MODELS TO SELECT CHEMOTHERAPY, ENDOCRINE THERAPY OR BIOLOGIC AGENT TO ADVANCE TO PHASE III TRIALS FOR ADVANCED BREAST CANCER

BY XINQUN CHEN

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, PhD

And Approved by

_____

_____

_____

_____

New Brunswick, New Jersey

May, 2015

ABSTRACT OF THE DISSERTATION

Models to Select Chemotherapy, Endocrine Therapy or Biologic Agent to Advance to Phase III Trials for Advanced Breast Cancer

by XINQUN CHEN

Dissertation Director:

Professor Weichung Joe Shih

In order to improve the success rate of oncology phase III trials giving promising results in previous phase II studies, the Exponential-Gamma statistical model based on Bayesian framework is adopted and applied into the oncology drug trial for patients with the advanced breast cancer to provide insight of the likelihood that experimental regimen will provide a clinical benefit such as prolongation of survival compared with standard therapy in subsequent phase III clinical trials to make a phase III go/no go decision. The Bayesian statistical model is a hybrid Bayesian/frequentist approach where the phase III test is still in the classical (frequentist) framework, and the preceding phase II or phase III studies are used to evaluate the probability of success of such a phase III test by a Bayesian approach. The information extracted from advanced breast cancer clinical trials from 1990 to 2012 and survival data from phase II or III studies of the experimental and

control regimens are used to model the survival hazard distributions for the two regimens. In addition, the existing statistical method is expanded to include two new models, the Weibull-Inverse Gamma model and Piece-wise Exponential model to derive the expected power. We study and explore retrospectively the consistence and inconsistence between calculated expected powers and the significance of actual subsequent phase III study results with endpoints of progression free survival and overall survival, and evaluate the validity of the expected power model in predicting the likelihood of successful phase III trials.

Based on our experience in advanced breast cancer, an expected power of greater than 0.59 with the Exponential-Gamma model, and of greater than 0.64 with the Weibull-Inverse Gamma model provide a reasonable base to proceed to a phase III study. However, due to limitation of data, we cannot evaluate the validity of the Piece-wise Exponential method, nor provide suggestion of cut-off value of expected power for later phase III studies.

Preface

This dissertation is organized as following:

Chapter 1 describes the background information of advanced breast cancer and cancer drug development. Then the power model based on a Bayesian framework first proposed by Chen, et al is introduced (Chen TT, 2000).

Chapter 2 introduces the methodology of modeling expected power for the subsequent phase III trial, which include the Exponential-Gamma model, the Weibull-Inverse Gamma Model and the Piece-wise Exponential Model.

Chapter 3 presents the detail derivation of distribution for modelling hazards ratio for experimental and control regimen based on the information from historical clinical trials of advanced breast cancer, and computation of expected power for phase III clinical trials with endpoint of progression free survival.

Chapter 4 compares the expected powers calculated from different methods with the actual subsequent phase III results. The assumptions and applications of the three models are also addressed.

Chapter 5 discusses and explores the consistence and inconsistence between calculated expected powers and the significance of actual subsequent phase III study results with endpoints of progression free survival, and evaluates the validity of the expected power model in predicting the likelihood of successful phase III trials. In addition we present the detail derivation of distribution for modelling hazards ratio for experimental and control regimen based on the information from historical clinical trials of advanced breast cancer,

and computation of expected power for phase III clinical trials using endpoint of overall survival.

Chapter 6 summarizes the conclusion and future works.

Acknowledgements

First and foremost, I would like to express my deep appreciation and gratitude to my advisor, Dr. Shih for his guidance and encouragement over the years as I started from an idea to a completed dissertation. His guidance and insight from research method to statistical methodology have made this a thoughtful and rewarding journey.

I would also like to thank my other dissertation committee members for their thoughtful comments and support in my doctoral research. Thanks to Dr. Moore for his thought-provoking suggestions on the expansion of the expected power models. I also greatly appreciate Dr. Wong for her critical thoughts and insight on the breast cancer treatment and oncology drug development. My gratitude is further extended to Dr. Ohman-Strickland who reviewed my dissertation and provided me valuable advices.

In addition, I would like to express my sincere thanks to my other School of Public Health faculty members for their teaching and advices.

I also want to thank my colleagues and supervisors from my full time work for their supporting with a flexible schedule, and helping me to cover the workload.

Last, but not least, I would like to thank my husband for standing beside me throughout the challenge. He has been my encouragement for continuing to improve my knowledge and move my study forward. I also thank my wonderful children: Peter and Katelyn, for always making me smile, being patient, and understanding on those weekends when I was studying and working instead of playing games with them.

TABLE OF CONTENTS

List of Tables

List of Figures

1 Introduction

Breast cancer represents 14.1% of all new cancer cases in the U.S. Breast cancer is one of the most common cancers, and a leading cause of cancer death in women (Howlader N, 2009). According to the Surveillance, Epidemiology and End Results (SEER) Program of the National Cancer Institute (NCI), it is estimated that 226,870 women were diagnosed with and 39,510 women would die of breast cancer in 2012 (Albain KS, 2005).

The overall 5-year relative survival for 2002-2008 from 18 SEER geographic areas was 89.0%. Relative survival measures the survival of the cancer patients in comparison to the general population to estimate the effect of cancer. Table 1-1 shows the stage distribution and 5-year relative survival by stage at diagnosis for 2002-2008 for all races and females with breast cancer.

Table 1-1 Stage distribution and 5-year relative survival by stage at diagnosis for 2002-2008, all races, females

Stage at Diagnosis	Stage Distribution (%)	5-year Relative Survival (%)
Localized (confined to primary site)	60	98.4
Regional (spread to regional lymphnodes)	33	83.9
Distant (cancer has metastasized)	5	23.8
Unknown (unstaged)	2	50.7

Despite advances in the clinical management of breast cancer, patients who develop metastatic disease (cancer that spreads throughout the body) are generally not curable. Patients who experience relapse and develop distant metastasis have a median life expectancy of 2 to 3 years (Albain KS, 2005, 2008; Hortobagyi GN, 2005). New treatments to achieve longer survival while control symptoms and minimizing toxicity are imperative.

A typical oncology drug development starts from a phase I dose escalation trial, then a phase II proof of concept trial. If a promising efficacy is shown in the phase II trial, usually in short term endpoint, such as response rate and progression free survival, a phase III confirmatory trial will be initiated to demonstrate its treatment effect on overall survival or progression free survival. Full marketing approval by the FDA has typically required two positive phase III trials (Hirschfeld S, 2002). However, historically, not all the therapeutic regimens require the phase III testing. Since 1992, there has been a mechanism for accelerated marketing approval of cancer drugs on the basis of surrogate end points, such as response rates in phase II trials (Hirschfeld S, 2002; Shulman SR, 1999). FDA may approve the drug based on an effect on a surrogate endpoint that is reasonably likely to predict clinical benefit for serious or life-threatening illnesses. However, a phase III trial is still needed for several reasons. A drug approved under the accelerated approval regulations is on condition that the manufacturer conducts clinical studies to verify and describe the actual clinical benefit (Hirschfeld S, 2002). In addition the accelerated marketing approval mechanism is only available for an indication where the new therapy comparing with the existing standard or providing therapy where none exists, and not for the second- and third-line indications for the major cancers.

To advance an agent from phase II to pivotal phase III study requires a strategic decision based on many considerations, such as importance of the target mechanism of action, pharmacokinetic properties, market size, competition, surrogates for response, and activity and toxicity in phase I and II trials, etc. With regard to the activity and toxicity in phase I and II trials, Roberts demonstrated the cancer drug development using matrix of biologic activity by toxicity before starting a phase III trial (Roberts Jr TG, 2003). The

biologic activities can be classified from strong evidence of tumor regression to no biologic activity. The toxicity levels include significant and modest toxicity compared with standard care. This is in line with FDA's major criterion for efficacy, which is the evidence of patient benefit with a favorable risk-benefit profile (Hirschfeld S, 2002). The investigational treatments with strong evidence of tumor regression and modest toxicity are termed superstars, while treatments with little biologic activity but with modest or significant toxicity, are termed castaway. The superstar and castaway treatments typically do not enter phase III trial initially (Roberts Jr TG, 2003). The superstars have favorable activity and toxicity profiles compared with available treatment. These treatments become approved on their phase I or phase II clinical results and may not require phase III testing for initial FDA approval. The other two treatment classifications, the incrementalists and trade-offs, with strong or modest evidence of tumor regression but significant toxicity, or modest evidence of tumor regression and modest toxicity, typically require phase III testing to confirm efficacy.

The success rate of oncology Phase III trials is only about 40% despite of promising results in phase II studies (Shulman SR, 1999). The overall phase I to FDA approval (2004-2011) for oncology was 6.7% vs. all other therapeutic areas combined yielding 12.1%. This two times difference is driven primarily by the big drop in Phase III success for oncology trials (Thomas D, 2012). Phase III trials are expensive and time-consuming. In view of this high rate of failure, the advance from Phase II to Phase III requires a strategic decision to increase the success rate.

A power model based on a Bayesian framework was derived and applied by Chen, et al in selecting chemotherapy regimens to phase III trials for extensive-stage small-cell lung

cancer and verified by Freidlin, et al in extensive-stage non-small-cell lung cancer (Chen TT, 2000; Freidlin B, 2003). In this report we adopt their approach and apply it to the oncology drug trial for patients with the advanced breast cancer to provide insight of the likelihood that experimental regimen will provide a clinical benefit such as prolongation of survival compared with standard therapy in subsequent phase III clinical trials to make a phase III go/no go decision.

We first introduce the methodology of modeling expected power for the subsequent phase III trial in Chapter 2 Methods, then in Chapter 3 we present the detail derivation of distribution for modelling hazards ratio for experimental and control regimen based on the information from historical clinical trials of advanced breast cancer, and computation of expected power for phase III clinical trials with endpoint of progression free survival, in Chapter 4 we compare the expected powers calculated from different methods with the actual subsequent phase III results. The assumptions and applications of the model are also addressed. In Chapter 5 Discussion, we study and explore the consistence and inconsistence between calculated expected powers and the significance of actual subsequent phase III study results with endpoints of progression free survival, and evaluate the validity of the expected power model in predicting the likelihood of successful phase III trials. In addition we present the detail derivation of distribution for modelling hazards ratio for experimental and control regimen based on the information from historical clinical trials of advanced breast cancer, and computation of expected power for phase III clinical trials using endpoint of overall survival. In Chapter 6 we summarize the conclusion and future works.

2 Method

In conventional oncology drug development, early phase clinical trials evaluate safety and identify evidence of biological drug activity, such as tumor shrinkage. Endpoints for later phase efficacy studies commonly evaluate whether a drug provide a clinical benefit such as prolongation of survival or an improvement in symptoms.

Response rates are frequently used as a surrogate endpoint in phase II survival trials as the early determination of the antitumor activity of the new therapy. However cytostatic drug may not produce tumor shrinkage as cytotoxic drug. Therefore response rate may not be a valid endpoint to evaluate the effectiveness of this type of new drug.

Overall survival is a universally accepted direct measure of benefit. Overall survival is defined as the time from randomization until death from any cause. It can be easily and precised measured. Survival is considered the most reliable cancer endpoint, and when studies can be conducted to adequately access survival, it is usually the preferred endpoint (Guidance, CDER 2007). However, it usually requires larger studies and much longer follow up time to reach event. There are less statistical significances reported in overall survival than those for progression free survival. The potential reasons could be in many phase III studies, patients are usually allowed to cross over to other cancer treatment after disease progression, which will underestimate the potential statistical significance in comparison of overall survival. Besides, the subsequent cancer therapy can potentially confound survival analysis.

In fact, in majority of phase III studies of advanced breast cancer, the primary endpoint was time to progression (TTP) or progression free survival (PFS) instead of overall

survival (OS). The TTP was usually measured from randomization to the date of documented disease progression, with censoring at last visit date (if alive) or date of death. Progression free survival is defined as the time from randomization until objective tumor progression or death. Both TTP and PFS are not confounded by subsequent therapy. TTP and PFS are endpoints that depend on the time of tumor evaluation and does not necessarily represent the natural time when progression occurs. Irregularities in the frequency or interval of clinical and radiological assessments to document progression and missing data can introduce bias. The randomization trial design is critical to reduce the bias. PFS is not statistically validated as surrogate for survival in all settings; it's not precised measured (Guidance, CDER 2007). However, compared with TTP, PFS includes deaths and thus can be a better correlate to overall survival. Therefore the expected power model is applied to the endpoint of both OS and PFS in advanced breast cancer to evaluate the validity of the model and predict the likelihood of success in the subsequent phase III trial.

Phase III and II clinical trials of patients with only advanced breast cancer were searched. One of the reasons is that patients with advanced breast cancer are always treated with systemic therapies, but patients with early-stage breast cancer may not. In addition, time to death or disease progression information, specifically median survival time or progression free time, is important clinical trial result information to be incorporated into the expected power model. In the phase II clinical trials for patients with early-stage breast cancer, the information is usually unavailable because patients were usually not followed long enough until the occurrence of the corresponding event on half of the patient population could be observed.

We apply the expected power model in clinical trials of patients with advanced breast cancer to explore and validate the model with regard to whether or not based on the model we can better predict the likelihood of success of the future phase III trials with historical clinical data. Literature search of phase III clinical trials for patients with advanced breast cancer is conducted first, we then find out the corresponding historical phase II or phase III trials identified in the later phase III trials. The detail information to identify those trials is provided in the section 2.1 and 2.2. In section 2.3, we introduce the Bayesian conjugate statistical model, and in section 2.4 the Piece-wise Exponential model. By extracting the phase II or phase III data from the literature to build the distributions of hazard with those models, we can retrospectively obtain expected power for the subsequent phase III clinical trials for patients with advance breast cancer. In section 2.3.1.4 and 2.3.2.4, we explain the method used in this report to obtain patient time from the publication.

2.1 Phase III trials

The phase III trials for patients with advanced breast cancer were identified through a search of PubMed with publication dates from 1990 through 2012.

2.2 Phase II and phase III trials

For each of the phase III trials in patients with advanced breast cancer, we attempt to identify the phase II or phase III studies that were referred in the subsequent phase III trials, and based on which the later phase III trials were initiated. The phase II or phase III studies are identified through a review of published references. Information is sought on the dates of the phase II studies, the number of patients, the treatment regimen, the

response rate, the median progression free survival (PFS) time and overall survival (OS) time, number of progression events and number of death at the time of the phase II or phase III data analysis.

2.3 The Bayesian conjugate statistical model

One of the major objectives of phase III trial is to detect a clinical difference of treatment effects of interest. The statistical power is the probability of obtaining such a statistical result. Its calculation is based on the assumed distributions of survivals for the experimental and control regimens. It is critical that the clinical trial is adequately powered so that the actual effect can be identified.

In patients with extensive-stage small cell lung cancer, Chen *et al.* developed a model to provide assistance in selecting chemotherapy regimens from Phase II studies for subsequent study in phase III randomized studies (Chen TT, 2000). The model incorporates the historical data to calculate the expected power for the potential subsequent phase III trial. Both the past phase III experience of the control regimen and the survival data from the preceding observational phase II or phase III studies of the experimental regimen are used to calculate the expected power of the future potential phase III trial. This expected power can then be utilized in making the decision of whether the regimen will go or not go in phase III testing. Chen *et al.* suggested that an expected power of 0.55 or higher can be used as an indication for taking the regimen into phase III testing for the Small Cell Lung Cancer. Freidlin *et al* used Chen *et al*'s model and 0.55 power guidance for the extensive-stage non-small-cell lung cancer and get consistent result that the use of the expected power model provides an important enhancement to the screening of new therapies. They both concluded regimens with an

expected power of >=0.55 may be good candidates for testing in Phase III trials (Chen TT, 2000; Freidlin B, 2003).

We adopt their method with Exponential-Gamma model and apply it to the oncology clinical trials for advanced breast cancer. In addition, we expand the existing Exponential-Gamma model to two new methods, the Weibull-Inverse Gamma model and Piece-wise Exponential model. The survival data from the preceding phase II or III studies of the experimental regimen are used to model the hazard distribution for experimental regimen group. Instead of the overall past phase III experiences of the control regimens as in Chen *et al*'s model, the survival data from preceding phase II or III studies of control regimen are used to model the hazard distribution for the corresponding control group separately for each subsequent phase III trial in this report.

The Bayesian conjugate statistical model is a hybrid Bayesian/frequentist approach where the phase III test is still in the classical (frequentist) framework, and the preceding phase II or phase III studies are used to evaluate the probability of success of such a phase III test by a Bayesian approach.

In Bayesian probability theory, if the posterior distributions are in the same family as the prior probability distribution, the prior and posterior are then called conjugate distribution. The use of conjugate priors allows expression for posterior distribution to be derived in closed form therefore it is an algebraic convenience. The hyperparameters of a conjugate prior distribution often are considered as corresponding to having observed a certain number of pseudo-observations with properties specified by the parameters, which provide intuition how the likelihood function updates a prior distribution. For the likelihood of Exponential distribution with model parameter λ

(rate), the conjugate prior distribution is the Gamma distribution with prior hyperparameters of shape parameter a and scale parameter b. The interpretation of the Gamma hyperparameters is a observations that sum to b. For Weibull distribution with known shape parameter β, the conjugate prior distribution for the unknown Weibull scale parameter θ is Inverse Gamma distribution with prior hyperparameters of shape parameter a and scale parameter b. The interpretation of Inverse-Gamma hyperparameters is a obervations with sum b of the β'th power of each observation (Fink D, 1997.).

The detail methods will be introduced in the following section 2.3.1 the Exponential-Gamma model, 2.3.2 the Weibull-Inverse Gamma model.

2.3.1 The Exponential-Gamma model

In the Exponential-Gamma model, we assume the time to event data follow exponential distribution. The exponential distribution is the probability distribution that describes the time between events in a Poisson process, i.e. a process in which events occur continuously and independently at a constant average rate. It is the continuous analogue of the geometric distribution, and it has the key property of being memoryless.

The probability density function of an exponential distribution is

$$f(x;\lambda)=\begin{cases}\lambda e^{-\lambda x} & x \geq 0\\ 0 & x < 0\end{cases}$$

The mean of an exponentially distributed random variable X with rate parameter λ is given by $E[X]=\frac{1}{\lambda}$

The median is given by $Median[X]=\frac{\ln 2}{\lambda}$

The standard deviation of X is given by $SD[X] = \frac{1}{\lambda}$, which is equal to the mean.

There are three steps in building the Exponential-Gamma model.

2.3.1.1 Step 1: Derivation of Gamma prior distributions of hazard for experimental and control regimens

The model assumes that survival times on the control (c) and experimental (e) arms follow exponential distributions with hazards λ_c and λ_e, respectively. Exponential distribution assumes a constant risk over time, so the hazard is $\lambda(t) = \lambda$ for all time t. The corresponding survival function is $S(t) = \exp^{-\lambda t}$ The density function is $f(t) = \lambda\exp^{-\lambda t}$ for all t.

Under the Bayesian framework, the hazard λ is considered as a random variable. The conjugate prior distribution for exponential distribution of λ is Gamma distribution. Gamma distribution is a two-parameter family of continuous probability distributions with a shape parameter a, and a scale parameter b, while a>0 and b>0.

G (λ| a, b) with density $f(\lambda) = \begin{cases} \frac{\lambda^{a-1}}{\Gamma(a)b^a}\exp\left(-\frac{\lambda}{b}\right) & where\ \lambda > 0 \\ 0 & otherwise \end{cases}$

The mean of λ is given by $E(\lambda) = ab$

The variance of λ is given by $VAR(\lambda) = ab^2$

The hazard λ_c and λ_e are modeled as Gamma distributions with parameters $G(a_c, b_c)$ and $G(a_e, b_e)$, respectively. These parameters are used to formally model the historical information and uncertainty about the effectiveness of the treatment.

For Control group, a gamma distribution $G(a_c, b_c)$ is formed with parameter a_c and b_c on the basis of historical trials. The a_c represents amount of information (the number of

event) on which the prior trial is based and assumed to be a fixed number. The $1/b_c$ represents the total patient-time of survival (until event or censoring) in preceding experience with the treatment.

Experimental group information about failure rate available at the time of planning the phase III trial can also be specified as Gamma probability distribution as $G(a_e, b_e)$. The prior information are specified by setting a_e and b_e so that the mean of distribution λ_e approximately equals to the mean of distribution λ_c but with a given probabilities that median survival time of experimental regimen is greater than median survival time of control regimen a certain length of time.

2.3.1.2 Step 2: Update of Gamma prior distribution of hazard of experimental regimen by incorporating information from trials preceding to the subsequent phase III trials

Prior to conducting a phase III study the above parameters are updated in the following way:

At the time of planning a phase III trial,

$$a_e' = a_e + d_e$$

$$b_e' = 1/(1/b_e + T_e)$$

a_e is the shape parameter of Gamma distribution of hazard for experimental regimen

b_e is the scale parameter of Gamma distribution of hazard for experimental regimen

d_e is number of events observed in the preceding phase II or phase III trials of experimental regimen group

T_e is the sum of the survival times in months (until event or last follow-up) observed in the preceding phase II or phase III trials of experimental regimen group.

a_e' is the shape parameter of Gamma distribution of hazard rate of experimental regimen after updating by the preceding phase II or phase III trials

b_e' is the scale parameter of Gamma distribution of hazard rate for experimental regimen after updating by the preceding phase II or phase III trials.

These simple updating rules are a result of complementarity of the exponential distribution of survival and the Gamma prior distribution of the hazard parameter.

Conjugate prior G (λ; a, b) with density $f(\lambda) = \frac{\lambda^{a-1}}{\Gamma(a)b^a}\exp\left(-\frac{\lambda}{b}\right)$

Likelihood function with censoring $L(\lambda \mid t_1, t_{2,\ldots,} t_n) = \prod_{i=1}^{n} \lambda^{\delta_i} \exp(-\lambda t_i)$

δ = 1 if event and δ = 0 if censor
t_i = time to event or censor for subject i

Posterior distribution

$f(\lambda) \propto L(\lambda)$ G (λ; a, b)

$= \lambda^d \exp(-\lambda \sum_{i=1}^{n} t_i) \frac{\lambda^{a-1}}{\Gamma(a)b^a}\exp\left(-\frac{\lambda}{b}\right)$

$= \frac{\lambda^{a+d-1}}{\Gamma(a)b^a}\exp(-\lambda(\frac{1}{b} + \sum_{i=1}^{n} t_i)$

$\propto$ G (λ; a+d, $1/(\frac{1}{b}$+T))

d = $\sum_{i=1}^{n} \delta_i$ is total number of event

T = $\sum_{i=1}^{n} t_i$ is sum of survival times until event or last follow-up observed

a+d is total number of event, $\frac{1}{b}$ + T is total survival times until event or last follow-up after updating the prior distribution.

2.3.1.3 Step 3: Compute the expected power for a subsequent phase III trial

The power of a statistical test is the probability that the test will reject the null hypothesis if the null hypothesis is false. In the clinical trials of advanced breast cancer, assume the exponential distribution for the survival distribution of experimental and control regimens, the hazard ratio is constant as $\lambda_e^{'}/\lambda_c$ over time.

Start with basic sample size calculation $\Delta = Z_{1-\alpha/2}\sqrt{V_0} + Z_{1-\beta}\sqrt{V_1}$, where Δ=log(λ_c)-log($\lambda_{e'}$), V_0 is variance under null hypothesis, and V_1 is variance under alternative null hypothesis. Assume hazards are proportional, randomization is even, and censoring distributions are same for different treatment groups, the power for specific hazard values of $\lambda_e^{'}$ and λ_c can be approximated by

$$Power = \phi^{-1}\left(\frac{\log\left(\frac{\lambda_c}{\lambda_e^{'}}\right)}{\sqrt{\frac{4}{d}}} - z_{1-\alpha/2}\right)$$

ϕ^{-1} is the inverse of the standard normal distribution

$z_{1-\alpha/2}$ is the upper $100(1-\alpha/2)$ percentile of the standard normal distribution

d is the total number of events expected in the subsequent phase III trial at the time of analysis.

For example, when d =256 as the total number of events to be observed in a phase III trial was used, the statistical power for detecting a 50% increase in median survival with a two-sized 5% significance level is 90%; similarly, with the same two-sized 5%

significance level and statistical power of 90%, we need 372 events to detect a 40% increase in median survival, and 611 events to detect a 30% increase, and 1264 events to detect a 20% increase in median survival.

On the other hand, instead of setting the $\lambda e^{'}$ and λc as fixed value, we want to incorporate prior survival information and consider λe' and λc as random variables with Gamma distribution as

$\lambda_c \sim G(a_c, b_c)$
$\lambda_e' \sim G(a_e', b_e')$

We model the hazard of event for the control and experimental regimen as two different Gamma distributions, incorporating the number of events and sum of time to censor or event into the model parameters. If consider the Gamma distribution for experimental regimen as prior distribution, with the likelihood function of hazard, the posterior Gamma distribution for hazard of experimental regimen can be modeled as,

$$p(\lambda_e' \mid a_e, b_e, t_1, ..., t_n) \propto L(\lambda_e \mid t_1, ..., t_n).P(\lambda_e \mid a_e, b_e)$$

while t is time to censor or event for subject i, i from 0 to n, $T_e = \sum_1^n t_i$ and a_e, b_e are constant.

Average this quantity with regard to the Gamma distribution $G(a_e', b_e')$ of λ_e' and $G(a_c, b_c)$ of λ_c described above, we get the expected power as

$$E(Power) = E_{\lambda_c}\left(E_{\lambda_e}(\phi^{-1}(\frac{\log(\frac{\lambda_c}{\lambda_e'})}{\sqrt{\frac{4}{d}}} - z_{1-\alpha/2})) \right)$$

Apply the Monte Carlo integration, which is $E[g(x)] = \int_a^b g(x).p(x)dx \approx \frac{1}{N}\sum_1^N g(x_i)$

The averages can be approximated with 100,000 samples.

The following steps will be conducted:

(i) Draw n samples from distribution G(a_c, b_c) for control regimen;

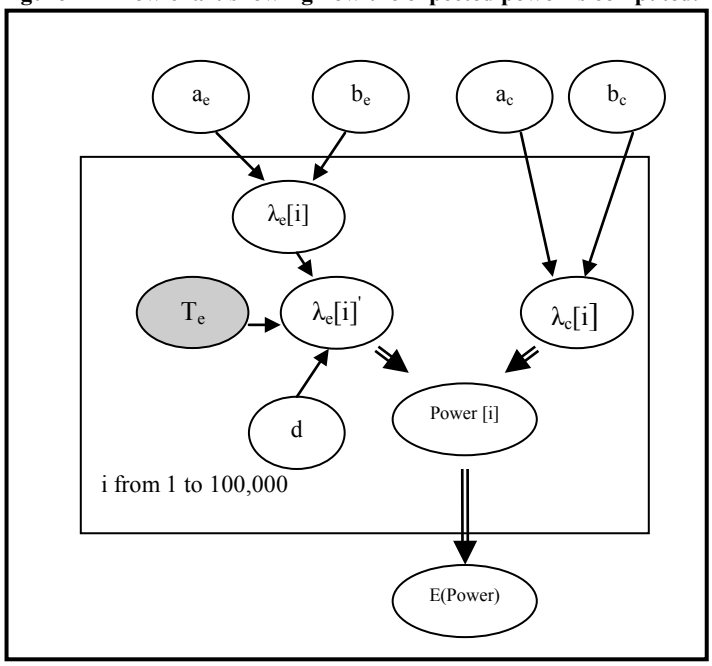
(ii) Draw n samples from prior distribution G(a_e, b_e) for experimental regimen;

(iii) Update the pair of a_e and b_e and generate n pairs of λ_{ei}' with distribution of Gamma(a_e', b_{ei}') with i from 1~n, a_e'= a_e+d, b_e'= 1/($\frac{1}{b_e}$ + T_e), a_e, b_e, d and T_e are all fixed.

(iv) Pick n randomly distributed points (λ_{e1},λ_{c1}), (λ_{e2},λ_{c2}), (λ_{e3},λ_{c3}),…., (λ_{en},λ_{cn}) from f (λ_{ei}') ~ Gamma(a_e', b_e') and f (λ_{ci}) ~ Gamma(a_c, b_c), where a_c, b_c , a_e' and b_e' are fixed;

(v) Determine the average values of the function of power regarding to variable λ_e' and λ_c.

Figure 2-1 Flow chart showing how the expected power is computed.



Note: T_e and d are derived based on experimental regimen in preceding trials.
Power and E(Power) are power and expected power for subsequent phase III trial.

Above Figure 2-1 is the flow chart showing how the expected power is computed.

2.3.1.4 Method to derive the patient-time

We indirectly derive patient-time to event or last follow-up (inverse of b) based on (1) d number of event (information from literature) and (2) Median survival time (information from literature) by setting shape parameter a of Gamma distribution to d and the mean of the resulting Gamma distribution corresponding to the median survival time.

The patient time (T) and number of events (d) are used to update the prior Gamma distribution (a, b) of hazard rate, in order to get posterior Gamma distribution [a+d, $1/(\frac{1}{b}$ +T)] for experiment regimen. Letter a is shape parameter, b is scale parameter of prior Gamma distribution of experimental regimen.

2.3.2 The Weibull-Inverse Gamma Model

When modelling the survival time as exponential distribution, we assume the hazard rate is constant over time. However, many times the hazard rate of event is not unchanging during the disease procedure, it could vary, decrease or increase over time. Weibull distribution is a continuous probability distribution with two parameters, shape parameter β, and scale parameter θ. The second parameter of Weibull makes Weibull distribution more flexible in modelling time to event data in comparison with Exponential distribution.

The Probability density function of a Weibull random variable is:

$$f(x \mid \beta,\theta)=\begin{cases}\frac{\beta}{\theta}x^{\beta-1}\exp\left(-\frac{x^{\beta}}{\theta}\right) & where\ x\geq 0\\ 0 & otherwise\end{cases}$$

Where $\beta > 0$ and $\theta > 0$.

The mean of a Weibull distributed random variable X with shape parameter β, and scale parameter θ is given by $E[X] = \theta\,\Gamma(1+\frac{1}{\beta})$

The median is given by $Median\ [X] = \theta\,(\ln(2))^{\frac{1}{\beta}}$

The variance of X is given by $VAR[X] = \theta^{2}\,[\Gamma(1+\frac{2}{\beta}) - (\Gamma(1+\frac{1}{\beta}))^{2}]$

When the survival time is assumed to be Weibull distributed, the corresponding survival function is $S(t) = \exp\left(-\frac{t^{\beta}}{\theta}\right)$ for $t \geq 0$. The hazard function is $h(t) = \frac{f(t)}{S(t)} = \frac{\beta}{\theta} t^{\beta-1}$, which is a function of both shape parameter β and scale parameter θ.

