bayesian nor frequentist statisticians may be satisfied" with the decision rules based on this method [35]. We may alternatively consider a fully Bayesian approach which relies on the idea of *predictive probability* of obtaining a positive trial outcome at the end of the study, given the current estimates at an interim stage [18, 25].

In TTE framework, let us consider that a study comprising of n subjects has been designed to test the hypotheses $H_0 : \gamma \leq \gamma_0$ vs. $H_1 : \gamma = \gamma_1$, where $\gamma_1 = \gamma_0 + \varepsilon$ and $\varepsilon > 0$ is a clinical meaningful effect to be detected under the alternative hypothesis. Utilizing the same notations from Section 2.3.1, the MLE of γ at the planned end (at calendar time τ), denoted by $\widehat{\gamma}_K$, is given in Eq. (2.5). At the kth interim analysis, suppose that $n-m$ subjects had already experienced an event or were censored due to loss to follow-up. In addition, we suppose that the remaining m subjects were still active in the study without experiencing the event of interest. Then, the MLE of γ at the pre-planned end time τ can be equivalently expressed as:

$$\widehat{\gamma}_K = \frac{1}{\kappa} \log \left(\frac{\mathscr{T}_k + \mathscr{T}_{K-k}}{\mathscr{D}_k + \mathscr{D}_{K-k}} \right), \tag{2.13}$$

where $\mathscr{T}_k = \sum_{i=1}^{n-m} X_i^{\kappa}$ and $\mathscr{D}_k = \sum_{i=1}^{n-m} \delta_i$ are the quantities corresponding to subjects who had experienced an event or were censored due to loss to follow-up by the interim stage k; and $\mathscr{T}_{K-k} = \sum_{j=1}^{m} X_j^{\kappa}$ and $\mathscr{D}_{K-k} = \sum_{j=1}^{m} \delta_j$ is the final data information at stage K for those subjects who were active in the study at the interim stage k.

Since $\mathscr{T}_{K-k}$ and $\mathscr{D}_{K-k}$ are not observable at the interim stage k, suppose that $\tilde{\mathscr{T}}_{K-k}$ and $\tilde{\mathscr{D}}_{K-k}$ denote the predicted values of $\mathscr{T}_{K-k}$ and $\mathscr{D}_{K-k}$ at the interim stage k, respectively. Following Dmitreinko and Wang [18], the predictive probability of obtaining a positive trial outcome at the pre-planned end date of a clinical trial, given the data accumulated at the kth stage, is defined as:

$$P_k = \int \mathbb{1}_{(\mathscr{Q} > \eta^*)} \, d\tilde{P} \tag{2.14}$$

where $\mathscr{Q} = \text{Prob}\big(\gamma > \gamma_1 | \mathscr{T}_k, \tilde{\mathscr{T}}_{K-k}, \mathscr{D}_k, \tilde{\mathscr{D}}_{K-k}\big)$, η^* is a pre-specified threshold level of probability of a successful trial outcome, and $\tilde{P}$ is the joint posterior predictive distribution of $\tilde{\mathscr{T}}_{K-k}$ and $\tilde{\mathscr{D}}_{K-k}$.

The threshold level η^* is recommended to be set between 0.90 and 0.975 in the literature [25]. In practice, Dmitreinko and Wang [25] note that researchers can terminate a trial at an interim stage k to conclude efficacy (i.e., reject H_0 in the favor of H_1) if $P_k \geq \zeta$ for some pre-specified $\zeta \in [0.8, 1]$, and conclude futility if $P_k \leq \zeta'$ for some $\zeta' \in [0, 0.2]$.

Returning to our main problem, we let $\overline{\tilde{\mathscr{D}}}_{K-k} = \tilde{\mathscr{D}}_{K-k}/m$ and $\overline{\tilde{\mathscr{T}}}_{K-k} = \tilde{\mathscr{T}}_{K-k}/m$, and note from Eq. (2.13) that the estimated predicted MLE at the pre-planned study end can be re-expressed as:

$$\widehat{\tilde{\gamma}}_K = \frac{1}{\kappa} \log \left(\frac{\mathscr{T}_k + m\overline{\tilde{\mathscr{T}}}_{K-k}}{\mathscr{D}_k + m\overline{\tilde{\mathscr{D}}}_{K-k}} \right). \tag{2.15}$$

Given a threshold level η^*, we implement the following algorithm (using the asymptotic joint posterior predictive distribution of $\overline{\tilde{\mathscr{D}}}_{K-k}$ and $\overline{\tilde{\mathscr{T}}}_{K-k}$) at an interim stage k to compute the predictive probability of a positive outcome at the end of the study:

1. Obtain an estimate $\widehat{\gamma}_k$ of the parameter γ at the interim stage k, and use the MH algorithm to generate samples from the asymptotic posterior distribution of γ as:

$$\pi(\gamma | \widehat{\gamma}_k) \propto \pi(\widehat{\gamma}_k | \gamma) \pi(\gamma).$$

2. Let $\tilde{P} = \left(\overline{\tilde{\mathscr{D}}}_{K-k}, \overline{\tilde{\mathscr{T}}}_{K-k} \right)$. We can numerically obtain the asymptotic posterior predictive distribution of $\tilde{P}$ as:

$$\pi\big(\tilde{P} | \widehat{\gamma}_k\big) = \int \pi\big(\tilde{P} | \gamma\big) \pi(\gamma | \widehat{\gamma}_k) d\gamma.$$

   For plausible values of γ based on the asymptotic posterior distribution generated in Step 1, let $\vartheta = \exp\{\gamma\}$ be the scale parameter of the failure time distribution, and ϑ_c denote the scale parameter of the loss to follow-up distribution. Using the multivariate central limit theorem, it can be shown that the joint asymptotic distribution of $\overline{\tilde{\mathscr{D}}}_{K-k}$ and $\overline{\tilde{\mathscr{T}}}_{K-k}$ has the following

property:

$$\lim_{m\to\infty}\sqrt{m}\left(\begin{pmatrix}\overline{\tilde{\mathscr{D}}}_{K-k}\\ \overline{\tilde{\mathscr{T}}}_{K-k}\end{pmatrix}-\begin{pmatrix}\mu_{\tilde{\delta}}\\ \mu_{\tilde{X}}\end{pmatrix}\right)\xrightarrow{d}\text{Normal}\left(\begin{pmatrix}0\\ 0\end{pmatrix},\begin{pmatrix}\sigma_1^2 & \sigma_{12}\\ \sigma_{12} & \sigma_2^2\end{pmatrix}\right),$$

where:

$$\sigma_1^2=\mu_{\tilde{\delta}}\left(1-\mu_{\tilde{\delta}}\right)$$

$$\sigma_{12}=\vartheta^{\kappa}\xi-\mu_{\tilde{\delta}}\mu_{\tilde{X}}$$

$$\sigma_2^2=E_{B_1}\left(B_1^{2\kappa}\right)+2\vartheta^{2\kappa}\xi-\mu_{\tilde{X}}^2$$

$$\mu_{\tilde{\delta}}=1-E_{B_1}\left(\exp\left\{\left(\frac{B_1}{\vartheta}\right)^{\kappa}\right\}E_{\tilde{Z}_1}\left(\exp\left\{-\left(\frac{\tilde{Z}_1}{\vartheta}\right)^{\kappa}\right\}\right)\right)$$

$$\mu_{\tilde{X}}=E_{B_1}(B_1^{\kappa})+\vartheta^{\kappa}\mu_{\tilde{\delta}}$$

$$\xi=\mu_{\tilde{\delta}}+\frac{1}{\vartheta^{\kappa}}\left(E_{B_1}(B_1^{\kappa})-E_{B_1}\left(\exp\left\{\left(\frac{B_1}{\vartheta}\right)^{\kappa}\right\}E_{\tilde{Z}_1}\left(\tilde{Z}_1^{\kappa}\exp\left\{-\left(\frac{\tilde{Z}_1}{\vartheta}\right)^{\kappa}\right\}\right)\right)\right)$$

A detailed derivation of this result, and the pdfs of the random variables B_1 and $\tilde{Z}_1$ are presented in Appendix B.

3. For each plausible combination of $\overline{\tilde{\mathscr{D}}}_{K-k}$ and $\overline{\tilde{\mathscr{T}}}_{K-k}$ among the remaining subjects, estimate the MLE of γ at the end of the study using Eq. (2.15), and subsequently calculate the corresponding quantity $\mathscr{Q}^*=\text{Prob}\left(\gamma>\gamma_1|\mathscr{T}_k,\overline{\tilde{\mathscr{T}}}_{K-k},\mathscr{D}_k,\overline{\tilde{\mathscr{D}}}_{K-k}\right)$. Then the predictive probability of a successful outcome can be computed as:

$$P_k=\int\mathbb{1}_{(\mathscr{Q}^*>\eta^*)}d\tilde{P},$$

where $\tilde{P}$ is the asymptotic joint posterior predictive distribution of $\overline{\tilde{\mathscr{D}}}_{K-k}$ and $\overline{\tilde{\mathscr{T}}}_{K-k}$.

The above algorithm based on the asymptotic posterior distribution of $\widehat{\gamma}_k$ and the asymptotic joint

posterior predictive distribution of $\overline{\widetilde{\mathscr{D}}}_{K-k}$ and $\overline{\widetilde{\mathscr{T}}}_{K-k}$ can be implemented in software such as `R` and `SAS`. Alternatively, a purely simulation-based algorithm for calculating the predictive probability may also be adopted using `R2OpenBUGS` package in `R` [26], and it is outlined as below:

1. Using the data available at the interim stage k, generate the posterior distribution of the parameter γ, denoted by $\gamma|\widehat{\gamma}_k$, using the Weibull model (available as `dweib` function) in `R2OpenBUGS` package in `R`.

2. Create a sequence of length S of the plausible values of the parameter in the posterior distribution generated in Step 1, and perform the following steps:

   (a) For each value γ_j in the sequence, generate 'predicted' survival data for the remaining subjects (using appropriate truncated distributions) at the pre-planned end of the study. Let us denote the vectors of predicted survival times and the corresponding survival status as $\boldsymbol{X}_{m;pred}$ and $\boldsymbol{\delta}_{m;pred}$, respectively.

   (b) Let $\boldsymbol{X}_{(n-m);obs}$ and $\boldsymbol{\delta}_{(n-m);obs}$ denote the data for the $n-m$ subjects completely observed by the interim stage k. Using the observed and predicted data, generate the posterior distribution of the parameter at study end using `R2OpenBUGS`, and determine whether

$$\text{Prob}\big(\gamma > \gamma_1 | \boldsymbol{\delta}_{(n-m);obs}, \boldsymbol{\delta}_{m;pred}, \boldsymbol{X}_{(n-m);obs}, \boldsymbol{X}_{m;pred}\big) > \eta^*. \tag{2.16}$$

   (c) Repeat Steps 2a and 2b for a large number of times (say $d = 10,000$), and the probability of success corresponding to a particular value of γ_j is the proportion of d predicted samples satisfying Eq. (2.16). In this manner, we are able to obtain an estimate of $\text{Prob}\big(\text{Positive Outcome}|\gamma_j\big)$.

3. Finally, we estimate the predicted probability of a positive trial outcome by:

$$P_k = \sum_{j=1}^{S} \text{Prob}\big(\text{Positive Outcome}|\gamma_j\big)\text{Prob}\big(\gamma_j|\widehat{\gamma}_k\big)$$

Once the predictive probability of a positive outcome is computed, researchers make a decision

regarding early termination in consultation with the DSMB overseeing the clinical trial.

2.3.4 Repeated Significance Testing for Group Sequential Design

Before discussing the RST approach in the context of single-arm studies for testing the hypothesis in Eq. (2.4), let us introduce some additional notations used in this section. Let $K \in \mathbb{N}^+$ be the total number of "looks" to be performed in the group sequential framework, and ℓ_k $(k = 1, \ldots, K)$ denote the calendar time corresponding to the kth look. Also, let $\Delta_g = \gamma_g - \gamma_0$, and $\mathscr{I}_{g,k}$ represent the corresponding statistical information at the kth look time under the hypothesis H_g $(g = 0, 1)$ stated in Eq. (2.4). Furthermore, let $Z_{g,k}$ denote the MLE-based parametric test statistic at the kth look time ℓ_k under H_g. Putting everything together, it follows that the observed test statistic

$$\widehat{Z}_{g,k} = \widehat{\Delta}_g \sqrt{\widehat{\mathscr{I}}_{g,k}} \dot{\sim} \text{Normal}\Big(\Delta_g \sqrt{\mathscr{I}_{g,k}}\ ,\ 1\Big),$$

at the kth interim look under the hypothesis H_g.

For a GSD plan with desired level of significance α, simply rejecting the null hypothesis at the kth interim look if $Z_{0,k} > Z_{1-\alpha} = \Phi^{-1}(1-\alpha)$ would result in the inflation of Type-I error [17]. To address this issue, several methods have been proposed in the literature, such as Pocock [30], O'Brien and Fleming [31], Wang and Tsiatis [32], Lan and DeMets [33], Hwang, Shih and DeCani [34], and Jennison and Turnbull [35], among others, to construct appropriate rejection boundaries. Suppose $\mathscr{U}_k$ and $\mathscr{L}_k$ $(k = 1, \ldots, K)$ denote the efficacy and futility stopping boundaries for a GSD at the kth look time, respectively. The rejection boundaries for testing both efficacy as well as futility can be constructed in the following manner:

1. To maintain an overall Type-I rate α, we allocate the local Type-I error rate α_k $(k = 1, \ldots, K)$ for each of the k looks such that $\sum_{k=1}^{K} \alpha_k = \alpha$. In a similar fashion, we can maintain an overall Type-II error β by allocating local Type-II error rates β_k for each look such that $\sum_{k=1}^{K} \beta_k = \beta$ [17, 18, 30, 31].

2. The efficacy and futility stopping boundaries $\mathscr{U}_k$ and $\mathscr{L}_k$ $(k = 1, \ldots, K-1)$ can be successively

calculated such that

$$\text{Prob}\big(\mathscr{L}_1 < \text{Z}_{0,1} < \mathscr{U}_1, \ldots, \mathscr{L}_{k-1} < \text{Z}_{0,k-1} < \mathscr{U}_{k-1}, \text{Z}_{0,k} \geq \mathscr{U}_k\big) = \alpha_k \tag{2.17}$$

$$\text{Prob}\big(\mathscr{L}_1 < \text{Z}_{1,1} < \mathscr{U}_1, \ldots, \mathscr{L}_{k-1} < \text{Z}_{1,k-1} < \mathscr{U}_{k-1}, \text{Z}_{1,k} < \mathscr{L}_k\big) = \beta_k \tag{2.18}$$

3. In the event trial continues to its pre-planned end, we must have $\mathscr{L}_K = \mathscr{U}_K$ to ensure a decision at the final look time ℓ_K.

From a theoretical perspective, we can follow Mütze et al. [40] to construct rejection boundaries using the asymptotic joint multivariate distribution of the test statistics at each look in order to implement the above mentioned RST approach for single-arm studies. Since our proposed Wald's test is based on the MLE of γ, therefore, following Scharfstein et al. [41], the vector of Wald test statistics $\big(Z_{g,1}, \ldots, Z_{g,K}\big)$ under the hypothesis H_g follows an asymptotic joint multivariate normal distribution with mean $\boldsymbol{\mu}_g \in \mathbb{R}^K$ and $K \times K$-dimensional covariance matrix $\boldsymbol{\Sigma}_g$ expressed as below:

$$\boldsymbol{\mu}_g = \begin{pmatrix} \Delta_g \sqrt{\mathscr{I}_{g,1}} \\ \Delta_g \sqrt{\mathscr{I}_{g,2}} \\ \vdots \\ \Delta_g \sqrt{\mathscr{I}_{g,K}} \end{pmatrix} \quad ; \quad \boldsymbol{\Sigma}_g = \begin{pmatrix} \sigma_{g,1,1} & \sigma_{g,1,2} & \cdots & \sigma_{g,1,K} \\ \sigma_{g,2,1} & \sigma_{g,2,2} & \cdots & \sigma_{g,2,K} \\ \vdots & \vdots & \ddots & \vdots \\ \sigma_{g,K,1} & \sigma_{g,K,2} & \cdots & \sigma_{g,K,K} \end{pmatrix} \tag{2.19}$$

where $\sigma_{g,k_1,k_2} = \sigma_{g,k_2,k_1} = \sqrt{\mathscr{I}_{g,k_1} / \mathscr{I}_{g,k_2}}$ for $1 \leq k_1 \leq k_2 \leq K$. This is often referred as the *canonical joint distribution* with information levels $\big\{\mathscr{I}_{g,1}, \ldots, \mathscr{I}_{g,K}\big\}$ [17].

Once the joint multivariate distribution of the test statistics is known, numerical recursive integration, such as described by Armitage, McPherson and Rowe [42], can be employed to determine the efficacy and futility stopping boundaries.

The allocation of the local error rates can be done using the error spending approach, initially proposed by Lan and DeMets [33], which is extremely flexible in the sense that it enables researchers and DSMB to change the number and timing of interim looks [17, 18]. In this

manuscript, we consider the error spending functions proposed by Hwang, Shih and DeCani (HSD) [34], and Jennison and Turnbull (JT) [35]. Let $\mathscr{F}_k \in (0,1)$ denote the fraction of information observed by the kth interim look ($k = 1,\ldots,K-1$), and $\mathscr{F}_K = 1$. The HSD α-spending function [34] is defined as:

$$\alpha(\mathscr{F}_k) = \begin{cases} \dfrac{\alpha \cdot \left(1 - e^{-\psi \mathscr{F}_k}\right)}{1 - e^{-\psi}} & \text{if } \psi \neq 0 \\ \alpha \cdot \mathscr{F}_k & \text{if } \psi = 0 \end{cases} \tag{2.20}$$

On the other hand, the JT α-spending function [35] is given as:

$$\alpha(\mathscr{F}_k) = \alpha \cdot (\mathscr{F}_k)^{\psi}. \tag{2.21}$$

The parameter ψ in Eqs. (2.20) and (2.21) influences the amount of cumulative error rate spent at each look, and hence influence the amount of local error rate to be spent at each look. Both of the above-mentioned error spending functions provide approximations to Pocock [30], and O'Brien and Fleming [31] methods for constructing stopping boundaries for designs with pre-specfied look times [18]. The HSD α-spending function approximates the O'Brien and Fleming boundaries when $\psi = -4$ or $\psi = -5$, whereas JT α-spending function approximates O'Brien and Fleming boundaries when $\psi = 3$. Both HSD and JT α-spending functions approximate the Pocock stopping boundaries when $\psi = 1$. The rejection boundaries for futility testing should be constructed using desired β-spending function obtained in Eqs. (2.20) and (2.21).

For testing hypotheses defined in Eq. (2.4) for a GSD using the RST approach, the power of the GSD with K looks with efficacy stopping boundaries $\mathscr{U}_1,\ldots,\mathscr{U}_K$ is defined as:

$$\text{Power} = 1 - \text{Prob}(Z_{1,1} < \mathscr{U}_1,\ldots,Z_{1,K} < \mathscr{U}_K). \tag{2.22}$$

Given the fraction of information $\mathscr{F}_k$ $(k = 1,\ldots,K)$ available at the kth interim look, we observe that the corresponding statistical information can be expressed in terms of the information at the

final look as $\mathscr{I}_{g,k} = \mathscr{F}_k \times \mathscr{I}_{g,K}$, where $\mathscr{F}_k \in (0,1)$ and $\mathscr{F}_K = 1$ at the final look time ℓ_K. As noted in Jennison and Turnbull [17] and Mütze et al. [40] , the mean of the joint asymptotic multivariate normal distribution of the test statistics expressed in Eq. (2.19) only depends on the maximum information $\mathscr{I}_{g,K}$ as:

$$\boldsymbol{\mu}_g = \left(\Delta_g \sqrt{\mathscr{F}_1 \mathscr{I}_{g,K}} \quad \Delta_g \sqrt{\mathscr{F}_2 \mathscr{I}_{g,K}} \quad \dots \quad \Delta_g \sqrt{\mathscr{I}_{g,K}} \right)^T, \tag{2.23}$$

and the entries of the $K \times K$ covariance matrix $\boldsymbol{\Sigma}_g$ are given as $\sigma_{g,k_1,k_2} = \sigma_{g,k_2,k_1} = \sqrt{\mathscr{F}_{k_1} / \mathscr{F}_{k_2}}$ for $1 \leq k_1 \leq k_2 \leq K$.

For power and sample size calculations, researchers specify the effect ε that they are interested in detecting under H_1 as: $\varepsilon = \gamma_1 - \gamma_0$. Consequently, the mean of the *canonical joint distribution* under H_1 can be expressed as:

$$\boldsymbol{\mu}_1 = \left(\varepsilon \sqrt{\mathscr{F}_1 \mathscr{I}_{1,K}} \quad \varepsilon \sqrt{\mathscr{F}_2 \mathscr{I}_{1,K}} \quad \dots \quad \varepsilon \sqrt{\mathscr{I}_{1,K}} \right)^T.$$

This implies that $\mathscr{I}_{1,K}$ is the only missing piece in Eq. (2.22), and it can be obtained by numerically solving this equation. Once we do so, the required sample size for a K look single-arm GSD plan with Type-I error rate α and desired power $1-\beta$ can be obtained by equating $\mathscr{I}_{1,fix} = \mathscr{I}_{1,K}$, where $\mathscr{I}_{1,fix}$ denotes the information under the alternative hypothesis H_1 for a fixed sample design with the same operating characteristics [40].

In Section A.3 of Appendix A, we have outlined a simulation-based approach for constructing GSD plans for single-arm studies with TTE primary endpoints. This algorithm follows the general framework implemented in the software `PASS` [11] for designing GSD plans for two-arm studies with TTE endpoints, and offers flexibility to incorporate numerous user-defined options (such as accrual and follow-up times, different accrual patterns, custom look times, binding or non-binding futility, etc.) in its calculations.

2.4 Simulations and Examples

In this section, we present some simulations studies and examples to demonstrate the methods proposed in this manuscript. Statistical software `R` (Version 3.6.3) was used to perform all computations and simulations presented in this section.

2.4.1 Performance of the Wald's Test Statistic

We conducted simulation studies to examine the performance of the Wald's test statistic (presented in Section 2.3.2) in terms of the empirical Type-I error and power under a wide range of scenarios. For these simulations, the nominal Type-I error rate and power were set to be equal to 5% and 90%, respectively. The maximum accrual time was assumed to be fixed at $\omega = 3$ months. In addition, varying values of the administrative censoring time ($\tau = 4, 7, 9$ months), the common shape parameter ($\kappa = 0.25, 0.50, 1.00, 2.00, 5.00$), expected loss to follow-up rate ($\upsilon = 0\%, 10\%, 20\%, 30\%$), power parameter of the accrual distribution ($\varphi = 0.1 -$ early; $1.0 -$ uniform; $5.0 -$ late), and effect size ($\Delta = 1.2, 1.4, 1.6, 1.8, 2.0$) were considered. Using Wu's [10] notation, the effect size was defined as $\Delta = (M_1/M_0)^{\kappa}$ with the median survival time under the null hypothesis assumed to be fixed at $M_0 = 1$ month. A total of 10,000 simulations were performed after computing the required sample size in each scenario, and these results are presented in Tables 2.1, 2.2, and 2.3. We summarize the main findings as below:

1. In most of the cases, the empirical Type-I error was maintained within the nominal level. It is worth noting that the empirical Type-I error tends to exceed the nominal level primarily when either: (i) the effect size Δ is very small ($= 1.2$) in magnitude, or (ii) the common shape parameter κ is small ($= 0.25$) which results in larger magnitudes of the median under the alternative hypothesis.

2. Empirical power was observed to be very close to the nominal levels. By referring to Table 2.1, one might argue that the proposed test is under-powered in most of the scenarios. It must be noted, however, that our proposed test is based on the asymptotic approximations, whereas

most of the sample sizes reported in Table 2.1 are less than 30. For moderate to large sample sizes, we can observe that the empirical power is very close to the desired nominal level.

3. Keeping all other parameters fixed, it appears that the required sample size is bounded below for a large value of the shape parameter κ, or corresponding values of large follow-up times.

Based on our simulation results, it appears that the desired study sample size is moderate to large when there is at least some level (greater than 10% or so) of censoring due to loss to follow-up, and the effect size is moderate to large in magnitude. In such scenarios, the proposed asymptotic test statistic is most appropriate as it preserves the nominal Type-I error rate and power within our desired levels.

Since the shape parameter of the Weibull distribution determines the shape of the underlying hazard function, we also conducted a simulation study to assess the effect of its misspecification on the Type-I error and power. For this purpose, we consider that a fixed sample study was originally designed to detect an effect size $\Delta = 1.2$ and $\Delta = 1.6$ assuming exponential survival times (i.e. $\kappa = 1$). Using varying values of other design parameters, we assessed the empirical Type-I error and power when the true shape parameter was in fact 0.75 (decreasing hazard) and 1.25 (increasing hazard). Corresponding results have been reported in Table 2.4. We summarize main findings of this simulation study as below:

1. When the effect size is small ($\Delta = 1.2$), the empirical Type-I error tends to exceed the nominal level in most of the cases. On the other hand, when the effect size is larger in magnitude ($\Delta = 1.6$), the empirical Type-I error was maintained within the nominal level except the setting with late subject accrual and shorter follow-up time. This is also consistent with the findings of our simulation studies described above.

2. Empirical power is significantly affected due to misspecification of the shape parameter of the Weibull distribution. More specifically, we observe that the fixed sample study is over-powered (under-powered) when the true shape parameter is larger (smaller) than the one used at the design stage of the study.

Table 2.1: Empirical Type-I error and power of the MLE-based test for hypothetical studies with maximum accrual time $\omega = 3$ months and administrative censoring time $\tau = 4$ months under varying values of the power parameter φ, loss to follow-up rate υ, shape parameter κ, and effect size $\Delta = (M_1/M_0)^\kappa$ with $M_0 = 1$ month

φ	υ	κ	$\Delta = 1.2$			$\Delta = 1.4$			$\Delta = 1.6$			$\Delta = 1.8$			$\Delta = 2.0$		
			n	α	$1-\beta$	n	α	$1-\beta$	n	α	$1-\beta$	n	α	$1-\beta$	n	α	$1-\beta$
0.1	0%	0.25	469	0.0570	0.9011	153	0.0532	0.8976	86	0.0519	0.9009	61	0.0539	0.9068	47	0.0536	0.9057
		0.50	386	0.0490	0.9036	124	0.0473	0.9110	69	0.0432	0.9201	48	0.0413	0.9220	37	0.0444	0.9279
		1.00	295	0.0502	0.9009	91	0.0476	0.8998	49	0.0460	0.9072	33	0.0392	0.9176	25	0.0398	0.9222
		2.00	262	0.0538	0.8856	77	0.0519	0.8774	40	0.0438	0.8797	26	0.0453	0.8802	19	0.0417	0.8835
		5.00	259	0.0547	0.8877	76	0.0455	0.8777	39	0.0492	0.8797	25	0.0461	0.8767	18	0.0458	0.8744
	10%	0.25	487	0.0530	0.8900	158	0.0572	0.8996	89	0.0528	0.9014	62	0.0561	0.9017	48	0.0534	0.9032
		0.50	406	0.0514	0.9021	130	0.0471	0.9051	72	0.0463	0.9112	50	0.0451	0.9227	38	0.0398	0.9209
		1.00	319	0.0494	0.8956	98	0.0460	0.9047	53	0.0418	0.9104	35	0.0424	0.9129	27	0.0417	0.9248
		2.00	290	0.0503	0.8915	86	0.0448	0.8899	44	0.0458	0.8838	29	0.0450	0.8877	21	0.0411	0.8903
		5.00	288	0.0499	0.8934	85	0.0489	0.8872	44	0.0430	0.8892	28	0.0448	0.8879	20	0.0441	0.8737
	20%	0.25	510	0.0609	0.8931	165	0.0570	0.8981	92	0.0597	0.8975	64	0.0604	0.9011	50	0.0588	0.9008
		0.50	431	0.0499	0.9017	137	0.0470	0.9120	76	0.0524	0.9145	52	0.0509	0.9213	40	0.0459	0.9228
		1.00	350	0.0446	0.9018	107	0.0487	0.9046	57	0.0444	0.9086	38	0.0411	0.9139	29	0.0433	0.9204
		2.00	326	0.0477	0.8962	96	0.0498	0.8947	50	0.0467	0.8899	32	0.0459	0.8879	23	0.0429	0.8927
		5.00	323	0.0483	0.8941	95	0.0449	0.8884	49	0.0457	0.8829	32	0.0420	0.8932	23	0.0429	0.8937
	30%	0.25	541	0.0592	0.8885	174	0.0617	0.8893	97	0.0655	0.8942	67	0.0592	0.9040	52	0.0600	0.8993
		0.50	465	0.0522	0.9054	146	0.0517	0.9047	80	0.0519	0.9081	55	0.0537	0.9184	42	0.0568	0.9213
		1.00	390	0.0515	0.9006	118	0.0507	0.9028	63	0.0453	0.9125	42	0.0467	0.9156	31	0.0462	0.9224
		2.00	371	0.0493	0.8962	110	0.0463	0.8922	57	0.0476	0.8945	36	0.0461	0.8895	26	0.0455	0.8888
		5.00	369	0.0510	0.9002	109	0.0486	0.8940	56	0.0436	0.8967	36	0.0418	0.8944	26	0.0480	0.8925
1.0	0%	0.25	505	0.0589	0.8908	166	0.0599	0.8895	94	0.0563	0.8942	66	0.0606	0.8991	52	0.0544	0.8978
		0.50	439	0.0524	0.9010	143	0.0498	0.9105	80	0.0487	0.9094	56	0.0508	0.9118	44	0.0501	0.9215
		1.00	352	0.0467	0.9054	111	0.0419	0.9052	61	0.0449	0.9098	42	0.0407	0.9230	32	0.0377	0.9318
		2.00	290	0.0528	0.8909	88	0.0482	0.8930	47	0.0424	0.8974	31	0.0425	0.8990	23	0.0407	0.9051
		5.00	268	0.0510	0.8878	80	0.0473	0.8863	41	0.0434	0.8858	27	0.0467	0.8862	20	0.0449	0.8834
	10%	0.25	523	0.0645	0.8848	171	0.0573	0.8897	96	0.0593	0.8923	67	0.0611	0.8911	53	0.0589	0.8915
		0.50	458	0.0541	0.9012	148	0.0536	0.9025	83	0.0531	0.9103	58	0.0522	0.9125	45	0.0532	0.9177
		1.00	374	0.0486	0.9016	118	0.0463	0.9021	64	0.0451	0.9180	44	0.0388	0.9223	34	0.0394	0.9290
		2.00	317	0.0480	0.8961	96	0.0452	0.9007	51	0.0442	0.8979	33	0.0407	0.9002	25	0.0405	0.9074
		5.00	296	0.0493	0.8924	88	0.0472	0.8936	46	0.0439	0.8928	30	0.0460	0.8944	21	0.0394	0.8766
	20%	0.25	546	0.0629	0.8905	177	0.0660	0.8880	100	0.0580	0.8812	69	0.0604	0.8849	54	0.0612	0.8841
		0.50	483	0.0540	0.8990	155	0.0545	0.8978	86	0.0556	0.9085	60	0.0536	0.9111	46	0.0566	0.9093
		1.00	402	0.0473	0.9020	126	0.0454	0.9126	69	0.0440	0.9156	47	0.0458	0.9150	35	0.0477	0.9228
		2.00	350	0.0506	0.8999	106	0.0446	0.9022	56	0.0443	0.9034	36	0.0412	0.8956	27	0.0378	0.9110
		5.00	331	0.0501	0.8998	98	0.0472	0.8917	51	0.0454	0.8894	33	0.0408	0.8927	24	0.0443	0.8949
	30%	0.25	576	0.0685	0.8817	186	0.0656	0.8830	104	0.0668	0.8841	72	0.0654	0.8809	56	0.0693	0.8836
		0.50	515	0.0557	0.8943	164	0.0577	0.8943	91	0.0572	0.8910	63	0.0558	0.9064	48	0.0567	0.9092
		1.00	440	0.0516	0.8976	136	0.0483	0.9088	74	0.0441	0.9118	50	0.0506	0.9131	38	0.0499	0.9214
		2.00	394	0.0490	0.8996	118	0.0473	0.9025	62	0.0472	0.9011	41	0.0481	0.9061	30	0.0446	0.9027
		5.00	377	0.0447	0.8987	112	0.0504	0.8976	58	0.0505	0.8971	37	0.0442	0.8932	27	0.0450	0.8917
5.0	0%	0.25	549	0.0630	0.8787	181	0.0635	0.8770	103	0.0682	0.8746	72	0.0623	0.8824	57	0.0607	0.8847
		0.50	514	0.0620	0.8888	169	0.0577	0.8908	96	0.0639	0.8961	67	0.0568	0.8869	53	0.0568	0.8963
		1.00	454	0.0563	0.9028	148	0.0487	0.8991	83	0.0492	0.9059	58	0.0518	0.9114	45	0.0488	0.9126
		2.00	376	0.0486	0.9028	119	0.0415	0.9062	66	0.0411	0.9152	45	0.0391	0.9220	35	0.0364	0.9272
		5.00	303	0.0497	0.8992	92	0.0425	0.8912	49	0.0408	0.8967	32	0.0419	0.9004	24	0.0360	0.8995
	10%	0.25	567	0.0665	0.8764	186	0.0658	0.8797	105	0.0680	0.8806	74	0.0597	0.8778	58	0.0648	0.8768
		0.50	532	0.0588	0.8856	174	0.0651	0.8917	98	0.0612	0.8891	69	0.0592	0.8968	54	0.0648	0.8946
		1.00	473	0.0543	0.8972	153	0.0559	0.8983	86	0.0565	0.9068	60	0.0552	0.9196	46	0.0522	0.9140
		2.00	398	0.0459	0.9026	126	0.0472	0.9105	69	0.0468	0.9084	47	0.0415	0.9156	36	0.0430	0.9275
		5.00	330	0.0475	0.8970	100	0.0424	0.8983	53	0.0416	0.8978	35	0.0401	0.8983	26	0.0369	0.9023
	20%	0.25	589	0.0715	0.8745	192	0.0670	0.8837	109	0.0675	0.8778	76	0.0698	0.8736	59	0.0697	0.8713
		0.50	555	0.0653	0.8789	180	0.0649	0.8895	102	0.0648	0.8886	71	0.0633	0.8872	55	0.0582	0.8851
		1.00	498	0.0571	0.8923	160	0.0545	0.8975	89	0.0565	0.9006	62	0.0574	0.9120	48	0.0524	0.9054
		2.00	427	0.0509	0.9060	134	0.0437	0.9052	73	0.0476	0.9057	50	0.0438	0.9165	38	0.0447	0.9174
		5.00	364	0.0521	0.8988	110	0.0463	0.8976	58	0.0435	0.8990	38	0.0417	0.8962	28	0.0435	0.8965
	30%	0.25	618	0.0724	0.8733	201	0.0719	0.8733	113	0.0704	0.8752	79	0.0762	0.8757	61	0.0700	0.8775
		0.50	585	0.0695	0.8782	189	0.0660	0.8852	106	0.0629	0.8857	73	0.0681	0.8826	57	0.0708	0.8793
		1.00	531	0.0587	0.8966	169	0.0565	0.8932	94	0.0559	0.8966	65	0.0622	0.9006	50	0.0581	0.9059
		2.00	465	0.0505	0.9011	145	0.0467	0.9053	79	0.0509	0.9074	53	0.0504	0.9122	40	0.0532	0.9126
		5.00	407	0.0503	0.9032	123	0.0464	0.9007	65	0.0450	0.9029	43	0.0490	0.9089	31	0.0480	0.8963

Note: Results are based on a total of 10,000 simulations.

