BAYESIAN STATISTICAL ANALYSIS

IN A PHASE II DOSE-FINDING TRIAL WITH SURVIVAL ENDPOINT

IN PATIENTS WITH B-CELL CHRONIC LYMPHOCYTIC LEUKEMIA

By SHIANSONG LI

A dissertation submitted to the

School of Public Health

Rutgers, The State University of New Jersey

In partial fulfillment of the requirements

For the degree of

Doctor of Public Health

Written under the direction of

Professor Weichung Joe Shih

And approved by

_____

_____

_____

_____

New Brunswick, New Jersey

May, 2014

ABSTRACT OF THE DISSERTATION

BAYESIAN STATISTICAL ANALYSIS

IN A PHASE II DOSE-FINDING TRIAL WITH SURVIVAL ENDPOINT

IN PATIENTS WITH B-CELL CHRONIC LYMPHOCYTIC LEUKEMIA

By SHIANSONG LI

Dissertation Director:

Professor Weichung Joe Shih

Bayesian approaches have been widely used in designing, monitoring and analyzing clinical studies in recent years. We utilize Bayesian parametric and non-parametric statistical methods in interim monitoring and decision-making for a phase II dose-finding trial with survival endpoint. The objective of the clinical trial is to find an optimal treatment schedule at the end of the study for planning future studies, using Bayesian decision rules. The primary efficacy outcome is time to progression. Binomial-Beta model and Exponential-Gamma model are included in parametric methods. Non-parametric methods include Bayesian life-table, Beta process model, Dirichlet process model and Gibbs sampling. Simulations are conducted for each of the statistical methods under 9 different scenarios including truncated exponential entry time, and the probability of a treatment-schedule being chosen is calculated based on 1,000 simulation studies. Finally, these different statistical methods are used to find optimal treatment-schedule among 3 arms in the phase II CLL clinical trial using the most recent unblinded data.

ACKNOWLEDGMENT

I wish to express my gratitude and sincere appreciation to Dr. Weichung Joe Shih, my dissertation research advisor, for his academic guidance and supervision on this dissertation. Thanks also go to other members of my dissertation committee: Dr. Sandra Echeverria, for her invaluable advice and suggestions on the rationale and writing of the dissertation; Dr. Sinae Kim, for her guidance and expertise on simulations and non-parametric process models; Dr. Xiaolong Luo, who originally introduced the CLL clinical trial with Bayesian adaptive design to be used for this dissertation, for his perspective and invaluable recommendations during the construction of the dissertation.

Special gratitude is extended to Dr. Arlene Swern, who is my supervisor at Celgene Corporation, for her patience and support for this dissertation. I am also indebted to Celgene's CLL-009 clinical study team, Drs. Jay Mei and Jennie Zhang, for their permission and support in using the data to illustrate the statistical methodologies. Last but not least, I am obliged to Dr. Alan Wu for his expertise and assistance in R coding and simulations.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

1. INTRODUCTION

1.1. Background

Chronic lymphocytic leukemia (CLL) is the most prevalent adult leukemia in the Western world (National Cancer Institute, 2010). CLL patients who relapse following purine-analog or bendamustine-based treatment have poor prognosis (Wendtner et al., 2011). These patients have limited treatment options and new agents or treatment strategies are needed.

Both safety and efficacy are major concerns of CLL patients in clinical trials, so dose selection and treatment management of biological agents are important and critical in such disease population. Ferrajoli et al. (2008) used a dose escalation scheme starting with 10 mg/day of lenalidomide given for 21 days in the first 28-day cycle followed by titration upward by 5 mg increments every 28 days to a maximum dose of 25 mg daily. In the Ferrajoli study, time to best response was 6 months in 11 patients and 9 months in 3 patients. The study demonstrated the starting dose of 10 mg resulted in 32% overall response rate with no episodes of tumor lysis syndrome (TLS) and a tumor flare reaction (TFR) rate of 30%. To achieve clinical efficacy, a higher starting dose such as previously reported by Ferrajoli et al. (2008) may be needed.

To identify a safe and clinically active starting dose, a phase II study to evaluate lenalidomide at different starting dose levels had been recently conducted in the setting of relapsed or refractory B-cell CLL (Wendtner et al., 2011, 2012a, 2012b). In this phase II,

multicenter, randomized, double-blind, parallel-group trial, three different starting dose administration schedules of 5 mg, 10 mg and 15 mg daily respectively were evaluated. These starting doses of 5 mg, 10 mg and 15 mg were followed by a step-wise dose escalation every 28 days to a maximum dose of 25 mg daily as tolerated to allow for fewer escalations to reach a higher dose. The study had n = 104 subjects enrolled in the three dose-administration schedules.

The objectives of this Phase II clinical trial (Wendtner et al., 2011, 2012a, 2012b) were to evaluate the efficacy and safety of different lenalidomide dose administration schedules in subjects with relapsed or refractory B-cell CLL. The study endpoints include type and frequency of toxicities, response rate and time to response, time to progression, progression-free survival (PFS) and overall survival (OS). The trial used Bayesian adaptive design in interim monitoring both efficacy and safety outcomes.

1.2. Trial Design

1.2.1. Randomization

It was reported by Wendtner et al. (2011, 2012a, 2012b) that subjects met all eligibility criteria specified in the protocol to be randomized for the clinical trial. Initially subjects were randomized (1:1:1) in a double-blind fashion to the 5 mg, 10 mg, and 15 mg starting dose administration schedules (denoted as schedule A, B and C, respectively). The randomization procedure was conducted by a validated interactive voice response system (IVRS). Subjects that initially were allocated to three different administration schedules escalated dose every 28 days, based on individual subject tolerability, as follows:

Administration Schedule A: 5 mg →10 mg →15 mg →20 mg →25 mg/daily

Administration Schedule B: 10 mg →15 mg →20 mg →25 mg/daily

Administration Schedule C: 15 mg →20 mg →25 mg/daily

Toxicities requiring dose interruption or modification followed detailed guidelines in the protocol. Subjects unable to escalate to the 25 mg maximum dose due to toxicity might continue with the highest dose achieved in the previous cycle or as indicated in the dose interruption/modification guideline in the trial.

Subjects continued in the clinical trial and received study drug until discontinuation from the study for any of the following reasons: disease progression; unacceptable AEs at the discretion of the investigators, subject withdrawal consent, subject lost to follow-up, death, protocol violation and other reasons that in the judgment of the investigator might rule out continuation of study drug. For those subjects who discontinued study drug for reasons other than disease progression or withdrawal of consent, study visits might continue every 28 days to assess response until documentation of disease progression or until a new CLL therapeutic administration schedule was started, whichever came first.

The protocol (Wendtner et al., 2011, 2012a, 2012b) stated that, after 18 subjects completed one 28 day cycle, an interim analysis was to be conducted to review the safety of each starting dose arms. Subsequent interim analyses to review the safety and progression rate of each starting dose arms occurred at 13-week intervals. After review of each interim analysis, an administration schedule might be dropped according to interim decision rules. At any time accrual to a starting dose administration schedule was

also stopped if unacceptable toxicities were observed in that schedule. If an administration schedule was dropped then all new subjects would be randomized equally into the remaining schedules.

1.2.2. Bayesian Adaptive Design

The trial protocol specified using a Bayesian approach to evaluate both efficacy and toxicity outcomes. This was a Bayesian adaptive design which considered modeling efficacy and toxicity outcomes to stop randomization of less promising schedules (Bekele and Shen, 2005; Ji and Bekele, 2009) following interim analysis. The efficacy outcome was probability of progression. The safety outcomes were several critical toxicities. Accrual to a starting dose schedule was stopped if unacceptable toxicities or progression rate were observed in that schedule. If an administration schedule was dropped then all new subjects would be randomized equally into the remaining schedules. If no schedule was dropped in any interim, all schedule arms would continue for a final decision making. The detailed interim monitoring and stopping rules using Bayesian posterior probabilities are presented in Section 2.1 of Chapter 2.

1.3. Data Collection

1.3.1. Response Assessment

Tumor response was assessed according to the International Workshop on Chronic Lymphocytic Leukemia (iwCLL) guidelines for the diagnosis and treatment of CLL (Hallek, 2008), including complete response (CR), partial response (PR), stable disease

(SD) and progressive disease (PD). Evaluation of response was performed after 3 cycles of therapy and every 4 weeks thereafter.

An independent Response Adjudication Committee (iRAC) performed a blinded, independent assessment of response (including the development of PD) prior to database lock. The iRAC adjudicated response data was used in the efficacy analysis for the trial. Probability of progression (development of PD) was estimated using the protocol-specified method in Chapter 2 and would also be evaluated by other Bayesian statistical methods in Chapter 3. Time to progression from randomization was used for survival analysis to estimate and compare the probability of progression for all dose schedules.

1.3.2. Toxicities Assessment

Based on initial clinical data from other trials, the most important toxic events for treatment in patients with relapsed or refractory CLL include grade ≥ 2 Tumor Lysis Syndrome (TLS), grade 4 neutropenia and/or thrombocytopenia, and febrile neutropenia. For this B-cell CLL trial, toxicity was characterized into various types with varying degrees of severity. Three major types of toxicities were assigned by different weight for clinical importance and a toxicity score was defined as weighted average of rates of toxicities in expression (2.1) in Chapter 2. Based on pre-simulation of the study, weight $w_3 = 1.0$ was chosen for grade ≥ 2 TLS, $w_2 = 0.6$ for grade 4 neutropenia and/or thrombocytopenia, and $w_1 = 0.1$ for febrile neutropenia. The ordering of w_l reflected the toxicity severity level which represented the PIs expert opinion and might take on any

positive values. Because only relative magnitudes of the toxicity weights mattered, any scores from the PIs could be scaled to be between 0 and 1.

Since multiple toxicities were being considered, a means of reducing the dimensionality of the toxicity parameters was introduced. The protocol used the approach in Bekele and Thall (2004) and Bekele et al. (2010) which incorporated medical knowledge into the dimension reduction process and assigned toxicity severity weights characterizing the importance of each type of toxicity.

1.4. Objectives of the Dissertation

The phase II dose-finding study with lenalidomide (Wendtner et al., 2011, 2012a, 2012b) in patients with B-cell CLL has been conducted by Celgene Corporation since 2009. The company has completed several interim analyses to evaluate efficacy and safety of different dose schedules. The sample size at each interim analysis was small, and no administration schedule has been stopped according to interim stopping rules. In order to establish an optimal dose regimen for larger phase III study, the protocol proposed decision rules on final data to choose optimal regimen. The company is preparing for final database lock in later 2014.

Bayesian statistics has become more and more popular in statistical literature in recent years. However, Bayesian approaches and their applications to real clinical trial data are very limited in pharmaceutical industry (Brannath et al., 2009; Scala and Glimm, 2011). The primary objective of the dissertation is to apply several different Bayesian statistical methods to the most recent unblinded data from the phase II clinical trial (Wendtner et

al., 2011, 2012a, 2012b) as of January 2014 and to evaluate its efficacy parameter – probability of progression. These Bayesian approaches can be classified into Bayesian parametric and non-parametric methods. To the best of my knowledge, systematic exploration of these Bayesian methods and their applications to clinical trials with survival endpoint have not appeared in literatures.

Another objective of the dissertation is to conduct simulations and compare the results from these Bayesian methods. Applying several different Bayesian methods in one clinical trial has not been seen in literature. In addition, using Bayesian posterior probability to quantify clinical outcomes is intuitively acceptable and easily understood by clinicians. Thus the Bayesian approaches offer new views and perspectives to clinical researchers.

Since no treatment schedule has been dropped so far during this clinical trial, I will not consider the interim analyses and related issues in the dissertation. Instead, I will focus on the Bayesian statistical methods and apply them to the January 2014 unblinded data. The statistical methods evaluate the probability of progression and compare dose schedules according to efficacy parameter - progression rate so that an optimal schedule can be recommended for future studies. The same statistical methodology can be applied to safety parameters such as toxicities and to each interim dataset as well as to other similar studies.

The remainder of the thesis is organized as follows. In Chapter 2, I will briefly review the protocol pre-specified method. Chapter 3 introduces other Bayesian statistical

approaches. Simulation results from different statistical methods are presented in Chapter 4. Chapter 5 presents results from the CLL clinical trial by several Bayesian approaches, followed by discussion and conclusion in Chapter 6.

2. REVIEW OF PROTOCOL PRE-SPECIFIED METHOD

The phase II CLL clinical trial was a schedule-administration (or treatment strategy) design. The sponsor was interested in evaluating whether various intra-patient dose-escalation schemes were safe and effective while monitoring safety and efficacy. Specifically, while treatment started at various doses ranging from very low doses (5 mg/daily) to moderately low (15 mg/daily), the goal was to perform intra-patient dose escalation (every 28 days as tolerated) until a maximum dose of 25 mg/daily was achieved. The decision to escalate was based on how well the subject tolerates the lower dose levels. While these dose escalations were taking place, toxicity and the progression rates were monitored.

The trial protocol pre-specified a Bayesian method that modeled both toxicity and disease progression using an extension methods developed by Ji and Bekele (2009). Toxicity and disease progression were major events to be monitored so time-to-event models were used. These models were not used to characterize time-to-event outcomes but to compare toxicity risk and progression rates among treatment schedules so as to help decision makers in selecting dose schedule(s) for a larger phase III study.

2.1. Decision Rules

Subjects were randomized to one of the three administration schedules. Interim analyses occurred at 13-week intervals to monitor both toxicity and disease progression. **Interim decision rules:** accrual into any of the three administration schedules was stopped if:

- The probability that schedule had the smallest progression rate was less than 0.05 or
- The probability that schedule had the highest toxicity score was greater than 0.90 or
- The probability that schedule's toxicity score was more than 0.20 was greater than 0.90 or
- The probability that schedule had a more than 5% ≥ Grade 2 TLS rate was greater than 0.90.

Let $\theta_{r,k}$ denote the probability of progression for the *k*th administration schedule, $\theta_{r,(-k)}$ the probability of progression for administration schedule(s) other than *k*, and $\theta_{t(l),k}$ the probability of toxicity type *l* for the *k*th administration schedule. Toxicity type and weight were described in subsection 1.3.2. Dimension reduction is achieved by calculating the toxicity score for the *k*th administration schedule as

$$\psi_k = \sum_{l=1}^{L} w_l \theta_{t(l),k} \qquad (2.1)$$

where w_l is the weight of toxicity type *l*. Excessive toxicity is defined as when toxicity score is over cutoff $\psi_0 = 0.20$ for this specific trial. The ψ_0 is called targeted average toxicity score and is determined by similar approach in Bekele and Thall (2004) and Bekele et al. (2008).

Based on the decision rules, while the study was ongoing, the *k*th administration schedule was declared unacceptable if, for the posterior probabilities,

$$\Pr(\theta_{r,k} < \theta_{r,(-k)} \mid data) < 0.05 \text{ or}$$

$$\Pr(\psi_k > \psi_{(-k)} \mid data) > 0.90 \text{ or}$$

$$\Pr(\psi_k > 0.20 \mid data) > 0.90 \text{ or}$$

$$\Pr(\theta_{t(3),k} > 0.05 \mid data) > 0.90$$

If an administration schedule was dropped then all new subjects would be randomized equally into the remaining schedules. If no schedule was dropped in any interim, all schedule arms would continue for a final decision making.

Final decision rules: at the end of the study, the schedule arm with the highest probability of having the lowest progression rate and having lower than cutoff $\xi = 0.90$ probability of excessive toxicity would be chosen; i.e., arm *k* would be selected for phase III study if

$$\left.\begin{array}{l} Max(\Pr(\theta_{r,k} < \theta_{r,(-k)} \mid data)), \\ \text{and} \\ \Pr(\psi_k > \psi_0 \mid data) < \xi \end{array}\right\} \qquad (2.2)$$

where $\xi = 0.90$ was chosen to reflect the PIs wish the algorithm to behave conservatively and was guided by preliminary computer simulations (Bekele et al., 2008; Bekele et al., 2010).

As noted previously, in this dissertation I will use the January 2014 unblinded data to study Bayesian methods and to focus on the decision rule for disease progression which is to select a dose schedule with $Max(\Pr(\theta_{r,k} < \theta_{r,(-k)} \mid data))$. The same methodology can be applied to decision rule for toxicity $\Pr(\psi_k > \psi_0 \mid data) < \xi$ and to any interim data.

2.2. Protocol Pre-specified Bayesian Method

The following briefly reviews the basis of the methodology given in the study protocol (Ji and Bekele, 2009).

2.2.1. Discrete Time Hazard, Survival and Likelihood

The protocol uses a so-called ordinal sequential model with discrete time hazards to define the likelihood for progression and toxicity. For ease of following presentation, we suppress the administration schedule (A, B or C) and event (progression or toxicity) as well as patient subscripts. For a given patient, let Y represent either the time-interval in which an event is observed or the last follow-up visit (whichever is less); $Y = 1,..., J$. Let d be a failure (or censoring) indicator where $d = 1$ indicating a failure event and $d = 0$ indicating a censored observation. Ignoring censoring for the moment, note that for this

equally spaced interval and each interval being one unit time, the discrete time hazard for the *j*th interval is equal to

$$h_j = \Pr(Y = j \mid Y \geq j) = \frac{\Pr(Y = j)}{\Pr(Y \geq j)} = \frac{\theta_j}{S_j}, \quad j = 1,..., J$$

where S_j is survival function at interval *j* (probability of event at the beginning or beyond interval *j*) and θ_j is probability of event at "tiny" interval *j* for discrete-time case which is also called probability mass function.

Moreover note that

$$1 - h_j = \frac{S_j - \theta_j}{S_j} = \frac{S_{j+1}}{S_j}$$

This allows us to exploit a recursive relationship between the hazard function during interval *j* and the probability of event beyond interval *j* via

$$S_j - \theta_j = (1 - h_j) S_j = S_{j+1}$$

Since $S_1 = 1$, we then have

$$S_2 = 1 - h_1$$

$$S_3 = (1 - h_1)(1 - h_2)$$

$$\vdots$$

$$S_j = \prod_{i=1}^{j-1} (1 - h_i)$$

So the contribution to the likelihood of an event that occurs in the *j*th interval is

$$\Pr(Y = j, d = 1) = \theta_j = h_j S_j = h_j \prod_{i=1}^{j-1} (1 - h_i)$$

The contribution to the likelihood for a censored observation having maximum follow-up in the *j*th interval is the discrete time survival function:

$$\Pr(Y = j, d = 0) = \Pr(Y \geq j) = S_j = \prod_{i=1}^{j-1} (1 - h_i)$$

We can write the two contributions in a single expression so contribution to the likelihood for an observation is:

$$L_j = h_j{}^d S_j = h_j{}^d \prod_{i=1}^{j-1} (1 - h_i)$$

2.2.2. Likelihood and Probability of Event

The likelihood of progression event that occurs in the *j*th interval for the *k*th schedule arm can be constructed as described in Section 2.2.1. We only need to add subscript *r* to represent progression event, $t(l)$ to represent toxicity event of type *l* (for *l* = 1, …*L*), and *k* to represent *k*th schedule arm.

For purposes of study monitoring, independence among the various toxicity types and between these toxicities and progression is assumed. Overall progression rate for the *k*th schedule arm at j intervals having elapsed is

$$\theta_{r,k} = 1 - \prod_{i=1}^{j} (1 - h_{r,i,k}) \qquad (2.3)$$

and j intervals toxicity rate for the lth type of toxicity for the kth schedule arm is

$$\theta_{t(l),k} = 1 - \prod_{i=1}^{j} (1 - h_{t(l),i,k}) \qquad (2.4)$$

Note that the θ in (2.3) and (2.4) doesn't have time interval subscript j and indicates overall event rate up to interval j, but θ_j indicates event rate at interval j. Apparently the overall event rate can be obtained through (2.3) or (2.4) when hazard or survival is known. One can also simply accumulate $\theta_i's$ for $i \le j$ to get the overall event rate when $\theta_i's$ are known.

2.2.3. Prior Distributions of Hazard Rate

An informative prior represents existing knowledge of parameter(s) of interest, which can come either from historical clinical trial data, or from expert opinions. There is no firm guideline on method of prior elicitation. Usually it is desirable to have investigators to provide prior data and to weight it so that statisticians can derive clinically meaningful and computationally feasible prior distribution. The prior distribution is then incorporated into the current study and help make decisions. The appropriate weight or prior sample size controls its impact on current study. Non-informative prior can also be used if no prior information is available.

When the study was designed, the clinical team consulted expert physicians regarding the rate of progression for the patient population in B-cell CLL. The following clinical priors for hazard rate of progression and hazard rate of toxicity event are set for the study:

$$h_{r,j,k} \sim Beta(0.025,0.975),$$

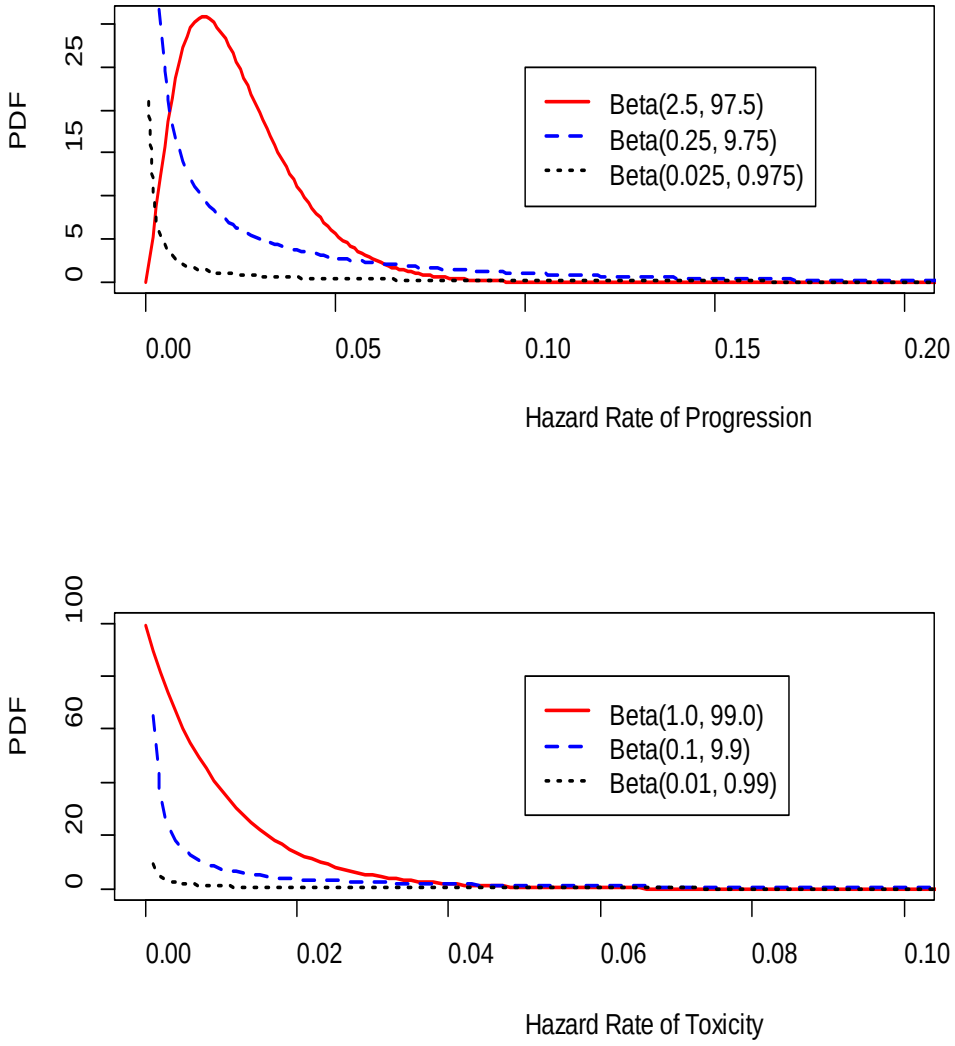
and

$$h_{t(l),j,k} \sim Beta(0.01,0.99)$$

where l = 1,2,3, and j = 1,…, J, and k = 1,2,3.

The expected hazard rate in a given time interval is 2.5% for the prior of progression event and is 1.0% for toxicity event. The prior sample size is set to be 1 since the prior information is weak according to expert physicians. Other priors with the same expected rate but increasing sample size are plotted in Figure 1 (plot_beta.R). With very small sample size, prior plays very limited role in posterior estimation. Compared to other priors, *Beta*(0.025, 0.975) or *Beta*(0.01, 0.99) distribution is so disperse that they reflect weak information about the prior hazard rate.

Figure 1: Prior Probability Density Functions of Hazard Rate



2.2.4. Posterior Distributions of Hazard and Probability of Event

Posterior distributions of hazard of progression and toxicity events in each time interval are also beta distributions since the data is binomially distributed with conjugate priors. Section 3.2.1 will detail the posterior distribution for binomially distributed data. If we replace with the posterior distributions of hazard of progression or toxicity event in the right-hand side of (2.3) and (2.4), we get posterior probability of progression or toxicity event. If posterior probabilities of toxicity events are utilized in the right-hand side of (2.1), we have posterior toxicity score. By applying decision rules of (2.2) we can make final dose schedule selection.

2.2.5. Relationship Between Event and Survival Probabilities

Some methods to be presented focus on event probability and some focus on survival probability. As a general note, event probability $\theta(t)$ and survival probability $S(t)$ at time *t* are complementary, i.e., $\theta(t) = 1 - S(t)$. Event probability during interval $j = (t_{j-1}, t_j]$ is $\theta_j = S_{j-1} - S_j$ where $S_j = S(t_j)$.

