Journal of Reliability and Statistical Studies (ISSN: 0974-8024)
Vol. 1, Issue 1(2008) 6-17

A BAYES COMPARISON OF TWO DIFFERENT CANCER THERAPIES UNDER THE ASSUMPTION OF WEIBULL SURVIVAL MODEL OR ITS SUBFAMILY

Puja Arora[1] and S.K. Upadhyay[2]
Department of Statistics
DST Centre for Interdisciplinary Mathematical Sciences
Banaras Hindu University, Varanasi-221 005, India
E Mail: 1. puja.stats@gmail.com, 2. skupadhyay@gmail.com

Abstract

The paper considers a group of patients suffering from leukemia B non-small lung cancer. Such patients are generally suggested to undergo for either radiotherapy or chemotherapy followed by radiotherapy. The objective of the paper is to compare the two therapies based on survival functions of the patients assuming Weibull survival model for each therapy. The paper further examines the feasibility of a subfamily of Weibull model, namely the exponential distribution, for a date set available from a clinical trial experiment. This feasibility is judged based on Bayes information criterion by comparing the Weibull model with its subfamily. The model compatibility study with the data based on posterior p-values has also been given to ensure the suitability of the two models. Finally, the recommendations are made accordingly.

Key words: Weibull Distribution, Shape Parameter, Exponential Distribution, Survival Function, Non-informative Prior, p-value, Chi-square Discrepancy Measure, Bayes Information Criterion

1. Introduction

Carcinoma is an important disease that causes large number of deaths around the world. The biggest problem with the disease is that it is often not diagnosed at an early stage precluding the chances of cure of the patients. At advanced stages of the disease, the medical practitioners often have limited scope to relieve the patient. Earlier the patients were used to be treated with radiotherapy but later on they were given chemotherapy followed by radiotherapy. The interest therefore centres among the medical practitioners that which of the two therapies provide better results in terms of the survival of the patients. Such studies have been considered earlier by a number of researchers when patient is suffering from advanced stages of carcinoma in different organs. The studies were mainly carried out by medical practitioners or statisticians working with the medical data and often focused on randomized, prospective, retrospective, multi centre clinical trials. Broadly speaking the analyses used both classical and Bayesian methodologies although the former dominate the latter paradigm. Since the studies are numerous covering a variety of cancerous forms, it is not possible to provide an exhaustive list of references and, therefore, we shall be focusing primarily on the studies based on survival data. It is to be noted that the data for such comparisons often come in the form of survival times and, among the various approaches, a better therapy can be suggested to practitioners on the basis of comparison between the corresponding survival functions.

Among the earlier classical developments, one can refer to Kaplan and Meier (1958), Mantel (1966), etc. for estimating and comparing survival functions based on Kaplan-Meier estimate or log rank test. Cox (1972) proposed the concept of proportional hazards model for quantifying the effects of covariates on the survival times. A systematic review of literature can be had from Qian (1994), Lawless (2002), Kalbfleisch and Prentice (2002) among others. On the Bayesian front Sinha and Dey (1997) is an important review article that provides a number of practical problems and correspondingly provides an updated list of related developments. The other important references include Ibrahim et al. (2001), Gelman (2004), etc. Most of these references are intended towards comparing survival functions obtained on the basis of various modelling assumptions for the data.

The present paper considers a comparison of survival functions obtained from two

therapies assuming Weibull models for the corresponding survival times so that one can be in a position to say that which therapy is better. The Weibull distribution has a wide range of applicability especially in lifetime data analysis perhaps because of its virtue of versatility or flexibility. The p.d.f. of two-parameter Weibull distribution can be written as

$$f(t|\ \alpha,\beta) = \frac{\beta}{\alpha}\left(\frac{t}{\alpha}\right)^{\beta-1} \exp\left[-\left(\frac{t}{\alpha}\right)^{\beta}\right], \quad \alpha,\beta > 0 \tag{1}$$

where α is the scale parameter and β determines the shape of the distribution. The corresponding survival function can be given by

$$S(t|\ \alpha,\beta\ ) = \exp\left[-\left(\frac{t}{\alpha}\right)^{\beta}\right],\ \alpha,\beta > 0 \tag{2}$$

The Weibull distribution encompasses monotonically increasing (for β >1), decreasing (for β <1), and constant (for β =1) failure rate and, as such, the model has been successfully used to describe both initial failures as well as the failures due to remission or aging (see Lawless (2002)). One of the biggest advantages with the Weibull model is the availability of closed form survival function, which makes the inferences related to the model quite easy although the non-availability of sufficient statistics poses some problem in comparison to those situations where the existence of the same is guaranteed.