Table 2.2: Empirical Type-I error and power of the MLE-based test for hypothetical studies with maximum accrual time $\omega = 3$ months and administrative censoring time $\tau = 7$ months under varying values of the power parameter φ, loss to follow-up rate υ, shape parameter κ, and effect size $\Delta = (M_1/M_0)^\kappa$ with $M_0 = 1$ month

φ	υ	κ	$\Delta = 1.2$			$\Delta = 1.4$			$\Delta = 1.6$			$\Delta = 1.8$			$\Delta = 2.0$		
			n	α	$1-\beta$	n	α	$1-\beta$	n	α	$1-\beta$	n	α	$1-\beta$	n	α	$1-\beta$
0.1	0%	0.25	437	0.0536	0.9006	142	0.0502	0.9069	80	0.0515	0.9102	56	0.0495	0.9115	43	0.0444	0.9134
		0.50	345	0.0446	0.9065	110	0.0430	0.9098	61	0.0422	0.9196	42	0.0416	0.9285	32	0.0371	0.9300
		1.00	269	0.0556	0.8894	81	0.0467	0.8866	43	0.0468	0.8967	28	0.0445	0.8947	21	0.0437	0.8944
		2.00	258	0.0526	0.8832	76	0.0524	0.8828	39	0.0508	0.8759	25	0.0436	0.8727	18	0.0415	0.8718
		5.00	258	0.0580	0.8764	76	0.0491	0.8797	39	0.0472	0.8800	25	0.0475	0.8741	18	0.0444	0.8707
	10%	0.25	456	0.0547	0.9031	147	0.0526	0.9086	83	0.0529	0.9080	57	0.0477	0.9069	45	0.0525	0.9189
		0.50	366	0.0473	0.9032	116	0.0438	0.9102	64	0.0437	0.9204	44	0.0429	0.9283	33	0.0421	0.9300
		1.00	295	0.0482	0.8919	89	0.0492	0.8997	47	0.0491	0.9013	31	0.0405	0.9046	23	0.0447	0.9092
		2.00	287	0.0500	0.8920	85	0.0481	0.8928	44	0.0448	0.8846	28	0.0458	0.8768	20	0.0434	0.8749
		5.00	287	0.0509	0.8922	85	0.0487	0.8861	44	0.0480	0.8864	28	0.0438	0.8744	20	0.0432	0.8756
	20%	0.25	480	0.0564	0.9003	154	0.0542	0.9051	86	0.0574	0.9032	60	0.0561	0.9122	46	0.0529	0.9088
		0.50	393	0.0489	0.9030	123	0.0475	0.9047	67	0.0473	0.9161	46	0.0466	0.9217	35	0.0474	0.9244
		1.00	329	0.0487	0.9004	98	0.0557	0.8944	51	0.0485	0.8969	34	0.0454	0.9002	25	0.0411	0.9060
		2.00	323	0.0508	0.8905	95	0.0465	0.8966	49	0.0436	0.8927	31	0.0440	0.8829	23	0.0438	0.8901
		5.00	323	0.0512	0.9006	95	0.0504	0.8880	49	0.0421	0.8835	31	0.0441	0.8843	23	0.0449	0.8907
	30%	0.25	511	0.0625	0.8956	163	0.0534	0.9013	90	0.0574	0.9009	62	0.0581	0.9037	48	0.0562	0.9017
		0.50	429	0.0482	0.9010	133	0.0503	0.9076	72	0.0527	0.9117	49	0.0504	0.9201	37	0.0483	0.9229
		1.00	373	0.0449	0.8987	111	0.0489	0.8955	58	0.0446	0.9033	38	0.0488	0.8996	28	0.0460	0.9141
		2.00	369	0.0499	0.8947	109	0.0477	0.8965	56	0.0453	0.8930	36	0.0463	0.8905	26	0.0472	0.8908
		5.00	369	0.0488	0.8994	109	0.0455	0.8997	56	0.0494	0.8949	36	0.0469	0.8887	26	0.0452	0.8889
1.0	0%	0.25	455	0.0534	0.8973	148	0.048	0.9014	84	0.0493	0.9069	58	0.0483	0.9085	46	0.0498	0.9151
		0.50	367	0.0482	0.9027	117	0.0467	0.9133	65	0.0419	0.9187	45	0.0401	0.9272	35	0.0421	0.9256
		1.00	282	0.0508	0.8925	86	0.0466	0.8932	46	0.0430	0.9019	31	0.0422	0.9090	23	0.0375	0.9183
		2.00	258	0.0495	0.8787	76	0.0484	0.8844	39	0.0524	0.8755	25	0.0450	0.8755	18	0.0433	0.8633
		5.00	258	0.0547	0.8821	76	0.0524	0.8717	39	0.0479	0.8732	25	0.0496	0.8737	18	0.0442	0.8743
	10%	0.25	473	0.0549	0.8946	153	0.0565	0.8988	86	0.0514	0.9106	60	0.0548	0.9046	47	0.0511	0.9092
		0.50	387	0.0490	0.9081	123	0.0488	0.9065	68	0.0426	0.9149	47	0.0475	0.9228	36	0.0395	0.9255
		1.00	307	0.0513	0.8958	93	0.0476	0.8994	50	0.0457	0.9068	33	0.0423	0.9075	25	0.0414	0.9203
		2.00	287	0.0503	0.8856	85	0.0465	0.8767	44	0.0428	0.8827	28	0.0447	0.8800	20	0.0434	0.8764
		5.00	287	0.0540	0.8861	85	0.0504	0.8848	44	0.0461	0.8875	28	0.0467	0.8805	20	0.0410	0.8813
	20%	0.25	497	0.0559	0.8934	160	0.0574	0.8948	90	0.0597	0.9024	62	0.0567	0.9052	48	0.0579	0.9009
		0.50	413	0.0493	0.9024	131	0.0483	0.9094	72	0.0479	0.9165	49	0.0510	0.9204	38	0.0449	0.9238
		1.00	338	0.0488	0.8974	102	0.0480	0.9050	54	0.0471	0.9041	36	0.0460	0.9115	27	0.0400	0.9134
		2.00	323	0.0488	0.8909	95	0.0465	0.8884	49	0.0471	0.8897	32	0.0413	0.8890	23	0.0446	0.8906
		5.00	323	0.0507	0.8979	95	0.0508	0.8931	49	0.0486	0.8897	31	0.0457	0.8806	23	0.0450	0.8878
	30%	0.25	528	0.0600	0.8952	169	0.0584	0.8934	94	0.0622	0.8986	65	0.0582	0.9044	50	0.0597	0.9018
		0.50	448	0.0481	0.9053	140	0.0486	0.9067	77	0.0502	0.9118	52	0.0534	0.9147	40	0.0521	0.9211
		1.00	380	0.0478	0.9050	114	0.0465	0.9016	60	0.0463	0.9059	40	0.0491	0.9092	29	0.0486	0.9092
		2.00	369	0.0496	0.9007	109	0.0503	0.8904	56	0.0496	0.8957	36	0.0504	0.8966	26	0.0465	0.8934
		5.00	369	0.0488	0.8981	109	0.0442	0.8954	56	0.0452	0.8935	36	0.0443	0.8915	26	0.0460	0.8943
5.0	0%	0.25	473	0.0537	0.8960	154	0.0568	0.894	87	0.0534	0.8992	61	0.0527	0.906	48	0.0503	0.9089
		0.50	391	0.0489	0.9086	126	0.0462	0.9162	70	0.0451	0.9187	49	0.0442	0.9259	38	0.0442	0.9303
		1.00	299	0.0449	0.8972	93	0.0451	0.9076	50	0.0443	0.9105	34	0.0392	0.9227	26	0.0377	0.9260
		2.00	259	0.0526	0.8841	76	0.0517	0.8831	40	0.0455	0.8813	26	0.0462	0.8800	19	0.0428	0.8858
		5.00	258	0.0505	0.8790	76	0.0535	0.8731	39	0.0474	0.8774	25	0.0473	0.8678	18	0.0426	0.866
	10%	0.25	491	0.0523	0.8917	160	0.0505	0.8973	90	0.0529	0.8986	63	0.0594	0.9052	49	0.0575	0.9017
		0.50	411	0.0527	0.9023	132	0.0457	0.9086	73	0.0464	0.9119	51	0.0477	0.9237	39	0.0455	0.9250
		1.00	322	0.0467	0.9040	99	0.0451	0.9035	54	0.0449	0.9098	36	0.0394	0.9191	27	0.0437	0.9235
		2.00	287	0.0525	0.8852	85	0.0498	0.8858	44	0.0464	0.8889	28	0.0475	0.8828	21	0.0431	0.8933
		5.00	287	0.0536	0.8888	85	0.0444	0.8874	44	0.0465	0.8869	28	0.0434	0.8793	20	0.0421	0.8828
	20%	0.25	514	0.0597	0.8927	166	0.061	0.8963	93	0.0512	0.8983	65	0.0548	0.8999	50	0.0544	0.8945
		0.50	436	0.0497	0.9024	139	0.0475	0.9113	77	0.0459	0.9109	53	0.0514	0.9148	41	0.0522	0.9237
		1.00	352	0.0471	0.8984	108	0.0444	0.9030	58	0.0430	0.9112	39	0.0432	0.9173	29	0.0445	0.9205
		2.00	323	0.0498	0.8967	95	0.0509	0.8891	49	0.0418	0.8897	32	0.0444	0.8837	23	0.0431	0.8893
		5.00	323	0.0538	0.8926	95	0.0533	0.8895	49	0.0468	0.8905	31	0.0467	0.8833	23	0.0403	0.8863
	30%	0.25	545	0.0590	0.8894	175	0.0619	0.8912	98	0.0616	0.8961	68	0.0666	0.896	52	0.0617	0.8992
		0.50	469	0.0548	0.9008	148	0.0539	0.9068	81	0.0535	0.9110	56	0.0556	0.9138	43	0.0578	0.0919
		1.00	392	0.0495	0.9010	119	0.0474	0.9079	63	0.0467	0.9114	42	0.0467	0.9119	32	0.0448	0.9199
		2.00	369	0.0485	0.9000	109	0.0505	0.8982	56	0.0473	0.8973	36	0.0426	0.8933	26	0.0447	0.8968
		5.00	369	0.0473	0.9026	109	0.0479	0.8928	56	0.0477	0.8945	36	0.0491	0.8913	26	0.0463	0.8941

Note: Results are based on a total of 10,000 simulations.

Table 2.3: Empirical Type-I error and power of the MLE-based test for hypothetical studies with maximum accrual time $\omega = 3$ months and administrative censoring time $\tau = 9$ months under varying values of the power parameter φ, loss to follow-up rate υ, shape parameter κ, and effect size $\Delta = (M_1/M_0)^\kappa$ with $M_0 = 1$ month

φ	υ	κ	$\Delta = 1.2$			$\Delta = 1.4$			$\Delta = 1.6$			$\Delta = 1.8$			$\Delta = 2.0$		
			n	α	$1-\beta$	n	α	$1-\beta$	n	α	$1-\beta$	n	α	$1-\beta$	n	α	$1-\beta$
0.1	0%	0.25	410	0.0535	0.9052	133	0.0509	0.9068	74	0.0437	0.9166	52	0.0499	0.9260	40	0.0465	0.9238
		0.50	315	0.0460	0.9055	99	0.0475	0.9047	54	0.0452	0.9128	37	0.0396	0.9222	28	0.0401	0.9282
		1.00	260	0.0537	0.8848	77	0.0489	0.8766	40	0.0533	0.8829	26	0.0438	0.8773	19	0.0406	0.8773
		2.00	258	0.0546	0.8813	76	0.0485	0.8768	39	0.0486	0.8763	25	0.0456	0.8752	18	0.0436	0.8728
		5.00	258	0.0564	0.8832	76	0.0511	0.8769	39	0.0460	0.8760	25	0.0451	0.8736	18	0.0464	0.8753
	10%	0.25	429	0.0510	0.9065	138	0.0481	0.9065	77	0.0509	0.9130	53	0.0428	0.9155	41	0.0486	0.9177
		0.50	338	0.0515	0.9086	105	0.0430	0.9105	57	0.0449	0.9106	39	0.0439	0.9220	30	0.0393	0.9340
		1.00	288	0.0543	0.8867	85	0.0435	0.8934	44	0.0443	0.8931	29	0.0430	0.8916	21	0.0437	0.8871
		2.00	287	0.0514	0.8863	85	0.0502	0.8872	44	0.0455	0.8895	28	0.0426	0.8861	20	0.0403	0.8797
		5.00	287	0.0534	0.8886	85	0.0504	0.8876	44	0.0510	0.8844	28	0.0455	0.8767	20	0.0445	0.8750
	20%	0.25	454	0.0529	0.9038	145	0.0497	0.9061	81	0.0542	0.9127	56	0.0506	0.9208	43	0.0512	0.9191
		0.50	366	0.0461	0.9055	113	0.0422	0.9063	61	0.0440	0.9164	41	0.0426	0.9164	31	0.0473	0.9222
		1.00	323	0.0493	0.8958	96	0.0478	0.8919	49	0.0478	0.8905	32	0.0461	0.8883	23	0.0427	0.8882
		2.00	323	0.0501	0.8903	95	0.0499	0.8914	49	0.0440	0.8931	31	0.0446	0.8853	23	0.0426	0.8949
		5.00	323	0.0509	0.8939	95	0.0462	0.8869	49	0.0451	0.8902	31	0.0411	0.8839	23	0.0444	0.8877
	30%	0.25	486	0.0552	0.8957	154	0.0556	0.8995	85	0.0534	0.9077	58	0.0541	0.9123	45	0.0537	0.9120
		0.50	404	0.0444	0.9052	124	0.0430	0.9078	67	0.0490	0.9123	45	0.0473	0.9179	34	0.0468	0.9244
		1.00	369	0.0458	0.8958	109	0.0497	0.8930	56	0.0470	0.8957	36	0.0464	0.8946	26	0.0461	0.8913
		2.00	369	0.0440	0.8947	109	0.0452	0.8964	56	0.0469	0.8903	36	0.0450	0.8885	26	0.0476	0.8913
		5.00	369	0.0510	0.8946	109	0.0521	0.9015	56	0.0466	0.8945	36	0.0519	0.8938	26	0.0473	0.8991
1.0	0%	0.25	419	0.0482	0.9005	136	0.0471	0.9099	76	0.0457	0.9114	53	0.0450	0.9168	41	0.0454	0.9228
		0.50	325	0.0467	0.9013	103	0.0448	0.9093	56	0.0422	0.9140	39	0.0372	0.9279	30	0.0386	0.9356
		1.00	262	0.0523	0.8880	78	0.0490	0.8867	41	0.0474	0.8893	27	0.0438	0.8919	20	0.0425	0.8991
		2.00	258	0.0566	0.8865	76	0.0508	0.8836	39	0.0511	0.8800	25	0.0467	0.8711	18	0.0399	0.8694
		5.00	258	0.0546	0.8887	76	0.0531	0.8795	39	0.0486	0.8732	25	0.0463	0.8756	18	0.0403	0.8654
	10%	0.25	438	0.0495	0.9026	141	0.0486	0.9051	79	0.0493	0.9111	55	0.0506	0.9165	43	0.0510	0.9224
		0.50	347	0.0485	0.9081	109	0.0444	0.9142	59	0.0454	0.9138	40	0.0426	0.9210	31	0.0406	0.9322
		1.00	289	0.0498	0.8899	86	0.0507	0.8917	45	0.0484	0.8933	29	0.0465	0.8878	22	0.0436	0.9008
		2.00	287	0.0494	0.8871	85	0.0458	0.8819	44	0.0476	0.8866	28	0.0425	0.8838	20	0.0444	0.8707
		5.00	287	0.0530	0.8933	85	0.0490	0.8933	44	0.0445	0.8843	28	0.0452	0.8817	20	0.0393	0.8837
	20%	0.25	463	0.0544	0.9011	148	0.0567	0.9038	82	0.0510	0.9099	57	0.0528	0.9087	44	0.0532	0.9194
		0.50	375	0.0466	0.9058	117	0.0447	0.9056	63	0.0468	0.9158	43	0.0483	0.9221	33	0.0395	0.9300
		1.00	324	0.0486	0.8927	96	0.0460	0.8941	50	0.0449	0.9035	32	0.0458	0.8925	24	0.0434	0.9076
		2.00	323	0.0481	0.8970	95	0.0502	0.8889	49	0.0440	0.8866	31	0.0429	0.8843	23	0.0463	0.8922
		5.00	323	0.0516	0.8872	95	0.0468	0.8888	49	0.0470	0.8853	31	0.0446	0.8898	23	0.0436	0.8881
	30%	0.25	495	0.0577	0.8932	157	0.0561	0.9018	87	0.0553	0.9024	60	0.0576	0.9043	46	0.0536	0.9124
		0.50	412	0.0499	0.9005	127	0.0469	0.9073	68	0.0483	0.9110	46	0.0505	0.9172	35	0.0461	0.9239
		1.00	370	0.0477	0.9011	109	0.0448	0.8955	57	0.0488	0.9013	37	0.0484	0.8989	27	0.0445	0.8992
		2.00	369	0.0504	0.8943	109	0.0490	0.8978	56	0.0464	0.8963	36	0.0465	0.8973	26	0.0453	0.8888
		5.00	369	0.0461	0.8935	109	0.0467	0.8986	56	0.0445	0.8914	36	0.0470	0.8960	26	0.0492	0.8909
5.0	0%	0.25	428	0.0526	0.9000	139	0.0488	0.9071	78	0.0521	0.9112	54	0.0471	0.9096	42	0.0466	0.9198
		0.50	335	0.0401	0.9111	106	0.0449	0.9117	59	0.0441	0.9252	40	0.0408	0.9260	31	0.0370	0.9309
		1.00	264	0.0570	0.8896	79	0.0490	0.8840	42	0.0433	0.8905	28	0.0441	0.9048	20	0.0431	0.8947
		2.00	258	0.0555	0.8791	76	0.0513	0.8786	39	0.0465	0.8717	25	0.0442	0.8793	18	0.0431	0.8706
		5.00	258	0.0516	0.8826	76	0.0489	0.8806	39	0.0454	0.8717	25	0.0473	0.8763	18	0.0427	0.8699
	10%	0.25	447	0.0525	0.9005	144	0.0528	0.9077	81	0.0519	0.9142	56	0.0524	0.9139	44	0.0492	0.9191
		0.50	356	0.0444	0.9097	112	0.0423	0.9107	61	0.0451	0.9146	42	0.0414	0.9209	32	0.0435	0.9305
		1.00	291	0.0494	0.8916	87	0.0458	0.8980	46	0.0482	0.8962	30	0.0429	0.8935	22	0.0413	0.9033
		2.00	287	0.0487	0.8988	85	0.0485	0.8824	44	0.0464	0.8845	28	0.0430	0.8872	20	0.0442	0.8764
		5.00	287	0.0525	0.8902	85	0.0495	0.8895	44	0.0466	0.8836	28	0.0427	0.8825	20	0.0399	0.8701
	20%	0.25	471	0.0551	0.8978	151	0.0519	0.8990	84	0.0516	0.9064	58	0.0518	0.9093	45	0.0491	0.9114
		0.50	383	0.0438	0.9042	120	0.0479	0.9133	65	0.0477	0.9094	44	0.0446	0.9178	34	0.0447	0.9307
		1.00	326	0.0463	0.8972	97	0.0463	0.8921	50	0.0462	0.8855	33	0.0469	0.8991	24	0.0444	0.9040
		2.00	323	0.0541	0.8951	95	0.0492	0.8912	49	0.0472	0.8867	31	0.0425	0.8829	23	0.0451	0.8874
		5.00	323	0.0477	0.8918	95	0.0459	0.8874	49	0.0463	0.8872	31	0.0454	0.8805	23	0.0449	0.8912
	30%	0.25	503	0.0595	0.8986	160	0.0574	0.8937	89	0.0548	0.9027	61	0.0581	0.9120	47	0.0566	0.9073
		0.50	420	0.0535	0.8987	130	0.0462	0.9142	70	0.0479	0.9086	47	0.0521	0.9135	36	0.0448	0.9249
		1.00	370	0.0498	0.8986	110	0.0469	0.8974	57	0.0489	0.9015	37	0.0461	0.9011	27	0.0457	0.8996
		2.00	369	0.0519	0.8974	109	0.0462	0.8976	56	0.0497	0.9010	36	0.0473	0.8966	26	0.0452	0.8950
		5.00	369	0.0535	0.8995	109	0.0440	0.8963	56	0.0448	0.8943	36	0.0484	0.8952	26	0.0483	0.8903

Note: Results are based on a total of 10,000 simulations.

Based on our findings, we feel important to emphasize that the proposed method is most appropriate when adequate data is available from historical control studies for reliably estimating the shape parameter of the Weibull distribution. If such estimate of the shape parameter is not available, it is recommended to incorporate adaptive features, such as sample size reassessment, into our study design.

Table 2.4: Effect of misspecification of the shape parameter on Type-I error and power of the MLE-based test when a fixed sample study was designed assuming exponential survival times ($\kappa = 1$)

φ	υ	True κ	$\Delta = 1.2$						$\Delta = 1.6$					
			$\tau = 4$		$\tau = 7$		$\tau = 9$		$\tau = 4$		$\tau = 7$		$\tau = 9$	
			α	$1-\beta$	α	$1-\beta$	α	$1-\beta$	α	$1-\beta$	α	$1-\beta$	α	$1-\beta$
0.1	0%	0.75	0.0466	0.6716	0.0507	0.6814	0.0503	0.6854	0.0425	0.6796	0.0417	0.6984	0.0449	0.6883
		1.25	0.0488	0.9823	0.0527	0.9743	0.0546	0.9714	0.0429	0.9837	0.0463	0.9738	0.0466	0.9649
	10%	0.75	0.0466	0.6820	0.0478	0.7043	0.0479	0.7017	0.0439	0.6954	0.0446	0.7042	0.0433	0.6919
		1.25	0.0491	0.9815	0.0511	0.9727	0.0545	0.9745	0.0445	0.9839	0.0456	0.9747	0.0467	0.9682
	20%	0.75	0.0499	0.6846	0.0507	0.7000	0.0515	0.6976	0.0431	0.6863	0.0446	0.6949	0.0517	0.6942
		1.25	0.0489	0.9799	0.0492	0.9731	0.0465	0.9760	0.0456	0.9826	0.0427	0.9751	0.0489	0.9724
	30%	0.75	0.0478	0.6918	0.0470	0.7016	0.0465	0.6989	0.0493	0.6909	0.0466	0.7031	0.0475	0.6962
		1.25	0.0464	0.9817	0.0489	0.9800	0.0504	0.9737	0.0459	0.9835	0.0454	0.9763	0.0479	0.9760
1.0	0%	0.75	0.0453	0.6824	0.0507	0.6953	0.0540	0.6867	0.0446	0.6895	0.0435	0.7088	0.0410	0.6830
		1.25	0.0496	0.9852	0.0544	0.9774	0.0555	0.9691	0.0449	0.9887	0.0482	0.9796	0.0514	0.9707
	10%	0.75	0.0467	0.6862	0.0471	0.6911	0.0534	0.6890	0.0463	0.6854	0.0431	0.7065	0.0399	0.6839
		1.25	0.0456	0.9841	0.0509	0.9806	0.0519	0.9735	0.0459	0.9874	0.0437	0.9831	0.0460	0.9733
	20%	0.75	0.0532	0.6850	0.0464	0.6986	0.0501	0.6983	0.0477	0.6906	0.0448	0.7100	0.0421	0.6878
		1.25	0.0461	0.9816	0.0526	0.9793	0.0518	0.9793	0.0469	0.9860	0.0458	0.9812	0.0460	0.9749
	30%	0.75	0.0548	0.6845	0.0463	0.7087	0.0508	0.7038	0.0532	0.6869	0.0517	0.7066	0.0452	0.6938
		1.25	0.0504	0.9830	0.0493	0.9789	0.0487	0.9765	0.0437	0.9847	0.0467	0.9795	0.0485	0.9765
5.0	0%	0.75	0.0505	0.6812	0.0506	0.7021	0.0547	0.6762	0.0542	0.7053	0.0409	0.7151	0.0451	0.6871
		1.25	0.0499	0.9819	0.0554	0.9813	0.0562	0.9713	0.0488	0.9852	0.0509	0.9882	0.0461	0.9722
	10%	0.75	0.0568	0.6898	0.0470	0.7036	0.0525	0.6908	0.0569	0.6946	0.0424	0.7217	0.0413	0.6849
		1.25	0.0507	0.9816	0.0522	0.9839	0.0515	0.9746	0.0522	0.9842	0.0470	0.9870	0.0452	0.9741
	20%	0.75	0.0579	0.6865	0.0449	0.7093	0.0487	0.6961	0.0593	0.6984	0.0449	0.7153	0.0462	0.6906
		1.25	0.0565	0.9809	0.0502	0.9854	0.0460	0.9741	0.0532	0.9816	0.0453	0.9850	0.0438	0.9762
	30%	0.75	0.0636	0.6857	0.0473	0.7046	0.0520	0.6925	0.0655	0.6997	0.0488	0.7036	0.0439	0.6976
		1.25	0.0580	0.9791	0.0475	0.9830	0.0474	0.9817	0.0559	0.9829	0.0470	0.9859	0.0438	0.9774

Note: In this simulation study, we used nominal Type-I error = 5%, nominal power = 90%, maximum accrual time $\omega = 3$ months, and varying values of administrative censoring time τ (in months), power parameter φ, loss to follow-up rate υ, and true shape parameter κ. Effect size is defined as $\Delta = M_1/M_0$ with $M_0 = 1$ month. Results are based on a total of 10,000 simulations.

2.4.2 Futility Monitoring of Hypothetical Single-Arm Trials

To discuss the implementation of SC methods described in Section 2.3.3, we consider that a hypothetical fixed sample study is designed to obtain preliminary evidence of efficacy of a hypothetical drug X to treat some disease Y, and investigators may conduct unplanned futility testing based on the recommendation of the DSMB overseeing this study. Based on historical evidence, suppose that researchers believe that an improvement in the median progression free survival time (in months) by a factor of 1.5 warrants a further investigation of the experimental treatment in a larger phase III randomized trial. More specifically, suppose that investigators are interested in testing the hypotheses $H_0 : M \leq 1$ month vs. $H_1 : M > 1.5$ months with nominal Type-I error rate and power equal to 5% and 90%, respectively. For this hypothetical example, suppose that researchers anticipate a uniform accrual pattern $(\varphi = 1.00)$ during its accrual phase with the maximum enrollment time equal to 3 months, and the maximum follow-up time of 9 months afterwards. Furthermore, the shape parameter of the Weibull distribution was assumed to be $\kappa = 1.25$ (increasing hazard), and the expected loss to follow-up rate is around 15%.

Using the method described in Section 2.3.2, it was determined that a sample size $n = 40$ is required for a fixed sample design. Consider an interim analysis is requested at 1 month after the conclusion of accrual period (i.e., look time $\ell = 4$ months). To observe how futility testing statistics vary, we generated a sample of size $n = 40$ from each of the Weibull distributions with varying values of the underlying median parameter ($M = 0.75, 1.00, 1.25, 1.50, 2.00$). SC methods were applied to each of the five simulated data sets, and the results are summarized in Table 2.5. We make the following observations:

1. As expected, the number of events observed at the kth interim time point, denoted by $n_{e;k}$, is the largest for the data generated under the assumption of the smallest median survival time (0.75 months), and it decreases as the median survival time increases. The converse holds true for the number of active subjects, $n_{a;k}$, remaining in the study at interim analyses.

2. Using $\zeta = \zeta' = 0.80$ for conditional power calculations, we note that the futility index

corresponding to the data generated assuming the smallest median survival time (0.75 months) is the largest (0.9167). This implies that the observed treatment difference for this simulated data set is no longer consistent with the alternative hypothesis, and the study should be terminated at the first interim analysis. As expected, we observe that the futility index decreases as the median survival time increases. None of the other four cases indicates stopping due to lack of efficacy at the first interim analysis.

3. Predictive power is also observed to be the smallest for the first case with the smallest median survival time, and it increases with an increase in the median survival time. In other words, we observe a direct relationship between the median survival time and the predictive power.

4. We also computed the Bayesian predictive probability of a successful trial outcome at the end of the study using threshold probability $\eta^* = 0.95$. We computed these probabilities using both asymptotic, as well as purely simulation-based algorithms described in Section 2.3.3.3. Consistent with our test hypotheses, we observe that the predictive probability of a positive trial outcome is close to 0 for all four data sets generated with the underlying median time $M \leq 1.5$ months. The predictive probability of success for data generated using $M = 2.00$ months is well above 50% in both cases, and investigators may either conclude efficacy at this stage, or decide to continue the trial according to the recommendation of DSMB.

5. We observe that the magnitude of the predictive power and the Bayesian predictive probability computed through purely simulation-based approach using `R2OpenBUGS` in `R` tends to be larger than those computed using our proposed asymptotic method. This is due to the fact that the (interim) posterior densities generated using the purely simulation-based approach tend to have heavier tails with an increase in the median survival time, which in turn contributes to a more pronounced difference in the magnitude of these SC tests.

Table 2.5: Futility monitoring statistics calculated at look time $\ell = 4$ months for hypothetical data sets simulated using varying values of the median survival time for a study to test the hypotheses $H_0 : M \leq 1.00$ month vs. $H_1 : M > 1.50$ months

Median	Obs. Status			Statistics				Conditional Power			Predictive Power		Bayesian Pred. Prob.	
	$n_{e;k}$	$n_{c;k}$	$n_{a;k}$	$\widehat{\gamma}_k$	$\widehat{M}_k$	$\text{Var}(\widehat{\gamma}_k)$	$Z_{obs;k}$	$P_k(\gamma_0)$	$P_k(\gamma_1)$	Fut. Index	Asymptotic	OpenBUGS	Asymptotic	OpenBUGS
0.75	36	3	1	−0.0621	0.7009	0.0199	−2.5199	7.3426×10^{-5}	0.0832	0.9167	1.637×10^{-6}	0.0030	0.0000	0.0000
1.00	34	3	3	0.2398	0.9479	0.0212	−0.3670	0.0115	0.5552	0.4448	0.0219	0.2946	0.0000	0.0000
1.25	33	2	5	0.4426	1.1611	0.0225	0.9939	0.0948	0.8649	0.1354	0.3626	0.8417	6.063×10^{-14}	4.936×10^{-8}
1.50	28	2	10	0.6880	1.4841	0.0251	2.4907	0.3999	0.9845	0.0155	0.9435	0.9981	0.0046	0.0072
2.00	24	2	14	0.9633	1.9545	0.0294	3.9069	0.7727	0.9992	0.0008	0.9998	1.0000	0.6912	0.7371

Note: For this study, we assume uniform accrual pattern ($\varphi = 1.00$), maximum accrual time $\omega = 3$ months, administrative censoring time $\tau = 12$ months, common shape parameter $\kappa = 1.25$, and loss to follow-up rate $\upsilon = 0.15$. A total of $n = 40$ subjects were enrolled during the accrual phase.

Table 2.6: Futility monitoring statistics calculated at look time $\ell = 6$ months for hypothetical data sets simulated using varying values of the median survival time for a study to test the hypotheses $H_0 : M \leq 2.50$ month vs. $H_1 : M > 3.75$ months

Median	Obs. Status			Statistics				Conditional Power			Predictive Power		Bayesian Pred. Prob.	
	$n_{e;k}$	$n_{c;k}$	$n_{a;k}$	$\widehat{\gamma}_k$	$\widehat{M}_k$	$\text{Var}(\widehat{\gamma}_k)$	$Z_{obs;k}$	$P_k(\gamma_0)$	$P_k(\gamma_1)$	Fut. Index	Asymptotic	OpenBUGS	Asymptotic	OpenBUGS
2.00	16	3	2	0.9574	2.0403	0.0276	−1.2229	0.0002	0.0395	0.9605	1.115×10^{-4}	0.0943	0.0000	0.0000
2.50	14	3	4	1.2142	2.6375	0.0310	0.3041	0.0216	0.4094	0.5906	0.0687	0.6244	0.0000	0.0000
3.25	13	3	5	1.3965	3.1649	0.0348	1.2642	0.1441	0.7676	0.2324	0.4840	0.9303	1.601×10^{-5}	2.247×10^{-4}
3.75	11	2	8	1.5648	3.7450	0.0397	2.0288	0.3831	0.9326	0.0674	0.8820	0.9946	0.0318	0.0639
4.50	9	1	11	1.7541	4.5256	0.0472	2.7329	0.6579	0.9861	0.0139	0.9928	0.9999	0.4718	0.5178

Note: For this study, we assume a non-uniform accrual pattern ($\varphi = 1.25$), maximum accrual time $\omega = 3$ months, administrative censoring time $\tau = 12$ months, common shape parameter $\kappa = 1.50$, and loss to follow-up rate $\upsilon = 0.15$. A total of $n = 21$ subjects were enrolled during the accrual phase.