2.2.6. Bayesian Simulation

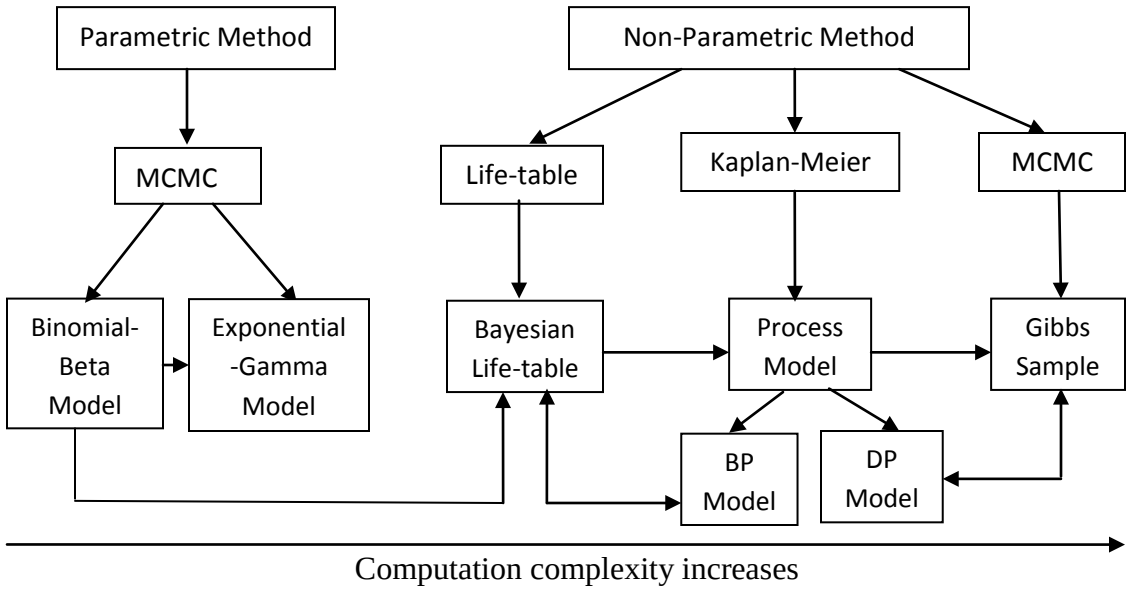
Bayesian inference uses simulation draws from (2.3) and (2.4) and will be further discussed in Section 3.1.2. Simulation results guide interim and final schedule arm selections following decision rules. In this dissertation, I will apply this protocol pre-specified method along with other Bayesian statistical methods presented in Chapter 3 to the January 2014 unblinded data from the CLL phase II clinical trial.

3. BAYESIAN STATISTICAL METHOD

The present thesis will apply several statistical methods to the estimation of event probability (progression or toxicity rate). Since the data is time to event variable, the survival probability is usually to be focused in many of these methods. The methodology can be classified as parametric and non-parametric methods. One of the Bayesian non-parametric methods is protocol pre-specified which was reviewed in Section 2.2. In this Chapter, we present the statistical methodology of the other five Bayesian models. Figure 2 describes relationship among these methods and traditional methods including MCMC (Markov chain Monte Carlo), life-table and Kaplan-Meier. Bayesian methods are basically extension of traditional methods by incorporating prior information. Figure 2 also shows some relationship among and complexity of these methods in terms of posterior computation. In the next Chapter, we will apply these methods to the B-cell CLL phase II clinical trial data.

1. Binomial-Beta model – Beta prior in conjunction with binomially distributed data
2. Exponential-Gamma model – Gamma prior in conjunction with exponentially distributed data
3. Protocol pre-specified Bayesian life-table method
4. Bayesian Beta process (BP) model
5. Bayesian Dirichlet process (DP) model
6. Monte Carlo Bayesian method or Gibbs sampler

Figure 2: Statistical Methods and Some Relationships



3.1. Bayesian Method Overview

The essential characteristic of Bayesian method is its explicit use of probability for quantifying uncertainty in statistical inferences. Bayesian analysis is to calculate and interpret the appropriate *posterior distribution* – the conditional probability distribution of the unobserved quantities of interest (*parameter* of interest) given the observed data.

In general, we use θ to denote unobservable vector quantities or population parameters of interest (unknown probability of event or survival in our case), y denote the observed data (outcome – number of event or time interval in which the event is observed). Bayesian inference about a parameter θ is made in terms of probability statement conditional on observed value of y which is denoted by $p(\theta \mid y)$. A general discussion of Bayesian methods in clinical trials is given by Spiegelhalter, et al. (2004).

3.1.1. Bayesian Rule

The core of Bayesian inference is to perform necessary computations to summarize $p(\theta \mid y)$ in appropriate ways. The *Bayesian rule* states that resulting posterior distribution of θ is proportional to the product of prior distribution $p(\theta)$ and data (or sampling) distribution $p(y \mid \theta)$:

$$p(\theta \mid y) = \frac{p(\theta, y)}{p(y)} = \frac{p(\theta)\, p(y \mid \theta)}{p(y)} \propto p(\theta)\, p(y \mid \theta) \qquad (3.1)$$

To evaluate the fit of the model, the dependence (or sensitivity) of conclusions on 'subjective' prior distribution should be examined and explored. Sampling distribution can play an important role in checking model assumptions. An applied Bayesian

statistician should be willing to apply Bayes' rule (3.1) under a variety of possible models (Gelman et al., 2004). In most cases, scientific judgment is necessary to specify both prior distribution and data distribution. In this research, we will utilize both Bayesian parametric models and non-parametric models to estimate probability of progression and probability of toxicity, which are our parameters of interest, and then use simulations in Chapter 5 to compare these parameters among administration schedule to guide decision-making.

3.1.2. Bayesian Inferences by Simulation of Posterior Distribution

Simulation is an important part of applied Bayesian analysis. We will use simulation to make random draws from joint posterior distribution of parameter(s) of interest. The basic structure of simulation draws is shown in Figure 3 (Gelman et al., 2004). We let $g = 1, 2, \ldots, G$ denote simulation draws and $(\theta_1^g, \theta_2^g, \ldots, \theta_k^g)$ the corresponding joint draw of parameters from their joint posterior distribution. Here we have a total of G simulations. Parameters $(\theta_1, \theta_2, \ldots, \theta_k)$ can represent probabilities of event (such as progression or toxicity) at each time interval from *1* to *k* for non-parametric statistical methods. After drawing random samples from the posterior probability distribution, we can make inferences such as mean, median and 95% posterior interval for the parameters of interest from these samples. We can also estimate the posterior probability of any event such as $\Pr(\theta_j < a)$ which usually answers some scientific questions for clinical trials and is actually one of the features of Bayesian statistics.

Histogram of a set of random draws from the distribution is a very useful and intuitive tool for summary statistics. Given a large enough sample, histogram can provide practically complete information about the density, such as mean, median, percentiles and other summary statistics thus providing estimates of any aspects of the distribution.

Figure 3: Structure of Simulation Draws From Joint Distribution of Parameters

Simulation draw	Parameters
g	$\theta_1, \theta_2, \ldots, \theta_k$
1	$\theta_1^1, \theta_2^1, \ldots, \theta_k^1$
2	$\theta_1^2, \theta_2^2, \ldots, \theta_k^2$
…	…
G	$\theta_1^G, \theta_2^G, \ldots, \theta_k^G$

3.1.3. Bayesian Decision Rules in Literature

Bayesian approaches to the design, monitoring and analysis of randomized clinical trials have received great attention in recent years. Suppose that a group sequential clinical trial is to compare two treatments and that the true treatment difference is summarized by a parameter θ, where large value of θ corresponds to superiority of the new treatment. For survival data with time-to-event endpoint, θ usually denotes a log hazard ratio. Under the Bayesian paradigm, at interim stage m, interim stopping rule are usually defined as (Spiegelhalter, Freedman and Parmar, 1994):

(1) stop the trial early for futility, if $P(\theta > \theta_L \mid Y_m) < \delta_1$;

(2) stop the trial early for overwhelming efficacy, if $P(\theta > \theta_U \mid Y_m) > \delta_2$;

where Y_m denotes the data accumulated up to stage m, θ_L denotes the lower bound of clinical equivalence, θ_U denotes the upper bound of equivalence which is also the amount of improvement or margin of benefit. In practice, θ_L and θ_U can be fixed constants for inferiority or superiority specified by clinician, and δ_1 and δ_2 are stopping boundaries and might set to be 2.5% and 97.5% respectively. Similar rules have been proposed by Berry (1985), Freedman and Spiegelhalter (1989), Thall and Simon (1994). These interim decision rules for futility or efficacy are based on posterior distribution. Most recently, Bayesian predictive power (PP) is used to guide interim treatment adaptations in confirmatory oncology trials (Schmidli et al., 2007; Brannath et al., 2009; Scala and Glimm, 2011). These decision rules use only efficacy endpoint(s) as a guideline but toxicity endpoint is not formally incorporated into the decision making.

In many circumstances, especially in oncology trials, treatments may be so unequal in their risk in terms of toxicity. Treatment schedule- or regimen-related severe toxicity occurs routinely and the case of multiple patient outcomes is quite common, so one needs to consider both benefit and risk (efficacy and toxicity) to decide whether to stop the trial early or to make interim treatment selections. Thall, Simon and Estey (1995, 1996) use Bayesian strategy for monitoring safety and efficacy in single-arm clinical trials, where response rate is the primary efficacy endpoint. Thall and Sung (1998) extend the application for monitoring multiple outcomes using Dirichlet distribution and specify the early stopping rules as follows for randomized controlled clinical trials:

(1) stop the trial early for toxicity, if $P(\theta_{t,E} > \theta_{t,C} + \theta_{t,0} \mid Y_m) > \delta_1$;

(2) stop the trial early for futility, if $P(\theta_{r,E} > \theta_{r,C} + \theta_{r,0} \mid Y_m) < \delta_2$;

(3) stop the trial early for overwhelming efficacy, if $P(\theta_{r,E} > \theta_{r,C} + \theta_{r,0} \mid Y_m) > \delta_3$

where $\theta_{t,E}$ and $\theta_{t,C}$ denote toxicity rates for experimental and control arms; $\theta_{r,E}$ and $\theta_{r,C}$ denote response rates for experimental and control arms; $\theta_{t,0} > 0$ and $\theta_{r,0} > 0$ denote maximum allowed increase in toxicity and desired increase (targeted improvement) in efficacy so they represent the trade-off between safety and efficacy (Conaway M. and Petroni G, 1996; Thall and Sung, 1998); δ_1, δ_2 and δ_3 are stopping boundaries. In these applications, both efficacy and safety outcomes are modeled as binary or multinomial data, and timing of the outcome is not considered. Bekele and Shen (2005) propose a Bayesian approach to a phase I/II dose-finding trial by jointly modeling a binary toxicity outcome and a continuous biomarker expression outcome via a bivariate continuous-

binary model, and their decision rules are based on the posterior distributions of both toxicity and activity. Basically, for multi-arm trials, stop enrollment in the specific administrative schedule if toxicity rate is unacceptably high or response rate is unacceptably low. Thall (2012) recently presents a utility-based clinical trial design and analysis which uses utility score to obtain a one dimensional criterion reflecting the relative importance of two or more outcomes for decision making. The approach to dose-finding is based on joint utilities of ordinal (response and toxicity) outcomes.

In oncology trials, response is actually a surrogate efficacy endpoint for time to progression, especially for short-term or phase II clinical studies. Disease progression is usually a primary efficacy endpoint for long-term or phase IIB/III studies and it's a time-to-event variable. Follmann and Albert (1999) address the case of a binary outcome defined in terms of a censored time-to-event variable. They assume a Dirichlet prior on discrete-time probabilities and approximate posterior distribution which is a mixture of Dirichlet processes by a data augmentation algorithm. Rosner (2005) takes a similar approach but uses Gibbs sampling to generate posteriors. Cheung and Thall (2002) propose an adaptive Bayesian method for continuous monitoring the probabilities of composite time-to-events in phase II trials. They use an approximate posterior to compute an early stopping criterion and extend to a 3-arm leukemia trial. In their proposed interim monitoring, arm *k* is dropped if it is inferior to the others in terms of response rate: $P(\theta_k < \max(\theta_{-k}) \mid data) > 0.90$, where θ_{-k} denotes response rate(s) for all arm(s) except the *k*th arm. Ji and Bekele (2009) present an outcome-adaptive randomization scheme for comparative trials in which the primary endpoint is a joint

efficacy/toxicity outcome. They extend the approximate Bayesian posterior model in Cheung and Thall (2002) to bivariate time-to-event outcomes by data augmentation algorithm or latent model.

The goal of phase II trial is not to obtain confirmatory comparative results, but to select one or more of the experimental treatments (or schedules) for subsequent evaluation. If the trial does not stop early, one may use any appropriate criteria using final data at the end of the study to determine if further evaluation is warranted in a phase III trial. The Bayesian approaches we present in this Chapter make a number of simplifying assumptions. In contrast to the bivariate model (Cheung and Thall, 2002), we do not consider jointly modeling efficacy and safety parameters but only combining evidence from efficacy and safety outcomes due to the reasons as discussed by Scala and Glimm (2011): (i) setting up a reasonable bivariate prior for progression and toxicity jointly in the absence of strong prior beliefs is a very difficult task; (ii) estimation of the joint probabilities of progression and toxicity and correlation between efficacy and safety outcomes are usually not of concern in practice; (iii) a relatively simple rule or methodology serves our purpose well enough as the statistical approach is used for selecting an administrative schedule to be continued into phase III trial and not for ultimate claim about efficacy of a treatment schedule.

The CLL clinical trial combines the efficacy and toxicity outcomes in decision making using criteria described in Section 2.1. In this dissertation, we only model the efficacy outcome which is probability of progression. Similar models can be applied to toxicity outcomes. These approaches model the parameters of interests directly without using

latent variable so as to straighten the applications. The rest of this Chapter describes these Bayesian parametric and non-parametric models for any administration schedule.

3.2. Bayesian Parametric Method

3.2.1. Binomial-Beta Model

Our aim is to estimate probability of progression θ (from now on we will only mention progression event, and toxicity event will apply to the same statistical methodology) at a time point of or the end of a clinical trial for the population under study. This is a typical problem in clinical trials such as estimating the probability of failure (or success) in a population.

The outcome or data y from the clinical trial is the number of failures (patients having progressive disease - PD) from a total of n patients at risk in a given time period. In this section, we assume that we only know patient's disease status (PD or not) at a given time but we don't know the exact time of PD. We simply consider binomially distributed data first. In section 3.2.2, we will incorporate time of PD into modeling.

Data distribution - binomial model

$$p(y \mid \theta) \propto \theta^{y}(1-\theta)^{n-y}$$

or

$$y \mid \theta \sim Bin(y \mid n, \theta) \tag{3.2}$$

where θ is event rate or probability of event (unknown).

Prior distribution

$$p(\theta) \propto \theta^{\alpha-1}(1-\theta)^{\beta-1}$$

or

$$\theta \sim Beta(\alpha, \beta) \tag{3.3}$$

This is a *conjugate* prior for the binomial model, where $\alpha > 0, \beta > 0$ and $\alpha + \beta$ implies prior sample size.

Posterior distribution

$$p(\theta \mid y) \propto \theta^{\alpha+y-1}(1-\theta)^{\beta+n-y-1}$$

or

$$\theta \mid y \sim Beta(\alpha + y, \beta + n - y) \tag{3.4}$$

which has mean and variance,

$$E(\theta \mid y) = \frac{\alpha + y}{\alpha + \beta + n}$$

$$\text{var}(\theta \mid y) = \frac{(\alpha + y)(\beta + n - y)}{(\alpha + \beta + n)^2(\alpha + \beta + n + 1)}$$

We can think of the first parameter of a beta distribution (3.4) as the number of failures (or PDs) and the second parameter as the number of successes (non-PDs).

3.2.2. Exponential-Gamma Model

If our interest is to model time to progression, we can use exponential model. The problem then becomes to a typical survival analysis issue and survival or event probability at any time point can be estimated by methods of survival data – where outcome *y* is time to failure (or survival time). Note that in most cases throughout the thesis other than in exponential-gamma model here, the outcome *y* represents number of failures.

Data distribution - exponential model

$$p(y \mid \theta) = \theta \exp(-y\theta), \text{ for } y > 0$$

or

$$y \mid \theta \sim Exp(y \mid \theta) \tag{3.5}$$

where $\theta = 1/E(y \mid \theta)$ is called 'rate' or 'hazard rate' (unknown). Exponential distribution is a special case of gamma distribution and has a 'memoryless' property that makes it a natural model for survival or lifetime data.

The sampling distribution of *n* independent exponential observations, $y = (y_1,...,y_n)$, with constant rate θ, or likelihood of θ is:

$$p(y \mid \theta) = \theta^n \exp(-\theta \sum_{i=1}^{n} y_i), \text{ for } y_i > 0 \tag{3.6a}$$

For right-censored data, denote failure (or censoring) indicators by $(d_1,...,d_n)$ where $d_i = 0$ if y_i is right-censored and $d_i = 1$ if y_i is a failure time. It can be shown that

$$p(y \mid \theta) = \theta^{d} \exp(-\theta \sum_{i=1}^{n} y_i), \text{ for } y_i > 0 \qquad (3.6b)$$

where $d = \sum_{i=1}^{n} d_i$ which is number of failures.

Prior distribution

$$p(\theta) \propto \theta^{\alpha-1} \exp(-\beta\theta), \quad \theta > 0$$

or

$$\theta \sim Gamma(\alpha, \beta) \qquad (3.7)$$

This is a *conjugate* prior for the exponential model, where $\alpha > 0, \beta > 0$. Comparing (3.7) to (3.6a), this prior density is equivalent to $(\alpha - 1)$ prior exponential observations with total time β.

Note that the equivalent prior distribution for the mean of event time (or survival time) is inverse-gamma:

$$\mu = E(y \mid \theta) = \frac{1}{\theta} \sim Inv - gamma(\alpha, \beta)$$

where α represents number of events and β represents total follow-up time for all patients. Mean of survival time (FU time) $E(\mu) = \frac{\beta}{\alpha - 1}$, and Median = Mean * log(2).

In survival analysis, statistical inference on median of survival time is usually of interest. Here hazard rate is our parameter of interest.

Posterior distribution

For right-censored data with likelihood in (3.6b), the posterior distribution is

$$p(\theta \mid y) \propto \theta^{\alpha+d-1} \exp\{-(\beta + \sum_{i=1}^{n} y_i)\theta\}$$

or

$$\theta \mid y \sim Gamma(\alpha + d, \beta + \sum_{i=1}^{n} y_i) \tag{3.8}$$

which has mean and variance,

$$E(\theta \mid y) = \frac{\alpha + d}{\beta + \sum_{i=1}^{n} y_i}$$

$$\text{var}(\theta \mid y) = \frac{\alpha + d}{(\beta + \sum_{i=1}^{n} y_i)^2}$$

We can think of the first parameter of a gamma distribution (3.8) as the number of events and the second parameter as the total follow-up time of all patients including those with censored events.

The *Beta* distribution (3.4) only involves modeling event rate and does not consider the time to event feature. The *Gamma* distribution (3.8) actually models hazard rate which considers the event time (or survival time) and includes censored data.

In most applications of survival analysis, we either model survival function $S(t)$ (see

Dirichlet process method in Section 3.3.3) or cumulative hazard function $H(t)$ (see Beta process method in Section 3.3.2). Note that there is one-to-one relationship between the two: $S(t) = \exp(-H(t))$. We present these nonparametric methods in the next Section.

3.3. Bayesian Nonparametric Method

Parametric method relies on restrictive parametric specifications and may limits the scope and type of inference. Nonparametric method has been widely used for survival data analysis, providing flexible modeling. Nonparametric Bayesian approach also allows the incorporation of prior information into the traditional models.

Nonparametric Bayesian data analysis has become part of the survival analysis following the introductory work of Ferguson (1973) on Dirichlet process (DP) and the denomination of Beta process by Hjort (1990). Kalbfleisch (1978) models the cumulative hazard function as a gamma process and provides the estimation of regression parameters in semi-parametric modeling for survival data. Rolin (1997) provides a review of the posterior distribution of hazard function (and survival function) in models with censored observations and without explanatory variables. Muller and Quintana (2004) provide a recent review of nonparametric Bayesian inference including density estimation using Dirichlet process, regression, survival analysis and hierarchical models. Review of nonparametric Bayesian inference in survival analysis including proportional hazards and multivariate survival data can also be found in Sinha and Dey (1997, 1998) and Dey et al. (1998). As extension, nonparametric Bayesian mixture models have been developed since the pioneer work of Antoniak (1974) on mixtures of Dirichlet processes (MDP). In terms of simulation and recent computational development, Gelfand and Smith (1990), Kuo and Smith (1992), and Arjas and Gasbarra (1994) use MCMC simulation for right censored data in nonparametric Bayesian inference.

In Chapter 2, we described protocol pre-specified Bayesian life-table method which incorporates the Bayesian paradigm into classic life-table. In this section, we first briefly review the classic nonparametric approaches to right censored survival data – life-table method and Kaplan-Meier method. The Bayesian approaches are generally extensions to the traditional nonparametric methods. We then present two other Bayesian nonparametric methods, both of which belong to a class of *random probability measures* (RPM): Beta process and Dirichlet process. As a more flexible Monte Carlo simulation method, Gibbs sampler will be reviewed later. We will also discuss the possible relationships and advantages/disadvantages among these methods.

Traditional Life-Table Estimator

Based on numbers of patients alive or censored at each time interval, several statistics can be computed:

Effective Sample Size (N_j): this is the number of cases that entered the respective interval alive, minus half of the number of cases lost/withdrawal or censored in the respective interval.

Conditional Probability of Failure: this probability is computed as the ratio of the number of cases failing in the respective interval, divided by the effective sample size in the interval. Denote as $\hat{p}_j = y_j / N_j$ for j'th interval.

Estimate of Cumulative Probability of Surviving (Survival Function, $\hat{S}_j$): this is the cumulative proportion of cases surviving up to the respective interval. Since the

probabilities of survival are assumed to be independent across the intervals, this probability is computed by multiplying out the probabilities of survival across all previous intervals. The resulting function is also called the *survivorship* or *survival function*. Denote the estimate of survival function at the beginning of the *j*'th interval as

$$\hat{S}_j = \prod_{i=1}^{j-1}(1-\hat{p}_i) \tag{3.9}$$

Estimate of Probability Density ($\hat{f}_j$): this is the estimated probability of failure in the respective interval, computed per unit of time, that is:

$$\hat{f}_j = (\hat{S}_j - \hat{S}_{j+1}) \Big/ L_j$$

where L_j is the width of the respective interval.

Estimate of Hazard Rate ($\hat{h}_j$): the hazard rate (the term was first used by Barlow, 1963) is defined as the conditional probability per unit of time that a case that has survived to the beginning of the respective interval will fail in that interval.

$$\hat{h}_j = \frac{\hat{f}_j}{\hat{S}_j} = \frac{(\hat{S}_j - \hat{S}_{j+1})/L_j}{\hat{S}_j} = (1-\frac{\hat{S}_{j+1}}{\hat{S}_j})/L_j = \left[1-\frac{\prod_{i=1}^{j}(1-\hat{p}_i)}{\prod_{i=1}^{j-1}(1-\hat{p}_i)}\right] \Big/ L_j = \frac{\hat{p}_j}{L_j}$$

Kaplan-Meier Product-Limit Estimator

Rather than classifying the observed survival times into a life table, we can estimate the survival function directly from the continuous failure times. To allow for possible ties in the data, suppose that the event occurs at *K* distinct times $t_1 < t_2 < ... < t_K$, and that at

time t_i there are y_i events observed. Let N_i be the number of individuals at risk at time t_i:

$$\hat{S}(t) = \prod_{t_i \le t} (1 - \frac{y_i}{N_i}) \ , \quad if \ t \ge t_1 \tag{3.10}$$

and $\hat{S}(t) = 1$, if $0 = t_0 \le t < t_1$. In (3.10), $\hat{S}(t)$ is the estimated survival function at time $t \in [t_i, t_{i+1})$ or actually at the event time t_i until next event occurs. This estimate of the survival function is also called the *product-limit estimator*, and was first proposed by Kaplan and Meier (1958). The product-limit estimator is a step function with jumps at the observed event time.

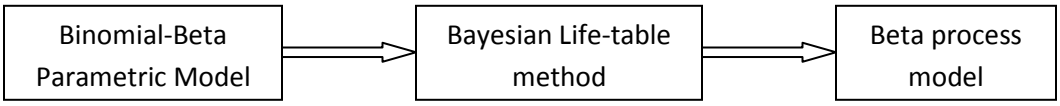
The advantage of the Kaplan-Meier product-limit method over the life table method for analyzing survival and failure time data is that the resulting estimates do not depend on the grouping of the data (into a certain number of time intervals). Actually, the product-limit method and the life table method are identical if the intervals of the life table contain at most one observation.

Dirichlet process and Beta process deal with any set of intervals and assume a distribution function to each of these intervals. Dirichlet process is used to estimate survival function. Beta process is to estimate cumulative hazard function. Both processes incorporate conjugate prior information.

3.3.1. Bayesian Life-table Method

We presented protocol-specified method in Chapter 2 and reviewed the traditional life-table method. We call the protocol-specified approach as a Bayesian life-table method since it's basically life-table but incorporates a beta prior at each interval. It is classified as one of the nonparametric methods in this Chapter. We have a brief summary of the method here, and then illustrate its relationships with Binomial-Beta parametric model (presented in Section 3.2.1) and Beta process model.

Figure 4: Relationships Among Beta Models



To simplify the notation on subscript, only one administration-schedule group is used here.

j: time interval, $1 \le j \le J$,

n_j: effective sample size at risk during interval j,

y_j: number of failures during interval j,

p_j: conditional probability of failure during interval j (discrete-time hazard),

Data:

$$y_j \mid p_j \sim Bin(n_j, p_j)$$

$$E(y_j) = n_j * p_j$$

Prior:

$$p_j \sim Beta(\alpha, \beta)$$

$$E(p_j) = \frac{\alpha}{\alpha + \beta}$$

Posterior:

$$p_j \mid y_j \sim Beta(\alpha + y_j, \beta + n_j - y_j)$$

$$E(p_j \mid y_j) = \frac{\alpha + y_j}{\alpha + \beta + n_j}$$

Traditional life-table method uses $E(\hat{p}_j) = p_j$, and cumulative probability of failure to interval J is:

$$\theta_J = 1 - \prod_{j=1}^{J} (1 - p_j)$$

or

$$\hat{\theta}_J = 1 - \prod_{j=1}^{J} (1 - \hat{p}_j)$$

where $\hat{p}_j = \frac{y_j}{n_j}$

Bayesian life-table method incorporates prior information so p_j is replaced by posterior estimator $\tilde{p}_j$ given data y_j:

$$\tilde{\theta}_J = 1 - \prod_{j=1}^{J} (1 - \tilde{p}_j \mid y_j)$$

where $\tilde{p}_j \mid y_j = \dfrac{\alpha + y_j}{\alpha + \beta + n_j}$.

The Binomial-Beta parametric model doesn't consider individual event time and combine all intervals (treat the whole study as one large interval: $0 \sim \infty$) to estimate the overall probability of failure:

$$\hat{\theta}_J = \hat{p}_J = \frac{\sum_{j=1}^{J} y_j}{\sum_{j=1}^{J} n_j} = \frac{Y}{N}$$

where $Y \mid p_J \sim Bin(Y \mid N, p_J)$.

If the same prior $p_j \sim Beta(\alpha, \beta)$ is used for each interval and probability of progression (failure event) is independent among intervals, we can prove that posterior of $p_J \mid y_1, y_2, ..., y_J$ at the last interval:

$$p_J \mid y_1, y_2, ..., y_J \sim Beta(\alpha + \textstyle\sum_{j=1}^{J} y_j, \beta + \sum_{j=1}^{J} n_j - \sum_{j=1}^{J} y_j)$$

It's the same as if we get posterior from *j*th interval, and use it as prior of the (*j*+1)th interval, where $j = 1,2,..., J$. At the last interval, one can also get the above posterior.