It can be seen from the hazard function, the survival rate is proportional to the time of the power of β-1. So that this shape parameter β can be interpreted directly as follows:

β < 1 indicates the failure rate decreases over time.

β = 1 indicates the failure rate is constant over time.

β > 1 indicates the failure rate increases with time.

The Weibull distribution is related to Exponential distribution as it becomes to an Exponential distribution when shape parameter β equals to 1. In the framework of Bayesian approach, the Exponential distribution has its conjugate prior distribution as Gamma distribution. However, when both parameters of the Weibull distribution are considered as random variables, the result obtained by Soland stated that the Weibull distribution does not have a conjugate continuous joint prior distribution (Soland R,

1966). Therefore, in this report, the shape parameter of Weibull distribution of time to event is considered as fixed instead of model it with a distribution. The value of Weibull shape parameter is derived based on information extracted from the Kaplan Meier curve in the literature.

We apply the Weibull-Inverse Gamma model to fit the time to event data and obtain distribution of hazard. The model assumes that survival times on the control (c) and experimental (e) arms follow Weibull distributions with shape parameter β_c and β_e, and scale parameter θ_c and θ_e, respectively. There are three steps in building the Weibull-Inverse Gamma model when the shape parameter of Weibull distribution is considered as a fixed constant.

2.3.2.1 Step 1: Derivation of Inverse Gamma prior distributions of Weibull scale parameter for experimental and control regimens

The scale parameter of Weibull distribution is considered as a random variable in this method while the shape parameter is assumed as a fixed constant derived from Kaplan Meier curve. Under the Bayesian framework, the conjugate prior distribution for scale parameter θ is Inverse Gamma distribution. Inverse Gamma distribution is a two-parameter family of continuous probability distributions with shape parameter a, and scale parameter b, while a>0 and b>0.

IG (θ|a, b) with density $f(\theta) = \begin{cases} \frac{b^a}{\Gamma(a)\theta^{a+1}} \exp\left(-\frac{b}{\theta}\right) & where\ \theta > 0 \\ 0 & otherwise \end{cases}$

The mean of an Inverse Gamma distributed random variable θ with shape parameter a, and scale parameter b is given by $E(\theta) = \frac{b}{a-1}$

The variance of X is given by $VAR(\theta) = \frac{b^2}{(a-1)^2(a-2)}$

The Weibull scale parameter θ_c and θ_e for patients with control and experimental regimens are modeled as Inverse Gamma distributions with parameters IG(a_c, b_c) and IG(a_e, b_e), respectively. These parameters are used to formally model the distribution of hazard by incorporating the available information on the treatment effects from previous experiences and taking into account the uncertainty about the effectiveness of the treatment the same time.

Summary statistics data from previous clinical trials which include the corresponding experimental or control regimen were available for use as historical data. For Control group, an inverse gamma prior distribution IG(a_c, b_c) is formed with shape parameter a_c and scale parameter b_c on the basis of historical trials data. The a_c represents amount of information (the number of event) on which the prior trial is based and assumed to be a fixed number. The b_c represents the total patient-time of survival (until event or censoring) raised to the βth power (Weibull shape parameter) from preceding experience with the treatment.

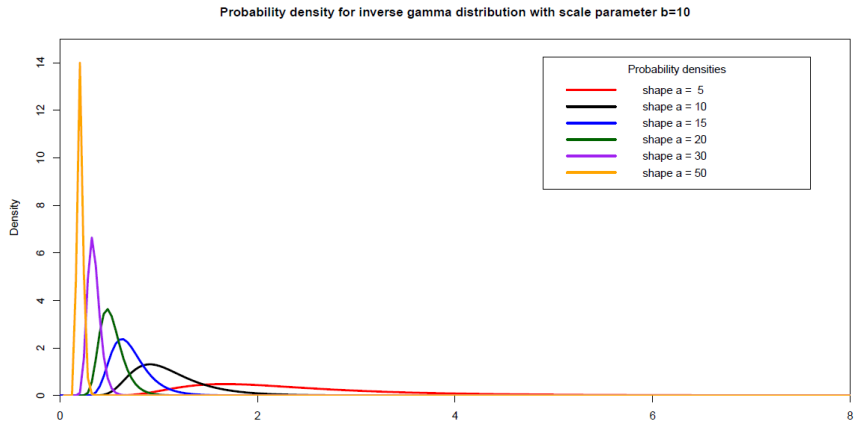
For experimental group information about failure rate available at the time of planning the phase III trial can also be specified as inverse gamma probability distribution as IG(a_e, b_e). Instead of setting the mean of hazard λ_e distribution equals to the mean of hazard λ_c distribution to derive prior Gamma distribution shape and scale parameters a_e and b_e in the Gamma-Exponential model, we let the variance of the prior Inverse Gamma distribution for Weibull scale parameter θ_e equals to the variance of the prior Inverse Gamma distribution for Weibull scale parameter θ_c in order to define the prior Inverse

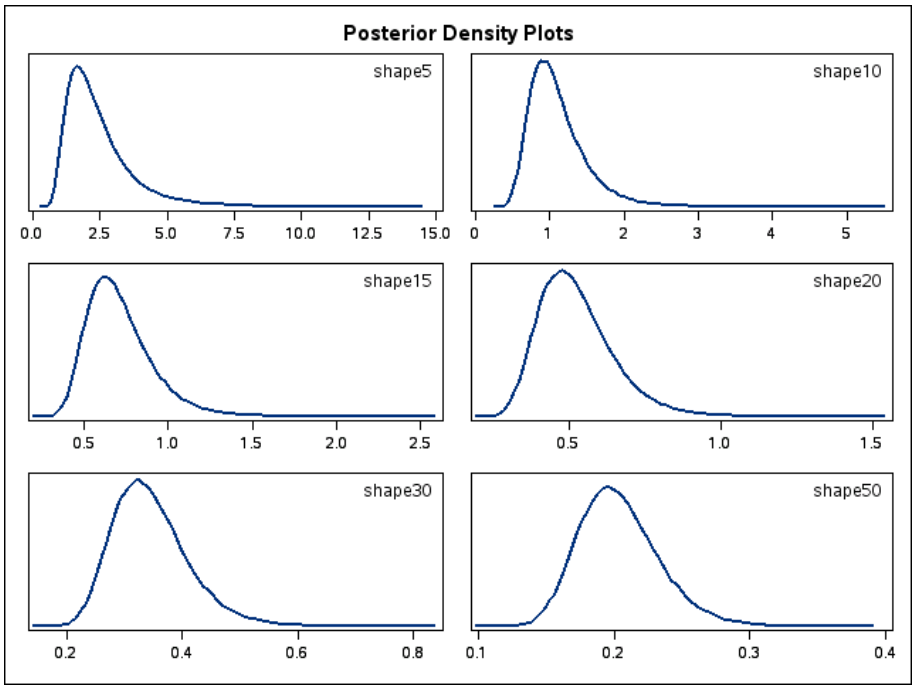
Gamma distribution parameter a_e and b_e in the Inverse Gamma-Weibull model. A given probabilities that median survival time for patients with experimental regimen is greater than median survival time for patients with control regimen a certain length of time is specified in order to derive the prior Inverse Gamma distribution parameter a_e and b_e.

We modify the algorithm used to derive the Inverse Gamma prior distribution IG (a_e, b_e) for experimental group based on the Inverse Gamma prior distribution IG (a_c, b_c) for control group from the equal means in the Exponential-Gamma model to the equal variances. The Inverse Gamma distribution is right skewed, and the varying shape parameter has an impact on the shape of inverse gamma distribution, as shape parameter a get large, the inverse gamma distribution gradually resembles a normal distribution.

Figure 2-2 is the plot of the probability densities for the inverse gamma distribution with different values of shape parameter a and constant scale parameter of 10.

Figure 2-2 Probability densities of the Inverse Gamma distribution with different values of shape parameter a and constant scale parameter of 10

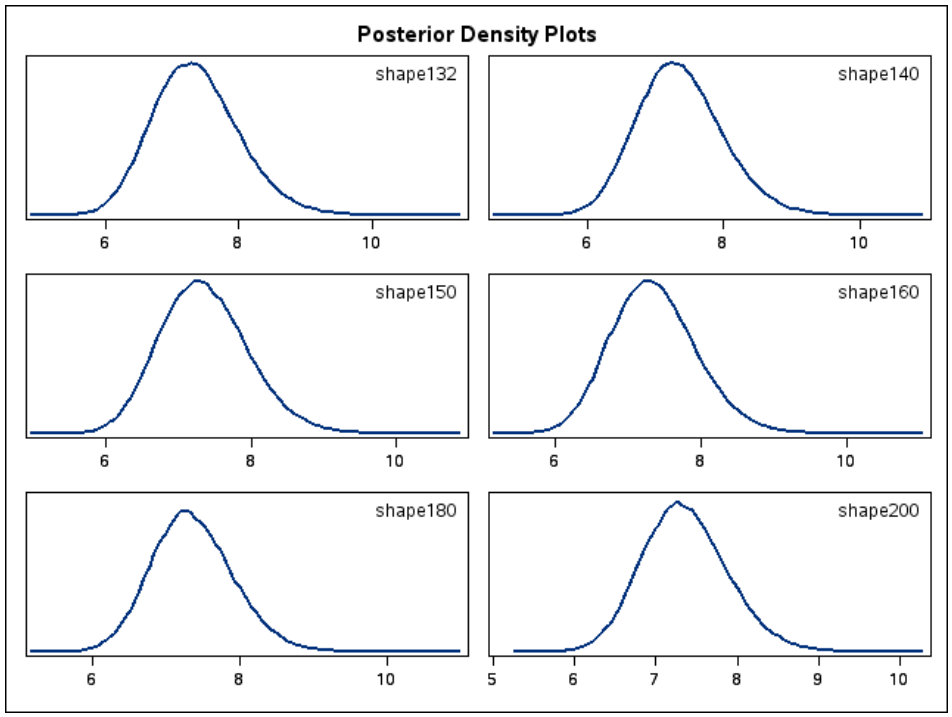




From the probability density function graph, we can visualize the impact of the inverse gamma shape parameter a on the shape of the inverse gamma distribution. When the shape parameter alpha is 30, the probability density function approximates to a normal distribution. As a_c is set to represents the number of event in the control group, the values of a_c are within a range of 103 to 369 according to table 3-3. Because the normal distribution is symmetric around the mean, if the means of two normal distributions are set to be equal, the area under curve for both normal distributions will balance around the same mean. Therefore we cannot set the prior Inverse Gamma distribution parameter a_e and b_e to meet both criteria: the mean of the distribution of Weibull scale parameter θ_e approximately equals to the mean of the distribution of Weibull distribution scale parameter θ_c, with given probabilities that median survival time of experimental regimen is greater than median survival time of control regimen a certain length of time.

Several inverse gamma probability density graphs are plotted from SAS procedure MCMC with the same mean of 7.36, but with different values of shape parameter from 132 to 200. From the graphs, all the density function plots are close to a bell shape and symmetric around the same mean of 7.36. Thus we cannot find one plot that overall density is higher than any other density plot with certain percentage of chances.

Figure 2-3 Probability densities of the Inverse Gamma distribution with different values of shape parameter a and mean of 7.36



2.3.2.2 Step 2: Update of Inverse Gamma prior distribution of Weibull scale parametr of experimental regimen by incorporating information from trials preceding to the subsequent phase III trials

At the time of planning a phase III trial, the prior inverse gamma shape and scale parameter are updated in the following way:

$$a_e' = a_e + d_e$$

$$b_e' = b_e + \sum_{i=1}^{n} t_i^{\beta}$$

a_e is shape parameter in inverse gamma prior distribution of Weibull scale parameter for experimental regimen

b_e is scale parameter in inverse gamma prior distribution of Weibull scale parameter for experimental regimen

d_e is number of events of interest observed in the preceding phase II or phase III trials of experimental regimen group.

t_i is the survival times in months (until event or last follow-up) observed in the preceding phase II or phase III trials of experimental regimen group.

β is the known Weibull shape parameter, which is a fixed constant

n is the sample size in the historical trials.

a_e' is the shape parameter in inverse gamma posterior distribution of Weibull scale parameter of experimental regimen after updating by the information from preceding phase II or phase III trials.

b_e' is the scale parameter in inverse gamma posterior distribution of Weibull scale parameter of experimental regimen after updating by the information from preceding phase II or phase III trials.

These simple updating rules are a result of complementarity of the Weibull distribution of survival time with known shape parameter and the Inverse Gamma prior distribution of the Weibull scale parameter.

Conjugate prior IG (θ |a, b) with density

$$f(\theta)=\begin{cases}\dfrac{b^{a}}{\Gamma(a)\theta^{a+1}}\exp\left(-\dfrac{b}{\theta}\right) & where\ \theta>0, a>0\ and\ b>0\\ 0 & otherwise\end{cases}$$

Suppose that data $t_1,\ldots t_n$ are independently and identically distributed from Weibull process, where the scale parameter θ is unknown and the shape parameter β is known.

The likelihood function with censoring $L(\theta \mid t_{1,} t_{2,\ldots,} t_n, \beta) \propto \prod_{i=1}^{n} \theta^{-\delta_i} \exp(-\dfrac{\sum_{i=1}^{n} t_i^{\beta}}{\theta})$

δ = 1 if event and δ = 0 if censor
t_i = time to event or censor for subject i

The sufficient statistics are n, the number of data points, and $\sum_{i=1}^{n} t_i^{\beta}$ the sum of the data raised to the βth power. As a function of θ, the equation above is proportional to an inverted Gamma distribution of θ. The conjugate prior, IG (θ|*a*, b) is an inverted Gamma distribution with hyperparameters, shape parameter a, and scale parameter b, where a, b>0.

IG (θ|a, b) with density f(θ|a, b) = $\begin{cases}\dfrac{b^{a}}{\Gamma(a)\theta^{a+1}}\exp\left(-\dfrac{b}{\theta}\right) & where\ \theta>0\\ 0 & otherwise\end{cases}$

The posterior distribution f(θ|a',b') is an inverse Gamma distribution with hyperparameters a' = a + n, and b' = b + $\sum_{i=1}^{n} t_i^{\beta}$

$$f(\theta \mid a', b') \propto L(\theta \mid t_{1,} t_{2,\ldots,} t_n, \beta)\, \mathrm{f}\,(\theta \mid \mathrm{a, b})$$

$$= \prod_{i=1}^{n} \theta^{-\delta_i} \exp\left(-\frac{\sum_{i=1}^{n} t_i^{\beta}}{\theta}\right) \frac{b^a}{\Gamma(a)\theta^{a+1}} \exp\left(-\frac{b}{\theta}\right)$$

$$= \frac{b^a}{\Gamma(a)} \prod_{i=1}^{n} \theta^{-\delta_i-(a+1)} \exp\left(-\frac{b+\sum_{i=1}^{n} t_i^{\beta}}{\theta}\right)$$

$$= \frac{b^a}{\Gamma(a)} \theta^{-d-(a+1)} \exp\left(-\frac{b+\sum_{i=1}^{n} t_i^{\beta}}{\theta}\right)$$

$$\propto \text{IG}\left(\theta;\ a+d,\ b+\sum_{i=1}^{n} t_i^{\beta}\right)$$

$d = \sum_{i=1}^{n} \delta_i$ is total number of event

t_i is the survival times in months (until event or censoring) observed in the preceding phase II or phase III trials of experimental regimen group.

β is the known Weibull shape parameter, which is a fixed constant.

n is the sample size in the historical trials.

2.3.2.3 Step 3: Compute the expected power for a subsequent phase III trial

The method introduced in section 2.3.1.3 Compute the expected power for a subsequent phase III trial in the Exponential-Gamma model can also be applied to Weibull-Inverse Gamma model. Assume hazards are proportional, randomization is even, and censoring distributions are same for different treatment groups, the power for hazard λ_e' and λ_c can be approximated by

$$Power = \phi^{-1}\left(\frac{\log\left(\frac{\lambda_c}{\lambda_e'}\right)}{\sqrt{\frac{4}{d}}} - z_{1-\alpha/2} \right)$$

ϕ^{-1} is the inverse of the standard normal distribution

$z_{1-\alpha/2}$ is the upper $100(1-\alpha/2)$ percentile of the standard normal distribution

d is the total number of events expected in the subsequent phase III trial at the time of analysis.

While in the case of exponential distributed survival time, the hazard $h(t_i) = \lambda$ is constant over time, if survival time is distributed in Weibull process the hazard $h(t_i)$ equals to $\frac{\beta}{\theta} t_i^{\beta-1}$, which could change over time.

Under the proportional hazards model, the hazard of event at time t for the ith individual is given by $h_i(t)=\exp^{\beta x_i} h_0(t)$, where x_i is the value of x for the i^{th} individual. Consequently, the hazard at time t for an individual in group A is $h_0(t)$, and that for an individual in group B is $\varphi\, h_0(t)$, where $\varphi = \exp(\beta)$. The quantity of β is then the logarithm of the ratio of the hazard for an individual in group B, to that of an individual in group A.

It's assumed that the survival times for the individuals in group A have a Weibull distribution with scale parameter θ and shape parameter β. Using equation above, the hazard function for individuals in this group is $h_0(t)$, where $h_0(t)=\frac{\beta}{\theta} t^{\beta-1}$, the hazard function for those in group B is $\varphi\ \frac{\beta}{\theta} t^{\beta-1}$, that is the hazard function for a Weibull distribution with scale parameter $\frac{\theta}{\varphi}$ and shape parameter β. Therefore if survival times of

individuals in one group have a Weibull distribution with shape parameter β, and the hazard of event at time t for an individuals in the second group is proportional to that of an individuals in the first, the survival times of those in the second group will also have a Weibull distribution with shape parameter β.

We have $h_c(t)=\frac{\beta}{\theta_c}t^{\beta-1}$, $h_e(t)=\frac{\beta}{\theta_e}t^{\beta-1}$, thus, for the hazard ratio at any time point between an individual in the control treatment and an individual in active treatment group is constant and equal to $\frac{\theta_e}{\theta_c}$.

We model the hazard ratio of event for the control and experimental regimen as ratio of two Weibull scale parameters, each with an Inverse Gamma distribution. The Inverse Gamma shape and scale parameters are derived by incorporating the number of events and the sum of the time to event or censor raised to the βth power into the model parameters. If consider the Inverse Gamma distribution of Weibull scale parameter for experimental regimen as prior distribution, with the likelihood function of scale parameter, the posterior Inverse Gamma distribution for Weibull scale parameter can be modeled as,

$$p(\theta_e'|a_e,b_e,\beta,t_1,...,t_n) \propto L(\theta_e|t_1,...,t_n,\beta).P(\theta_e|a_e,b_e)$$

while t is time to censoring or event for subject i, i from 1 to n, $T_e=\sum_{i=1}^{n}t_i^{\beta}$ and a_e, b_e are constant.

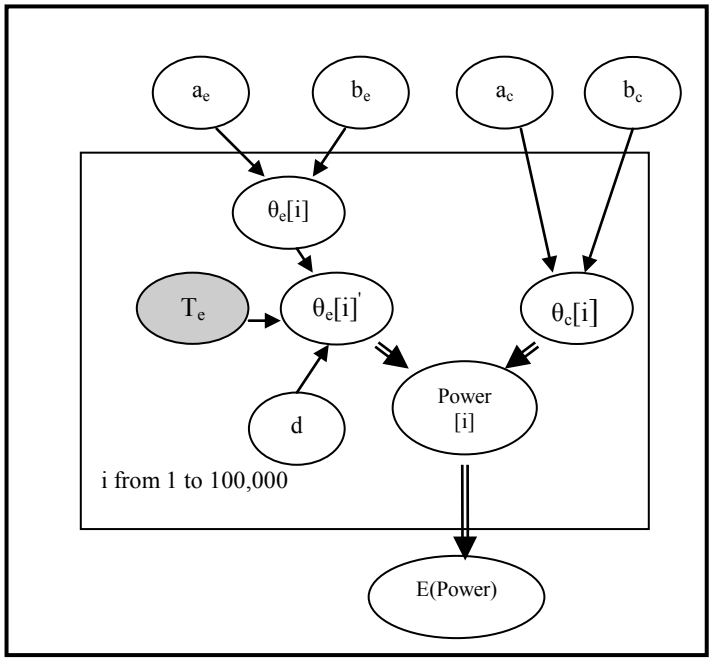
Average this quantity with regard to the prior Inverse Gamma distribution IG(a_e', b_e') of Weibull scale parameter for experimental group and prior IG(a_c, b_c) of Weibull scale parameter for control group described above, we get the expected power as

$$E(Power) = E_{\lambda_c}\left(E_{\lambda_e}(\phi^{-1}(\frac{\log(\frac{\theta_e}{\theta_c})}{\sqrt{\frac{4}{d}}} - z_{1-\alpha/2})) \right)$$

Apply the Monte Carlo integration, which is $E[g(x)] = \int_a^b g(x).p(x)dx \approx \frac{1}{N}\sum_1^N g(x_i)$

The averages can be approximated with 100,000 samples.

Figure 2-4 Flow chart showing how the expected power is computed



Note: T_e and d are derived based on experimental regimen in preceding trials.
Power and E(Power) are power and expected power for subsequent phase III trial.

Above Figure 2-4 is the flow chart showing how the expected power is computed.

(i) Draw n samples from distribution Inverse Gamma(a_c, b_c) for control regimen;

(ii) Draw n samples from prior distribution Inverse Gamma(a_e, b_e) for experimental regimen;

(iii) Update the pair of a_e and b_e and generate n pairs of θ_{ei}' with distribution of Inverse Gamma(a_e', b_{ei}') with i from 1~n, $a_e' = a_e + d$, $b_e' = b + \sum_{i=1}^{n} t_i^{\beta}$, a_e, b_e, d and β are all fixed.

(iv) Pick n randomly distributed points (θ_{e1},θ_{c1}), (θ_{e2}, θ_{c2}), (θ_{e3}, θ_{c3}),…., (θ_{en}, θ_{cn}) from f (θ_{ei}') ~ Inverse Gamma(a_e', b_e') and f (θ_{ci}) ~ Inverse Gamma(a_c, b_c), where a_c, b_c , a_e' and $b_e'$are fixed;

(v) Determine the average values of the function of power regarding to variable θ_e' and θ_c.

2.3.2.4 Method to derive the patient-time raised to the β^{th} power

We indirectly derive patient-time to event or last follow-up raised to the β^{th} power based on (1) d Number of event (information from literature); (2) Median survival time (information from literature); (3) The Weibull shape parameter β by setting shape parameter of Inverse Gamma distribution to number of events d and the mean of the resulting Inverse Gamma distribution corresponding to the median of Weibull distributed time.

The patient time raised to the β^{th} power (T) and number of events (d) are used to update the prior Inverse Gamma distribution (a, b) of Weibull scale parameter, in order to get posterior Inverse Gamma distribution [a+d, $b+\sum_{i=1}^{n} t_i^{\beta}$] for experiment regimen. Letter a is shape parameter, b is scale parameter of prior Inverse Gamma distribution of experimental regimen.

2.4 Piece-wise Exponential Model

In addition to the Bayesian conjugate statistical models, we also propose to conduct a piecewise exponential representation of the original survival data. The hazard regression in Piece-wise Exponential model can be linked with estimation schemes based on the Poisson likelihood.

Parametric models for time to event data are not flexible to accommodate the various patterns of non-constant hazards due to the restriction of the distribution assumptions. Piece-wise Exponential approach can be used to approximate nonparametric models to better fit the real time to event data, and retain assumption of parametric distribution of data. It is a flexible, semi-parametric strategy.

Consider partitioning the whole duration in the time to event data into J intervals with cut points $0 = \tau_0 < \tau_1 <\ldots< \tau_J = \infty$. We will define the j-th interval as $[\tau_{J-1}, \tau_J)$, extending from the (j - 1) boundary to the j-th and including the former but not the later.

We will then assume that the baseline hazard is constant within each interval, so that

$$\lambda_0(t) = \lambda_j \text{ for t in } [\tau_{J-1}, \tau_J).$$

In another word, the hazard rate is allowed to differ in different time intervals but is assumed to be constant within any given time interval. The basic idea of the Piece-wise Exponential model is therefore that the proportional hazard assumption holds at least over short periods of time such that all time-varying effects can be treated as piecewise constant. Thus, we model the baseline hazard $\lambda_0(t)$ using J parameters $\lambda_1,\ldots, \lambda_J$, each representing the risk for the reference group (or individual) in one particular interval.

Since the risk is assumed to be piece-wise constant, the corresponding survival function is often called a piece-wise exponential.

One of the key issues with the Piece-wise Exponential model involves careful choice of the cut points and appropriate number of time intervals to be used. The number of time intervals is something up to analysts to use their discretion. Although any number of time periods can be chosen, it is important to recognize that there is always a tradeoff to be made. If one chooses a large number of time periods, then we get a better approximation of the unknown baseline hazard but we have to estimate a larger number of coefficients and this may cause problems. Alternatively, if one chooses a small number of time periods, then there will be fewer estimation problems but the approximation of the baseline hazard will be worse. The bottom line is when choosing the number of time intervals, there should be time of event of interest or censoring falls into each of the divided intervals. Otherwise the estimate may not be reasonable.

Holford (1980) and Laird and Oliver (1981) introduced that log-linear models for the cell means of contingency tables with Poisson data are exactly equivalent to log-linear hazard models for survival data, when specify (a) a piecewise exponential survival distribution and (b) categorical covariates. The likelihoods for Poisson contingency table data and piecewise exponential survival data are also equivalent. This last equivalence has the important implication that the two likelihoods can be used interchangeably for deriving maximum likelihood estimates, their asymptotic variances, and for calculating likelihood ratio statistics and their asymptotic sampling distributions.

Assume that we observe t_i, the total time lived by the i-th individual, and d_i, a death indicator that takes the value one if the individual died and zero otherwise. We will now

define analogous measures for each interval that individual i goes through. During the process, some pseudo-observations are created for each combination of individual and interval.

First we create measures of exposure. Let t_{ij} denote the time lived by the i-th individual in the j-th interval, that is, between τ_{j-1} and τ_j. If the individual lived beyond the end of the interval, so that $t_i > \tau_j$, then the time lived in the interval equals the width of the interval and $t_{ij} = \tau_j - \tau_{j-1}$. If the individual died or was censored in the interval, i.e. if $\tau_{j-1} < t_i < \tau_j$, then the time lived in the interval is $t_{ij} = t_i - \tau_{j-1}$, the difference between the total time lived and the lower boundary of the interval.

Let d_{ij} take the value one if individual i dies in interval j and zero otherwise. Let j(i) indicate the interval where t_i falls, i.e. the interval where individual i died or was censored. Functional notation is used to emphasize that this interval will vary from one individual to another. If t_i falls in interval j(i), say, then d_{ij} must be zero for all $j < j(i)$ (i.e. all prior intervals) and will equal d_i for $j = j(i)$, (i.e. the interval where individual i was last seen).

Then, the Piece-wise Exponential model may be fitted to data by treating the death indicators d_{ij}'s as if they were independent Poisson observations with means

$$\mu_{ij} = t_{ij} \lambda_{ij}$$

where t_{ij} is the exposure time as defined above and λ_{ij} is the hazard for individual i in interval j. Taking logs in this expression, and recalling that the hazard rates satisfy the proportional hazards model, we obtain

$$\log\mu_{ij} = \log t_{ij} + \alpha_j + x_i'\beta$$

where $\alpha_j = \log \lambda_j$.

Thus, the Piece-wise Exponential proportional hazards model is equivalent to a Poisson log-linear model for the pseudo observations, one for each combination of individual and interval, where the death indicator is the response and the log of exposure time enters as an offset.

3 Derivation of distribution for modelling hazard ratio and computation of expected power with endpoint of progression free survival

Literatures of phase III trials with patients in advanced breast cancer are searched first, then the previous phase II or phase III trials referred in the subsequent phase III trials which led to the phase III trials are identified. The power model is built upon the survival information extracted from preceding phase II or phase III clinical trials. The retrospectively calculated expected powers are then compared with the actual phase III trial results to evaluate that the validity of the expected power model in predicting the likelihood of successful phase III trials.

3.1 Phase III trials

A search of randomized phase III clinical trial for advanced breast cancer from PubMed reveals that there are 57 trials with publication date from 1990 to 2012, in which 12 trials from 1990 to 1999, and 45 from 2000 to 2012. The search from PubMed uses species of humans and article types of clinical trial and randomized controlled trial with key words of metastatic breast cancer, advanced breast cancer, chemotherapy, endocrine therapy and phase III. Trials with cross-over design in the nature of design, interim analysis, early terminated or without reaching median time for progression free survival, or overall survival, or with objective solely for evaluating safety or quality of life, comparison among same regimen but different formulations or dosage levels are excluded.

The search above showed that only 21 of 57 trials have progression free survival as one of the efficacy end points, in which 18 trials set progression free survival as primary efficacy endpoint. Among the 21 studies, the experimental arms of 9 trials show statistically significant progression free survival advantages in comparison with the

corresponding control arms. The rate of success was 43% (95% confidence interval [CI], 21% to 65%). The median progression free survival time differences comparing experimental regimen with control regimen in those successful trials ranged from 1.6 to 6.5 months, with median difference of median progression free survival time of 2.3 months. This search and review also showed that in majority of phase III randomized studies of advanced breast cancer the experimental regimen is not proved to be superior to the control regimen.

Among the nine studies with successful progression survival results, the hazard ratios of progression survival and 95% confidence interval were published for 8 studies. We integrate the different treatment effects from the 8 studies. Since the sample size, patient selection and interventions are different for each study, we assume that there is a distribution of treatment effects to incorporate the individual differences, while still considered the effects are related to each other. Random effect model in the Meta data analysis is applied to the analysis of the hazard ratios of progression free survival. Comprehensive Meta-Analysis (CMA) 2.0 is used to compute the effect size and the variance (Borenstein M, 2009). The summary effect size measured in hazard ratio is 0.65 with 95% confidence interval of [0.55, 0.77], which corresponds to 54% increase in median progression free survival time. According to the power computation section 2.3.1.3 formula, the subsequent study need to observe 227 events in order to be adequately powered (90%) to detect the 54% increase in median survival time.

Figure 3-2 shows flow chart of the searched the 57 phase III clinical trials and the information extracted from those trials.