We consider another hypothetical study in which investigators are interested in testing the hypotheses $H_0 : M \leq 2.50$ months vs. $H_1 : M > 3.75$ months using a fixed sample design with nominal Type-I error rate 5%, and power 80%. We assume a non-uniform accrual pattern ($\varphi = 1.25$) during the accrual phase spanning 3 months, and a follow-up period of 9 months. The common shape parameter is assumed to be $\kappa = 1.50$, and the proportion of censoring due to loss to follow-up is around 15%. In this case, we found that a sample size of $n = 21$ is needed to conduct a fixed sample study. Suppose an interim analysis is requested for testing futility after 3 months following the accrual phase (i.e, look time $\ell = 6$ months). In this case, we generated a survival data of size $n = 21$ from each of the Weibull distributions with the underlying median parameter ($M = 2.00, 2.50, 3.25, 3.75, 4.00$), and the corresponding results are presented in Table 2.6. Comparing to our previous example, we note similar patterns in the number of events, futility index, predictive power, and Bayesian predictive probability. While there is an agreement in the conclusions based on the conditional power and the predictive power tests, we note that the Bayesian predictive probability is large enough to warrant continuation of the trial (or to conclude efficacy) only for the data simulated using the largest median parameter of the Weibull distribution.

2.4.3 Group Sequential Design Plans for a Hypothetical Single-Arm Trial

During public health emergencies, it may seem reasonable to construct sequential design plans to expedite the process of drug testing and evaluation in early phase clinical trials. Suppose we are interested in constructing GSD plans for the first example discussed in Section 2.4.2 for obtaining preliminary evidence of efficacy and futility of a hypothetical drug X to treat some disease Y. We compare and contrast various aspects of Pocock and O'Brien-Fleming GSD plans using 10,000 runs of the simulation-based algorithm outlined in Section A.3 of Appendix A. Throughout this section, we used Hwang-Shih-DeCani error spending function [34] to obtain approximations of α- and β- spending functions for Pocock and O'Brien-Fleming plans.

First, we present a GSD plan to evaluate both efficacy and futility at three equally-spaced look times, and the corresponding results are reported in Table 2.7. We can clearly observe that

the efficacy boundaries $\mathscr{U}_k$ (presented on the test statistic scale) for the Pocock plan (1.876) is much smaller than that for the O'Brien-Fleming plan (2.651) at the first interim analysis because it allocates a much larger portion of the Type-I error at the first look. We observe a similar pattern in the futility boundaries $\mathscr{L}_k$ for both plans (0.936 for Pocock plan vs. -0.096 for O'Brien-Fleming plan). As a consequence, the stopping probabilities are much higher for the Pocock plan than the respective probabilities for the O'Brien-Fleming design at the first look, and the converse is true at the last look. If the study continues to its pre-planned end date, we observe that the average number of events needed to detect a difference using the Pocock plan are larger than those for the O'Brien-Fleming plan (37.7662 for Pocock plan vs. 34.9959 for O'Brien-Fleming plan for futility testing). Consistent with the literature, we also note that the Pocock plan requires a relatively smaller number of expected events under the alternative hypothesis. This difference in the number of required events may not seem apparent due to relatively small overall sample sizes needed for both plans.

Second, we present equally-spaced GSD plans for evaluating efficacy and futility (with 1 skip) in Table 2.8. In comparison to designs provided in Table 2.7, we note that the probability of stopping under the null hypothesis during the first interim analysis drops considerably for both plans (0.0702 for Pocock plan vs. 0.0504 for O'Brien-Fleming plan). At the second look time, however, we note that these stopping probabilities are very high as the efficacy and futility boundaries are much closer. In these plans, we may clearly observe that the expected number of events under the alternative hypothesis using the Pocock plan is lower than that for the O'Brien-Fleming design.

Third, we present results corresponding to unequally-spaced efficacy-only GSD plans. The efficacy stopping boundaries in this design are similar to the results for the design with 1-futility skip presented in Table 2.8. We note that the stopping probability under H_0 at any particular look time is essentially equal to the local Type-I error rate spent at that particular look. We observe a considerable difference in the magnitude of stopping probabilities under the alternative hypothesis in Tables 2.8 and 2.9. Furthermore, the cumulative stopping probability under the alternative hypothesis is slightly larger than the desired power.

Table 2.7: Comparison of Pocock and O'Brien-Fleming GSD plans with three equally-spaced looks for evaluating both efficacy and futility in a hypothetical study to test the hypotheses $H_0 : M \leq 1.00$ month vs. $H_1 : M > 1.50$ months

Type	k	ℓ_k	Efficacy						Futility					
			$\bar{n}_{e;H_0}$	α_k	$\sum_k \alpha_k$	$\mathscr{U}_k$	$p_{k;H_0}$	$\sum_k p_{k;H_0}$	$\bar{n}_{e;H_1}$	β_k	$\sum_k \beta_k$	$\mathscr{L}_k$	$p_{k;H_1}$	$\sum_k p_{k;H_1}$
	1	4	31.9950	0.0224	0.0224	1.876	0.8532	0.8532	27.2450	0.0406	0.0406	0.936	0.8247	0.8247
Pocock	2	8	35.3529	0.0161	0.0385	1.792	0.1116	0.9648	37.1768	0.0291	0.0698	1.408	0.1321	0.9568
$(n = 43)$	3	12	35.0085	0.0115	0.0500	1.626	0.0335	0.9983	37.7662	0.0209	0.0906	1.626	0.0421	0.9989
			$E(n_{e;H_0}) = 32.4758$						$E(n_{e;H_1}) = 29.0115$					
O'Brien &	1	4	29.7752	0.0026	0.0026	2.651	0.4732	0.4732	25.3258	0.0052	0.0052	−0.096	0.4438	0.4438
Fleming	2	8	33.4332	0.0099	0.0125	2.150	0.3464	0.8196	34.2722	0.0197	0.0249	0.832	0.3596	0.8034
$(n = 40)$	3	12	32.8010	0.0375	0.0500	1.579	0.1804	1.0000	34.9959	0.0748	0.0997	1.579	0.1966	1.0000
			$E(n_{e;H_0}) = 31.5882$						$E(n_{e;H_1}) = 30.4441$					

Note: Using nominal α = 5%, power = 90%, uniform accrual pattern ($\varphi = 1$), maximum accrual time $\omega = 3$ months, administrative censoring time $\tau = 12$ months, common shape parameter $\kappa = 1.25$, and loss to follow-up rate $\upsilon = 0.15$, these GSD plans were generated on the basis of 10,000 simulations.

Table 2.8: Comparison of Pocock and O'Brien-Fleming GSD plans with three equally-spaced looks for evaluating efficacy and futility (with 1 skip) in a hypothetical study to test the hypotheses $H_0 : M \leq 1.00$ month vs. $H_1 : M > 1.50$ months

Type	k	ℓ_k	Efficacy						Futility					
			$\bar{n}_{e;H_0}$	α_k	$\sum_k \alpha_k$	$\mathscr{U}_k$	$p_{k;H_0}$	$\sum_k p_{k;H_0}$	$\bar{n}_{e;H_1}$	β_k	$\sum_k \beta_k$	$\mathscr{L}_k$	$p_{k;H_1}$	$\sum_k p_{k;H_1}$
	1	4	29.7752	0.0224	0.0224	1.914	0.0702	0.0702	25.3258	–	–	–	0.7507	0.7507
Pocock	2	8	33.9777	0.0161	0.0385	1.757	0.8960	0.9662	34.6611	0.0767	0.0767	1.436	0.2025	0.9532
$(n = 40)$	3	12	32.5178	0.0115	0.0500	1.626	0.0325	0.9987	34.9145	0.0230	0.0997	1.626	0.0439	0.9971
			$E(n_{e;H_0}) = 32.5765$						$E(n_{e;H_1}) = 27.6649$					
O'Brien &	1	4	29.7752	0.0026	0.0026	2.651	0.0504	0.0504	25.3258	–	–	–	0.4386	0.4386
Fleming	2	8	33.9327	0.0099	0.0125	2.150	0.7716	0.8220	34.2804	0.0249	0.0249	0.839	0.3596	0.8034
$(n = 40)$	3	12	32.7955	0.0375	0.0500	1.580	0.1780	1.0000	34.9919	0.0748	0.0997	1.580	0.1966	1.0000
			$E(n_{e;H_0}) = 33.5207$						$E(n_{e;H_1}) = 30.3145$					

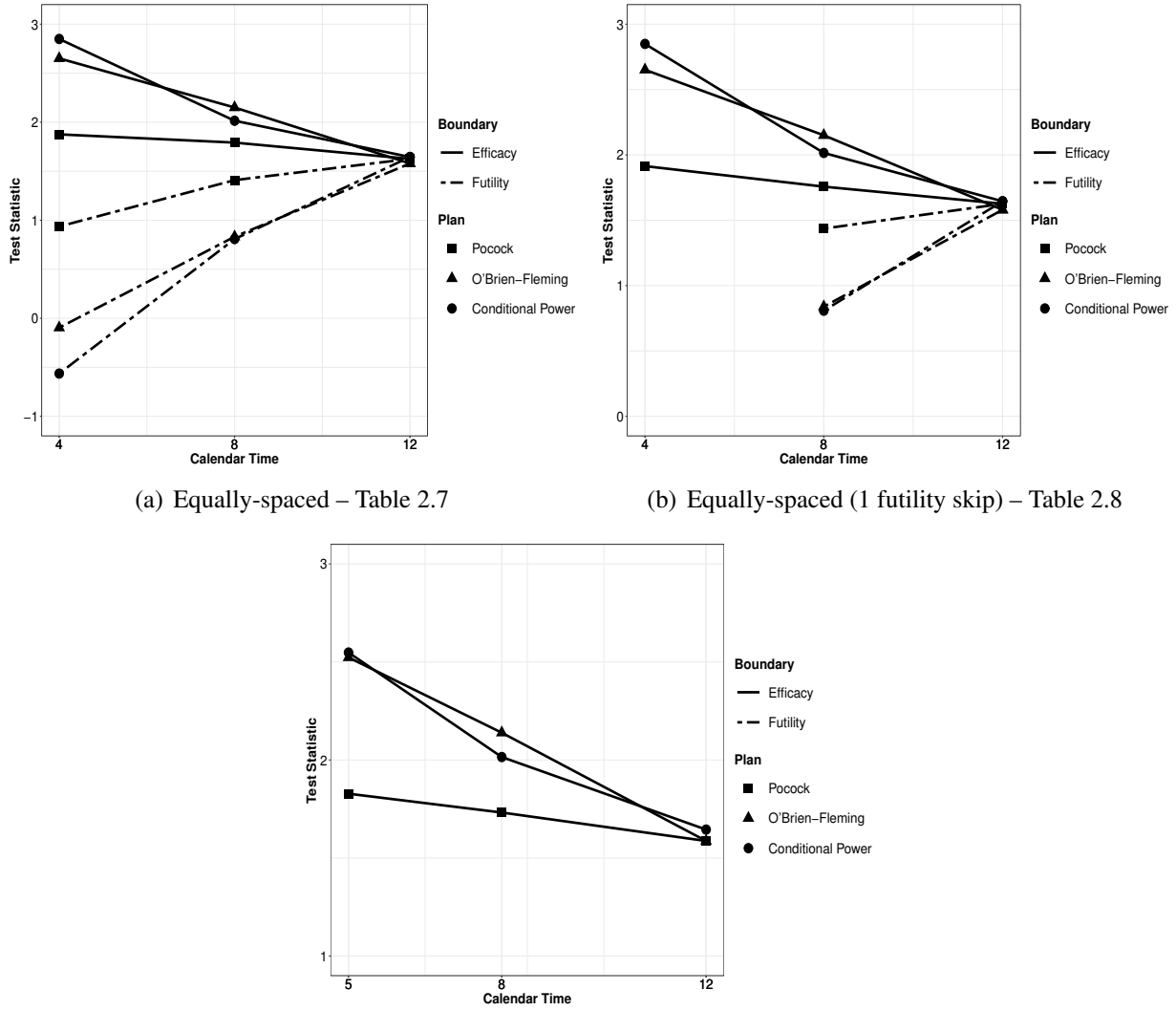
Note: Using nominal α = 5%, power = 90%, uniform accrual pattern ($\varphi = 1$), maximum accrual time $\omega = 3$ months, administrative censoring time $\tau = 12$ months, common shape parameter $\kappa = 1.25$, and loss to follow-up rate $\upsilon = 0.15$, these GSD plans were generated on the basis of 10,000 simulations.

Table 2.9: Comparison of Pocock and O'Brien-Fleming GSD plans for evaluating efficacy at three unequally-spaced look times in a hypothetical study to test the hypotheses $H_0 : M \leq 1.00$ month vs. $H_1 : M > 1.50$ months

Type	k	ℓ_k	Efficacy						Futility					
			$\bar{n}_{e;H_0}$	α_k	$\sum_k \alpha_k$	$\mathscr{U}_k$	$p_{k;H_0}$	$\sum_k p_{k;H_0}$	$\bar{n}_{e;H_1}$	β_k	$\sum_k \beta_k$	$\mathscr{L}_k$	$p_{k;H_1}$	$\sum_k p_{k;H_1}$
	1	5	33.5831	0.0270	0.0270	1.828	0.0270	0.0270	30.7088	–	–	–	0.8374	0.8374
Pocock	2	8	34.9208	0.0115	0.0385	1.732	0.0115	0.0385	35.8542	–	–	–	0.0433	0.8807
$(n = 41)$	3	12	34.9592	0.0115	0.0500	1.587	0.0115	0.0500	36.1618	–	–	–	0.0216	0.9023
			$E(n_{e;H_0}) = 34.5196$						$E(n_{e;H_1}) = 31.5821$					
O'Brien &	1	5	33.5831	0.0040	0.0040	2.523	0.0040	0.0040	30.7088	–	–	–	0.6122	0.6122
Fleming	2	8	34.8619	0.0085	0.0125	2.139	0.0085	0.0125	35.4464	–	–	–	0.1912	0.8034
$(n = 41)$	3	12	34.9054	0.0375	0.0500	1.587	0.0500	0.0500	35.9359	–	–	–	0.0979	0.9013
			$E(n_{e;H_0}) = 33.5908$						$E(n_{e;H_1}) = 32.6423$					

Note: Using nominal α = 5%, power = 90%, uniform accrual pattern ($\varphi = 1$), maximum accrual time $\omega = 3$ months, administrative censoring time $\tau = 12$ months, common shape parameter $\kappa = 1.25$, and loss to follow-up rate $\upsilon = 0.15$, these GSD plans were generated on the basis of 10,000 simulations.

Pictorial comparison of stopping boundaries for the Pocock and O'Brien-Fleming designs for all three scenarios is shown in Figure 2.1. We have also shown efficacy and futility boundaries obtained using $\zeta = \zeta' = 0.5$ in the conditional power stopping boundaries in Eqs. (2.10) and (2.11). Consistent with Davis and Hardy [43], we observe that the conditional power boundaries with $\zeta = \zeta' = 0.5$ closely resemble the O'Brien-Fleming boundaries. Since the parameters ζ and ζ' are typically chosen around 0.80, the O'Brien-Fleming testing procedure is more likely to trigger an early stopping of a clinical trial in comparison to the approach based on conditional power [18].

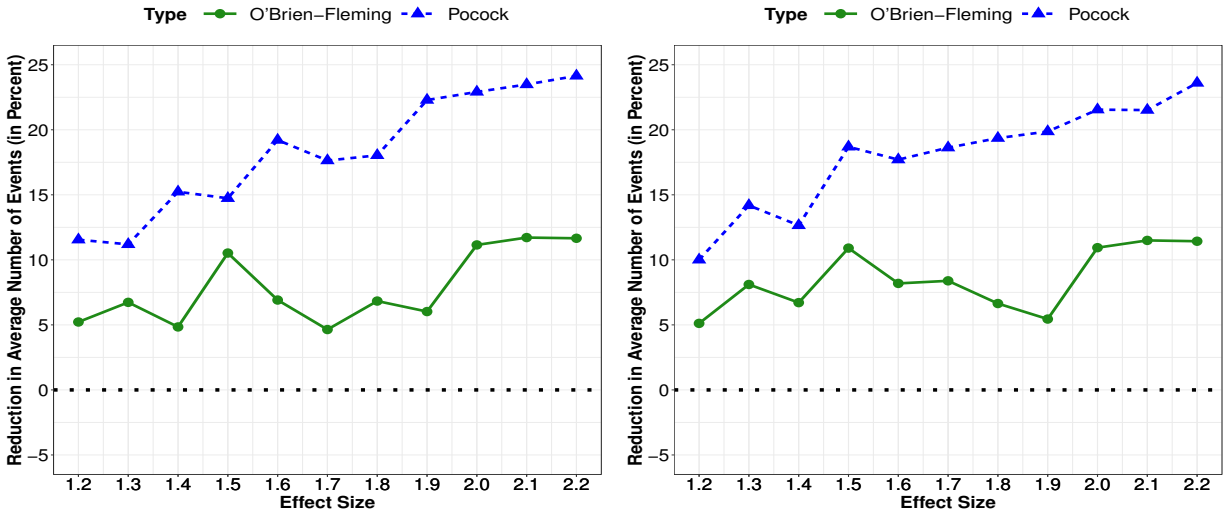


(a) Equally-spaced – Table 2.7

(b) Equally-spaced (1 futility skip) – Table 2.8

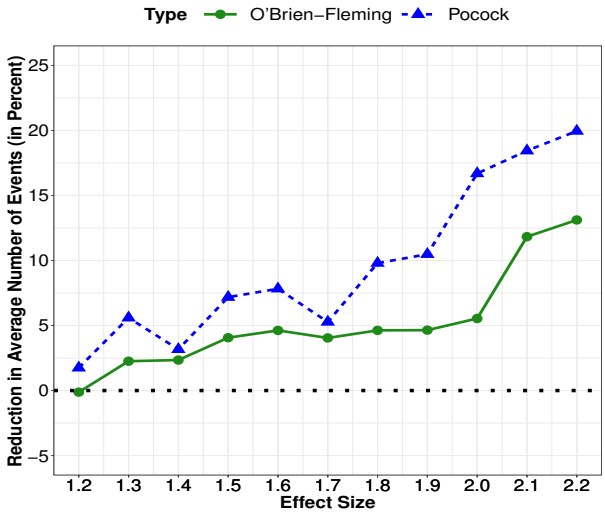
(c) Unequally-spaced efficacy-only – Table 2.9

Figure 2.1: Stopping boundaries for the GSD plans in Tables 2.7–2.9



(a) Equally-spaced – Table 2.7

(b) Equally-spaced (1 futility skip) – Table 2.8



(c) Unequally-spaced efficacy-only – Table 2.9

Figure 2.2: Percent reduction in the average number of events for the Pocock and O'Brien-Fleming GSD plans, in comparison to a fixed sample design, in Tables 2.7–2.9

To illustrate usefulness of the RST approach, we compared the relative benefit of the group sequential designs to that of the fixed designs in terms of the percent reduction in the expected (average) number of events. For each of the three examples discussed in this section, we plotted the percent reduction in the expected (average) number of events needed for the Pocock and O'Brien-Fleming plans against varying values of the effect size in Figure 2.2. It can be observed that the

percent reduction in the average number of events needed under the Pocock and O'Brien-Fleming plans relative to the fixed design increases with an increase in the desired effect size. In addition, we note that the Pocock plan offers greater percent reduction in the average number of events in comparison to the O'Brien-Fleming plan. This is primarily due to the fact that the Pocock's design allocates relatively larger magnitudes of the Type-I and Type-II error rates at the earlier interim analyses, which in turn may trigger early stopping of a clinical trial with a smaller number of events.

2.5 Discussion

In this manuscript, we attempted to address some outstanding issues pertaining to the design of single-arm phase II studies with TTE primary endpoints using Weibull distribution. First, we presented a parametric Wald's test statistic for designing fixed-sample single-arm studies which can incorporate various flexible options including accrual patterns and expected loss to follow-up rate. Our proposed method is able to maintain the desired Type-I error and power in most cases, and it was demonstrated through extensive simulations that the proposed asymptotic method performs the best when we are interested in designing single-arm phase II studies with moderate to large sample sizes. The methods proposed by Phadnis [6] and Wu [10] appear to be reasonable alternatives when the affordable sample sizes are small, or exact calculations are desired.

Second, we discussed the application of three SC methods (conditional power, predictive power, Bayesian predictive probability) for futility monitoring. To our knowledge, these methods have not been studied previously in the context of single-arm studies with TTE primary endpoints, and therefore our work on these methods offers additional tools to aid decision making during interim analyses in such studies with TTE outcomes.

Third, we discussed the construction of GSD plans using the RST approach for evaluating efficacy and futility for single-arm phase II studies in the TTE framework. We also outlined a simulation-based approach which follows a similar algorithm as the one presented by Phadnis and Mayo [29], and seems reasonable to implement especially when designing single-arm studies with

moderate to large sample sizes. While the stopping boundaries can be constructed in terms of the test statistic in Section 2.3.2, this algorithm also permits us to obtain results in terms of median survival time which may be more relevant to clinical researchers. In addition, it allows us to incorporate flexible options such as futility skips, number of looks, and error spending function, among others.

In this manuscript, we restricted our attention towards the construction of GSD plans when the parametric Weibul model is appropriate for modeling TTE data in an early phase study. If such assumption does not hold true, non-parametric methods based on increments of the log-rank test statistics can be alternatively utilized to construct such study designs. For instance, we can utilize the adaptive one-sample log-rank test proposed by Schmidt et al. [44] to compare the survival curve of the patients under treatment to some pre-specified reference survival curve. They demonstrated that their proposed method protects the study power and reduces the average sample number under the null hypothesis in comparison to a fixed sample design.

Sequential designs allow researchers to perform interim analyses to obtain overwhelming evidence of efficacy or futility of an experimental treatment that warrants early termination before the pre-planned end date without compromising its operating characteristics [17, 18, 28]. As emphasized by Dmitreinko et al. [18], there are two key distinctions between the SC and RST methods: (1) the RST approach depends on the pre-specified error spending function dictating the characteristic of sequential tests and the subsequent decision-making, whereas the SC methods are tied to final outcome of the trial in the sense that the decision to 'curtail' sampling depends of likelihood of positive or negative outcome if trial continues to the planned end date; and (2) the SC methods are "aimed toward predictive inferences, whereas repeated significance tests focus on currently available data." Due to safety or ethical concerns, it is not uncommon to implement the SC methods in fixed trial designs, upon the recommendation of DSMB, in a post-hoc manner without explicitly adjusting for repeated testing [18].

Throughout this manuscript, it is assumed that a reasonable estimate of the shape parameter of the Weibull distribution is known from historical studies. It is quite possible to encounter real-life

situations where sufficient amount of historical data is unavailable (for instance, in studies involving rare diseases, or lack of standard treatments for treating certain conditions), and thus hindering our ability to obtain a reliable estimate. Since the shape parameter of the Weibull model determines the shape of the underlying hazard function, it is critical to incorporate the best available estimate into our sample size calculations. If no reliable estimate of the shape parameter is available, we suggest designing the study assuming exponential survival times (i.e. shape parameter $\kappa = 1$). At the interim stage, we can obtain an estimate of the nuisance shape parameter based on the accumulated data, and then recalibrate our study design accordingly. Further investigation is needed to study necessary adjustments related to the methods discussed in this manuscript.

As we realize, knowledge regarding the posterior distribution of a parameter of interest is a prerequisite to performing the predictive power or the Bayesian predictive probability calculations for futility monitoring. To generate the posterior distribution, however, we also need to elicit appropriate prior information in the context of a given problem. In the literature, there does not appear a wider consensus on the appropriate choices of prior distributions [18, 25, 45]. Ideally, clinicians and statisticians should work together to identify appropriate priors for the relevant parameters in the model. In situations where it is not possible, statisticians may consider doing a sensitivity analysis by calculating the predictive power and the Bayesian predicted probability under different choices of the prior distribution.

In this manuscript, simulation-based approaches have been proposed to construct GSD plans, and to perform calculations related to predictive power/probability for futility monitoring. It is recommended to perform a sufficiently large number of simulations to minimize random noise in the reported results. Although we have presented our results up to four significant digits, a more formal investigation shall be conducted in the future to determine the number of simulations needed to report results with such a degree of precision.

Chapter 3

Bayesian Predictive Probability for Single-Arm Clinical Trials with a Time-to-Event Endpoint using Weibull Distribution with Unknown Shape Parameter

Abstract

Bayesian predictive probability is an important method for conducting interim analyses to obtain preliminary evidence of efficacy or futility of an experimental treatment warranting early termination of a clinical trial. In the context of single-arm clinical trials with time-to-event endpoints following Weibull distribution, we discuss the calculation of the Bayesian predictive probability when the shape parameter of the Weibull distribution is unknown. Based on the data accumulated at an interim stage, we propose two approaches which rely on the posterior mode or the entire posterior distribution of the shape parameter. To account for uncertainty in the shape parameter, it is recommended to incorporate its entire posterior distribution in our calculations. In this manuscript, we also explore the utility of internal pilot study (IPS) approach for reestimating sample size at the interim. Although IPS approach can help rescue the study power, it is noted that the adjusted sample size can be as much as twice the initially planned sample size, which may put substantial practical constraints to continue the study.

3.1 Introduction

Single-arm clinical trails are often carried out in the early phases of oncology drug development to evaluate safety and to obtain preliminary evidence of therapeutic effect of new cancer treatments [3,

5]. In such trials, tumor response rate (TRR) or objective response rate (ORR) has popularly been used as the primary endpoint to identify any potential of biological drug activity that is assessed in terms of tumor shrinkage [5, 46]. As noted by Rubinstein [5], many phase II clinical trials are now being designed to assess the promise of molecularly targeted agents which may not necessarily improve TRR or ORR, but instead yield an improvement in other time-to-event (TTE) endpoints such as progression-free survival (PFS) or overall survival (OS). This manuscript deals with some unaddressed planning aspects concerning single-arm phase II clinical trials with TTE endpoints.

A limited number of options based on the log-rank test and its weighted versions are available in the literature for designing single-arm clinical trials with TTE endpoints. Some of the existing approaches include the ones proposed by Finkelstein et al. [7], Kwak and Jung [8], Sun et al. [9], Wu [10], and Phadnis [6]. Among these approaches, the method proposed by Phadnis [6] is appropriate when the subject survival times are assumed to follow the Weibull distribution, and it can be used for calculating the required sample size while adjusting for administrative censoring along with an ad-hoc inflation for random loss to follow-up. Most recently, Waleed et al. [47] proposed a parametric maximum likelihood estimate (MLE) test, based on the asymptotic approximation of the scale parameter of the Weibull distribution, whose variance component can account for the expected loss to follow-up rate and different accrual patterns (early, late, or uniform accrual). It is worth mentioning that both methods (Phadnis [6], Waleed et al. [47]) assume that a reliable estimate of the shape parameter of the Weibull distribution is known from historical studies.

Due to ethical and practical considerations, single-arm oncology trials are often conducted via Simon's two-stage approach which allows researchers to obtain early evidence of futility of an experimental treatment, and consequently terminating the study in consultation with the Data Safety Monitoring Board (DSMB) overseeing the clinical trial [15, 17, 16]. Alternatively, stochastic curtailment (SC) methods can be employed to decide whether to continue or 'curtail' sampling beyond an interim analysis based on the likelihood of a positive or negative outcome if the trial were to continue to its pre-planned end [17, 16, 18]. Conditional power (Lan, Simon and Halprin [19], Andersen [20]), predictive power (Spiegelhalter et al. [21]), and Bayesian pre-

dictive probability (Herson [22], Geisser [24], Dmitrienko and Wang [25]) are the most popularly used SC methods. These methods have been well-studied for normal and binary endpoints, and implemented in various statistical software including `R` [26] and `SAS` [27]. Very recently, Waleed et al. [47] studied these SC methods in the context of single-arm oncology trials with TTE primary endpoints. More specifically, they presented mathematical development of these methods when the parametric Weibull model is appropriate for modeling survival data derived from such studies, and different censoring mechanisms and accrual patterns are under consideration. A limitation of the work by Waleed et al. [47] is that a reliable estimate of the shape parameter of the Weibull distribution is assumed to be known, which may not hold true such as in the case of studies related to rare diseases.

When reliable estimates of any nuisance parameters, such as the shape parameter of the Weibull distribution for modeling survival data, are unavailable during the planning phase of a study, adaptations to the sample size can be incorporated using the estimates of nuisance parameters obtained using the data accumulated at an interim stage [48, 49, 50]. Besides other advantages, such adaptive features in the study design enhance statistical efficiency of a clinical trial [48]. In this manuscript, we aim to build upon the framework developed by Waleed et al. [47] by considering the scenario when adequate historical data is not available to obtain a reasonably accurate estimate of the shape parameter. More specifically, the objective of this manuscript is two-fold: first, we discuss adaptation to the sample size for single-arm phase II trials with TTE endpoints via implementation of the internal pilot study (IPS) approach proposed by Wittes and Brittain [49] and, secondly, we present calculation of the Bayesian predictive probability (BPP) for efficacy or futility testing based on the data accumulated at the interim.

This manuscript is organized in the following order. After presenting a brief review of the fixed sample design of Waleed et al. [47] in Section 3.2, we discuss sample size reestimation at a prospectively planned interim stage, and calculation of the Bayesian predictive probability in single-arm phase II clinical trials with TTE endpoints following the Weibull distribution with unknown shape parameter. We present some simulation studies and examples to demonstrate the

proposed approaches in Section 3.3. Finally, in Section 3.4, we present a discussion on the contents presented in this manuscript.

3.2 Methods

3.2.1 Notation and Preliminaries

Suppose that a total of n subjects are accrued during the enrollment period of a single-arm phase II clinical trial with a TTE endpoint. Due to practical constraints, administrative censoring is incorporated at a pre-specified calendar time τ, when all active subjects in the study are censored and the resulting data are analyzed. For the ith subject, suppose E_i denotes its calendar time of accrual into the study; Y_i denotes the amount of time from E_i to the calendar time of event; C_i denotes the amount of time from E_i to the time of loss to follow-up, and $Z_i := \min(\max(0, \tau - E_i), C_i)$ denotes the amount of time to being lost to follow-up or administrative censoring. We assume that the loss to follow-up is unrelated to the event of interest, that is, non-informative of the survival process, and $\{Y_i, Z_i, i = 1, \ldots, n\}$ are independent and identically distributed. In summary, we have n pairs of data $\{(X_i, \delta_i), i = 1, \ldots, n\}$, where $X_i := \min(Y_i, Z_i)$ is the subject's survival time, and $\delta_i := \mathbb{1}_{(Y_i < Z_i)}$ is the corresponding survival status which equals 1 if $Y_i < Z_i$ and 0 otherwise.

The event time Y_i is assumed to follow the Weibull distribution having shape parameter κ and scale parameter θ with the probability density function (pdf) expressed as below:

$$f_{Y_i}(y) = \frac{\kappa}{\theta^\kappa} y^{\kappa-1} \exp\left\{-\left(\frac{y}{\theta}\right)^\kappa\right\}, \qquad \text{where: } y > 0, \kappa > 0, \theta > 0. \tag{3.1}$$

The Weibull distribution is flexible in the sense that it allows us to handle different shapes of the underlying hazard function. More specifically, the hazard function is constant, increasing or decreasing when the shape parameter κ is equal to, greater, or less than 1, respectively [36]. In their proposed method, Waleed et al. [47] assume that a reasonably accurate estimate of the shape parameter κ of the Weibull(κ, θ) distribution is known from historical studies. The random loss to follow-up time C_i also follows the Weibull distribution having the same shape parameter κ and

scale parameter η. To accommodate anticipated loss to follow-up rate υ, it can be conveniently verified, following Wan [38], that $\eta = \theta\left(\frac{1-\upsilon}{\upsilon}\right)^{1/\kappa}$ ensures the loss to follow-up rate υ.

Suppose that ω represents the maximum calendar time of accrual into the study. The accrual time E_i is assumed to follow a rather general form of uniform distribution, with an additional power parameter φ, having the following pdf (at a realized value e of the random variable E_i):

$$f_{E_i}(e) = \frac{\varphi e^{\varphi-1}}{\omega^{\varphi}}, \qquad \text{where: } e \in [0,\omega],\ \varphi > 0. \tag{3.2}$$

In addition to incorporating uniform accrual pattern with $\varphi = 1$, the above choice of accrual distribution is flexible in the sense that it enables us to incorporate very early (late) accrual patterns by choosing φ that is very small (large) in magnitude.

Under the aforementioned assumptions, it can be conveniently verified that the pdf of $Z_i = \min(\max(0,\tau - E_i), C_i)$ having four parameters $\kappa > 0, \eta > 0, \omega > 0$, and $\varphi > 0$ is given as:

$$f_{Z_i}(z) = \begin{cases} \frac{\kappa}{\eta^{\kappa}} z^{\kappa-1} \exp\left\{-\left(\frac{z}{\eta}\right)^{\kappa}\right\} & \text{if } z \in [0, \tau-\omega) \\ \left(\frac{\varphi}{\tau - z} + \frac{\kappa}{\eta^{\kappa}} z^{\kappa-1}\right)\left(\frac{\tau - z}{\omega}\right)^{\varphi} \exp\left\{-\left(\frac{z}{\eta}\right)^{\kappa}\right\} & \text{if } z \in [\tau-\omega, \tau] \\ 0 & \text{otherwise} \end{cases} \tag{3.3}$$

3.2.2 Fixed Sample Design

Before discussing sample size reestimation using the IPS approach, we present a brief overview of the fixed sample design, proposed by Waleed et al. [47], for designing single-arm phase-II clinical trials with TTE endpoints. Since covariates are commonly introduced [36] into the parametric survival models through the scale parameter as $\theta = \exp\{\boldsymbol{\gamma}^T \boldsymbol{x}\}$, where: $\boldsymbol{x} = (1, x_1, \ldots, x_k)^T$ and $\boldsymbol{\gamma}^T = (\gamma_0, \gamma_1, \ldots, \gamma_k)$ are the vectors of $k+1$ covariates and the corresponding parameters, respectively, we can define the two alternatives as: $H_0 : \theta \leq \theta_0$ versus $H_1 : \theta > \theta_0$. When no covariates other than the experimental treatment administered to the subjects are introduced into the model, the scale parameter can be expressed as $\theta = \exp\{\gamma\}$. Thus, our hypotheses can be equivalently

expressed as:

$$\begin{aligned} H_0 &: \gamma \leq \gamma_0 \\ H_1 &: \gamma > \gamma_0 \end{aligned} \tag{3.4}$$

It is straightforward to verify that the MLE of γ, denoted by $\widehat{\gamma}$, is given as:

$$\widehat{\gamma} = \frac{1}{\kappa} \log\left(\frac{\overline{X^{\kappa}}}{\overline{\delta}}\right), \qquad \text{where: } \overline{X^{\kappa}} = \frac{1}{n}\sum_{i=1}^{n} X_i^{\kappa} \text{ and } \overline{\delta} = \frac{1}{n}\sum_{i=1}^{n} \delta_i. \tag{3.5}$$

Since it appears analytically intractable to obtain the exact distribution of $\widehat{\gamma}$ due to the underlying correlation between a subject's survival time and the corresponding survival status, Waleed et al. [47] relied on asymptotic calculations to construct a parametric MLE-based statistic for testing the hypotheses in Eq. (3.4). Without loss of generality, they showed that

$$\lim_{n\to\infty} \sqrt{n}\,(\widehat{\gamma} - \gamma) \xrightarrow{d} \text{Normal}\left(0, \sigma^2 = \frac{1}{\kappa^2 \mu_{\bar{\delta}}}\right), \tag{3.6}$$

where $\mu_{\bar{\delta}} = 1 - E_{Z_1}\left(\exp\left\{-\left(\frac{Z_1}{\theta}\right)^{\kappa}\right\}\right)$, $\theta = \exp\{\gamma\}$, and $Z_1 = \min(\tau - E_1, C_1)$.