The Weibull model is perhaps the richest one as far as the inferential developments are concerned both with regard to classical and Bayesian paradigms. Lawless (2002) is an important text which systematically describes the classical developments based on the model both in the context of engineering and medical applications. A few other important references include Kalbfleisch and Prentice (2002), Lee and Wang (2003), etc. On the Bayesian front, a systematic accountability can be seen in Martz and Waller (1982) and, more recently, in Singpurwalla (2006). The other important references include Gelman et al. (2003), Ibrahim et al. (2004), etc. although a number of research papers on Weibull distribution appeared regularly in various journals.

The distribution is, in general, not too straightforward to deal with. The classical developments on the model mostly relied on large sample approximations or empirical results. The problem with the Bayesian inference lies in the involvement of integrals in the posterior based inferences, which are difficult to solve analytically and, as such, require specialized techniques of Bayesian computation (see, for example, Upadhyay et al. (2001), and, more recently, Gamerman and Lopes (2006)). This last reference advocated the use of sample based approaches in Bayesian computation because of their several inherent advantages. A few such advantages may include the straightforwardness of the procedures to deal with censored data problems and routine inferential development for some nonlinear functions of the model parameters.

The Weibull distribution becomes straightforward if one is confronted with a situation where shape parameter β can be taken to be unity. The resulting distribution becomes one-parameter exponential and inferential developments based on it are routine. This is equivalent to say that an experimenter tests β against unity for the given data set and goes for the exponential model if the hypothesis is accepted. Such problems have been considered earlier by a number of authors in both classical and Bayesian paradigms. The most frequently used classical tool for testing β against unity is based on the likelihood ratio test. The earlier cited references do provide enough material on testing β against unity for the Weibull model. For Bayesians, the obvious technique can be based on the evaluation of Bayes factor which is a bit difficult when the priors are non-informative and the data are compounded with censoring mechanism (see, for example, Upadhyay and Mukherjee (2008)). The problem of testing β=1 can also be visualized as that of model comparison where one can use, for example, the Bayes information criterion (BIC) (see, for example, Schwarz (1978)) for drawing the necessary

conclusion. No doubt, this measure is comparatively easy and provides answers parallel to that based on Bayes factor.

A model comparison is justified among the compatible models only where compatibility is referred to mean that all the models under consideration do provide an adequate representation to the given data. Therefore, we first propose a compatibility study of the exponential and Weibull models and then provide a model comparison study to pick up an appropriate model. For studying the compatibility, we have used the posterior Bayes factor based on Bayesian version of chi-square discrepancy measure.

The plan of the paper is as follows. The next section provides a detailed model formulation starting from the likelihood to prior and then to the posterior. The section also provides a brief review discussion of the Gibbs sampler algorithm so that we may be able to generate samples from the concerned posterior. The posterior corresponding to exponential distribution when β =1 has also been commented briefly. Section 3 provides a brief review discussion of model compatibility and comparison tools that have been employed in the paper. Section 4 considers a real data set obtained from clinical trials on the two therapies and provides numerical results of our proposed study plan. The section first considers comparison of the two therapies assuming Weibull model for the survival times and then considers examining the possibility of taking β =1 so that feasibility of exponential model for the data can be seen. Results based on exponential modelling assumption have also been given for completeness. Finally, a brief conclusion is given in Section 5.

2. Model Formulation

To begin with let us consider a group of n patients who have undergone treatments for certain disease. These n patients can also be considered prospectively especially when one is indulged with experiments involving clinical trials. It is to be noted that the present paper primarily considers a group of advanced stage caner patients treated with either radiotherapy (RT) or chemotherapy followed by radiotherapy (CT+RT). We further assume that out of n patients receiving a particular therapy, survival times x_i (i =1,..., r) for r patients are observed completely whereas for remaining n-r patients, we simply have the information that j^{th} patient left the study at the censoring time, say c_j (j=r+1,…,n). If we consider x's to follow Weibull model with parameters (α, β), the corresponding likelihood function (LF) can be written as

$$L(\underline{x}|\alpha,\beta) \propto \frac{(\beta)^r}{(\alpha)^{r\beta}} \prod_{i=1}^{r} (x_i)^{\beta-1} \exp\left[-\sum_{i=1}^{r}\left(\frac{x_i}{\alpha}\right)^{\beta}\right] \prod_{j=r+1}^{n} \Pr(x_j > c_j),\ \alpha,\beta>0 \tag{3}$$

where $\underline{x}$ is used to denoted the available information on survival times. The term $\Pr(x_j > c_j)$ is the survival function at c_j corresponding to the Weibull model and it is available in a nice closed form. This last term occurs in (3) because of the censored data and it can be completely removed if there had been no censoring and survival times for all the n patients are observed completely, that is, r itself becomes n.