Parameter θ_J is more appropriate to use life-table method to estimate. Parameter p_J has the property of "additive" and it doesn't consider the individual event time. The estimator of the above p_J is an estimate of the overall probability of failure event. As

discussed earlier, if one combines all intervals or treat the whole study as one interval, then θ_J and p_J are the same.

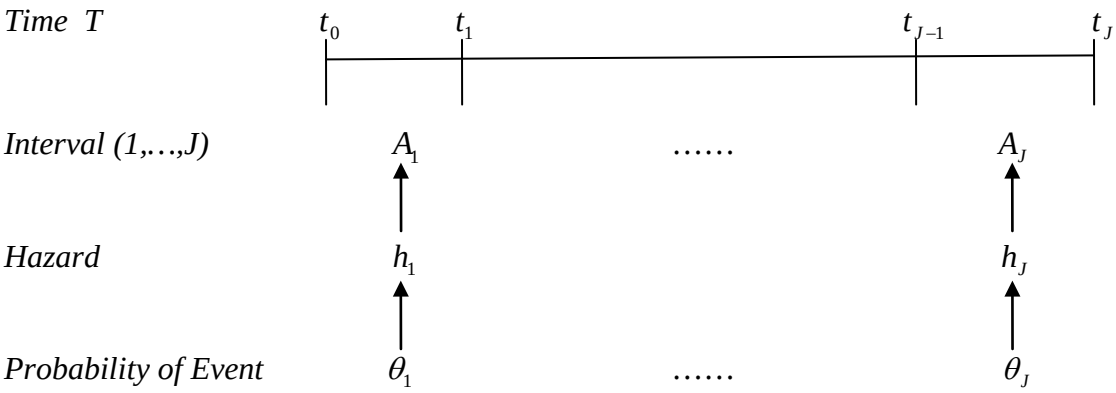
Like Bayesian life-table method, Beta process model assumes beta priors for ΔH_j (increment of cumulative hazard) or h_j (hazard at time t_j or at the *j*th interval). We present Beta process model in next subsection.

3.3.2. Beta Process Model

Another approach to modeling survival data is to provide a prior distribution for the cumulative hazard function. Hjort (1990) introduces a Beta process prior on cumulative hazard function $H(t)$. Hjort assumes independent, beta-distributed priors for h_j (hazard at time t_j) or for ΔH_j (increment of cumulative hazard, see below for discrete case). This generates a beta process with independent increments for the cumulative hazard function $H(t)$. Full Bayesian inference for a model with a Beta process prior for the $H(t)$ using Gibbs sampling can be found in Damien et al. (1996).

Survival data in practice is commonly grouped with some grid of intervals, it is more convenient and often sufficient to use a discretized version of the beta process (Hjort, 1990; Sinha, 1997; Ibrahim et al., 2001) for survival data subject to right censoring. We focus on discrete-time case of the beta process below. Note that one consequence of the discreteness is that the relationship between cumulative hazard and survival with $H(t) = -\ln[S(t)]$ for continuous-time case does not hold anymore.

Figure 5: Chart of Time, Interval and Parameters of Interest



Hjort (1990) uses the term *beta processes* to describe processes to produce cumulative hazard rates whose increments are independent and approximately beta distributed. The principal is that if H is a beta process priori, then it still is a posteriori.

Data distribution

Let $A_j = [t_{j-1}, t_j)$, $j = 1,..., J$ be a series of nonoverlapping intervals with discrete random variable T that takes the values $0 = t_0 < t_1 < t_2 < ... < t_J$ (Figure 5). Let hazard rate and cumulative hazard rate be

$$h_j = \Pr\{T = t_j \mid T \geq t_j\} = \frac{\Pr(T = t_j)}{\Pr(T \geq t_j)} = \frac{f_j}{S_j},$$

$$H_j = \sum_{i=0}^{j} h_i$$

Note that S and f can be recovered from knowledge of hazard rate:

$$S_j = \prod_{i=0}^{j} (1 - h_i),$$

$$f_j = \left[\prod_{i=0}^{j-1} (1 - h_i) \right] h_j$$

Note that notation $f_j = \Pr(T = t_j)$ with t_j representing a discrete-time point is little different from that in Chapter 2 where $\theta_j = \Pr(Y = j)$ with *j* denoting tiny interval *j*.

We assume data is right censored with observations $(T_i, d_i), i = 1,...,n$, where T_i is failure time or censoring time and d_i is failure indicator for individual *i*. Consider the failure counting process *Y* and the number-at-risk process *N*, given by

$$Y_j = \sum_{i=1}^{n} I\{t_{j-1} < T_i \le t_j, d_i = 1\},$$

$$N_j = \sum_{i=1}^{n} I\{T_i \ge t_j\}$$

where Y_j and N_j denote number of failures at interval *j* and number at risk at time t_j, respectively.

Hjort (1990) proves that the likelihood of what is observed can be written

$$L(data) = \prod_{i=0}^{j} \left[(1 - h_i)^{N_i - Y_i} h_i^{Y_i} \right] \tag{3.11}$$

So the nonparametric maximum likelihood (ML) estimator of $h(.)$ is given by

$$\hat{h}_j = Y_j / N_j$$

and

$$\hat{H}_j = \sum_{i=0}^{j} \frac{Y_i}{N_i} \quad \text{and} \quad \hat{S}_j = \prod_{i=0}^{j} (1 - \frac{Y_i}{N_i})$$

These are the proper discrete-time analogues of the Nelson-Aalen estimator and the Kaplan-Meier estimator.

Beta process prior

In order to construct a class of nonparametric Bayesian estimators for *H* (and for *h* and *S*), Hjort assumes that a beta process prior for *H*,

$$H \sim BP(N_0, H_0) \tag{3.12}$$

has independent summands

$$h_j \sim Beta\{N_{0j} h_{0j}, N_{0j}(1 - h_{0j})\} \tag{3.13}$$

where $E(h_j) = h_{0j}$ is prior guess whereas $Var(h_j) = h_{0j}(1 - h_{0j}) / (N_{0j} + 1)$ is prior uncertainty, and N_{0j} is a measure of confidence around the prior guess h_{0j} (or prior sample size).

Beta process posterior distribution

Given censored data set (T_i, d_i) and beta process prior in (3.12) which constitutes a natural class of conjugate prior distributions, the posterior is (Hjort, 1990)

$$H_j \mid data \sim BP\left\{ N_0 + N, \sum_{i=0}^{j} \frac{N_{0i} h_{0i} + Y_i}{N_{0i} + N_i} \right\} \tag{3.14}$$

where $h_{0i} = dH_{0i}$.

Furthermore, the nonparametric Bayesian estimator of *H* is given by

$$\hat{H}_j = E\{H_j \mid data\} = \sum_{i=0}^{j} \frac{N_{0i} h_{0i} + Y_i}{N_{0i} + N_i} \tag{3.15}$$

Note that $N_0 \to 0$, then the above is Nelson-Aalen estimator; $N_0 \to \infty$, then the $\hat{H}$ simply becomes the prior guess H_0. The combined sample size for the posterior is $N_0 + N$, N_0 of them having hazard $h_0 = dH_0$ and N of them having hazard Y/N.

The conditional variance of H_j is useful when constructing Bayesian confidence band for *H*.

Moreover, survival estimate is given by

$$\hat{S}_j = \prod_{i=0}^{j} \left[1 - \frac{N_{0i} dH_{0i} + Y_i}{N_{0i} + N_i} \right]$$

When $N_0 \to 0$, the above is Kaplan-Meier estimator; $N_0 \to \infty$, then the $\hat{S}$ simply becomes the prior guess S_0.

Note that another perfect way to estimate *H* is to simulate realizations of *H* from the posterior distribution as described by Ibrahim et al. (2001). Ibrahim et al. (2001) present the posterior for h_j in discrete-time case as

$$h_j \mid data \sim Beta\{N_{0j} h_{0j} + Y_j, N_{0j}(1 - h_{0j}) + N_j - Y_j\}, \; j = 1, \ldots, J \tag{3.16}$$

In Chapter 5, we will use simulation to draw posterior samples from (3.16) and then transform the simulation results to the estimates of survival or probability of event for the CLL clinical trial data.

Bayesian estimator of the survival function

At any time t where $t_j \le t < t_{j+1}$, the Bayesian estimator of the survival function under squared-error loss is given by (Klein and Moeschberger, 2007):

$$\tilde{S}_{BP}(t) = \exp\left\{-\sum_{i=1}^{j}\frac{N_0[H_0(t_i)-H_0(t_{i-1})]}{N_0+N_i} - \frac{N_0[H_0(t)-H_0(t_j)]}{N_0+N_{j+1}}\right\} * \prod_{i:t_i\le t}\left[1-\frac{N_0 h_0(t_i)+Y_i}{N_0+N_i}\right]^{\delta_i}$$
$$\text{for } t_j \le t < t_{j+1},\ j = 0,...,J \qquad (3.17)$$

For large N (data sample size) or relatively $N_0 \to 0$, the Bayesian estimator (3.17) reduces to Kaplan-Meier estimator. For small N or when $N_0 \to \infty$, the Bayesian estimator is close to the prior guess at $S(t)$.

Notes about Beta process model and its Bayesian estimator of survival

From prior distribution in (3.13): $h_j \sim Beta\{N_{0j}h_{0j}, N_{0j}(1-h_{0j})\}$, $j = 1,...,J$ and posterior distribution in (3.16), we can see the similarity and difference between Beta process model and Bayesian life-table method. Beta distribution prior can be considered as a special case of beta process prior. The life-table method arbitrarily classifies the time intervals while beta process model intervals are based on actual event or censor time. In the protocol pre-defined Bayesian life-table method, clinical priors *Beta*(0.025,

0.975) and *Beta*(0.01, 0.99) are used for hazards of progression and toxicity at each time interval. These prior distributions assume that prior sample size $N_{0j} = 1$ and prior guess of hazard $h_{0j} = 0.025$ or $h_{0j} = 0.01$, where $j = 1,2,..., J$. The prior *Beta*(0.01, 0.99) corresponds to the Beta process prior parameters of $H_0(t) = 0.01t$ and $N_0(t) = 1$. On the other hand, as we already shown, Binomial-Beta parametric model is a special case of Bayesian life-table method when we treat the whole study as a single interval in which case event time is not considered into modeling.

Results from Bayesian estimator in (3.17) may not exactly match results calculated by posterior samples from Beta process model (3.16) due to the following reasons: (1) Beta process model assumes small time intervals. In each such tiny interval filling in between observed life times, the posterior cumulative hazard has, approximately, a beta distribution (Hjort, 1990; Damien et al., 1996). In practice, intervals may not be short enough in clinical trials. (2) For discrete lifetime data, after drawing posterior samples, we define cumulative hazard function by $H(t) = \sum_{t_j \le t} h(t_j)$ but the relationship of $S(t) = \exp\{-H(t)\}$ for this definition no longer holds true (Klein and Moeschberger, 2007). We may need to re-define the cumulative hazard for discrete lifetime as $H(t) = -\sum_{t_j \le t} \ln[1 - h(t_j)]$ (Cox and Oakes, 1984) so that the relationship of $S(t) = \exp\{-H(t)\}$ for continuous-time is preserved for discrete-time data. Alternatively we can directly use $S(t) = \prod_{t_j \le t} [1 - h(t_j)]$ as in life-table method for survival estimate.

3.3.3. Dirichlet Process Model

The Dirichlet process is perhaps the most important and popular prior process in modern day nonparametric Bayesian inference (Ibrahim et al., 2001). Ferguson (1973, 1974) introduces the Dirichlet process (DP) as one of the *random process measures* (RPM). Susarla and Van Ryzin (1976) and Ferguson and Phadia (1979) discuss inference with a DP prior in the context of survival data. For a review of related approaches applying the DP see Ferguson et al. (1992).

DP model is to estimate survival function directly while BP is to model hazard rate. With conjugate prior and right censored survival data, posterior distribution of the survival function $S(t)$ follows Dirichlet process with parameter function α.

Dirichlet distribution

Dirichlet distribution is conjugate prior distribution for the parameters of the multinomial distribution. The probability density of the Dirichlet distribution for variables $\theta = (\theta_1,...,\theta_J)$ with parameters $\alpha = (\alpha_1,...,\alpha_J)$ is defined by

$$f(\theta_1,...,\theta_{J-1}) = \frac{\Gamma(N_0)}{\prod_{j=1}^{J}\Gamma(\alpha_j)}\left(\prod_{j=1}^{J}\theta_j^{\alpha_j-1}\right)$$

where $\theta_j \geq 0, j = 1,...,J$ with $\sum_{j=1}^{J}\theta_j = 1$ and $N_0 = \sum_{j=1}^{J}\alpha_j$ representing prior sample size.

Mean and variance of θ_j are:

$$E(\theta_j) = \frac{\alpha_j}{N_0}$$

$$Var(\theta_j) = \frac{E(\theta_j)[1 - E(\theta_j)]}{N_0 + 1}$$

Dirichlet process prior

On a positive real line X (Figure 5) or a sample space, a stochastic process is said to be a Dirichlet process (prior) if, for any set of disjoint intervals $A_1, ..., A_J$ or any partition of the sample space, the joint distribution of the (prior) probabilities $\Pr[X \in A_1] = \theta_1$, …, $\Pr[X \in A_J] = \theta_J$ has a $(J-1)$-dimentional Dirichlet distribution with parameters $[\alpha(A_1), ..., \alpha(A_J)]$. This property must hold for any set of intervals and any J (Klein and Moeschberger, 2007). Hence, different partition of line X or sample space leads to Dirichlet distribution with different parameters.

We assume prior distribution of survival function $S(t)$ follows

$$S(t) \sim DP(N_0, S_0(t))$$

or

$$S(t) \sim DP(\alpha)$$

where the parameter function α is determined by N_0 and $S_0(t)$ and usually takes the form of $\alpha([t, \infty)) = N_0 S_0(t)$. Note that $S_0(t)$ is prior expectation of survival function and N_0 is a measure of how much weight to put on the prior (or sample size of prior). So the meaning of the parameter function $\alpha([t, \infty)) = N_0 S_0(t)$ can be interpreted as the number of patients survived at or beyond time *t*, and

$$\alpha([0,\infty)) = \sum_{j=1}^{J} \alpha_j = N_0$$

where α_j represents number of failures at *j*th interval A_j.

Mean and variance of the prior of $S(t)$ which follows Dirichlet process are:

$$E[S(t)] = \frac{\alpha([t,\infty))}{\alpha[0,\infty)} = \frac{N_0 S_0(t)}{N_0} = S_0(t)$$

$$Var[S(t)] = \frac{S_0(t)[1-S_0(t)]}{N_0+1}$$

Assume a Dirichlet process prior with parameters $S_0(t) = \exp(-0.1t)$ and $N_0 = 5$ (or $N_0 = 1$), we will simulate or draw samples from this form of Dirichlet process prior and show how prior sample paths look like and how they change with the increase (or decrease) with prior sample size in Chapter 5.

Data distribution

The right-censored survival data consists of *J* distinct times (or *J* time intervals), denoting as $0 = t_0 < t_1 < ... < t_J = \infty$ (Figure 5). At time t_j, $j = 0,...,J$, let N_j be the number of patients at risk, y_j be the number of events (d_j event indicator: $d_j = 1$ if $y_j > 0$ and $d_j = 0$ if $y_j = 0$) and λ_j the number of right-censored observations.

Posterior distribution

Combining the survival data distribution with the Dirichlet process prior, it can be shown (Ibrahim et at., 2001; Klein and Moeschberger, 2007) that the posterior distribution of survival function $S(t)$ is also Dirichlet.

Without considering censoring, the parameter R of the posterior distribution is the original parameter α from the prior distribution plus a point mass of one at points where events occur, i.e., for any interval (*a*, *b*) and the *k*th subject's study time $T_k, k = 1,..., N$,

$$R((a,b)) = \alpha((a,b)) + \sum_{k=1}^{N} I[d_k = 1, a < T_k < b] = \alpha((a,b)) + y((a,b)) \qquad (3.18)$$

where $I[]$ is the indicator function and $y((a,b)) = \sum_{k=1}^{N} I[d_k = 1, a < T_k < b]$ is number of failures observed in interval (*a*, *b*). Original proof can be found in Ferguson (1973).

Considering right-censored observations, posterior distribution of the survival function is a mixture Dirichlet process (MDP) in which case MCMC scheme is to be used for posterior inference. We will refer to Gibbs sampling in Kuo and Smith (1992) in subsection 3.3.4.

Bayesian estimator of the survival function

Susarla and Ryzin (1976) present a Bayesian nonparametric estimator – the posterior expected values of the survival distribution, assuming a Dirichlet process prior (Ferguson, 1973). The Bayesian estimator of the survival function under squared-error loss (Susarla and Ryzin, 1976, 1978a, 1978b; Klein and Moeschberger, 2007) is:

$$\tilde{S}_{DP}(t) = \frac{\alpha(t,\infty) + N_{j+1}}{\alpha(0,\infty) + N} \prod_{k=0}^{j} \frac{\alpha(t_k,\infty) + N_{k+1} + \lambda_k}{\alpha(t_k,\infty) + N_{k+1}}$$

$$= \frac{N_0 S_0(t) + N_{j+1}}{N_0 + N} \prod_{k=0}^{j} \frac{N_0 S_0(t_k) + N_{k+1} + \lambda_k}{N_0 S_0(t_k) + N_{k+1}} \qquad (3.19)$$

$$\text{for } \; t_j \le t < t_{j+1},\ j = 0,\ldots,J$$

Susarla and Ryzin (1976, 1978b) use notation $N^+(t)$ instead of N_{j+1} to denote the number of patients at risk at time $> t$ in (3.19). The asymptotic properties of this Bayesian estimator are discussed in Susarla and Ryzin (1978a, 1978b). For large *N* (data sample size) or very small N_0 (prior sample size), Susarla and Ryzin (1976) show that the Bayesian estimator (3.19) reduces to Kaplan-Meier estimator. For small *N*, the prior information dominates and the Bayesian posterior estimator is close to the prior guess $S_0(t)$. So the Bayesian estimator can be considered as weighted average of the Kaplan-Meier estimator and the prior guess.

Note that the above Bayesian estimator, when t is close to the end of the study, the number at risk N_{k+1} can be smaller than λ_k. In extreme case, if $N_{k+1} = 0$ at the end, the prior information dominates the estimate and may result in unreasonable large value of the estimate. In general, a better way is to utilize MCMC Gibbs sampling method, which is discussed in next subsection.

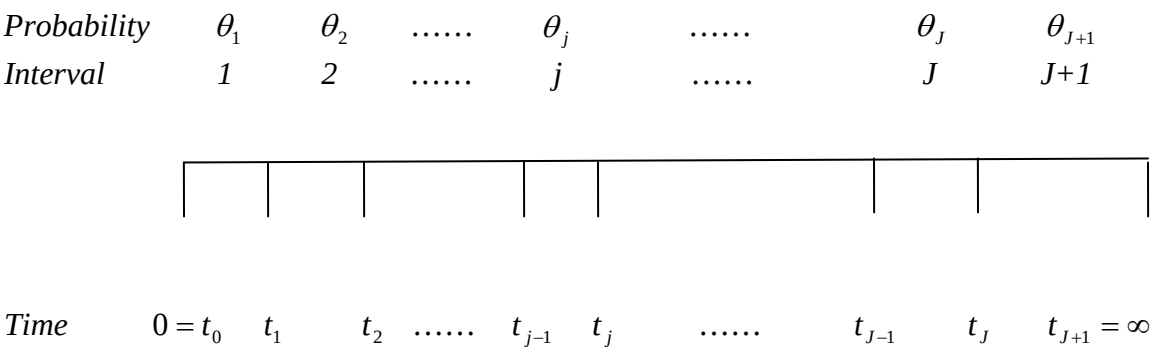
3.3.4. MCMC Bayesian Method or Gibbs Sampler

Gibbs sampling or a *Gibbs sampler* is a Markov chain Monte Carlo (MCMC) algorithm for obtaining a sequence of random samples from the joint probability distribution of two or more random variables. This sequence can be used to approximate the joint distribution; to approximate the marginal distribution of one of the variables, or some subset of the variables (for example, the unknown parameters or latent variables). Gibbs sampling is applicable when the joint distribution is not known explicitly or is difficult to sample from directly. It can be shown (for example, Gelman et al., 2004) that the sequence of samples constitutes a Markov chain, and the stationary distribution of that Markov chain is just the sought-after joint distribution.

Kuo and Smith (1992) discuss Dirichlet process priors with interval censored survival data and demonstrate the posterior distribution is to be a mixture of Dirichlet processes (Antoniak, 1974). Kuo and Smith (1992) use Gibbs sampler to generate samples from the posterior distribution. Gibbs sampler approach is more flexible than previous two nonparametric process models. In this subsection we utilize this MCMC method for right-censored data, even though it's a general approach for sampling posterior distribution and can be easily extended to interval-censored data. We start with data distribution and prior distribution which are similar to those presented in the Dirichlet process method. For posterior distribution, this method imputes failure time of censored observations conditional on observed failures and re-distributes these censored observations to later individual time intervals. Because the method is nonparametric, it can be easily used in situations where hazards cross or are suspected to cross.

Kuo and Smith (1992) method is also introduced in Ibrahim et al. (2001), Rosner (2005), and Klein and Moeschberger (2007). One can use the random samples from the posterior distribution of the survival function to calculate data summaries and make statistical inference. In this subsection, we present the Gibbs sampling algorithm for survival distribution with Dirichlet process prior and right-censored data introduced in subsection 3.3.3. In next Chapter, we will use Gibbs sampler in our B-cell CLL clinical trial data and also compare the numerical results from this method to those obtained directly from Bayesian estimator (3.19).

Figure 6: Chart of Time, Interval and Parameters of Interest



Data distribution

Let $0 < t_1 < ... < t_J$ be J time points.

y_j : number of events in the time interval $(t_{j-1}, t_j]$,

λ_j : number of right-censored observations at time t_j,

$S_j = S(t_j)$: survival function at time t_j,

So the likelihood function is $L(y, \lambda \mid S) \propto \prod_{j=1}^{J+1} (S_{j-1} - S_j)^{y_j} S_j^{\lambda_j}$.

Let $\theta_j = S_{j-1} - S_j$ be parameter of interest which is probability of event for interval *j* for $j = 1,..., J$ and $\theta_{J+1} = S_J$, where interval *j* is for time interval $(t_{j-1}, t_j]$ and

$S_j = S(t_j) = P(T > t_j) = 1 - \sum_{i=1}^{j} \theta_i$. Note that $\sum_{j=1}^{J+1} \theta_j = 1$.

Prior distribution

We specify a Dirichlet process prior for the $\theta^{'}s$ and thus joint prior distribution of $\theta^{'}s$ (probability of event) is a Dirichlet distribution with density function

$$f(\theta_1,..., \theta_J) \propto \prod_{j=1}^{J+1} (\theta_j)^{\alpha_j - 1}$$

or

$$\theta \sim Dirichlet(\alpha_1, \alpha_2,..., \alpha_{J+1})$$

where parameters $\alpha_j = N_0[S_0(t_{j-1}) - S_0(t_j)]$ for *j=1,…,J+1* with $S_0(t)$ indicating prior survival function, where $S_0(t_{J+1}) = 0$ and N_0 is prior sample size or weight of prior.

To generate a sample of θ from $Dirichlet(\alpha_1, \alpha_2,..., \alpha_{J+1})$ distribution, one can use the definition by generating independent gamma random variables. In **R**'s MCMC package, *rdirichlet(alpha)* function can be used directly to generate a random vector sample of θ following Dirichlet distribution where *alpha* is vector of shape parameters.

We denote the first realization of **θ** from the Dirichlet prior as θ^0 and will use it as the starting values of the following Gibbs sampling.

Posterior distribution approximation via MCMC

Gibbs sampling approach to Bayesian estimation approximates the posterior distribution via Monte Carlo simulation. The basic idea here is that we treat censored observations as missing data. First, we impute event times for each censored observation, then use the new event times to update the parameter of Dirichlet distribution and so are the sample draws of θ's. This process repeats over and over, and the procedure converges to a realization of θ drawn from the posterior distribution of θ given the data (Gelfand and Smith, 1990).

Let $\lambda_j > 0$ be the number of right-censored observations at time t_j or the beginning of the $(j+1)$th interval $(t_j, t_{j+1}]$. These censored observations will eventually fall in intervals $(t_h, t_{h+1}]$ with $h \geq j$. Denote the number of failures λ_j that might fall in these intervals as $Z_{j+1,j},...,Z_{J+1,j}$, so $\lambda_j = \sum_{h=j+1}^{J+1} Z_{h,j}$.

The posterior full conditionals for $\theta = (\theta_1, \theta_2, ..., \theta_{J+1})$ given the *Z*'s and the data is an updated Dirichlet distribution, and the posterior full conditional for the *Z*'s given $\theta = (\theta_1, \theta_2, ..., \theta_{J+1})$ and the data is a multinomial distribution. Suppose at the *i*th iteration of the Gibbs sampler, we have the realization $\theta^i = (\theta_1^i, \theta_2^i, ..., \theta_{J+1}^i)$ with

$\sum_{j=1}^{J+1}\theta_j^i = 1$. Practically the sampler starts from θ^0 as described above from prior distribution.