Let $\widehat{\sigma}^2$ denote the MLE plugged-in estimate of σ^2. Under H_0, the Wald's test statistic is

$$Z_{stat} = \frac{\widehat{\gamma} - \gamma_0}{\widehat{\sigma}/\sqrt{n}} \;\dot{\sim}\; \text{Normal}(0,1). \tag{3.7}$$

For a given Type-I error rate α, we reject the null hypothesis when the observed test statistic $\widehat{Z}_{stat} > Z_{1-\alpha}$, where $Z_{1-\alpha} = \Phi^{-1}(1-\alpha)$ represents the upper α-quantile of the standard normal distribution.

For sample size calculations, researchers specify a clinically meaningful difference $\varepsilon > 0$ that they are interested in detecting under the alternative hypothesis $\gamma_1 = \gamma_0 + \varepsilon$. The required sample size to detect the difference ε using the Wald's test statistic in Eq. (3.7) with Type-I error α and power $1-\beta$ satisfies: $n = \sigma_1^2 \cdot \left(\frac{Z_{1-\beta}+Z_{1-\alpha}}{\varepsilon}\right)^2$, where $\sigma_1^2 \equiv \sigma^2(\gamma_1)$ is the plug-in estimator of σ^2

under H_1, $Z_{1-\beta} = \Phi^{-1}(1-\beta)$, $Z_{1-\alpha} = \Phi^{-1}(1-\alpha)$, and $\Phi^{-1}(\cdot)$ denotes the inverse cumulative distribution function (cdf) of the standard normal distribution. To compute the required sample size, numerical integration can be used to calculate σ_1^2.

3.2.3 Sample Size Reestimation

A limitation of the fixed sample design, proposed by Waleed et al. [47], is that a reliable estimate of the shape parameter of the Weibull distribution is assumed to be known from historical studies. Sometimes due to the unavailability of adequate historical data, this assumption may not hold true for designing studies related to small populations, such as in the case of rare diseases. It has been demonstrated that gross misspecification of the shape parameter can have an adverse effect on the study power [47]. To tackle this limitation, it is worthwhile to consider adaptation to the study sample size. Since the variance of $\widehat{\gamma}$ provided in Eq. (3.6) depends on the shape parameter, we discuss the implementation of the internal pilot study (IPS) approach, proposed by Wittes and Brittain [49], to readjust the sample size at a prospectively planned time point [17, 49, 50]. The IPS approach is carried out by adjusting the desired sample size based on an estimate of the nuisance shape parameter obtained using the data available at the pre-specified interim stage [49]. More specifically, it requires the following steps in our context:

1. Based on the best available estimate of the shape parameter κ during the planning phase, obtain an initial estimate of the required sample size n.
2. Let p denote the proportion of events, or complete observations (which includes events as well as censored observations) at which we intend to adjust the study sample size. At that time point, we obtain an estimate $\widehat{\kappa}_{\text{new}}$ of the shape parameter from the accumulated data, and subsequently obtain a new estimate of the desired sample size, say n_{new}, using the variance $\sigma^2_{1,\text{int}} \equiv \sigma^2(\gamma_1|\kappa = \widehat{\kappa}_{\text{new}})$ as an estimate of σ^2 under H_1. The number of additional subjects to be enrolled, say n_{add}, is: $n_{\text{add}} = \max(n, n_{\text{new}}) - n$. Therefore, the updated sample size is: $N = n + n_{\text{add}}$.
3. The final analysis is conducted using the data for all N subjects.

It must be noted that the above IPS approach is restricted in the sense that it only permits upward adjustment to the sample size [49, 50]. In our context, this restriction seems reasonable because the TTE data for all subjects, regardless of their survival status, contributes to the estimation of the shape parameter during an interim analysis [18].

3.2.4 Bayesian Predictive Probability

Bayesian predictive probability is a fully Bayesian SC method which can be utilized to calculate the predictive probability of obtaining a positive trial outcome if the clinical trial were to continue to its pre-planned end, conditional on the data accumulated at the interim stage [18, 25]. In our context, suppose that n subjects are enrolled in a single-arm phase II clinical trial designed to test the hypotheses $H_0 : \gamma \leq \gamma_0$ vs. $H_1 : \gamma > \gamma_1$ $(= \gamma_0 + \varepsilon)$, where: $\varepsilon > 0$ is a clinical meaningful effect to be detected under H_1. At the interim stage k, suppose that the survival data corresponding to $n-m$ subjects is fully observed, that is, $n-m$ subjects had already experienced an event or were lost to follow-up, and the remaining m subjects were still active participants in the study. Let $\boldsymbol{X}_k = (X_1, \ldots, X_{n-m})$ and $\boldsymbol{D}_k = (\delta_1, \ldots, \delta_{n-m})$, respectively, denote the vectors of survival times and the corresponding survival status' for the $n-m$ subjects fully observed at the stage k. Similarly, $\boldsymbol{X}_{K-k} = (X_{n-m+1}, \ldots, X_n)$ and $\boldsymbol{D}_{K-k} = (\delta_{n-m+1}, \ldots, \delta_n)$ denote the vectors of survival times and status for the active subjects that are to be observed between the stage k and trial end stage K, respectively. Since $\boldsymbol{X}_{K-k}$ and $\boldsymbol{D}_{K-k}$ are not observable at the interim stage k, suppose that $\tilde{\boldsymbol{X}}_{K-k}$ and $\tilde{\boldsymbol{D}}_{K-k}$ denote the corresponding predicted vectors. When the shape parameter of the Weibull distribution is known, the Bayesian predictive probability of a successful trial outcome is expressed as follows:

$$P_k = \int \mathbb{1}_{(\mathscr{Q} > \eta^*)} \, \pi(\gamma | \kappa, \widehat{\gamma}_k) d\gamma \tag{3.8}$$

where: $\mathscr{Q} = \text{Prob}\big(\gamma > \gamma_1 | \boldsymbol{X}_k, \tilde{\boldsymbol{X}}_{K-k}, \boldsymbol{D}_k, \tilde{\boldsymbol{D}}_{K-k}, \kappa\big)$, η^* is some pre-specified threshold level of probability of a successful trial outcome, and $\pi(\gamma | \kappa, \widehat{\gamma}_k)$ is the posterior distribution of γ based on the

data accumulated at the interim stage k. Also, $\mathbb{1}_{(\mathscr{Q}>\eta^*)}$ is an indicator variable equal to 1 if $\mathscr{Q} > \eta^*$, and 0 otherwise.

When the shape parameter κ is assumed to be known, a fully simulation-based algorithm, outlined by Waleed et al. [47], can be easily implemented to calculate the Bayesian predictive probability. In practice, Dmitreinko and Wang [25] recommended setting the threshold level η^* between 0.90 and 0.975, and after consultation with DSMB, trial can be terminated to conclude efficacy if $P_k \geq \zeta$ for some pre-specified $\zeta \in [0.8, 1]$, or to conclude futility if $P_k \leq \zeta'$ for some $\zeta' \in [0, 0.2]$.

3.2.4.1 Calculating Bayesian Predictive Probability when κ is Unknown

When a reliable estimate of κ is not available from historical studies, an independent joint prior specification of γ and κ is typically considered [51]. For this purpose, we assume that the prior distributions $\gamma \sim \text{Normal}\left(\mu_0, \sigma_0^2\right)$ and $\kappa \sim \text{Gamma}(\alpha_0, \beta_0)$. Suppose that d denotes the total number of observed events, that is, $d = \sum_{i=1}^{n} \delta_i$, then the joint posterior distribution of (γ, κ) can be expressed as below:

$$\begin{aligned}
\pi(\gamma, \kappa|\text{Data}) &\propto L(\gamma, \kappa|\text{Data})\ \pi\left(\gamma|\mu_0, \sigma_0^2\right)\ \pi(\kappa|\alpha_0, \beta_0) \\
&= \left[\prod_{i=1}^{n} f_{Y_i}(x_i|\gamma, \kappa)^{\delta_i} S_{Y_i}(x_i|\gamma, \kappa)^{1-\delta_i}\right] \pi\left(\gamma|\mu_0, \sigma_0^2\right) \pi(\kappa|\alpha_0, \beta_0) \\
&\propto \kappa^{\alpha_0+d-1} \exp\left\{(\kappa-1)\sum_{i=1}^{n} \delta_i \log(x_i) - e^{-\gamma\kappa} \sum_{i=1}^{n} x_i^{\kappa} - \frac{1}{2}\left(\frac{\gamma-\mu_0}{\sigma_0}\right)^2 - \gamma\kappa d - \kappa\beta_0\right\}
\end{aligned} \tag{3.9}$$

Although it is analytically intractable to obtain a closed form for the joint posterior distribution of (γ, κ), it can be conveniently verified that the conditional posterior distributions $\pi(\gamma|\kappa, \text{Data})$ and $\pi(\kappa|\gamma, \text{Data})$ are log-concave, and Gibbs sampling can be implemented using statistical software, such as `R2OpenBUGS` package available in `R` [26, 51, 52].

When a reliable estimate of the shape parameter is not available, we may consider two

approaches to calculate Bayesian predictive probability of a successful trial outcome, in Eq. (3.8), using the simulation-based approach in Waleed et al [47]. The first approach is to update the shape parameter κ used at the design stage with the mode, say κ_{mode}, of the posterior distribution of κ generated at the interim. This updated value of κ is subsequently used to generate a large number of predicted data sets for the active subjects remaining in the study, and then the fully observed and predicted data are used to calculate our desired quantity as defined below:

$$P_{k,\kappa_{\text{mode}}} = \int \mathbb{1}_{(\mathscr{Q}_{\text{mode}} > \eta^*)}\ \pi(\gamma|\kappa,\widehat{\gamma}_k)d\gamma \tag{3.10}$$

where $\mathscr{Q}_{\text{mode}} = \text{Prob}\big(\gamma > \gamma_1 | \boldsymbol{X}_k, \tilde{\boldsymbol{X}}_{K-k}, \boldsymbol{D}_k, \tilde{\boldsymbol{D}}_{K-k}, \kappa_{\text{mode}}\big)$.

The second approach to obtain predictive probability of successful trial outcome is to calculate it as a weighted average over the entire posterior distribution $\pi(\kappa|\widehat{\kappa}_k)$ as below:

$$P_{k,\kappa_{\text{full}}} = \int P_k \cdot \pi(\kappa|\widehat{\kappa}_k)\ d\kappa \tag{3.11}$$

where P_k is defined in Eq. (3.8).

3.2.4.2 Prior Elicitation for the Shape Parameter

We briefly discuss elicitation of appropriate Gamma(shape = α_0, rate = β_0) priors for the shape parameter κ. For this purpose, we consider two approaches similar to the ones proposed by Mayo and Gajewski [53] for eliciting appropriate beta priors in the context of the beta-binomial model. Let $F(\kappa|\alpha_0,\beta_0)$ denote the cumulative distribution function of the Gamma prior, and therefore $F^{-1}(q|\alpha_0,\beta_0)$ is the corresponding $q\cdot 100$th percentile. Suppose that $\mathscr{W}_{100(1-x)}$ represents the required width of the $100(1-x)$th percent probability interval for the gamma prior. We may choose $x = 5\%$ or 10% as available from the historical data. Using the best available knowledge about κ, statisticians may consider the following approaches to elicit prior parameters:

1. *Mode method:* Suppose that the best available estimate of the shape parameter, say κ_{prior}, is considered to be the mode of the prior distribution. For a specified width $\mathscr{W}_{100(1-x)}$, the

unknown parameter α_0 and β_0 can be obtained by simultaneously solving the following system of equations:

$$\text{Median}(\pi(\kappa|\alpha_0,\beta_0)) = \frac{\alpha_0 - 1}{\beta_0} = \kappa_{\text{prior}}$$
$$F^{-1}\left(1-\frac{x}{2}|\alpha_0,\beta_0\right) - F^{-1}\left(\frac{x}{2}|\alpha_0,\beta_0\right) = \mathscr{W}_{100(1-x)}$$

2. *Mean method:* This approach differs from the first approach in the sense that it assumes κ_{prior} to be the mean of the prior distribution. Therefore, the unknown parameter α_0 and β_0 can be obtained by solving the two simultaneous equations:

$$E(\pi(\kappa|\alpha_0,\beta_0)) = \frac{\alpha_0}{\beta_0} = \kappa_{\text{prior}}$$
$$F^{-1}\left(1-\frac{x}{2}|\alpha_0,\beta_0\right) - F^{-1}\left(\frac{x}{2}|\alpha_0,\beta_0\right) = \mathscr{W}_{100(1-x)}$$

It must be noted that small values of the specified width $\mathscr{W}_{100(1-x)}$ yield a more informative prior for the shape parameter. Since closed form solutions are not possible using either approaches, we need to implement numerical methods to obtain the values of α_0 and β_0.

3.3 Simulations and Examples

We present some simulations studies and examples to demonstrate the methods discussed in this manuscript. Statistical software `R` (Version 3.6.3) was used to perform all computations and simulations. Due to their computationally intensive nature, Bayesian predictive probability calculations were done using the high-performance computing (HPC) facilities operated by the Center for Research Computing at the University of Kansas.

Table 3.1: Effect of misspecification of the shape parameter κ on the empirical Type-I error and power for the MLE-based test, for testing $H_0 : M \leq 1$ month vs. $H_1 : M > 1.5$ months, for a study designed assuming exponential survival times ($\kappa = 1$) when the true $\kappa = 0.75$ or 1.25

φ	υ	κ	$\tau = 7$			$\tau = 9$		
			n	α	$1-\beta$	n	α	$1-\beta$
0.1	0%	0.75	–	0.0459	0.6813	–	0.0454	0.6888
		1.00	55	0.0486	0.8892	54	0.0486	0.8859
		1.25	–	0.0484	0.9690	–	0.0508	0.9689
	10%	0.75	–	0.0456	0.6887	–	0.0471	0.6900
		1.00	60	0.0440	0.8954	59	0.0464	0.8891
		1.25	–	0.0447	0.9731	–	0.0421	0.9727
	20%	0.75	–	0.0465	0.6906	–	0.0484	0.6953
		1.00	67	0.0492	0.8962	66	0.0446	0.8910
		1.25	–	0.0451	0.9761	–	0.0436	0.9739
	30%	0.75	–	0.0465	0.7090	–	0.0482	0.7031
		1.00	76	0.0446	0.8988	75	0.0504	0.8967
		1.25	–	0.0478	0.9793	–	0.0472	0.9780
1.0	0%	0.75	–	0.0425	0.6768	–	0.0469	0.6876
		1.00	57	0.0463	0.8914	54	0.0473	0.8855
		1.25	–	0.0479	0.9732	–	0.0498	0.9672
	10%	0.75	–	0.0440	0.6906	–	0.0469	0.6957
		1.00	62	0.0462	0.8980	60	0.0404	0.8945
		1.25	–	0.0448	0.9773	–	0.0442	0.9741
	20%	0.75	–	0.0440	0.6915	–	0.0489	0.6980
		1.00	69	0.0432	0.9020	67	0.0470	0.8984
		1.25	–	0.0441	0.9760	–	0.0448	0.9748
	30%	0.75	–	0.0469	0.6851	–	0.0468	0.6954
		1.00	77	0.0465	0.9008	76	0.0481	0.9000
		1.25	–	0.0491	0.9790	–	0.0485	0.9767
5.0	0%	0.75	–	0.0422	0.6808	–	0.0445	0.6806
		1.00	60	0.0456	0.9033	55	0.0469	0.8894
		1.25	–	0.0514	0.9793	–	0.0491	0.9733
	10%	0.75	–	0.0431	0.6859	–	0.0431	0.6814
		1.00	65	0.0447	0.9116	61	0.0474	0.8966
		1.25	–	0.0487	0.9789	–	0.0454	0.9742
	20%	0.75	–	0.0471	0.6847	–	0.0494	0.6862
		1.00	71	0.0445	0.9017	67	0.0467	0.8933
		1.25	–	0.0484	0.9800	–	0.0473	0.9733
	30%	0.75	–	0.0466	0.6882	–	0.0455	0.6963
		1.00	79	0.0489	0.9003	76	0.0440	0.8987
		1.25	–	0.0459	0.9789	–	0.0446	0.9757

Note: In this simulation study, we used nominal Type-I error = 5%, nominal power = 90%, maximum accrual time ω = 3 months, and varying values of the administrative censoring time τ (in months), power parameter of the accrual distribution φ, and loss to follow-up rate υ. Results are based on a total of 10,000 simulations.

3.3.1 Sample Size Reestimation using the IPS Approach

Suppose that a fixed sample study is being designed, using the method in Section 3.2.2, to test the following alternatives about the median survival time $H_0 : M \leq 1$ month vs. $H_1 : M > 1.5$ months with the nominal Type-I error rate and power set to be 5% and 90%, respectively. Suppose that the enrollment period will span a total of 3 months, that is, $\omega = 3$ months. For this hypothetical example, varying values of the administrative censoring time ($\tau = 7, 9$ months), loss to follow-up rate ($\upsilon = 0\%, 10\%, 20\%, 30\%$), and power parameter of the accrual distribution (φ = 0.1 – early; 1.0 – uniform; 5.0 – late) were considered. In the absence of a reliably accurate estimate of the shape parameter, suppose that the fixed sample study was originally designed assuming exponential survival times, that is, $\kappa = 1$, and the corresponding sample sizes for all scenarios are reported in Table 3.1.

To examine the performance of the asymptotic test statistic by Waleed et al. [47] in terms of empirical Type-I error and power, a total of 10,000 simulations were performed after computing the required sample sizes under the assumption of exponential survival times. We note that the empirical Type-I error rate remains preserved within our desired nominal levels even when the shape parameter was misspecified at the design stage. On the other hand, the empirical power is significantly affected by misspecification of the shape parameter. More specifically, the fixed sample study tends to be under-powered (over-powered) if the true shape parameter κ was in fact smaller (larger) than 1.

To address this issue, suppose that the researchers plan to adjust the sample size using the IPS approach of Wittes and Brittain, outlined in Section 3.2.3, after a proportion of the n enrolled subjects, say p_E, have experienced an event. For the sake of demonstration, we assume $p_E = 30\%$ and 50%, and study the properties of our design in terms of the expected sample size, empirical Type-I error rate, and power. The corresponding results based on a total of 10,000 simulations are reported in Table 3.2. We summarize our findings as below:

1. If the true shape parameter was equal to 0.75 (decreasing hazard), the expected sample size obtained using the IPS approach is almost twice the sample size needed for a fixed study

design. Such a significant increase in the expected sample size is likely to put substantial practical constraints to carryout remainder of the study. We note that the expected sample size is also greater than the original fixed sample size n even if the assumption of exponential survival times ($\kappa = 1$) holds true. This increase in the expected sample size, however, is much more achievable from a practical perspective. Finally, we observe a minimal increase in the expected sample size when the true shape parameter was 1.25. As one might anticipate, the expected sample size with $p_E = 50\%$ is slightly smaller than that for $p_E = 30\%$.

2. As expected for the IPS approach [49, 50], the empirical Type-I error rate tends to be slightly inflated in almost all of the scenarios. In our context, this inflation in the Type-I error rate is more pronounced when subjects are accrued very late in the enrollment period (that is, $\varphi = 5.0$) where it can be as much as twice the desired level in some cases.
3. The desired threshold for the empirical power is achieved with the adjustment of the study sample size using the IPS approach. When the true shape parameter was 0.75, the study tends to be over-powered due to very large expected sample sizes.
4. For a majority of scenarios, the empirical Type-I error rate and power corresponding to $p_E = 50\%$ is slightly larger than that for $p_E = 30\%$. This is possibly explained due to the fact that, as we implement the IPS approach after observing more events, we get a smaller but more accurate estimate of the variance which results in slightly greater number of rejections under the null and alternative hypotheses.

We also investigated our design properties when the IPS approach is implemented after a proportion of the n enrolled subjects, say p_O, is fully observed. That is, p_O includes all those subjects who have either experienced an event or were lost to follow-up. The corresponding results are presented in Table 3.3. We note that the expected sample sizes are slightly larger in this case, but the overall trends in the empirical Type-I error rate and power are virtually similar to those observed using p_E.

Table 3.2: Properties of a study designed to test $H_0 : M \leq 1$ month vs. $H_1 : M > 1.5$ months, when sample size reestimation is done using the IPS approach after $p_E\%$ of the n enrolled subjects experienced the event of interest

φ	υ	True κ	$\tau = 7$						$\tau = 9$					
			$p_E = 30\%$			$p_E = 50\%$			$p_E = 30\%$			$p_E = 50\%$		
			E(N)	α	$1-\beta$	E(N)	α	$1-\beta$	E(N)	α	$1-\beta$	E(N)	α	$1-\beta$
0.1	0%	0.75	109	0.0585	0.9705	108	0.0627	0.9824	103	0.0613	0.9693	102	0.0598	0.9807
		1.00	66	0.0654	0.9105	63	0.0622	0.9113	64	0.0630	0.9041	61	0.0637	0.9082
		1.25	57	0.0639	0.8887	56	0.0643	0.8806	56	0.0633	0.8748	55	0.0636	0.8748
	10%	0.75	117	0.0509	0.9737	116	0.0511	0.9843	111	0.0570	0.9709	111	0.0591	0.9829
		1.00	71	0.0533	0.9125	68	0.0555	0.9114	69	0.0586	0.9119	67	0.0583	0.9119
		1.25	62	0.0599	0.8876	61	0.0623	0.8885	61	0.0632	0.8817	60	0.0649	0.8828
	20%	0.75	127	0.0517	0.9745	127	0.0520	0.9857	123	0.0436	0.9749	122	0.0495	0.9828
		1.00	78	0.0505	0.9135	75	0.0536	0.9193	77	0.0566	0.9178	74	0.0511	0.9153
		1.25	69	0.0534	0.8946	68	0.0552	0.8949	68	0.0526	0.8834	67	0.0545	0.8868
	30%	0.75	141	0.0412	0.9798	140	0.0472	0.9893	136	0.0482	0.9789	137	0.0503	0.9877
		1.00	87	0.0527	0.9191	85	0.0505	0.9156	86	0.0525	0.9217	83	0.0495	0.9174
		1.25	77	0.0524	0.8992	77	0.0498	0.8933	77	0.0526	0.8892	76	0.0494	0.8891
1.0	0%	0.75	120	0.0657	0.9804	116	0.0773	0.9876	111	0.0605	0.9757	107	0.0659	0.9864
		1.00	69	0.0574	0.9164	66	0.0624	0.9155	65	0.0606	0.9120	63	0.0634	0.9129
		1.25	59	0.0601	0.8952	58	0.0642	0.8882	56	0.0613	0.8856	55	0.0641	0.8797
	10%	0.75	127	0.0549	0.9818	123	0.0762	0.9884	120	0.0571	0.9795	115	0.0573	0.9874
		1.00	74	0.0535	0.9221	71	0.0586	0.9226	71	0.0556	0.9174	68	0.0622	0.9145
		1.25	63	0.0618	0.8914	63	0.0570	0.8915	62	0.0543	0.8883	61	0.0614	0.8857
	20%	0.75	137	0.0539	0.9825	134	0.0795	0.9898	129	0.0453	0.9838	125	0.0547	0.9883
		1.00	80	0.0517	0.9211	77	0.0595	0.9252	78	0.0513	0.9176	75	0.0562	0.9188
		1.25	70	0.0525	0.8958	70	0.0574	0.8965	68	0.0543	0.8924	68	0.0574	0.8913
	30%	0.75	150	0.0582	0.9857	144	0.0816	0.9902	143	0.0444	0.9826	139	0.0560	0.9881
		1.00	89	0.0496	0.9283	86	0.0539	0.9237	88	0.0521	0.9195	85	0.0487	0.9207
		1.25	78	0.0547	0.8988	78	0.0476	0.8955	77	0.0507	0.8986	77	0.0525	0.9027
5.0	0%	0.75	123	0.1119	0.9789	121	0.1613	0.9826	111	0.0671	0.9761	108	0.0766	0.9844
		1.00	72	0.0587	0.9222	69	0.0638	0.9234	66	0.0628	0.9128	63	0.0604	0.9120
		1.25	62	0.0632	0.8999	61	0.0632	0.8989	57	0.0641	0.8838	56	0.0624	0.8873
	10%	0.75	130	0.1127	0.9826	129	0.1672	0.9835	120	0.0606	0.9778	117	0.0748	0.9845
		1.00	77	0.0597	0.9245	74	0.0573	0.9203	71	0.0545	0.9153	69	0.0561	0.9159
		1.25	67	0.0577	0.9014	66	0.0568	0.9017	63	0.0588	0.8887	62	0.0598	0.8873
	20%	0.75	140	0.1102	0.9836	138	0.1637	0.9845	129	0.0545	0.9814	127	0.0718	0.9891
		1.00	83	0.0539	0.9287	80	0.0558	0.9170	78	0.0557	0.9219	76	0.0515	0.9178
		1.25	72	0.0516	0.9078	72	0.0500	0.9088	69	0.0545	0.8921	68	0.0512	0.8928
	30%	0.75	152	0.0988	0.9811	149	0.1598	0.9792	141	0.0554	0.9814	140	0.0744	0.9883
		1.00	91	0.0565	0.9220	88	0.0531	0.9173	88	0.0500	0.9216	84	0.0524	0.9146
		1.25	80	0.0500	0.9035	80	0.0499	0.9058	77	0.0477	0.8959	77	0.0531	0.8955

Note: The fixed sample study was initially designed assuming exponential survival times ($\kappa = 1$). In this simulation study, we used nominal Type-I error = 5%, nominal power = 90%, maximum accrual time ω = 3 months, and varying values of the administrative censoring time τ (in months), power parameter of the accrual distribution φ and loss to follow-up rate υ. Results are based on a total of 10,000 simulations.

3.3.2 Bayesian Predictive Probability for Hypothetical Single-Arm Trials

Suppose that a fixed sample study is designed to evaluate whether an experimental treatment yields an improvement in the median PFS time (in months). More specifically, suppose that researchers

Table 3.3: Properties of a study designed to test $H_0 : M \leq 1$ month vs. $H_1 : M > 1.5$ months, when sample size reestimation is done using the IPS approach after $p_O\%$ of the n enrolled subjects have been observed

φ	υ	True κ	$\tau = 7$						$\tau = 9$					
			$p_O = 30\%$			$p_O = 50\%$			$p_O = 30\%$			$p_O = 50\%$		
			E(N)	α	$1-\beta$	E(N)	α	$1-\beta$	E(N)	α	$1-\beta$	E(N)	α	$1-\beta$
0.1	0%	0.75	109	0.0585	0.9705	108	0.0627	0.9824	103	0.0613	0.9693	102	0.0598	0.9807
		1.00	66	0.0654	0.9105	63	0.0622	0.9113	64	0.0630	0.9041	61	0.0637	0.9082
		1.25	57	0.0639	0.8887	56	0.0643	0.8806	56	0.0633	0.8748	55	0.0636	0.8748
	10%	0.75	119	0.0478	0.9719	117	0.0513	0.9844	113	0.0532	0.9710	112	0.0579	0.9806
		1.00	72	0.0528	0.9118	69	0.0513	0.9137	70	0.0601	0.9131	68	0.0541	0.9160
		1.25	62	0.0557	0.8918	61	0.0599	0.8888	61	0.0580	0.8842	60	0.0618	0.8818
	20%	0.75	131	0.0458	0.9728	129	0.0472	0.9838	126	0.0477	0.9723	124	0.0467	0.9824
		1.00	80	0.0531	0.9161	77	0.0535	0.9180	79	0.0542	0.9119	76	0.0544	0.9149
		1.25	69	0.0554	0.8957	68	0.0549	0.8944	68	0.0522	0.8909	67	0.0546	0.8842
	30%	0.75	147	0.0497	0.9726	144	0.0450	0.9836	141	0.0471	0.9723	139	0.0498	0.9834
		1.00	91	0.0482	0.9237	87	0.0518	0.9196	89	0.0518	0.9178	86	0.0497	0.9219
		1.25	79	0.0513	0.8999	77	0.0540	0.8932	78	0.0552	0.8996	76	0.0544	0.8899
1.0	0%	0.75	120	0.0617	0.9805	115	0.0814	0.9896	111	0.0643	0.9758	107	0.0605	0.9840
		1.00	69	0.0616	0.9212	66	0.0548	0.9198	65	0.0609	0.9104	62	0.0591	0.9127
		1.25	59	0.0636	0.8940	58	0.0614	0.8898	56	0.0690	0.8793	55	0.0603	0.8793
	10%	0.75	129	0.0561	0.9796	125	0.0656	0.9872	120	0.0592	0.9773	116	0.0576	0.9854
		1.00	75	0.0520	0.9200	72	0.0604	0.9167	72	0.0571	0.9137	69	0.0548	0.9120
		1.25	64	0.0527	0.8947	63	0.0549	0.8952	62	0.0586	0.8876	61	0.0597	0.8817
	20%	0.75	140	0.0503	0.9789	136	0.0606	0.9861	132	0.0476	0.9774	128	0.0480	0.9854
		1.00	83	0.0542	0.9236	79	0.0555	0.9293	80	0.0534	0.9244	76	0.0559	0.9157
		1.25	71	0.0533	0.9028	70	0.0551	0.9026	69	0.0545	0.8969	68	0.0568	0.8915
	30%	0.75	155	0.0480	0.9789	151	0.0585	0.9872	147	0.0442	0.9803	144	0.0521	0.9869
		1.00	92	0.0498	0.9256	89	0.0493	0.9257	91	0.0498	0.9260	87	0.0545	0.9239
		1.25	79	0.0497	0.9048	78	0.0509	0.9015	78	0.0502	0.9021	77	0.0482	0.8953
5.0	0%	0.75	125	0.1168	0.9812	121	0.1618	0.9837	111	0.0673	0.9779	109	0.0742	0.9827
		1.00	72	0.0624	0.9225	69	0.0555	0.9221	66	0.0593	0.9129	63	0.0608	0.9146
		1.25	62	0.0630	0.9019	61	0.0589	0.9001	57	0.0638	0.8906	56	0.0646	0.8825
	10%	0.75	133	0.1045	0.9800	129	0.1518	0.9869	122	0.0536	0.9799	117	0.0643	0.9832
		1.00	78	0.0515	0.9249	75	0.0562	0.9237	72	0.0580	0.9206	70	0.0569	0.9161
		1.25	67	0.0591	0.9082	66	0.0595	0.9062	63	0.0577	0.8959	62	0.0603	0.8904
	20%	0.75	145	0.0962	0.9812	140	0.1370	0.9843	134	0.0508	0.9791	130	0.0600	0.9849
		1.00	85	0.0557	0.9276	82	0.0558	0.9219	80	0.0551	0.9194	77	0.0525	0.9222
		1.25	73	0.0512	0.9025	72	0.0553	0.9098	69	0.0513	0.8925	68	0.0520	0.8924
	30%	0.75	160	0.0915	0.9799	154	0.1193	0.9844	150	0.0518	0.9794	144	0.0513	0.9838
		1.00	95	0.0515	0.9300	91	0.0483	0.9210	91	0.0507	0.9220	88	0.0521	0.9197
		1.25	81	0.0473	0.9074	80	0.0506	0.9070	78	0.0512	0.8998	77	0.0537	0.8973

Note: The fixed sample study was initially designed assuming exponential survival times ($\kappa = 1$). In this simulation study, we used nominal Type-I error = 5%, nominal power = 90%, maximum accrual time ω = 3 months, and varying values of the administrative censoring time τ (in months), power parameter of the accrual distribution φ and loss to follow-up rate υ. Results are based on a total of 10,000 simulations. Observed subjects are defined as those who have either experienced the event of interest or were censored by the interim look time.

are interested in testing the hypotheses $H_0 : M \leq 1$ month vs. $H_1 : M > 1.5$ months with nominal Type-I error rate and power equal to 5% and 90%, respectively. For this hypothetical study, investigators anticipate a uniform accrual pattern ($\varphi = 1$) with maximum enrollment time $\omega = 3$ months, and administrative censoring time $\tau = 12$ months. The expected loss to follow-up rate

is $\upsilon = 15\%$. Assuming exponentially distributed survival times, $n = 62$ subjects are enrolled to conduct this fixed sample study with the given characteristics.

On the recommendation of the DSMB, suppose that an interim analysis based on Bayesian predictive probability is to be conducted one month after the conclusion of the accrual period (that is, calendar time $\ell = 4$ months). Suppose that we use threshold level $\eta^* = 95\%$, and the following decision rules for the Bayesian predictive probability P_k: conclude efficacy (futility) if $P_k \geq 0.80$ ($P_k \leq 0.20$), or decide to continue the trial if $P_k \in (0.20, 0.80)$. For the sake of demonstration, we generated a sample of size $n = 62$ from each of the Weibull distributions with different underlying median parameter ($M = 0.75, 1.00, 1.25, 1.50, 2.00$) and true shape parameter ($\kappa = 0.75, 1.00, 1.25$). For this example, a non-informative normal prior was used for γ, and $\kappa \sim \text{Gamma}(11, 10)$ which has a unit mode. For calculating BPP using the simulation-based approach, a total of 1000 predicted data sets were generated for active subjects at the interim, and the corresponding results are summarized in Table 3.4. We make the following observations:

1. As anticipated, the number of observed events at the interim analysis, denoted by $n_{e;k}$, is the largest for the data generated assuming the smallest median survival time (0.75 months), and it decreases as the assumed median survival time increases. The converse holds true for the number of active subjects $n_{a;k}$. The number of subjects lost to follow-up, $n_{c;k}$, slightly exceeded the expected loss to follow-up rate $\upsilon = 15\%$.