The Bayesian formulation next requires appropriate priors for the parameters. If we have enough information that can help us to go for informative prior, it may certainly be preferred over all other choices. Otherwise it is better to stick to non-informative or vague priors. In this paper we prefer taking a vague choice on the lines of Upadhyay et al. (2001) (see also Singpurwalla (2006)). The priors considered by the authors are

$$g(\alpha,\beta) \propto (\alpha.\beta)^{-1} \tag{4}$$

Combining the LF with the prior via Bayes theorem yields the posterior that can be written upto proportionality as

$$p(\alpha,\beta|\underline{x}) \propto \frac{(\beta)^{r-1}}{(\alpha)^{r\beta+1}} \prod_{i=1}^{r} (x_i)^{\beta-1} \exp\left[-\sum_{i=1}^{r}\left(\frac{x_i}{\alpha}\right)^{\beta}\right] \prod_{j=r+1}^{n} \exp\left[-\left(\frac{c_j}{\alpha}\right)^{\beta}\right] \tag{5}$$

The posterior given in (5) can be analyzed by any of the various available techniques (see, for example, Upadhyay et al. (2001)). The solution is not that difficult, as we often require solving only one-dimensional integral whether the interest focuses on joint posterior or the marginal posterior. There are several other approximate techniques (see, for example, Gamerman and Lopes (2006)) which can equally well be applied to obtain the desired inferences from the posterior given in (5). We, however, advocate the use of sample based approaches, in particular the Gibbs sampler, simply because of its inherent ease.

Before we provide a brief discussion of the implementation of Gibbs sampler algorithm on the posterior (5), let us briefly review the algorithm itself. The Gibbs sampler algorithm is a Markovian updating scheme that proceeds by generating from various full conditionals specified upto proportionality from the joint posterior, the latter also needs to be specified upto proportionality only. In order to run the algorithm some initial values are assigned at the beginning to the generating variates and then the chain proceeds in a cyclic order using the most recent values of all other variates. The details about the algorithm, its necessary implementation, and the convergence diagnostic issues can be found in Smith and Roberts (1993) and Upadhyay et al. (2001) among others. The algorithm can be implemented either by means of a single long run of the chain or by means of multiple chains of long run and then outcomes can be picked up once the convergence is assured in the generating chain. In a single long run of the chain the outcomes can be picked up from equidistant positions to avoid serial correlation among the generating variates. Similarly, for parallel chains the outcomes can be taken from the same relative positions after the convergence is assured (see, for example, Smith and Roberts (1993)). The final selected outcomes can be regarded as random samples from the joint posterior with components as the random samples from the corresponding marginal posteriors.

The Gibbs sampler algorithm has an apparent advantage when one is interested in the posterior of some non-linear function of the original variates. The analytical derivation of this posterior is often difficult. The Gibbs sampler algorithm suggests that samples from such a posterior can be easily obtained by replacing each parameter in the nonlinear function with the corresponding sample. Thus sample-based estimates can be easily derived once the final samples are made available from the corresponding posteriors. In case of censoring Gibbs sampler can be routinely extended without any extra burden. We apply the scheme on the concerned posterior in a usual way, treating the censored observations as further unknowns. The rest of the developments are same except that new full conditionals are introduced corresponding to the unknown censored data. That is, the full conditionals corresponding to unknown parameters will be same as would have been obtained had there been no censoring. The full conditionals corresponding to independent censored data are, however, the parent sampling distributions truncated in the appropriate regions (see, for example, Upadhyay et al. (2001)). Thus the unknown censored data can be generated as independent draws from the truncated parent sampling distribution.

The implementation of the algorithm for the posterior (5) is, therefore, quite straightforward. We simply need to think for the full conditionals of α and β. We also need to think for the full conditionals corresponding to unknown censored data. It can be shown that the full conditional of α reduces to gamma distribution after a simple transformation whereas that of β can be shown to be log-concave. The full conditional corresponding to censored data, say x_j ($>c_j$), is truncated Weibull distribution in the region (c_j, ∞). An apparent advantage of this scheme is that it can be used to assess the unknown censored data exactly the way it does provide information on unknown α and β. That is, once the convergence monitoring is done on all the unknowns, the samples from the generated chains can be used to study the desired

features of interest. We skip a detailed discussion of the Gibbs sampler implementation to the posterior corresponding to the Weibull model and refer to Upadhyay et al. (2001) for the same. The authors have systematically detailed every little step on its implementation (see also Singpurwalla (2006)).