An iteration of the Gibbs sampler consists of imputing failures for censored observations from multinomial distribution, combining the prior with the observed and imputed failures to update the parameters of the Dirichlet distribution, and generating a random sample from the Dirichlet distribution (Rosner, 2005). The Gibbs sampling algorithm follows procedures below in detail (Kuo and Smith, 1992; Ibrahim et al., 2001; Klein and Moeschberger, 2007):

(1) Sample $Z_{j+1,j}^{i+1},...,Z_{J+1,j}^{i+1}$ from a multinomial distribution with sample size λ_j and parameters $\rho_{j+1}^i,...,\rho_{J+1}^i$:

$$(Z_{j+1,j}^{i+1},...,Z_{J+1,j}^{i+1}) \sim Multinomial(\lambda_j;\rho_{j+1}^i,...,\rho_{J+1}^i) \tag{3.20}$$

where $$\rho_h^i = \frac{\theta_h^i}{\sum_{h=j+1}^{J+1}\theta_h^i}, h = j+1,...,J+1$$

(2) Having sampled the random variables *Z*'s, we revise the number of failures at time t_h by $y_h + \sum_{j=1}^{J} Z_{h,j}^{i+1}$ and update the Dirichlet distribution parameters:

$$R_h^{i+1} = \alpha_h + y_h + \sum_{j=1}^{J} Z_{h,j}^{i+1}, h = 1,2,...,J+1 \tag{3.21}$$

Then sampling $\theta^{i+1} = (\theta_1^{i+1},\theta_2^{i+1},...,\theta_{J+1}^{i+1})$ from Dirichlet distribution with parameters $(R_1^{i+1},R_2^{i+1},...,R_{J+1}^{i+1})$:

$$\theta^{i+1} \sim Dirichlet(R_1^{i+1},R_2^{i+1},...,R_{J+1}^{i+1}) \tag{3.22}$$

(3) By running *S* parallel chains, after the *i*th iteration assuming we start with θ^0, we have $\theta_{1s}^i, \theta_{2s}^i, ..., \theta_{J+1,s}^i$ and $R_{1s}^i, R_{2s}^i, ..., R_{J+1,s}^i$ for $s = 1, ..., S$ where *S* and *i* have been chosen to achieve convergence to smooth estimates. Typically *i* is relatively small, of the order of 10 or 20, and *S* is of the order of 1,000 – 10,000. The posterior mean estimate of θ_h is, then given by:

$$\tilde{\theta}_h = S^{-1} \sum_{s=1}^{S} \frac{R_{hs}^i}{\sum_{j=1}^{J+1} R_{js}^i} \tag{3.23}$$

Note the clinical meaning of the parameter function *R* of the Dirichlet process in (3.22) and (3.23): R_h represents the number of events at interval *h*; $\sum_{h=1}^{J+1} R_h$ is total sample size; and θ_h is the probability of event at interval *h*.

The posterior estimator of θ_h from the Gibbs sampler (3.23) is based on the fact that posterior distribution of θ_h for $h = 1,2,..., J+1$ can be approximated by a beta random variable with parameters R_h and $\sum_{k \neq h} R_k$ (Ibrahim et al., 2001; Klein and Moeschberger, 2007):

$$\theta_h \sim Beta(R_h, \sum_{k \neq h} R_k)$$

An alternative approach to estimate θ_h from the posterior distribution is to use the empirical distribution function of the simulated values of θ: $\theta_{hs}^i, s = 1,2,..., S$ (Klein and Moeschberger, 2007). One can make inference about θ based on these *S* samples. The

precisions of these two approaches are different. The standard error of the means from (3.23) is much smaller than that from this alternative approach.

To estimate the survival function at t_j, we accumulate the θ_i for $i > j$, so

$$\hat{S}_j = \sum_{i>j} \tilde{\theta}_i = 1 - \sum_{i \le j} \tilde{\theta}_i \tag{3.24}$$

The overall probability of event up to time t_j is to accumulate the θ_i for $i \le j$.

With the parameters of the posterior Dirichlet distribution in (3.21) at the *i*th iteration, one can also estimate the posterior distribution of the survival probability $S(t)$ at any time t in the following way (Rosner, 2005):

$$S(t) \mid data \sim Beta(\sum_{h:t_h \ge t} R_h^i, \sum_{h:t_h < t} R_h^i) \tag{3.25}$$

By drawing random samples from the above Beta distribution at time t, say, 2 or 3 years, one can make inference on posterior survival distribution.

4. SIMULATIONS

Chapter 3 describes Bayesian parametric and non-parametric statistical methods to evaluate accumulative event rate for survival endpoint. In this Chapter, simulation results from these methods are presented. The purpose of simulation is to ensure each of the methods provides valid and realistic results with given scenarios and samples so as to clear the path to apply these methods to the phase II CLL clinical trial data in Chapter 5.

Simulations are conducted using efficacy endpoint which is probability of progression. The same simulation approach can be applied to safety endpoints and is not included here. The operating characteristics of the design of the phase II trial (the design's average behavior - probability of being selected for each arm) under different scenarios are summarized by 1,000 simulation studies (500 studies for Gibbs sampler method) per scenario. The protocol considered 6 scenarios under different levels of probability of progression for each arm (from high and medium to low) and we also consider these similar scenarios in simulation.

Generating Simulation Data

During clinical trials with survival type of endpoint, patient's enrollment is staggered. For the *i*th patient, $i = 1,2,...,n,$ let X_i be the entry time (in calendar time) with density function *g*, let T_i be the failure time with density function *f*, and let C_i be the censoring time with distribution function 1 - *H*. Note that T_i and C_i are measured from X_i and assume that X_i, T_i, and C_i are mutually independent for all patients.

At any calendar time *t* when the data is reviewed, for the *i*th patient, the observation time is

$$Y_i(t) = \min(T_i, C_i, (t - X_i)^+)$$

and censoring status is

$$\Delta_i(t) = I_{\{T_i \le \min(C_i, t - X_i)\}}$$

where $\Delta = 1$ indicating a failure and $\Delta = 0$ being censored.

The accumulative failure rate by time *t* is (Li, Shih and Wang, 2005):

$$\theta(t) = P\{T_i \le \min(C_i, t - X_i)\} = \int_0^t \left[\int_0^{t-y} H(u) f(u) du \right] g(y) dy \qquad (4.1)$$

Enrollment can be modeled as exponential, poisson, uniform, or truncated exponential distribution (Lachin and Foulkes, 1986), etc.

Assume entry time is uniformly distributed, both failure time and censoring time are exponentially distributed, i.e.,

$$g(y) = \frac{1}{b}\ ,\ \ f(u) = \lambda e^{-\lambda u}\ ,\ H(u) = e^{-\eta u}$$

Then

$$\theta(t) = \int_0^b \left[\int_0^{t-y} H(u) f(u) du \right] g(y) dy$$

$$= \frac{\lambda}{(\eta+\lambda)} \left[1 - \left(\frac{1}{(\eta+\lambda)b} \right) \left(e^{-(\eta+\lambda)(t-b)} - e^{-(\eta+\lambda)t} \right) \right] \qquad (4.2)$$

Note that when $t \to \infty$, $\theta(t) \to \frac{\lambda}{\eta+\lambda}$, which is the case when only considering failure time and censoring time.

We can also model the enrollment with the family of truncated exponential distribution (Lachin and Foulkes, 1986; Li, Shih and Wang, 2005) indexed by γ

$$g(y) = \begin{cases} \frac{\gamma e^{-\gamma y}}{1 - e^{-\gamma b}}, 0 \le y \le b; \gamma \ne 0. \\ \frac{1}{b}, 0 \le y \le b; \gamma = 0. \end{cases}$$

The family includes entry rate being uniform ($\gamma = 0$), concave ($\gamma < 0$), or convex ($\gamma > 0$).

In the case $\gamma \ne 0$,

$$\theta(t) = \int_0^b \left[\int_0^{t-y} H(u) f(u) du \right] g(y) dy$$

$$= \frac{\lambda}{(\eta+\lambda)} \left\{ 1 - \left(\frac{\gamma}{(\eta+\lambda-\gamma)} \frac{1}{(1-e^{-\gamma b})} \right) \left(e^{-(\eta+\lambda)(t-b)-\gamma b} - e^{-(\eta+\lambda)t} \right) \right\} \qquad (4.3)$$

Derive as follows:

$$\theta(t) = \int_0^b \left[\int_0^{t-y} H(u) f(u) du \right] g(y) dy$$

$$= \int_0^b \left[\int_0^{t-y} \lambda e^{-(\eta+\lambda)u} du \right] \frac{\gamma e^{-\gamma y}}{1-e^{-\gamma b}} dy$$

$$= \int_0^b \frac{\lambda}{\eta+\lambda} \left[1 - e^{-(\eta+\lambda)(t-y)} \right] \frac{\gamma e^{-\gamma y}}{1-e^{-\gamma b}} dy$$

$$= \int_0^b \frac{\lambda}{\eta+\lambda} \frac{\gamma e^{-\gamma y}}{1-e^{-\gamma b}} dy - \int_0^b \frac{\lambda}{\eta+\lambda} \frac{\gamma}{1-e^{-\gamma b}} e^{-(\eta+\lambda)t} e^{(\eta+\lambda-\gamma)y} dy$$

$$= \frac{\lambda}{\eta+\lambda} - \frac{\lambda}{\eta+\lambda} \frac{\gamma}{1-e^{-\gamma b}} e^{-(\eta+\lambda)t} \frac{1}{\eta+\lambda-\gamma} \left(e^{(\eta+\lambda-\gamma)b} - 1 \right)$$

$$= \frac{\lambda}{\eta+\lambda} \left\{ 1 - \left(\frac{\gamma}{\eta+\lambda-\gamma} \frac{1}{1-e^{-\gamma b}} \right) \left(e^{-(\eta+\lambda)(t-b)-\lambda b} - e^{-(\eta+\lambda)t} \right) \right\}$$

In simulation, in order to generate enrollment time y which follows truncated exponential distribution

$$g(y) = \frac{\gamma e^{-\gamma y}}{1-e^{-\gamma b}}, 0 \le y \le b; \gamma \ne 0$$

we start with CDF of y which is

$$G(y) = \frac{1-e^{-\gamma y}}{1-e^{-\gamma b}}$$

Let $z = G(y)$, so

$$y = -\frac{1}{\gamma} \ln\left(1 - z(1-e^{-\gamma b})\right)$$

where $z \sim U(0,1)$.

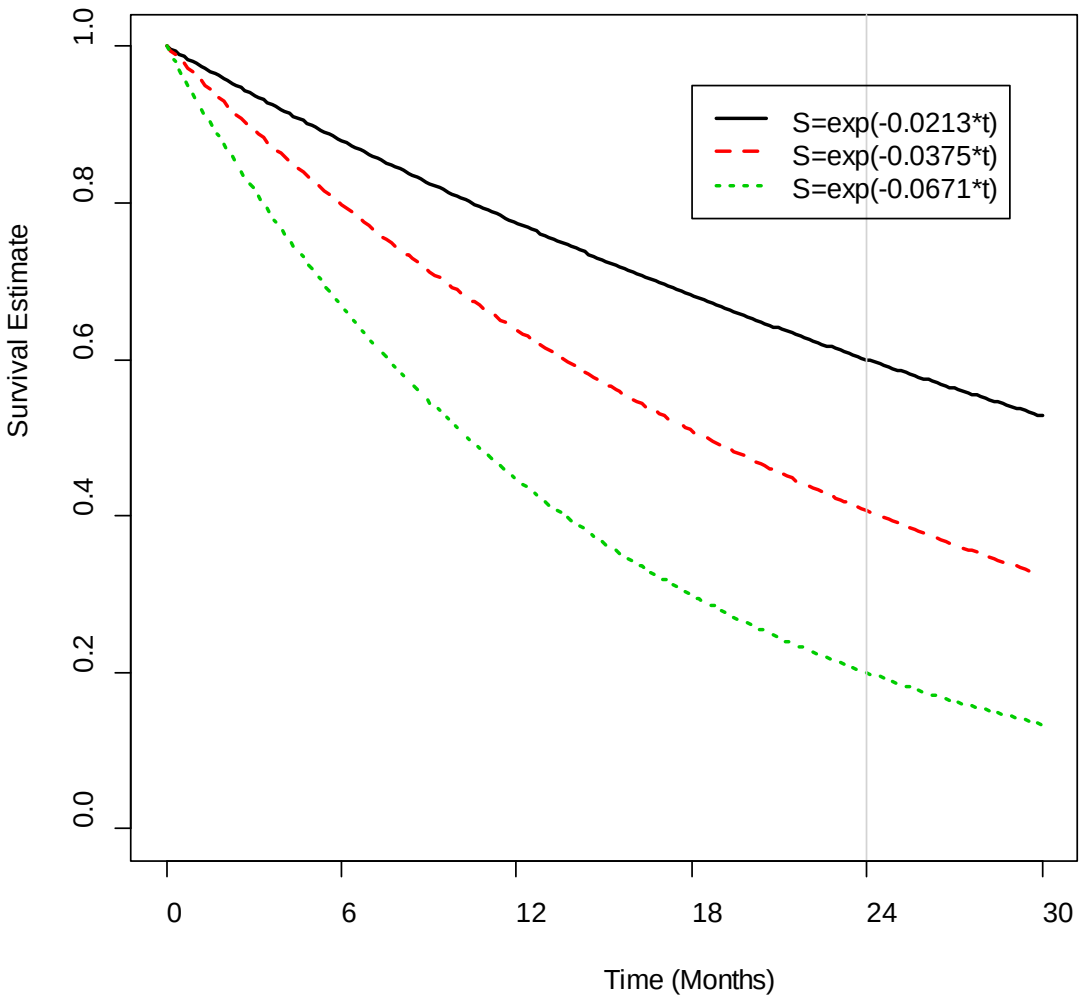
The following first 6 simulation scenarios are based on (4.2) – uniform distribution for enrollment and exponential distribution for event and censor time. The additional 3 simulation scenarios (scenarios 7, 8 and 9) are based on (4.3) – truncated exponential distribution for enrollment and exponential distribution for event and censor time.

The parameter of exponential distribution for time to progression event is estimated as follows: according to preliminary data (one of the interim analyses) from the real clinical trial data, survival rates at month 24 for 3 arms were approximately 0.20, 0.40 and 0.60. Assume event time is exponentially distributed and the corresponding accumulative event rates are 0.80, 0.60 and 0.40, then the parameters are

$\lambda_1 = 0.0213, \lambda_2 = 0.0375, \lambda_3 = 0.0671$, respectively for the *i*th arm ($i = 1,2,3$) showing in Figure 7 (Simul_data_plot_exp.R).

We generate 1,000 simulation studies (replicates). Each simulation study includes 3 arms and each arm has 35 subjects. Assume failure time and censoring time are independently exponentially distributed. Recruitment (enrollment) period is $b = 12$ months. We have simulated enrollment in several different way, including uniform, exponential, and truncated exponential distributions, but only present results here from exponential distribution (Scenarios 1 to 6) and truncated exponential distribution (Scenarios 7 to 9).

Figure 7: Exponentially Distributed Survival Function



4.1. Cases with Uniform Entry

Simulation data is generated as follows for cases with uniform entry rate:

The simulation data is generated by:

$$T_i \sim \exp(\lambda_i),\ C_i \sim \exp(\eta_i),\ X_i \sim uniform(0,b)$$

So observation time is

$$Y_i(t) = \min(T_i, C_i, (t - X_i)^+)$$

and censoring status is

$$\Delta_i(t) = I_{\{T_i \le \min(C_i, t - X_i)\}}$$

for the *i*th patient, $i = 1,2,...,n.$

Scenario 1: The Null Case (case 11: medium, medium, medium)

The null case assumes the probabilities of event are equivalent for all 3 arms,

$$\theta_1 = \theta_2 = \theta_3$$

where $\theta_i, i = 1,2,3$ denotes probability of progression (or hazard of progression) for the *i*th schedule arm. The hazard rates

$$\lambda = 0.0375,\ \eta = 0.010$$

are chosen for failure time and censoring time. Per (4.2), accumulative failure rate is $\theta(24) = 0.45$ at month 24 and $\theta(36) = 0.60$ at month 36.

The parameters used to generate simulation data are shown in Table 1.

Summary statistics for the 1,000 simulation studies is listed in Table 2.

For each of the Bayesian statistical method, the following priors are assumed:

Method 1 – Beta Model: event rate ~ $Beta(0.025,0.975)$ for all arms;

Method 2 – Gamma Model: hazard ~ $Gamma(1,10)$ for all arms;

Method 3 – Bayesian Life-Table: event rate ~ $Beta(0.025,0.975)$ for all arms and at all intervals;

Method 4 – Beta Process Model: prior is determined by 2 parameters: cumulative hazard $H_0 = h_0 t$ where $h_0 = 0.025$ and prior weight $N_0 = 1$;

Method 5 – Gibbs Sampler: prior is determined by 2 parameters: survival $S_0 = e^{-h_0 t}$ where $h_0 = 0.025$ and prior weight $N_0 = 1$.

For each of the Bayesian statistical method, 10,000 posterior samples for each of the 3 arms from each simulation study are drawn and compared to determine which arm is selected for the simulation study. Then 1,000 simulation studies (500 studies for Gibbs sampling method) are conducted and simulation results for the probability of being chosen for each of the 3 arms are summarized is Table 3 for the null case scenario.

Table 1: Parameters of Simulation Data for the Null Case

b	λ	η	t	$\theta(t)$
12	0.0375	0.010	24	0.45
12	0.0375	0.010	36	0.60

Table 2: Summary Statistics for Simulation Studies for the Null Case

	Schedule A	Schedule B	Schedule C
N	35	35	35
Number of events, mean (SD)	20.855 (2.978)	20.889 (2.885)	20.767 (2.851)
Number of censored, mean (SD)	14.145 (2.978)	14.111 (2.885)	14.233 (2.851)
Proportion of events, mean (SD)	0.596 (0.0851)	0.597 (0.0824)	0.593 (0.0815)
Proportion of censored, mean (SD)	0.404 (0.0851)	0.403 (0.0824)	0.407 (0.0815)

Table 3: Simulation Result: Probability of Being Chosen for Each Schedule by Different Statistical Methods for the Null Case

	Schedule A	Schedule B	Schedule C	Simulation run time for 1000 **studies**
Method 1 (Beta model): Based on prob. of PD event	0.324	0.334	0.342	7 minutes (1000*10000)
Method 2 (Gamma model): Based on hazard rate	0.345	0.330	0.325	6 minutes (1000*10000)
Method 3 (non-parametric – Bayesian life-table)	0.325	0.346	0.329	1.4 hours (1000*10000)
Method 4 (non-parametric – Beta process model)	0.324	0.346	0.330	4.4 hours (1000*10000)
Method 5 (non-parametric – Dirichlet process & Gibbs model)	0.318	0.350	0.332	~ 40 hours (500*10000*15)

Scenario 2: The Alternative Case (case 21: low, medium, high)

For the alternative case which assumes the event rates are different at $\theta(t) =$(low, medium, high) levels for 3 arms, the hazard rates

$$\lambda_1 = 0.0213, \lambda_2 = 0.0375, \lambda_3 = 0.0671,\ \eta = 0.010$$

are chosen for failure time of each arm and for censoring time. Per (4.2), accumulative failure rates $\theta(24)$ at month 24 and $\theta(36)$ at month 36 for each of the 3 arms are listed in Table 4.

The parameters used to generate simulation data are shown in Table 4.

Summary statistics for the 1,000 simulation studies is listed in Table 5.

Simulation results from different statistical methods for the probability of being chosen for each of the 3 arms are summarized is Table 6 for scenario 2.

Table 4: Parameters of Simulation Data for the Alternative Case

Schedule	*b*	λ	η	*t*	$\theta(t)$
A	12	0.0213	0.010	24	0.30
B		0.0375			0.45
C		0.0671			0.65
A	12	0.0213	0.010	36	0.40
B		0.0375			0.60
C		0.0671			0.80

Table 5: Summary Statistics for Simulation Studies for the Alternative Case

	Schedule A	Schedule B	Schedule C
N	35	35	35
Number of events, mean (SD)	14.384 (3.004)	20.889 (2.885)	27.150 (2.402)
Number of censored, mean (SD)	20.616 (3.004)	14.111 (2.885)	7.850 (2.402)
Proportion of events, mean (SD)	0.411 (0.0858)	0.597 (0.0824)	0.776 (0.0686)
Proportion of censored, mean (SD)	0.589 (0.0858)	0.403 (0.0824)	0.224 (0.0686)

Table 6: Simulation Result: Probability of Being Chosen for Each Schedule by Different Statistical Methods for the Alternative Case

	Schedule A	Schedule B	Schedule C	Simulation run time for 1000 **studies**
Method 1 (Beta model): Based on prob. of PD event	0.944	0.055	0.001	10 minutes (1000*10000)
Method 2 (Gamma model): Based on hazard rate	0.956	0.043	0.001	9 minutes (1000*10000)
Method 3 (non-parametric – Bayesian life-table)	0.947	0.052	0.001	5.8 hours (1000*10000)
Method 4 (non-parametric – Beta process model)	0.947	0.052	0.001	8.0 hours (1000*10000)
Method 5 (non-parametric – Dirichlet process & Gibbs model)	0.820	0.164	0.016	~ 40 hours (500*10000*15)

Scenario 3: The Alternative Case (case 31: medium, medium, high)

For the alternative case which assumes the event rates are different at $\theta(t) =$(medium, medium, high) levels for 3 arms, the hazard rates

$$\lambda_1 = 0.0375, \lambda_2 = 0.0375, \lambda_3 = 0.0671,\ \eta = 0.010$$

are chosen for failure time of each arm and for censoring time. Per (4.2), accumulative failure rates $\theta(24)$ at month 24 and $\theta(36)$ at month 36 for each of the 3 arms are listed in Table 7.

The parameters used to generate simulation data are shown in Table 7.

Summary statistics for the 1,000 simulation studies is listed in Table 8.

Simulation results from different statistical methods for the probability of being chosen for each of the 3 arms are summarized is Table 9 for scenario 3.

Table 7: Parameters of Simulation Data for the Alternative Case

Schedule	*b*	λ	η	*t*	$\theta(t)$
A	12	0.0375	0.010	24	0.45
B		0.0375			0.45
C		0.0671			0.65
A	12	0.0375	0.010	36	0.60
B		0.0375			0.60
C		0.0671			0.80

Table 8: Summary Statistics for Simulation Studies for the Alternative Case

	Schedule A	Schedule B	Schedule C
N	35	35	35
Number of events, mean (SD)	20.855 (2.978)	20.889 (2.885)	27.150 (2.402)
Number of censored, mean (SD)	14.145 (2.978)	14.111 (2.885)	7.850 (2.402)
Proportion of events, mean (SD)	0.596 (0.0851)	0.597 (0.0824)	0.776 (0.0686)
Proportion of censored, mean (SD)	0.404 (0.0851)	0.403 (0.0824)	0.224 (0.0686)

Table 9: Simulation Result: Probability of Being Chosen for Each Schedule by Different Statistical Methods for the Alternative Case

	Schedule A	Schedule B	Schedule C	Simulation run time for 1000 **studies**
Method 1 (Beta model): Based on prob. of PD event	0.489	0.499	0.012	10 minutes (1000*10000)
Method 2 (Gamma model): Based on hazard rate	0.512	0.483	0.005	9 minutes (1000*10000)
Method 3 (non-parametric – Bayesian life-table)	0.493	0.499	0.008	2 hours (1000*10000)
Method 4 (non-parametric – Beta process model)	0.494	0.498	0.008	5.5 hours (1000*10000)
Method 5 (non-parametric – Dirichlet process & Gibbs model)	0.482	0.470	0.048	~ 40 hours (500*10000*15)

Scenario 4: The Alternative Case (case 41: low, low, high)

For the alternative case which assumes the event rates are different at $\theta(t) =$(low, low, high) levels for 3 arms, the hazard rates

$$\lambda_1 = 0.0213, \lambda_2 = 0.0213, \lambda_3 = 0.0671,\ \eta = 0.010$$

are chosen for failure time of each arm and for censoring time. Per (4.2), accumulative failure rates $\theta(24)$ at month 24 and $\theta(36)$ at month 36 for each of the 3 arms are listed in Table 10.

The parameters used to generate simulation data are shown in Table 10.

Summary statistics for the 1,000 simulation studies is listed in Table 11.

Simulation results from different statistical methods for the probability of being chosen for each of the 3 arms are summarized is Table 12 for scenario 4.

Table 10: Parameters of Simulation Data for the Alternative Case

Schedule	*b*	λ	η	*t*	$\theta(t)$
A	12	0.0213	0.010	24	0.30
B		0.0213			0.30
C		0.0671			0.65
A	12	0.0213	0.010	36	0.40
B		0.0213			0.40
C		0.0671			0.80

Table 11: Summary Statistics for Simulation Studies for the Alternative Case

	Schedule A	Schedule B	Schedule C
N	35	35	35
Number of events, mean (SD)	14.384 (3.004)	14.485 (2.892)	27.150 (2.402)
Number of censored, mean (SD)	20.616 (3.004)	20.515 (2.892)	7.850 (2.402)
Proportion of events, mean (SD)	0.411 (0.0858)	0.414 (0.0826)	0.776 (0.0686)
Proportion of censored, mean (SD)	0.589 (0.0858)	0.586 (0.0826)	0.224 (0.0686)

Table 12: Simulation Result: Probability of Being Chosen for Each Schedule by Different Statistical Methods for the Alternative Case

	Schedule A	Schedule B	Schedule C	Simulation run time for 1000 **studies**
Method 1 (Beta model): Based on prob. of PD event	0.509	0.491	0.000	10 minutes (1000*10000)
Method 2 (Gamma model): Based on hazard rate	0.503	0.497	0.000	5 minutes (1000*10000)
Method 3 (non-parametric – Bayesian life-table)	0.508	0.492	0.000	2 hours (1000*10000)
Method 4 (non-parametric – Beta process model)	0.508	0.492	0.000	5.5 hours (1000*10000)
Method 5 (non-parametric – Dirichlet process & Gibbs model)	0.502	0.498	0.000	~ 40 hours (500*10000*15)

Scenario 5: The Alternative Case (case 51: low, low, medium)

For the alternative case which assumes the event rates are different at $\theta(t) =$(low, low, medium) levels for 3 arms, the hazard rates

$$\lambda_1 = 0.0213, \lambda_2 = 0.0213, \lambda_3 = 0.0375,\ \eta = 0.010$$

are chosen for failure time of each arm and for censoring time. Per (4.2), accumulative failure rates $\theta(24)$ at month 24 and $\theta(36)$ at month 36 for each of the 3 arms are listed in Table 13.

The parameters used to generate simulation data are shown in Table 13.

Summary statistics for the 1,000 simulation studies is listed in Table 14.

Simulation results from different statistical methods for the probability of being chosen for each of the 3 arms are summarized is Table 15 for scenario 5.