2. In all cases, the data dominates the assumed prior of κ in the sense that the mode of the posterior distribution is pulled towards the true value of κ used for the underlying data distributions.

3. Irrespective of the true shape parameter κ, the predictive probability of a positive trial outcome calculated using either approach is close to 1 for the data sets generated with underlying median equal to 2.00 months, which suggests us to conclude efficacy of the experimental treatment. We can draw appropriate conclusions for the rest of scenarios in accordance with the decision rules specified above.

   It is worth noting that there are some circumstances in which the two approaches lead us to different conclusions at the interim stage. As an example, consider the data simulated assum-

ing the underlying median survival time equal to 0.75 months and true shape parameter 1.25. We note that the BPP calculated using the posterior mode and the entire posterior distribution of κ is 0.0406 (conclude futility) and 0.3012 (continue trial), respectively. Since the posterior mode of $\pi(\kappa|\gamma, \widehat{\kappa}_k)$ is 1.3268 suggesting increasing hazard (that is, shorter survival times), it would be reasonable to take a rather conservative approach if ethically permissible, and decide to continue the trial instead of terminating it at the interim. This example demonstrates that it is of vital importance to take an informed decision by considering all the factors before making a final conclusion.

4. In general, the predictive probability calculated by incorporating the entire conditional posterior distribution $\pi(\kappa|\gamma, \widehat{\kappa}_k)$ tends to be greater than the one computed using its mode. In our reported results in Table 3.4, we observe some scenarios where the BPP calculated using the latter approach is larger in magnitude. For instance, for the data generated assuming the true κ to be 1.25 and the underlying median equal to 1.25 months, the BPP is 0.9243 and 0.8364 for the posterior mode based and the full distribution approach, respectively. This discrepancy can be explained by the fact that the BPP corresponding to some κ greater than the mode (1.3282) of $\pi(\kappa|\gamma, \widehat{\kappa}_k)$ was smaller than 0.9243 (BPP corresponding to the posterior mode). As a consequence, a smaller value of the BPP was obtained when a weighted average of the predictive probabilities was computed over the entire posterior distribution of κ. Since these calculations were conducted by simulating only 1000 predicted data sets for the active subjects at the interim, it is recommended to generate a larger number of predicted data sets to minimize the effect of such randomness in our results. Even though the BPP values under the two approaches are quantitatively different, it must be noted that they qualitatively suggest the same decision of stopping the trial due to a high predictive probability of a successful trial outcome.

Table 3.4: Bayesian predictive probability calculated at the interim look time $\ell = 4$ months for data sets generated using varying values of the median survival time and true shape parameter κ for a study designed to test the hypotheses $H_0 : M \leq 1.00$ month vs. $H_1 : M > 1.50$ months

True κ	Underlying Dist. Median	Observed Status			Mode of $\pi(\kappa\|\gamma, \widehat{\kappa}_k)$	Bayesian Predictive Probability	
		$n_{e;k}$	$n_{c;k}$	$n_{a;k}$		Using the mode of $\pi(\kappa\|\gamma, \widehat{\kappa}_k)$	Using entire $\pi(\kappa\|\gamma, \widehat{\kappa}_k)$
0.75	0.75	38	13	11	0.9248	6.113×10^{-5}	2.152×10^{-3}
	1.00	33	13	16	0.9212	0.0832	0.0612
	1.25	30	12	20	0.8781	0.3429	0.4356
	1.50	28	12	22	0.9509	0.7133	0.7546
	2.00	21	12	29	0.8533	0.9988	0.9989
1.00	0.75	39	14	9	1.0695	4.012×10^{-4}	6.057×10^{-2}
	1.00	37	13	12	1.1578	0.0204	0.1503
	1.25	33	13	16	1.1592	0.4823	0.5329
	1.50	29	13	20	1.0905	0.6589	0.9160
	2.00	22	12	28	1.0459	0.9999	0.9993
1.25	0.75	43	14	5	1.3268	0.0406	0.3012
	1.00	39	14	9	1.3382	0.1431	0.5015
	1.25	34	13	15	1.3282	0.9243	0.8364
	1.50	32	13	17	1.3979	0.9999	0.9733
	2.00	21	12	29	1.2392	1.0000	0.9999

Note: For this study, we assume uniform accrual pattern ($\varphi = 1.00$), maximum accrual time $\omega = 3$ months, administrative censoring time $\tau = 12$ months, and loss to follow-up rate $\upsilon = 15\%$. The fixed sample study was designed assuming exponential survival times, and $n = 62$ subjects were enrolled during the accrual phase.

We consider another hypothetical study in which investigators are interested in testing the alternatives $H_0 : M \leq 2.50$ months vs. $H_1 : M > 3.75$ months at a 5% level of significance with 80% power. For this example, we assume a non-uniform accrual pattern ($\varphi = 1.25$) during the accrual phase spanning $\omega = 3$ months, and a follow-up period of 9 months (i.e., $\tau = 12$ months). Assuming exponential survival times ($\kappa = 1$), the sample size needed for the fixed sample design is $n = 50$ subjects when the expected loss to follow-up rate $\upsilon = 15\%$. Suppose that an interim analysis using the BPP approach is requested at the calendar time $\ell = 6$ months. We generated a sample of size $n = 50$ subjects from each of the Weibull distributions with different values of the underlying median parameter ($M = 2.00, 2.50, 3.25, 3.75, 4.50$) and true shape parameter ($\kappa = 0.50, 1.00, 1.50$). We consider the same threshold η^* and decision rules as used in the previous example, and the corresponding results are presented in Table 3.5. Comparing these results to our previous example, we can make similar observations regarding the subject survival status, posterior mode of $\pi(\kappa|\gamma, \widehat{\kappa}_k)$, and BPP. When the true shape parameter is 0.50 (decreasing hazard), our results suggest that the trial can be stopped for futility for each of the five data sets. On the other hand, when the true

shape parameter is 1.00 (constant hazard) or 1.50 (increasing), the trial can be: (i) terminated to conclude futility (efficacy) for the data generated from distributions with the underlying median $M \leq 2.50$ ($M = 4.50$), and (ii) continued to its pre-planned end in the remaining scenarios.

Table 3.5: Bayesian predictive probability calculated at the interim look time $\ell = 6$ months for data sets generated using varying values of the median survival time and true shape parameter κ for a study designed to test the hypotheses $H_0 : M \leq 2.50$ months vs. $H_1 : M > 3.75$ months

True κ	Underlying Dist. Median	Observed Status			Mode of $\pi(\kappa\|\gamma, \widehat{\kappa}_k)$	Bayesian Predictive Probability	
		$n_{e;k}$	$n_{c;k}$	$n_{a;k}$		Using the mode of $\pi(\kappa\|\gamma, \widehat{\kappa}_k)$	Using entire $\pi(\kappa\|\gamma, \widehat{\kappa}_k)$
0.50	2.00	27	6	17	0.5575	0	4.127×10^{-5}
	2.50	27	6	17	0.5911	0	1.240×10^{-4}
	3.25	23	6	21	0.5628	5.267×10^{-3}	0.0618
	3.75	23	6	21	0.6144	0.0432	0.0692
	4.50	23	6	21	0.6441	0.1409	0.1028
1.00	2.00	35	6	9	1.0778	0	4.933×10^{-4}
	2.50	32	6	12	1.0875	1.916×10^{-4}	2.299×10^{-3}
	3.25	27	6	17	1.0353	0.0774	0.1628
	3.75	24	6	20	1.0263	0.4222	0.5539
	4.50	22	6	22	1.0568	0.8332	0.8509
1.50	2.00	39	7	4	1.5251	2.880×10^{-4}	0.1049
	2.50	36	6	8	1.5062	5.140×10^{-4}	0.1514
	3.25	30	6	14	1.4927	0.2911	0.3945
	3.75	26	6	18	1.4147	0.7214	0.7356
	4.50	20	6	24	1.3170	0.9941	0.9864

Note: For this study, we assume non-uniform accrual pattern ($\varphi = 1.25$), maximum accrual time $\omega = 3$ months, administrative censoring time $\tau = 12$ months, and loss to follow-up rate $\upsilon = 15\%$. The fixed sample study was designed assuming exponential survival times, and $n = 50$ subjects were enrolled during the accrual phase.

In both examples discussed above, we assumed $\kappa \sim \text{Gamma}(11, 10)$ which has a unit mode, and an equal-tailed 95% probability interval of width 1.2899. We also studied the impact of different priors for κ on our calculations. For this purpose, we consider five different gamma priors with a unit mode, and width of equal-tailed 95% probability interval to be 0.1, 0.5, 1.0, 2.0, and 5.0. These priors are denoted by $\mathscr{P}_i$ ($i = 1, \ldots, 5$), where $\mathscr{P}_1$ is the most informative gamma prior having the smallest width (= 0.1) and $\mathscr{P}_5$ is the most non-informative gamma prior having the largest width (= 5.0). The parameters for the appropriate gamma priors were computed by numerically solving the two simultaneous equations in Section 3.2.4.2, and these priors are given as below:

$\mathscr{P}_1 : \text{Gamma}(\alpha_0 = 1538.4076, \beta_0 = 1537.4076)$

$\mathscr{P}_2 : \text{Gamma}(\alpha_0 = 63.2778, \beta_0 = 62.2778)$

$\mathscr{P}_3 : \text{Gamma}(\alpha_0 = 17.1526, \beta_0 = 16.1526)$

$\mathscr{P}_4$: Gamma($\alpha_0 = 5.5452, \beta_0 = 4.5452$)

$\mathscr{P}_5$: Gamma($\alpha_0 = 2.0917, \beta_0 = 1.0917$)

Table 3.6: Comparison of Bayesian predictive probability, calculated at the interim look time $\ell = 4$ months, for different priors for the shape parameter κ and data sets simulated using varying values of the underlying median survival time in a study designed to test the hypotheses $H_0 : M \leq 1.00$ month vs. $H_1 : M > 1.50$ months

Approach	True κ	Prior for κ	Underlying Median of Simulated Data Distribution				
			0.75	1.00	1.25	1.50	2.00
Using the mode of $\pi(\kappa\|\gamma, \widehat{\kappa}_k)$	0.75	$\mathscr{P}_1$	0	0.0157	0.3455	0.6670	0.9972
		$\mathscr{P}_2$	3.384×10^{-6}	0.0691	0.4726	0.7474	0.9954
		$\mathscr{P}_3$	8.045×10^{-5}	0.0833	0.3828	0.7524	0.9956
		$\mathscr{P}_4$	4.057×10^{-6}	0.0667	0.3965	0.7333	0.9989
		$\mathscr{P}_5$	1.188×10^{-4}	0.0887	0.4169	0.7250	0.9999
		$\mathscr{P}_U$	1.314×10^{-4}	0.0881	0.4466	0.7222	0.9998
	1.00	$\mathscr{P}_1$	1.176×10^{-5}	0.0068	0.4382	0.9858	1.0000
		$\mathscr{P}_2$	3.287×10^{-4}	0.0071	0.1952	0.9279	1.0000
		$\mathscr{P}_3$	3.748×10^{-4}	0.0222	0.4567	0.8191	1.0000
		$\mathscr{P}_4$	3.662×10^{-4}	0.0927	0.6336	0.6527	0.9988
		$\mathscr{P}_5$	3.428×10^{-3}	0.0937	0.4965	0.9747	0.9984
		$\mathscr{P}_U$	1.099×10^{-3}	0.0060	0.8380	0.8465	0.9968
	1.25	$\mathscr{P}_1$	0	0.0030	0.5016	0.9428	1.0000
		$\mathscr{P}_2$	1.753×10^{-3}	0.1221	0.4235	0.9012	1.0000
		$\mathscr{P}_3$	0.0272	0.7472	0.9260	0.9872	1.0000
		$\mathscr{P}_4$	0.0295	0.4164	0.8905	0.9539	1.0000
		$\mathscr{P}_5$	0.1727	0.4159	0.8523	1.0000	1.0000
		$\mathscr{P}_U$	0.1728	0.7687	0.8777	1.0000	1.0000
Using the entire $\pi(\kappa\|\gamma, \widehat{\kappa}_k)$	0.75	$\mathscr{P}_1$	2.409×10^{-5}	0.0371	0.3755	0.7254	0.9976
		$\mathscr{P}_2$	1.013×10^{-3}	0.0509	0.4002	0.7416	0.9977
		$\mathscr{P}_3$	1.638×10^{-3}	0.0575	0.4527	0.7869	0.9982
		$\mathscr{P}_4$	1.853×10^{-3}	0.0606	0.3862	0.7714	0.9988
		$\mathscr{P}_5$	2.393×10^{-3}	0.0630	0.4070	0.7728	0.9986
		$\mathscr{P}_U$	3.104×10^{-3}	0.0737	0.4309	0.7376	0.9989
	1.00	$\mathscr{P}_1$	1.726×10^{-4}	0.0143	0.4553	0.9515	0.9999
		$\mathscr{P}_2$	0.0176	0.0586	0.4266	0.9008	0.9998
		$\mathscr{P}_3$	0.0537	0.1439	0.5281	0.9056	0.9997
		$\mathscr{P}_4$	0.0667	0.1575	0.6203	0.9051	0.9990
		$\mathscr{P}_5$	0.0849	0.2369	0.5955	0.9243	0.9989
		$\mathscr{P}_U$	0.0756	0.2356	0.5983	0.9305	0.9969
	1.25	$\mathscr{P}_1$	0	0.0094	0.5235	0.9093	1.0000
		$\mathscr{P}_2$	0.0319	0.1391	0.5912	0.9130	1.0000
		$\mathscr{P}_3$	0.0206	0.4452	0.7620	0.9616	1.0000
		$\mathscr{P}_4$	0.3686	0.5846	0.8727	0.9869	1.0000
		$\mathscr{P}_5$	0.4315	0.6309	0.8715	0.9844	1.0000
		$\mathscr{P}_U$	0.4228	0.6315	0.8661	0.9838	1.0000

Note: For this study, we assume a uniform accrual pattern ($\varphi = 1.00$), maximum accrual time $\omega = 3$ months, administrative censoring time $\tau = 12$ months, and loss to follow-up rate $\upsilon = 15\%$. The fixed sample study was designed assuming exponential survival times, and $n = 62$ subjects were enrolled during the accrual phase.

Table 3.7: Comparison of Bayesian predictive probability, calculated at the interim look time $\ell = 6$ months, for different priors for the shape parameter κ and data sets simulated using varying values of the underlying median survival time in a study designed to test the hypotheses $H_0 : M \leq 2.50$ month vs. $H_1 : M > 3.75$ months

Approach	True κ	Prior for κ	Underlying Median of Simulated Data Distribution				
			2.00	2.50	3.25	3.75	4.50
Using the mode of $\pi(\kappa\|\gamma, \widehat{\kappa}_k)$	0.50	$\mathscr{P}_1$	3.733×10^{-4}	7.818×10^{-4}	0.0811	0.1203	0.1424
		$\mathscr{P}_2$	5.733×10^{-6}	9.868×10^{-5}	0.0614	0.0508	0.1504
		$\mathscr{P}_3$	0	1.734×10^{-5}	0.0094	0.0564	0.0094
		$\mathscr{P}_4$	0	0	0.0247	0.0163	0.0802
		$\mathscr{P}_5$	0	0	0.0135	0.0438	0.0321
		$\mathscr{P}_U$	0	0	0.0021	0.0474	0.0948
	1.00	$\mathscr{P}_1$	0	9.997×10^{-7}	0.0458	0.3969	0.8238
		$\mathscr{P}_2$	0	3.442×10^{-4}	0.0509	0.4782	0.8650
		$\mathscr{P}_3$	0	2.961×10^{-5}	0.0547	0.4777	0.8173
		$\mathscr{P}_4$	0	4.896×10^{-7}	0.0705	0.3751	0.7449
		$\mathscr{P}_5$	0	3.869×10^{-7}	0.1464	0.5333	0.7483
		$\mathscr{P}_U$	0	1.256×10^{-5}	0.0668	0.4766	0.7169
	1.50	$\mathscr{P}_1$	0	0	0.2513	0.8614	0.9975
		$\mathscr{P}_2$	0	2.646×10^{-6}	0.0645	0.7316	0.9982
		$\mathscr{P}_3$	0	3.555×10^{-4}	0.3139	0.7384	0.9951
		$\mathscr{P}_4$	0.0107	0.0246	0.3323	0.7949	0.9741
		$\mathscr{P}_5$	0.1102	0.1129	0.4515	0.8839	0.9673
		$\mathscr{P}_U$	0.1089	0.3826	0.4539	0.8793	0.9789
Using the entire $\pi(\kappa\|\gamma, \widehat{\kappa}_k)$	0.50	$\mathscr{P}_1$	7.565×10^{-4}	0.0017	0.0779	0.0919	0.1326
		$\mathscr{P}_2$	1.305×10^{-4}	4.062×10^{-4}	0.0521	0.0709	0.0971
		$\mathscr{P}_3$	7.274×10^{-5}	1.600×10^{-4}	0.0502	0.0711	0.0941
		$\mathscr{P}_4$	2.730×10^{-5}	1.835×10^{-4}	0.0519	0.0899	0.1267
		$\mathscr{P}_5$	8.571×10^{-5}	1.396×10^{-4}	0.0989	0.0906	0.1403
		$\mathscr{P}_U$	1.536×10^{-5}	2.032×10^{-4}	0.0988	0.1076	0.1405
	1.00	$\mathscr{P}_1$	0	5.503×10^{-5}	0.1038	0.4419	0.8377
		$\mathscr{P}_2$	1.767×10^{-6}	4.276×10^{-4}	0.1535	0.5185	0.8408
		$\mathscr{P}_3$	8.179×10^{-5}	0.0016	0.1671	0.5357	0.8690
		$\mathscr{P}_4$	2.973×10^{-4}	0.0047	0.1801	0.5282	0.8451
		$\mathscr{P}_5$	6.112×10^{-4}	0.0052	0.1806	0.5641	0.8410
		$\mathscr{P}_U$	0.0011	0.0051	0.2023	0.5551	0.8182
	1.50	$\mathscr{P}_1$	0	0	0.2468	0.8007	0.9945
		$\mathscr{P}_2$	3.224×10^{-4}	6.007×10^{-4}	0.1634	0.7088	0.9915
		$\mathscr{P}_3$	0.0573	0.0753	0.2751	0.7098	0.9909
		$\mathscr{P}_4$	0.2344	0.2921	0.5169	0.7649	0.9805
		$\mathscr{P}_5$	0.3149	0.4088	0.5666	0.7891	0.9869
		$\mathscr{P}_U$	0.3738	0.4239	0.6127	0.8277	0.9784

Note: For this study, we assume a non-uniform accrual pattern ($\varphi = 1.25$), maximum accrual time $\omega = 3$ months, administrative censoring time $\tau = 12$ months, and loss to follow-up rate $\upsilon = 15\%$. The fixed sample study was designed assuming exponential survival times, and $n = 50$ subjects were enrolled during the accrual phase.

We also considered a flat uniform prior, denoted by $\mathscr{P}_U$, for the shape parameter. Results for both examples are presented in Tables 3.6 and 3.7. As discussed previously, calculations performed by incorporating the entire posterior distribution of κ tends to yield larger BPP than the posterior

mode based approach. Moreover, it is unsurprising to note that, for each of the 15 simulated data sets, the Bayesian predictive probability tends to be the smallest when the most aggressive/informative gamma prior $\mathscr{P}_1$ is used, and it increases with the use of a more vague prior as it incorporates larger values of κ and therefore longer survival times. In addition, as one might expect, our results corresponding to the vague (non-informative) gamma prior $\mathscr{P}_5$ are mostly similar to the ones obtained using a flat uniform prior $\mathscr{P}_U$. We observe some discrepancies in BPP calculated by using the posterior mode based approach. For instance, in Table 3.6, consider the scenario where simulated data was generated assuming true $\kappa = 0.75$ and the underlying median survival time of 1.00 month, we note that the BPP calculated using the posterior mode based approach for a vague prior turns out to be smaller the one computed using relatively aggressive priors. To avoid such anomalies in our calculations, it is also recommended to use the entire posterior distribution of κ for calculating the BPP.

3.4 Discussion

The fixed sample method, proposed by Waleed et al. [47], for designing single-arm phase II clinical trials with TTE endpoints is appropriate under the assumption that a reliable estimate of the shape parameter is known from some historical studies. Recently, Phadnis et al [37] demonstrated that a reasonable estimate of the shape parameter can be obtained from historical studies with at least 50 subjects and censoring rate close to 20%. There are real life situations, such as studies involving rare diseases, where adequate historical data may not be available for obtaining an estimate of the shape parameter. When no prior information about the shape parameter is available, we could design fixed sample studies assuming exponential survival times as done for the traditional methods available in the literature, and subsequently consider an adjustment to the study sample size using the data accumulated at some pre-specified stage. In this manuscript, we explored the utility of the IPS approach, proposed by Wittes and Brittain [49], for sample size reestimation at an interim stage. It was demonstrated that the power of the study is indeed rescued in our context. We noted that the adjusted sample size using the IPS approach can be more than twice the initially planned

sample size if the shape parameter is grossly misspecified at the design stage, and this may put serious practical constraints to continue with the remaining study. In the future, it would be of interest to compare different sample size reestimation procedures, such as based on the conditional power or Bayesian predictive probability.

In this manuscript, we also discussed the calculation of the Bayesian predictive probability for conducting interim analysis for single-arm phase II clinical trials with TTE endpoints following Weibull distribution with unknown shape parameter. Based on the data accumulated at the interim stage, we propose to generate posterior distributions for both shape as well as scale parameter of the Weibull distribution. The predictive probability of a successful trial outcome can be calculated by either using the posterior mode or the entire posterior distribution of the shape parameter. Although we observed that the BPP calculated using the two approaches tends to differ quantitatively, they yield same qualitative conclusion at the interim stage in most scenarios. It is worth pointing out that the mode based approach may not be appropriate in some circumstances, for instance when the posterior of the shape parameter is flatter or it has heavier tails. Therefore, to appropriately account for uncertainty in the shape parameter, it is recommended to incorporate its entire posterior distribution in our calculations.

Bayesian predictive probabilities are advantageous in the sense that they are easily interpretable, and they can be incorporated in fixed sample designs in a post-hoc manner without an explicit adjustment for repeated significance testing [18, 25, 45]. In this paper, we have used a gamma prior for the shape parameter, as suggested by Ibrahim, Chen and Sinha [51] in the context of the Weibull model. A statistician should utilize any available historical data, and work closely with the clinicians to identify appropriate priors applicable in their area of research. Interested readers can find a discussion on the choice of suitable priors in Dmitrienko and Wang [25].

In comparison to other SC methods such as conditional power and predictive power, a limitation of the Bayesian predictive probabilities is that they require much more computationally intensive calculations due to repeated sampling of the predicted survival data, for the active subjects at the interim, from the posterior predictive distribution [45]. These calculations are becoming increas-

ingly manageable with the advent of sophisticated high performance computing capabilities, and therefore Bayesian predictive probabilities can be utilized to better inform decision making at an interim stage.

Chapter 4

SCM.SurvWeibull: An R Package for Stochastic Curtailment Methods in Single-Arm Clinical Trials with a Survival Endpoint using Weibull Distribution

Abstract

We introduce an R package, SCM.SurvWeibull, to design single-arm clinical trials with a time-to-event endpoint following the Weibull distribution, and to implement stochastic curtailment methods for efficacy or futility testing purposes. When the shape parameter of the Weibull distribution is known from historical studies, the package implements functions to calculate the required sample size of a trial, and executes simulations to evaluate its operating characteristics in terms of the empirical Type-I error and power. In addition, it implements three stochastic curtailment methods (conditional power, predictive power, Bayesian predictive probability) to aid decision making based on data accumulated at an interim stage. When the shape parameter is unknown, this package can also be employed to compute Bayesian predictive probability by incorporating either the posterior mode or the entire posterior distribution of the shape parameter generated at the interim stage. We provide an overview of the methods implemented in the package. Some examples are also presented to demonstrate its usage and functionalities. The package will be made available to the scientific community on the Comprehensive R Archive Network (CRAN) in the near future.

4.1 Introduction

In the recent times, time-to-event (TTE) endpoints such as progression-free survival (PFS) or overall survival (OS) are being considered in the early phase single-arm clinical trials to obtain preliminary evidence of therapeutic effect of novel treatments [5]. To design such trials, a limited number of options primarily based on the log-rank test and its weighted versions, such as Finkelstein et al. [7], Kwak and Jung [8], Sun et al. [9], Wu [10], and Phadnis [6], are available. The methods proposed by Phadnis [6] and Wu [10] have been implemented in the statistical software `PASS` [11] and `nQuery` [12]. Among these, Phadnis' method [6] is appropriate when survival times are assumed to follow the Weibull distribution with known shape parameter (from historical studies), and it can incorporate administrative censoring and an ad-hoc inflation for random loss to follow-up. When the Weibull distribution is appropriate for modeling survival data, Waleed et al. [47] recently proposed a parametric maximum likelihood estimate (MLE) based test which can incorporate flexible design features such as administrative censoring, expected loss to follow-up rate, and different accrual patterns (early, late, or uniform accrual). Thus far, the method proposed by Waleed et al. [47] is not available in any widely used software.

Due to ethical, financial and administrative reasons, group sequential designs (GSD) or stochastic curtailment (SC) methods are often incorporated in clinical trials at an interim stage to obtain early evidence of efficacy or futility of an experimental treatment [17, 18]. In comparison to GSD plans, SC methods are flexible in the sense that it is not uncommon to incorporate them in a fixed sample design in a post-hoc manner without adjusting for repeated testing at the planning phase [18]. Three most popular SC methods are: conditional power (Lan, Simon and Halprin [19], Andersen [20]), predictive power (Spiegelhalter et al. [21]), and Bayesian predictive probability (Herson [22], Geisser [24], Dmitrienko and Wang [25]). Different SC methods for normal and binary endpoints have been studied previously, and implemented in various statistical software such as `PASS` [11], `R` [26], and `SAS` [27]. Conditional power and predictive power calculations based on the two-sample log-rank test are available in `PASS` [11]. For the two-sample scenario in the TTE framework, conditional power can also be calculated using the `CP` package available in `R`

[26, 54]. In the context of single-arm clinical trials with TTE endpoints, it appears that none of the three SC methods have been implemented in any statistical software.

For the benefit of scientific community, we have develpoed an `R` package, `SCM.SurvWeibull`, to design single-arm clinical trials with a TTE endpoint following a Weibull distribution, and to implement the above-mentioned three SC methods in the same context. This package is based on the work of Waleed et al. [47] when the Weibull distribution is appropriate for modeling survival data derived from a single-arm clinical trial, and different design features (such as accrual patterns, loss to follow-up rate, administrative censoring, etc.) are under consideration. It is worth pointing out that the methods proposed by Waleed et al. [47] assume that a reasonably accurate estimate of the shape parameter of the Weibull distribution is available from historical studies. When adequate historical data is not available for estimating the shape parameter, Bayesian predictive probability calculations based on the work of Waleed et al. [55] have also been implemented in the package.

The primary objective of this manuscript is to highlight the main functionalities and usage of the package. For the sake of completeness, we provide a brief theoretical overview of the methods proposed by Waleed et al. [47, 55] in Section 4.2 which have been implemented in this package. We present various functions available in the package and demonstrate their usage with some examples in Section 4.3. Some concluding remarks are presented in Section 4.4.

4.2 Methods

4.2.1 Notation and Mathematical Preliminaries

Suppose that n subjects are enrolled in a single-arm phase II clinical trial with a TTE primary endpoint. We further suppose that the maximum accrual time and final study time on the calendar time scale are denoted by ω and τ, respectively. For the ith subject:

- The time of enrollment into the clinical trial, E_i, is assumed to follow a general form of uniform distribution with an additional power parameter φ with the following probability

density function (pdf) :

$$f_{E_i}(a) = \frac{\varphi a^{\varphi-1}}{\omega^{\varphi}}, \qquad \text{where: } a \in [0,\omega],\ \varphi > 0. \tag{4.1}$$

It is worth noting that a uniform accrual pattern can be incorporated by setting $\varphi = 1$, and very early (late) accrual pattern can be incorporated by choosing $\varphi \to 0$ ($\varphi \gg 1$).

In the package, users may also specify E_i to follow the truncated exponential distribution with rate parameter λ having the pdf:

$$f_{E_j}(a) = \frac{\lambda e^{-\lambda a}}{1-e^{-\lambda\omega}}, \qquad \text{where: } a \in [0,\omega], \lambda > 0. \tag{4.2}$$

- The amount of time to event, Y_i, is assumed to follow the Weibull distribution having shape parameter κ and scale parameter θ with its pdf expressed as below:

$$f_{Y_i}(y) = \frac{\kappa}{\theta^{\kappa}} y^{\kappa-1} \exp\left\{-\left(\frac{y}{\theta}\right)^{\kappa}\right\}, \qquad \text{where: } y > 0, \kappa > 0, \theta > 0. \tag{4.3}$$

- The amount of time to random loss to follow-up, C_i, also follows the Weibull distribution with the same shape parameter κ and scale parameter $\eta = \theta\left(\frac{1-\upsilon}{\upsilon}\right)^{1/\kappa}$ that ensures the expected loss to follow-up rate υ [38, 47].

- When $E_i \sim$ Gen-Uniform$(0,\omega,\varphi)$, the amount of time from enrollment to being lost to follow-up or administrative censoring, $Z_i := \min(\max(0,\tau-E_i), C_i)$, has the pdf:

$$f_{Z_i}(z) = \begin{cases} \dfrac{\kappa}{\eta^{\kappa}} z^{\kappa-1} \exp\left\{-\left(\dfrac{z}{\eta}\right)^{\kappa}\right\} & \text{if } z \in [0, \tau-\omega) \\ \left(\dfrac{\varphi}{\tau-z} + \dfrac{\kappa}{\eta^{\kappa}} z^{\kappa-1}\right)\left(\dfrac{\tau-z}{\omega}\right)^{\varphi} \exp\left\{-\left(\dfrac{z}{\eta}\right)^{\kappa}\right\} & \text{if } z \in [\tau-\omega, \tau] \\ 0 & \text{otherwise} \end{cases} \tag{4.4}$$

On the other hand, when $E_i \sim \text{Trunc-Exp}(\lambda, \omega)$, the pdf of Z_i can be expressed as:

$$f_{Z_i}(z) = \begin{cases} \dfrac{\kappa}{\eta^\kappa} z^{\kappa-1} \exp\left\{-\left(\frac{z}{\eta}\right)^\kappa\right\} & \text{if } z \in [0, \tau-\omega) \\ \left(\dfrac{\exp\left\{-\left(\frac{z}{\eta}\right)^\kappa\right\}}{1-e^{-\lambda\omega}}\right)\left(\left(1-e^{-\lambda(\tau-z)}\right)\dfrac{\kappa z^{\kappa-1}}{\eta^\kappa} + \lambda e^{-\lambda(\tau-z)}\right) & \text{if } z \in [\tau-\omega, \tau] \\ 0 & \text{otherwise} \end{cases} \tag{4.5}$$

The random loss to follow-up is assumed to be non-informative of the survival process, and therefore $\{Y_i, Z_i, i = 1, \ldots, n\}$ are independent and identically distributed random variables. Thus, we have n pairs of data $\{(X_i, \delta_i), i = 1, \ldots, n\}$, where $X_i := \min(Y_i, Z_i)$ is the survival time for the ith subject, and $\delta_i := \mathbb{1}_{(Y_i < Z_i)}$ is its survival status which equals 1 if $Y_i < Z_i$, and 0 otherwise.

4.2.2 Fixed Sample Design

In their work, Waleed et al. [47] assume that adequate historical data is available to provide a reasonably accurate estimate of the shape parameter of the Weibull distribution. Suppose that $M = \theta(\log(2))^{1/\kappa}$ denotes the median of the Weibull(κ, θ) distribution. Then, our hypotheses can be expressed as:

$$\begin{array}{ccccc} H_0 : M \leq M_0 & & H_0 : \theta \leq \theta_0 & & H_0 : \gamma \leq \gamma_0 \\ & \equiv & & \equiv & \\ H_1 : M > M_0 & & H_1 : \theta > \theta_0 & & H_1 : \gamma > \gamma_0 \end{array} \tag{4.6}$$

The above equivalence holds because covariates are usually introduced [36] into the parametric survival models through the scale parameter as $\theta = \exp\{\boldsymbol{\gamma}^T \boldsymbol{x}\}$, where: $\boldsymbol{x} = (1, x_1, \ldots, x_k)^T$ and $\boldsymbol{\gamma}^T = (\gamma_0, \gamma_1, \ldots, \gamma_k)$ respectively denote the vectors of $k+1$ covariates and the corresponding parameters. Further, in the context of a single-arm trial with a TTE endpoint, we assume that the experimental treatment administered to the subjects is the only covariate introduced into the

model so that $\theta = \exp\{\gamma\}$.

To test the hypotheses in Eq. (4.6), Waleed et al. [47] proposed a Wald's test statistic based on the asymptotic approximation of the MLE of the parameter γ. The sample size needed for a single-arm clinical trial with a Weibull distributed TTE endpoint to detect a clinically meaningful difference ε with Type-I error α and power $1-\beta$ satisfies:

$$n = \sigma_1^2 \cdot \left(\frac{\Phi^{-1}(1-\beta) + \Phi^{-1}(1-\alpha)}{\varepsilon} \right)^2, \tag{4.7}$$

where $\sigma_1^2 = \left[\kappa^2 \left(1 - E_{Z_1} \left(\exp\left\{ - \left(\frac{Z_1}{\exp\{\gamma_1\}} \right)^{\kappa} \right\} \right) \right) \right]^{-1}$ is the plug-in estimator of σ^2 under $H_1 : \gamma > \gamma_1 (= \gamma_0 + \varepsilon)$, and $\Phi^{-1}(\cdot)$ denotes the inverse cumulative distribution function (cdf) of the standard normal distribution.

Depending on the choice of the accrual distribution, the appropriate pdfs of Z_i are provided in Eqs. (4.4) and (4.5). Numerical methods can be implemented to calculate the value of σ_1^2.

4.2.3 Stochastic Curtailment Methods

In this section, we briefly touch upon three SC methods: conditional power, predictive power (mixed Bayesian-frequentist approach), and Bayesian predictive probability (purely Bayesian approach), which can be employed to determine if there is sufficient evidence in favor of the null hypothesis to 'curtail' sampling beyond an interim analysis [17, 18].