Figure 3-1 Impact of treatment on disease progression by study hazard ratio and 95% CL

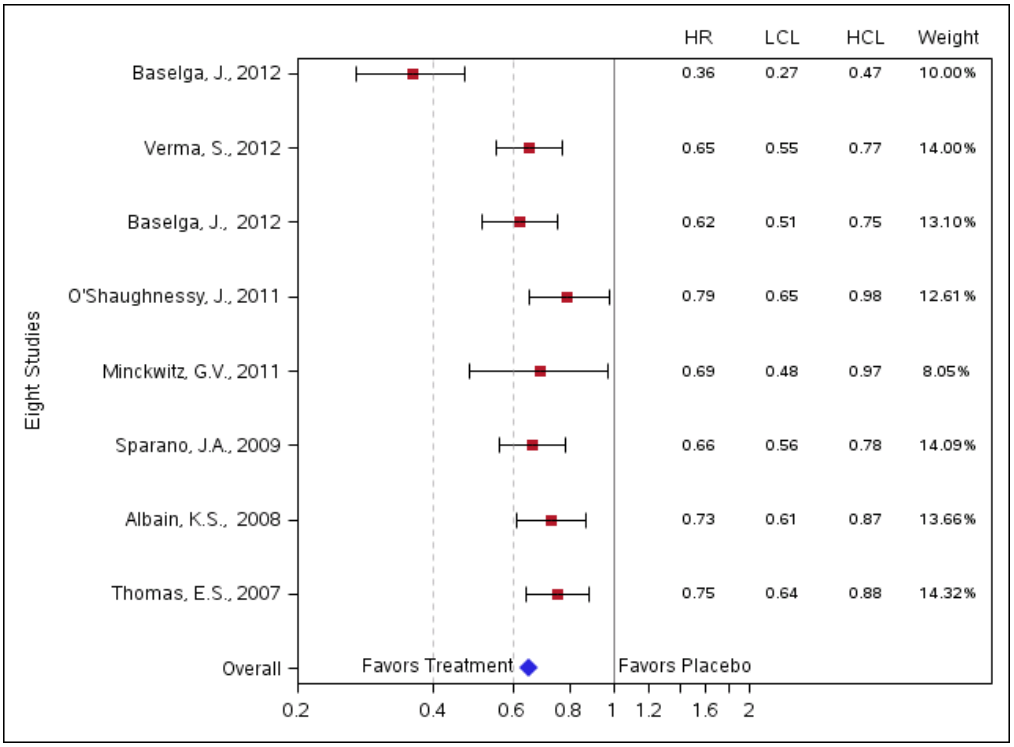
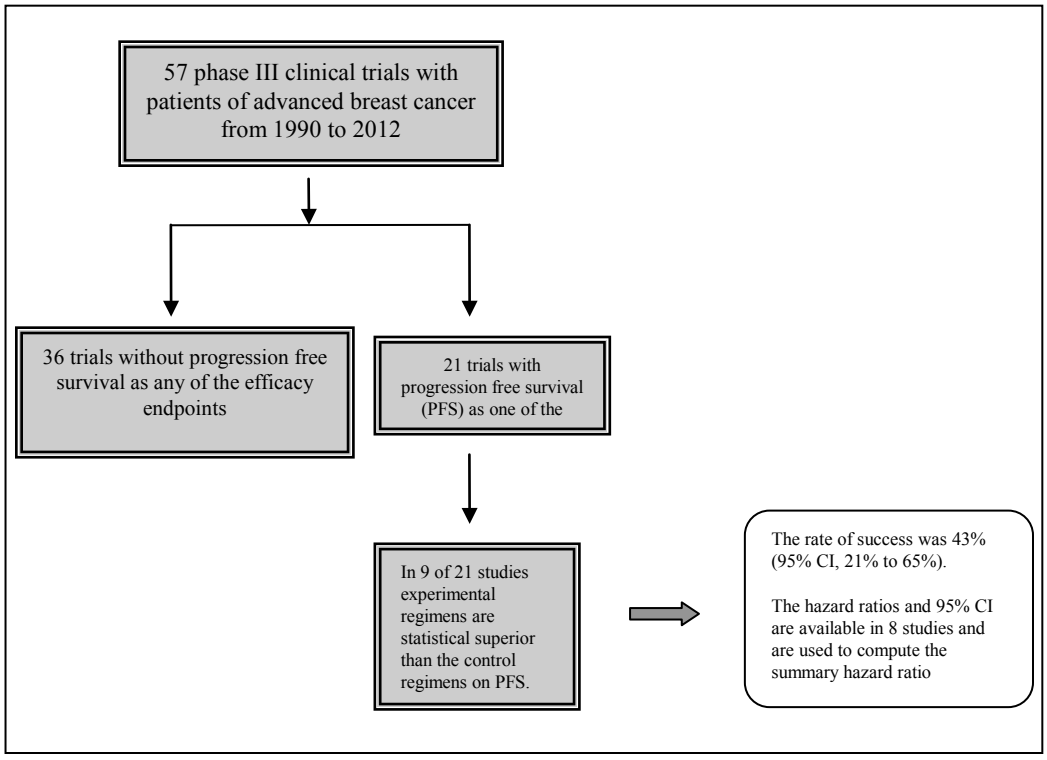


Figure 3-2 Flow chart of the searched phase III clinical trials



3.2 Phase II and phase III trials

We identify 7 phase II or phase III studies that tested a regimen subsequently studied in a phase III trial with PFS as one of efficacy endpoints (Allouache D, 2005; Demiray M, 2005; Fountzilas G, 2000; Bontenbal M, 2005; Mansutti M, 2008; Bunnell CA, 2006; Baselga J, 2003). The information from these phase II or phase III trial studies is summarized in Table 3-1. One subsequent phase III study is initiated based on two historical phase II studies (Allouache D, 2005; Demiray M, 2005).

A median of 46 patients were treated in each of the seven previous phase II or phase III studies. The number of patients treated in different phase II or phase III studies varied greatly, ranging from 24 to 136 patients. The range of median progression free survival time is from 3.8 to 12.3 months.

Table 3-1 Phase II or phase III studies of advanced breast cancer identified as preceding studies for subsequent phase III trial

Authors of phase II or phase III studies	No. of patients in phase II or phase III studies	Phase II or phase III regimen	Response rate	Median PFS (Month)
Allouache D, et al ;	35	Gemcitabine	40	7.2
Demiray M, et al	24	and Paclitaxel	41.7	9.6
Fountzilas G, et al.	39	Docetaxel and Gemcitabine	36	7
Bontenbal M, et al.	109*	Doxorubicin and Docetaxel	58	8
Mansutti M, et al	136*	Paclitaxel, Epirubicin and Capecitabine (TEX)	67	12.3
Bunnell CA, et al.	50	Ixabepilone and Capecitabine	30	3.8
Baselga J, et al.	46	Arzoxifene	40.5	10.7

* The studies done by **Bontenbal M** et al and **Mansutti M** et al are phase III studies. The other studies are phase II studies.

Table 3-2 shows that the number of patients, treatment regimens for the experimental and control groups, progression free survival (PFS), and overall survival (OS) information for the each of the subsequent six phase III studies. The order of studies separated by solid line is consistent with the order of studies separated by solid line in Table 3-2, which indicate each of the six subsequent phase III studies (Albain KS, 2008;

Chan S, 2009; Sparano JA, 2009; Hatschek T, 2012; Thomas ES, 2007; Deshmane V, 2007).

Table 3-2 Six Phase III trials of advanced breast cancer which are the subsequent phase III trials after the preceding trials listed in Table 3-1.

Authors of phase III studies	No. of patients in phase III studies (Experimental/ Control)	Phase III regimen (Experimental/ Control)	Progression Free Survival (Experimental/Control)			Overall Survival (Experimental/Control)		
			Number of Event	Median (Month)	p-value	Number of Event	Median (Month)	p-value
Albain KS, et al	266/263	Paclitaxel plus Gemcitabine with Paclitaxel monotherapy	246/247	5.9/3.9	0.0005	182/195	18.6/15.8	0.0489
Chan S, et al	153/152	Gemcitabine plus Docetaxel with Capecitabine plus Docetaxel	151/142	8.05/7.98	0.121*	119/115	19.29/21.45	0.983
Sparano JA, et al	373/378	PLD + docetaxel with docetaxel	NA/NA	9.8/7	0.000001	NA/NA	20.5/20.6	0.81
Hatschek T, et al	144/143	Epirubicin and paclitaxel with with without capecitabine (ET)	NA/NA	12.40/10.80	0.84*	NA/NA	29.7/26	0.22
Thomas ES , et al	375/377	ixabepilone plus capecitabine with capecitabine	310/329	5.8/4.2	0.0003*	NA/NA	12.9/11.1	Not significant
Deshmane V, et al	165/173	Arzoxifene With Tamoxifen	113/94	4/7.5	0.007†*	27/18	NA/17.1	0.157

† p-value is over 0.5 since the data is in the direction opposite that specified by the test.
* PFS is primary efficacy endpoint in the phase III study.

Response rates are frequently used as primary endpoint in phase II survival trials to assess the likelihood that the experimental regimen will increase survival over standard treatment in a phase III trial.

Figure 3-3 shows the regression line for the median progression free survival time of patients treated with a particular regimen in the subsequent phase III trial by the median progression free survival time of patients in the preceding phase II or phase III studies with the same regimen. The R square is 0.127, with the estimated slope of 0.368 and p-value of 0.431. There are one outlier in the figure, point (10.7, 4), which affects the correlation between prior and posterior PFS. It was from the PFS in the phase III study comparing arzoxifene with tamoxifen (RR) (Deshmane V, 2007).

Figure 3-3 Median progression free survival time in the subsequent phase III trials by median progression free survival time in the previous phase II or phase III studies

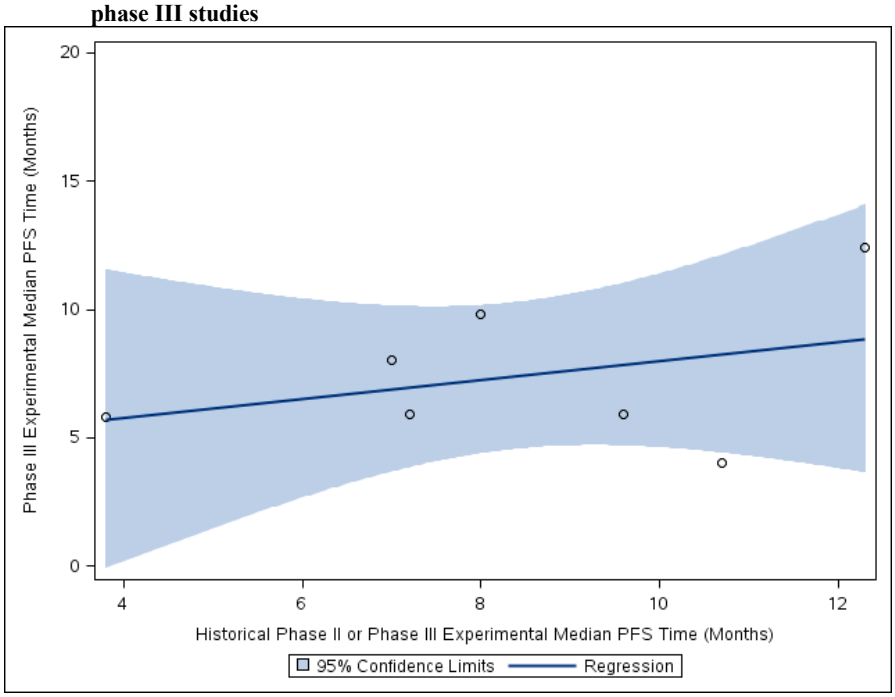
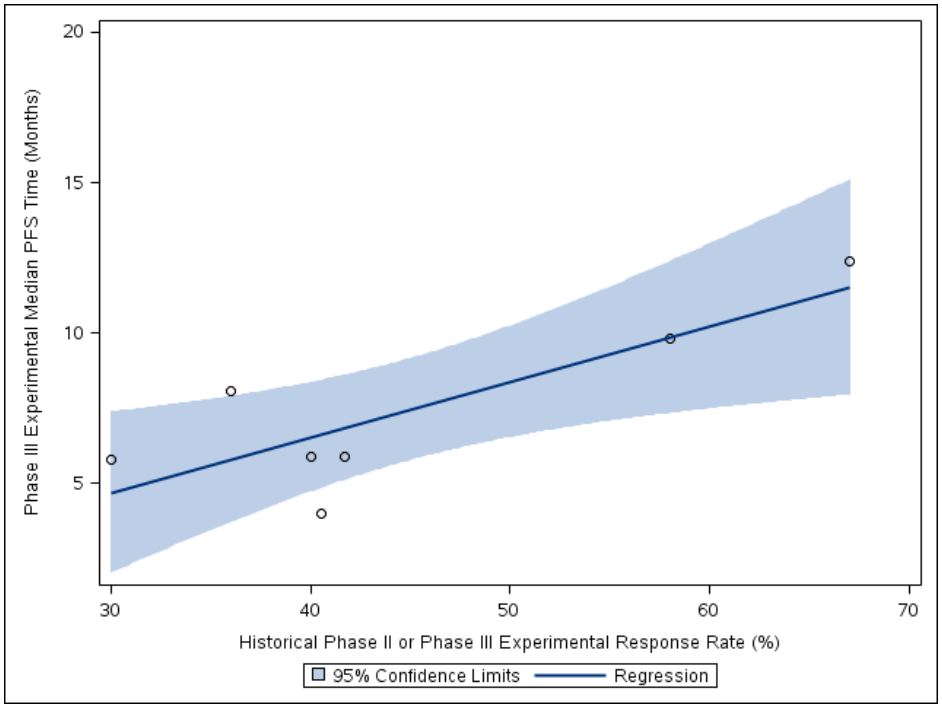


Figure 3-4 shows the regression line for the median progression free survival time of patients treated with a particular regimen in the subsequent phase III trial by the response rate of patients in the preceding phase II or phase III studies with the same regimen. The R square is 0.69, with the estimated slope 0.184 and p-value of 0.02. All the points are closely distributed around the regression line.

The median progression free survival of patients treated in the preceding phase II or phase III studies is plotted versus the median survival of patients treated on the experimental arm of the corresponding subsequent phase III trial in Figure 3-5. The least-squares regression lines are given in the figure, and the Pearson correlation coefficient is 0.917, with the estimated slope 1.76 and p-value =0.01. The PFS of 10.7 months for preceding phase II arzoxifene study, which is shown as an outlier in Figure 3-3, is not included in this figure because the median overall survival information for arzoxifene

regimen is not available in the subsequent phase III study comparing arzoxifene with tamoxifen (RR) due to large proportion of censored values (Deshmane V, 2007).

Figure 3-4 Median progression free survival time in the subsequent phase III trials by response rate in the previous phase II or phase III studies



In addition, in Figure 3-6, the median survival of patients treated on the experimental arm of subsequent phase III trial by response rate in the same regimen in patients treated in the preceding phase II or phase III studies shows similar correlation with the Pearson correlation coefficient is 0.895, and the estimated slope 0.347 and p-value = 0.016.

Figure 3-5 Median overall survival time in the subsequent phase III trials by median progression free survival time in the previous phase II or phase III studies

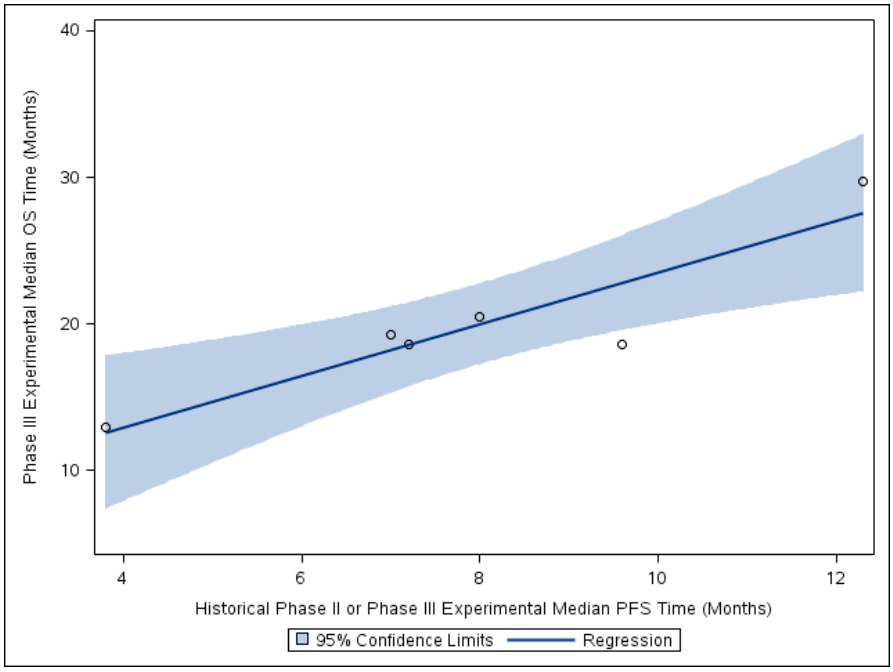
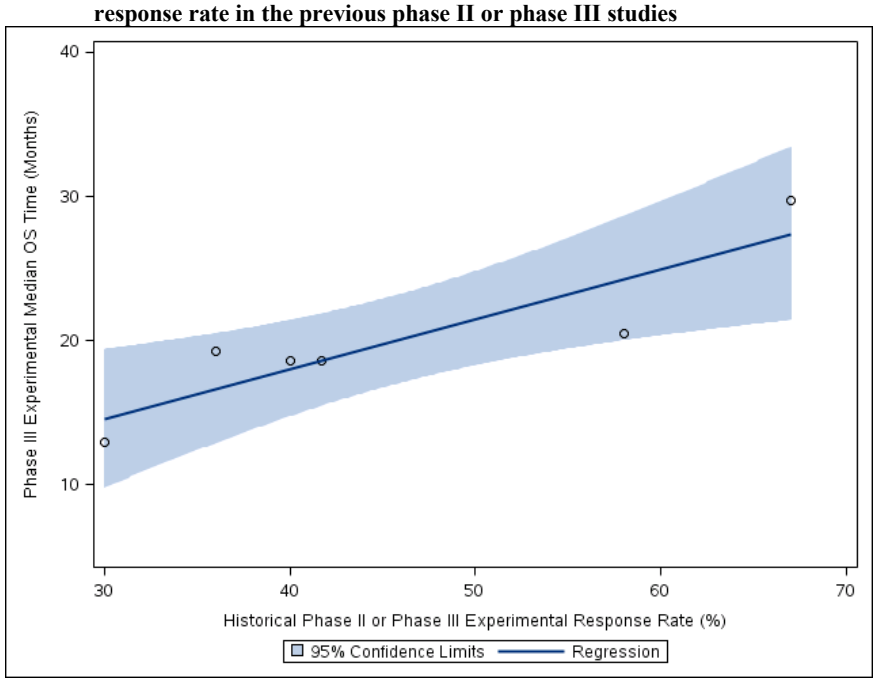


Figure 3-6 Median overall survival time in the subsequent phase III trials by response rate in the previous phase II or phase III studies



The median progression survival time was slightly longer in three of the six phase III studies than that in the previous trials that tested the same regimen (range of differences, 0.1–1.8 months). In other three studies the median progression free survival time in phase III is 1.3, 3.7 and 6.7 months less than that in preceding phase II trials, respectively. A nonparametric paired Wilcoxon signed rank test comparing the overall progression free survival of patients in the seven Phase II or Phase III trials with that of patients treated on the experimental arms of the subsequent phase III studies indicated that the median progression free survival times of these groups were insignificantly different with p-value of 0.81.

The response rates in previous phase II or phase III trials have a linear correlation with the phase III PFS results in the six studies. In addition, the previous PFS time also show weak correlation with subsequent PFS time in the phase III studies with correlation coefficient of 0.356, but not linear. Given the small sample size in the above regression analysis and the limits of information which response rate can convey in the oncology studies, the progression free survival information instead of response rates in previous phase II or phase III studies are utilized in the predictive model in this report. This is consistent with FDA recommendation on the endpoints supporting approvals in oncology. In the early 1980s, the FDA determined that cancer drug approval should be based on more direct evidence of clinical benefit, such as improvement in survival, improvement in a patient's quality of lift (QOL), improved physical functioning, or improved tumor-related symptoms, which may not always be predicted by, or correlate with, objective response rate (ORR) (Guidance, CDER 2007).

To retrospectively test how well the statistical model estimates the outcome of the phase III trials from previous historical survival data, we analyze the 7 phase II or III studies that gave rise to 6 subsequent phase III trials of the same regimen. The expected power is the usual statistical power averaged with regard to the size of the treatment difference anticipated on the basis of the median progression free survival observed in the preceding historical phase II or III trial, the number of events observed in the preceding historical phase II or III trial, and the distribution of median progression free survival anticipated in the phase III trial for the control group. The detail calculations are shown below.

3.3 Results from the Bayesian conjugate statistical model

3.3.1 The Exponential-Gamma model

3.3.1.1 Gamma prior distribution of hazard of progression for control regimen

For control regimen, six Gamma distributions are formed with parameter a_c and b_c on the basis of historical phase II or III trials referred in the six individual subsequent phase III trials. The a_c represents amount of information (the number of events) on which the prior trial is based and assumed to be a fixed number. The $1/b_c$ represented the total patient-time of survival (until event or censoring) in previous experience with the treatment. The distribution of hazard of disease progression for control regimen is modeled separately for different subsequent phase III clinical trials. Giving a_c, make mean of the Gamma distribution equal to log(2)/median survival time to get b_c for each trial.

3.3.1.2 Gamma prior distribution of hazard of progression for experimental regimen

Experimental group information about hazard available at the time of planning the phase III trial can also be specified as Gamma probability distribution.

For an experimental regimen, before the Phase III study, the probability of obtaining a positive result in a Phase III trial is set to 0.43 because only 9 (43%) of the 21 Phase III trials demonstrated a statistically significant improvement with regard to the endpoint of PFS. Prior information are specified by setting a_e and b_e to give an expected median (m_e) approximately equal to that expected of the control treatment but with the probability of 43% that m_e is greater than median PFS time of control group plus median PFS time difference of 2.3 months between experimental and control regimen groups.

$$\int_0^{\infty} \lambda \frac{1}{\Gamma(a_e) b_e^{a_e}} \lambda^{a_e - 1} \exp\left(-\frac{\lambda}{b_e}\right) d\lambda = \ln 2/m_c \quad (1)$$

$$\int_0^{\lambda_2} \frac{1}{\Gamma(a_e) b_e^{a_e}} \lambda^{a_e - 1} \exp\left(-\frac{\lambda}{b_e}\right) d\lambda = 0.43 \text{ and } \lambda_2 = \ln 2/(m_c + 2.3) \quad (2)$$

Use SAS program to get a_e and b_e from equation (1) and (2).

3.3.1.3 Gamma posterior distribution of hazard of progression for experimental regimen

We then update of Gamma prior distribution of hazard for PFS in experimental regimen group by incorporating information (number of event and the total patient-time to event or censoring) from trials preceding to the subsequent phase III trials, on which the phase III trials were initiated based.

3.3.1.4 Computation of expected power and results

After model the Gamma distribution of hazard for the experimental and control regimen based on the historical data from literature, we use the experimental data to update the existing distribution, and retrospectively calculate the expected power for the six phase III trials based on the formula in 2.3.1.3. We assume a 1:1 allocation ratio for experimental and control regimen, total number of event of 227 to be observed in a phase III trial in order to be adequately powered (90%) to detect the 54% increase in median time to progression. The derivation of total number of event is presented in section 3.1.

The control regimen information for each phase III studies are presented in Table 3-3 (Paridaens R, 2000; Smith RE, 1999; Bishop JF, 1999; O'Shaughnessy J, 2002; Chan S, 1999; Mansutti M, 2008; Blum JL, 1999; Buzdar A, 2002). The number of events and patient-time are used to model the Gamma distribution of hazard for control regimen for each individual phase III studies, respectively.

In the subsequent phase III trial comparing Paclitaxel plus Gemcitabine with Paclitaxel monotherapy, there are three historical trials for Paclitaxel monotherapy regimen and two historical trials for Paclitaxel plus Gemcitabine regimen referred in the report, respectively (Paridaens R, 2000; Smith RE, 1999; Bishop JF, 1999; Allouache D, 2005; Demiray M, 2005). The Gamma distribution of hazard for control regimen was modeled by using the sum of number of events and sum of patient time in the three historical studies with control regimen. Similarly, the number of events and patient time for two preceding studies with experimental regimen were added up and utilized to update the Gamma distribution for experimental regimen.

Table 3-3 List of treatment regimen, number of PFS events, Median PFS time and Patient-time in previous historical phase II or phase III studies

	Authors phase II or phase III/phase III	Control Regimen	Numer of Events (a_c) /Number of Patients	Median PFS (Month) of Control Regimen	Patient -time (Month) of Control Regimen ($1/b_c$)
1	**Paridaens R, et al**	Paclitaxel Monotherapy	160/166	3.9	898
	Smith RE, et al		241/278	6.3	2187
	Bishop JF, et al		103/107	5.3	785
2	**O'Shaughnessy J, et al**	Capecitabine plus Docetaxel	237/255	6.1	2083
3	**Chan S, et al**	Docetaxel	132/161	5.9	1120.7
4	**Mansutti M, et al**	Epirubicin and Paclitaxel	111/135	9.8	1564.7
5	**Blum JL, et al**	Capecitabine	135/162	3.06	594.5
6	**Buzdar A, et al**	Tamoxifen	369/673	7.5	3989

We calculate the expected power for six phase III studies (Albain KS, 2008; Chan S, 2009; Sparano JA, 2009; Hatschek T, 2012; Thomas ES, 2007; Deshmane V, 2007). The Gamma parameters used to model the Gamma distribution of hazard for both experimental and control groups are presented in Table 3-4.

In Table 3-5, the information of median PFS and number of events of the 7 previous phase II or III studies and median survival time and p-value in the subsequent phase III trial are presented.

The median expected power for the six phase III trials based on the corresponding previous phase II or III trials was 0.505 (range, 0.29-0.78). The two experimental regimens that yield the high expected powers (0.78 and 0.59) were followed by phase III studies that show a statistically significant difference between experimental regimen and control regimen arms (Albain KS, 2008; Sparano JA, 2009). Likewise, the experimental regimen that yield the lowest expected powers (0.29) was followed by phase III study that show a statistically insignificant difference between experimental regimen and control

regimen arms (Chan S, 2009). There are two experimental regimens with expected power of 0.41 and 0.42. One of the subsequent phase III trials is successful, one is not.

Table 3-4 List of treatment regimen in subsequent phase III studies, and corresponding Gamma parameters used to model the Gamma distribution of hazard for both experimental and control groups

Experiment Vs. Control Treatment in Phase III	a_c	$1/b_c$	a_e	$1/b_e$	d_e	T_e	Power
Paclitaxel plus Gemcitabine with Paclitaxel Monotherapy	160	898	1.25	7.03	31	319	0.93
	241	2187	2	18.18	31	319	0.28
	103	785	1.7	13	31	319	0.56
	160	898	1.25	7.03	21	286	0.99
	241	2187	2	18.18	21	286	0.69
	103	785	1.7	13	21	286	0.87
	504*	**3870***	**1.7**	**13**	**52***	**605***	**0.78**
Gemcitabine plus Docetaxel with Capecitabine plus Docetaxel	**237**	**2083**	**1.9**	**16.72**	**30**	**299**	**0.29**
PLD + Docetaxel with Docetaxel	**132**	**1120.7**	**1.85**	**15.75**	**103**	**1184.9**	**0.59**
Epirubicin and Paclitaxel with and without Capecitabine	**111**	**1564.7**	**3.2**	**45.24**	**110**	**1946**	**0.42**
Ixabepilone plus Capecitabine with Capecitabine	**135**	**594.5**	**1.05**	**4.64**	**47**	**255.84**	**0.41**
Arzoxifene with Tamoxifen	**369**	**3989**	**2.4**	**25.97**	**26**	**396**	**0.62**

* sum of events and sum of patient time were used separately from historical trials with control regimen and experimental regimen to model Gamma distribution.

Four of the six subsequent phase III trials showed statistically significant difference between the PFS of patients on the experimental arm and that of patients on the control arm. However, in one of the five statistically significant trials, the PFS of patients in experimental arm are 3.5 months less than that of patients in control arm (Deshmane V, 2007). The p-value is in fact over 0.5 since the data is in the direction opposite that specified by the test. However, the expected power based on previous experiences is as high as 0.62. In this trial, the effect demonstrated by Arzoxifene in subsequent phase III study was lower than anticipated on the basis of previous studies. Compared with the PFS of 4 months (95% CI 3.4–5.6 months) with 36% of the patients censored in subsequent phase III study, the PFS for patients assigned to Arzoxifene who were deemed tamoxifen

sensitive in the preceding phase II studies were 10.7 months (95% CI 8.6–16.8 months) with 43% of the patients censored in the European study (Baselga J, 2003; Deshmane V, 2007). The response rate (CR + PR) in the current study (23.6%) also was less than those observed in the preceding phase II studies (40.5% among tamoxifen-sensitive patients in the European trial). Although there were some modest differences in patient and tumor characteristics among the phase II study and the subsequent phase III study, no differences appear sufficiently large to account for the lower-than-anticipated time-to-event parameters noted in the present study (Deshmane V, 2007). The PFS for the experimental regimen is 4 months, which is surprisingly low comparing with previous PFS result in the historical phase II European study, which was 10.7 months (Baselga J, 2003; Deshmane V, 2007).

Figures 3-7 to 3-12 shows the Gamma distribution of hazard for control regimen in black line, the Gamma prior distribution of hazard for experimental regimen in red line, and the Gamma posterior distribution of hazard for experimental regimen in green line for each of six subsequent treatment comparisons in the corresponding phase III studies.