As already mentioned, the Weibull distribution reduces to one-parameter exponential distribution when the shape parameter β becomes unity. In this case if we consider the prior for α proportional to α^{-1}, the corresponding posterior can be easily reduced from (5) by putting β =1. The posterior after a simple reciprocal transformation can be written as

$$p_e(\lambda|\underline{x}) \propto (\lambda)^{r-1} \exp\left[-\lambda\left(\sum_{i=1}^{r}(x_i)+\sum_{j=r+1}^{n}(c_j)\right)\right] \tag{6}$$

where $\lambda = \alpha^{-1}$. Obviously, (6) is gamma density with shape parameter r and scale parameter $\left(\sum_{i=1}^{r} x_i + \sum_{j=r+1}^{n} c_j\right)^{-1}$. Therefore, the posterior (6) corresponding to exponential model can be easily managed for any desired inferences. We shall, however, use samples from (6) to draw the needed inferences and employ the same strategy that has been discussed for Weibull distribution with censored data situation.

3. Model Compatibility and Comparison

Model compatibility study is meant to see if a model under consideration does provide a good fit to the data in hand and, therefore, provides a valid reason for considering a model. A number of tools have been suggested for studying compatibility of a model in both classical and Bayesian frameworks. An important approach in classical paradigm is to use tail area probability or better known as the p-value based on a goodness of fit test and to replace the unknown parameter(s), if any, involved in the process by some good estimates usually the maximum likelihood (see, for example, Lawless (2002)). Bayesian paradigm offers a number of possibilities for checking model compatibility, the most important being the one based on predictive simulation ideas. The idea suggests that if the observed data and the data predicted from the model exhibit some kind of similarities, the model under consideration can be considered compatible with the observed data (see, for example, Gelman et al. (1996)). Bayesians have also defined a number of versions of p-values analogously to the classical approach but they have suggested integrating out the unknown parameter(s) by some of its possible distributions. These versions are referred to as the prior, posterior, conditional, or partial posterior predictive p-values. Each of these measures has their own merits or demerits but we do not go it to the details of these various aspects due to space restriction. Gelman et al. (1996), Bayarri and Berger (1998), and Upadhyay and Mukherjee (2008), etc. are some important references for a detailed discussion of these ideas. For the purpose of our illustration, we consider the use of posterior predictive p-value based on an important classical discrepancy measure in spite of the fact that the measure has invited a few shortcomings too. We simply use it because of its ease and also because of the fact that our compatibility study requires only a tentative answer and the final answer will be based on the result of model comparison. Moreover, as pointed out by a number of authors, the posterior predictive p-value can be used at least for a preliminary check of model compatibility (see, for example, Upadhyay and Peshwani (2008)). Before we proceed further, let us review it briefly on the lines of Upadhyay et al. (2001) (see also Gelman et al. (1996)).

Let the observed data be denoted by x and the predictive data by y, D is the measure of discrepancy between the samples and population values and $f(.|\theta)$ be assumed model for the data. Then the Bayesian posterior predictive p- value can be defined as

$$p = \Pr[(D_2 \geq D_1|f,x)] = \int \Pr(D_2 \geq D_1|f,\theta)\, p(\theta|f,x)\, d\theta \tag{7}$$

where $p(\theta \mid f, x)$ is the posterior distribution of θ under the model f, D_1 and D_2 are the measures of discrepancy corresponding to the observed and the predictive data, respectively. Equation (7) can be regarded as the classical p-value averaged over the posterior distribution of θ under the model f. If we assume, for example, chi- square as a measure of discrepancy (Gelman et al. (1996)), we can write

$$D = \sum_{i=1}^{n} \frac{(d_i - E(d_i \mid \theta))^2}{V(d_i \mid \theta)}, \tag{8}$$

where d_i (i=1,…, n) is the i^{th} observation in the considered data set and n is the corresponding sample size. It is to be noted that chi-square discrepancy measure is arbitrarily chosen for illustration only; one can similarly define p-values based on other discrepancy measures as well. Thus using (8) in (7), the posterior predictive p-values corresponding to chi-square discrepancy measure can be easily obtained. Our conclusion based on the evaluated p-value will simply be 'larger the p- value, better is the compatibility of the considered model with the observed data' (see also Upadyay and Peshwani (2008)). The integration in (7) can often be a major difficulty in evaluating the p-value and the situation worsens with the increasing dimensionality of θ. The situation can, however, be easily managed if one resorts to sample based approaches for simulating the posterior $p(\theta \mid f, x)$ and then evaluates the sample-based estimate of the corresponding p-value. Upadhyay et al. (2001) have provided details of the various steps involved in the evaluation of (7) using sample based approaches.