4.2.3.1 Conditional Power

Conditional power is a purely frequentist SC approach which relies on "predicting the distribution of the final outcome given the data already observed in the study [18]." For a single-arm clinical trial designed to test the hypotheses in Eq. (4.6) with an α level of significance, the conditional power [17, 18, 19, 47] at an interim stage $k = 1, \ldots, K-1$ is given as:

$$P_k(\gamma) = \Phi\left(\frac{Z_k \sqrt{\mathscr{I}_k} + (\gamma - \gamma_0)(\mathscr{I}_K - \mathscr{I}_k) - Z_{1-\alpha}\sqrt{\mathscr{I}_K}}{\sqrt{\mathscr{I}_K - \mathscr{I}_k}} \right), \tag{4.8}$$

where Z_k is the observed test statistic at the interim stage k, $Z_{1-\alpha} = \Phi^{-1}(1-\alpha)$ is the upper α-quantile of the standard normal distribution, and $\mathscr{I}_k$ ($\mathscr{I}_K$) denote the information at the interim stage k (trial end at stage K).

To compute the conditional power for a fixed sample design using Eq. (4.8), the final information $\mathscr{I}_K$ can be obtained using the variance formula in the asymptotic distribution of the parameter γ in Waleed et al. [47], and $\mathscr{I}_k = \mathscr{F}_k \times \mathscr{I}_K$ where $\mathscr{F}_k$ is the proportion of the total information $\mathscr{I}_K$ available at the interim stage k.

Following decision rules have been recommended in the literature [17, 18, 19] for early termination of a clinical trial using the conditional power approach:

- Reject H_0 in favor of H_1 at the stage k if $P_k(\gamma_0) \geq \zeta$ for some $\zeta \in [0.80, 1]$ as it implies that the test is unlikely to accept H_0 given the accumulated data, even if H_0 is assumed to be true.

- Fail to reject H_0 at stage k if the *futility index* $1 - P_k(\gamma_1) \geq \zeta'$, for some $\zeta' \in [0.80, 1]$. This is because a large value of the *futility index* implies that the test is unlikely to reject H_0 given the current data at stage k, even if H_1 is true.

4.2.3.2 Predictive Power

Predictive power is a mixed Bayesian-frequentist SC method in which the conditional power (frequentist component) is averaged over the posterior distribution (Bayesian component) of the parameter γ given its estimate $\widehat{\gamma}_k$ at stage k [21]. It is given as:

$$P_k = \int P_k(\gamma)\pi(\gamma|\widehat{\gamma}_k, \kappa)d\gamma \tag{4.9}$$

where $P_k(\gamma)$ denotes the conditional power function in Eq. (4.8), and $\pi(\gamma|\widehat{\gamma}_k, \kappa)$ is the posterior distribution of γ given its estimate $\widehat{\gamma}_k$ at stage k.

As suggested by Waleed et al. [47], posterior samples from $\pi(\gamma|\widehat{\gamma}_k, \kappa)$ can be generated by either implementing the Metropolis-Hastings (MH) algorithm, or a purely simulation based via `R2OpenBUGS` package available in `R` [26, 52]. Similar to the decision rules for conditional power,

clinical trials can be terminated early to reject H_0 (fail to reject H_0) if the predictive power is greater than 0.80 (less than 0.20) [17, 18].

4.2.3.3 Bayesian Predictive Probability

As the name suggests, Bayesian predictive probability is a fully Bayesian SC approach "which relies on the idea of *predictive probability* of obtaining a positive trial outcome at the end of the study, given the current estimates at an interim stage [18, 25, 47]." At an interim stage k of a study designed to test the alternatives $H_0 : \gamma \leq \gamma_0$ vs. $H_1 : \gamma > \gamma_1 (= \gamma_0 + \varepsilon)$, where $\varepsilon > 0$ is a clinically meaningful difference, suppose that $n - m$ subjects had already experienced an event or were lost to follow-up, and the remaining m subjects were still active in the study. The MLE of γ at the final time τ can be expressed as:

$$\widehat{\gamma}_K = \frac{1}{\kappa} \log \left(\frac{\mathscr{T}_k + \mathscr{T}_{K-k}}{\mathscr{D}_k + \mathscr{D}_{K-k}} \right) = \frac{1}{\kappa} \log \left(\frac{\mathscr{T}_k + m\overline{\mathscr{T}}_{K-k}}{\mathscr{D}_k + m\overline{\mathscr{D}}_{K-k}} \right). \tag{4.10}$$

where $\mathscr{T}_k = \sum_{i=1}^{n-m} X_i^{\kappa}$ and $\mathscr{D}_k = \sum_{i=1}^{n-m} \delta_i$ are the quantities for subjects who had experienced an event or were lost to follow-up; and $\overline{\mathscr{T}}_{K-k} = m^{-1} \sum_{j=1}^{m} X_j^{\kappa}$ and $\overline{\mathscr{D}}_{K-k} = m^{-1} \sum_{j=1}^{m} \delta_j$ correspond to the active subjects observed between the interim stage k (at calendar time ℓ_k), and the final stage K (at calendar time τ).

Since the random vector $P = \left(\overline{\mathscr{T}}_{K-k}, \overline{\mathscr{D}}_{K-k} \right)$ is not observable at the interim stage k, we assume that $\tilde{P}$ represents its predicted value. Then the *predictive probability* of a positive trial outcome at the final study time τ is expressed as [18, 25, 47]:

$$P_k = \int \mathbb{1}_{(\mathscr{Q} > \eta^*)} \, d\tilde{P} \tag{4.11}$$

where $\mathscr{Q} = \text{Prob}\Big(\gamma > \gamma_1 | \mathscr{T}_k, \tilde{\overline{\mathscr{T}}}_{K-k}, \mathscr{D}_k, \tilde{\overline{\mathscr{D}}}_{K-k}, \kappa \Big)$, η^* is some pre-specified threshold probability of a successful trial outcome, and $\tilde{P}$ is the posterior predictive distribution of $\overline{\mathscr{T}}_{K-k}$ and $\overline{\mathscr{D}}_{K-k}$. A detailed derivation of the asymptotic posterior predictive distribution of $\tilde{P}$ is provided in Waleed et al [47].

Let $\boldsymbol{X}_{(n-m);obs}$ and $\boldsymbol{\delta}_{(n-m);obs}$ denote the vectors containing the survival times and the corresponding survival status for the $n-m$ subjects observed by the interim look time ℓ_k, respectively. Similarly, $\boldsymbol{X}_{m;pred}$ and $\boldsymbol{\delta}_{m;pred}$ denote the predicted survival times and survival status for the m active subjects, respectively. The Bayesian predictive probability of a successful trial outcome in Eq. (4.11) can be equivalently defined as follows:

$$P_k = \int \mathbb{1}_{(\mathscr{Q}^* > \eta^*)}\, \pi(\gamma|\widehat{\gamma}_k, \kappa) d\gamma \tag{4.12}$$

where: $\mathscr{Q}^* = \text{Prob}\big(\gamma > \gamma_1 | \boldsymbol{X}_{(n-m);obs}, \boldsymbol{X}_{m;pred}, \boldsymbol{\delta}_{(n-m);obs}, \boldsymbol{\delta}_{m;pred}, \kappa\big)$, η^* is some pre-specified threshold probability of a successful trial outcome, and $\pi(\gamma|\kappa, \widehat{\gamma}_k)$ is the posterior distribution of γ generated using the data accumulated at the interim stage k.

Algorithm 1 has been implemented in the package to calculate Bayesian predictive probability defined in Eq. (4.11). Algorithm 2 has been implemented in the package to calculate Bayesian predictive probability, defined in Eq. (4.12), using the purely simulation-based approach via `R2OpenBUGS` package [52] in `R` [26].

Algorithm 1 Bayesian predictive probability using the asymptotic posterior predictive distribution of $\tilde{P}$ (Waleed et al. [47])

1: Use the MH algorithm to generate the asymptotic posterior distribution of γ as:

$$\pi(\gamma|\widehat{\gamma}_k, \kappa) \propto \pi(\widehat{\gamma}_k|\gamma, \kappa)\pi(\gamma).$$

2: Use numerical integration to obtain and normalize the asymptotic posterior predictive distribution of $\tilde{P}$ given as:

$$\pi\big(\tilde{P}|\widehat{\gamma}_k, \kappa\big) = \int \pi\big(\tilde{P}|\gamma, \kappa\big)\pi(\gamma|\widehat{\gamma}_k, \kappa) d\gamma.$$

3: For each combination of $\tilde{\overline{\mathscr{D}}}_{K-k}$ and $\tilde{\overline{\mathscr{T}}}_{K-k}$ in the support of $\pi\big(\tilde{P}|\widehat{\gamma}_k, \kappa\big)$:

(i) Generate the posterior distribution of γ at final stage K.

(ii) Calculate the quantity $\mathscr{Q} = \text{Prob}\Big(\gamma > \gamma_1 \,|\, \mathscr{T}_k, \tilde{\overline{\mathscr{T}}}_{K-k}, \mathscr{D}_k, \tilde{\overline{\mathscr{D}}}_{K-k}, \kappa\Big)$.

4: Estimate the Bayesian predictive probability of a successful trial outcome as:

$$P_k = \int \mathbb{1}_{(\mathscr{Q} > \eta^*)}\, d\tilde{P} \approx \sum_{\tilde{P}} \mathbb{1}_{(\mathscr{Q} > \eta^*)} \pi\big(\tilde{P}|\widehat{\gamma}_k, \kappa\big).$$

Algorithm 2 Bayesian predictive probability using the purely simulation-based approach via `R2OPenBUGS` package in `R` (Waleed et al. [47])

1: For the data accumulated by the kth interim stage (at look time ℓ_k):

(i) Use `R2OpenBUGS` to generate samples from $\pi(\gamma|\widehat{\gamma}_k,\kappa)$ via Gibbs sampling.

(ii) Normalize $\pi(\gamma|\widehat{\gamma}_k,\kappa)$.

(iii) For the support set $\mathscr{X}=(\gamma_{\min},\gamma_{\max})$ of $\pi(\gamma|\widehat{\gamma}_k,\kappa)$, create a sequence of γ of length H as: $\gamma_1=\gamma_{\min}<\gamma_2<\cdots<\gamma_{\max}=\gamma_H$

2: **for** $h=1$ to H **do**

3: **for** $i=1$ to I **do**

4: **for** $j=1$ to m **do**

5: Generate predicted survival time and survival status for the jth subject in the ith predicted sample corresponding to the hth value of γ as:

$$\tilde{Y}_{h,i,j}=\text{Trunc-Weibull}\big(\text{shape }=\kappa,\text{scale }=e^{\gamma_h},\min=\ell_k-E_j\big)$$

$$\tilde{C}_{h,i,j}=\text{Trunc-Weibull}\left(\text{shape }=\kappa,\text{scale }=e^{\gamma_h}\left(\frac{1-\upsilon}{\upsilon}\right)^{1/\kappa},\min=\ell_k-E_j\right)$$

$$\tilde{X}_{h,i,j}=\min\left(\max\left(0,\tau-E_j\right),\tilde{Y}_{h,i,j},\tilde{C}_{h,i,j}\right)$$

$$\tilde{\delta}_{h,i,j}=\mathbb{1}_{\left(\tilde{Y}_{h,i,j}<\min\left(\max\left(0,\tau-E_j\right),\tilde{C}_{h,i,j}\right)\right)}$$

6: **end for**

7: Using the ith full data that comprises of the observed subjects and the ith predicted data for the active subjects, generate the posterior distribution of γ and determine whether

$$\text{Prob}\big(\gamma>\gamma_1|\boldsymbol{X}_{(n-m);obs},\boldsymbol{X}_{m;pred_i},\boldsymbol{\delta}_{(n-m);obs},\boldsymbol{\delta}_{m;pred_i}\big)>\eta^*.$$

8: **end for**

9: Estimate Prob(Successful Outcome$|\gamma_h$) as the proportion of I predicted samples which satisfy the condition in Step 7.

10: **end for**

11: Estimate the Bayesian predictive probability of a successful trial outcome as:

$$P_k=\int\mathbb{1}_{(\mathscr{Q}^*>\eta^*)}\,\pi(\gamma|\widehat{\gamma}_k,\kappa)d\gamma\approx\sum_{h=1}^{H}\text{Prob(Successful Outcome}|\gamma_h)\pi(\gamma_h|\widehat{\gamma}_k,\kappa).$$

In practice, the threshold level η^* is usually set between 0.90 and 0.975. After consultation with the Data Safety Monitoring Board (DSMB) overseeing the clinical trial, researchers may terminate the trial at an interim stage k to reject H_0 if $P_k \geq \zeta$ for some $\zeta \in [0.8, 1]$, and conclude futility (i.e., fail to reject H_0) if $P_k \leq \zeta'$ for some $\zeta' \in [0, 0.2]$. See Dmitreinko and Wang for further details [25].

Both Algorithms 1 and 2 assume that a reliable estimate of the shape parameter κ is available from some historical studies. Very recently, Phadnis et al. [37] demonstrated that the shape parameter of the Weibull distribution is reasonably accurate when it is estimated from a historical data comprising of at least 50 subjects and censoring rate close to 20%. When a sufficiently large historical dataset is not available as prescribed by Phadnis et al. [37], such as in the case of clinical trials concerning small population sizes, an independent joint prior specification of γ and κ can be considered [51]. Assuming the prior distributions $\gamma \sim \text{Normal}\big(\mu_0, \sigma_0^2\big)$ and $\kappa \sim \text{Gamma}(\alpha_0, \beta_0)$, Waleed et al. proposed two approaches for calculating the Bayesian predicted probability [55].

In the first approach, the predicted data for active subjects is generated using the posterior mode, κ_{mode}, of $\pi(\kappa|\gamma, \text{Data})$ generated at the interim stage. Thus, the Bayesian predictive probability is defined as below:

$$P_{k,\kappa_{\text{mode}}} = \int \mathbb{1}_{(\mathscr{Q}_{\text{mode}} > \eta^*)}\ \pi(\gamma|\kappa, \widehat{\gamma}_k) d\gamma \tag{4.13}$$

where $\mathscr{Q}_{\text{mode}} = \text{Prob}\big(\gamma > \gamma_1 | \boldsymbol{X}_{(n-m);obs}, \boldsymbol{X}_{m;pred}, \boldsymbol{\delta}_{(n-m);obs}, \boldsymbol{\delta}_{m;pred}, \kappa_{\text{mode}}\big)$. Implementation of this approach is essentially the same as the one presented in Algorithm 2 with κ replaced with the posterior mode κ_{mode} in Step 5.

In the second approach, the Bayesian predictive probability is calculated as a weighted average over the entire posterior distribution $\pi(\kappa|\widehat{\kappa}_k)$ as expressed below:

$$P_{k,\kappa_{\text{full}}} = \int P_k \cdot \pi(\kappa|\widehat{\kappa}_k)\ d\kappa \tag{4.14}$$

where P_k is defined in Eq. (4.12). Implementation of this approach requires the execution of Steps

2 – 11 in Algorithm 2 for all κ in the support of $\pi(\kappa|\widehat{\kappa}_k)$, and taking weighted average of the results.

4.3 Description of the Package and Examples

The `devtools` package in `R` was used to build the `SCM.SurvWeibull` package that will be made available on the Comprehensive R Archive Network (CRAN) in the future [56, 57]. By default, the `SCM.SurvWeibull` package imports the following `R` packages for its successful execution: `R2OpenBUGS` [52], `survival` [58], `stats` [59], `RGeode` [60], `mvtnorm` [61], `truncdist` [62]. In the current version, it contains seven core functions to implement the methods proposed by Waleed et al. [47, 55]: `n_calc`, `designProps`, `stats_calc`, `CP_calc`, `PP_calc`, `BPP_calc`, `BPP_shape_calc`. In this section, we demonstrate the usage of these functions with some hypothetical examples.

n_calc

This function is used to calculate the sample size needed for a single-arm clinical trial with various design features to detect some clinically meaningful difference with a desired Type-I error rate α and power $1-\beta$. The three parameters of interest in Eq. (4.6) are available as the input: `type =` "`median`", "`scale`", or "`gamma`". Since the shape parameter of the Weibull distribution is assumed to be known, the user can input the values under the null and alternative hypotheses for any parameter of interest, and they are automatically converted to the parameter γ for internal calculations to be consistent with Waleed et al [47]. The user can input the desired distribution for the accrual pattern as: `dist.accrual =` "`GenUnif`" for the generalized-uniform distribution, or "`TExp`" for the truncated-exponential distribution. The input `param.accrual` represents the power (rate) parameter of the accrual pattern following the generalized-uniform (truncated-exponential) distribution.

Example 1: Suppose that we are interested in testing the alternatives $H_0 : M < 1$ month vs.

$H_1 : M \geq 1.50$ months at 5% level of significance and 90% power. Further suppose that the fixed sample study is envisioned with a maximum accrual period of 3 months and study end time at 12 months. Also suppose that a uniform accrual pattern is anticipated, the expected loss to follow-up rate is 15%, and the known shape parameter of the Weibull distribution is 1.25. The required sample size can be computed as below:

```
> # Load the package into R environment
> library("SCM.SurvWeibull")
>
> # Define all relevant inputs for our design
> param    <- "median"            # Parameter to be tested
> H0_val   <- 1.00                # Value of parameter under H0
> H1_val   <- 1.50                # Value of parameter under H1
> kappa    <- 1.25                # Shape parameter of the Weibull dist.
> alpha    <- 0.05                # Desired Type-I error
> pow      <- 0.90                # Desired power
> accType <- "GenUnif"            # Type of accrual distribution
> omega    <- 3.00                # Maximum accrual time
> phi      <- 1.00                # Additional parameter of accrual dist.
> tau      <- 12.00               # Final study time
> nu       <- 0.15                # Expected loss to follow-up rate
>
> # Compute the desired sample size and print result
> n <- SCM.SurvWeibull::n_calc(type = param, H0 = H0_val, H1 = H1_val,
+ shape = kappa, sig.level = alpha, power = pow, dist.accrual = accType,
+ max.accrual = omega, param.accrual = phi, study.end = tau, nu = nu)
>
```

```
> print(n)
   [1] 40
```

designProps

This function executes simulations to evaluate the operating characteristics of the proposed method (Waleed et al. [47]) in terms of the empirical Type-I error rate and power. In comparison to the function `n_calc`, this function requires two additional inputs: the number of simulations to be executed, `n.sim`, and the study sample size, `n.size`. If `n.size` is not provided as an input, this function automatically computes the desired sample size and executes the simulation study. The function outputs a data frame with the sample size (`n`), empirical Type-I error rate (`T1.error`), and power (`power`).

Example 2: Suppose that we want to compute the sample size, and execute 10,000 simulations to evaluate the empirical Type-I error rate and power for the same study described in Example 1. This can be achieved as follows:

```
> set.seed(413412)
> MyProps <- SCM.SurvWeibull::designProps(n.sim = 10000, n.size = NA,
+ type = param, H0 = H0_val, H1 = H1_val, shape = kappa,
+ sig.level = alpha, power = pow, dist.accrual = accType,
+ max.accrual = omega, param.accrual = phi, study.end = tau, nu = nu)
>
> print(MyProps)
        n   T1.error    power
   [1] 40     0.0427   0.8834
```

For the above example, suppose that we now want to evaluate the operating characteristics

when the study sample size `n.size = 25`. From the results below, we note that the empirical Type-I error rate is controlled with the desired level, but the study is grossly under-powered for this sample size.

```
> set.seed(413412)
> MyProps_25 <- SCM.SurvWeibull::designProps(n.sim = 10000, n.size = 25,
+ type = param, H0 = H0_val, H1 = H1_val, shape = kappa,
+ sig.level = alpha, power = pow, dist.accrual = accType,
+ max.accrual = omega, param.accrual = phi, study.end = tau, nu = nu)
>
> print(MyProps_25)
        n   T1.error    power
   [1] 25     0.0415   0.7306
```

stats_calc

This function fits the parametric Weibull model on an input dataset, and outputs a data frame with the fitted values of all three parameters (M as `median`, θ as `scale`, γ as `gamma`) in Eq. (4.6), shape of the fitted Weibull distribution (`shape`), estimated variance of γ (`var.est`), and the Wald's test statistic (`z.stat`) in Waleed et al [47]. It is important to note that the names of the columns containing the subject survival times and the corresponding survival status in the input data frames should be `SurvTime` and `SurvStatus`, respectively, for all functions requiring an input dataset in the package.

Example 3: For the fixed sample study in Example 1, we simulate a hypothetical survival data corresponding to the interim look time equal to 4 months with the assumption that the true value of the median survival time is 2 months. For simulating hypothetical survival datasets in our context,

we have also created a function `example_data` in the package. We can implement the function `stats_calc` as below:

```
> set.seed(413412)
>
> # Simulate data at time = 4 months assuming the median time = 2 months
> MyData <- SCM.SurvWeibull::example_data(n.size = 40, type = param,
+ value = 2, shape = kappa, dist.accrual = accType, max.accrual = omega,
+ param.accrual = phi, nu = nu, study.end = tau, look.time = 4)
>
> # Obtain fitted values of the parameters and Wald's test statistic
> MyStats <- SCM.SurvWeibull::stats_calc(interim.data = MyData,
+ type = param, H0 = H0_val, shape = kappa, dist.accrual = accType,
+ max.accrual = omega, param.accrual = phi, nu = nu, look.time = 4)
>
> print(MyStats)
          gamma     scale    median     shape     var.est   z.stat
   [1] 0.9633494  2.620459  1.954509  1.257842  0.02942184  3.90688
```

The large value of the test statistic in the above output suggests that the study might be terminated early to reject the null hypothesis.

CP_calc

As the name suggests, this function can be used to calculate and output the conditional power under three scenarios ($H_0, H_1,$ MLE of γ) for an input dataset.

Example 4: For the study design and hypothetical data generated at the look time 4 months, we

compute and output results of the `CP_calc` function as below:

```
> # Calculate conditional power for the data generated in Example 3
> MyCP <- SCM.SurvWeibull::CP_calc(interim.data = MyData, type = param,
+ H0 = H0_val, H1 = H1_val, shape = kappa, sig.level = alpha, power = pow,
+ dist.accrual = accType, max.accrual = omega, param.accrual = phi,
+ nu = nu, study.end = tau, look.time = 4)
>
> print(MyCP)
        Scenario        CP
   [1]        H0    0.7728
   [2]        H1    0.9992
   [3]       MLE    1.0000
```

Large values of the conditional power under all three scenarios suggests early termination of the clinical trial to reject the null hypothesis.

PP_calc

This function can be used to calculate the predictive power for an input dataset. The posterior distribution of γ can be generated using the input `sim.method = "bugs"` (Gibbs sampling using `R2OpenBUGS` package) or `sim.method = "mh"` (implementation of the MH algorithm). The user must also specify the number of equally-spaced γ, `n.gamma`, to be considered in the support of its posterior distribution. The output of this function consists of a list object with three components:

1. a numeric variable `PP` containing the value of the predictive power
2. a data frame having the following five columns:

   (i) `gamma`: the sequence of γ in $\pi(\gamma|\widehat{\gamma}_k, \kappa)$

(ii) `post.pdf`: the value of posterior density $\pi(\gamma|\widehat{\gamma}_k, \kappa)$ for each γ

(iii) `std.post.pdf`: the normalized posterior density for each γ

(iv) `delta`: the corresponding values $\Delta = \gamma - \gamma_0$

(v) `cp`: the corresponding value of conditional power

3. a list object `Post_Data` containing the output of Gibbs sampling using `R2OpenBUGS` package, or the posterior samples generated using the MH-algorithm.

Example 5: We continue working with the hypothetical survival data generated in Example 3 at look time 4 months, and calculate the predictive power using the `bugs` approach:

```
>
> sim.method <- "bugs"       # Method for generating the posterior samples
> n.iter     <- 110000       # Number of posterior samples to be generated
> n.burn     <- 10000        # Number of burn-in samples
> n.thin     <- 10           # Thinning parameter
> n.gamma    <- 100          # Number of gamma to be considered
> MyModel    <- "C:/Mdl.txt" # Path of the Weibull model in bugs format
>
> # Calculate predictive power for the data generated in Example 3
> MyPP <- SCM.SurvWeibull::PP_calc(interim.data = MyData, type = param,
+ H0 = H0_val, H1 = H1_val, shape = kappa, sig.level = alpha, power = pow,
+ dist.accrual = accType, max.accrual = omega, param.accrual = phi,
+ nu = nu, study.end = tau, look.time = 4, sim.method, n.iter, n.burn,
+ n.thin, n.gamma, model.path = MyModel)
>
```

```
> print(MyPP$PP)       # Print value of the predictive power
   [1] 0.9998
```

The large value of the predictive power also suggests an early termination of the trial to conclude efficacy of the experimental treatment.

BPP_calc

This function is used to calculate the Bayesian predictive probability of a successful trial outcome based on the data accumulated at an interim stage. In addition to the same inputs as the `PP_calc` function, users must also specify the number of predicted datasets, `n.pred.data`, to be generated when computing the Bayesian predictive probability using the `bugs` approach. In this case, the output of this function consists of a list object with five components:

1. a numeric variable `BPP` containing the value of the Bayesian predictive probability
2. `seq.gamma`: a numeric vector containing the sequence of γ in $\pi(\gamma|\widehat{\gamma}_k, \kappa)$
3. `std.pdf.gamma`: a numeric vector containing the normalized posterior density for each γ in the vector `seq.gamma`
4. `reject.prob`: an `n.pred.data` $\times$ `n.gamma` matrix containing the posterior probability of rejecting H_0 for each of the predicted sample for each γ in the vector `seq.gamma`
5. `BPP.gamma`: a numeric vector containing the probability of successful trial outcome for each γ in the vector `seq.gamma`

Example 6: We continue with the inputs defined in Example 5, and compute the Bayesian predictive probability using the `bugs` approach for the data generated in Example 3.

```
> eta         <- 0.95   # Threshold probability of a successful outcome
```

```
> n.pred.data <- 250   # Number of predicted data sets to be generated
> MyProgress  <- "Yes" # Whether or not print progress of the algorithm
>
> # Calculate the Bayesian predictive probability using R2OpenBUGS
> MyBPP <- SCM.SurvWeibull::BPP_calc(interim.data = MyData, type = param,
+ H0 = H0_val, H1 = H1_val, shape = kappa, dist.accrual = accType,
+ max.accrual = omega, param.accrual = phi, nu = nu, study.end = tau,
+ look.time = 4, sim.method, n.iter, n.burn, n.thin, n.gamma,
+ model.path = MyModel, eta, n.pred.data, print.progress = MyProgress)
>
> print(MyBPP$BPP)    # Print value of the Bayesian predictive probability
   [1] 0.7526
```

In consultation with the DSMB overseeing the clinical trial, the study investigators may decide to terminate the trial at the interim look time to reject the null hypothesis.

BPP_shape_calc

This function is appropriate for computing the Bayesian predictive probability when no reasonably accurate estimate of the shape parameter is available from historical data. In this case, we also generate the posterior distribution of the shape parameter at the interim stage, and subsequently compute the predictive probability using either of the two approaches proposed by Waleed et al [55]. A user can implement the posterior mode based and weighted average approach by inputting `shape.method ="mode"` and `"dist"`, respectively. The output of this function consists of a list object with six components:

1. a numeric variable `BPP` containing the value of the Bayesian predictive probability
2. `seq.gamma`: a numeric vector containing the sequence of γ in $\pi(\gamma|\kappa, \text{Data})$

3. `std.pdf.gamma`: a numeric vector containing the normalized posterior density for each γ in the vector `seq.gamma`

4. `seq.shape`: a numeric vector containing the sequence of κ in $\pi(\kappa|\gamma, \text{Data})$

5. `std.pdf.shape`: a numeric vector containing the normalized posterior density for each κ in the vector `seq.shape`

6. `BPP.shape`: a numeric vector (with same length as `seq.shape`) containing the Bayesian predictive probability of a successful trial outcome corresponding to each κ in the vector `seq.shape`

We feel important to point out that when no reasonable estimate of the shape parameter is available, researchers can calculate the study sample size at the design stage under the assumption of exponential survival times ($\kappa = 1$), and then reestimate it at the interim stage, if necessary.

Example 7: Suppose that the shape parameter was unknown while designing the study in Example 1, and investigators calculated the required sample size assuming exponential survival times ($\kappa = 1$). We compute the Bayesian predictive probability using the weighted average approach for a hypothetical dataset at look time 4 months generated under the assumption that the true median survival time and shape parameter of the Weibull distribution were 2.00 months and 1.25, respectively.

```
> # Compute the initial sample size of the study in Example 1 assuming
> # exponential survival time, and printing the result
> n.new <- SCM.SurvWeibull::n_calc(type = param, H0 = H0_val, H1 = H1_val,
+ shape = 1.00, sig.level = alpha, power = pow, dist.accrual = accType,
+ max.accrual = omega, param.accrual = phi, study.end = tau, nu = nu)
>
```

```
> print(n.new)
   [1] 62
>
> set.seed(413412)
>
> # Simulate data at time = 4 months with M = 2 months and κ = 1.25
> NewData <- SCM.SurvWeibull::example_data(n.size = 62, type = param,
+ value = 2, shape = 1.25, dist.accrual = accType, max.accrual = omega,
+ param.accrual = phi, nu = nu, study.end = tau, look.time = 4)
>
> # Path of the bugs Weibull model for generating samples for γ and κ
> mod_int    <- "C:/model_interim.txt"
>
> # Compute Bayesian predictive prob.  using the full dist.  approach
>
> MyBPP_full <- SCM.SurvWeibull::BPP_shape_calc(interim.data = NewData,
+ type = param, H1 = H1_val, shape = 1.00, dist.accrual = accType,
+ max.accrual = omega, param.accrual = phi, nu = nu, study.end = tau,
+ look.time = 4, sim.method, n.iter = 11000, n.burn = 1000, n.thin = 1,
+ n.gamma = 30, gamma.init = 0, n.shape = 30, shape.init = 1.00,
+ shape.method = "dist", model.interim = mod_int , model.final = MyModel,
+ eta, n.pred.data, print.progress = MyProgress)
>
> # Print value of the Bayesian predictive probability
> print(MyBPP_full$BPP)
   [1] 0.9999
```

For the posterior mode based approach (`shape.method = "mode"`), the Bayesian predictive probability turned out to be 1.0000. Since the Bayesian predictive probability is almost 1 in both cases, the trial can be terminated at the interim look time to reject the null hypothesis.

4.4 Conclusions

The `SCM.SurvWeibull` package offers a free and accessible implementation of the methods proposed by Waleed et al. [47, 55]. Based on the MLE based test by Waleed et al. [47], the package can be used to design single-arm clinical trials with a TTE endpoint following a Weibull distribution with a variety of design features, such as expected loss to follow-up, administrative censoring, accrual patterns, etc. When the shape parameter of the Weibull distribution is known, the package enables users to implement three SC methods (conditional power, predictive power, Bayesian predictive probability) to aid decision making based on the data collected at an interim stage. In addition, when the shape parameter is unknown, the two approaches proposed by Waleed et al. [55] are also implemented to calculate the Bayesian predictive probability of a successful trial outcome. In the current version, only numerical results are output by the package. In future upgraded versions, we plan to build upon some of the functions by adding graphical features, such as plots of the required study sample size vs unconditional power, effect size vs. study sample size, and effect size vs. conditional power, etc.

Chapter 5

Conclusions and Future Direction

In this dissertation, a parametric MLE-based method was proposed for designing flexible single-arm clinical trials with time-to-event endpoints following a Weibull distribution with known shape parameter. The proposed method can be utilized to incorporate various design features including different accrual patterns and expected loss to follow-up rate, and we demonstrated that it maintains the desired Type-I error rate and power in most scenarios. Furthermore, we presented mathematical development of three stochastic curtailments, namely conditional power, predictive power and Bayesian predictive probability, which can be employed for decision making regarding early termination of a clinical trial at an interim stage. We also extended our work on Bayesian predictive probability by addressing the case when the shape parameter of the Weibull distribution is not known from historical studies. An `R` package was also developed to implement these methods for the use of wider scientific community.

We feel important to reemphasize some issues related to the work presented in this dissertation:

1. The proposed asymptotic method performs the best for designing clinical trials with moderate to large sample sizes. If the affordable study sample sizes are small or exact calculations are desired, one may consider the methods proposed by Phadnis [6] and Wu [10].

2. The proposed MLE-based method is appropriate when a reasonably accurate estimate of the shape parameter of the Weibull distribution is available from historical studies. If its estimate is not available for some reasons, we may design the study assuming exponentially distributed survival times (i.e., $\kappa = 1$), and reestimate the sample size using an estimate of the shape parameter based on data collected at an interim stage, if necessary.

3. Although Bayesian predictive probabilities are easily interpretable, a drawback is that there calculations are extremely time-consuming, especially when both parameters of the Weibull distribution are unknown, due to repeated sampling of the predicted survival data [45].

4. Unfortunately, there does not appear to be a wider consensus on the appropriate choices of prior distributions [18, 25, 45]. Therefore, clinicians and statistician should work closely to identify appropriate priors for the relevant parameters in the context of their specific research questions.

Listed below are some of the avenues which could be explored in future research:

1. In our context, the design of single-arm clinical trials based on a fully Bayesian approach could also be considered, and compared with the proposed parametric MLE-based method.

2. In this dissertation, we explored the utility of the internal pilot study approach for reestimating sample size at an interim stage. Future studies could also explore sample size reestimation based on different metrics, such as conditional power and Bayesian predictive probability.

3. Since multiple treatments can also be evaluated in the early phase clinical trials, the methods proposed in this dissertation could be generalized to allow for the introduction of covariates into the scale parameter of the Weibull distribution.

4. Though the Weibull distribution is a good choice for designing single arm trials owing to the mathematical expressions having closed forms, the underlying assumption is of hazards that increase from 0 to infinity, or, decrease from infinity to 0. For some diseases where this is not true (e.g. hazard increases from 0 to a constant), other distributions may be considered by adopting the general framework of mathematical derivations and asymptotic results discussed in this dissertation.

References

[1] Clinical Trials Information for Patients and Caregivers. National Cancer Institute Website. Modified February 4, 2020. Accessed July 30, 2020. https://www.cancer.gov/about-cancer/treatment/clinical-trials/.

[2] The Drug Development Process. U.S. Food and Drug Administration Website. Created January 4, 2018. Accessed July 30, 2020. https://www.fda.gov/patients/learn-about-drug-and-device-approvals/drug-development-process/.

[3] Evans SR. Clinical trial structures. *J Exp Stroke Transl Med.* 2010;3:8–18.

[4] U. S. Food and Drug Administration. Clinical Trial Endpoints for the Approval of Cancer Drugs and Biologics. 2018.