Table 13: Parameters of Simulation Data for the Alternative Case

Schedule	*b*	λ	η	*t*	$\theta(t)$
A	12	0.0213	0.010	24	0.30
B		0.0213			0.30
C		0.0375			0.45
A	12	0.0213	0.010	36	0.40
B		0.0213			0.40
C		0.0375			0.60

Table 14: Summary Statistics for Simulation Studies for the Alternative Case

	Schedule A	Schedule B	Schedule C
N	35	35	35
Number of events, mean (SD)	14.384 (3.004)	14.485 (2.892)	20.767 (2.851)
Number of censored, mean (SD)	20.616 (3.004)	20.515 (2.892)	14.233 (2.851)
Proportion of events, mean (SD)	0.411 (0.0858)	0.414 (0.0826)	0.593 (0.0815)
Proportion of censored, mean (SD)	0.589 (0.0858)	0.586 (0.0826)	0.407 (0.0815)

Table 15: Simulation Result: Probability of Being Chosen for Each Schedule by Different Statistical Methods for the Alternative Case

	Schedule A	Schedule B	Schedule C	Simulation run time for 1000 **studies**
Method 1 (Beta model): Based on prob. of PD event	0.491	0.494	0.015	6 minutes (1000*10000)
Method 2 (Gamma model): Based on hazard rate	0.500	0.492	0.008	5 minutes (1000*10000)
Method 3 (non-parametric – Bayesian life-table)	0.498	0.487	0.015	1 hours (1000*10000)
Method 4 (non-parametric – Beta process model)	0.498	0.487	0.015	10 hours (1000*10000)
Method 5 (non-parametric – Dirichlet process & Gibbs model)	0.454	0.466	0.080	~ 40 hours (500*10000*15)

Scenario 6: The Alternative Case (case 61: low, medium, medium)

For the alternative case which assumes the event rates are different at $\theta(t) =$(low, medium, medium) levels for 3 arms, the hazard rates

$$\lambda_1 = 0.0213, \lambda_2 = 0.0375, \lambda_3 = 0.0375\,,\ \eta = 0.010$$

are chosen for failure time of each arm and for censoring time. Per (4.2), accumulative failure rates $\theta(24)$ at month 24 and $\theta(36)$ at month 36 for each of the 3 arms are listed in Table 16.

The parameters used to generate simulation data are shown in Table 16.

Summary statistics for the 1,000 simulation studies is listed in Table 17.

Simulation results from different statistical methods for the probability of being chosen for each of the 3 arms are summarized is Table 18 for scenario 6.

Table 16: Parameters of Simulation Data for the Alternative Case

Schedule	*b*	λ	η	*t*	$\theta(t)$
A	12	0.0213	0.010	24	0.30
B		0.0375			0.45
C		0.0375			0.45
A	12	0.0213	0.010	36	0.40
B		0.0375			0.60
C		0.0375			0.60

Table 17: Summary Statistics for Simulation Studies for the Alternative Case

	Schedule A	Schedule B	Schedule C
N	35	35	35
Number of events, mean (SD)	14.384 (3.004)	20.889 (2.885)	20.767 (2.851)
Number of censored, mean (SD)	20.616 (3.004)	20.515 (2.885)	14.233 (2.851)
Proportion of events, mean (SD)	0.411 (0.0858)	0.597 (0.0824)	0.593 (0.0815)
Proportion of censored, mean (SD)	0.589 (0.0858)	0.403 (0.0824)	0.407 (0.0815)

Table 18: Simulation Result: Probability of Being Chosen for Each Schedule by Different Statistical Methods for the Alternative Case

	Schedule A	Schedule B	Schedule C	Simulation run time for 1000 **studies**
Method 1 (Beta model): Based on prob. of PD event	0.895	0.047	0.058	7 minutes (1000*10000)
Method 2 (Gamma model): Based on hazard rate	0.920	0.040	0.040	6 minutes (1000*10000)
Method 3 (non-parametric – Bayesian life-table)	0.910	0.044	0.046	1 hours (1000*10000)
Method 4 (non-parametric – Beta process model)	0.910	0.044	0.046	10 hours (1000*10000)
Method 5 (non-parametric – Dirichlet process & Gibbs model)	0.718	0.140	0.142	~ 40 hours (500*10000*15)

4.2. Cases with Truncated Exponential Entry

Simulation data is generated as follows for cases with truncated exponential entry:

1,000 simulation studies are generated. Each simulation study includes 3 arms and each arm has 35 subjects. Recruitment period is $b = 12$ months and assume patient enrollment is truncated exponential. Assume failure time and censoring time are independently exponentially distributed.

Scenario 7: The Null Case (case 11a: medium, medium, medium)

For the null case which assumes the event rates are equivalent for 3 arms, the hazard rates

$$\lambda = 0.0375,\ \eta = 0.010,\ \gamma = -0.0003$$

are chosen for failure time, censoring time and entry time for all arms. Per (4.3), accumulative failure rate is $\theta(24) = 0.45$ at month 24 and $\theta(36) = 0.60$ at month 36.

The parameters used to generate simulation data are shown in Table 19.

Summary statistics for the 1,000 simulation studies is listed in Table 20.

Simulation results from different statistical methods for the probability of being chosen for each of the 3 arms are summarized is Table 21 for scenario 7.

Table 19: Parameters of Simulation Data for the Null Case

b	λ	η	γ	t	$\theta(t)$
12	0.0375	0.010	-0.0003	24	0.45
12	0.0375	0.010	-0.0003	36	0.60

Table 20: Summary Statistics for Simulation Studies for the Null Case

	Schedule A	Schedule B	Schedule C
N	35	35	35
Number of events, mean (SD)	20.854 (2.979)	20.888 (2.886)	20.766 (2.850)
Number of censored, mean (SD)	14.146 (2.979)	14.112 (2.886)	14.234 (2.850)
Proportion of events, mean (SD)	0.596 (0.0851)	0.597 (0.0825)	0.593 (0.0814)
Proportion of censored, mean (SD)	0.404 (0.0851)	0.403 (0.0825)	0.407 (0.0814)

Table 21: Simulation Result: Probability of Being Chosen for Each Schedule by Different Statistical Methods for the Null Case

	Schedule A	Schedule B	Schedule C	Simulation run time for 1000 **studies**
Method 1 (Beta model): Based on prob. of PD event	0.323	0.335	0.342	7 minutes (1000*10000)
Method 2 (Gamma model): Based on hazard rate	0.345	0.330	0.325	6 minutes (1000*10000)
Method 3 (non-parametric – Bayesian life-table)	0.325	0.346	0.329	1.1 hours (1000*10000)
Method 4 (non-parametric – Beta process model)	0.324	0.346	0.330	4.6 hours (1000*10000)
Method 5 (non-parametric – Dirichlet process & Gibbs model)	0.320	0.348	0.332	~ 40 hours (500*10000*15)

Scenario 8: The Alternative Case (case 21a: low, medium, high)

For the alternative case which assumes the event rates are different at $\theta(t) =$(low, medium, high) levels for 3 arms, the hazard rates

$$\lambda_1 = 0.0213, \lambda_2 = 0.0375, \lambda_3 = 0.0671,$$

$$\eta = 0.010, \quad \gamma = -0.0003$$

are chosen for failure time, censoring time, and entry time for each arm. Censoring time and entry time are assumed the same for each of the 3 arm. Per (4.3), accumulative failure rates $\theta(24)$ at month 24 and $\theta(36)$ at month 36 for each of the 3 arms are listed in Table 22.

The parameters used to generate simulation data are shown in Table 22.

Summary statistics for the 1,000 simulation studies is listed in Table 23.

Simulation results from different statistical methods for the probability of being chosen for each of the 3 arms are summarized is Table 24 for scenario 8.

Table 22: Parameters of Simulation Data for the Alternative Case

Schedule	*b*	λ	η	γ	*t*	$\theta(t)$
A	12	0.0213	0.010	-0.0003	24	0.30
B		0.0375				0.45
C		0.0671				0.65
A	12	0.0213	0.010	-0.0003	36	0.42
B		0.0375				0.60
C		0.0671				0.78

Table 23: Summary Statistics for Simulation Studies for the Alternative Case

	Schedule A	Schedule B	Schedule C
N	35	35	35
Number of events, mean (SD)	14.382 (3.002)	20.888 (2.886)	27.147 (2.400)
Number of censored, mean (SD)	20.618 (3.002)	14.112 (2.886)	7.853 (2.400)
Proportion of events, mean (SD)	0.411 (0.0858)	0.597 (0.0825)	0.776 (0.0686)
Proportion of censored, mean (SD)	0.589 (0.0858)	0.403 (0.0825)	0.224 (0.0686)

Table 24: Simulation Result: Probability of Being Chosen for Each Schedule by Different Statistical Methods for the Alternative Case

	Schedule A	Schedule B	Schedule C	Simulation run time for 1000 **studies**
Method 1 (Beta model): Based on prob. of PD event	0.943	0.056	0.001	7 minutes (1000*10000)
Method 2 (Gamma model): Based on hazard rate	0.956	0.043	0.001	6 minutes (1000*10000)
Method 3 (non-parametric – Bayesian life-table)	0.947	0.052	0.001	1.2 hours (1000*10000)
Method 4 (non-parametric – Beta process model)	0.947	0.052	0.001	~10 hours (1000*10000)
Method 5 (non-parametric – Dirichlet process & Gibbs model)	0.820	0.164	0.016	~40 hours (500*10000*15)

Scenario 9: The Alternative Case (case 31a: low, medium, medium)

For the alternative case which assumes the event rates are different at $\theta(t) =$(low, medium, medium) levels for 3 arms, the hazard rates

$$\lambda_1 = 0.0213, \lambda_2 = 0.0375, \lambda_3 = 0.0375,$$

$$\eta = 0.010, \quad \gamma = -0.0003$$

are chosen for failure time, censoring time, and entry time for each arm. Censoring time and entry time are assumed the same for each of the 3 arm. Per (4.3), accumulative failure rates $\theta(24)$ at month 24 and $\theta(36)$ at month 36 for each of the 3 arms are listed in Table 25.

The parameters used to generate simulation data are shown in Table 25.

Summary statistics for the 1,000 simulation studies is listed in Table 26.

Simulation results from different statistical methods for the probability of being chosen for each of the 3 arms are summarized is Table 27 for scenario 9.

Table 25: Parameters of Simulation Data for the Alternative Case

Schedule	*b*	λ	η	γ	*t*	$\theta(t)$
A	12	0.0213	0.010	-0.0003	24	0.30
B		0.0375				0.45
C		0.0375				0.45
A	12	0.0213	0.010	-0.0003	36	0.42
B		0.0375				0.60
C		0.0375				0.60

Table 26: Summary Statistics for Simulation Studies for the Alternative Case

	Schedule A	Schedule B	Schedule C
N	35	35	35
Number of events, mean (SD)	14.382 (3.002)	20.888 (2.886)	20.766 (2.849)
Number of censored, mean (SD)	20.618 (3.002)	14.112 (2.886)	14.234 (2.849)
Proportion of events, mean (SD)	0.411 (0.0858)	0.597 (0.0825)	0.593 (0.0814)
Proportion of censored, mean (SD)	0.589 (0.0858)	0.403 (0.0825)	0.407 (0.0814)

Table 27: Simulation Result: Probability of Being Chosen for Each Schedule by Different Statistical Methods for the Alternative Case

	Schedule A	Schedule B	Schedule C	Simulation run time for 1000 **studies**
Method 1 (Beta model): Based on prob. of PD event	0.895	0.048	0.057	9 minutes (1000*10000)
Method 2 (Gamma model): Based on hazard rate	0.920	0.040	0.040	9 minutes (1000*10000)
Method 3 (non-parametric – Bayesian life-table)	0.910	0.044	0.046	1.2 hours (1000*10000)
Method 4 (non-parametric – Beta process model)	0.910	0.044	0.046	~10 hours (1000*10000)
Method 5 (non-parametric – Dirichlet process & Gibbs model)	0.718	0.140	0.142	~40 hours (500*10000*15)

Table 28: Summary of Simulation Result by Bayesian Life-Table Method

Scenario*	Operating Characteristics	Schedule A	Schedule B	Schedule C
Uniform entry				
Scenario 1 (MMM)	True Prob. of PD	0.60	0.60	0.60
	Prob. of Being Selected	0.325	0.346	0.329
Scenario 2 (LMH)	True Prob. of PD	0.40	0.60	0.80
	Prob. of Being Selected	0.947	0.052	0.001
Scenario 3 (MMH)	True Prob. of PD	0.60	0.60	0.80
	Prob. of Being Selected	0.493	0.499	0.008
Scenario 4 (LLH)	True Prob. of PD	0.40	0.40	0.80
	Prob. of Being Selected	0.508	0.492	0.000
Scenario 5 (LLM)	True Prob. of PD	0.40	0.40	0.60
	Prob. of Being Selected	0.498	0.487	0.015
Scenario 6 (LMM)	True Prob. of PD	0.40	0.60	0.60
	Prob. of Being Selected	0.910	0.044	0.046
Truncated exponential entry				
Scenario 7 (MMM)	True Prob. of PD	0.60	0.60	0.60
	Prob. of Being Selected	0.325	0.346	0.329
Scenario 8 (LMH)	True Prob. of PD	0.40	0.60	0.80
	Prob. of Being Selected	0.947	0.052	0.001
Scenario 9 (LMM)	True Prob. of PD	0.40	0.60	0.60
	Prob. of Being Selected	0.910	0.044	0.046

*Note: L is for low, M for medium and H for High.

Summary of Simulations

As discussed at the beginning of this Chapter, the purpose of simulation is to confirm that these Bayesian approaches are valid and applicable in dose-finding studies. These Bayesian methods use the posterior probability to quantify efficacy (and safety) outcomes and help clinicians make decisions on choosing the most appropriate schedule(s) to advance to next phase of clinical studies.

In most of the scenarios, method 5 (Dirichlet process model and Gibbs sampling) seems reducing the differences among 3 schedules and showing about 10%-20% differences in probability of being chosen, as compared to simulation results from other methods. This is probably due to the imputation of number of failures from previous censored observations which leads to closer total number of failures among 3 schedules at a later time. Method 5 also takes much longer running time in simulations. The objective of the dissertation is not trying to pick the best models and further research may be needed to understand the characteristics of these different methods which drive these differences.

Bayesian life-table method and Beta process model are almost identical in simulation results. Here we take the protocol pre-specified Bayesian life-table method as an example, and the simulation results for all 9 scenarios are summarized in Table 28.

Below we enclose the R codes for each of the Bayesian methods used for simulation.

In next Chapter, we will apply all these methods to the most recent unblinded data from the phase II B-cell CLL clinical trial.

5. RESULTS

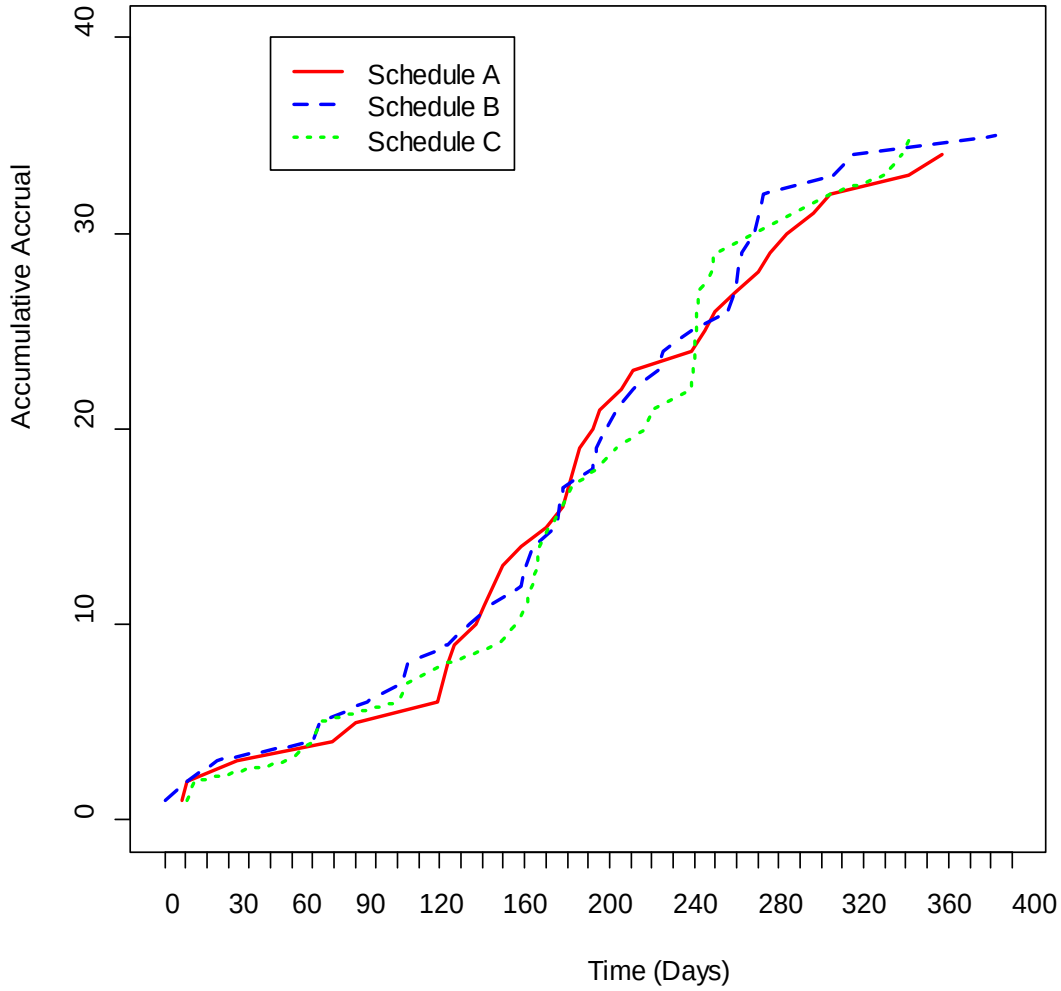
Bayesian inference has two significant features as compared to Frequentist approach. One is that the conclusion is made in term of probability and the concept of probability is easy for clinicians to understand and accept. Another feature is that Bayesian approach can use prior belief (or prior knowledge) and incorporate it into the data model from the trial itself to make inference.

This chapter summarizes the results from each of the Bayesian statistical methods described in previous chapters, based on the most recent unblinded interim data from the phase II CLL clinical trial dated on January, 2014. Only efficacy outcome (probability of progression) is discussed and presented in this dissertation. Efficacy results from all three administration schedules are compared. Conclusions and recommendations will be made in terms of which administration schedule is chosen for planning future phase III clinical study.

Generally speaking, Schedule B had lower proportion of progression disease (PD) but Schedules A and C had similar proportion of PDs (Table 29). Patients enrollment (Figure 8) was slow during the first 3 months and then stabilized after month 4. This enrollment pattern can be considered approximately uniform (Scenario 6 in Chapter 4) or truncated exponential with $\gamma < 0$ but close to zero (Scenario 9 in Chapter 4). Both Scenarios 6 and 9 in Chapter 4 simulated the (low, medium, medium) probability of PD for 3 schedule arms and the results were very similar.

Below we present the results from each of the Bayesian statistical methods introduced in Chapters 2 and 3 using the most recent unblinded data from the CLL clinical trial.

Figure 8: Patient Enrollment Over Time



5.1. Bayesian Parametric Method

5.1.1. Binomial-Beta Model

Histogram of simulation draws is intuitive and a very useful tool in helping make Bayesian inference. We focus on efficacy outcomes including data distribution, prior and posterior distributions. In some cases, we use histogram to illustrate the outcomes.

Data distribution

As of January 2014, the study has enrolled 104 patients and all of them have had at least one response assessment. Out of 104 patients, 59 had PDs (progressive diseases). The proportions of PDs [95% confidence interval] for 3 administration schedules are summarized in Table 29: 61.8% [45.4%, 78.1%]] for administration-schedule A; 48.6% [32.0%, 65.1%]] for administration-schedule B; 60.0% [43.8%, 76.2%]] for administration-schedule C. Frequentist would conclude that there is no statistically significant difference in the probability of PD among 3 schedules since these confidence intervals are overlapped. Bayesian analysis makes additional inference about the probability of Schedule B being selected via comparing posterior samples.

Prior distribution

Clinical prior *Beta*(0.025, 0.075) is used for the probability of progression for all schedules and it's conjugated prior to binomially data (details see Chapter 2).

Posterior distribution

Posterior distribution has closed form for Binomial-Beta model. The posterior probability distribution is

$$p(\theta \mid y) \propto \theta^{\alpha+y-1}(1-\theta)^{\beta+n-y-1}.$$

Posterior distributions of the probability of progression for 3 administration-schedules are shown in Figure 9 (plot_beta.R) and histogram in Figure 10 (probcalc_beta.R).

Figure 9: Posterior Probability Density Function of Progression Rate

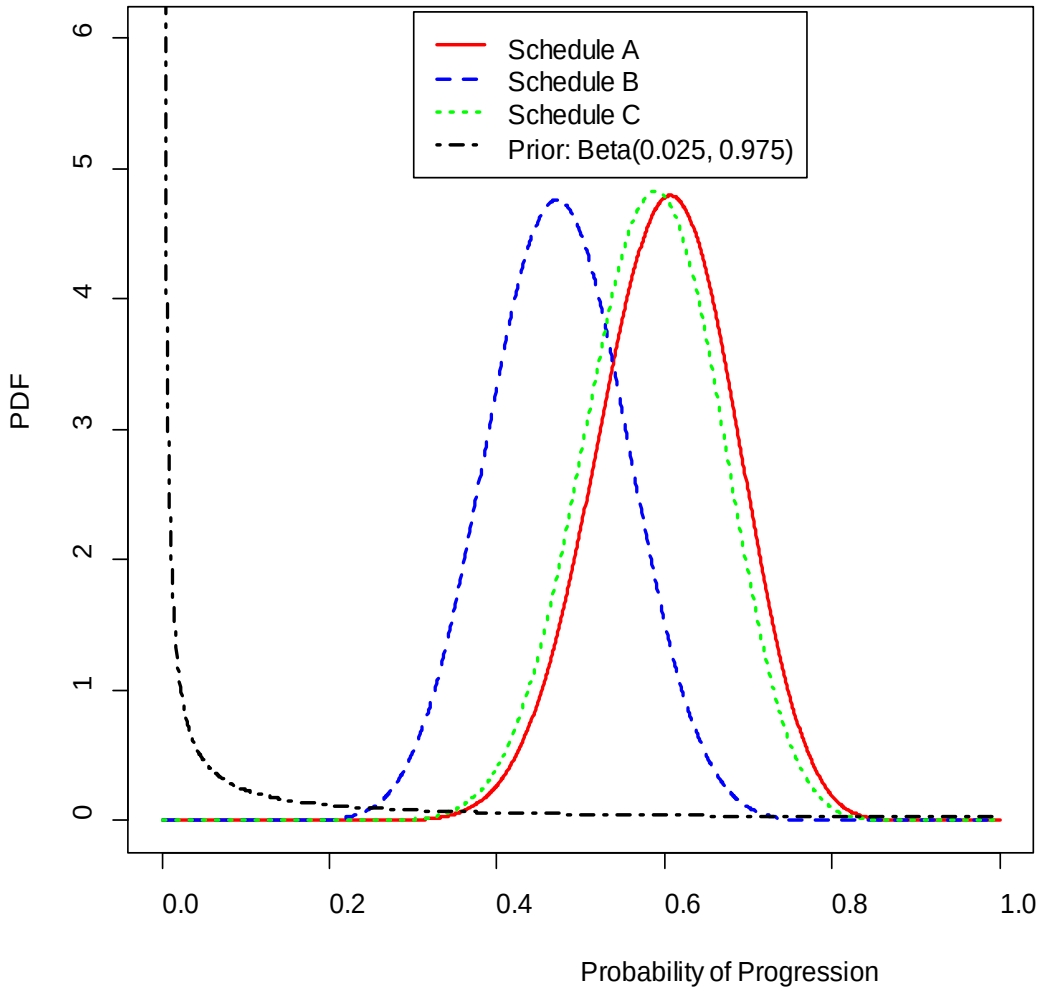
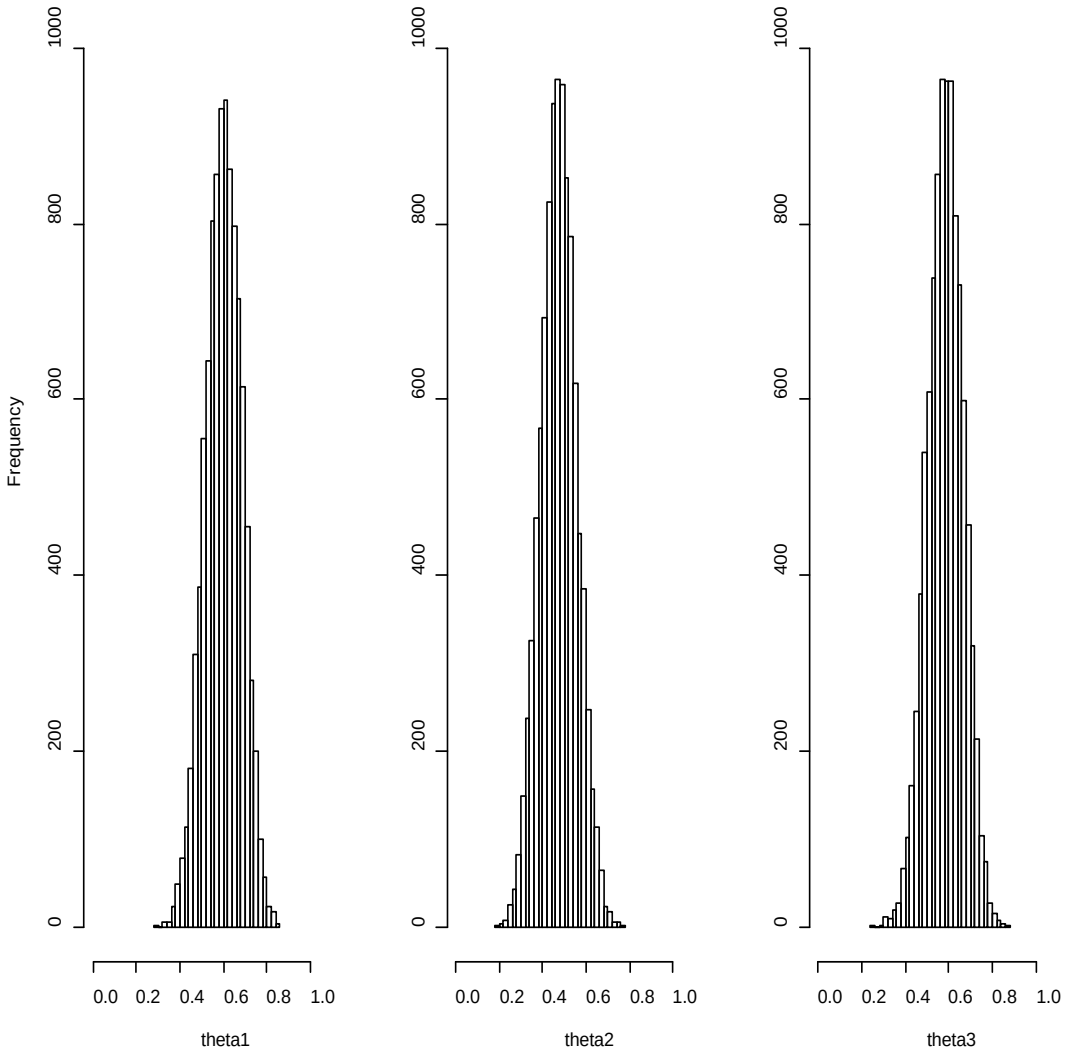


Figure 10: Histogram of Posterior Samples of Probability of Progression



Calculation and comparison of posterior distributions

At the end of the study, one administration-schedule will be selected and advanced to a phase III clinical trial. The administration-schedule with the highest probability of having the lowest progression rate and having lower than a 90% probability of excessive toxicity score will be chosen. We use the calculation of posterior probability of progression as an example to illustrate the comparisons.