Figure 3-7 Probability density for Gamma distribution – Paclitaxel plus Gemcitabine with Paclitaxel monotherapy

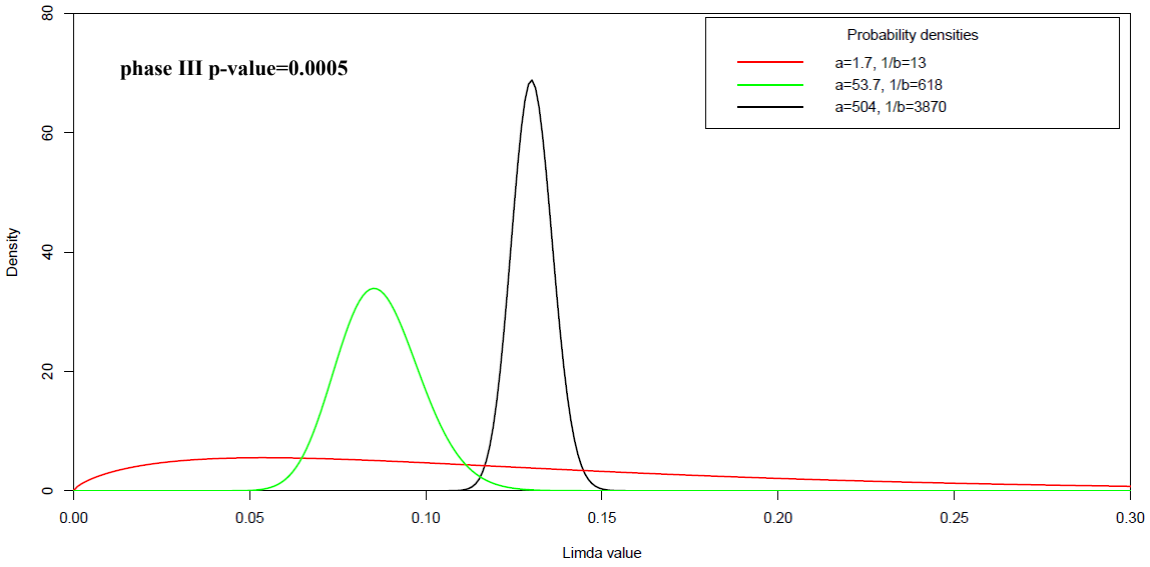


Figure 3-8 Probability density for Gamma distribution - Gemcitabine plus Docetaxel with Capecitabine plus Docetaxel

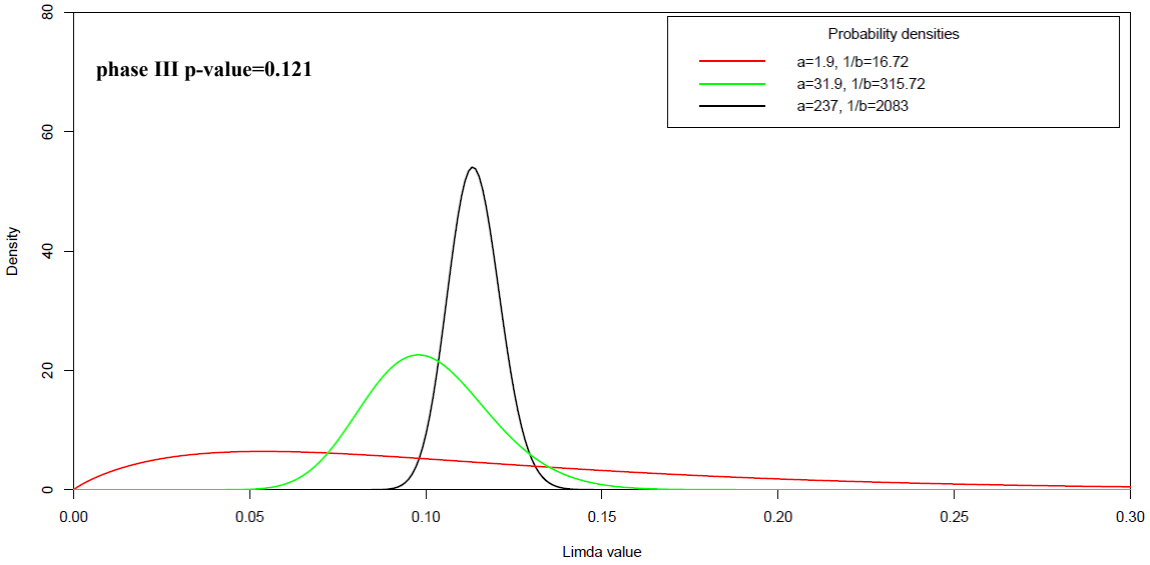


Figure 3-9 Probability density for Gamma distribution - PLD + Docetaxel with Docetaxel

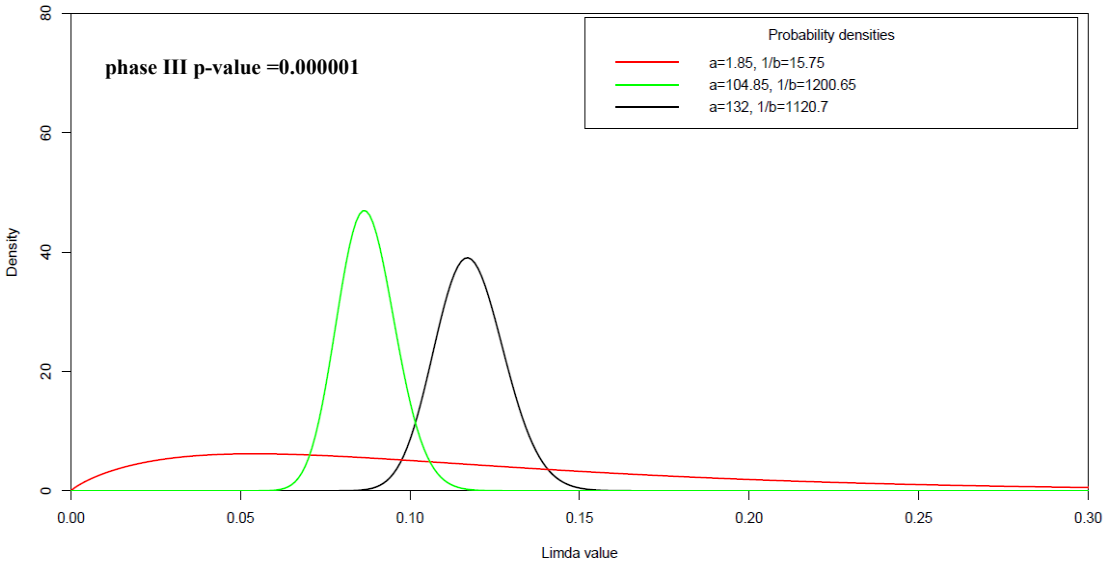


Figure 3-10 Probability density for Gamma distribution - Epirubicin and Paclitaxel with with without Capecitabine

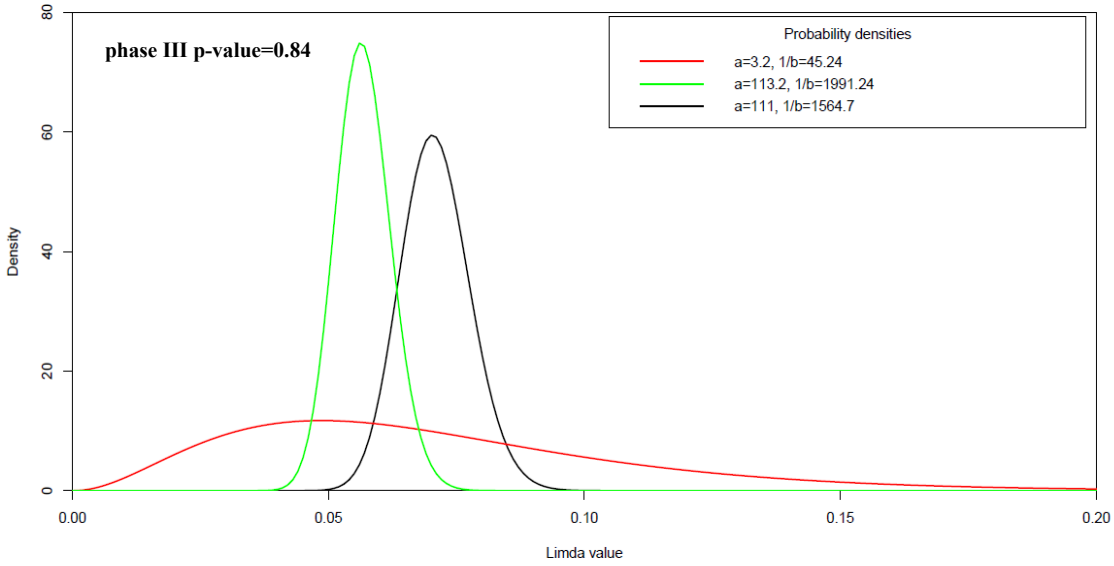


Figure 3-11 Probability density for Gamma distribution - Ixabepilone plus Capecitabine with Capecitabine

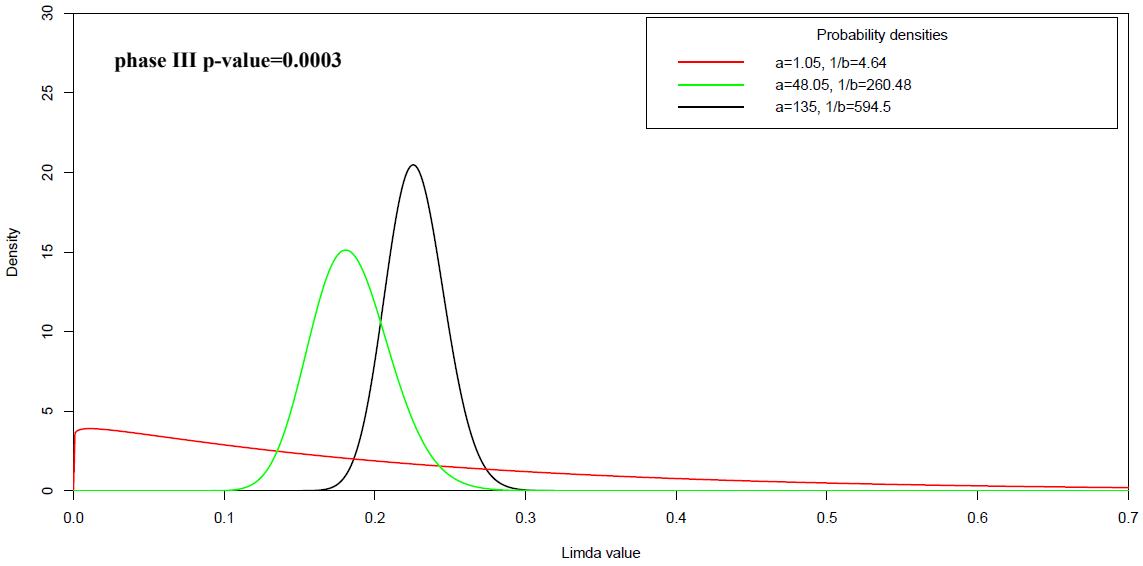


Figure 3-12 Probability density for Gamma distribution - Arzoxifene with Tamoxifen

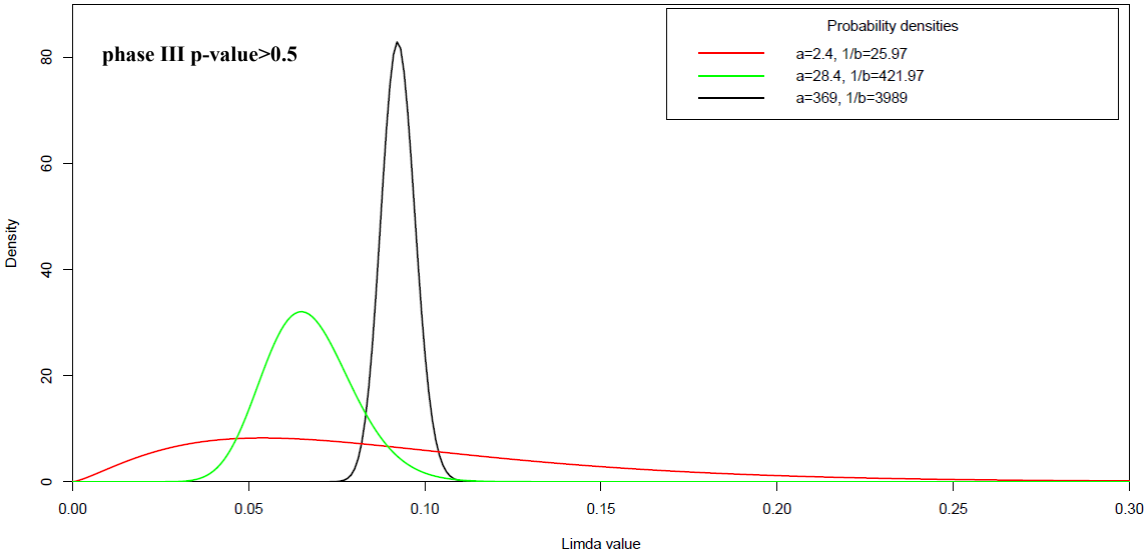


Table 3-5 List of expected power for future phase III studies, previous historical phase II or phase III studies, and results of PFS in subsequent actual phase III trials from Exponential-Gamma Model

Std. #	Authors of phase II or phase III/phase III	Experiment Vs. Control Treatment in Phase III	Observed Preceding Study		Power †	Subsequent Phase III			
			Number of Events/ Number of Patients	Median PFS (Month) of Experimental		Median PFS Month (Experimental vs. Control)	Percentage Increase in Median PFS Time	Number of Even (Experimental v Control)	p-Value
1	**Allouache D, et al ; Demiray M, et al /Albain KS, et al**	Paclitaxel plus Gemcitabine with Paclitaxel Monotherapy	31/35; 21/24	7.2 9.6	0.78	5.9/3.9	0.51	246/247	0.0005
2	**Fountzilas G, et al; /Chan S, et al (2009)**	Gemcitabine plus Docetaxel with Capecitabine plus Docetaxel	30/39	7	0.29	8.05/7.98	0.009	151/142	0.121*
3	**Bontenbal M, et al. /Sparano JA, et al**	PLD + Docetaxel with Docetaxel	103/109	8	0.59	9.8/7	0.4	NA/NA	0.000001
4	**Mansutti M, et al /Hatschek T, et al**	Epirubicin and Paclitaxel with and without Capecitabine	110/136	12.3	0.42	12.40/10.80	0.15	NA/NA	0.84*
5	**Bunnell CA, et al. /Thomas ES, et al**	Ixabepilone plus Capecitabine with Capecitabine	47/50	3.8	0.41	5.8/4.2	0.38	310/329	0.0003*
6	**Baselga J, et al /Deshmane V, et al**	Arzoxifene with Tamoxifen	26/46	10.7	0.62	4/7.5	-0.46	113/94	0.007‡*

† The power is calculated based on total number of events corresponding to 227 events to detect 54% increase in median PFS time.
‡ p-value is over 0.5 since the data is in the direction opposite that specified by the test.
* PFS is primary efficacy endpoint in the phase III study.

3.3.2 The Weibull-Inverse Gamma model

3.3.2.1 Inverse Gamma prior distribution of Weibull scale parameter for control regimen

For control regimen, Inverse Gamma distributions are formed with parameter a_c and b_c on the basis of historical phase II or III trials referred in the six individual subsequent phase III trials. The a_c represents amount of information (the number of event of interest) in which the prior trial included and is assumed to be a fixed constant. The b_c represented the sum of the β'th power of each patient-time of survival (until event or censoring) in previous experience with the treatment. The distribution of Weibull scale parameter for disease progression of control regimen is modeled separately for each control regimen subsequent tested in phase III clinical trials. Giving a_c, make mean of the Inverse Gamma distribution equals to median survival time/$\log(2)^{1/\beta}$ to get b_c for each trial.

Weibull distribution gives a distribution for which the failure rate is proportional to a power of time. The β is the Weibull shape parameter, which is that power plus one, so this parameter can be interpreted as how the failure rate changes over time.

In the Weibull-Inverse Gamma model, the Weibull shape parameter is consider as known fixed number. The values of β are obtained through information extracted from the Kaplan Meier curve. The detail procedures are shown as below: (1) pick two points from the corresponding Kaplan Meier curve. The time point selection is arbitrary, but in order to extract information to the fullest from the main survival curve, we pick one point with the survival probability in the range of 0.2 to 0.4, and another point with survival probability in 0.6 to 0.8; (2) With two pairs of time and survival probability, we can

easily find the parameters of a Weibull distribution using the Weibull2 function in the R Hmisc package.

Table 3-6 shows the two time points, the survival probability at those two points from preceding studies with the control regimens, and calculated Weibull β parameter. One of the prior clinical trials does not include the Kaplan Meier curve so we have to exclude this study from the Weibull-Inverse Gamma model.

Table 3-6 List of two time points with corresponding PFS probabilities from prior studies for each study with control regimen

	Authors of phase II or phase III/phase III	Control Regimen	Time (month)	Survival Probability	Weibull β parameter
1	Paridaens R, et al Smith RE, et al Bishop JF, et al	Paclitaxel Monotherapy	2.13	0.736	1.22
			7.27	0.254	
			4.45	0.638	0.91
			13.3	0.295	
			3.06	0.688	1.40
			8.72	0.198	
2	O'Shaughnessy J, et al	Capecitabine plus Docetaxel	3.71	0.686	1.47
			8.72	0.266	
3	Chan S, et al	Docetaxel	4.93	0.721	1.95
			8.91	0.279	
5	Blum JL, et al	Capecitabine	1.51	0.735	0.96
			7.24	0.252	
6	Buzdar A, et al	Tamoxifen	3.52	0.671	0.69
			20.4	0.264	

Note: For the phase III study 4, the prior clinical trial with control regimen Epirubicin and Paclitaxel does not present Kaplan Meier plot for time to event, so we do not include it into the Weibull-Inverse Gamma model.

Based on the values of shape parameter for different studies, we can see some cases of the change of survival probability increase over time, such as the values of β as 0.69 and 0.96, some cases of the change of the survival probability decrease over time, such as the values of β as 1.22 and 1.40.

3.3.2.2 Inverse Gamma prior distribution of Weibull scale parameter for experimental regimen

For experimental group information about hazard available at the time of planning the phase III trial can also be modelled through an Inverse Gamma probability distribution.

For an experimental regimen, before the Phase III study, the probability of obtaining a positive result in a phase III trial is set to 0.43 because only 9 (43%) of the 21 phase III trials demonstrated a statistically significant improvement with regard to the endpoint of PFS. Prior information are specified by setting a_e and b_e so that the variance of the prior Inverse Gamma distribution for experimental regimen equals to the variance of the prior Inverse Gamma distribution for control regimen but with the probability of 43% that median PFS time of experimental group m_e is greater than median PFS time of control group m_c plus median PFS time difference of 2.3 months between experimental and control regimen groups.

$$\frac{b_c^{\,2}}{(a_c-1)^2(a_c-2)}=\frac{b_e^{\,2}}{(a_e-1)^2(a_e-2)} \quad (1)$$

$$\int_0^{\theta_2}\frac{b_e^{\,a_e}}{\Gamma(a_e)\theta^{a_e+1}}\exp\left(-\frac{b_e}{\theta}\right)d\theta=0.57 \text{ and } \theta_2=\frac{(m_c+2.3)}{\ln(2)^{\frac{1}{\beta_c}}} \quad (2)$$

The β_c in above equation (2) is the Weibull shape parameter estimated from the results of individual clinical trial with control regimen. The Weibull models allow for the possibility of time varying hazard rates, but we assume a constant hazard ratio between experimental and control regimen groups. Therefore we use the same shape parameter for the experimental group as that for control group. The values of shape parameter β are

presented in previous table 3-6. We incorporate the median survival time into the distribution of Weibull scale parameter θ by setting the mean of the prior distribution of Weibull scale parameter corresponding to the median survival time given the fixed value of Weibull shape parameter β.

Use SAS program to get a_e and b_e from equation (1) and (2).

3.3.2.3 Inverse Gamma posterior distribution of Weibull scale parameter for experimental regimen group

We update the prior Inverse Gamma distribution of Weibull scale parameter for experimental regimen group by incorporating information of number of event of interests and sum of the β'th power of each patient-time to event or censoring from trials preceding to the subsequent phase III trials, on which the phase III trials were initiated based. These information was elicited from six preceeding phase II or phase III studies with the corresponding experimental regimen. The scale parameter β for experimental group is set to the same as the scale parameter for the corresponding control group.

3.3.2.4 Computation of expected power and results

After build the Inverse Gamma distributions, which are the prior distributions of Weibull scale parameter for the experimental and control regimen, based on the historical data from literature, we use the experimental data to update the existing distributions. Then we obtained one posterior distribution of Weibull scale parameter. According to the section 2.3.2.3, in the Weibull-Inverse Gamma model the hazard ratio at any time point between an individual in the control treatment and an individual in active treatment group is

constant and equals to $\frac{\theta_e}{\theta_c}$. Thus we model the hazard ratio of event for the control and experimental regimen as ratio of two Weibull scale parameters, each with an Inverse Gamma distribution to retrospectively calculate the expected power for the five phase III trials based on the formula in 2.3.2.3. We make the same assumptions as those of in Exponential-Gamma model, a 1:1 allocation ratio for experimental and control regimen in the potential subsequent phase III study, total number of event of 227 to be observed in the phase III trial in order to be adequately powered (90%) to detect the 54% increase in median time to progression. The derivation of total number of event is presented in section 3.1.

In the subsequent phase III trial comparing Paclitaxel plus Gemcitabine with Paclitaxel monotherapy, there are three historical trials for Paclitaxel monotherapy regimen and two historical trials for Paclitaxel plus Gemcitabine regimen referred in the report, respectively (Paridaens R, 2000; Smith RE, 1999; Bishop JF, 1999; Allouache D, 2005; Demiray M, 2007). The Inverse Gamma distribution of Weibull scale parameter for control regimen is modeled by using the sum of number of events and sum of patient time to the β^{th} power in the three historical studies with control regimen. Similarly, the number of events and patient time for two preceding studies with experimental regimen are added up and utilized to update the prior Inverse Gamma distribution for experimental regimen.

We calculate the expected power for five phase III studies. The Inverse Gamma parameters used to model the Inverse Gamma distribution of hazard for both experimental and control groups are presented in Table 3-7.

In Table 3-8, the information of median PFS and number of events of the 6 previous phase II or III studies and median survival time and p-value in the subsequent phase III trial are presented.

The median expected power for the five phase III trials based on the corresponding previous phase II or III trials was 0.66 (range, 0.51-0.98). The three experimental regimens that yield the highest expected powers (0.75, 0.66 and 0.98) were followed by phase III studies that show a statistically significant difference between experimental regimen and control regimen arms (Albain KS, 2008; Sparano JA, 2009; Thomas ES, 2007).

Table 3-7 List of treatment regimen in subsequent phase III studies, and corresponding Inverse Gamma parameters used to model the Inverse Gamma distribution of Weibull scale parameter for experimental and control groups

	Experiment Vs. Control Treatment in Phase III	a_c	b_c	β	a_e	b_e	d_e	T_e	Power
1	**Paclitaxel plus Gemcitabine with Paclitaxel Monotherapy**	160	837.4	1.22	396	3285.1	31	291.7	0.94
		241	2261.9	0.91	441	5620.2	31	291.7	0.58
		103	702.4	1.4	206	2006.3	31	291.7	0.75
		160	837.4	1.22	396	3285.1	21	287.2	0.96
		241	2261.9	0.91	441	5620.4	21	287.2	0.64
		103	702.4	1.4	206	2006.3	21	287.2	0.84
		504*	3801.7*	1.11**	1009	10790.3	52*	578.9*	**0.75**
2	**Gemcitabine plus Docetaxel with Capecitabine plus Docetaxel**	237	1847.2	1.47	441	4707.1	30	263.7	**0.62**
3	**PLD + Docetaxel with Docetaxel**	132	932.7	1.95	249	2433.9	103	984.7	**0.66**
5	**Ixabepilone plus Capecitabine with Capecitabine**	135	600.7	0.96	404	3140.8	47	256.1	**0.98**
6	**Arzoxifene with Tamoxifen**	369	4694.6	0.69	621	10272.0	26	455.0	**0.51**

* sum of events and sum of patient time to the β^{th} power were used separately from historical trials with control regimen and experimental regimen to model Inverse Gamma distribution.

** The β takes the value of weighted βs from three clinical trials.

Likewise, one experimental regimen Arzoxiene that yields the lowest expected powers (0.51) of five studies was followed by phase III study comparing Arzoxifene with

Tamoxifen in which the experimental regimen fail to show a statistically significant superiority than control regimen arm (Deshmane V, 2007).

Another experimental regimen with Gemcitabine plus Docetaxel as experimental regimen yields the second to the lowest expected powers (0.62) was followed by phase III study comparing Gemcitabine plus Docetaxel with Capecitabine plus Docetaxel, in which there is no statistically significant difference between experimental regimen and control regimen arms (Chan S, 2009).

Figures 3-13 to 3-17 shows the Inverse Gamma distribution of Weibull scale parameter for control regimen in blue line, the Inverse Gamma prior distribution of Weibull scale parameter for experimental regimen in red line, and the Inverse Gamma posterior distribution for experimental regimen in green line for each of five subsequent treatment comparisons in the corresponding phase III studies.

Figure 3-13 Probability density of Inverse Gamma distribution - Paclitaxel plus Gemcitabine with Paclitaxel Monotherapy (phase III p-value=0.0005)

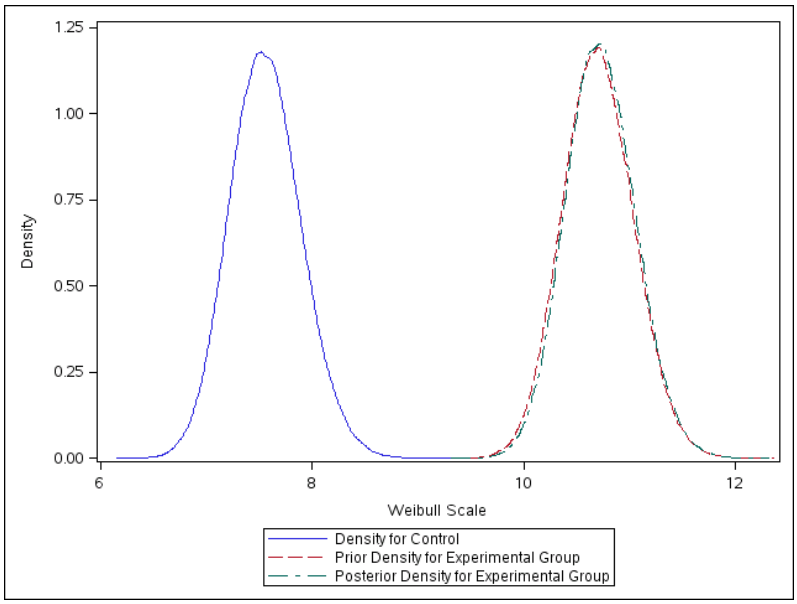


Figure 3-14 Probability density of Inverse Gamma distribution - Gemcitabine plus Docetaxel with Capecitabine plus Docetaxel (phase III p-value=0.121)

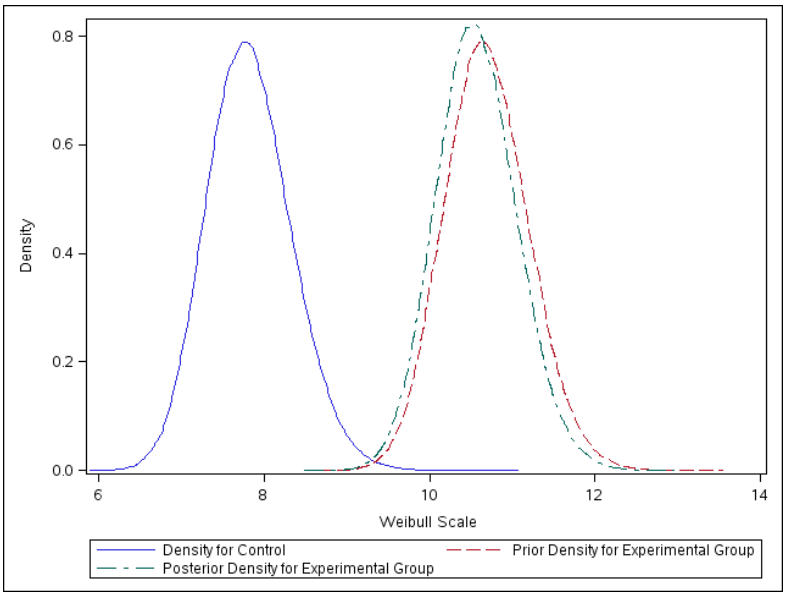


Figure 3-15 Probability density of Inverse Gamma distribution - PLD + Docetaxel with Docetaxel (phase III p-value=0.000001)

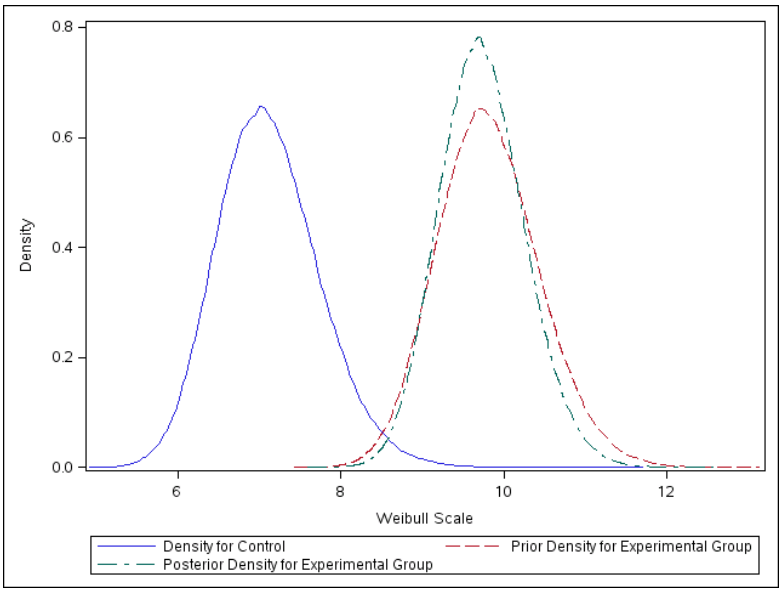


Figure 3-16 Probability density of Inverse Gamma distribution - Ixabepilone plus Capecitabine with Capecitabine (phase III p-value=0.0003)

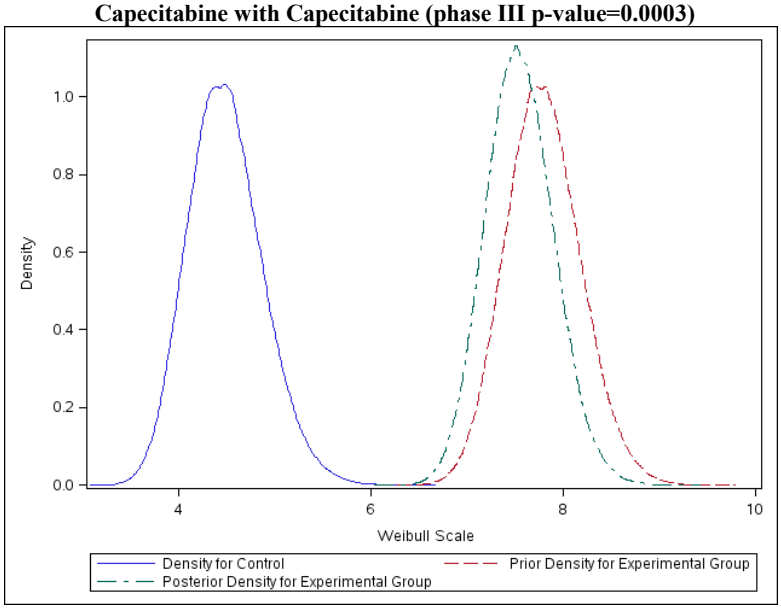
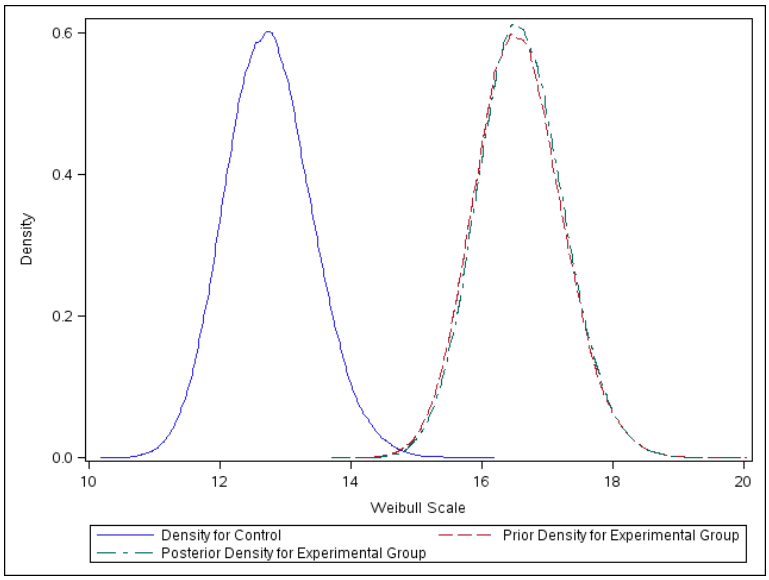


Figure 3-17 Probability density of Inverse Gamma distribution - Arzoxifene with Tamoxifen (phase III p-value > 0.5)



As we can see from the Table 3-8, the lowest values of calculated power obtained from Weilbul-Inverse Gamma model are 0.51 and 0.62, respectively, corresponding to the phase III trials comparing experimental regimen Gemcitabine plus Docetaxel with control regimen Capecitabine plus Docetaxel, and experimental regimen Arzoxifene with control regimen Tamoxifen. Those two phase III trials are the only two trials among the five phase III trials we studied that the experimental regimens were not statistically superior to the control regimens in the endpoint of progression free survival. The calculated powers for the other three phase III trials, in which the experimental regimens were proved to be statistically superior to the control regimens with regard to endpoint of progression free survival, are higher than the previous two, which are 0.66, 0.75 and 0.98.