[5] Rubinstein L. Phase II design: history and evolution. *Chin Clin Oncol.* 2014;3.

[6] Phadnis MA. Sample size calculation for small sample single-arm trials for time-to-event data: Logrank test with normal approximation or test statistic based on exact chi-square distribution? *Contemp Clin Trials Commun.* 2019;15.

[7] Finkelstein DM, Muzikansky A, Schoenfeld DA. Comparing survival of a sample to that of a standard population. *J Natl Cancer Inst.* 2003;95:1434–1439.

[8] Kwak M, Jung SH. Phase II clinical trials with time-to-event endpoints: optimal two-stage designs with one-sample log-rank test. *Stat Med.* 2014;33:2004–2016.

[9] Sun X, Peng P, Tu D. Phase II cancer clinical trials with a one-sample log-rank test and its corrections based on the Edgeworth expansion. *Contemp Clin Trials.* 2011;32:108–113.

[10] Wu J. Sample size calculation for the one-sample log-rank test. *Pharm Stat.* 2015;14:26–33.

[11] PASS Power Analysis and Sample Size Software Kaysville, Utah: NCSS, LLC; 2021. https://www.ncss.com/software/pass/.

[12] nQuery 8 Power and Sample Size for Group Sequential Trials Cork, Ireland: "Statsols" (Statistical Solutions Limited); 2021. https://www.statsols.com/nquery/.

[13] Narula SC, Li FS. Sample size calculations in exponential life testing. *Technometrics.* 1975;17:229–231.

[14] Owzar K, Jung SH. Designing phase II studies in cancer with time-to-event endpoints. *Clin Trials.* 2008;5:209–221.

[15] Simon R. Optimal two-stage designs for phase II clinical trials. *Control Clin Trials.* 1989;10:1–10.

[16] Kunz CU, Kieser M. Curtailment in single-arm two-stage phase II oncology trials. *Biom J.* 2012;54:445–456.

[17] Jennison C, Turnbull BW. *Group Sequential Methods With Applications to Clinical Trials.* Boca Raton, FL: Chapman & Hall/CRC, 2000.

[18] Dmitrienko A, Koch GG. *Analysis of Clinical Trials Using SAS: A Practical Guide, Second Edition.* Cary, NC: SAS Institute Inc., 2017.

[19] Lan KKG, Simon R, Halperin M. Stochastically curtailed tests in long-term clinical trials. *Commun. Stat. Part C: Sequential Analysis.* 1982;1:207–219.

[20] Andersen PK. Conditional power calculations as an aid in the decision whether to continue a clinical trial. *Control Clin Trials.* 1987;8:67–74.

[21] Spiegelhalter DJ, Freedman LS, Blackburn PR. Monitoring clinical trials: Conditional or predictive power? *Control Clin Trials.* 1986;7:8–17.

[22] Herson J. Predictive probability early termination plans for phase II clinical trials. *Biometrics.* 1979;35:775–783.

[23] Choi SC, Pepple PA. Monitoring clinical trials based on predictive probability of significance. *Biometrics.* 1989;45:317–323.

[24] Geisser S. On the curtailment of sampling. *Can J Stat.* 1992;20:297–309.

[25] Dmitrienko A, Wang MD. Bayesian predictive approach to interim monitoring in clinical trials. *Stat Med.* 2006;25:2178–2195.

[26] R Core Team. R: A language and environment for statistical computing. Vienna, Austria: R Foundation for Statistical Computing; 2017. https://www.R-project.org/.

[27] SAS 9.4. Cary, NC: SAS Institute; 2017. https://sas.com/.

[28] Proschan MA, Lan KKG, Wittes JT. *Statistical monitoring of clinical trials.* New York, NY: Springer, 2006.

[29] Phadnis MA, Mayo MS. Group sequential design for time-to-event data using the concept of proportional time. *Stat Methods Med Res.* 2020;29:1867–1890.

[30] Pocock SJ. Group sequential methods in the design and analysis of clinical trials. *Biometrika.* 1977;64:191–199.

[31] O'Brien PC, Fleming TR. A multiple testing procedure for clinical trials. *Biometrics.* 1979;35:549–556.

[32] Wang SK, Tsiatis AA. Approximately optimal one-parameter boundaries for group sequential trials. *Biometrics.* 1987;43:193–199.

[33] Lan KKG, DeMets DL. Discrete sequential boundaries for clinical trials. *Biometrika.* 1983;70:659–663.

[34] Hwang IK, Shih WJ, De Cani JS. Group sequential designs using a family of type I error probability spending functions. *Stat Med.* 1990;9:1439–1445.

[35] Jennison C, Turnbull BW. Statistical approaches to interim monitoring of medical trials: a review and commentary. *Statist. Sci.* 1990;5:299–317.

[36] Klein JP, Moeschberger ML. *Survival Analysis: Techniques for Censored and Truncated Data.* New York, NY: Springer, 2003.

[37] Phadnis MA, Sharma P, Thewarapperuma N, Chalise P. Assessing accuracy of Weibull shape parameter estimate from historical studies for subsequent sample size calculation in clinical trials with time-to-event outcome. *Contemp Clin Trials Commun.* 2020;17.

[38] Wan F. Simulating survival data with predefined censoring rates for proportional hazards models. *Stat Med.* 2017;36:838–854.

[39] Casella G, Berger RL. *Statistical Inference.* Belmont, CA: Duxbury Press, 2002.

[40] Mütze T, Glimm E, Schmidli H, Friede T. Group sequential designs for negative binomial outcomes. *Stat Methods Med Res.* 2019;28:2326–2347.

[41] Scharfstein DO, Tsiatis AA, Robins JM. Semiparametric efficiency and its implication on the design and analysis of group-sequential studies. *J. Am. Stat. Assoc.* 1996;92:1342–1350.

[42] Armitage P, McPherson CK, Rowe BC. Repeated significance tests on accumulating data. *J. Royal Stat. Soc. Series A (General).* 1969;132:235–244.

[43] Davis BR, Hardy RJ. Data monitoring in clinical trials: the case for stochastic curtailment. *J Clin Epidemiol.* 1994;47:1033–1042.

[44] Schmidt R, Faldum A, Kwiecien R. Adaptive designs for the one-sample log-rank test. *Biometrics.* 2018;74:529–537.

[45] Saville BR, Connor JT, Ayers GD, Alvarez J. The utility of Bayesian predictive probabilities for interim monitoring of clinical trials. *Clin Trials.* 2014;11:485–493.

[46] U. S. Food and Drug Administration. Clinical trial endpoints for the approval of cancer drugs and biologics. 2018.

[47] Waleed M, He J, Phadnis MA. Some design considerations incorporating early futility for single-arm clinical trials with time to event primary endpoints using Weibull distribution. *Pharm Stat.* 2021;20:610–644.

[48] U. S. Food and Drug Administration. Adaptive design clinical trials for drugs and biologics. 2019.

[49] Wittes J, Brittain E. The role of internal pilot studies in increasing the efficiency of clinical trials. *Stat Med.* 1990;9:65–72.

[50] Friede T, Kieser M. Sample size recalculation in internal pilot study designs: a review. *Biom J.* 2006;48:537–555.

[51] Ibrahim JG, Chen MH, Sinha D. *Bayesian Survival Analysis.* New York, NY: Springer, 2001.

[52] Thomas N. R2OpenBUGS: Running OpenBUGS from R. 2020. R package version 3.2. https://cran.r-project.org/web/packages/R2OpenBUGS/.

[53] Mayo MS, Gajewski BJ. Bayesian sample size calculations in phase II clinical trials using informative conjugate priors. *Control Clin Trials.* 2004;25:157–167.

[54] Kuehnapfel A. CP: Conditional Power Calculations. R package version 1.6. https://cran.r-project.org/web/packages/CP/.

[55] Waleed M, He J, Phadnis MA. Bayesian predictive probability for single-arm clinical trials with a time-to-event endpoint using Weibull distribution with unknown shape parameter. Under Review.

[56] Wickham H. *R Packages: Organize, Test, Document, and Share Your Code.* Sebastopol, CA: O'Reilly Media, 2015.

[57] Wickham H, Hester J, Chang W, RStudio, R Core Team. devtools: Tools to Make Developing R Packages Easier. 2020. R package version 2.3.2. https://cran.r-project.org/web/packages/devtools/.

[58] Therneau TM, Lumley T, Elizabeth A, Cynthia C. survival: Survival Analysis. 2021. R package version 3.2. https://cran.r-project.org/web/packages/survival/.

[59] R Core Team. stats: The R Stats Package. 2021. R package version 4.1. https://stat.ethz.ch/R-manual/R-devel/library/stats/html/00Index.html.

[60] Rimella L. RGeode: Geometric Density Estimation. 2017. R package version 0.1.0. https://cran.r-project.org/web/packages/RGeode/.

[61] Genz A, Bretz F, Miwa T, et al. mvtnorm: Multivariate Normal and t Distributions. 2020. R package version 1.1. https://cran.r-project.org/web/packages/mvtnorm/.

[62] Novomestky F, Nadarajah S. truncdist: Truncated Random Variables. 2016. R package version 1.0. https://cran.r-project.org/web/packages/truncdist/.

Appendix A

Supplementary Materials for Chapter 2

A.1 Probability Density Functions for the Special Cases of the Censoring Variable

1. When the ith subject is accrued into the study at the calendar time 0, that is, $E_i = 0$, and there is no loss to follow-up, i.e., $C_i \to \infty$, then $Z_i = \tau$ is a degenerate random variable with its entire mass located at τ.

2. When all n subjects are accrued into the study at the calendar time 0, that is, $E_i = 0$ for all $i = 1, \ldots, n$, and $C_i \sim \text{Weibull}(\kappa, \eta)$, then $Z_i = \min(C_i, \tau)$ follows a truncated Weibull distribution with parameters $\kappa > 0$ and $\eta > 0$ having the pdf:

$$f_{Z_i}(z) = \begin{cases} \dfrac{\dfrac{\kappa}{\eta^\kappa} z^{\kappa-1} \exp\left\{-\left(\dfrac{z}{\eta}\right)^\kappa\right\}}{1 - \exp\left\{-\left(\dfrac{\tau}{\eta}\right)^\kappa\right\}} & \text{if } z \in [0, \tau] \\ 0 & \text{otherwise.} \end{cases}$$

3. When $E_i \sim \text{Gen-Uniform}(0, \omega, \varphi)$, and there is no loss to follow-up, i.e., $C_i \to \infty$ for all $i = 1, \ldots, n$, then the pdf of $Z_i = \tau - E_i$ is given as:

$$f_{Z_i}(z) = \begin{cases} \dfrac{\varphi(\tau - z)^{\varphi-1}}{\omega^\varphi} & \text{if } z \in [\tau - \omega, \tau] \\ 0 & \text{otherwise} \end{cases}$$

A.2 Metropolis–Hastings Algorithm for Generating Posterior Samples from $\pi(\gamma|\hat{\gamma})$

In order to generate samples from the posterior distribution of γ given by $\pi(\gamma|\widehat{\gamma}) \propto \pi(\widehat{\gamma}|\gamma)\pi(\gamma)$, we assume a flat normal prior $\pi(\gamma) \sim \text{Normal}\left(\mu = 0, \sigma^2 = 5^2\right)$, and $\pi(\widehat{\gamma}|\gamma)$ is defined in Eq. (2.6). We implement the MH algorithm to generate posterior samples from $\pi(\gamma|\widehat{\gamma})$ as follows:

1. Initialize $\gamma^{(1)} = 0$.

2. At the i-th iteration of the algorithm, perform the following steps:

   i. Draw a new candidate y from Normal $\left(\mu = \gamma^{(i)}, \sigma^2 = \dfrac{1}{\kappa^2 \mu_{\bar{\delta}}^{(i)} n}\right)$ distribution, where: $\mu_{\bar{\delta}}^{(i)} = 1 - E_Z\left(\exp\left\{-\left(\dfrac{Z}{\theta^{(i)}}\right)^{\kappa}\right\}\right)$, $\theta^{(i)} = \exp\left\{\gamma^{(i)}\right\}$, $Z = \min(A, C) = \min(\max(0, \tau - E), C)$, $E \sim \text{Gen-Uniform}(0, \omega, \varphi)$, and $C \sim \text{Weibull}\left(\kappa, \theta^{(i)}\left(\dfrac{1-\upsilon}{\upsilon}\right)^{1/\kappa}\right)$.

   ii. Compute $\pi(\gamma = \widehat{\gamma}|y)$ and $\pi\left(\gamma = \widehat{\gamma}|\gamma^{(i)}\right)$ as:

$$\pi(\gamma = \widehat{\gamma}|y) \propto \pi(\widehat{\gamma}|y)\pi(y).$$
$$\pi\left(\gamma = \widehat{\gamma}|\gamma^{(i)}\right) \propto \pi\left(\widehat{\gamma}|\gamma^{(i)}\right)\pi\left(\gamma^{(i)}\right).$$

   iii. Compute the acceptance probability $\alpha\left(\gamma^{(i)}, y\right)$ as follows:

$$\alpha\left(\gamma^{(i)}, y\right) = \min\left\{\frac{g(y)q\left(\gamma^{(i)}, y\right)}{g\left(\gamma^{(i)}\right)q\left(y, \gamma^{(i)}\right)}, 1\right\} = \min\left\{\frac{\pi(\widehat{\gamma}|y)\pi(y)q\left(\gamma^{(i)}, y\right)}{\pi\left(\widehat{\gamma}|\gamma^{(i)}\right)\pi\left(\gamma^{(i)}\right)q\left(y, \gamma^{(i)}\right)}, 1\right\},$$

   where $q\left(\gamma^{(i)}, y\right)$ is the density of Normal $\left(y, \frac{\sigma^2(y)}{n}\right)$ distribution evaluated at $\gamma^{(i)}$, and $q\left(y, \gamma^{(i)}\right)$ is the density of Normal $\left(\gamma^{(i)}, \frac{\sigma^2(\gamma^{(i)})}{n}\right)$ distribution evaluated at y.

iv. Draw a uniform random number u between 0 and 1. If $u < \alpha\left(\gamma^{(i)}, y\right)$, then

$$\text{set } \gamma^{(i+1)} = y,$$

$$\text{else set } \gamma^{(i+1)} = \gamma^{(i)}.$$

3. Repeat steps 2(i)–2(iv) a large number of times, say 11000, and exclude the 1000 burn-in samples from the generated posterior samples to obtain the estimated posterior distribution.

A.3 Simulation Algorithm for Constructing Group Sequential Design Plans for Single-Arm Studies with a Time-to-Event Endpoint

The following simulation algorithm can be implemented in statistical software such as `R` [26], `SAS` [27], etc., to construct GSD plans to evaluate both efficacy and futility (binding) during interim analyses for single-arm clinical trials with a time-to-event primary endpoint.

1. After consultation with the researchers, specify the following inputs: **(i)** desired global Type-I error rate α along with the parameter ψ_α for the α-spending function; **(ii)** Type-II error rate β and parameter ψ_β for our choice of the β-spending function; **(iii)** total number of looks K to be performed, and the corresponding information fractions $\mathscr{F}_k$ ($k = 1, \ldots, K$); **(iv)** median survival time under the null hypothesis to obtain log of the scale parameter, $\log(\theta_0) = \gamma_0$, for the failure time distribution; **(v)** clinically meaningful effect $\varepsilon > 0$ to be detected under the alternative hypothesis; **(vi)** a reasonable estimate of the common shape parameter κ of the Weibull distribution for the failure and random censoring times; **(vii)** expected loss to follow-up rate υ; **(viii)** maximum accrual time ω; **(ix)** power parameter φ of the accrual distribution; **(x)** final look time $\ell_K = \tau$; and **(xi)** a sufficiently large number of simulations S (≥ 10000) to be performed.

2. For the user-defined information fractions $\mathscr{F}_k$, determine the corresponding calendar time of the kth interim analyses as: $\ell_k = \tau \cdot \mathscr{F}_k$. Based on these information fractions and the pre-

specified parameters ψ_α and ψ_β, the local Type-I error rate α_k and Type-II error rate β_k to be spent at the kth interim analysis are determined using the pre-specified choices of spending function in Eqs. (2.20) and (2.21) of the manuscript.

3. Initialize sample size $n = n^{(0)}$.

4. For all n subjects in the sth simulated sample, generate the corresponding calendar times of accrual into the study, denoted by $e_{i,s}$ $(i = 1, \ldots, n; s = 1, \ldots, S)$, from the Gen-Uniform$(0, \omega, \varphi)$ distribution.

5. Under the null hypothesis $H_0 : \gamma \leq \gamma_0$, generate the amount of time from $e_{i,s}$ to failure $(y_{i,s;H_0})$ as well as loss to follow-up $(c_{i,s;H_0})$ for the ith subject in the sth sample $(i = 1, \ldots, n; s = 1, \ldots, S)$ from Weibull$(\kappa, \theta_0 = e^{\gamma_0})$ and Weibull$\left(\kappa, \eta = \theta_0 \left(\frac{1-\upsilon}{\upsilon}\right)^{1/\kappa}\right)$ distributions, respectively.

   Repeat the same process under the alternative hypothesis $H_1 : \gamma = \gamma_1$ $(= \gamma_0 + \varepsilon$ with $\varepsilon > 0)$ to generate the times to failure $(y_{i,s;H_1})$ and loss to follow-up $(c_{i,s;H_1})$.

6. For each of the K interim analyses to be performed at calendar times ℓ_k $(k = 1, \ldots, K)$, use the data generated in Steps 4 and 5 to prepare appropriate survival data for all n subjects within the sth sample under the hypothesis H_g, where $s = 1, \ldots, S$, and $g = 0, 1$. Let $x_{i,s,k;H_g}$ and $\delta_{i,s,k;H_g}$, respectively, denote the survival time and survival status for the ith subject within the sth sample at the calendar time ℓ_k for the kth interim analysis generated under H_g. The components of the pair $\left(x_{i,s,k;H_g}, \delta_{i,s,k;H_g}\right)$ of simulated survival data are obtained as follows:

   - $x_{i,s,k;H_g} := \min\left(y_{i,s;H_g}, z_{i,s,k;H_g}\right)$, where $z_{i,s,k;H_g} := \min\left(c_{i,s;H_g}, \max\left(0, \ell_k - e_{i,s}\right)\right)$ is the amount of time from accrual into the study to being censored at the look time ℓ_k, and $y_{i,s;H_g}$ is defined in Step 5

   - $\delta_{i,s,k;H_g} := \mathbb{1}_{\left(y_{i,s;H_g} < z_{i,s,k;H_g}\right)}$ is the corresponding survival status at the look time ℓ_k.

7. Let $\widehat{\gamma}_{s,k;H_g}$ denote the estimated value of γ obtained by fitting the parametric Weibull model (using the `survreg` function available in the `survival` [58] package in `R`) on the appropriate

sth sample at the kth look (generated in Step 6) under the hypothesis H_g. We construct the simulated efficacy and futility stopping boundaries (denoted by $\widehat{\mathscr{U}_k}$ and $\widehat{\mathscr{L}_k}$, respectively) as follows:

(a) At the first look time ℓ_1, obtain the sorted vector of the estimated values $\widehat{\boldsymbol{\gamma}}_{1;H_g} = \left\{\widehat{\gamma}_{s,1;H_g}\right\}_{s=1}^{S}$ under the hypothesis H_g. The estimated stopping boundary for efficacy $\widehat{\mathscr{U}_1}$ and futility $\widehat{\mathscr{L}_1}$ at the look time ℓ_1 is determined such that:

$$\text{Prob}(\widehat{\boldsymbol{\gamma}}_{1;H_0} \geq \widehat{\mathscr{U}_1}) = \alpha_1$$
$$\text{Prob}(\widehat{\boldsymbol{\gamma}}_{1;H_1} \leq \widehat{\mathscr{L}_1}) = \beta_1$$

(b) For $k = 1, \ldots, K-1$, only those samples simulated under the hypothesis H_g progress from the kth look to the $(k+1)$st look which yielded estimated values of parameter $\widehat{\gamma}_{s,k;H_g}$ such that $\widehat{\mathscr{L}_k} < \widehat{\gamma}_{s,k;H_g} < \widehat{\mathscr{U}_k}$. For the samples available at $(k+1)$st look, repeat the same procedure as mentioned for the 1st look to compute the estimated efficacy and futility stopping boundaries as:

$$\text{Prob}(\widehat{\boldsymbol{\gamma}}_{k+1;H_0} \geq \widehat{\mathscr{U}_{k+1}}) = \alpha_{k+1}$$
$$\text{Prob}(\widehat{\boldsymbol{\gamma}}_{k+1;H_1} \leq \widehat{\mathscr{L}_{k+1}}) = \beta_{k+1}$$

8. Once we determine the estimated values of the final stopping boundaries $\widehat{\mathscr{U}_K}$ and $\widehat{\mathscr{L}_K}$ (using Step 7) corresponding to the initial sample size $n = n^{(0)}$, update the sample size $n = n^{(new)}$ for the next iteration of simulation algorithm as follows: increase (decrease) the candidate sample size n if $\widehat{\mathscr{U}_K} > \widehat{\mathscr{L}_K}$ ($\widehat{\mathscr{U}_K} < \widehat{\mathscr{L}_K}$) during the previous iteration.

9. Repeat Steps 4 – 8 until we find a value of sample size n such $\widehat{\mathscr{L}_K} < \widehat{\mathscr{U}_K}$ for $n-1$, and $\widehat{\mathscr{L}_K} \geq \widehat{\mathscr{U}_K}$ for n. This is the required sample size for our single-arm study in the GSD framework.

10. Since $\widehat{\mathscr{L}_K} \geq \widehat{\mathscr{U}_K}$ for our chosen sample size n, we need to make adjustment in the desired

power to ensure $\widehat{\mathscr{L}_K} = \widehat{\mathscr{U}_K}$ at the final look time ℓ_K. Keeping our chosen sample size n as fixed, we do the following:

(a) Deduct very small quantities from the global Type-II error rate β, and perform steps 4 – 8 using a duly updated β-spending function.

(b) Repeat Step 10a until we find some β^* smaller than the pre-specified β such that $\widehat{\mathscr{L}_K} = \widehat{\mathscr{U}_K}$, preferably up to 4 decimal places.

The above simulation algorithm can be used to construct appropriate GSD plan for a single-arm study with sample size n and power $1-\beta^*$ for testing both efficacy and futility of an experimental treatment during all of the K interim analyses.

To incorporate some other design features (such as futility skips, non-binding futility, etc.), few adjustments in the above algorithm are necessary:

1. The above simulation approach assumes binding futility rules in its implementation. To incorporate non-binding futility rules instead, we simply need to make sure in Step 7 that all those samples at the kth look, generated under the null hypothesis H_0, progress to the $(k+1)$st look for which $\widehat{\gamma}_{s,k;H_0} < \widehat{\mathscr{U}_k}$ (where $k = 1,\ldots,K-1$).

2. If the researcher desires futility skips during early interim analyses, we can incorporate such a requirement by allocating no local Type-II error rate at the corresponding initial looks.

3. This algorithm simplifies considerably when we are interested in constructing GSD plans with efficacy boundaries only. In such a scenario, there are no calculations for the futility boundaries, and the desired sample size n is the one which ensures at least power $1-\beta$ at the final look time K.

Expected Number of Events and Stopping Probabilities

It is also of interest to calculate the expected number of events and stopping probabilities during interim analyses. Let $P_{k;H_g}$ denote the probability of stopping the trial at the kth look ($k = 1,\ldots,K$)

under the hypothesis H_g $(g = 0, 1)$. During the kth interim analysis, these probabilities can be easily computed within our simulation algorithm as:

$$P_{k;H_g} = 1 - \text{Prob}\left(\widehat{\mathscr{L}_k} < \widehat{\boldsymbol{\gamma}}_{k;H_g} < \widehat{\mathscr{U}_k}\right) \tag{A.1}$$

For GSD plans with efficacy boundaries only, note that $P_{k;H_0}$ equals to the local error rate α_k spent at the kth look time, and $P_{k;H_1}$ is the proportion of samples generated under H_1 which yield $\widehat{\gamma}_{s,k;H_1} > \widehat{\mathscr{U}_k}$.

Under the hypothesis H_g, suppose that $\bar{n}_{e,k;H_g}$ denotes the average number of events observed across available simulated samples during the kth look. The expected number of events under the hypothesis H_g, denoted by $E\left(n_{e;H_g}\right)$, can be easily computed as:

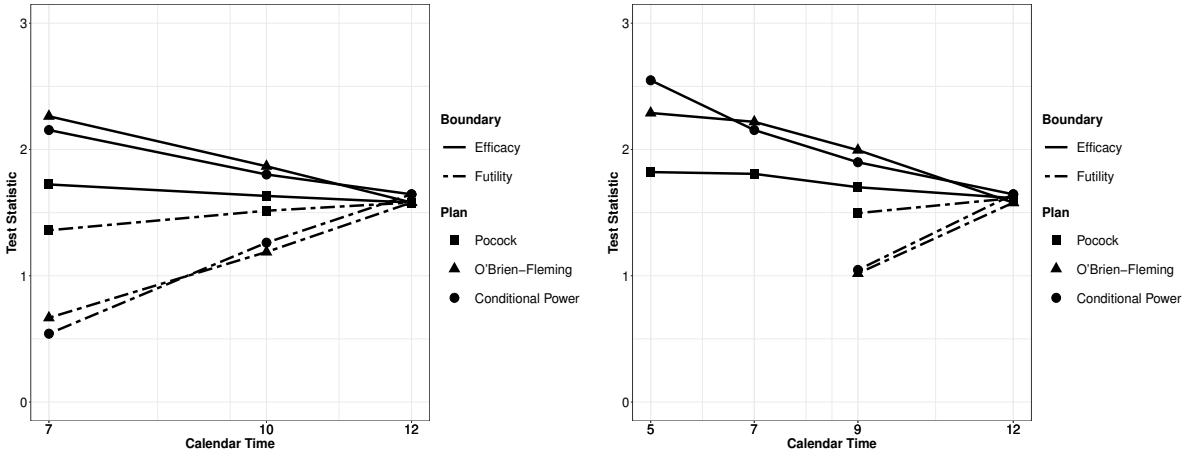
$$E\left(n_{e;H_g}\right) = \sum_{k=1}^{K-1}\left(\bar{n}_{e,k;H_g} \cdot P_{k;H_g}\right) + \bar{n}_{e,K;H_g} \cdot \left(1 - \sum_{k=1}^{K-1} P_{k;H_g}\right) \tag{A.2}$$

Examples

For the hypothetical study discussed in Section 2.4.3 of the manuscript, we present two additional examples to compare and contrast aspects of the Pocock and O'Brien-Fleming GSD plans. First, we present a GSD plan to evaluate both efficacy and futility at three unequally-spaced look times (at time = 7, 10, 12 months), and the results are summarized in Table A.1. We observe that the basic underlying differences between Pocock and O'Brien-Fleming plans are essentially similar to the GSD plan with equally-spaced looks discussed in the manuscript. It must also be noted that the stopping probabilities during the first interim analysis for this unequally-spaced are larger than the ones observed for the GSD plan with equally-spaced look times. This is primarily due to the fact that the timing of the first look in this design is later (time = 7 months) than those in Table 2.7 of the manuscript (time = 4 months). This amounts to a larger information fraction available at the first interim analysis, and consequently tighter stopping boundaries, which in turn results in higher stopping probabilities under both hypotheses.

Second, we implement a 4-look unequally-spaced design with 2 initial futility skips, and the corresponding results are summarized in Table A.2. Similar to our previous findings, we note that the O'Brien-Fleming plan yields stringent efficacy boundaries at the earlier looks than the Pocock plan. Under the null hypothesis, stopping probabilities are very small at the first two looks due to futility skips, which yields the largest stopping probabilities at the third look. We again observe that the expected number of events for the Pocock plan under the alternative hypothesis is comparatively smaller than the one for the O'Brien-Fleming design.

Pictorial comparison of the stopping boundaries for the Pocock and O'Brien-Fleming designs for these two scenarios is shown in Figure A.1. We have also shown the efficacy and futility boundaries obtained using $\zeta = \zeta' = 0.5$ in the conditional power stopping boundaries. Once again, we observe that the conditional power boundaries with $\zeta = \zeta' = 0.5$ closely resemble the O'Brien-Fleming boundaries. Since the parameters ζ and ζ' are typically chosen around 0.80, the O'Brien-Fleming testing procedure is more likely to trigger an early stopping of a clinical trial in comparison to the conditional power approach.



(a) Unequally-spaced – Table A.1 (b) Unequally-spaced (2 futility skips) – Table A.2

Figure A.1: Stopping boundaries for the GSD plans in Tables A.1–A.2

Table A.1: Comparison of Pocock and O'Brien-Fleming GSD plans with three unequally-spaced looks for evaluating efficacy and futility in a hypothetical study to test the hypotheses $H_0 : M \leq 1.00$ month vs. $H_1 : M > 1.50$ months

Type	k	ℓ_k	Efficacy						Futility					
			$\bar{n}_{e;H_0}$	α_k	$\sum_k \alpha_k$	$\mathscr{U}_k$	$p_{k;H_0}$	$\sum_k p_{k;H_0}$	$\bar{n}_{e;H_1}$	β_k	$\sum_k \beta_k$	$\mathscr{L}_k$	$p_{k;H_1}$	$\sum_k p_{k;H_1}$
	1	7	33.9326	0.0350	0.0350	1.723	0.9542	0.9542	33.2677	0.0697	0.0697	1.360	0.9470	0.9470
Pocock	2	10	32.5458	0.0098	0.0447	1.632	0.0316	0.9858	34.9226	0.0195	0.0892	1.515	0.0348	0.9818
$(n = 40)$	3	12	32.7887	0.0053	0.0500	1.580	0.0142	1.0000	34.9560	0.0105	0.0997	1.580	0.0182	1.0000
			$E(n_{e;H_0}) = 33.8725$						$E(n_{e;H_1}) = 33.3560$					
O'Brien &	1	7	33.9326	0.0087	0.0087	2.263	0.7633	0.7633	33.2677	0.0173	0.0173	0.667	0.7541	0.7541
Fleming	2	10	32.9485	0.0165	0.0252	1.867	0.1523	0.9156	34.8833	0.0330	0.0503	1.188	0.1426	0.8967
$(n = 40)$	3	12	32.5687	0.0248	0.0500	1.580	0.0844	1.0000	34.9894	0.0494	0.0997	1.580	0.1033	1.0000
			$E(n_{e;H_0}) = 33.6676$						$E(n_{e;H_1}) = 33.6759$					

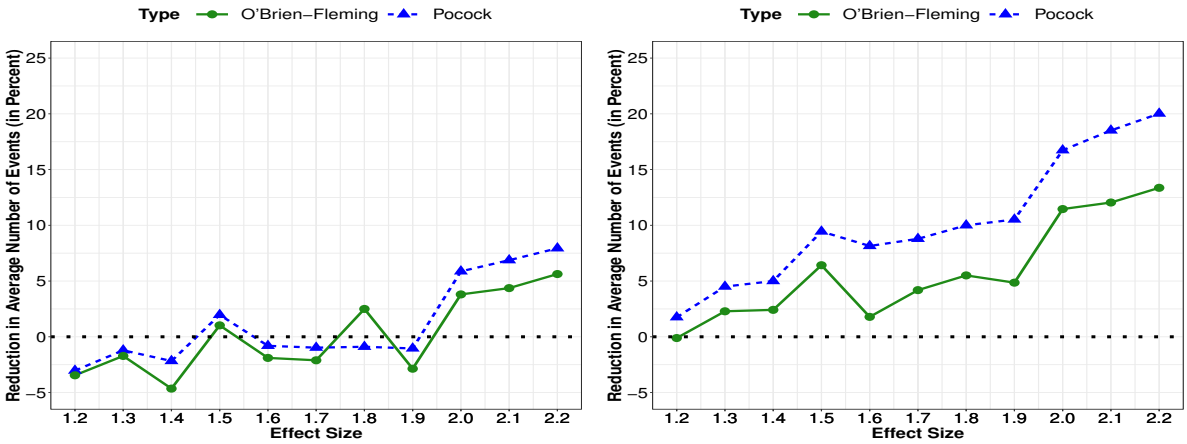
Note: Using nominal α = 5%, power = 90%, uniform accrual pattern ($\varphi = 1$), maximum accrual time $\omega = 3$ months, administrative censoring time $\tau = 12$ months, common shape parameter $\kappa = 1.25$, and loss to follow-up rate $\upsilon = 0.15$, these GSD plans were generated on the basis of 10,000 simulations.

Table A.2: Comparison of Pocock and O'Brien-Fleming GSD plans for evaluating efficacy and futility (with 2 skips) at four unequally-spaced look times in a hypothetical study to test the hypotheses $H_0 : M \leq 1.00$ month vs. $H_1 : M > 1.50$ months

Type	k	ℓ_k	Efficacy						Futility					
			$\bar{n}_{e;H_0}$	α_k	$\sum_k \alpha_k$	$\mathscr{U}_k$	$p_{k;H_0}$	$\sum_k p_{k;H_0}$	$\bar{n}_{e;H_1}$	β_k	$\sum_k \beta_k$	$\mathscr{L}_k$	$p_{k;H_1}$	$\sum_k p_{k;H_1}$
Pocock ($n = 40$)	1	5	32.7709	0.0270	0.0270	1.821	0.1806	0.1806	29.9567	–	–	–	0.8355	0.8355
	2	7	33.7870	0.0080	0.0350	1.807	0.0132	0.1938	34.7751	–	–	–	0.0331	0.8686
	3	9	33.8632	0.0068	0.0417	1.701	0.7872	0.9810	35.2831	0.0831	0.0831	1.496	0.1011	0.9697
	4	12	32.5105	0.0083	0.0500	1.615	0.0186	0.9996	35.1452	0.0165	0.0996	1.615	0.0294	0.9991
			$E(n_{e;H_0}) = 33.6392$						$E(n_{e;H_1}) = 30.8119$					
O'Brien & Fleming ($n = 40$)	1	5	32.7772	0.0040	0.0040	2.289	0.0487	0.0487	29.9474	–	–	–	0.6042	0.6042
	2	7	33.8737	0.0047	0.0087	2.219	0.0062	0.0549	34.2289	–	–	–	0.1281	0.7323
	3	9	33.9583	0.0091	0.0178	1.995	0.8168	0.8717	34.8876	0.0355	0.0355	1.018	0.1210	0.8533
	4	12	32.6882	0.0322	0.0500	1.579	0.1283	1.0000	35.0381	0.0642	0.0997	1.579	0.1467	1.0000
			$E(n_{e;H_0}) = 33.7373$						$E(n_{e;H_1}) = 31.8404$					

Note: Using nominal α = 5%, power = 90%, uniform accrual pattern ($\varphi = 1$), maximum accrual time $\omega = 3$ months, administrative censoring time $\tau = 12$ months, common shape parameter $\kappa = 1.25$, and loss to follow-up rate $\upsilon = 0.15$, these GSD plans were generated on the basis of 10,000 simulations.