Simulation is used to compare the posterior probabilities of progression among 3 administration-schedules. In order to make comparisons, 10,000 posterior samples are drawn from each of the 3 posterior distributions simultaneously (probcalc_beta.R). Simulation results show that Schedule A has 10.74% probability of having the lowest probability of PD, and Schedule B has 74.92% probability of having the lowest probability of PD, and Schedule C has 14.34% probability of having the lowest probability of PD.

Recall the decision criteria in (2.2),

$$\left.\begin{array}{c} Max(\Pr(\theta_{r,k} < \theta_{r,(-k)} \mid data)), \\ \text{and} \\ \Pr(\psi_k > \psi_0 \mid data) < \xi \end{array}\right\}$$

In this dissertation, we only consider efficacy criterion of $Max(\Pr(\theta_{r,k} < \theta_{r,(-k)} \mid data))$ so Schedule B is selected to advance to phase III study according to Binomial-Beta model.

Note that this method is to estimate the overall probability of progression without considering time to progression or adjusting for censoring.

5.1.2. Exponential-Gamma Model

Data distribution

Consider PD as an event and time to progression (TTP) as a time-to-event variable, as of January 2014 data, there are 59 PDs out of 104 patients. The total follow-up time for all 104 patients is 1053 months (total of TTP including censored observations), so there is about 1 event per 10 months. This information can also be used as clinical prior for Exponential-Gamma model. We assume TTP is exponentially distributed. Number of PDs and total follow-up time for each of the 3 arms are used as input data in this model (Table 29).

Table 29: Number (%) of Patients with PD and Follow-up Time

Administration-schedule	**Schedule A**	**Schedule B**	**Schedule C**
PD, n/N (%) [95% CI]	21/34 (61.8%) [45.4%, 78.1%]	17/35 (48.6%) [32.0%, 65.1%]	21/35 (60.0%) [43.8%, 76.2%]
Total Follow-up Time (months)	349	355	349

Prior distribution

Clinical prior *Gamma*(1, 10) is used for hazard rate of progression for all arms and it's conjugated prior to exponentially distributed data. Compared to data samples, this prior is considered similar to non-informative priors such as *Gamma*(1, 0) or *Gamma*(0, 1) in posterior determination since it only represents a very small sample.

Posterior distribution

Posterior distribution has closed form for Exponential-Gamma model. The posterior probability distribution is

$$p(\theta \mid y) \propto \theta^{\alpha+n-1} \exp\{-(\beta + \sum_{i=1}^{n} y_i)\theta\} \ .$$

Note that here we model the hazard rate of progression instead of probability of progression. Posterior distributions of the hazard rates for 3 administration-schedules as of January 2014 unblinded data are shown in Figure 11 (plot_gamma.R).

Figure 12 is histogram of simulation results for all 3 schedule arms (probcalc_gamma.R).

Figure 11: Posterior Probability Density Function of Hazard of Progression

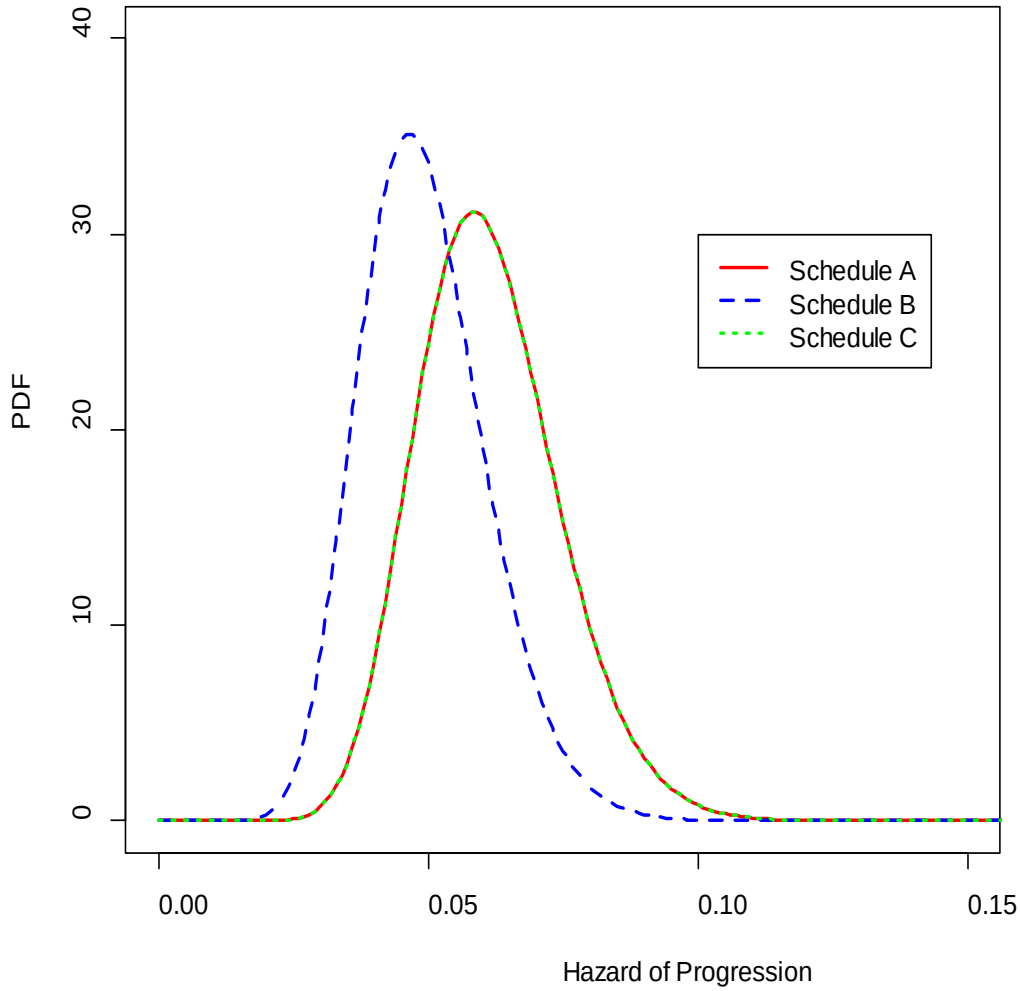
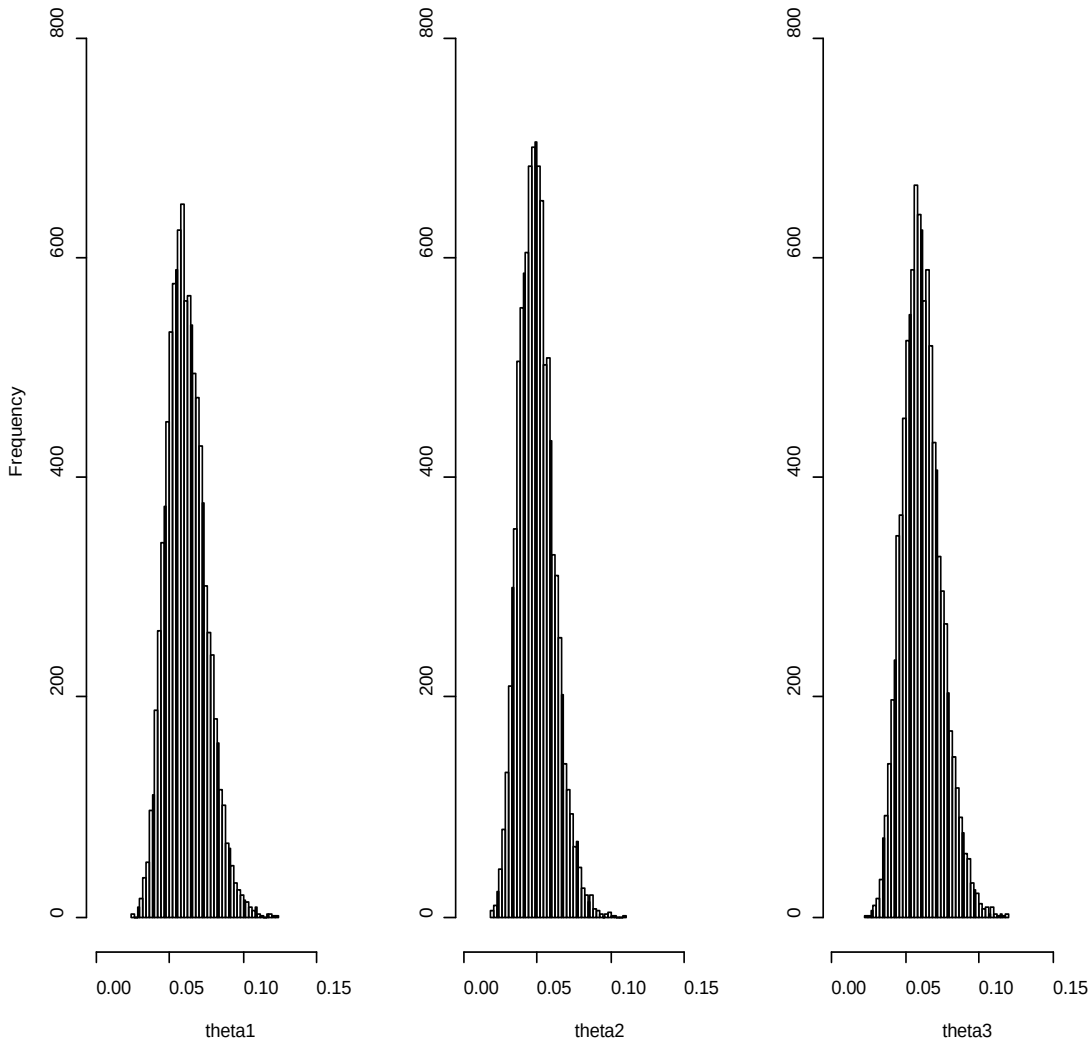


Figure 12: Histogram of Posterior Samples of Hazard of Progression



Calculation and comparison of posterior distributions

Simulation is used to compare the posterior probabilities of the hazard rate of progression among 3 administration-schedule arms. In order to make comparisons, 10,000 posterior samples are drawn from each of the 3 posterior distributions simultaneously. The administration-schedule arm with the highest probability of having the lowest hazard rate of progression is chosen to advance to phase III study if the criterion on toxicity score for that arm is also met.

Based on the January 2014 interim unblinded data, simulation results show that Schedule A has 18.25% probability of having the lowest hazard rate of PD, and Schedule B has 63.15% probability of having the lowest hazard rate of PD, and Schedule C has 18.60% probability of having the lowest hazard rate of PD (probcalc_gamma.R). Schedule B is selected to advance to phase III study according to Exponential-Gamma model.

Note that the Exponential-Gamma model is to estimate the hazard rate of progression while the Binomial-Beta model is to estimate the overall probability of progression. When we use hazard to approximate the probability of progression which will be discussed later, Schedule A has 16.64% probability of having the lowest probability of PD, and Schedule B has 62.81% probability of having the lowest probability of PD, and Schedule C has 20.55% probability of having the lowest probability of PD (probcalc_gamma.R).

5.2. Bayesian Nonparametric Method

We first briefly discuss the results from traditional nonparametric life-table and Kaplan-Meier methods, and then present the results from Bayesian nonparametric methods.

Life-Table Method

The life-table method specifies an arbitrary time interval. Event rate and survival probability are calculated for each time interval. The CLL clinical trial data has been updated every 3 months since DMC meetings were held every 3 months before unblinding. We use interval of 3-month for the calculations. Estimate of cumulative probability of PD at month 24 is listed in Table 30.

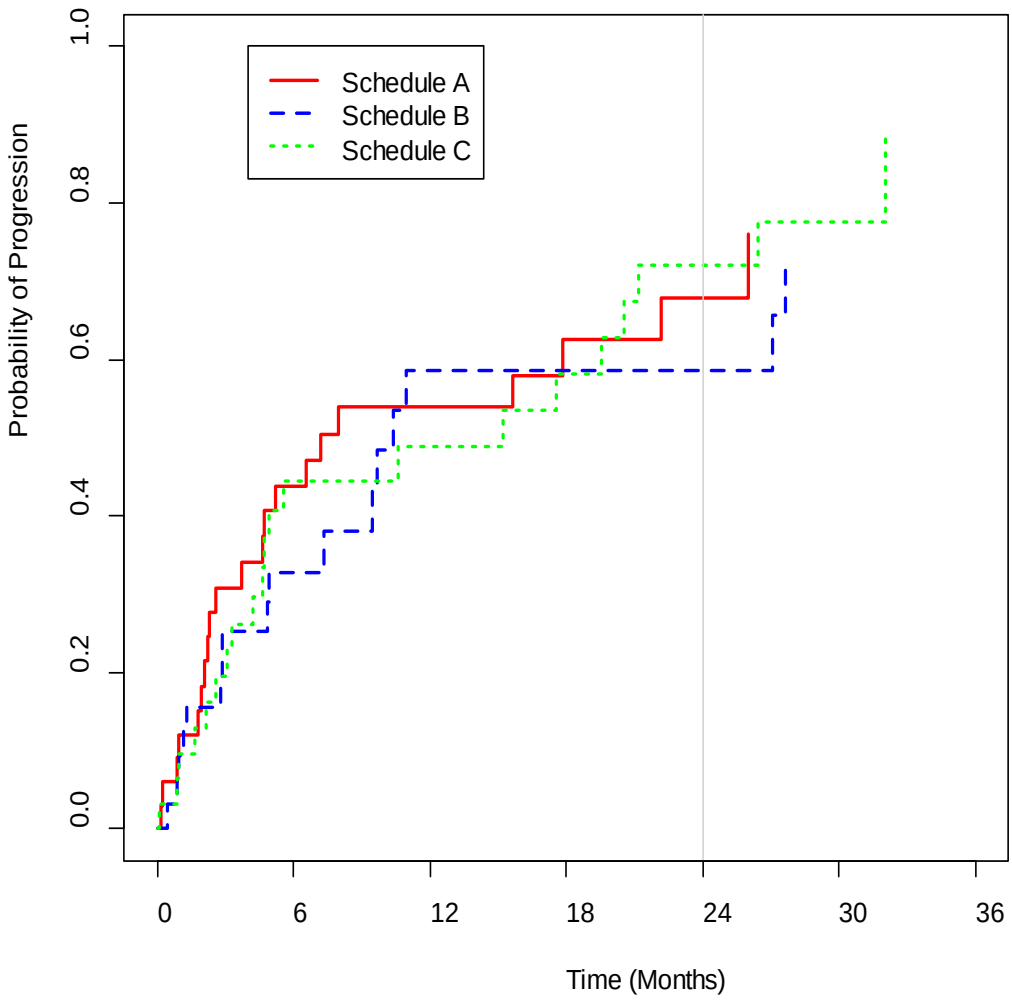
Kaplan-Meier Method

The Kaplan-Meier (KM) estimate of survival probability and KM curve is the most popular nonparametric method used in clinical trials with survival endpoint. Figure 13 is the KM curve from January 2014 unblinded interim data. Estimate of cumulative probability of PD at month 24 is listed in Table 30.

Table 30: Cumulative Probability of PD at Month 24

Method	Cumulative Probability (SE) of PD at Month 24		
	Schedule A	**Schedule B**	**Schedule C**
Life-Table	0.684 (0.094)	0.578 (0.105)	0.714 (0.094)
Kaplan-Meier	0.679 (0.093)	0.587 (0.105)	0.721 (0.093)

Figure 13: Kaplan-Meier Curve



5.2.1. Bayesian Life-Table Method

Prior distribution

See Chapter 2.

Input Data

In life-table method, the following data summary from each time-interval in each of the 3 arms of the CLL clinical trial is used as input data to estimate survival or failure: number of PDs, number of censored patients, and effective sample size. Nine intervals are used: [0, 3), [3, 6), [6, 9), [9, 12), [12, 15), [15, 18), [18, 21), [21, 24), [24, 27). Data summary for the 9 time-intervals is listed in Table 31.

Table 31: Input Data Summary for Each Time Interval in Bayesian Life-table Method

	Schedule A	**Schedule B**	**Schedule C**
Number of PDs	(10,4,3,0,0,2,0,1,1)	(8,2,1,4,0,0,0,0,0)	(6,7,0,1,0,2,2,1,1)
Number of censored	(2,1,2,0,0,3,0,1,2)	(4,6,2,0,0,1,1,0,0)	(5,2,2,0,1,0,0,0,1)
Effective sample size	(33,21.5,16,12,12, 10.5,7,6.5,4)	(33,20,14,12,8, 7.5,6.5,6,6)	(32.5,23,14,13, 11.5,11,9,7,5.5)

Calculation and comparison of posterior distribution

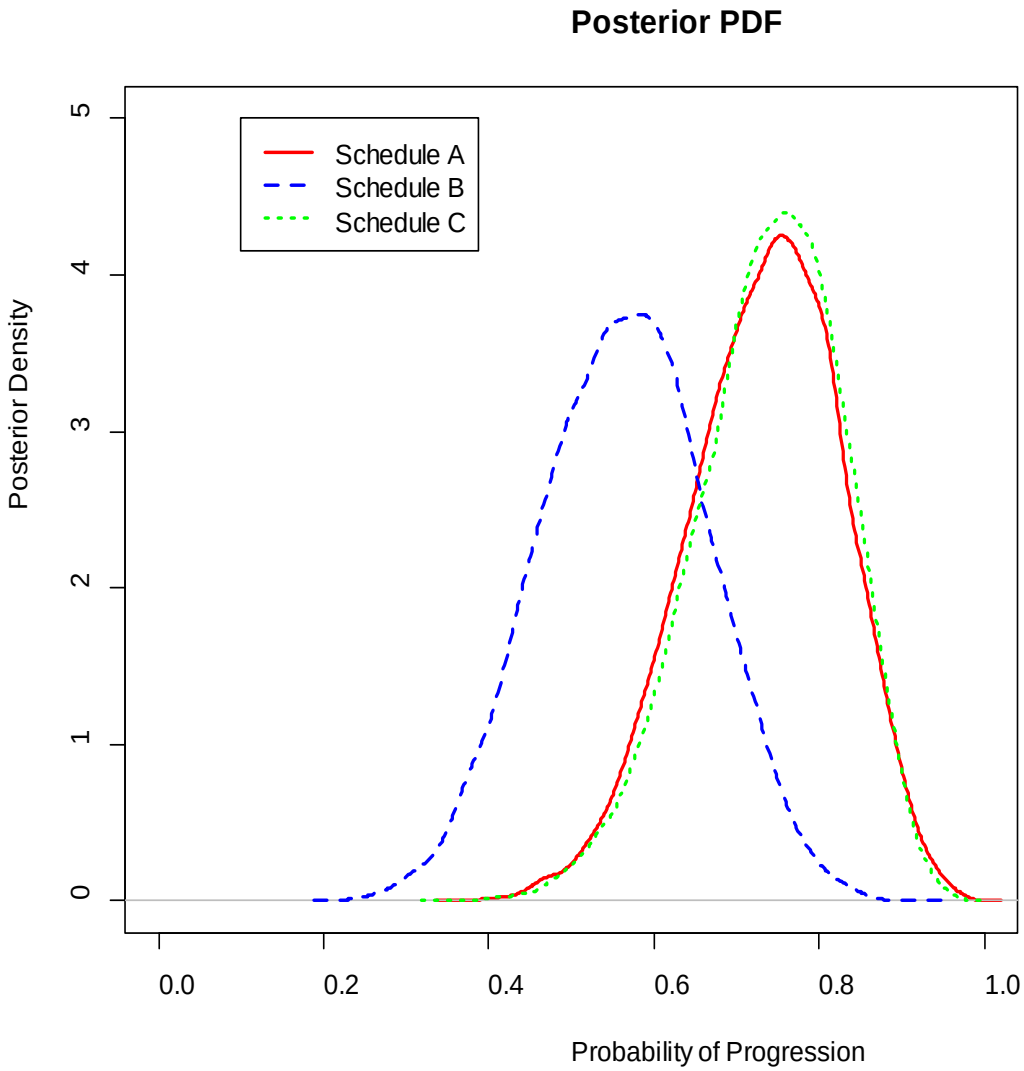
Posterior distribution for probability of PD for each time interval has closed form according to Binomial-Beta model. Posterior cumulative probability of PD see formula (2.3) in Chapter 2 with prior *Beta*(0.025,0.975) for each time interval. Cumulative probability of PD at month 24 is our target parameter.

To compare the cumulative probability of PD at month 24 for 3 schedule arms, simulation of posterior is conducted. Use Bayesian life-table method described in Chapter 2, incorporating prior *Beta*(0.025, 0.975) for hazard of progression for each time interval, and taking 10,000 posterior samples from each arm then making head-to-head comparisons.

Based on the January 2014 interim unblinded data, simulation results show that Schedule A has 8.96% probability of having the lowest probability of PD, and Schedule B has 82.78% probability of having the lowest probability of PD, and Schedule C has 8.26% probability of having the lowest probability of PD (Bayesian_LT.R). Schedule B is selected to advance to phase III study according to Bayesian life-table method.

Posterior density functions of the probability of PD at year 2 for 3 administration-schedules as of January 2014 unblinded data are shown in Figure 14 (Bayesian_LT.R).

Figure 14: Posterior Probability Density Function of Progression Rate



5.2.2. Beta Process Model

Beta Process Prior

As discussed in subsection 3.3.2, ten sample paths are randomly drawn from Beta process prior with $H_0(t) = 0.1t$ and different prior sample size ($N_0 = 1,\ N_0 = 5,\ N_0 = 25$). The path curves are shown in Figure 15. With the increase of prior sample size, the sample paths become closer to the prior guess $H_0(t)$ (BP_prior_paths.R).

Note that Beta process prior with parameters $H_0(t) = 0.1t$ and $N_0 = 1$ corresponds with prior distribution *Beta*(0.1, 0.9). Both represent a distribution with constant hazard rate of 0.1 and prior sample size of 1. Similarly, Beta process prior with parameters $H_0(t) = 0.01t$ and $N_0 = 1$ (BP_prior_paths2.R for $H_0(t) = 0.01t$) corresponds with *Beta*(0.01, 0.99) (Figure 1). In this sense, Bayesian life-table prior can be viewed as a special case of Beta process prior. Beta process prior for cumulative hazard can be in more complicated forms than exponential distribution specified here.

Input Data

The input of the data from the CLL clinical trial is the same as for the Bayesian life-table method (see Table 31).