Based on the our experiences of calculating the expected powers for those five studies, the successful phase III trials correspond to higher calculated powers, with the borderline of calculated power between 0.62 and 0.66.

Table 3-8 List of expected power for future phase III studies, previous historical phase II or phase III studies, and results of PFS in subsequent actual phase III trials from Weibull-Inverse Gamma Model

Std. #	Authors of phase II or phase III/phase III	Experiment Vs. Control Treatment in Phase III	Observed Preceding Study		Power †	Subsequent Phase III			
			Numer of Events/ Number of Patients	Median PFS (Month) of Experimental		Median PFS Month (Experimental vs. Control)	Percentage Increase in Median PFS Time	Number of Events (Experimental vs. Control)	p-Value
1	**Allouache D, et al ; Demiray M, et al /Albain KS, et al**	Paclitaxel plus Gemcitabine with Paclitaxel Monotherapy	31/35; 21/24	7.2 9.6	0.75	5.9/3.9	0.51	246/247	0.0005
2	**Fountzilas G, et al; /Chan S, et al (2009)**	Gemcitabine plus Docetaxel with Capecitabine plus Docetaxel	30/39	7	0.62	8.05/7.98	0.009	151/142	0.121*
3	**Bontenbal M, et al. /Sparano JA, et al**	PLD + Docetaxel with Docetaxel	103/109	8	0.66	9.8/7	0.4	NA/NA	0.000001
5	**Bunnell CA, et al. /Thomas ES, et al**	Ixabepilone plus Capecitabine with Capecitabine	47/50	3.8	0.98	5.8/4.2	0.38	310/329	0.0003*
6	**Baselga J, et al /Deshmane V, et al**	Arzoxifene with Tamoxifen	26/46	10.7	0.51	4/7.5	-0.46	113/94	0.007‡*

† The power is calculated based on total number of events corresponding to 227 events to detect 54% increase in median PFS time.

‡ p-value is over 0.5 since the data is in the direction opposite that specified by the test.

* PFS is primary efficacy endpoint in the phase III study.

3.4 Derivation of hazard ratio from Piece-wise Exponential model

3.4.1 Step 1 Reconstructing the data from published Kaplan Meier survival curves

The statistical results, which usually are reported in the publication for oncology phase II or III clinical trials with time to event outcome, are median time to events, log-rank statistics and Cox hazard ratio. These do not constitute the sufficient statistics required for secondary analyses, such as meta-analysis. The use of the statistics requires strong assumptions that may not hold in reality and have not been adequately tested.

In order to enhance the quality of secondary data analyses, a method was proposed by Patricia Guyot, et al to derive a close approximation to the original individual patient time to event data from which the published Kaplan Meier survival graph was generated (Guyot P, 2012). With the assistance of digital software they read in the coordinates of the Kaplan Meier curves from the published graph and utilized the information of numbers at risk, often published at four or five time points under the x-axis of the KM graph, and total number of events, where available, to reconstruct the Kaplan-Meier data for each arm. The iterative numerical methods are utilized in this approach to solve the inverted Kaplan Meier equations in order to obtain consistent results and make the best use of the information available.

The reproducibility and accuracy of survival probabilities, median survival times and hazard ratios based on reconstructed KM data was assessed by comparing published statistics (survival probabilities, medians and hazard ratios) with statistics based on repeated reconstructions by multiple observers (Guyot P, 2012). Based on information on total number of events and numbers at risk they found a mean error of -0.103% (95%CI:-0.260; 0.055) for survival probabilities. This means that if the original survival

probability estimate was 50%, we would expect survival probability based on reconstructed data to be 49.897% (95% CI: 49.740: 50.055). The authors concluded that there is therefore no significant systematic error. For median, the ME on the log scale was 0.011 (95%CI: 0.004; 0.018). By taking the exponentials of these values, they obtained that the mean error is on average a factor of exp(0.011), or 1.1% (95%CI: 0.4%; 1.8%). With regarding to hazards ratio, based on full information they obtained a ME on the log scale of 0.008 (95%CI:-0.015; 0.030). By taking the exponentials again, we can infer that if the original HR is 1.5, or 0.667 for its inverse, then we would expect to obtain a reconstructed HR of 1.512, or 0.661 for its inverse. The confidence intervals for the ME span zero, indicating no statistically significant systematic error. The MAE on hazard ratio is 0.017 (95%CI: 0.002; 1.222), in another word, if the original HR was 1.5, or 0.667 for its inverse, we would expect the reconstructed HR would be within a factor or exp (0.017) = 1.017 either side of the original values, i.e. 1.475 or 1.525, or 0.656 or 0.678 for its inverse.

Furthermore, the reliability of the reconstructed data depends on two related elements: the quality of the initial input and the level of information provided by the publication. The figure extracted from the .pdf should not be of low quality (for instance, blurry figure and/or poor numerical axis scale); otherwise the user may struggle to extract accurate data via the digitizing software. Moreover, the extracted data needed to run the algorithm should be consistent and sufficient. One rule of thumb is that since the survival probability is decreasing over time, the survival probability at later time point should always be less than or equal to the survival probabilities at previous time points.

The reconstructed individual patient time to event data enable us to test the survival model assumption, and practice survival data analysis in different methods. We call the reconstructed data as individual time to event data in a sense that we obtain the individual patient time to event data and survival probabilities, however, this is not the true individual patient level data because of the lack of the information on other covariates which have effects on the survival also.

3.4.2 Step 2 Model the distribution of hazard ratio with semi-parametric model

In this report we have applied the Exponential-Gamma and the Weibull-Inverse Gamma model separately into estimation of the expected power for six phase III trials. The six phase III trials are selected solely because the information of total number of event of interest during the clinical trial and median survival time were available and reported in the publications of the preceding trials, based on what the later six phase III trials were initiated. In addition to the two model based methods, we propose the third method to estimate the distribution of hazard ratio instead of individual distribution of hazard for experimental and control group with a semi-parametric model. We will derive a normal distribution of natural logarithm of hazard ratio by fitting the reconstructed survival data regenerated from Kaplan Meier curve to a Piecewise Exponential model. In the setup of the Piecewise model, we have assumed that the hazard rate varies across time periods but that the effect of the covariates is the same.

The power can be approximated as

$$Power = \phi^{-1}\left(\frac{\log(HR)}{\sqrt{\frac{4}{d}}} - z_{1-\alpha/2}\right)$$

ϕ^{-1} is the inverse of the standard normal distribution

$z_{1-\alpha/2}$ is the $100(1-\alpha/2)$ upper percentile of the standard normal distribution

d is the total number of events expected in the subsequent phase III trial at the time of analysis.

$\log(HR) \sim \varphi(\mu, \sigma^2)$

In the meantime, we will examine the application of the proportional hazard model in addition to the Piecewise Exponential model.

Among the trials included in this report for estimation of expected power, there are five previous phase II or phase III trials (Demiray M, 2005; Bontenbal M, 2005; Paridaens R, 2000; Bishop JF, 1999; Chan S, 1999), in which number of subjects at risk is presented in the Kaplan Meier graph and total number of progression events are reported in the publications. Those five clinical trials are part of the previous trials based on which two phase III trials were initiated later on (Albain KS, 2008; Sparano JA, 2009).

Clinical Trial 1: Phase III trial Gemcitabine Plus Paclitaxel Versus Paclitaxel Monotherapy in Patients With Metastatic Breast Cancer and Prior Anthracycline Treatment (Albain KS, 2008)

Two prior clinical trials with regimen Gemcitabine plus Paclitaxel (Allouache D, 2005; Demiray M, 2005) and three prior clinical trials with regimen Paclitaxel (Paridaens R, 2000; Smith RE, 1999; Bishop JF, 1999) were referred in the publication of the later phase III clinical trial comparing Gemcitabine plus Paclitaxel versus Paclitaxel monotherapy in patients with metastatic breast cancer and prior anthracycline treatment. The information of total number of disease progression and median survival times is

available in the publications of those five clinical trials, which make it possible to apply the Exponential-Gamma model and the Weibull-Inverse Gamma model into estimation of the expected power for the subsequent phase III trial as shown in previous section in this report.

After taking a further look at the publications, we find out that the Kaplan Meier graph with number of subjects at different time points and total number of disease progression are available and reported in two of the trials with regimen Paclitaxel (Paridaens R, 2000; Bishop JF, 1999), which is the control regimen in the later Phase III trial.

Below Figure 3-18 and Figure 3-19 are the two Kaplan Meier graphs for those two studies.

Figure 3-18 PFS on first-line treatment with paclitaxel (solid line) or with doxorubicin (dashed line)

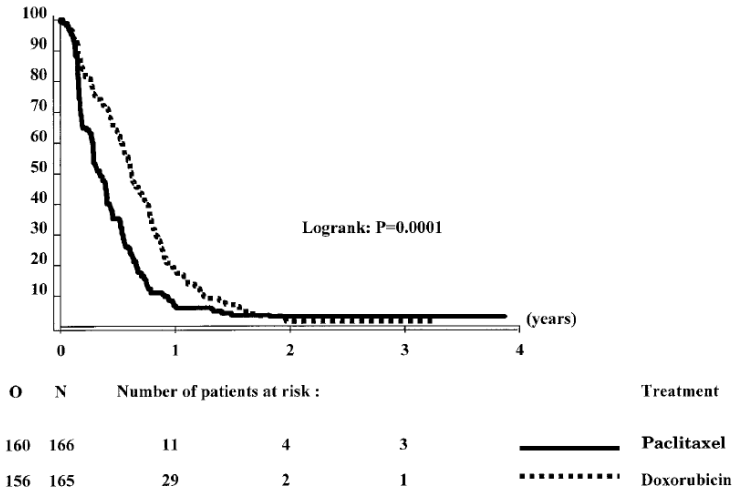
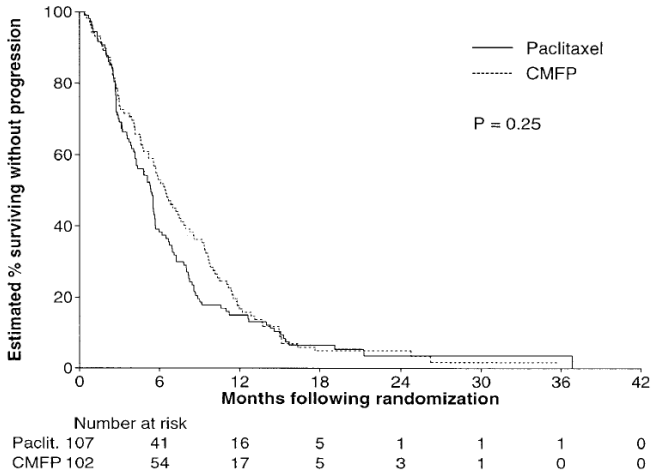


Fig 1. PFS on first-line treatment with paclitaxel (solid line) or with doxorubicin (dashed line).

We use digital software (http://www.digitizeit.de/) to read in the coordinates of the KM curves from the published graph. The R-code for the algorithm is then used to reconstruct the Kaplan-Meier data (Guyot P, 2012).

Figure 3-19 Progression free survival by treatment arm

Fig 2. Progression-free survival by treatment arm.



In addition, Kaplan Meier graph without number of subjects at different time points are published for the trial with regimen Gemcitabine plus Paclitaxel as shown in Figure 3-20 (Demiray M, 2005). Total number of disease progression was reported.

Figure 3-20 Kaplan Meier analysis of time to progression

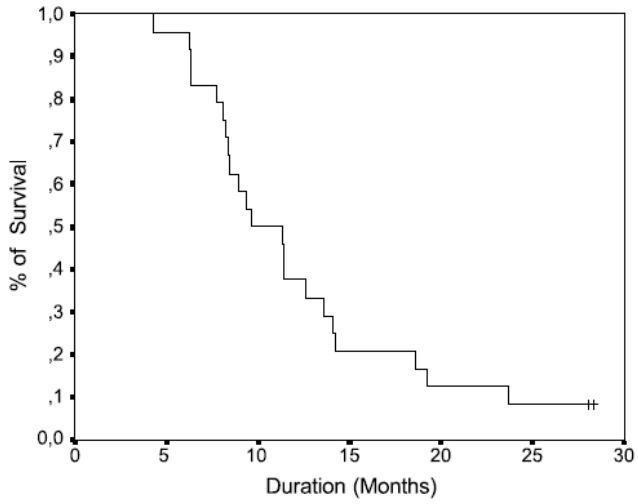


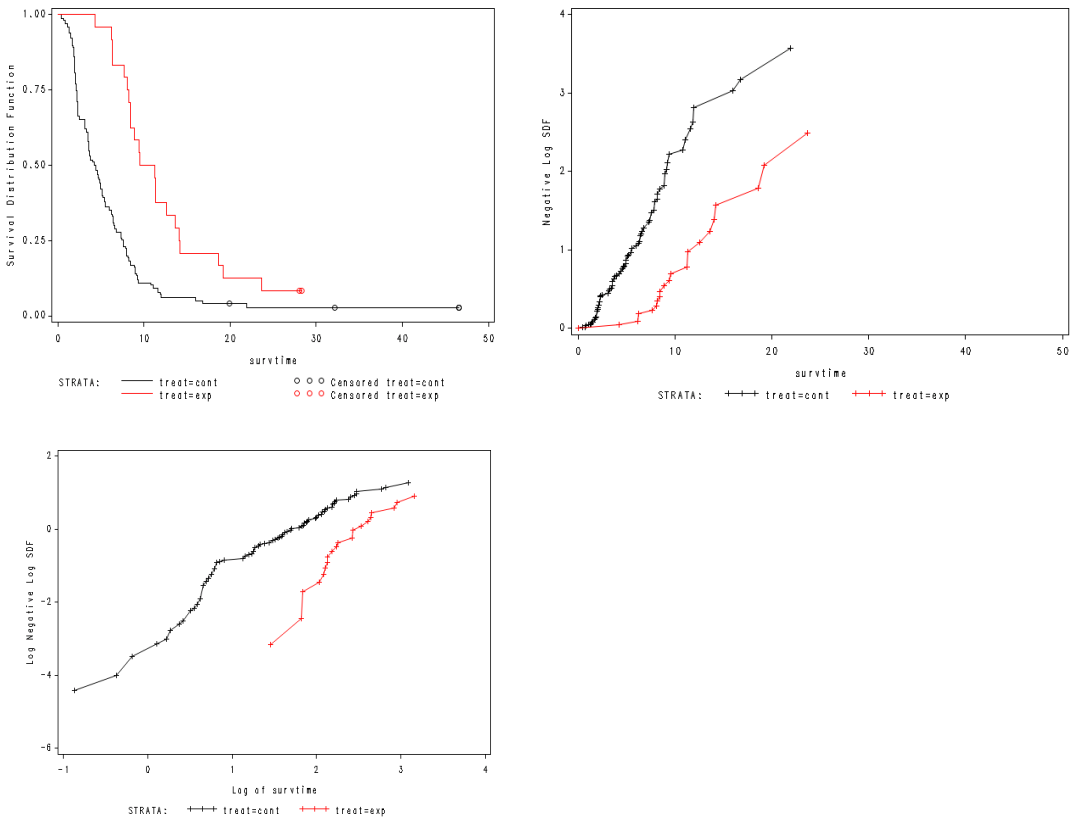
FIG. 1. Kaplan–Meier analysis of time to progression. Median TTP = 9.6 months.

Total 24 subjects were enrolled and evaluated in this trial. As shown in the Figure 3-20 two ticks cross the line are two or at least two subjects censored near the end of study. There are total 20 step-downs in the Kaplan Meier curve, in which each step-down

represents at least one occurrence of event of interest. Based on the information, it's not difficult to identify the two time points, at 6.28 months and 11.35 months respectively, when event happened on two subjects. Therefore, we are able to reconstruct the individual patient level data for time to event without the number of subjects at risk at different time points.

We combine the individual patient time to event data of experimental group with the data from the first trial with the control regimen. There are 24 subjects from experimental group and 166 subjects from control group. The Kaplan Meier graph of the two groups is shown in the Figure 3-21 below.

Figure 3-21 Kaplan Meier graph



When we graph the survival function versus survival time, if assume a constant hazard, $h(t) = \lambda$, which implies an exponential distribution of survival times, $s(t)=e^{-\lambda t}$, the -log(survival) versus the survival time should be a straight line; if the predictors satisfy the proportional hazard assumption, then the shapes of the curves should be basically the same, and the separation between the curves should remain proportional across analysis time. Similarly, the graph of the log(-log(survival)) versus log of survival time graph should result in parallel curves if the predictor is proportional. According to Figure 3-21 the survival probabilities by treatment strata, the two curves of -log(survival) versus the survival time are close to straight line from the beginning until 15 months for control group while it is flat for experimental group from beginning until 8 months then it's close to straight line until 15 months. The constant hazard assumption does not hold well, the proportional hazard assumption does not hold well either if we only have covariate of treatment regimen group in the model. Due to the lack of information for other potential covariates, we cannot model the survival with either exponential or proportional hazard model.

We test the proportional hazards assumption by generating the time dependent covariate, which is a function of survival time, and including it in the model.

Output 4-1 Statistics of Proportional Hazard Model

Type 3 Tests			
Effect	**DF**	**Wald Chi-Square**	**Pr > ChiSq**
Treatment	1	20.0085	<.0001
Treatment_Survtime	1	11.6095	0.0007

The time dependent covariate is highly significant, which is the strong evidence against proportional hazards.

We then fit the individual patient level time to event data into Piece-wise Exponential model with treatment group (Exponential and Control groups) and period (0 or 1) as covariates. The Piece-wise Exponential model is also a proportional hazard (PH) model as its basic hazard rate can be specified in the following way: $h(t,X)=h_0(t)e^{X\beta}$. The main difference is that the baseline hazard rate is allowed to vary in different time periods but remain constant within each period.

The whole period presented in Kaplan Meier graph is divided into two periods with cut point of 14 months which is the middle point of the whole period for experimental group in the preceding study. Period 0 is defined as time period from 0 to 14 months; period 1 is defined as time period from 14 months and after.

The tabulations reported number of events and person month by treatment group and period shown in Table 3-9.

Table 3-9 Advance breast cancer event of interest and exposure time by treatment group and period.

Treatment	Period	Event	Person Month	log(Person Month)
Control	0	156	861.32	6.75847
Control	1	5	144.19	4.97113
Experimental	0	17	249.72	5.52034
Experimental	1	5	48.18	3.87494

Since SAS GENMOD procedure fits Poisson regression models, we can use it to fit Piece-wise Exponential model. The Piecewise Exponential model can be written as

$$\log\mu_{ij}=\log t_{ij} + \alpha_j + \text{Treatment}_i\,\beta_1+\text{Period}_j\,\beta_2$$

where $\alpha_j = \log \lambda_j$, μ_{ij} is mean of Poisson event of interest observations, t_{ij} is the exposure time for subject i during interval j.

Output 3-2 shows the deviance of 8.379 on 1 degree of freedom from the Piece-wise Exponential model, which is significant. The deviance can be large when all appropriate covariates are not included in the model or when distribution assumptions are not correct or both. In the Poisson regression model, we have to assume variance of response equals to mean of response. When the response has greater variability than when expected, standard errors for the regression parameters will be too small.

Output 3-2 Goodness of Fit for the Piecewise Exponential model

Criteria For Assessing Goodness Of Fit			
Criterion	**DF**	**Value**	**Value/DF**
Deviance	1	8.3790	8.3790
Scaled Deviance	1	8.3790	8.3790
Pearson Chi-Square	1	12.1273	12.1273
Scaled Pearson X2	1	12.1273	12.1273
Log Likelihood		664.8470	
Full Log Likelihood		-13.4550	
AIC (smaller is better)		32.9099	

We adjust the standard errors to account for this extra variability in the response by multiplying the standard errors with over dispersion parameter.

Output 4-3 SAS output with over dispersion from Piece-wise Exponential model

Analysis Of Maximum Likelihood Parameter Estimates								
Parameter		**DF**	**Estimate**	**Standard Error**	**Wald 95% Confidence Limits**		**Wald Chi-Square**	**Pr > ChiSq**
Intercept		1	-3.5741	1.2967	-6.1156	-1.0326	7.60	0.0058
Treatment	0	1	0.7604	0.7916	-0.7911	2.3119	0.92	0.3368
Treatment	1	0	0.0000	0.0000	0.0000	0.0000	.	.

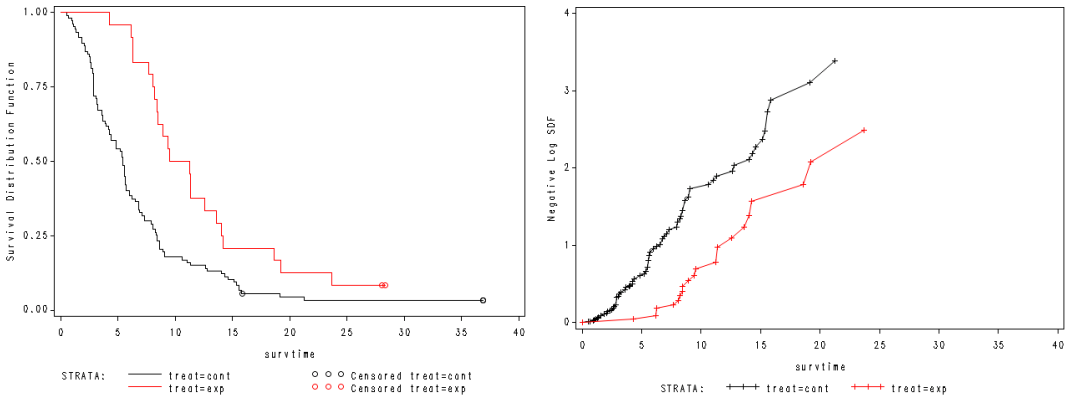
Period	0	1	1.0814	1.1327	-1.1386	3.3014	0.91	0.3397
Period	1	0	0.0000	0.0000	0.0000	0.0000	.	.
Scale		0	3.4824	0.0000	3.4824	3.4824		

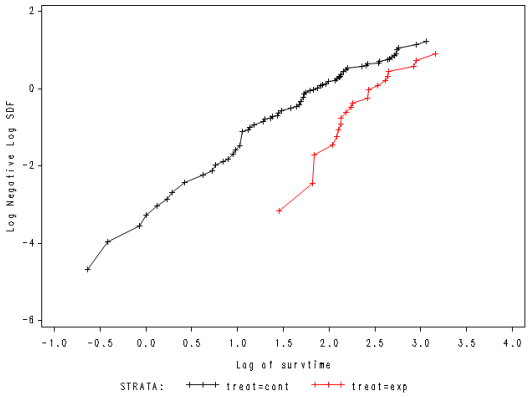
Note: The scale parameter was estimated by the square root of Pearson's Chi-Square/DOF.

The estimate of coefficient is 0.7604 with standard error of 0.7916 for the treatment effect. The standard error increases 3.48 times comparing with the model without adjustment of over dispersion. Assume normal distribution of natural logarithm of hazard ratio and apply the power formula in section 3.4.2, the predicted power is 0.73.

We then pool the individual patient time to event data of experimental group with the data from the second trial with the control regimen. There are 24 subjects from experimental group and 107 subjects from control group. The Kaplan Meier graph of the two groups is shown in the Figure 3-22 below.

Figure 3-22 Kaplan Meier graph





From the Kaplan Meier graph by treatment strata, the two curves of -log(survival) versus the survival time are close to straight line for control group while it is flat for experimental group from beginning until 8 months then it's close to straight line until 15, the proportional hazard assumption does not hold well since the distance between two log(-log(survival)) versus log of survival time curves is decreasing over the period if we only have covariate of treatment regimen group in the model. Due to the lack of information for other potential covariates, we cannot model the survival with the either exponential or proportional hazard model.

We test the proportional hazards assumption by generating the time dependent covariate, which is a function of survival time, and including it in the model. The time dependent covariate is highly significant, which is the strong evidence against proportional hazards.

Output 4-4 Statistics of Proportional Hazard Model

Type 3 Tests			
Effect	**DF**	**Wald Chi-Square**	**Pr > ChiSq**
Treatn	1	14.0135	0.0002
Treatn_survtime	1	8.2588	0.0041

We then fit the individual patient level time to event data into Piece-wise Exponential model with treatment and period as covariates.

The whole period reported in the clinical trial also is divided into two periods with cut point of 14 months which is the middle point of the whole period for experimental group in the preceding study. Period 0 is defined as period from 0 to 14 months; period 1 is period from 14 months and after.

Similarly, the Piecewise Exponential model can be written as

$$\log\mu_{ij}=\log t_{ij} + \alpha_j + \text{Treatment}_i\,\beta_1+\text{Period}_j\,\beta_2$$

where $\alpha_j = \log \lambda_j$, μ_{ij} is mean of Poisson death observations, t_{ij} is the exposure time for subject i during interval j.

The tabulations reported number of deaths and person month by treatment group and period shown in Table 3-10.

Table 3-10 Advance breast cancer event of interest and exposure time by treatment group and period

Treatment	Period	Event	Person Month	log(Person Month)
Control	0	659.140	93	6.49094
Control	1	91.285	10	4.51399
Experimental	0	249.720	17	5.52034
Experimental	1	48.180	5	3.87494

Output 3-5 shows the deviance of 1.1669 on 1 degree of freedom for the Piecewise Exponential model, which is insignificant, indicating that the Piecewise Exponential model provides an adequate fit for the breast cancer data.

Output 4-5 Goodness of fit for the Piece-wise Exponential model

Criteria For Assessing Goodness Of Fit			
Criterion	**DF**	**Value**	**Value/DF**
Deviance	1	1.1669	1.1669
Scaled Deviance	1	1.1669	1.1669
Pearson Chi-Square	1	1.2639	1.2639
Scaled Pearson X2	1	1.2639	1.2639
Log Likelihood		375.1860	
Full Log Likelihood		-9.9289	
AIC (smaller is better)		25.8578	
AICC (smaller is better)		.	
BIC (smaller is better)		24.0166	

Output 3-6 shows parameter estimates for Piecewise Exponential model with covariates of treatment and period. The parameter estimate of 0.6167, corresponding to a hazard ratio of $e^{0.6171}$=1.85, which is comparing the hazards of disease for control groups with experimental group. The standard error for the parameter estimate is 0.24.

Assume normal distribution of natural logarithm of hazard ratio and apply the power formula in section 4.4.2, the predicted power is 0.9.

Output 4-6 SAS output from Piece-wise Exponential model

Analysis of Maximum Likelihood Parameter Estimates								
Parameter		**DF**	**Estimate**	**Standard Error**	**Wald 95% Confidence Limits**		**Wald Chi-Square**	**Pr > ChiSq**
Intercept		1	-2.6733	0.3165	-3.2935	-2.0530	71.36	<.0001
Treatment	0	1	0.6167	0.2351	0.1559	1.0775	6.88	0.0087
Treatment	1	0	0.0000	0.0000	0.0000	0.0000	.	.
Period	0	1	0.0801	0.2755	-0.4599	0.6201	0.08	0.7713
Period	1	0	0.0000	0.0000	0.0000	0.0000	.	.
Scale		0	1.0000	0.0000	1.0000	1.0000		

Clinical Trial 2: Pegylated Liposomal Doxorubicin Plus Docetaxel Significantly Improves Time to Progression Without Additive Cardiotoxicity Compared With Docetaxel Monotherapy in Patients With Advanced Breast Cancer Previously Treated With Neoadjuvant-Adjuvant Anthracycline Therapy: Results From a Randomized Phase III Study (Sparano JA, 2009)

One prior clinical trial with regimen Pegylated Liposomal Doxorubicin plus Docetaxel (Bontenbal M, 2005) and one prior clinical trials with regimen Docetaxel (Chan S, 1999) were referred in the publication of the later phase III clinical trial comparing pegylated liposomal Doxorubicin plus Docetaxel versus Docetaxel monotherapy in patients with metastatic breast cancer previously treated with neoadjuvant-adjuvant anthracycline therapy. The information of total number of events and median survival times are available in the publications of those two clinical trials, which make it possible to apply the Exponential-Gamma model and the Weibull-Inverse Gamma model into estimation of the expected power.