For the two examples discussed in this section, we plotted the percent reduction in the expected (average) number of events needed for the Pocock and O'Brien-Fleming plans against varying values of the effect size in Figure A.2. It can be observed that the percent reduction in the average number of events needed under the Pocock and O'Brien-Fleming plans relative to the fixed design increases with an increase in the desired effect size. In addition, we note that the Pocock plan offers greater percent reduction in the average number of events in comparison to the O'Brien-Fleming plan. This is primarily due to the fact that the Pocock's design allocates relatively larger magnitudes of the Type-I and Type-II error rates at the earlier interim analyses, which in turn may trigger early stopping of a clinical trial with relatively smaller number of events.



(a) Unequally-spaced – Table A.2

(b) Unequally-spaced (2 futility skips) – Table A.2

Figure A.2: Percent reduction in the average number of events for the Pocock and O'Brien-Fleming GSD plans, in comparison to a fixed sample design, in Tables A.1–A.2

Appendix B

Joint Asymptotic Distribution of $\overline{\tilde{\mathscr{D}}}_{K-k}$ and $\overline{\tilde{\mathscr{T}}}_{K-k}$

B.1 Mathematical Preliminaries

Let ℓ_k and τ denote the calendar time of performing interim analysis and final analysis of the study, respectively. At the interim look time ℓ_k, let ϑ and ϑ_c denote the scale parameter of the failure time and random censoring distributions, respectively. For the jth subject ($j = 1, \ldots, m$) remaining in the study at the interim look time ℓ_k, let $E_j \sim \text{Gen-Uniform}(0, \omega, \varphi)$ represent his/her accrual time into the study, then under the assumption $0 < \ell_k < \tau$, the maximum and minimum possible survival time for this subject in this study is $A_j = \max\left(0, \tau - E_j\right)$ and $B_j = \max\left(0, \ell_k - E_j\right)$, respectively. Therefore, the random variables representing his/her predicted survival time $\tilde{X}_j = \min\left(\tilde{Y}_j, \tilde{Z}_j\right)$ and the corresponding survival status $\tilde{\delta}_j$ are defined as below:

$$\tilde{X}_j = \begin{cases} \tilde{Y}_j & \text{if } \tilde{Y}_j \leq \tilde{Z}_j \\ \tilde{Z}_j & \text{if } \tilde{Y}_j > \tilde{Z}_j \end{cases} \quad ; \quad \tilde{\delta}_j = \begin{cases} 1 & \text{if } \tilde{Y}_j \leq \tilde{Z}_j \\ 0 & \text{if } \tilde{Y}_j > \tilde{Z}_j \end{cases} \quad ; \quad \tilde{Z}_j = \min\left(\tilde{C}_j, A_j\right)$$

$$\tilde{Y}_j \sim \text{Truncated-Weibull}\big(\text{shape} = \kappa, \text{scale} = \vartheta, \text{min} = B_j\big)$$

$$\tilde{C}_j \sim \text{Truncated-Weibull}\big(\text{shape} = \kappa, \text{scale} = \vartheta_c, \text{min} = B_j\big)$$

For the jth subject, we note that the pdf and cdf of $\tilde{Y}_j \sim \text{Weibull}\big(\kappa, \vartheta, B_j\big)$ are given as:

$$f_{\tilde{Y}_j}(y) = \frac{\kappa}{\vartheta^{\kappa}} y^{\kappa-1} \exp\left\{ \left(\frac{B_j}{\vartheta}\right)^{\kappa} - \left(\frac{y}{\vartheta}\right)^{\kappa} \right\}$$

$$F_{\tilde{Y}_j}(y) = 1 - \exp\left\{ \left(\frac{B_j}{\vartheta}\right)^{\kappa} - \left(\frac{y}{\vartheta}\right)^{\kappa} \right\}$$

The pdf and cdf of $\tilde{C}_j$ is similar to that of $\tilde{Y}_j$ with ϑ replaced by ϑ_c. The pdf of $\tilde{Z}_j$ is given as:

$$f_{\tilde{Z}_j}(z) = \begin{cases} \dfrac{\kappa}{\vartheta_c^{\kappa}} z^{\kappa-1} \exp\left\{ \left(\dfrac{B_j}{\vartheta_c}\right)^{\kappa} - \left(\dfrac{z}{\vartheta_c}\right)^{\kappa} \right\} & \text{if } z \in [B_j, \tau-\omega) \\ \left(\dfrac{\varphi}{\tau - z} + \dfrac{\kappa}{\vartheta_c^{\kappa}} z^{\kappa-1} \right) \left(\dfrac{\tau - z}{\omega} \right)^{\varphi} \exp\left\{ \left(\dfrac{B_j}{\vartheta_c}\right)^{\kappa} - \left(\dfrac{z}{\vartheta_c}\right)^{\kappa} \right\} & \text{if } z \in [\tau-\omega, \tau] \\ 0 & \text{otherwise} \end{cases}$$

The pdf of $B_j = \max\left(0, \ell_k - E_j\right)$ is given as:

$$f_{B_j}(b) = \begin{cases} \dfrac{\varphi\,(\ell_k - b)^{\varphi-1}}{\omega^{\varphi}} & \text{if } b \in [\ell_k - \omega, \ell_k] \\ 0 & \text{otherwise} \end{cases}$$

B.2 Derivation of the Joint Asymptotic Distribution

To derive the joint asymptotic distribution of $\overline{\tilde{\mathscr{D}}}_{K-k} = \dfrac{1}{m}\sum_{j=1}^{m} \tilde{\delta}_j$ and $\overline{\tilde{\mathscr{T}}}_{K-k} = \dfrac{1}{m}\sum_{j=1}^{m} \tilde{X}_j^{\kappa}$, we need to derive the mean and variance of $\overline{\tilde{\mathscr{D}}}_{K-k}$ and $\overline{\tilde{\mathscr{T}}}_{K-k}$, and their covariance.

B.2.1 Mean and Variance of $\overline{\tilde{\mathscr{D}}}_{K-k}$

Since $\tilde{\delta}_j$'s are i.i.d. random variables, we observe that

$$E\left(\overline{\tilde{\mathscr{D}}}_{K-k}\right) = E\left(\frac{1}{m}\sum_{j=1}^{m}\tilde{\delta}_j\right) = \frac{1}{m}\sum_{j=1}^{m} E\left(\tilde{\delta}_j\right) = \frac{1}{m}\sum_{j=1}^{m} E\left(\tilde{\delta}_1\right) = E\left(\tilde{\delta}_1\right).$$

By the law of iterated expectations, we note that

$$E\left(\tilde{\delta}_1\right) = E_{B_1}\left(E_{\tilde{Z}_1}\left(E_{\tilde{\delta}_1}\left(\tilde{\delta}_1 | B_1, \tilde{Z}_1\right) \middle| B_1 \right) \right). \tag{B.1}$$

To evaluate the innermost expectation in Eq. (B.1), we observe that

$$E_{\tilde{\delta}_1}\left(\tilde{\delta}_1 | B_1, \tilde{Z}_1\right) = 1 \cdot P\left(\tilde{Y}_1 \le \tilde{Z}_1\right) + 0 \cdot P\left(\tilde{Y}_1 > \tilde{Z}_1\right) = P\left(\tilde{Y}_1 \le \tilde{Z}_1\right) = F_{\tilde{Y}_1}\left(\tilde{Z}_1\right). \tag{B.2}$$

Now we plug Eq. (B.2) into Eq. (B.1) to have that

$$E\left(\tilde{\delta}_1\right) = E_{B_1}\left(E_{\tilde{Z}_1}\left(F_{\tilde{Y}_1}\left(\tilde{Z}_1\right)|\tilde{Z}_1\right)\right) = 1 - E_{B_1}\left(\exp\left\{\left(\frac{B_1}{\vartheta}\right)^{\kappa}\right\} E_{\tilde{Z}_1}\left(\exp\left\{-\left(\frac{\tilde{Z}_1}{\vartheta}\right)^{\kappa}\right\}\right)\right) \quad \text{(B.3)}$$

From here onward, let us denote $\mu_{\tilde{\delta}} = E\left(\tilde{\delta}_1\right)$. It can be easily verified that $E\left(\tilde{\delta}_1^2\right) = E\left(\tilde{\delta}_1\right)$, and we again use the fact that $\tilde{\delta}_j$'s are i.i.d. random variables to note that

$$\mathrm{Var}\left(\overline{\tilde{\mathscr{D}}}_{K-k}\right) = \mathrm{Var}\left(\frac{1}{m}\sum_{j=1}^{m}\tilde{\delta}_j\right) = \frac{1}{m^2}\sum_{j=1}^{m}\mathrm{Var}\left(\tilde{\delta}_i\right) = \frac{1}{m^2}\sum_{j=1}^{m}\mathrm{Var}\left(\tilde{\delta}_1\right) = \frac{\mathrm{Var}\left(\tilde{\delta}_1\right)}{m},$$

where:

$$\mathrm{Var}\left(\tilde{\delta}_1\right) = E\left(\tilde{\delta}_1^2\right) - \left(E\left(\tilde{\delta}_1\right)\right)^2 = \mu_{\tilde{\delta}} - \mu_{\tilde{\delta}}^2 = \mu_{\tilde{\delta}}\left(1-\mu_{\tilde{\delta}}\right). \quad \text{(B.4)}$$

B.2.2 Mean and Variance of $\overline{\tilde{\mathscr{T}}}_{K-k}$

Since $\tilde{X}_j$'s are also i.i.d. random variables, we also note that $E\left(\overline{\tilde{\mathscr{T}}}_{K-k}\right) = E\left(\tilde{X}_1^{\kappa}\right)$, and use the law of iterated expectations to observe that

$$E\left(\tilde{X}_1^{\kappa}\right) = E_{B_1}\left(E_{\tilde{Z}_1}\left(E_{\tilde{X}_1}\left(\tilde{X}_1^{\kappa}|B_1,\tilde{Z}_1\right)\middle|B_1\right)\right).$$

For fixed B_1 and $\tilde{Z}_1$, the innermost conditional expectation $E_{\tilde{X}_1}\left(\tilde{X}_1^{\kappa}|B_1,\tilde{Z}_1\right)$ is computed as:

$$\begin{aligned} E_{\tilde{X}_1}\left(\tilde{X}_1^{\kappa}|B_1,\tilde{Z}_1\right) &= E_{\tilde{X}_1}\left(\tilde{X}_1^{\kappa}|\tilde{Y}_1 \le \tilde{Z}_1\right)P\left(\tilde{Y}_1 \le \tilde{Z}_1\right) + E_{\tilde{X}_1}\left(\tilde{X}_1^{\kappa}|\tilde{Y}_1 > \tilde{Z}_1\right)P\left(\tilde{Y}_1 > \tilde{Z}_1\right) \\ &= E_{\tilde{Y}_1}\left(\tilde{Y}_1^{\kappa}|\tilde{Y}_1 \le \tilde{Z}_1\right)F_{\tilde{Y}_1}\left(\tilde{Z}_1\right) + \tilde{Z}_1^{\kappa}\left(1 - F_{\tilde{Y}_1}\left(\tilde{Z}_1\right)\right). \end{aligned} \quad \text{(B.5)}$$

Given that B_1 and $\tilde{Z}_1$ are fixed, we compute $E_{\tilde{Y}_1}\left(\tilde{Y}_1^{\kappa}|\tilde{Y}_1 \leq \tilde{Z}_1\right)$ by evaluating this expectation with respect to corresponding truncated Weibull distribution:

$$E_{\tilde{Y}_1}\left(\tilde{Y}_1^{\kappa}|\tilde{Y}_1 \leq \tilde{Z}_1\right) = \frac{1}{F_{\tilde{Y}_1}\left(\tilde{Z}_1\right)} \int_{B_1}^{\tilde{Z}_1} \tilde{y}_1^{\kappa} f_{\tilde{Y}_1}(\tilde{y}_1) d\tilde{y}_1 = \frac{1}{F_{\tilde{Y}_1}\left(\tilde{Z}_1\right)} \int_{B_1}^{\tilde{Z}_1} \tilde{y}_1^{\kappa} \left(\frac{\kappa}{\vartheta^{\kappa}}\right) \tilde{y}_1^{\kappa-1} \exp\left\{\left(\frac{B_1}{\vartheta}\right)^{\kappa} - \left(\frac{\tilde{y}_1}{\vartheta}\right)^{\kappa}\right\} d\tilde{y}_1$$

Using the substitution $u = \left(\frac{\tilde{y}_1}{\vartheta}\right)^{\kappa}$, we have that $du = \left(\frac{\kappa}{\vartheta^{\kappa}}\right) \tilde{y}_1^{\kappa-1} d\tilde{y}_1$, and $\tilde{y}_1^{\kappa} = u\vartheta^{\kappa}$. The above integral can now be written as:

$$E_{\tilde{Y}_1}\left(\tilde{Y}_1^{\kappa}|\tilde{Y}_1 \leq \tilde{Z}_1\right) = \frac{\vartheta^{\kappa}}{F_{\tilde{Y}_1}\left(\tilde{Z}_1\right)} \exp\left\{\left(\frac{B_1}{\vartheta}\right)^{\kappa}\right\} \int_{\left(\frac{B_1}{\vartheta}\right)^{\kappa}}^{\left(\frac{\tilde{Z}_1}{\vartheta}\right)^{\kappa}} ue^{-u} du \tag{B.6}$$

We evaluate the above integral using integration by parts by letting $x = u$ and $dy = e^{-u}du$, then $dx = du$ and $y = -e^{-u}$, and get

$$\begin{aligned}
\int_{\left(\frac{B_1}{\vartheta}\right)^{\kappa}}^{\left(\frac{\tilde{Z}_1}{\vartheta}\right)^{\kappa}} ue^{-u} du &= \left(-ue^{-u}\right)\Big|_{\left(\frac{B_1}{\vartheta}\right)^{\kappa}}^{\left(\frac{\tilde{Z}_1}{\vartheta}\right)^{\kappa}} + \int_{\left(\frac{B_1}{\vartheta}\right)^{\kappa}}^{\left(\frac{\tilde{Z}_1}{\vartheta}\right)^{\kappa}} e^{-u} du \\
&= -\left[\left(\frac{\tilde{Z}_1}{\vartheta}\right)^{\kappa} \exp\left\{-\left(\frac{\tilde{Z}_1}{\vartheta}\right)^{\kappa}\right\} - \left(\frac{B_1}{\vartheta}\right)^{\kappa} \exp\left\{-\left(\frac{B_1}{\vartheta}\right)^{\kappa}\right\}\right] - \left(e^{-u}\right)\Big|_{\left(\frac{B_1}{\vartheta}\right)^{\kappa}}^{\left(\frac{\tilde{Z}_1}{\vartheta}\right)^{\kappa}} \\
&= \left(\frac{B_1}{\vartheta}\right)^{\kappa} \exp\left\{-\left(\frac{B_1}{\vartheta}\right)^{\kappa}\right\} - \left(\frac{\tilde{Z}_1}{\vartheta}\right)^{\kappa} \exp\left\{-\left(\frac{\tilde{Z}_1}{\vartheta}\right)^{\kappa}\right\} \\
&\qquad - \exp\left\{-\left(\frac{\tilde{Z}_1}{\vartheta}\right)^{\kappa}\right\} + \exp\left\{-\left(\frac{B_1}{\vartheta}\right)^{\kappa}\right\}
\end{aligned} \tag{B.7}$$

By plugging Eq. (B.7) into Eq. (B.6), and doing some algebra yields

$$\begin{aligned}
E_{\tilde{Y}_1}\left(\tilde{Y}_1^{\kappa}|\tilde{Y}_1 \leq \tilde{Z}_1\right) &= \frac{1}{F_{\tilde{Y}_1}\left(\tilde{Z}_1\right)} \left[B_1^{\kappa} - \tilde{Z}_1^{\kappa} \exp\left\{\left(\frac{B_1}{\vartheta}\right)^{\kappa} - \left(\frac{\tilde{Z}_1}{\vartheta}\right)^{\kappa}\right\} - \theta^{\kappa} \exp\left\{\left(\frac{B_1}{\vartheta}\right)^{\kappa} - \left(\frac{\tilde{Z}_1}{\vartheta}\right)^{\kappa}\right\} + \vartheta^{\kappa}\right] \\
&= \frac{1}{F_{\tilde{Y}_1}\left(\tilde{Z}_1\right)} \left[B_1^{\kappa} - \tilde{Z}_1^{\kappa}\left(1 - F_{\tilde{Y}_1}\left(\tilde{Z}_1\right)\right) - \vartheta^{\kappa}\left(1 - F_{\tilde{Y}_1}\left(\tilde{Z}_1\right)\right) + \vartheta^{\kappa}\right] \\
&= \frac{1}{F_{\tilde{Y}_1}\left(\tilde{Z}_1\right)} \left[B_1^{\kappa} - \tilde{Z}_1^{\kappa}\left(1 - F_{\tilde{Y}_1}\left(\tilde{Z}_1\right)\right) + \vartheta^{\kappa} F_{\tilde{Y}_1}\left(\tilde{Z}_1\right)\right]
\end{aligned} \tag{B.8}$$

By plugging Eq. (B.8) into Eq. (B.5), we have that

$$E_{\tilde{Y}_1}\left(\tilde{Y}_1^{\kappa}|B_1,\tilde{Z}_1\right) = \frac{1}{F_{\tilde{Y}_1}\left(\tilde{Z}_1\right)}\left[B_1^{\kappa} - \tilde{Z}_1^{\kappa}\left(1-F_{\tilde{Y}_1}\left(\tilde{Z}_1\right)\right) + \vartheta^{\kappa}F_{\tilde{Y}_1}\left(\tilde{Z}_1\right)\right]F_{\tilde{Y}_1}\left(\tilde{Z}_1\right) + \tilde{Z}_1^{\kappa}\left(1-F_{\tilde{Y}_1}\left(\tilde{Z}_1\right)\right)$$

$$= B_1^{\kappa} + \vartheta^{\kappa}F_{\tilde{Y}_1}\left(\tilde{Z}_1\right)$$

Using the above expression, we obtain

$$E\left(\tilde{X}_1^{\kappa}\right) = E_{B_1}\left(E_{\tilde{Z}_1}\left(E_{\tilde{X}_1}\left(\tilde{X}_1^{\kappa}|B_1,\tilde{Z}_1\right)\middle|B_1\right)\right) = E_{B_1}\left(E_{\tilde{Z}_1}\left(B_1^{\kappa} + \vartheta^{\kappa}F_{\tilde{Y}_1}\left(\tilde{Z}_1\right)\middle|B_1\right)\right)$$

$$= E_{B_1}(B_1^{\kappa}) + \vartheta^{\kappa}\mu_{\tilde{\delta}} \tag{B.9}$$

Since $\tilde{X}_j$'s are i.i.d. random variables, we note that

$$\mathrm{Var}\left(\overline{\tilde{\mathscr{T}}}_{K-k}\right) = \frac{\mathrm{Var}\left(\tilde{X}_1^{\kappa}\right)}{m} = \frac{E\left(\tilde{X}_1^{2\kappa}\right) - \mu_{\tilde{X}}^2}{m},$$

where $\mu_{\tilde{X}} = E\left(\tilde{X}_1^{\kappa}\right)$.

Similar to our previous calculations, we find the innermost conditional expectation of $E\left(\tilde{X}_1^{2\kappa}\right) = E_{B_1}\left(E_{\tilde{Z}_1}\left(E_{\tilde{X}_1}\left(\tilde{X}_1^{2\kappa}|B_1,\tilde{Z}_1\right)\middle|B_1\right)\right)$ as:

$$E_{\tilde{X}_1}\left(\tilde{X}_1^{2\kappa}|B_1,\tilde{Z}_1\right) = E_{\tilde{Y}_1}\left(\tilde{Y}_1^{2\kappa}|\tilde{Y}_1 \le \tilde{Z}_1\right)F_{\tilde{Y}_1}\left(\tilde{Z}_1\right) + \tilde{Z}_1^{2\kappa}\left(1-F_{\tilde{Y}_1}\left(\tilde{Z}_1\right)\right) \tag{B.10}$$

For fixed B_1 and $\tilde{Z}_1$, we note that the substitution $u = \left(\frac{\tilde{y}_1}{\vartheta}\right)^{\kappa}$ yields

$$E_{\tilde{Y}_1}\left(\tilde{Y}_1^{2\kappa}|\tilde{Y}_1 \le \tilde{Z}_1\right) = \frac{1}{F_{\tilde{Y}_1}\left(\tilde{Z}_1\right)}\int_{B_1}^{\tilde{Z}_1} \tilde{y}_1^{2\kappa} f_{\tilde{Y}_1}(\tilde{y}_1)d\tilde{y}_1$$

$$= \frac{\vartheta^{2\kappa}}{F_{\tilde{Y}_1}\left(\tilde{Z}_1\right)}\exp\left\{\left(\frac{B_1}{\vartheta}\right)^{\kappa}\right\}\int_{\left(\frac{B_1}{\vartheta}\right)^{\kappa}}^{\left(\frac{\tilde{Z}_1}{\vartheta}\right)^{\kappa}} u^2 e^{-u}du \tag{B.11}$$

Using Eq. (B.7) and integration by parts, we find that

$$\int_{\left(\frac{B_1}{\vartheta}\right)^{\kappa}}^{\left(\frac{\tilde{Z}_1}{\vartheta}\right)^{\kappa}} u^2 e^{-u} du = \left(\frac{B_1}{\vartheta}\right)^{2\kappa} \exp\left\{-\left(\frac{B_1}{\vartheta}\right)^{\kappa}\right\} - \left(\frac{\tilde{Z}_1}{\vartheta}\right)^{2\kappa} \exp\left\{-\left(\frac{\tilde{Z}_1}{\vartheta}\right)^{\kappa}\right\} + 2\left[\left(\frac{B_1}{\vartheta}\right)^{\kappa} \exp\left\{-\left(\frac{B_1}{\vartheta}\right)^{\kappa}\right\}\right.$$
$$\left. - \left(\frac{\tilde{Z}_1}{\vartheta}\right)^{\kappa} \exp\left\{-\left(\frac{\tilde{Z}_1}{\vartheta}\right)^{\kappa}\right\} - \exp\left\{-\left(\frac{\tilde{Z}_1}{\vartheta}\right)^{\kappa}\right\} + \exp\left\{-\left(\frac{B_1}{\vartheta}\right)^{\kappa}\right\}\right] \tag{B.12}$$

Plugging Eq. (B.12) into Eq. (B.11), and doing some algebra yields

$$E_{\tilde{Y}_1}\left(\tilde{Y}_1^{2\kappa} | \tilde{Y}_1 \leq \tilde{Z}_1\right) = \frac{\left[B_1^{2\kappa} - \tilde{Z}_1^{2\kappa}\left(1 - F_{\tilde{Y}_1}\left(\tilde{Z}_1\right)\right)\right]}{F_{\tilde{Y}_1}\left(\tilde{Z}_1\right)} + \frac{2\vartheta^{2\kappa}}{F_{\tilde{Y}_1}\left(\tilde{Z}_1\right)}\left[\left(\frac{B_1}{\vartheta}\right)^{\kappa} - \left(\frac{\tilde{Z}_1}{\vartheta}\right)^{\kappa}\left(1 - F_{\tilde{Y}_1}\left(\tilde{Z}_1\right)\right) + F_{\tilde{Y}_1}\left(\tilde{Z}_1\right)\right] \tag{B.13}$$

Now we substitute Eq. (B.13) into Eq. (B.10) to obtain

$$\begin{aligned} E_{\tilde{X}_1}\left(\tilde{X}_1^{2\alpha} | B_1, \tilde{Z}_1\right) &= B_1^{2\kappa} - \tilde{Z}_1^{2\kappa}\left(1 - F_{\tilde{Y}_1}\left(\tilde{Z}_1\right)\right) + 2\vartheta^{2\kappa}\left[\left(\frac{B_1}{\vartheta}\right)^{\kappa} - \left(\frac{\tilde{Z}_1}{\vartheta}\right)^{\kappa}\left(1 - F_{\tilde{Y}_1}\left(\tilde{Z}_1\right)\right)\right. \\ &\quad \left. + F_{\tilde{Y}_1}\left(\tilde{Z}_1\right)\right] + \tilde{Z}_1^{2\kappa}\left(1 - F_{\tilde{Y}_1}\left(\tilde{Z}_1\right)\right) \\ &= B_1^{2\kappa} + 2\vartheta^{2\kappa}\left[F_{\tilde{Y}_1}\left(\tilde{Z}_1\right) + \frac{1}{\vartheta^{\kappa}}\left(B_1^{\kappa} - \tilde{Z}_1^{\kappa}\left(1 - F_{\tilde{Y}_1}\left(\tilde{Z}_1\right)\right)\right)\right] \\ &= B_1^{2\kappa} + 2\vartheta^{2\kappa}\left[F_{\tilde{Y}_1}\left(\tilde{Z}_1\right) + \frac{1}{\vartheta^{\kappa}}\left(B_1^{\kappa} - \tilde{Z}_1^{\kappa}\exp\left\{\left(\frac{B_1}{\vartheta}\right)^{\kappa} - \left(\frac{\tilde{Z}_1}{\vartheta}\right)^{\kappa}\right\}\right)\right] \end{aligned} \tag{B.14}$$

Using the above expression, $E\left(\tilde{X}_1^{2\kappa}\right)$ is now given as:

$$\begin{aligned} E\left(\tilde{X}_1^{2\kappa}\right) &= E_{B_1}\left(E_{\tilde{Z}_1}\left(E_{\tilde{X}_1}\left(\tilde{X}_1^{2\kappa} | B_1, \tilde{Z}_1\right) \middle| B_1\right)\right) \\ &= E_{B_1}\left(E_{\tilde{Z}_1}\left(B_1^{2\kappa} + 2\vartheta^{2\kappa}\left[F_{\tilde{Y}_1}\left(\tilde{Z}_1\right) + \frac{1}{\vartheta^{\kappa}}\left(B_1^{\kappa} - \tilde{Z}_1^{\kappa}\exp\left\{\left(\frac{B_1}{\vartheta}\right)^{\kappa} - \left(\frac{\tilde{Z}_1}{\vartheta}\right)^{\kappa}\right\}\right)\right] \middle| B_1\right)\right) \\ &= E_{B_1}\left(B_1^{2\kappa}\right) + 2\theta^{2\kappa}\xi \end{aligned} \tag{B.15}$$

where:

$$\xi = \mu_{\tilde{\delta}} + \frac{1}{\theta^{\kappa}}\left(E_{B_1}(B_1^{\kappa}) - E_{B_1}\left(\exp\left\{\left(\frac{B_1}{\vartheta}\right)^{\kappa}\right\}E_{\tilde{Z}_1}\left(\tilde{Z}_1^{\kappa}\exp\left\{-\left(\frac{\tilde{Z}_1}{\vartheta}\right)^{\kappa}\right\}\right)\right)\right). \tag{B.16}$$

Using Eq. (B.9) and (B.15), we have that

$$\mathrm{Var}\left(\tilde{X}_1^{\kappa}\right) = E_{B_1}\left(B_1^{2\kappa}\right) + 2\theta^{2\kappa}\xi - \mu_{\tilde{X}}^2. \tag{B.17}$$

B.2.3 Covariance between $\overline{\tilde{\mathscr{D}}}_{K-k}$ and $\overline{\tilde{\mathscr{T}}}_{K-k}$

Before computing the covariance, we note that $\mathrm{Cov}\left(\overline{\tilde{\mathscr{D}}}_{K-k}, \overline{\tilde{\mathscr{T}}}_{K-k}\right) = \mathrm{Cov}\left(\frac{\sum_{j=1}^{m}\tilde{\delta}_j}{m}, \frac{\sum_{j=1}^{m}\tilde{X}_j^{\kappa}}{m}\right) = \frac{\mathrm{Cov}\left(\tilde{\delta}_1, \tilde{X}_1^{\kappa}\right)}{m}$, and

$$\mathrm{Cov}\left(\tilde{\delta}_1, \tilde{X}_1^{\kappa}\right) = E\left(\tilde{\delta}_1\tilde{X}_1^{\kappa}\right) - E\left(\tilde{\delta}_1\right)E\left(\tilde{X}_1^{\kappa}\right) = E\left(\tilde{\delta}_1\tilde{X}_1^{\kappa}\right) - \mu_{\tilde{\delta}}\mu_{\tilde{X}} \tag{B.18}$$

Now we need to compute $E\left(\tilde{\delta}_1\tilde{X}_1^{\kappa}\right) = E_{B_1}\left(E_{\tilde{Z}_1}\left(E_{\tilde{X}_1}\left(\tilde{\delta}_1\tilde{X}_1^{\kappa}|B_1,\tilde{Z}_1\right)\middle|B_1\right)\right)$. For fixed B_1 and $\tilde{Z}_1$, we observe that $\tilde{X}_1^{\kappa} = \tilde{Y}_1^{\kappa}\cdot\mathbb{1}_{\left(\tilde{Y}_1\leq\tilde{Z}_1\right)} + \tilde{Z}_1^{\kappa}\cdot\mathbb{1}_{\left(\tilde{Y}_1>\tilde{Z}_1\right)}$, and $\tilde{\delta}_1 = \mathbb{1}_{\left(\tilde{Y}_1\leq\tilde{Z}_1\right)}$. That is,

$$\tilde{\delta}_1\tilde{X}_1^{\kappa} = \begin{cases} \tilde{Y}_1^{\kappa} & \text{if } \tilde{Y}_1 \leq \tilde{Z}_1 \\ 0 & \text{if } \tilde{Y}_1 > \tilde{Z}_1 \end{cases}$$

Using Eq. (B.8), we find that

$$\begin{aligned} E_{\tilde{X}_1}\left(\tilde{\delta}_1\tilde{X}_1^{\kappa}|B_1,\tilde{Z}_1\right) &= E_{\tilde{X}_1}\left(\tilde{\delta}_1\tilde{X}_1^{\kappa}|\tilde{Y}_1\leq\tilde{Z}_1\right)P\left(\tilde{Y}_1\leq\tilde{Z}_1\right) + E_{\tilde{X}_1}\left(\tilde{\delta}_1\tilde{X}_1^{\kappa}|\tilde{Y}_1>\tilde{Z}_1\right)P\left(\tilde{Y}_1>\tilde{Z}_1\right) \\ &= E_{\tilde{Y}_1}\left(\tilde{Y}_1^{\kappa}|\tilde{Y}_1\leq\tilde{Z}_1\right)F_{\tilde{Y}_1}\left(\tilde{Z}_1\right) \\ &= B_1^{\kappa} - \tilde{Z}_1^{\kappa}\left(1 - F_{\tilde{Y}_1}\left(\tilde{Z}_1\right)\right) + \theta^{\kappa}F_{\tilde{Y}_1}\left(\tilde{Z}_1\right) \end{aligned}$$

Thus we can verify that

$$E\left(\tilde{\delta}_1\tilde{X}_1^{\kappa}\right) = E_{B_1}\left(E_{\tilde{Z}_1}\left(B_1^{\kappa} - \tilde{Z}_1^{\kappa}\left(1 - F_{\tilde{Y}_1}\left(\tilde{Z}_1\right)\right) + \vartheta^{\kappa}F_{\tilde{Y}_1}\left(\tilde{Z}_1\right)\middle| B_1\right)\right) = \vartheta^{\kappa}\xi \tag{B.19}$$

where ξ is given in Eq. (B.16). Therefore, the covariance between $\overline{\tilde{\mathscr{D}}}_{K-k}$ and $\overline{\tilde{\mathscr{T}}}_{K-k}$ is:

$$Cov\left(\tilde{\delta}_1, \tilde{X}_1^{\kappa}\right) = \vartheta^{\kappa}\xi - \mu_{\tilde{\delta}}\mu_{\tilde{X}}. \tag{B.20}$$

Using Eqs. (B.3, B.4, B.9, B.17, B.20) , we conclude by virtue of the multivariate central limit theorem that the joint asymptotic distribution of $\overline{\tilde{\mathscr{D}}}_{K-k}$ and $\overline{\tilde{\mathscr{T}}}_{K-k}$ is given as:

$$\lim_{m\to\infty}\sqrt{m}\left(\begin{pmatrix}\overline{\tilde{\mathscr{D}}}_{K-k}\\ \overline{\tilde{\mathscr{T}}}_{K-k}\end{pmatrix} - \begin{pmatrix}\mu_{\tilde{\delta}}\\ \mu_{\tilde{X}}\end{pmatrix}\right) \xrightarrow{d} \text{Normal}\left(\begin{pmatrix}0\\0\end{pmatrix}, \begin{pmatrix}\sigma_1^2 & \sigma_{12}\\ \sigma_{12} & \sigma_2^2\end{pmatrix}\right),$$

where:

$$\sigma_1^2 = \mu_{\tilde{\delta}}\left(1-\mu_{\tilde{\delta}}\right),$$

$$\sigma_{12} = \vartheta^{\kappa}\xi - \mu_{\tilde{\delta}}\mu_{\tilde{X}},$$

$$\sigma_2^2 = E_{B_1}\left(B_1^{2\kappa}\right) + 2\vartheta^{2\kappa}\xi - \mu_{\tilde{X}}^2,$$

$$\mu_{\tilde{\delta}} = 1 - E_{B_1}\left(\exp\left\{\left(\frac{B_1}{\vartheta}\right)^{\kappa}\right\}E_{\tilde{Z}_1}\left(\exp\left\{-\left(\frac{\tilde{Z}_1}{\vartheta}\right)^{\kappa}\right\}\right)\right),$$

$$\mu_{\tilde{X}} = E_{B_1}(B_1^{\kappa}) + \vartheta^{\kappa}\mu_{\tilde{\delta}},$$

$$\xi = \mu_{\tilde{\delta}} + \frac{1}{\vartheta^{\kappa}}\left(E_{B_1}(B_1^{\kappa}) - E_{B_1}\left(\exp\left\{\left(\frac{B_1}{\vartheta}\right)^{\kappa}\right\}E_{\tilde{Z}_1}\left(\tilde{Z}_1^{\kappa}\exp\left\{-\left(\frac{\tilde{Z}_1}{\vartheta}\right)^{\kappa}\right\}\right)\right)\right).$$