Figure 15: Beta Process Prior Sample Paths with Different Prior Parameters

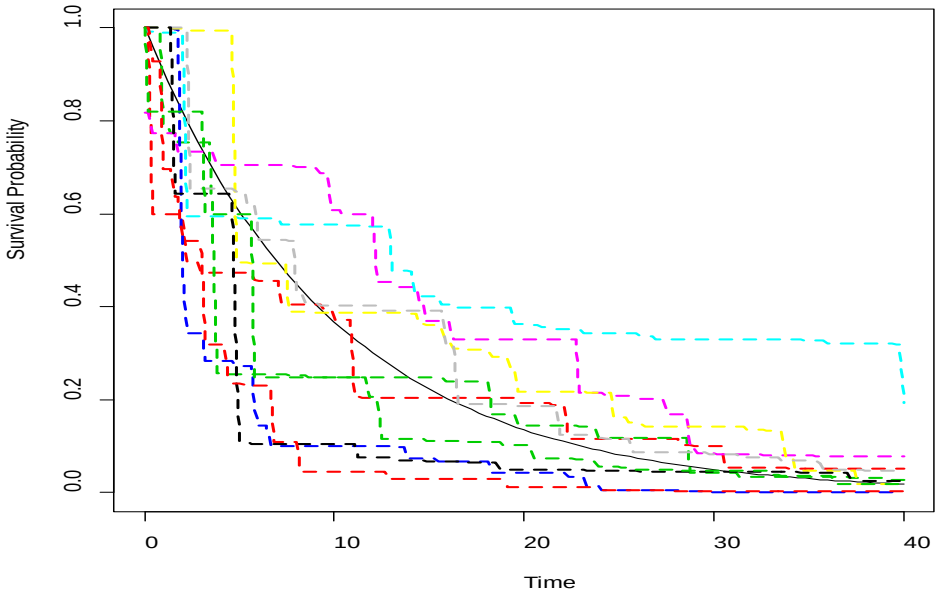


Figure 15A (above): Beta Process Prior with $H_0(t) = 0.1t$ and $N_0 = 1$

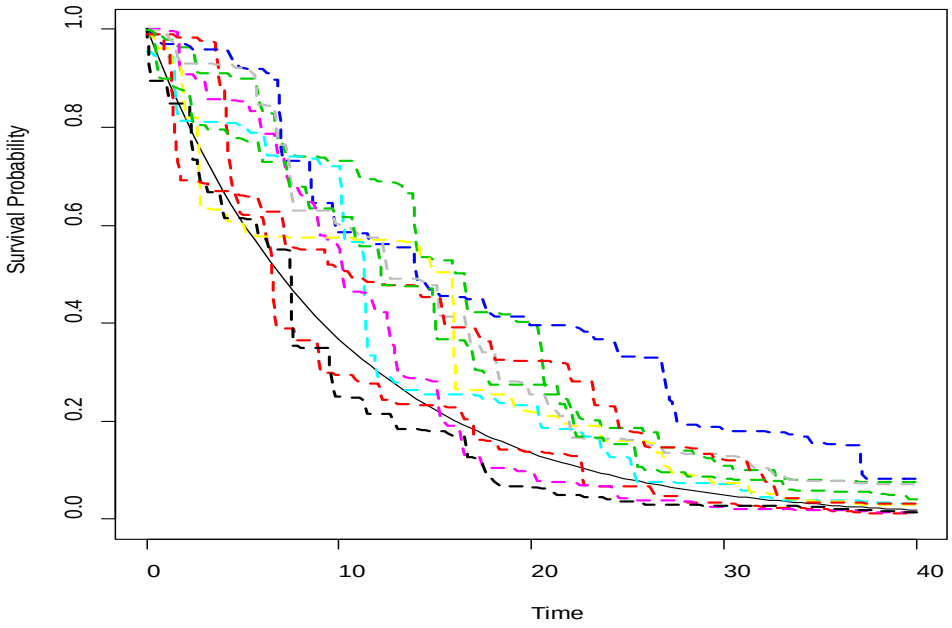


Figure 15B (above): Beta Process Prior with $H_0(t) = 0.1t$ and $N_0 = 5$

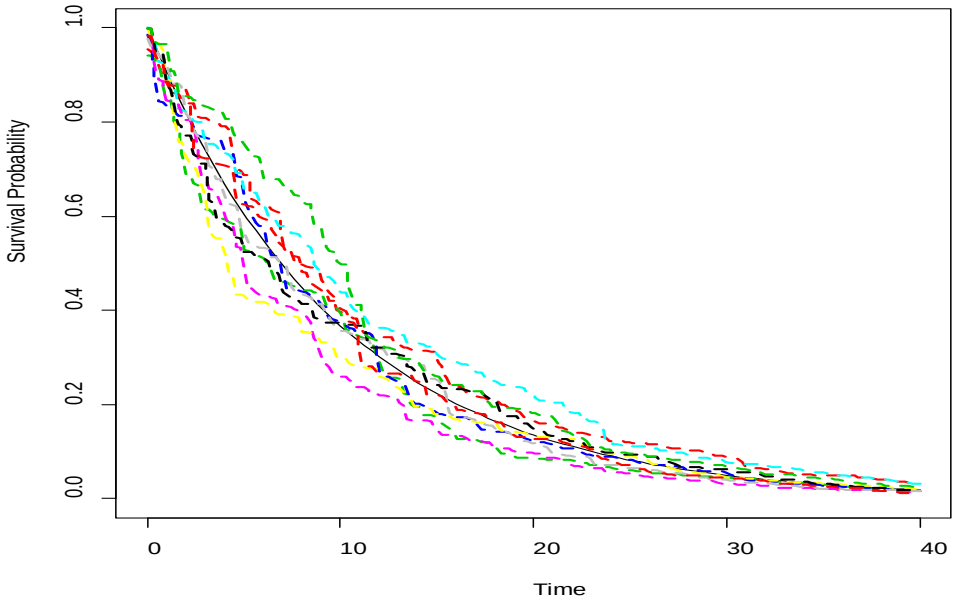


Figure 15C (above): Beta Process Prior with $H_0(t) = 0.1t$ and $N_0 = 25$

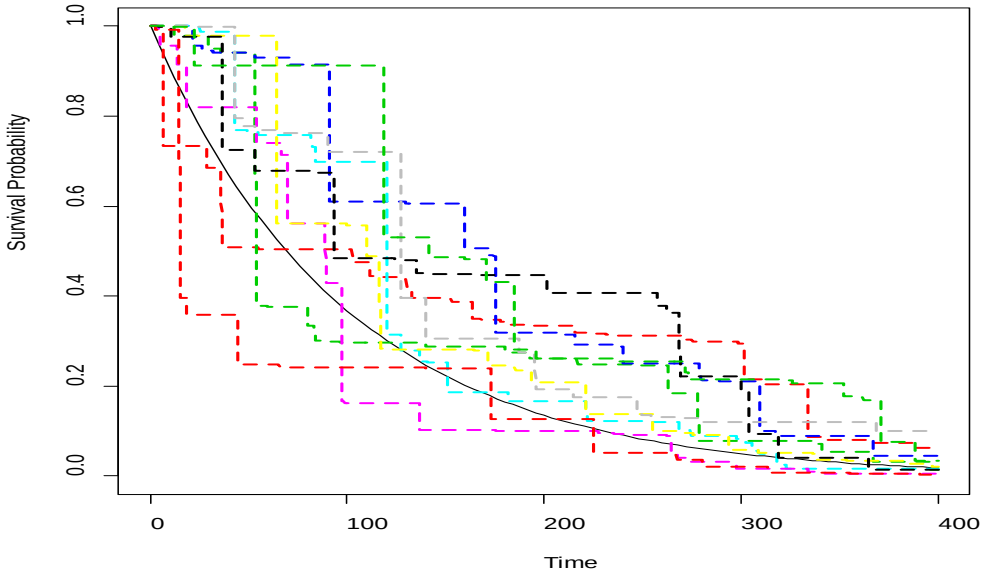


Figure 15D (above): Beta Process Prior with $H_0(t) = 0.01t$ and $N_0 = 1$

Beta Process Posterior

According to (3.16):

$$h_j \mid data \sim Beta\{N_{0j}h_{0j} + Y_j, N_{0j}(1-h_{0j}) + N_j - Y_j\},\ j = 1,...,J$$

In this application, we use Beta process prior with $H_0(t_j) = 0.1t_j$ and prior sample size $N_{0j} = 1,\ j = 1,...,J$ for each interval to draw posterior samples of hazard.

The posterior of survival is then calculated according to the following algorithm: after hazard samples are drawn from Beta process in (3.16), the cumulative hazard is calculated by $H(t) = -\sum_{t_j \le t} \ln[1 - h(t_j)]$ (Cox and Oakes, 1984). The estimated cumulative hazard is finally converted to posterior survival by $S(t) = \exp\{-H(t)\}$.

The posterior estimate of survival is based on 10,000 samples draws of hazard: take 10,000 samples of h_j, then calculate 10,000 samples of H_j and converted to 10,000 samples of $S_j = S(t_j)$. Statistical inference or comparison is based on posterior samples of survival.

Calculation and comparison of posterior distribution

To compare the cumulative probability of PD at month 24 for 3 schedule arms, simulation of posterior is conducted and head-to-head comparison of 10,000 posterior samples from each of the 3 arms is made.

Based on the January 2014 interim unblinded data, simulation results show that Schedule A has 8.96% probability of having the lowest probability of PD, and Schedule B has 82.78% probability of having the lowest probability of PD, and Schedule C has 8.26% probability of having the lowest probability of PD (BP_3arm_simul_update.R). These probabilities are exactly the same as from Bayesian life-table method. Schedule B is selected to advance to phase III study according to Beta process model.

Posterior density functions of the probability of PD at year 2 for 3 administration-schedules as of January 2014 unblinded data are shown the same as in Figure 14 from Bayesian life-table method (BP_3arm_simul_update.R).

5.2.3. Dirichlet Process Model and MCMC Gibbs Sampler

Dirichlet process prior

As discussed in subsection 3.3.3, ten random sample paths are drawn from Dirichlet process prior with $S_0(t) = \exp(-0.1t)$ and different prior sample size ($N_0 = 1$, $N_0 = 5$, $N_0 = 25$). The path curves are shown in Figure 16. With the increase of prior sample size, the sample paths become closer to the prior guess $S_0(t)$ (DP_prior_paths.R). Just like for other models, $N_0 = 1$ will be used for posterior estimation in CLL clinical trial for Dirichlet process model.

Note that Dirichlet process prior with parameters $S_0(t) = \exp(-0.1t)$ and $N_0 = 1$ corresponds with distribution *Beta*(0.1, 0.9). Both represent a distribution with constant hazard rate of 0.1 and sample size of 1. Similarly, Dirichlet process prior with parameters $S_0(t) = \exp(-0.025t)$ and $N_0 = 1$ corresponds with *Beta*(0.025, 0.975) which is used in Bayesian life-table method and Beta process model for probability of progression (Figure 1). In this sense, Dirichlet process prior and Beta process prior as well as Bayesian life-table prior all have similar properties in our application. Like Beta process prior, Dirichlet process prior for survival function can be in more complicated forms than exponential distribution.

Input Data

The input of the data from the CLL clinical trial is similar to the one for Bayesian life-table method or beta process model but includes more information beyond time intervals

at month 24 Table 32 summarizes input data for 12 time intervals: [0, 3), [3, 6), [6, 9), [9, 12), [12, 15), [15, 18), [18, 21), [21, 24), [24, 27), [27, 30), [30, 33), [33, ∞).

Table 32: Input Data Summary for Each Time Interval in Dirichlet Process Model

	Schedule A	**Schedule B**	**Schedule C**
Number of PDs	(10,4,3,0,0,2,0,1,1, 0,0,0)	(8,2,1,4,0,0,0,0,0, 2,0,0)	(6,7,0,1,0,2,2,1,1, 0,1,0)
Number of censored	(2,1,2,0,0,3,0,1,2, 0,1,1)	(4,6,2,0,0,1,1,0,0, 1,1,2)	(5,2,2,0,1,0,0,0,1, 2,1,0)

Figure 16: Dirichlet Process Prior Sample Paths with Different Prior Weight

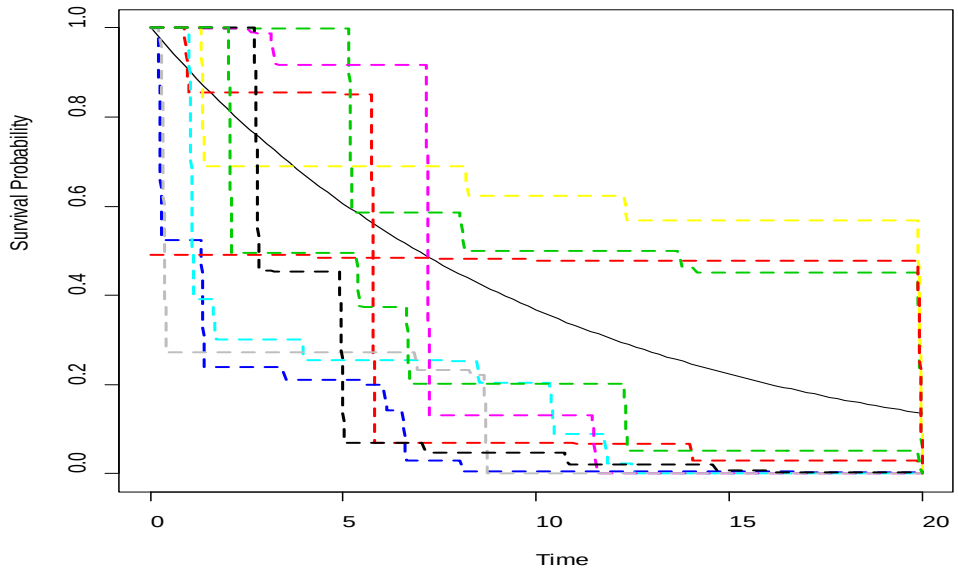


Figure 16A (above): Dirichlet Process Prior with $S_0(t) = \exp(-0.1t)$ and $N_0 = 1$

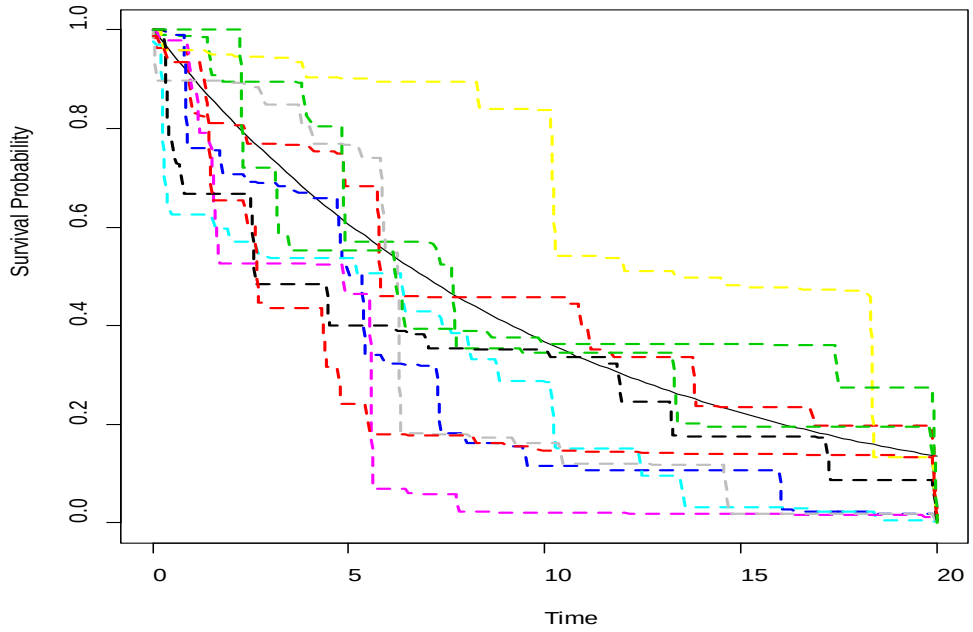


Figure 16B (above): Dirichlet Process Prior with $S_0(t) = \exp(-0.1t)$ and $N_0 = 5$

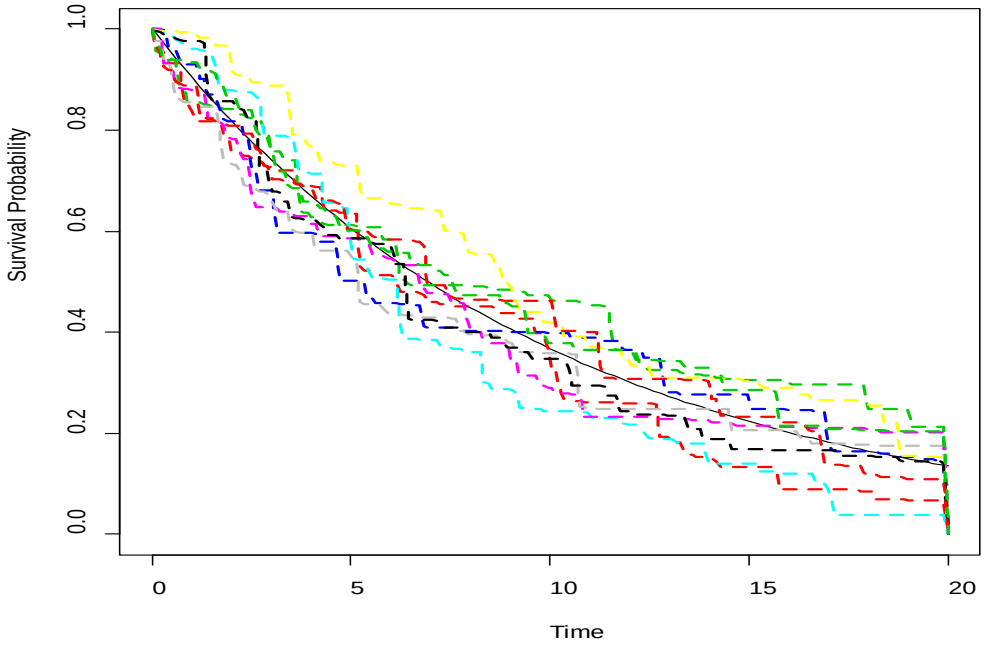


Figure 16C (above): Dirichlet Process Prior with $S_0(t) = \exp(-0.1t)$ and $N_0 = 25$

Dirichlet process posterior approximation by MCMC Gibbs sampler

Posterior distribution for right-censored data and Dirichlet process prior follows mixture Dirichlet process (MDP). We use MCMC Gibbs sampler method to draw posterior samples follows the 3 steps introduced in subsection 3.3.4 and make inference for the CLL clinical trial data.

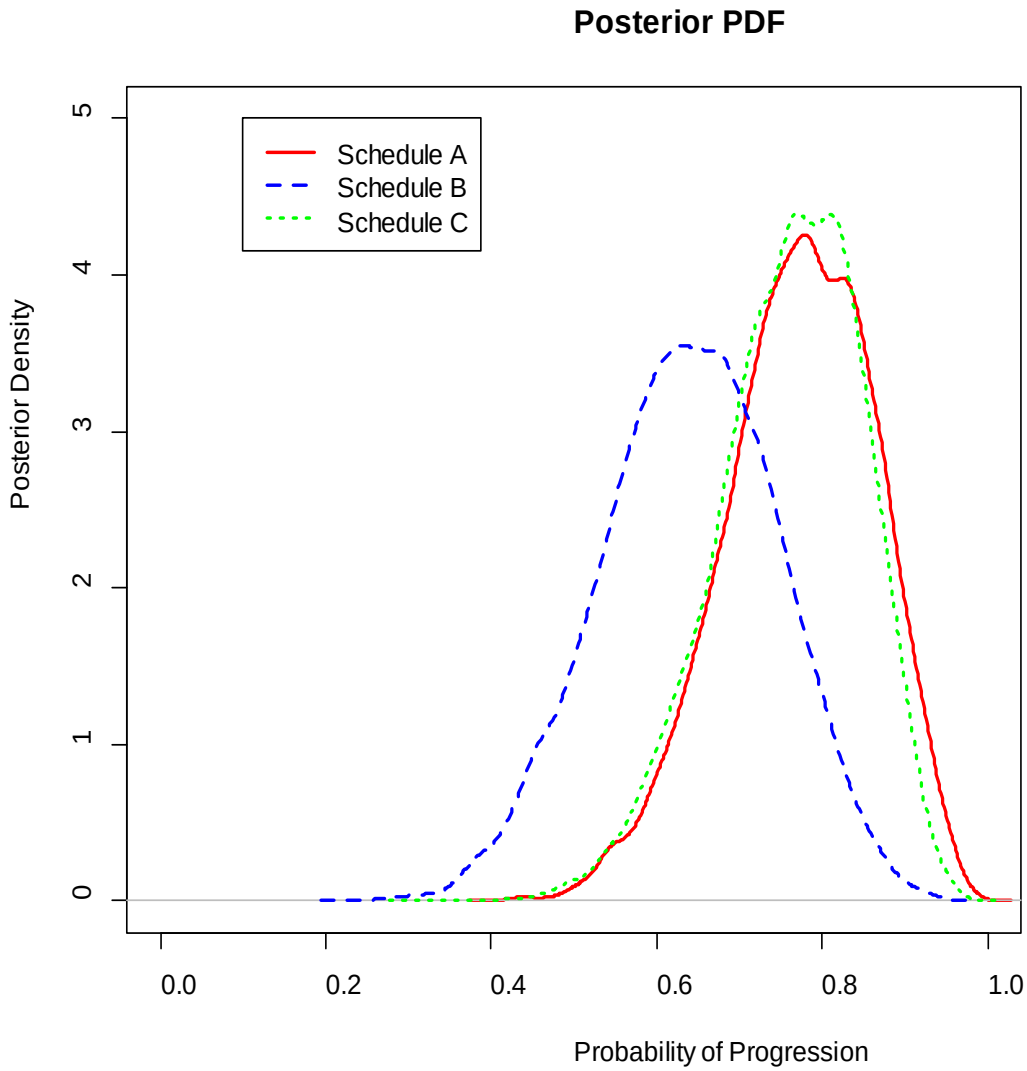
Calculation and comparison of posterior distribution

To compare the cumulative probability of PD at month 24 for 3 schedule arms, simulation of posterior is conducted and head-to-head comparison of 10,000 posterior samples from each of the 3 arms is made.

Based on the January 2014 interim unblinded data, simulation results show that Schedule A has 13.48% probability of having the lowest probability of PD, and Schedule B has 71.53% probability of having the lowest probability of PD, and Schedule C has 14.99% probability of having the lowest probability of PD (Gibbs_3arm_simul_update.R). Schedule B is selected to advance to phase III study according to Dirichlet process model by Gibbs sampler.

Posterior density functions of the probability of PD at year 2 for 3 administration-schedules as of January 2014 unblinded data are shown in Figure 17 (Gibbs_3arm_simul_update.R).

Figure 17: Posterior Probability Density Function of Progression Rate



5.2.4. Summary of Results from Different Bayesian Statistical Methods

Results from different Bayesian statistical methods are summarized in Table 33.

According to the decision rule for efficacy outcome specified in the protocol, all 5 methods conclude that Schedule B is the optimal arm and being selected to advance to future confirmatory clinical study.

Table 33: Summary of Bayesian Posterior Probability of Progression and Probability of a Schedule Being Chosen by Different Statistical Methods (Unblinded Interim Data as of January 2014)

	Binomial-Beta Model (1)	Exponential-Gamma Model (2)	Bayesian Life-table (3)	Beta Process Model (4)	Gibbs Sampler (5)
$\hat{\theta}_1$	0.600	0.631	0.733	0.733	0.771
$\hat{\theta}_2$	0.473	0.501	0.563	0.563	0.638
$\hat{\theta}_3$	0.584	0.610	0.738	0.738	0.759
$\Pr(\hat{\theta}_1 < \hat{\theta}_{-1)})$	0.1074	0.1664	0.0896	0.0896	0.1348
$\Pr(\hat{\theta}_2 < \hat{\theta}_{-2)})$	0.7492	0.6281	0.8278	0.8278	0.7153
$\Pr(\hat{\theta}_3 < \hat{\theta}_{-3)})$	0.1434	0.2055	0.0826	0.0826	0.1499

Note: $\hat{\theta}_i$ is posterior probability of progression for *i*th arm, $i = 1,2,3$. $\Pr(\hat{\theta}_i < \hat{\theta}_{-i})$ is the probability that the *i*th arm has the lowest probability of progression.

6. DISCUSSION AND CONCLUSION

6.1. Discussion of Statistical Methods

As discussed in previous chapters, the parametric models (Binominal-Beta and Exponential-Gamma models) assume data follows specific parametric distributions. In clinical trials, the assumption of binomial distribution for binary data is reasonable but Binominal-Beta model ignores the time to event information. Assumption of exponential distribution for time to event data is also generally acceptable but the constant hazard assumption is violated in some cases.

Since Exponential-Gamma parametric method is to model hazard rate $\lambda_i, i = 1,2,3$, the comparison of hazard rate from two different treatment arms is actually to model hazard ratio:

$$\Pr(\lambda_1 < \lambda_2) = \Pr(\frac{\lambda_1}{\lambda_2} < 1) = \Pr(HR < 1)$$

In this regard, the Exponential-Gamma method should lead to the same conclusion as the Cox proportional hazard model. Exponential distribution models the hazard rate, so we need to convert hazard rate of progression to cumulative hazard or probability of progression in order to compare results from other statistical models. The Exponential-Gamma parametric method in Table 33 computes and compares the cumulative hazard at $\bar{t}$:

$$H_i(\bar{t}_i) = \lambda_i * \bar{t}_i$$

where $\bar{t}_i$ is the average time to progression for the *i*th arm. If non-informative prior $Gamma(1,0)$ is used, the posterior distribution is actually only distribution of the data, so this cumulative hazard then approximately equals to the raw probability of progression:

$$E(H(\bar{t})) = E(\lambda \mid y) * \bar{t} = \frac{d}{\sum_{i=1}^{n} y_i} * \frac{\sum_{i=1}^{n} y_i}{n} = \frac{d}{n}$$

So methods (1) and (2) in Table 33 should lead to the same conclusion if cumulative hazard at $\bar{t}$ is used for method (2).

Compared to parametric methods, non-parametric models don't require parametric distributions so they are more popular for time to event variable in survival analysis. Note that method (1) is a special case of non-parametric methods (3) or (4). If only one large interval $[0, \infty)$ is used for Bayesian life-table method, then it simply becomes method (1).

Methods (3) and (4) are basically the same statistical models: Bayesian life-table method computes survival from hazard rate of all intervals, and Beta process model computes survival from cumulative hazard. But Beta process model is flexible for more complicated prior distribution other than Beta priors.

Method (5) uses Gibbs sampler for Dirichlet process model. The idea is to impute event time for right-censored observations and Gibbs sampling is a nice way for the imputation. But the implementation is computationally complicated and simulation running time is

long. Dirichlet process model considers all time intervals including those beyond the evaluation time point say at 24 month or beyond the end of the study which are unobservable, so the estimation of the event probability depends on how you group the future intervals. As discussed in Chapter 4, in most of the simulation scenarios, method (5) shows about 10%-20% differences in probability of being chosen of each schedule, as compared to simulation results from other methods. This is a future research topic to understand the characteristics of these different methods which drive these differences.

Model Selection

One common issue related to statistical analysis is *model selection*. We present results from several analysis models including both parametric and nonparametric methods, but it is difficult to give general advice on model selections.

In order to understand which model is more appropriate for a specific problem, first of all, determine the study objective for the posterior inference. For this dissertation, our aim is to provide a Bayesian computational tool to make administration-schedule selection and recommend an optimal administration-schedule for planning Phase III study. When comparing different statistical models, we need to understand there is no super-model. No model is perfect – scientific judgment plays an important role in applied Bayesian analysis. As long as the correct administration-schedule is selected, or the conclusions from different Bayesian models are consistent, then our primary objective is reached.

Secondly, we need to understand assumptions on which the model is based and whether these assumptions match the type or nature of the research data. Parametric models

require certain type of distribution of the data, but non-parametric models are generally more robust to data distributions.

Finally, one can rely on model checking and model fit statistics. Bayesian model checking and fitting (such as using of DIC and Bayesian Factor) has a long history in Bayesian statistics and has become a popular research topic recently. This feature of model selection is not the purposes of present dissertation and further investigation in this area including convergence of Gibbs sampling and the characteristics of these different methods is recommended for future work.

6.2. Discussion of Study Design

The purpose of the clinical trial is to characterize the risk/benefit profile of each starting dose administration schedules with respect to both toxicity and progression. Higher starting dose administration schedules may have a higher risk of toxicity early on, while lower starting dose administration schedules have lower but increasing risk of toxicity. Operationally, this means that early in the study, the initial doses administered are compared, while later in the study (as subjects receive higher doses) the administration schedules are compared.

A second aspect of this design is that early-onset hazards for toxicity but later-onset hazards for progression may be observed. Under this scenario toxicity and progression events could occur at different times thus making the risk/benefit assessments of the various administration schedules difficult early in the study. However, this clinical trial revealed that most progression events occurred during the first year of enrollment.

Since our major purpose of this research is to evaluate which treatment administration schedule should be carried to a larger phase III study, the decision is to be made in the later stage of the study so the impact of these scenarios is diminished.

An underlying assumption of many survival methods is treatment's proportional hazards over time. Since CLL patients may experience multiple hazards (or composite events) during the course of the study, many scenarios can occur, such as earlier hazard of toxicity for higher starting dose as well as earlier hazard of progression for lower starting dose. Hazard may cross over time among administration schedules for the CLL clinical trial. Bayesian nonparametric analyses provide a way to estimate survival when proportional hazard assumption is violated.

6.3. Discussion of Prior Selections

Choosing a prior precision parameter (N_0) requires some consideration. We also examined how much the priors might influence the posterior inference. In this clinical trial, relatively vague priors (or weak informative priors) are used. A fairly vague prior discounts the influence of prior data on posterior inference. One can increase the weight of prior if one strongly believe the historical data or prior information. Beta or Dirichlet process priors with total of all parameters equal to 1 which are generally considered as non-informative priors could be very informative in our case just as Rosner (2005) pointed out in their clinical trial, since the prior with a small sample size is applied to each time interval. If a trial consists of a large number of intervals or sample size for each interval is very small, the prior information could play an important role in the posterior, especially at the end of a clinical trial when number of subjects at risk

decreases. Dirichlet prior actually allows one to choose different prior weights over time. More complicated priors can be used if more information about the prior changes over time is available.