The Kaplan Meier graph with number of subjects at different time points are also available and reported in these two trials. Below Figure 3-23 and Figure 3-24 are the two Kaplan Meier graphs for those two studies.

Figure 3-23 Kaplan-Meier curve for time to progression. AT, Doxorubicin and Docetaxel; FAC, Fluorouracil, Doxorubicin, and Cyclophosphamide (Bontenbal M, 2005)

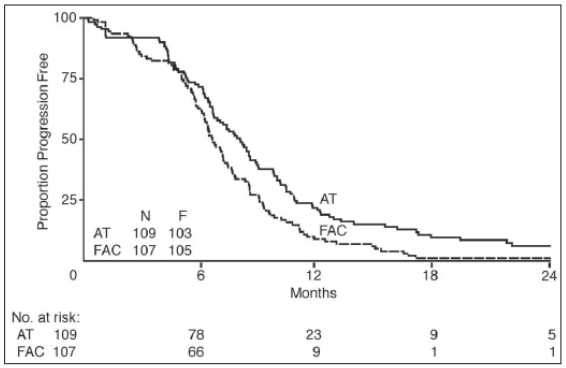
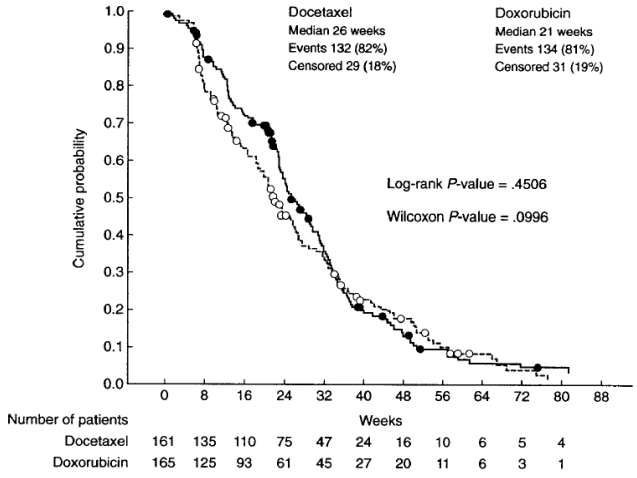


Fig 1. Kaplan-Meier curve for time to progression. AT, doxorubicin and docetaxel; FAC, fluorouracil, doxorubicin, and cyclophosphamide.

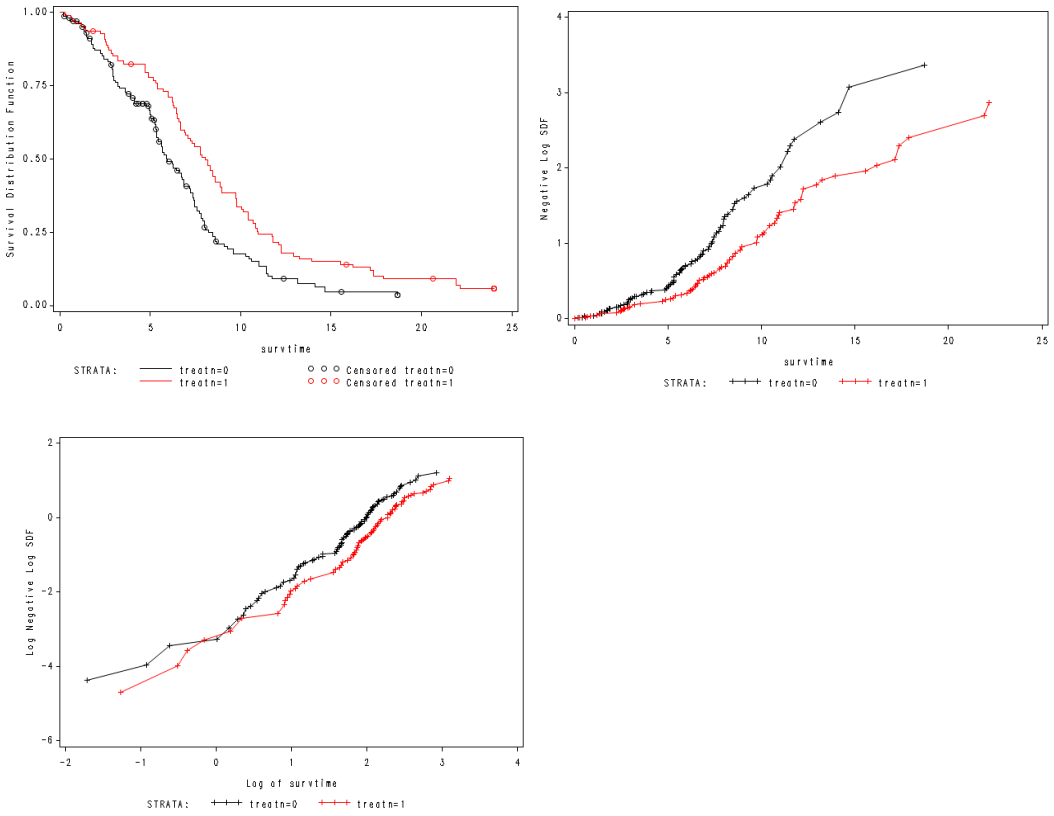
Figure 3-24 Kaplan-Meier estimate of cumulative probability of remaining free from disease progression in each treatment group (ITT population) (docetaxel, n=161, dark circle, doxorubicin, n=165, empty circle) (Chan S, 1999)



Digital software is utilized to read in the coordinates of the KM curves from the published graph. The R-code for the algorithm is then used to reconstruct the Kaplan-Meier data (Guyot P, 2012).

We combine the individual patient time to event data of experimental group with the control regimen. There are 109 subjects from experimental group and 161 subjects from control group. The Kaplan Meier graph of the two groups is shown in the Figure 3-25 below.

Figure 3-25 Kaplan Meier graph



As shown in Figure 3-25 the survival rate is presented by treatment strata, the two curves of -log(survival) versus the survival time are close to straight line even though vibration is getting bigger after 12 months. The proportional hazard assumption seems hold well since the distance between two log(-log(survival)) versus log of survival time curves are parallel except the two lines cross in a very short period of duration.

We then test the proportional hazards assumption by including time dependent covariate into the model.

Output 4-7 Ana lysis of Maximum Likelihood Estimates

Type 3 Tests			
Effect	**DF**	**Wald Chi-Square**	**Pr > ChiSq**
Treatn	1	5.9175	0.0150
Treatn_survtime	1	0.8910	0.3452

The time dependent covariate is insignificant, indicating the proportional hazard model with treatment group as covariate can fit the data. The Cox model is written as

$\log h_i(t) = \alpha(t) + \text{Treatment}_i \beta_1$

Where the baseline hazard function $\alpha(t) = \log h_0(t)$ is left unspecified.

Output 4-8 Analysis of Maximum Likelihood Estimates

Analysis of Maximum Likelihood Estimates										
Parameter		**DF**	**Parameter Estimate**	**Standard Error**	**Chi-Square**	**Pr > ChiSq**	**Hazard Ratio**	**95% Hazard Ratio Confidence Limits**		**Label**
Treatn	**0**	1	0.46264	0.13634	11.5140	0.0007	1.588	1.216	2.075	treatn 0

The SAS output from fitting proportion hazard model with covariate of treatment group indicate the parameter estimate is 0.46, which can transfer to the hazard ratio between control group over experimental group is $e^{0.46}$=1.588. The standard error for the parameter estimate is 0.136.

We can also fit the data with Piecewise Exponential model with treatment and period as covariates. The whole period reported in the clinical trial also is divided into two periods with cut point of 9 months which is the middle point of the whole period for experimental group in the preceding study. Period 0 is defined as period from 0 to 9 months; period 1 is period from 9 months and after.

The tabulations reported number of event and person month by treatment group and period shown in Table 3-11.

The SAS procedure Genmod was used to fit the data with Piecewise Exponential model with covariates of treatment group and period. Similarly the Piecewise Exponential model can be written as

$$\log\mu_{ij}=\log t_{ij} + \alpha_j + \text{Treatment}_{ij}\beta_1+\text{Period}_{ij}\beta_2$$

where $\alpha_j = \log \lambda_j$, μ_{ij} is mean of Poisson death observations, t_{ij} is the exposure time for subject i during interval j.

Table 3-11 Advance breast cancer event of interest and exposure time by treatment group and period

Treatment	Period	Events	Person Month	Log(Person Month)
Control	0	111	851.709	6.74724
Control	1	20	95.855	4.56283
Experimental	0	66	739.570	6.60607
Experimental	1	34	233.775	5.45436

According to the SAS output below, the deviance for this model is 0.0031 with df=1, which is insignificant, and indicate the model fit the breast cancer data. The maximum Likelihood parameter estimate for treatment group is 0.3745 with standard error of 0.136, which can transfer to the hazard ratio between control group over experimental group is $e^{0.3745}$=1.4543.

Output 4-9 Creteria for Assessing Goodness of Fit

Criteria For Assessing Goodness Of Fit			
Criterion	**DF**	**Value**	**Value/DF**
Deviance	1	0.0031	0.0031
Scaled Deviance	1	0.0031	0.0031
Pearson Chi-Square	1	0.0031	0.0031

Scaled Pearson X2	1	0.0031	0.0031
Log Likelihood		748.0844	
Full Log Likelihood		-11.3966	
AIC (smaller is better)		28.7931	

Output 4-10 Analysis of Maximum Likelihood Parameter Estimates

Analysis Of Maximum Likelihood Parameter Estimates								
Parameter		**DF**	**Estimate**	**Standard Error**	**Wald 95% Confidence Limits**		**Wald Chi-Square**	**Pr > ChiSq**
Intercept		1	-1.9331	0.1453	-2.2178	-1.6484	177.10	<.0001
Treatn	0	1	0.3745	0.1360	0.1080	0.6411	7.59	0.0059
Treatn	1	0	0.0000	0.0000	0.0000	0.0000	.	.
Period	0	1	-0.4807	0.1592	-0.7928	-0.1687	9.12	0.0025
Period	1	0	0.0000	0.0000	0.0000	0.0000	.	.
Scale		0	1.0000	0.0000	1.0000	1.0000		

Both treatment group and period are significant in the Piece-wise model. Assume normal distribution of natural logarithm of hazard ratio and apply the power formula in section 4.4.2, the predicted power is 0.726 for Piecewise exponential model and 0.856 from proportional Cox model.

3.4.3 Computation of expected power and results

In this section we applied an alternative method to estimate the predicted power given the full information of prior trial, which include Kaplan-Meier curve with number at risk at different time points and total number of events of interests during the whole study and follow up period. In the first step, we generate the individual subject patient level data, in another word, for each subject we derive the time to event or censoring, and the event indication (event or censoring). However, this is not truly the subject level data in a sense, except for the treatment group each subject belongs to, we are not able to obtain other

information of patient characteristics, which likely will have impact on the survival prognostics of subject. Since those information come from randomized clinical trials, we hope those characteristics can distribute between the experimental and control groups evenly, so that the hazard ratio we calculate is still valid event we do not control other characters in addition to treatment and time to event.

The calculated power for the first phase III study is 0.73 and 0.9, which are comparable with the calculated power of 0.78 from the Exponential-Gamma method, and calculated power of 0.75 from the Weibull-Inverse Gamma method. The second phase III study the derived power from Piecewise exponential model and Cox model is 0.73 and 0.86, respectively, both of those are higher than the derived power of 0.59 from the Exponential-Gamma method, and 0.66 from the Weibull-Inverse Gamma method . These high values of power results, at the minimum, for these two studies, support the claim that the powers calculated with multiple methods have indicated consistently a potential successful later phase III trial. Nevertheless, because of lacking the full information for other studies, we are not able to further apply this approach to more phase III studies, especially phase III studies with insignificant results in order to evaluate its validity.

4 Comparison of results from different methods with endpoint of progression free survival

In this report, a power model is implemented to predict the likelihood of success of future potential phase III studies based on the information from previous historical phase II or phase III studies. We apply three statistical methods to model either the hazard or the hazard ratio between experimental and control regimen groups to estimate the expected power through the power model. The proposed methods include the Exponential-Gamma model, the Weibull-Inverse Gamma Model and the Piece-wise Exponential model. The first two methods are Bayesian conjugate statistical model.

Different kinds of survival hazard models may be obtained by making different assumptions about the hazard function. In the Exponential-Gamma model, the exponential model assumes a constant hazard rate, while in the Weibull-Inverse Gamma model, the Weibull model assumes the risk monotonically increases or decreases over time. In the Piece-wise Exponential model, the hazard is not constant over time, but assumed to be the same in each period. This assumption allows us to estimate changes in the hazard rate over different periods.

Both exponential model and Weibull model are parametric proportional hazard models. Parametric hazard model for time to event data works best if the model fit the data. However, parametric models are not flexible to accommodate the various patterns of non-constant hazards due to the restriction of the distribution assumptions. Therefore, Piece-wise Exponential approach can be used to approximate nonparametric models to better fit the real time to event data, and retain assumption of parametric distribution of data. It is a flexible, semi-parametric model. In this model the proportional hazard

assumption holds at least over short periods of time such that all time-varying effects can be treated as piecewise constant.

We compared the results of these methods to predict the likelihood of success for future potential phase III clinical trials with population of advanced breast cancer patients in terms of prolonging the disease progression free time. All methods compared are implemented in SAS or R software.

The expected power calculated from above three statistical methods are shown in Table 4-1. We are able to obtain the expected power for all of the six trials listed with the Exponential-Gamma model, for five of the six trials with the Weibull-Inverse Gamma model, and only two studies with the Piece-wise Exponential model due to limit of information available from literature.

For study comparing Paclitaxel plus Gemcitabine with Paclitaxel Monotherapy, the expected powers are 0.78, 0.75, and 0.73/0.9 with the Exponential-Gamma model, the Weibull-Inverse Gamma model, and the Piece-wise Exponential model, respectively. The values of all the estimated expected power are high and consistent cross the three methods. It's the highest among all the powers with the Exponential-Gamma model and second to highest with the Weibull-Inverse Gamma model. The values of power are consistent with the subsequent successful phase III study.

For study comparing Gemcitabine plus Docetaxel with Capecitabine plus Docetaxel, the expected powers are 0.29 with the Exponential-Gamma model and 0.62 with the Weibull-Inverse Gamma model. The result with the Piece-wise Exponential model is not available. The two values are not close from two different methods. However, it's the lowest value among the powers calculated from the Exponential-Gamma model and

second to lowest value from the Weibull-Inverse Gamma model for all the studies. The later phase III study is not successful, which is consistent with the rank of the power values.

For study comparing PLD + Docetaxel with Docetaxel, the expected powers are 0.59, 0.66, and 0.73 with the Exponential-Gamma model, Weibull-Inverse Gamma model, and the Piece-wise Exponential model, respectively. The values are high and consistent cross the three methods. It's the third to the highest power from both the Exponential-Gamma model and the Weibull-Inverse Gamma model of all the studies. The values of power are consistent with the later successful phase III study.

For study comparing Epirubicin and Paclitaxel with and without Capecitabine, the expected power is only available with the Exponential-Gamma model. It is 0.42 and around the borderline 0.41 to 0.42 with which seems to differentiate a successful study from an unsuccessful one from the Exponential-Gamma model. The later phase III study is not successful.

For study comparing Ixabepilone plus Capecitabine with Capecitabine, and study comparing Arzoxifene with Tamoxifen, the expected power is only available with the Exponential-Gamma model and the Weibull-Inverse Gamma model methods. With the Exponential-Gamma model method, the expected power is 0.41 and 0.62 for the first and second studies, while with the Weibull-Inverse Gamma model method, the expected power is 0.98 and 0.51 for the two studies. The subsequent study comparing Ixabepilone plus Capecitabine with Capecitabine is successful, while the subsequent study comparing Arzoxifene with Tamoxifen is a failure. We will expect the estimated power should be higher for the successful study, and lower for the unsuccessful study, but the results from

the Exponential-Gamma model method contradict with this expectation. On the other hand the results from the Weibull-Inverse Gamma model method are consistent with the subsequent study results and resolve the problem of the Exponential-Gamma model method.

Other than the last two studies, the values of expected powers estimated from Exponential-Gamma model are consistent with the later phase III trial results.

Among five studies with expected power available from Weibull-Inverse Gamma model, the highest three values of expected power 0.98, 0.75 and 0.66 are for three experimental regimens in fact approved to lead to successful PFS outcomes in the subsequent phase III study. The other two lowest values of expected power 0.51 and 0.62 are for other two studies in which experimental regimen failed to prolong the time to disease progression comparing with control regimen.

As to the Piece-wise Exponential method, expected power is only available for two studies due to limit of data available. The expected powers are 0.73/0.9 and 0.73, respectively, which are consistent with the successful results from later actual phase III studies.

We also study the impact of number of events of progression observed in the preceding phase II or phase III trials on the expected power. As we have illustrated in Figure 4-1 to 4-11, the expected power to predict whether a regimen that appears promising in a phase II trial will be successful in a phase III trial depends on the number of events observed in the previous phase II or phase III trial and the median survival observed. In each of the eleven figures, the median survival time is set to be fixed.

Figure 4-1 Expected power by number of progression events in preceding historical study comparing Paclitaxel plus Gemcitabine with Paclitaxel Monotherapy with Exponential-Gamma model

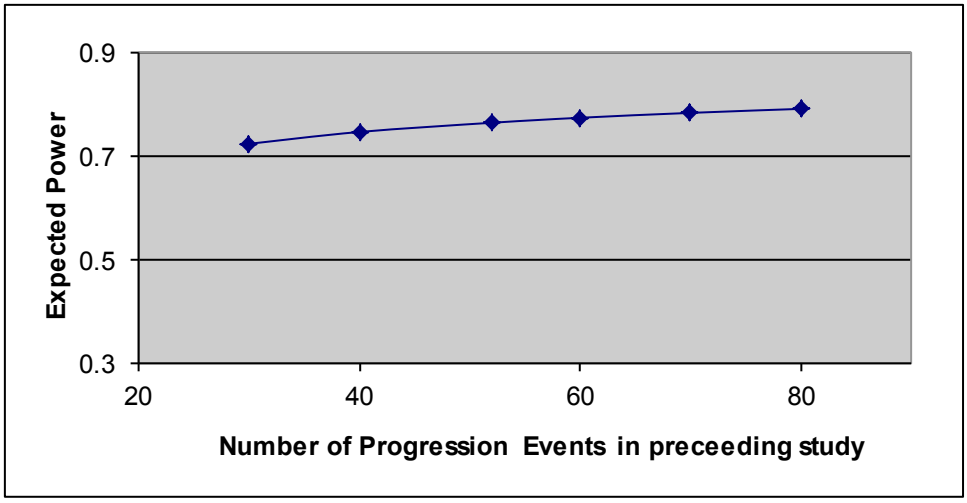


Figure 4-2 Expected power by number of progression events in preceding historical study comparing Gemcitabine plus Docetaxel with Capecitabine plus Docetaxel with Exponential-Gamma model

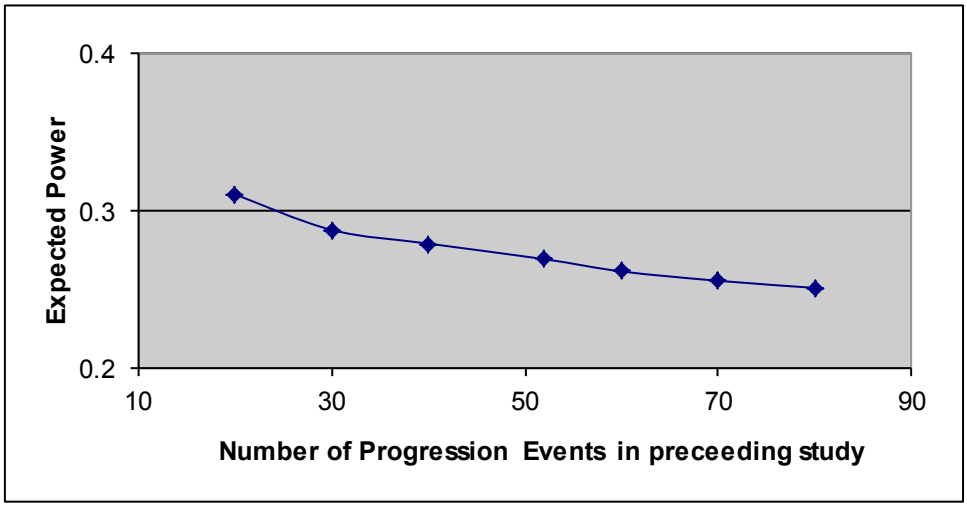


Figure 4-3 Expected power by number of progression events in preceding historical study comparing PLD + Docetaxel with Docetaxel with Exponential-Gamma model

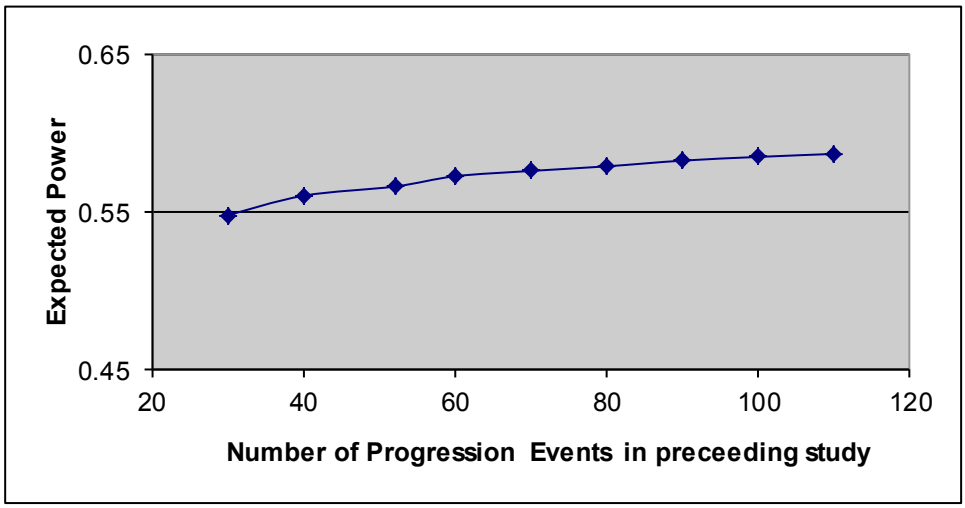


Figure 4-4 Expected power by number of progression events in preceding historical study comparing Epirubicin and Paclitaxel with and without Capecitabine with Exponential-Gamma model

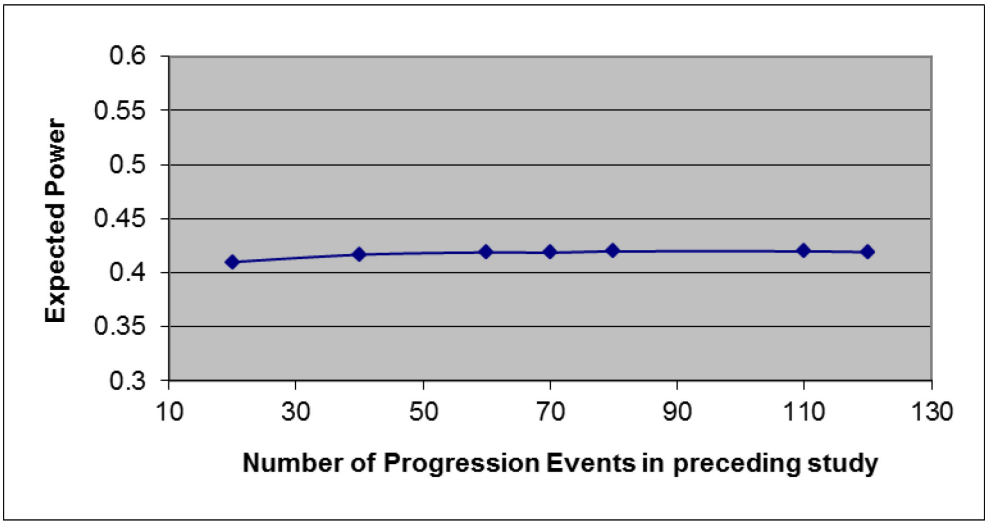


Figure 4-5 Expected power by number of progression events in preceding historical study comparing Ixabepilone plus Capecitabine with Capecitabine with Exponential-Gamma model

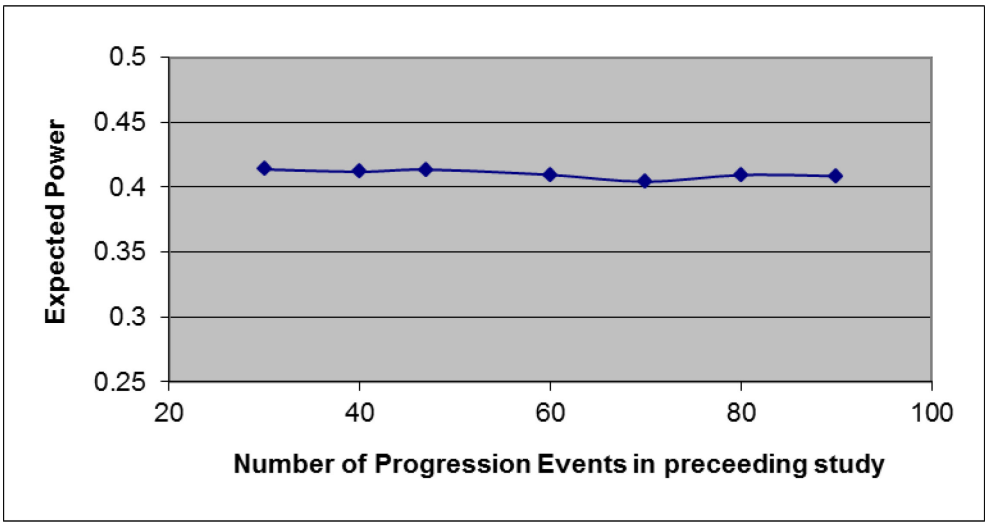


Figure 4-6 Expected power by number of progression events in preceding historical study comparing Arzoxifene with Tamoxifen with Exponential-Gamma model

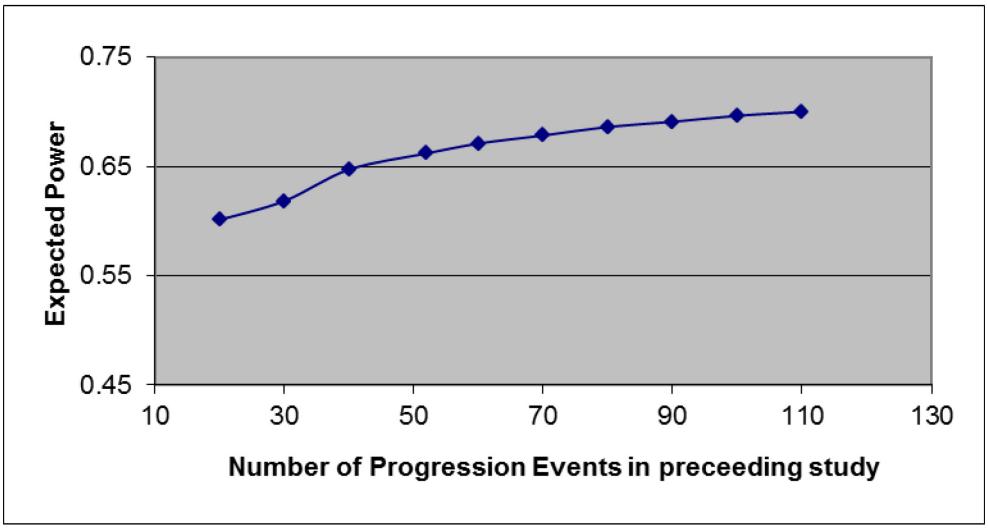


Figure 4-7 Expected power by number of progression events in preceding historical study comparing Paclitaxel plus Gemcitabine with Paclitaxel Monotherapy with Weibull-Inverse Gamma model

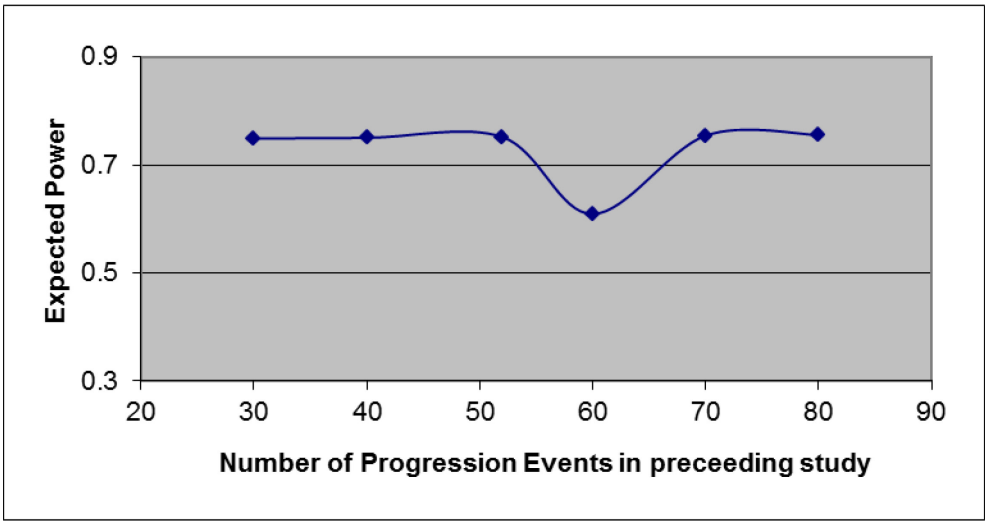


Figure 4-8 Expected power by number of progression events in preceding historical study comparing Gemcitabine plus Docetaxel with Capecitabine plus Docetaxel with Weibull-Inverse Gamma model

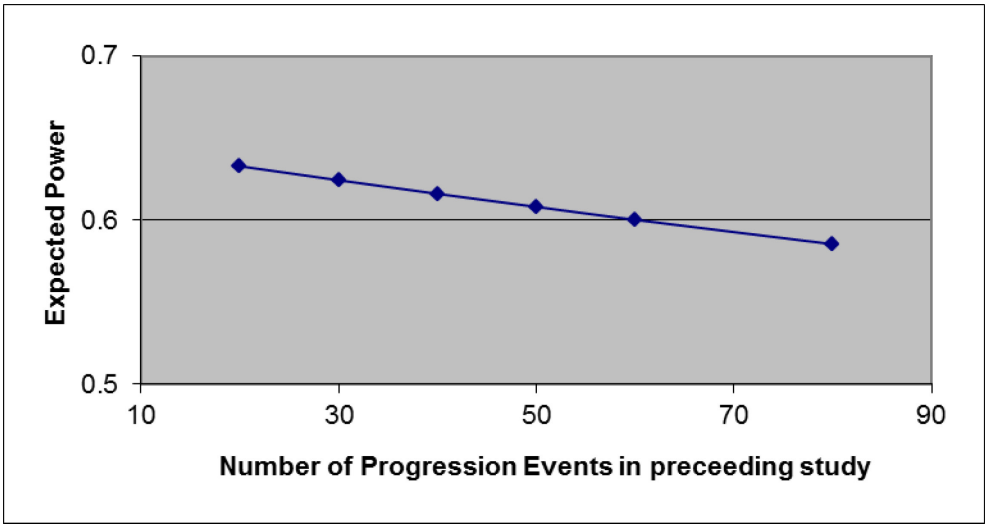


Figure 4-9 Expected power by number of progression events in preceding historical study comparing PLD + Docetaxel with Docetaxel with Weibull-Inverse Gamma model

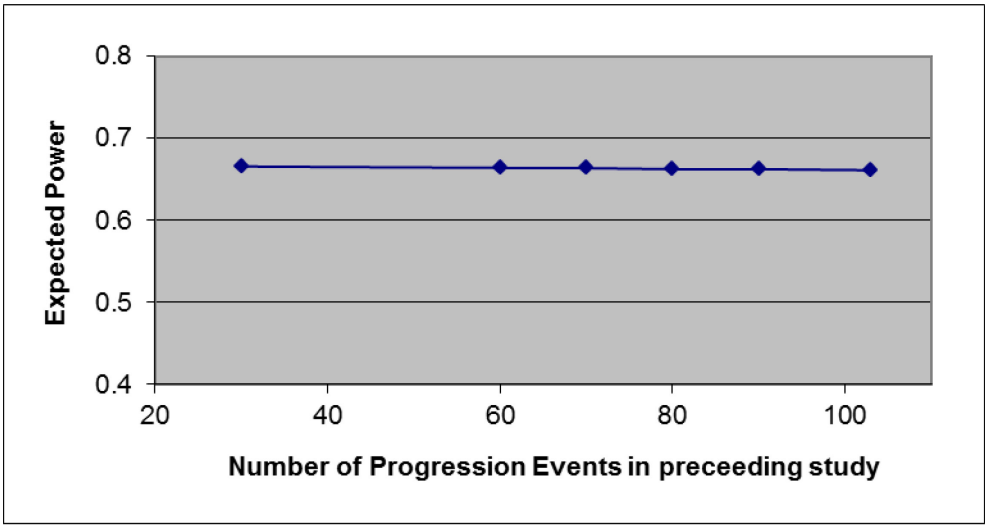


Figure 4-10 Expected power by number of progression events in preceding historical study comparing Ixabepilone plus Capecitabine with Capecitabine with Weibull-Inverse Gamma model

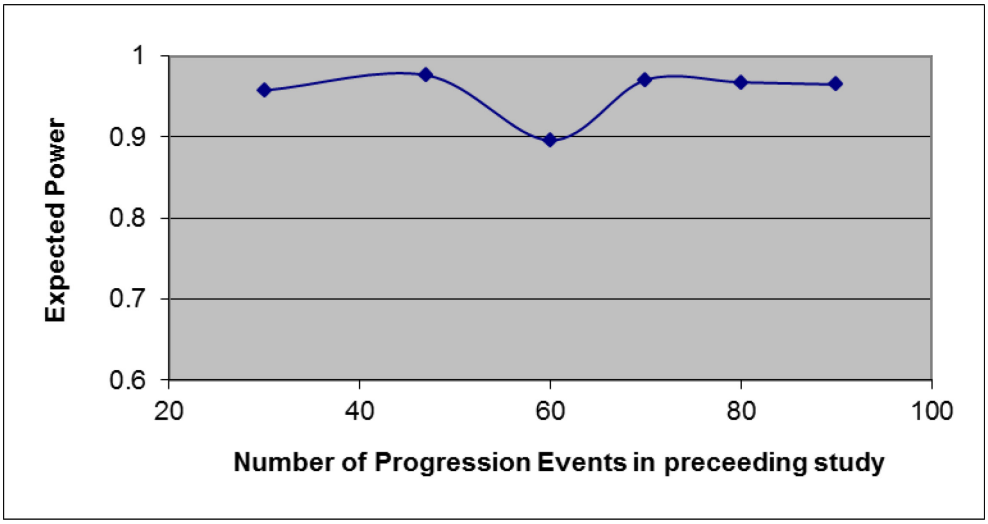
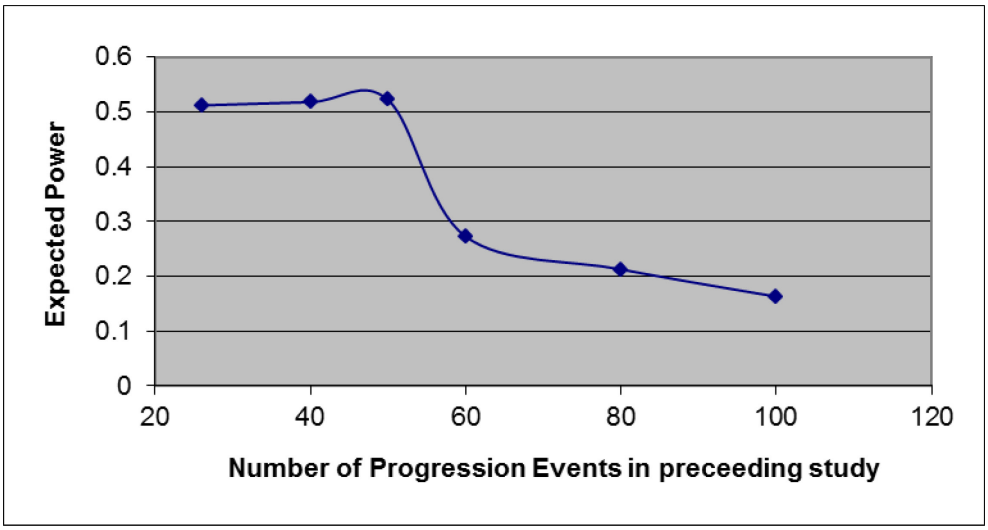


Figure 4-11 Expected power by number of progression events in preceding historical study comparing Arzoxifene with Tamoxifen with Weibull-Inverse Gamma model



For both Exponential-Gamma and Weibull-Inverse Gamma model, the expected power does not change dramatically with different number of events expected to be observed giving the fixed median survival time except in Figure 4-11, the expected power decrease more than 20% when number of events increase from 50 to 60.

Table 4-1 List of expected power for future phase III studies, previous historical phase II or phase III studies, and results of PFS in subsequent actual phase III trials from three models

Std. #	Authors of phase II or phase III/phase III	Experiment Vs. Control Treatment in Phase III	Observed Preceding Study		Power †			Subsequent Phase III		
			Numer of Events/ Number of Patients	Median PFS (Month) of Experimental	Exponential -Gamma	Weibull -Inverse Gamma	Piece-wise Exponential	Median PFS Month (Experimental vs. Control)	Percentage Increase in Median PFS Time	p-Value
1	**Allouache D, et al; Demiray M, et al / Albain KS, et al**	Paclitaxel plus Gemcitabine with Paclitaxel Monotherapy	31/35; 21/24	7.2 9.6	0.78	0.75	0.73/0.9	5.9/3.9	0.51	0.0005
2	**Fountzilas G, et al.; /Chan S, et al**	Gemcitabine plus Docetaxel with Capecitabine plus Docetaxel	30/39	7	0.29	0.62	NA	8.05/7.98	0.009	0.121*
3	**Bontenbal M, et al. /Sparano JA, et al**	PLD + Docetaxel with Docetaxel	103/109	8	0.59	0.66	0.73	9.8/7	0.4	0.000001
4	**Mansutti M, et al /Hatschek T, et al**	Epirubicin and Paclitaxel with and without Capecitabine	110/136	12.3	0.42	NA	NA	12.40/10.80	0.15	0.84*
5	**Bunnell CA, et al /Thomas ES , et al**	Ixabepilone plus Capecitabine with Capecitabine	47/50	3.8	0.41	0.98	NA	5.8/4.2	0.38	0.0003*
6	**Baselga J, et al /Deshmane V, et al**	Arzoxifene with Tamoxifen	26/46	10.7	0.62	0.51	NA	4/7.5	-0.46	0.007‡*

† The power is calculated based on total number of events corresponding to 227 events to detect 54% increase in median PFS time.
‡ p-value is over 0.5 since the data is in the direction opposite that specified by the test.
* PFS is primary efficacy endpoint in the phase III study.

5 Discussion

5.1 Progression Free Survival

In Chen et al's Extensive-Stage Small-Cell Lung Cancer (SCLC) paper and Freidlin et al's Extensive-Stage Non-Small-Cell Lung Cancer (NSCLC) paper, the statistical model calculating the expected power was based on the distributions of survival for both experimental and control regimen groups (Chen TT, 2000; Freidlin B, 2003). The distributions of survival for the control and experimental regimen groups were derived based on the number of event of interest and the median survival time of the control group from previous phase III or phase II studies, which were identified through the search of National Cancer Institute Cancer Therapy Evaluation Program database. Their analysis indicated that an expected power of greater than 0.55 for a particular phase II study appears to be a reasonable reference point toward estimating that the regimen is likely to statistically significantly prolong survival when compared with standard treatment in a phase III trial.

In addition in Chen et al's Extensive-Stage Small-Cell Lung Cancer (SCLC) paper and Freidlin et al's Extensive-Stage Non-Small-Cell Lung Cancer (NSCLC) paper, current standard treatment concept was used to model the distribution of hazard for the control group. The distribution of hazard for control regimen was derived based on the survival information of patients with control regimen in all available historical phase III trials. As a result it assumed that the underlining distributions of hazards for control regimen in all SCLC or NSCLC trials are the same. However, when we introduced the same concept into the advanced breast cancer trials, the results are various. For example, for the trials of comparing two first-line chemotherapies and the trials of comparing two second-line

chemotherapies, the distributions of hazard for the first-line and second-line control regimens are totally different. If we compare a promising second-line chemotherapy regimen with a first-line standard control regimen, it's possible that the power will be low since we used an optimistic regimen as control. Therefore the PFS data from preceding phase II or III studies of the control regimen arms are used to model the hazard distribution of control group individually for each of six subsequent phase III studies in this report for advanced breast cancer.

From our analysis, an expected power of greater than 0.59 for a particular investigation regimen with the Exponential-Gamma model maybe a reasonable reference point toward estimating that the regimen is likely to statistically significantly prolong PFS time when compared with individual control regimen in a phase III trial. However, the values of expected power under 0.59 still belongs to a grey area instead of a definite cut-off since this cut-off is data driven. The expected power of 0.42 was derived for a failed subsequent phase III trial (Hatschek T, 2012), while the expected power of 0.41 was obtained for a successful subsequent phase III trial (Thomas ES, 2007). In the phase III study comparing regimen Arzoxifene with Tamoxifen, the expected power is 0.62, which contradicted with the actual unsuccessful study result (Deshmane V, 2007).

With the Weibull-Inverse Gamma model, we resolved the above problem with the Experimental-Gamma model. Instead of considering the hazard rate as constant, Weibull hazard model allows more flexible hazard modelling. The calculated expected powers from Weibull-Inverse Gamma model are completely consistent with the phase III results with a cut-off between 0.62 and 0.66.

In addition, we also implement the semi-parametric Piece-wise Exponential model, the calculated expected powers are consistent with the phase III actual results. However, since the data are very limited, i.e., data are available for only two successful phase III trials we cannot evaluate the validity of this model, nor provide suggestion of cut-off value of expected power for later phase III studies.

In the future, we can apply those three statistical models together into more studies to validate the usefulness of the models in predicting the likelihood of future successful phase III trials. Based on our experience in advanced breast cancer, an expected power of greater than 0.59 with the Exponential-Gamma model, and of greater than 0.64 with the Weibull-Inverse Gamma model provide a reasonable base to proceed to a phase III study.

Since we are only able to find six phase III randomized trials out of total 57 phase III trials for advanced breast cancer population, in which PFS information (number of event of progression and median progression time) from the historical preceding studies are available, it bring up the issue that how the preceding phase II studies should be designed at first place so that it can provide more information to advise whether or not to advance to a bigger phase III trial. Besides, as Chen et al mentioned in their original paper that the use of historical data to predict the efficacy of a particular treatment poses several problems. Differences in patient selection, treatment regimens, and supportive care are unaccounted for when a current treatment regimen is compared with historical data based on previous regimens (Chen TT, 2000). Among the six subsequent phase III studies in this report, the patient population was compared and considered as no big differences from the preceding study patient population.

The literature search and analysis revealed that the go and no go decision making involves many considerations, such as importance of the target mechanism of action, pharmacokinetic properties, market size, competition, surrogates for response, and activity and toxicity in phase I and II trials, etc. It relies on preclinical data and early clinical information. However, majority of the breast cancer phase II studies are single arm studies with response rates as primary endpoint. Thus, a greater investment in phase II studies may be required, with bigger sample size, randomized control arm, and longer follow up time to allow comparison between different arms and make the time to progression and progression free information available. The improvement of the phase II design will lead to more informative decision making on advancing agent to phase III study, and ultimately save money on the mistakes. The models can be applied prospectively in assisting the go and no go decision to advance to a phase III trial for a promising regimen.

In a recent published phase II clinical trial, the cyclin-dependent kinase 4/6 inhibitor palbociclib in combination with letrozole was compared with letrozole alone as first-line treatment of oestrogen receptor-positive, HER2-negative, advanced breast cancer (Finn RR, 2015). In this open-label, randomised phase II study, postmenopausal women with advanced oestrogen receptorpositive and HER2-negative breast cancer who had not received any systemic treatment for their advanced disease were enrolled in two separate cohorts. At the time of the final analysis, 41 progression free survival events had occurred in the palbociclib plus letrozole group and 59 in the letrozole group. Median progression free survival was 10.2 months (95% CI 5.7-12.6) for the letrozole group and 20.2 months (95% CI 13.8-27.5) for the palbociclib plus letrozole group (HR 0.488, 95% CI 0.319-

0.718; one-sided p=0.0004). A phase III, double blind, placebo-controlled study (NCT01740427) in a similar patient population (n=650) with the aim of confirming the present phase 2 findings is now fully enrolled and ongoing.

We apply both the Exponential-Gamma model and the Weibull-Inverse Gamma model into this study to estimate the likelihood of success for the phase III study. The information of experimental regimen palbociclib plus letrozole group and control regimen letrozole group are extracted from the same phase II clinical trial (Finn RR, 2015). Since in this phase II studies, the two regimens have already been compared directly with individual patient data and hazard ratio is available, we will not apply the third method into this experimental regimen.

The calculated expected power is 0.95 with the Exponential-Gamma model and 0.80 with the Weibull-Inverse Gamma model. Comparing with the suggested cutoffs for the two methods, which are 0.59 and 0.64, respectively, we will expect that the phase III trial will be successful.

5.2 Overall Survival

Overall survival, as a universally accepted direct measurement of benefit for clinical trials of oncology, is also an important endpoint to evaluate the effectiveness of regimens in advanced breast cancer. In addition to PFS, the expected power model is also applied to the endpoint of OS in advanced breast cancer to evaluate the validity of the model and predict the likelihood of success in the subsequent phase III trial.

Literatures of Phase III trials with patients in advanced breast cancer are searched first, then the previous phase II or phase III trials referred in the subsequent phase III trials which lead to the phase III trials are identified. The power model is built upon the

survival information extracted from preceding phase II or phase III clinical trials. The retrospectively calculated expected powers are then compared with the actual phase III trial results to evaluate whether or not the expected power model can help predicting the likelihood of successful phase III trials.

A search of randomized Phase III clinical trial for advanced breast cancer from PubMed reveals that there are 57 trials with publication date from 1990 to 2012, in which 12 trials from 1990 to 1999, and 45 from 2000 to 2012. The search from PubMed using species of humans and article types of clinical trial and randomized controlled trial with key words of metastatic breast cancer, advanced breast cancer, chemotherapy, endocrine therapy and phase III. Trials with cross-over design in the nature of design, without reaching median time for progression free survival, or overall survival, interim analysis, early terminated, with objective solely for safety purpose or quality of life, same regimen with different formulation or dosage levels are excluded.

The search above find out that total 47 trials have overall survival as one of the efficacy endpoints, in which only among 12 the experimental arm show a statistically significant survival advantage relative to the control arm. The rate of success was 26% (95% confidence interval [CI], 13% to 39%). The median overall survival time differences comparing experimental regimen with control regimen in those twelve trials range from 0.92 to 8 months, with median of median overall survival time of 2.7 months. This search and review also show that among the majority of phase III randomized studies of advanced breast cancer, the experimental regimen is not proved to be superior to the control regimen. The median survival of patients treated on the control arm was 21.5 months for studies started during the period of 1990 through 1999; it was 20.6 months for

studies started during the period from 2000 through 2012, with p-value of 0.51 in t-test. The median survival time is 20.75 months for control group for all the 47 trials during the whole period.

Among the 12 studies with successful overall survival results, the hazard ratios of overall survival and 95% confidence interval were published for 6 studies. We integrated the different treatment effects from the 6 studies. Since the sample size, patient selection and interventions are different for each study, we assume that there is a distribution of treatment effects to incorporate the individual differences, while still considered the effects are related to each other. Random effect model in the Meta data analysis is applied to the analysis of the hazard ratios of overall survival. Comprehensive Meta-Analysis (CMA) 2.0 is used to compute the effect size and the variance (Borenstein M, 2009). The summary effect size measured in hazard ratio is 0.756 with 95% confidence interval of [0.691, 0.827], which corresponds to 32% increase in median overall survival time. According to the power computation section 2.3.1.3 formula, the subsequent study need to observe 537 deaths in order to be adequately powered (90%) to detect the 32% increase in median survival time.

Figure 5-1 Impact of treatment on death by study hazard ratio and 95% CL

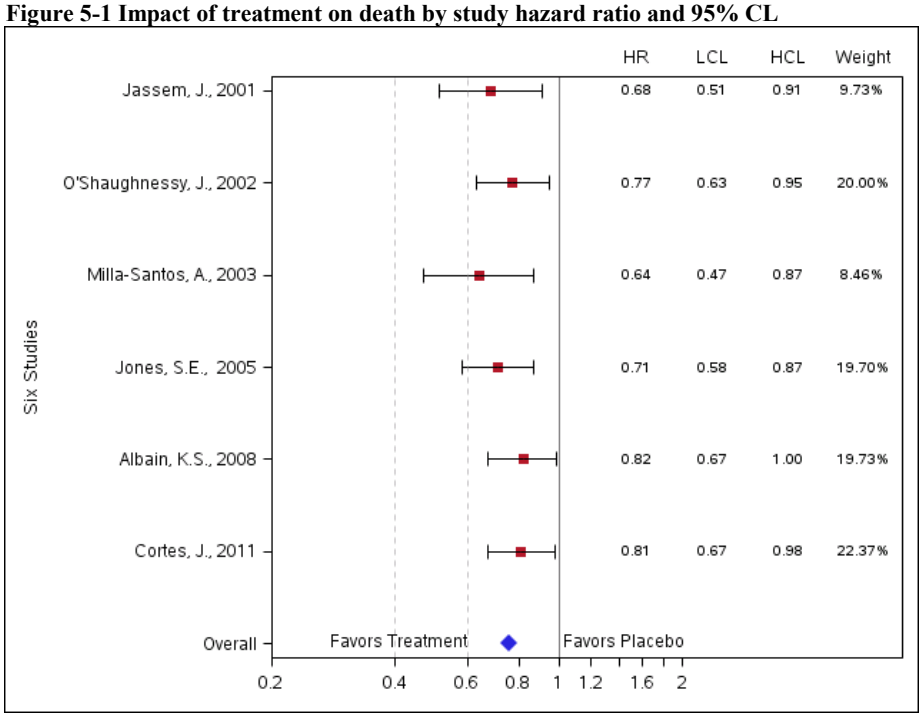
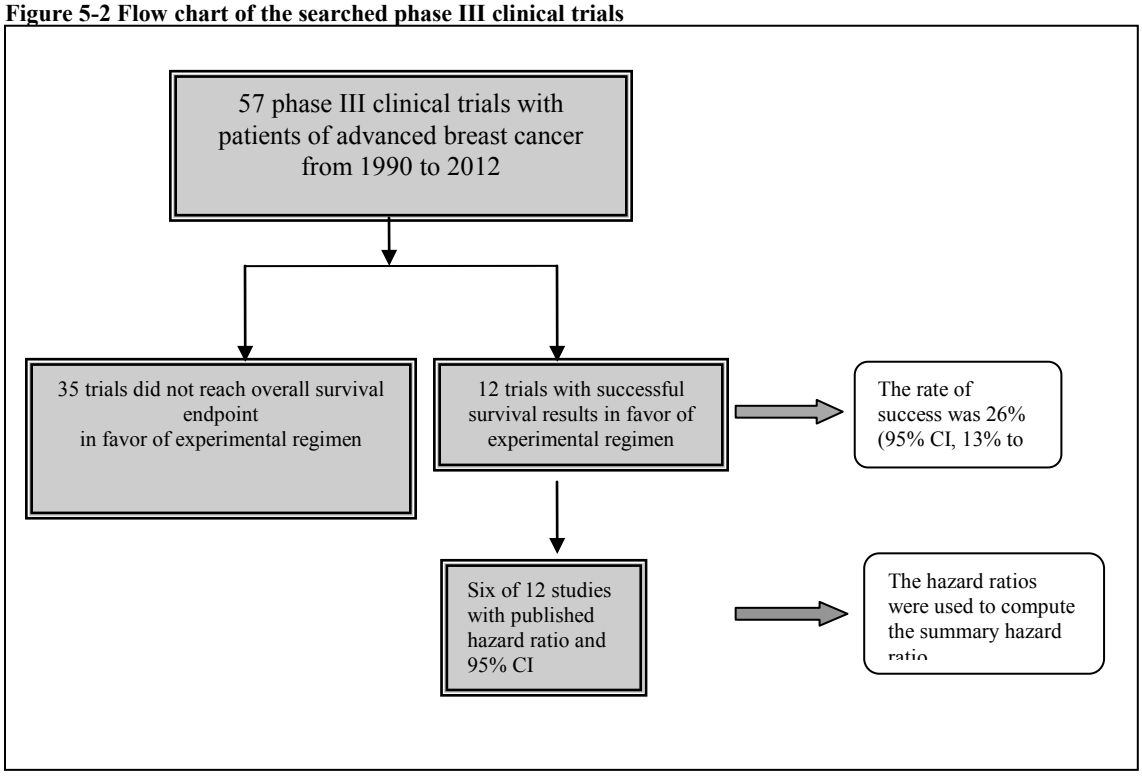


Figure 5-2 shows flow chart of the searched the 57 phase III clinical trials and the information extracted from those trials.

Figure 5-2 Flow chart of the searched phase III clinical trials



We identify 10 phase II or phase III studies (Dombernowsky P, 1998; Gershanovich M, 1998; Vahdat LT, 2009; Cortes J, 2010; Chevallier B, 1995; Allouache D, 2005; Demiray M, 2005; Brandi M, 2004; Fountzilas G, 2000; Blackstein M, 2002) that tested a regimen subsequently studied in a phase III trial. The information from these phase II or phase III trial studies is summarized in Table 5-1. Four of the subsequent phase III studies, each of them based on two or more phase II or phase III studies (Buzdar A, 2001; Cortes J, 2011; Albain KS, 2008; Chan S, 2009) are shown in Table 5-2. The phase III trial of Letrozole by Buzdar et al. was based on two preceding phase III trials.

Table 5-1 shows that a median of 46 patients were treated in each of the ten previous phase II or phase III studies. The number of patients treated in different phase II or phase III studies varied greatly, ranging from 24 to 551 patients. The median overall survival time is from 9 to 28 months.

Table 5-1 Phase II or phase III studies of advanced breast cancer identified as previous studies for subsequent phase III trials

Authors of phase II studies	No. of patients in phase II studies	Phase II regimen	Response rate	Median overall survival (Month)
Dombernowsky P, et al;	551	Letrozole	23.6	25.3
Gershanovich M, et al	155		19.5	28
Vahdat LT, et al.;	87	Eribulin	11.5	9
Cortes J, et al.	269	monotherapy	9.3	10.4
Chevallier B, et al.	34	Docetaxel	68	16.3
Allouache D, et al.;	35	Paclitaxel +	35	25.7
Demiray M, et al.	24	Gemcitabine	41.7	14.5
Brandi M, et al.;	53	Gemcitabine+	53	16.5
Fountzilas G, et al.	39	Docetaxel	36	12.7
Blackstein M, et al.	35	Gemcitabine	37.1	21.1

The studies done by Dombernowsky et al and Gershanovich et al are phase III studies.

Table 5-2 shows that the number of patients, treatment regimens for the experimental and control groups, and overall survival (OS) information for the each of the subsequent six phase III studies after the ten previous phase II or phase III studies. The order of studies separated by solid line in Table 5-1 is consistent with the order of studies separated by solid line in Table 5-2, which indicate each of the six subsequent phase III studies.

Response rates are frequently used as primary endpoint in phase II survival trials to assess the likelihood that the experimental regimen will increase survival over standard treatment in a phase III trial. However, the correlation of response rate in phase II trials to improved survival in phase III in solid tumor is uncertain until proven for any given agent and tumor (Pazdur R, 2000).

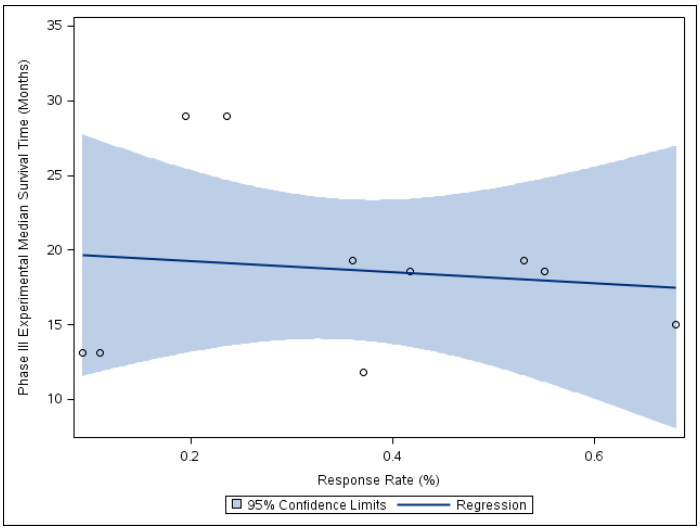
Table 5-2 Six Phase III trials of advanced breast cancer which are the subsequent phase III trials after the previous trials listed in Table 5-1

Authors of phase III studies	No. of patients in phase III studies (Experimental/ Control)	Phase III regimen (Experimental/ Control)	Overall survival (Experimental/Control)		
			Number of Event	Median (Month)	p-value
Buzdar A, et al	199/201	Letrozole vs. Megestrol Acetate	143/140	28.6/26.2	0.492
Cortes J. et al	508/254	Eribulin monotherapy vs. treatment of physician's choice	386/203	13.2/10.5	0.014 Primary†
Chan S, et al (1999)	161/165	Doctaxel with Doxorubicin	102/105	15/14	0.3893
Albain KS, et al	266/263	Paclitaxel plus Gemcitabine with Paclitaxel monotherapy	182/195	18.6/15.8	0.0489 Co-primary†
Chan S, et al (2009)	153/152	Gemcitabine plus Docetaxel with Capecitabine plus Docetaxel	119/115	19.29/21.45	0.983
Feher O, et al	198/199	Gemcitabine with Epirubicin	97/68	11.8/19.1	0.0004

† Primary means primary endpoint of efficacy analysis in the phase III studies. In the cases there is no primary specified in the table, the primary endpoint is response rate.

Figure 5-3 shows that the response rates in the ten Phase II or Phase III studies did not correlate with the median survival of patients treated with a particular regimen in the subsequent six phase III trial. The R square is 0.014, with the estimated slope -3.653 and p-value of 0.747.

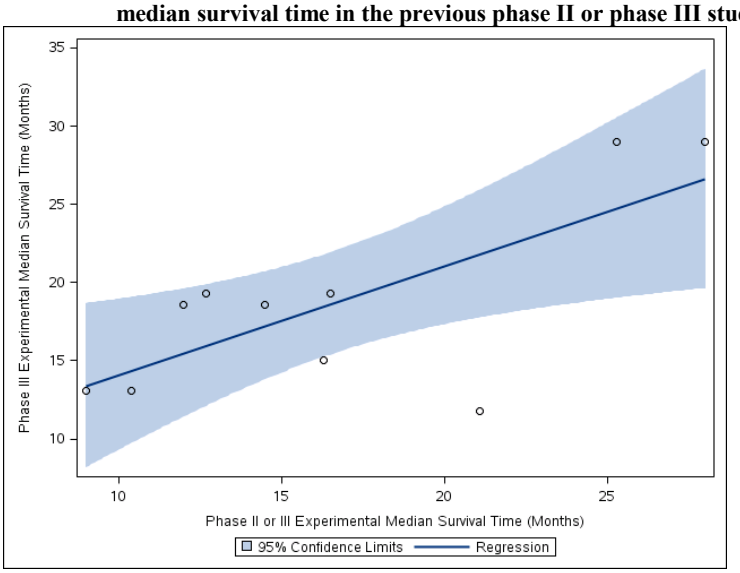
Figure 5-3 Median survival time in the subsequent phase III trials by the response rates in the previous Phase II or Phase III studies



The median survival of patients treated in the ten previous phase II or phase III studies is plotted versus the median survival of patients treated on the experimental arm of the corresponding subsequent six phase III trial in Figure 5-4. The least-squares regression lines are given in the figure, and the Pearson correlation coefficient is .723 (P = .018), with the estimated slope 0.696 and p-value = .018. The median survival of the patients treated on randomized phase III trials appeared to increase compared with the phase II studies because of the passage of time. The median survival times of patients were modestly longer in four of the six phase III studies than in the experimental arms of the previous trials that tested the same regimen (range of differences, 1–6.6 months) except in two cases the median survival time in phase III is 1.3 and 9.3 months less than those in preceding phase II trials. A nonparametric paired Wilcoxon signed rank test comparing the overall survival of patients in the ten Phase II or Phase III trials with that of patients treated on the experimental arms of the subsequent six phase III studies indicated that the

median survival times of these groups were insignificantly different with p-value of 0.131.