6.4. Conclusion

Bayesian approaches have become more and more popular in designing, monitoring and analyzing clinical trials in recent years. Health authorities encourage adaptive designs and the use of Bayesian methodology especially for early clinical studies such as phases I and II trials. In this dissertation, we used 5 different Bayesian statistical methods to evaluate and compare the probability of progression for 3 different administration schedules in a phase II clinical trial with B-cell CLL patients. According to the decision rule for efficacy outcome specified in the protocol, all 5 methods conclude that Schedule B is the optimal arm and being selected to advance to future confirmatory clinical study.

Since time to event outcome is evaluated in the CLL trial, we recommend Bayesian life-table method or Beta process model for the evaluation of such clinical endpoint among 5 Bayesian statistical methods presented in the research. Dirichlet process model with Gibbs sampling is computationally complex and running time is over 40 hours in simulation studies.

Of limitation, the conclusion of selecting Schedule B is based upon January 2014 unblinded data. Since this phase II clinical trial enrolled only 35 subjects per schedule arm, with more data coming in the next several months, the conclusion may be altered if the hazard of progression changes its course dramatically, even though it's very unlikely.

This dissertation research also demonstrates that Bayesian approaches to clinical studies are more flexible in decision making and the use of posterior probability is acceptable to clinicians and to health authorities. As applied statisticians in pharmaceutical or biotech industry, we should utilize different or alternative approaches in clinical research. Bayesian adaptive design in clinical trial is such one of the most efficient designs and popular research topics in recent years.

As of a final note, the B-cell CLL clinical trial is not all finished yet and final database lock is scheduled for late 2014. When final data is available, we will conduct another simulation using the newest incidence rate in Table 29 finally observed in the trial to understand the "power" (probability of selecting the optimal dose).

7. REFERENCES

Antoniak, C. E. (1974). Mixtures of Dirichlet processes with applications to nonparametric problems, *The Annals of Statistics* 2: 1152-1174.

Arjas, E. and Gasbarra, D. (1994). Nonparametric Bayesian inference from right censored survival data, using the Gibbs sampler, *Statistica Sinica* 4: 505-524.

Bekele, B.N. and Thall, P.F. (2004). Dose-finding based on multiple toxicities in a soft tissue sarcoma trial, *Journal of the American Statistical Association* 99: 26-35.

Bekele, B.N. and Shen, Y. (2005). A Bayesian approach to jointly modeling toxicity and biomarker expression in a phase I/II dose-finding trial, *Biometrics* 61: 344-354.

Bekele, B.N., Ji, Y., Shen, Y., and Thall, P.F. (2008). Monitoring late-onset toxicity in phase I trials using predicted risks, *Biostatistics* 9: 442-457.

Bekele, B.N., Li, Y. and Ji, Y. (2010). Risk-Group-Specific dose finding based on an average toxicity score, *Biometrics* 66: 541-548.

Berry, D.A. (1985). Interim analyses in clinical trials: Classical vs. Bayesian approaches, *Statistics in Medicine* 4: 521-526.

Brannath W., Zuber, E., Branson, M., Bretz, F., Gallo, P., Posch, M., Racine, A. (2009). Confirmatory adaptive designs with Bayesian decision tools for a targeted therapy in oncology, *Statistics in Medicine* 28: 1445-1463.

Cheung, Y.K. and Thall, P.F. (2002). Monitoring the rates of composite events with censored data in phase II clinical trials. *Biometrics* 58: 89-97.

Conaway, M. and Petroni, G. (1996). Design for phase II trials allowing for a trade-off between response and toxicity. *Biometrics* 52: 1375-1386.

Cox, D.R., and Oakes, D. (1984). *Analysis of Survival Data*, New York: Chapman and Hall.

Damien, P., Laud, P. and Smith, A. (1996). Implementation of Bayesian nonparametric inference using Beta processes, *Scandinavian Journal of Statistics* 23: 27-36.

Dey, D., Muller, P. and Sinha, D. (eds) (1998). *Practical Nonparametric and Semiparametric Bayesian Statistics*, Springer-Verlag.

Ferguson, T.S. (1973). A Bayesian analysis of some nonparametric problems, *The Annals of Statistics* 1: 209-230.

Ferguson, T.S. (1974). Prior distributions on spaces of probability measures, *The Annals of Statistics* 2: 615-629.

Ferguson, T.S. and Phadia, E. G. (1979). Bayesian nonparametric estimation based on censored data, *The Annals of Statistics* 7: 163-186.

Ferguson, T.S., Phadia, E.G. and Tiwari, R.C. (1992). Bayesian nonparametric inference, *in* M. Ghosh and P.K. Pathak (eds.), *Current Issues in Statistical Inference*: Essays in Honor of D. Basu, Haywar, California: IMS Lecture Notes – Monograph Series, Vol. 17.

Ferrajoli, A., Lee, B.N., Schlette, E., et al. (2008). Lenalidomide induces complete and partial responses in patients with relapsed and refractory chronic lymphocytic leukemia, *Blood* 111: 5291-5297.

Freedman, L.S. and Spiegelhalter, D.J. (1989). Comparison of Bayesian with group sequential methods for monitoring clinical trials, *Controlled Clinical Trials* 10: 357-367.

Gelfand, A. and Smith, A.F.M. (1990). Sampling-based approaches to calculating marginal densities, *Journal of the American Statistical Association* 85: 398-409.

Gelman, A., Carlin, J.B., Stern, H.S., and Rubin D.B. (2004). *Bayesian Data Analysis* (2nd Edition), Chapman & Hall/CRC.

Hjort, N.L. (1990). Nonparametric Bayes estimators based on beta processes in models for life history data, *The Annals of Statistics* 18: 1259-1294.

Ibrahim, J.G., Chen, M. and Sinha, D. (2001). *Bayesian Survival Analysis*, New York: Springer.

Ji, Y. and Bekele, N.B. (2009). Adaptive randomization for multi-arm comparative clinical trials based on joint efficacy/toxicity outcomes, *Biometrics* 65: 876-884.

Kalbfleisch, J.D. (1978). Nonparametric Bayesian analysis of survival time data, *Journal of the Royal Statistical Society, Series B* 40: 214-221.

Klein, J.P. and Moeschberger, M.L. (2003). *Survival Analysis Techniques for Censored and Truncated Data* (2nd Edition), Springer.

Kuo, L. and Smith, A.F.M. (1992). Bayesian Computations in Survival Models via the Gibbs Sampler, *in* J. P. Klein and P. K. Goel (eds.), *Survival Analysis: State of the Art*, 11-24. Dordrecht, The Netherlands: Kluwer Academic Publishers.

Li, G., Shih, W.J. and Wang, Y. (2005). Two-stage adaptive design for clinical trials with survival data, *Journal of Biopharmaceutical Statistics* 15: 707-718.

Rolin, J.M. (1992). Some useful properties of the Dirichlet process, *Technical Report 9207*, Center for Operations Research & Econometrics, Universite Catholique de Louvain.

Rosner, G.L. (2005). Bayesian monitoring of clinical trials with failure-time endpoints. *Biometrics* 61: 239-245.

Scala, L.D., and Glimm, E. (2011). Time-to-event analysis with treatment arm selection at interim, *Statistics in Medicine* 30: 3067-3081.

Schmidli H., Bretz, F. and Racine, A. (2007). Bayesian predictive power for interim adaptation in seamless phase II/III trials where the endpoint is survival up to some specified timepoint, *Statistics in Medicine* 26: 4925-4938.

Sinha, D. (1997). Time-discrete beta process model for interval-censored survival data, *Canadian Journal of Statistics* 25: 445-456.

Sinha, D. and Dey, D.K. (1998). Survival analysis using semiparametric Bayesian methods, *in* D. Dey, P. Muller, and D. Sinha (eds), *Practical Nonparametric and Semiparametric Bayesian Statistics*, New York: Springer-Verlag, pp. 195-211.

Spiegelhalter, D.J., Abrams, K.R., and Myles, J.P. (2004). *Bayesian Approaches to Clinical Trials and Health Care Evaluation,* New York, Wiley.

Spiegelhalter, D.J., Freedman, L.S., and Parmar, M.K.B. (1994). Bayesian approaches to randomized trials (with Discussion), *Journal of the Royal Statistical Society, Series A* 157: 357-416.

Susarla, V. and Van Ryzin, J. (1976). Nonparametric Bayesian estimation of survival curves from incomplete observations, *Journal of the American Statistical Association* 71: 897-902.

Susarla, V. and Van Ryzin, J. (1978a). Empirical Bayes estimation of a distribution (survival) function from right censored observations, *The Annals of Statistics* 6: 740-754.

Susarla, V. and Van Ryzin, J. (1978b). Large sample theory for a Bayesian nonparametric survival curve estimator based on censored samples, *The Annals of Statistics* 6: 755-768.

Thall, P.F. and Simon, R. (1994). Practical Bayesian guidelines for phase IIB clinical trials, *Biometrics* 50: 337-349.

Thall, P.F. and Sung, H.G. (1998). Some extensions and applications of a Bayesian strategy for monitoring multiple outcomes in clinical trials, *Statistics in Medicine* 17: 1563-1580.

Thall, P.F. and Cool, J.D. (2004). Dose-finding based on efficacy-toxicity trade-offs, *Biometrics* 60: 684-693.

Thall, P.F. (2009). Recent developments in Bayesian adaptive methods for clinical trials, *ASA Biopharmaceutical Section FDA-Industry Statistics Workshop*, Washington, D. C., September 23-25, 2009.

Thall, P.F. (2012). Utility-based clinical trial design and analysis, *ASA Biopharmaceutical Section FDA-Industry Statistics Workshop*, Washington, D. C., September 12-14, 2012.

Wendtner, C., Fraser, G., et al. (2011). Interim results for the safety and efficacy of different lenalidomide starting dose regimens in subjects with relapsed or refractory chronic lymphocytic leukemia (CC-5013-CLL-009 Study) [abstract], *Proceedings of the 53rd Annual Meeting and Exposition of the American Society of Hematology (ASH)*, San Diego, CA, December 10-13, 2011.

Wendtner, C., Hallek, M., et al. (2012a). Updated interim results of the safety and efficacy of different lenalidomide starting dose regimens in patients with relapsed or refractory (rel/ref) chronic lymphocytic leukemia (CC-5013-CLL-009 Study) [abstract],

Proceedings of the 54th Annual Meeting of the American Society of Hematology (ASH), Atlanta, GA, December 8-11, 2012.

Wendtner, C., Hallek, M., et al. (2012b). Updated interim results of the safety and efficacy of different lenalidomide starting dose regimens in patients with relapsed or refractory (rel/ref) chronic lymphocytic leukemia (CC-5013-CLL-009 Study) [Poster], *Poster presented at: the 54th Annual Meeting of the American Society of Hematology (ASH)*, Atlanta, GA, December 8-11, 2012.

8. APPENDIX

8.1. R Codings

Simulation for Binomial-Beta Model (Method 1):

```
rm(list=ls(all=TRUE)); start <- Sys.time (); library(Hmisc); library(foreign)
survmtd1 <- read.csv("H:/Jack/CLL 009/simulation/case 11/SURVMTD1.csv")
set.seed(123456); K=1000; # K = number of simulation studies; mark = matrix(NA,K,3)
for (k in 1:K)
{
        N=10000
        #define prior for progression as Beta(0.025, 0.975)
        alpha <- 0.025; beta <- 0.975
        n <- subset(survmtd1,study==k)[,6]
        nprog <- subset(survmtd1,study==k)[,3]
        #theta[i,j] is posterior dist for arm j from the ith posterior sample
        theta=matrix(NA,10000,3)
        for (i in 1:N)
        {
        for (j in 1:3)
                {
             theta[i,j] = rbeta(1,alpha+nprog[j], beta+n[j]-nprog[j])
        }
```

```
    }

    mark123<- tabulate(apply(theta,1,which.min),nbins=3)/N
    #print("mark123, probability that the Arm has the lowest progression rate =")
    #print(mark123)
    mark[k,] = mark123
}
method1 <- tabulate(apply(mark,1,which.max),nbins=3)/K
print(method1)
print(Sys.time () - start)
```

Simulation for Exponential-Gamma Model (Method 2):

```
rm(list=ls(all=TRUE)); start <- Sys.time (); library(Hmisc); library(foreign)
survmtd1 <- read.csv("H:/Jack/CLL 009/simulation/case 21/SURVMTD1.csv")
set.seed(123456); K=1000; # K = number of simulation studies; mark = matrix(NA,K,3)
for (k in 1:K)
{
    loop=10000
    #define prior for progression hazard
    alpha <- 1; beta <- 16
    ntime <- subset(survmtd1,study==k)[,4]
    nprog <- subset(survmtd1,study==k)[,3]
```

```
    n <- subset(survmtd1,study==k)[,6]


    #theta[i,j] is posterior dist for arm j from the ith posterior sample
    theta=matrix(NA,10000,3)
    for (i in 1:loop)
    {
        for (j in 1:3)
        {
        #theta[i,j] = (ntime[j]/n[j])*rgamma(1,alpha+nprog[j], beta+ntime[j])
        theta[i,j] = rgamma(1,alpha+nprog[j], beta+ntime[j])
        }
    }
    mark123<- tabulate(apply(theta,1,which.min),nbins=3)/loop
    #print("mark123, probability that the Arm has the lowest progression rate =")
    #print(mark123)
    mark[k,] = mark123
}
method2 <- tabulate(apply(mark,1,which.max),nbins=3)/K
print(method2)
print(Sys.time () - start)
```

Simulation for Bayesian Life-Table Method (Method 3):

```
rm(list=ls(all=TRUE));start <- Sys.time ();set.seed(123456);K=1000; # K = number of
simulation studies; mark = matrix(NA,K,3)
for (k in 1:K)
{
    loop=10000;    maxgrps=3;     maxtime=9
    y <- t(subset(survmtd3,study==k)[1:9,6:8])
    d <- t(subset(survmtd3,study==k)[1:9,3:5])
    theta <- matrix(0, nrow=loop, ncol=maxgrps)
    for (i in 1:loop)
    {
        for (j in 1:maxgrps)
        {
            temp.eff <- 1.0
            for (h in 1:maxtime)
            {
                a=0.025; b=0.975; event <- d[j,h]+a; nonevent <- y[j,h]-d[j,h]+b
                out1 <- rbeta(1, event, nonevent)
                temp.eff<-temp.eff*(1.0-out1)
            }
            theta[i,j]<-1-temp.eff
```

```
        }
    }
    mark123<- tabulate(apply(theta,1,which.min),nbins=3)/loop
    #print("mark123, probability that the Arm has lowest progression rate =")
    #print(mark123)
    mark[k,] = mark123
}
method3 <- tabulate(apply(mark,1,which.max),nbins=3)/K
print(method3)
print(Sys.time () - start)
```

Simulation for Beta Process Model (Method 4):

```
rm(list=ls(all=TRUE)),start <- Sys.time (),set.seed(123456),K=1000, # K = number of
simulation studies; mark = matrix(NA,K,3)
for (k in 1:K)
{
    loop=10000; maxgrps=3; maxtime=9
    for (i in 1:loop)
    {
        for (j in 1:maxgrps) # begin loop for groups
```

```
        {
            for (h in 1:maxtime)
            {
                h0=0.025;  N0=1;
                evt <- N0*h0 + d[j,h];  nonevt <- N0*(1-h0) + y[j,h] - d[j,h]
                hh[j,h] <- rbeta(1,evt, nonevt); logh[j,h] <- -log(1-hh[j,h])
            } # end loop of k
            chf[j,] <- cumsum(logh[j,]); surv[j,] <- exp(-chf[j,]); event[j,] <- 1-surv[j,]
        } # end loop for groups
        theta[i,] <- t(event[, maxtime])
    } # end loop of i
    mark123<- tabulate(apply(theta,1,which.min),nbins=3)/loop
    #print("mark123, probability that the Arm has the lowest progression rate =")
    #print(mark123)
    mark[k,] = mark123
}  #end loop of k
method4 <- tabulate(apply(mark,1,which.max),nbins=3)/K
print(method4)
```

Simulation for Dirichlet Process Model (Gibbs Sampler) (Method 5):

```
rm(list=ls(all=TRUE));start <- Sys.time ();set.seed(123456); library(MCMCpack);
library(VGAM); N=500; #N = number of simulation studies; mark = matrix(NA,N,3)
for (n in 1:N)
{
    N0 <- 1; interval <- length(tj); maxgrps <- 3; loop <- 10000;
    for (l in 1:maxgrps)
    {
        alpha[l,] <- N0*(exp(-0.1*tj_1)-exp(-0.1*tj))
        theta[l,] <- rdirichlet(1,alpha[l,]);   #this is a draw from the prior;
    }
for (m in 1:loop)
{
    for (k in 1:15)
    {
        for (j in 1:maxgrps)
        {
            for(i in 1:(interval-1))
            {
                Z[i,(i+1):interval]<rmultinom(1,size=lambda[j,i],prob=theta[j,-c(1:i)])
            }         #end loop of i
            sumZ[j,] <- colSums(Z,na.rm=T), R[j,] <- alpha[j,]+d[j,]+sumZ[j,]
```

```
            theta[j,] <- rdirichlet(1,R[j,]), thetacumu[j,] <- cumsum(theta[j,])
            thetaend[j,] <- thetacumu[j,(interval-1)]
        }           #end loop of j
    }               #end loop of k
    Finaltheta[m,] <- thetaend
}                   #end loop of m
mark123<- tabulate(apply(Finaltheta,1,which.min),nbins=3)/loop;  mark[n,] = mark123;
}  #end loop of n
method5 <- tabulate(apply(mark,1,which.max),nbins=3)/N
print(method5);   print(Sys.time () - start);
```

Part of R code for Binomial-Beta model in Chapter 5 (probcalc_beta.R):

```
N=10000; alpha <- 0.025; beta <- 0.975; n <- c(34,35,35); nprog <- c(21,17,21);

#theta[i,j] is posterior dist for arm j from the ith posterior sample

theta=matrix(NA,10000,3)

for (i in 1:N)

{

        for (j in 1:3)

        {

              theta[i,j] = rbeta(1,alpha+nprog[j], beta+n[j]-nprog[j])

        }

}

mark123<- tabulate(apply(theta,1,which.min),nbins=3)/N

print("mark123, probability that the Arm has the lowest progression rate =")

print(mark123)
```

Part of R code for Exponential-Gamma model in Chapter 5 (probcalc_gamma.R):

```
N=10000

#define prior for progression as gamma(1,10): assume 1 event per 10 month clinical
prior, as of January 2014, average 1 event per 10 months.

a <- 1; b <- 10; n <- c(34,35,35); nprog <- c(21,17,21); ntime <- c(349,355,349);

#theta[i,j] is posterior dist for arm j from the ith posterior sample
theta=matrix(NA,10000,3)
for (i in 1:N)
{
        for (j in 1:3)
        {
            #theta[i,j] = (ntime[j]/n[j])*rgamma(1,a+nprog[j], b+ntime[j])
            theta[i,j] = rgamma(1,a+nprog[j], b+ntime[j])
        }
}
mark123<- tabulate(apply(theta,1,which.min),nbins=3)/loop
print("mark123, probability that the Arm has the lowest progression rate =")
print(mark123)
```

Part of R code for Bayesian life-table method in Chapter 5 (Bayesian_LT.R):

```
Alpha0 <- 0.025; beta0 <- 0.975

EffProb <- matrix(0, nrow=loop, ncol=maxgrps)

for (i in 1:loop)

{

        for (j in 1:maxgrps)

        {

            temp.eff <- 1.0

                for (h in 1:maxtime)

                {

                    alpha <- EffEvents[j,h]+alpha0

                    beta <- N.eff[j,h]-EffEvents[j,h]+beta0

                    out <- rbeta(1, alpha, beta)

                    temp.eff<-temp.eff*(1.0-out)

                }

            EffProb[i,j]<-1-temp.eff

        }

}

mark123 <- table(apply(EffProb,1,which.min))/loop

print("mark123, probability that the Arm has the lowest progression rate =")

print(mark123)

par(mfrow=c(1,1))
```

```
plot(density(EffProb[,1]),type="l",  col="red", lwd=2, ylim=c(0,5), xlim=c(0, 1),
xlab="Probability of Progression", ylab="Posterior Density", main="Posterior Probability
Density Function of Progression")
lines(density(EffProb[,2]), lty=2,  col="blue", lwd=2)
lines(density(EffProb[,3]), lty=3,  col="green", lwd=2)
legend(0.1, 5, legend=c("Schedule A", "Schedule B", "Schedule C"), col=c("red",
"blue","green" ),lty=c(1,2,3),lwd=c(2,2,2))
```

Part of R code for Beta process prior in Chapter 5 (BP_prior_paths.R):

```
library(MCMCpack)
N0=1; t=seq(0,40,length=200)
H=0.1*t; HM=c(H[2:length],4.020); W=HM-H;
p=N0*W; q=N0*(1-W);
rbetaf <- array(NA, dim=c(10,length))
chf <- array(NA, dim=c(10,length))
surv <- array(NA, dim=c(10,length))
for(i in 1:length)
{
   rbetaf[,i] <- rbeta(10,p[i],q[i])
}
chf <- t(apply(rbetaf,1,cumsum))
surv <- exp(-chf)
curve(exp(-0.1*x),xlim=c(0,40),ylim=c(0,1),xlab="Time",
ylab="Survival Probability",main="Beta Process Prior with H0(t)=0.1t and N0=1")
lines(t,surv[1,],type="l",lty=2,lwd=2,col=2)
……
lines(t,surv[10,],lty=2,lwd=2,col=11)
```

Part of R code for Gibbs sampling in Chapter 5 (Gibbs_3arm_simul_update.R):

```
rm(list=ls(all=TRUE));start <- Sys.time ();
library(MCMCpack);library(VGAM);library(foreign)
for (n in 1:N)
{
    set.seed(123456+n); N0 <- 1
    tj_1   <- c(0,3,6, 9,12,15,18,21,24,27,30,33)
    tj     <- c(3,6,9,12,15,18,21,24,27,30,33,99999)
    interval <- length(tj);  maxgrps  <- 3;  loop <- 10000
    for (l in 1:maxgrps)
    {
    alpha[l,] <- N0*(exp(-0.1*tj_1)-exp(-0.1*tj))
    #alpha=N0*(S0(t(j-1))-S0(t(j))) is prior for number of deaths at each interval
    theta[l,] <- rdirichlet(1,alpha[l,])
    #this is a draw from the prior: prior prob of death at each interval
    }
for (m in 1:loop)
{
    for (k in 1:15)
```

```
    {
        for (j in 1:maxgrps)
        {
            for(i in 1:(interval-1))
            {
                Z[i,(i+1):interval]<-rmultinom(1,size=lambda[j,i],prob=theta[j,-c(1:i)])
            }         #end loop of i
            R[j,] <- alpha[j,]+d[j,]+colSums(Z,na.rm=T)
            theta[j,] <- rdirichlet(1,R[j,])
            thetacumu[j,] <- cumsum(theta[j,])
            thetaend[j,] <- thetacumu[j,(interval-3)]
        }              #end loop of j
    }              #end loop of k
    Finaltheta[m,] <- thetaend
}              #end loop of m
print("probability of progression for Arm A =")
print(mean(Finaltheta[,1]))
print("probability of progression for Arm B =")
print(mean(Finaltheta[,2]))
```

```
print("probability of progression for Arm C =")

print(mean(Finaltheta[,3]))

mark123<- tabulate(apply(Finaltheta,1,which.min),nbins=3)/loop

print("mark123, probability that the Arm has the lowest progression rate =")

print(mark123)

par(mfrow=c(1,1))

plot(density(Finaltheta[,1]),type="l",  col="red", lwd=2, ylim=c(0,5), xlim=c(0, 1),
xlab="Probability of Progression", ylab="Posterior Density", main="Posterior Probability
Density Function of Progression")

lines(density(Finaltheta[,2]), lty=2,  col="blue", lwd=2)

lines(density(Finaltheta[,3]), lty=3,  col="green", lwd=2)

legend(0.1, 5, legend=c("Schedule A", "Schedule B", "Schedule C"), col=c("red",
"blue","green" ),lty=c(1,2,3),lwd=c(2,2,2))

}  #end loop of n

print(Sys.time () - start)
```

8.2. UMDNJ eIRB Study Approval Documentation

Subject: UMDNJ eIRB: Study Approved

** This is an auto-generated email. Please do not reply to this email message.
The originating e-mail account is not monitored.
If you have questions, please contact your local IRB office or log into eIRB.umdnj.edu **

DHHS Federal Wide Assurance Identifier:
FWA00001861

IRB Chair Person: Nancy Fiedler

IRB Director: Donna Hoagland

Effective Date: 4/15/2013

eIRB Notice of IRB Determination

Study ID: Pro2013002897

Title: BAYESIAN STATISTICAL ANALYSIS IN A PHASE II DOSE-FINDING TRIAL WITH SURVIVAL ENDPOINT IN PATIENTS WITH B-CELL CHRONIC LYMPHOCYTIC LEUKEMIA

Principal Investigator:	Shiansong Li	**Study Coordinator:**	
Co-Investigator(s):	Weichung Shih	**Other Study Staff:**	There are no items to display

Review Type:	Non-Human	**Expedited Category:**	N/A	**Exempt Category:**	N/A

CURRENT SUBMISSION STATUS

Submission	Research Protocol/Study	**Submission**	Approved

Type:		**Status:**	
Determination Date:	4/15/2013	**Expiration Date:**	N/A

The activities described in this application does not meet the regulatory definition of human subjects research provided in 45 CFR 46.102. Therefore, this project does not require approval by the IRB as submitted. Please note that changes to the project must be submitted to the IRB for review prior to implementation to determine if the changes incorporate elements of human subjects research activities which require IRB oversight.

ALL APPROVED INVESTIGATOR(S) MUST COMPLY WITH THE FOLLOWING:

1. Conduct the project as submitted to the IRB.

2. **Amendments/Modifications/Revisions** : If you wish to change any aspect of this project, you are required to obtain IRB review and approval prior to implementation of these changes unless necessary to eliminate apparent immediate hazards to subjects.

3. **Unanticipated Problems**: Unanticipated problems involving risk to subjects or others must be reported to the IRB Office (45 CFR 46, 21 CFR 312, 812) as required, in the appropriate time as specified in the attachment online at: http://www.umdnj.edu/hsweb

4. **Protocol Deviations and Violations** : Deviations/violations of the project must be reported to the IRB Office (45 CFR 46, 21 CFR 312, 812) as required, in the appropriate time as specified in the attachment online at: http://www.umdnj.edu/hsweb

5. **Completion of Study:** If your school requires, notify the IRB when your study has been stopped for any reason.

6. The Investigator(s) did not participate in the review, discussion, or vote of this protocol.

7. Letter Comments: *There are no additional comments.*