Figure 5-4 Median survival time in the subsequent phase III trials by the median survival time in the previous phase II or phase III studies



Since median survival time in phase II trials have a good correlation with the phase III survival results, the survival information instead of response rates was utilized in the predictive model in this report. This is consistent with FDA recommendation on the endpoints supporting approvals in oncology. In the early 1980s, the FDA determined that cancer drug approval should be based on more direct evidence of clinical benefit, such as improvement in survival, improvement in a patient's quality of lift (QOL), improved physical functioning, or improved tumor-related symptoms, which may not always be predicted by, or correlate with, objective response rate (ORR).

To retrospectively test how well the statistical model estimates the outcome of the phase III trials from previous historical survival data, we analyze the 10 phase II or III studies that gave rise to 6 subsequent phase III trials of the same experimental regimen. The expected power is the usual statistical power averaged with the regard to the size of

the treatment difference anticipated on the basis of the median survival time observed in the preceding phase II or III trial, the number of events observed in the preceding phase II or III trial, and the distribution of median survival time anticipated in the phase III trial for the control group. Similar to the endpoint of PFS, the detail calculations are shown below.

5.2.1 Gamma prior distribution of hazard of death for control regimen

For Control group, a Gamma distribution is formed with parameter a_c and b_c on the basis of historical phase III trials. The a_c represents amount of information (the number of death) on which the prior trial is based and assumed to be a fixed number. The $1/b_c$ represented the total patient-time of survival (until death or censoring) in previous experience with the treatment.

5.2.2 Gamma prior distribution of hazard of death for experimental regimen

For experimental group information about hazard available at the time of planning the phase III trial can also be specified as Gamma probability distribution.

Before the Phase III study, the probability of obtaining a positive result in a Phase III trial is set to 0.26 because only 12 (26%) of the 47 Phase III trials demonstrated a statistically significant improvement with regard to the endpoint of OS. Prior information are specified by setting a_e and b_e to give an expected median survival time (m_e) approximately equal to the expected median survival time (m_c) of the control treatment but with the probability of 26% that m_e is greater than median OS time of control group plus median survival time difference of 2.7 months between experimental and control regimen groups.

$$\int_0^{\infty} \lambda \frac{1}{\Gamma(a_e) b_e^{a_e}} \lambda^{a_e - 1} \exp\left(-\frac{\lambda}{b_e}\right) d\lambda = \ln 2/m_c \quad (1)$$

$$\int_0^{\lambda_2} \frac{1}{\Gamma(a_e) b_e^{a_e}} \lambda^{a_e - 1} \exp\left(-\frac{\lambda}{b_e}\right) d\lambda = 0.26 \text{ and } \lambda_2 = \ln 2/(m_c + 2.7) \quad (2)$$

Use SAS program to get a_e and b_e from equation (1) and (2).

5.2.3 Gamma posterior distribution of hazard of death for experimental regimen

We then update of Gamma prior distribution of hazard for OS in experimental regimen group by incorporating information (number of event and the total patient-time to event or censoring) from trials preceding to the subsequent phase III trials, on which the phase III trials were initiated based.

5.2.4 Computation of expected power and results

After model the Gamma distribution of hazard for the experimental and control groups based on the historical data from literature, we use the experimental data to update the existing distribution, and retrospectively calculate the expected power for the six phase III trials based on the formula in 2.3.1.3. We assume a 1:1 allocation ratio for experimental and control regimen in the future subsequent phase III study, total number of deaths of 537 to be observed in order to be adequately powered (90%) to detect the 32% increase in median survival time.

The control regimen information for the first five phase III studies are presented in Table 5-3 (Dombernowsky P, 1998; Sparano JA, 2010; Vaughn CB, 1988; Smith RE, 1999; O'Shaughnessy J, 2002). The control information is not available for the sixth study comparing Gemcitabine with Epirubicin in the Table 5-2. The number of events

and patient-time are used to model the Gamma distribution of hazard for control regimen for each individual phase III studies, respectively.

It is mentioned in the report of phase III trial comparing Eribulin monotherapy with treatment of physician's choice that a PubMed literature review was conducted on Sept 23, 2010 to identify phase III studies in metastatic breast cancer pretreated with both an anthracycline and taxane (Cortes J, 2011). This search yielded 25 reports, of which only five reported phase III studies in this specific setting. Three of the five studies reported the median overall survival time (Barrios CH, 2010; Miller KD, 2005; Sparano JA, 2010). The overall survival data in one study were not mature since the majority of patients in both arms were in follow-up at data cut-off, which yielded a high censoring rate (>65%) in both arms (Barrios CH, 2010). The rest two studies both have Capecitabine as control regimen with median survival time of 14.5 and 15.6 months, respectively. In one study, around 40% of patients received 2 or more prior chemotherapy; at data cut, 38% of patients in the study had died, but the number of death for control regimen only was not reported (Miller KD, 2005). In the other study less than 20% of patients received 2 or more prior regimens in the metastatic setting; at the time of analysis, there were 450 deaths (74%) in the control regimen group (Sparano JA, 2010). The Gamma model for the treatment of physician's choice was derived based on this third study.

Table 5-3 List of treatment regimen, number of OS events, Median OS time and Patient-time in previous historical phase II or phase III studies

	Authors of phase II or phase III/phase III	Control Regimen	Numer of Events (a_c) /Number of Patients	Median OS (Month) of Control Regimen	Patient -time (Month) of Control Regimen ($1/b_c$)
1	**Dombernowsky P, et al**	Megestrol Acetate	128/188	21.5	3960
2	**Sparano JA, et al**	Physician's choice	450/612	15.6	10120
3	**Vaughn CB, et al**	Doxorubicin	57/59	8.5	695
4	**Smith RE, et al**	Paclitaxel monotherapy	176/278	21.1	5347.45
5	**O'Shaughnessy J, et al**	Capecitabine plus Docetaxel	184/256	14.5	3842
6	**NA**	Epirubicin	-	-	-

In the phase III study comparing Gemcitabine with Epirubicin, several studies were referred as previous studies for the control regimen Epirubicin, however none of those reported both median survival time and number of patient death information. Therefore we do not have data to build a Gamma distribution for the control regimen Epirubicin.

We calculate the expected power for five phase III studies. The Gamma parameters used to model the Gamma distribution of hazard for both experimental and control groups are presented in Table 5-4.

Table 5-4 List of treatment regimen in subsequent phase III studies, and corresponding Gamma parameters used to model the Gamma distribution of hazard for both experimental and control regimens

Experiment vs. Control Treatment in Phase III	a_c	$1/b_c$	a_e	$1/b_e$	d_e	T_e	Power
Letrozole with Megestrol Acetate	**128**	**3960**	**36**	**1117**	**103**	**3747**	**0.29**
					98	**3945**	**0.64**
Eribulin monotherapy with treatment of physician's choice	**450**	**10120**	**21.5**	**484**	**53**	**684**	**5.07E-10**
					170	**2546**	**7.08E-10**
Doctaxel with Doxorubicin	**57**	**695**	**9**	**110**	**14**	**321**	**>0.99**
Paclitaxel plus Gemcitabine with Paclitaxel monotherapy	**176**	**5347.45**	**35.8**	**1090**	**14**	**507**	**0.11**
					19	**391**	**0.00056**
Gemcitabine plus Docetaxel with Capecitabine plus Docetaxel	**184**	**3842**	**19.8**	**408**	**26**	**611**	**0.15**
					20	**360**	**0.004**

In Table 5-5, the information of median OS time and number of events of the 9 previous phase II or III studies and median survival time and p-value in the subsequent 5 phase III trials are presented.

The range of the expected power is from close to zero to >0.99. The two phase II trials that yield the highest expected powers (>0.99 and 0.64) were followed by phase III studies that failed to show a statistically significant difference between experimental regimen and control regimen arms (Buzdar A, 2001; Chan S, 1999).

In contrast to the high powers, the other two phase II trials that yield the lowest expected powers (close to zero) were followed by phase III study that show a statistically significant difference between experimental regimen and control regimens (Cortes J, 2011; Albain KS, 2008).

From table 5-5 we cannot identify any pattern in the relationship between the calculated expected power and the final results with regard to whether or not the experimental regimen can prolong the OS comparing with the control regimen.

Therefore, according to our analysis by applying the Exponential-Gamma model to the endpoint of overall survival in advanced breast cancer, the calculated expected power fails to predict the likelihood of success of the subsequent phase III trial. The inconsistence is considered as understandable because there are many lines of breast cancer treatment, in phase III trial, patients are usually allowed to cross over to other cancer treatment or receive many other types of cancer therapy after disease progression, which can confound the OS results.

Table 5-5 List of expected power for future phase III studies, previous historical phase II or phase III studies, and results of overall survival in subsequent Actual phase III trials

Std. #	Authors of phase II/phase III	Experiment Vs. Control Treatment in Phase III	Observed Preceding Study			Subsequent Phase III			
			Numer of Death/ Number of Patients	Median Survival (Month) of Experimental	Power †	Median Survival Month (Experimental vs. Control)/ Crossover	Percentage Increase in Median Survival Time	Number of death (Experimental vs. Control)	p-Value
1	**Dombernowsky P, et al; Gershanovich M, et al /Buzdar A, et al**	Letrozole vs. Megestrol Acetate	103/551; 98/155	25.3; 28	0.29; 0.64	29 vs 26	0.12	137/140	0.492
2	**Vahdat LT, et al Cortes J, et al /Cortes J, et al**	Eribulin monotherapy vs. treatment of physician's choice	53/87; 170/269	9; 10.4	5.07E-10; 7.08E-10	13.1 vs 10.6	0.24	274/148	0.014;
3	**Chevallier B, et al. /Chan S, et al (1999)**	Doctaxel with Doxorubicin	14/34	16.3	> 0.99	15 vs. 14	0.07	102/105	0.3893
4	**Allouache D, et al; Demiray M, et al /Albain KS, et al**	Paclitaxel plus Gemcitabine with Paclitaxel monotherapy	16/29 19/24	12 14.5	0.11; 0.00056	18.6 vs 15.8	0.11	182/195	<0.0489
5	**Brandi M, et al; Fountzilas G, et al; /Chan S, et al (2009)**	Gemcitabine plus Docetaxel with Capecitabine plus Docetaxel	26/53; 20/39	16.5; 12.7	0.15; 0.004	19.29 vs 21.45	-0.1	78/76	0.983

† The power is calculated based on total number of death corresponding to 537 deaths to detect 32% increase in median survival time

6 Conclusion

In this report, a power model is implemented to predict the likelihood of success of future potential phase III studies for advanced breast cancer based on the information from previous historical phase II or phase III studies. We apply three statistical methods to model either the hazard or the hazard ratio between experimental and control regimen groups to estimate the expected power through a power model. The proposed methods include the Exponential-Gamma model, the Weibull-Inverse Gamma Model and the Piece-wise Exponential model. The first two methods are Bayesian conjugate statistical model.

We compare the results of these methods with regarding to how well the methods can assist to predict the likelihood of success for future potential phase III clinical trials of advanced breast cancer in terms of prolonging the disease progression free time.

With the Exponential-Gamma model, the estimated powers of two of the six studies contradict with the results from the subsequent phase III studies. Other than the two studies, the values of expected powers estimated from Exponential-Gamma model are consistent with the later phase III trial results.

The values of expected power from the Weibull-Inverse Gamma model method are consistent with the subsequent study results and resolve the problem of the Exponential-Gamma model method.

The Piece-wise Exponential method can only be applied to two studies due to limit of data available. The calculated expected powers are consistent with the successful results from later actual phase III studies.

Based on our experience in advanced breast cancer, an expected power of greater than 0.59 with the Exponential-Gamma model, and of greater than 0.64 with the Weibull-Inverse Gamma model provide a reasonable base to proceed to a phase III study. However, due to limitation of data, we cannot evaluate the validity of the Piece-wise Exponential method, nor provide suggestion of cut-off value of expected power for later phase III studies. In the future, we can apply those three statistical models together into more studies to validate the usefulness of the models in predicting the likelihood of future successful phase III trials.

For both Exponential-Gamma and Weibull-Inverse Gamma model, the expected power does not change dramatically with different number of events expected to be observed giving the fixed median progression free survival time.

In addition, according to our analysis by applying the Exponential-Gamma model to the endpoint of overall survival in advanced breast cancer, the calculated expected powers fails to predict the likelihood of success of the subsequent phase III trial. The inconsistence is considered as understandable because there are many lines of breast cancer treatment, in phase III trial, patients are usually allowed to cross over to other cancer treatment or receive many other types of cancer therapy after disease progression, which can confound the OS results.

Appendix

A.1 Authors and sponsors for phase II and phase III trials with the experimental regimen and endpoint of progression free survival

Std. #	Experiment Vs. Control Treatment in Phase III	Authors, phase II or phase III/phase III	Sponsor
1	Paclitaxel plus Gemcitabine with Paclitaxel Monotherapy	Allouache D, et al; Demiray M, et al /Albain KS, et al	Eli Lilly and Company; Unknown /Eli Lilly and Company
2	Gemcitabine plus Docetaxel with Capecitabine plus Docetaxel	Fountzilas G, et al; /Chan S, et al (2009)	Hellenic Cooperative Oncology Group /Eli Lilly and Company
3	PLD + Docetaxel with Docetaxel	Bontenbal M, et al. /Sparano JA, et al	Aventis Pharma /Johnson & Johnson Pharmaceutical Research & Development
4	Epirubicin and Paclitaxel with and without Capecitabine	Mansutti M, et al /Hatschek T, et al	Unknown /Bristol-Myers Squibb, Pfizer, Roche, the Research Funds at Radiumhemmet, the Swedish Cancer Society, the Swedish Breast Cancer Association (BRO) and ALF/FOU research funds at the Karolinska Institutet and Stockholm County Council
5	Ixabepilone plus Capecitabine with Capecitabine	Bunnell CA, et al /Thomas ES , et al	Bristol-Myers Squibb /Bristol-Myers Squibb
6	Arzoxifene with Tamoxifen	Baselga J, et al /Deshmane V, et al	Eli Lilly & Co /Eli Lilly & Co

A.2 Authors and sponsors for phase II and phase III trials with the experimental regimen and endpoint of overall survival

Std. #	Experiment Vs. Control Treatment in Phase III	Authors, phase II or phase III/phase III	Sponsor
1	Letrozole vs. Megestrol Acetate	Dombernowsky P, et al; Gershanovich M, et al /Buzdar A, et al	Grant from Novartis Pharma AG; Novartis Pharma AG /Novartis Pharmaceuticals Corp
2	Eribulin monotherapy vs. treatment of physician's choice	Vahdat LT, et al Cortes J, et al (2010) /Cortes J, et al (2011)	In part by Eisai Medical Research Inc.; Eisai Medical Research /Eisai.
3	Doctaxel with Doxorubicin	Chevallier B, et al. /Chan S, et al (1999)	Rhoˆne-Poulenc Rorer /Rhoˆne-Poulenc Rorer
4	Paclitaxel plus Gemcitabine with Paclitaxel monotherapy	Allouache D, et al; Demiray M, et al /Albain KS, et al	Eli Lilly and Company; Unknown /Eli Lilly and Company
5	Gemcitabine plus Docetaxel with Capecitabine plus Docetaxel	Brandi M, et al; Fountzilas G, et al; /Chan S, et al (2009)	Unknown; Hellenic Cooperative Oncology Group /Eli Lilly and Company

References

Albain KS, Nag SM, Calderillo-Ruiz G, et al: Gemcitabine plus paclitaxel versus paclitaxel monotherapy in patients with metastatic breast cancer and prior anthracycline treatment: Final results of a global phase III study. J Clin Oncol 26:3950-3957, 2008

Albain KS, Salazar J, Pienkowski T, et al. Reducing the global breast cancer burden: The importance of patterns of care research. Clin Breast Cancer 6:412-420, 2005

Allouache D, Gawande SR, Tubiana-Hulin M, et al. First-line therapy with gemcitabine and paclitaxel in locally, recurrent or metastatic breast cancer: A phase II study. BMC Cancer 5:151-160, 2005

Barrios CH, Liu MC, Lee SC, et al. Phase III randomized trial of sunitinib versus capecitabine in patients with previously treated HER2-negative advanced breast cancer. *Breast Cancer Res Treat* 2010;**121:** 121–31.

Baselga J, Campone M, et al. Everolimus in Postmenopausal Hormone-Receptor–Positive Advanced Breast Cancer. N Engl J Med 2012; 366:520-529.

Baselga J, Cortés J, et al. Pertuzumab plus Trastuzumab plus Docetaxel for Metastatic Breast Cancer. n engl j med 366;2 nejm.org january 12, 2012

Baselga J, Llombart-Cussac A, Bellet M, et al. Randomized, double-blind, multicenter trial comparing two doses of arzoxifene (LY353381) in hormonesensitive advanced or metastatic breast cancer patients. Ann Oncol 14:1383-1390, 2003.

Bishop JF, Dewar J, Toner GC, Smith J, Tattersall M, et al. Initial Paclitaxel Improves Outcome Compared With CMFP Combination Chemotherapy as Front-Line Therapy in Untreated Metastatic Breast Cancer. Journal of Clinical Oncology, Vol 17, No 8 (August), 1999: pp 2355-2364.

Blackstein M, et al. Gemcitabine as First-Line Therapy in Patients with Metastatic Breast Cancer: A Phase II Trial. Oncology 2002;62-:2-8.

Blum JL, Jones SE, Buzdar AU, et al. Multicenter phase II study of capecitabine in paclitaxel refractory metastatic breast cancer. J Clin Oncol 17:485-493, 1999.

Bontenbal M, Creemers GJ, Braun HJ, et al. Phase II to III study comparing doxorubicin and docetaxel with fluorouracil, doxorubicin, and cyclophosphamide as first-line chemotherapy in patients with metastatic breast cancer: Results of a Dutch Community Setting Trial for the Clinical Trial Group of the Comprehensive Cancer Centre. J Clin Oncol 23:7081-7088, 2005.

Borenstein M, Hedges LV, Higgins JPT, Rothstein HR. Introduction to Meta-Analysis. 2009 John Wiley & Sons, Ltd.

Brandi M, et al. Novel Association With Gemcitabine and Docetaxel as Salvage Chemotherapy in Metastatic Breast Cancer Previously Treated With Anthracyclines: Results of a Multicenter Phase II Study. Seminars in Oncology, Vol 31, No 2, Suppl 5 (April), 2004: pp 13-19.

Bunnell CA, Klimovsky J, Thomas E. Final efficacy results of a phase I/II trial of ixabepilone in combination with capecitabine in patients with metastatic breast cancer (MBC) previously treated with a taxane and an anthracycline. J Clin Oncol 24:568s, 2006 (suppl; abstr 10511)

Buzdar A, et al. Phase III, Multicenter, Double-Blind, Randomized Study of Letrozole, an Aromatase Inhibitor, for Advanced Breast Cancer Versus Megestrol Acetate. Journal of Clinical Oncology, Vol 19, No 14 (July 15), 2001: pp 3357-3366.

Buzdar A, Hayes D, El-Khoudary A, et al. Phase III randomized trial of droloxifene and tamoxifen as first-line endocrine treatment of ER/PgRpositive advanced breast cancer. Breast Cancer Res Treat 73:161-175, 2002.

Chan S, Friedrichs K, Noel D, Pinter T, Belle SV, et al. Prospective Randomized Trial of Docetaxel Versus Doxorubicin in Patients With Metastatic Breast Cancer. Journal of Clinical Oncology, Vol 17, No 8 (August), 1999: pp 2341-2354.

Chan S, Romieu G, Huober J, Delozier T, Tubiana-Hulin M, et al. Phase III Study of Gemcitabine Plus Docetaxel Compared With Capecitabine Plus Docetaxel for Anthracycline-Pretreated Patients With Metastatic Breast Cancer. J Clin Oncol 27:1753-1760 © 2009.

Chen TT, et al. A Model to Select Chemotherapy Regimens for Phase III Trials for Extensive-Stage Small-Cell Lung Cancer. Journal of the National Cancer Institute, Vol. 92, No. 19, October 4, 2000.

Chevallier B, Fumoleau P, et al. Docetaxel is a Major Cytotoxic Drug for the Treatment of Advanced Breast Cancer: A Phase II Trial of the Clinical Screening Cooperative Group of the European Organization for Research and Treatment of Cancer. J Clin Oncol 13:314-322.1995.

Cortes J, Vahdat L, Blum JL, et al. Phase II study of the halichondrin B analog eribulin mesylate in patients with locally advanced or metastatic breast cancer previously treated with an anthracycline, a taxane, and capecitabine. J Clin Oncol 2010;28: 3922-28.

Cortes J, O'Shaughnessy J, et al. Eribulin monotherapy versus treatment of physician's choice in patients with metastatic breast cancer (EMBRACE): a phase 3 open-label randomised study. Lancet 2011; 377: 914–23.

Demiray M, Kurt E, Evrensel T, et al: Phase II study of gemcitabine plus paclitaxel in metastatic breast cancer patients with prior anthracycline exposure. Cancer Invest 23:386-391, 2005.

Deshmane V, Krishnamurthy S, Melemed AS, Peterson P, and Buzdar AU. Phase III Double-Blind Trial of Arzoxifene Compared With Tamoxifen for Locally Advanced or Metastatic Breast Cancer. J Clin Oncol 25:4967-4973. © 2007.

Dombernowsky P, et al. Letrozole, a new oral aromatase inhibitor for Advanced Breast Cancer: Double-Blind Randomized trial Showing a Dose Effect and Improved Efficacy and Tolerability Compared With Megestrol . Journal of Clinical Oncology, Vol 16, No 2 (Feb), 1998: pp 453-461.

Feher O, et al. First-Line gemcitabine versus epirubicin in postmenopausal women aged 60 or older with metastatic breast cancer: a multicenter, randomized, phase III study. Annals of Oncology 16:899-908, 2005.

Fink, D. (1997). "A Compendium of Conjugate Priors". *DOE contract 95-831*.

Finn RS, Crown JP, Lang I, etc. The cyclin-dependent kinase 4/6 inhibitor palbociclib in combination with letrozole versus letrozole alone as first-line treatment of oestrogen receptor-positive, HER2-negative, advanced breast cancer (PALOMA-1/TRIO-18): a randomised phase 2 study. Lancet Oncol 2015; 16: 25–35.

Fountzilas G, Nicolaides C, Bafaloukos D, et al. Docetaxel and gemcitabine in anthracyclineresistant advanced breast cancer: A Hellenic Cooperative Oncology Group Phase II study. Cancer Invest 18:503-509, 2000.

Freidlin, B, et al. A Model to Select Regimens for Phase III Trials for Patients with Advanced-Stage Non-Small Cell Lung Cancer. Clinical Cancer Research Vol. 9, 917–922, March 2003.

Gershanovich M, et al. Letrozole, a new oral aromatase inhibitor: Randomized trial comparing 2.5 mg daily, 0.5 mg daily and aminoglutethimide in postmenopausal women with advanced breast cancer. Annals of Oncology 9: 639-645, 1998.

Guidance for Industry Clinical Trail Endpoints for the Approval of Cancer Drugs and Biologics. U.S. Department of Health and Human Services. Food and Drug Administration Center for Drug Evaluation and Research (CDER) Center for Biologics Evaluation and Research (CBER) May 2007 Clinical/Medical.

Guyot P, Ades AE, et al. BMC Medical Research Methodology 2012, 12:9.

Hatschek T, Carlsson L, Einbeigi Z, Lidbrink E, Linderholm B, et al. Individually tailored treatment with epirubicin and paclitaxel with or without capecitabine as first-line chemotherapy in metastatic breast cancer: a randomized multicenter trial. Breast Cancer Res Treat (2012) 131:939–947.

Hirschfeld S, Pazdur R. Oncology drug development: United States Food and Drug Administration perspective. Crit Rev Oncol Hematol 42:137-143, 2002 (8)
Holford TR. The analysis of rates and survivorship using log-linear models. *Biometrics* **36** 299-305. 1980.

Hortobagyi GN, Salazar J, Pritchard K, et al: The global breast cancer burden: Variations in epidemiology and survival. Clin Breast Cancer 6:391-401, 2005.

Howlader N, Noone AM, Krapcho M, Neyman N, Aminou R, Altekruse SF, Kosary CL, Ruhl J, Tatalovich Z, Cho H, Mariotto A, Eisner MP, Lewis DR, Chen HS, Feuer EJ, Cronin KA (eds). *SEER Cancer Statistics Review, 1975-2009 (Vintage 2009 Populations)*, National Cancer Institute.

Jassem J, Pienkowski T, et al. Doxorubicin and paclitaxel versus fluorouracil, doxorubicin, and cyclophosphamide as first-line therapy for women with metastatic breast cancer: final results of a randomized phase III multicenter trial. J Clin Oncol. 2001 Mar 15;19(6):1707-15.

Jones SE, Erban J, et al. Randomized phase III study of docetaxel compared with paclitaxel in metastatic breast cancer. J Clin Oncol 2005 Aug 20;23(24):5542-51.

Laird N. and Olivier D. (1981). Covariance analysis of censored survival data using log-linear analysis technique. *J. Amer. Statist. Assoc.* **76** 231-240.

Mansutti M, Cavazzini G, Lorusso V, Boni C et al (2008) Randomized, multicenter, phase III trial of docetaxel plus epirubicin (ET) with or without capecitabine (X) as first-line therapy for stage IV breast cancer (BC). J Clin Oncol 26:49s (May 20 suppl; abstr 1034).

Minckwitz GV, Schwedler K, et al. Trastuzumab beyond progression: overall survival analysis of the GBG 26/BIG 3-05 phase III study in HER2-positive breast cancer. Eur J Cancer. 2011 Oct;47(15):2273-81. Epub 2011 Jul 7. Eur J Cancer. 2011 Oct;47(15):2273-81.

Milla-Santos A, Miller L, et al. Anastrozole versus tamoxifen as first-line therapy in postmenopausal patients with hormone-dependent advanced breast cancer: a prospective, randomized, phase III study. Am J Clin Oncol. 2003 Jun;26(3):317-22.

Miller KD, Chap LI, Holmes FA, et al. Randomized phase III trial of capecitabine

compared with bevacizumab plus capecitabine in patients with previously treated metastatic breast cancer. J Clin Oncol 2005; **23:** 792–99.

O'Shaughnessy J, Miles D, Vukelja S, Moiseyenko V, Ayoub JP, Cervantes G, et al. Superior Survival With Capecitabine Plus Docetaxel Combination Therapy in Anthracycline-Pretreated Patients With Advanced Breast Cancer: Phase III Trial Results. J Clin Oncol, Vol 20, No 12 (June 15), 2002: pp 2812-2823.

O'Shaughnessy J, Schwartzberg LS, et al. A randomized phase III study of iniparib (BSI-201) in combination with gemcitabine/carboplatin (G/C) in metastatic triple-negative breast cancer (TNBC). Journal of Clinical Oncology, 2011 ASCO Annual Meeting Abstracts. Vol 29, No 15_suppl (May 20 Supplement), 2011: 1007

Paridaens R, Biganzoli L, Bruning P, Klijn JGM, Gamucci T, Houston S, Coleman R, Schachter J, et al. Paclitaxel Versus Doxorubicin as First-Line Single-Agent Chemotherapy for Metastatic Breast Cancer: A European Organization for Research and Treatment of Cancer Randomized Study With Cross-Over. Journal of Clinical Oncology, Vol 18, No 4 (February), 2000: pp 724-733.

Pazdur R. Response rates, survival, and chemotherapy trials. J Natl Cancer Inst 92:1552~1553, 2000.

Roberts Jr TG, Lynch Jr TJ, and Chabner BA. The Phase III Trial in the Era of Targeted Therapy: Unraveling the "Go or No Go" Decision. Journal of Clinical Oncology, Vol 21, No 19 (October 1), 2003: pp 3683-3695.

Smith RE, Brown AM, Eleftherios P. Mamounas, et al. Randomized Trial of 3-Hour Versus 24-Hour Infusion of High-Dose Paclitaxel in Patients With Metastatic or Locally Advanced Breast Cancer: National Surgical Adjuvant Breast and Bowel Project Protocol B-26. Journal of Clinical Oncology, Vol 17, No 11 (November), 1999: pp 3403-3411.

Shulman S.R., Wood-Armany MJ. Accelerating access to cancer drugs. J Biolaw Bus 2:38-44, 1999.

Soland R. Use of the Weibulll Distribution in Bayesian Decision Theory. Research Analysis Corporaton, McLean, VA, Report no. RAC-TP-225, 1966.

Sparano JA, Makhson AN, Semiglazov VF, et al. Pegylated Liposomal Doxorubicin Plus Docetaxel Significantly Improves Time to Progression Without Additive Cardiotoxicity Compared With Docetaxel Monotherapy in Patients With Advanced Breast Cancer Previously Treated With Neoadjuvant-Adjuvant Anthracycline Therapy: Results From a Randomized Phase III Study. J Clin Oncol 27:4522-4529. © 2009

Sparano JA, Vrdoljak E, Rixe O, et al. Randomized phase III trial of ixabepilone plus

capecitabine versus capecitabine in patients with metastatic breast cancer previously treated with an anthracycline and a taxane. *J Clin Oncol* 2010; **28:** 3256–63.

Thomas D. Oncology Clinical Trials – Secrets of Success. Inside BIO Industry Analysis. Feb 24, 2012.

Thomas ES, Gomez HL, Li RK, Chung HC, Fein LE, Chan VF, Jassem J, et al. Ixabepilone Plus Capecitabine for Metastatic Breast Cancer Progressing After Anthracycline and Taxane Treatment. J Clin Oncol 25:5210-5217. © 2007.

Vahdat LT, Pruitt B, Fabian CJ, et al. Phase II study of eribulin mesylate, a halichondrin B analog, in patients with metastatic breast cancer previously treated with an anthracycline and a taxane.J Clin Oncol 2009; 27: 2954–61.

Vaughn CB, Green SJ, O'Bryan R, et al. VP-16 + Adriamycin vs. Adriamycin alone in advanced adenocarcinoma of the breast: Phase II, A randomized trial—A Southwest Oncology Group study. Med Pediatr Oncol 16:312-319, 1988.

Verma S, Miles D, et al. Trastuzumab Emtansine for HER2-Positive Advanced Breast Cance. N Engl J Med 2012;367:1783-91.

Vita

Xinqun Chen

2015 Dr.PH in Biostatistics, School of Public Health, Rutgers, The State University of New Jersey

1999 – 2001 MPH in Biostatistics, The George Washington University, Washington DC

1989 – 1994 Bachelor of Medicine in Nursing, Tianjin Medical University, China