In performing compatibility study, it is often seen that a number of models are found compatible with the data in hand. The question, therefore, arises which model should be finally considered for the data. The question, although difficult, can only be answered if one performs some sort of comparison among the competitive models and then accordingly recommends a model. It is to be noted that the results of model compatibility study can never be used for comparing the models rather it can be used only to check if the assumed model is compatible with the data or not. Before we proceed further, let us make a simple comment on parsimony principle which recommends a model which is simplest. Undoubtedly, this principle is quite useful and advocated by a number of authors but sometimes, while recommending a model according to this principle, the experimenter may loose some of the important inferential aspects (see, for example, Upadhyay and Mukherjee (2008)). In the present paper, we shall focus on BIC although a number of other sophisticated tools can also be used for the desired comparison.

3.1 Bayes Information Criterion

The BIC also known as Schwarz criterion is a well-known criterion for comparing the models. According to this criterion, a model is recommended if it minimizes the term given by

$$BIC = -2 (\log (L(\hat{\theta}))) + p \log (n) \tag{9}$$

where $L(\hat{\theta})$ denotes the maximized likelihood function corresponding to a model indexed with the parameter θ, n denotes the total number of observations and p is the dimension of the concerned model. First term supports the more complex model and second term supports a simpler model having low dimensions. It is obvious from (9) that BIC is free from any prior information and it penalizes the complexity of the model according to its dimension. It is a consistent measure in the sense that the probability of selecting the correct model tends to unity as the number of observations approaches to infinity although it suffers from a disadvantage that it is a valid measure only for a well-behaved model. The quantity $\hat{\theta}$ in (10) can be replaced by posterior mode if the prior is vague. Similarly, an extension of BIC to censored data problems is routine if one employs sample-based approaches, in particular the Gibbs sampler, and replaces the corresponding censored data with their estimates obtained through Gibbs run.

4. Numerical Illustration

For numerical illustration, we considered a real data set on survival times of patients with stage III non-small cell lung cancer (NSCLC). The data were the results of phase III clinical trial conducted by Cancer and Leukemia group B (CALGB) in United States of America from May 1984 to May 1987 in the form of five interim analyses. The objective of the study included the comparison of two cancer therapies, that is, CT+RT on one hand and RT alone on the other. These clinical trials mostly used log rank tests and Kaplan-Meier plot (see, for example, Lawless (2002)) for the comparison of two therapies at each interim analysis. After the trial stopped enrolling the new patients, enrolled patients were followed up until the summer of 1992 and this data set was finally analyzed by Li (1994), Qian et al. (1996), among others. The complete description of the entire study is given in Li (1994) (see also Qian (1994)). We shall use the final data of 1992 to carry the analysis proposed in this paper. The data set in an ordered form is summarized in Table 1 where asterisk with an observation denotes the censored value.

Therapy	Survival times in days
RT	0.27, 0.37, 1.23, 1.27, 1.27, 2.27, 2.30, 2.40, 2.60, 2.73, 2.87, 2.93, 2.97, 3.37, 3.57, 3.63, 4.23, 4.40, 4.50, 4.83, 5.33, 6.00, 6.10, 6.10, 6.77, 6.87, 6.90, 7.17, 7.50, 7.57, 7.63, 7.67, 8.13, 8.30, 8.53, 8.57, 8.90, 9.50, 9.67, 10.13, 10.27, 10.47, 10.53, 10.67, 10.67, 10.83, 12.63, 12.67, 12.77, 13.10, 13.23, 14.20, 15.00, 15.20, 15.33, 15.83, 16.10, 16.23, 16.87, 17.50, 18.10, 19.73, 19.77, 19.93, 21.43, 23.30, 23.40, 31.20*, 31.93, 32.90, 42.47*, 44.13, 45.40, 62.50*, 64.87*, 73.43*, 83.77* (n=77, r=71)
CT+RT	0.20, 1.83, 2.70, 3.13, 3.90, 3.97, 4.03, 4.50, 5.03, 5.20, 5.93, 6.07, 6.27, 6.33, 6.47, 6.57, 6.70, 7.00, 7.00, 7.20, 7.47, 7.53, 7.97, 8.33, 8.73, 9.03, 9.43, 9.47, 9.50, 9.80, 10.03, 10.10, 10.97, 11.40, 11.67, 12.03, 12.83, 13.30, 13.73, 14.07, 14.57, 15.57, 16.40, 16.53, 16.53, 16.87, 17.23, 17.47, 18.13, 18.53, 18.93, 19.03, 19.07, 20.47, 20.67, 21.20, 23.00, 23.43, 28.83, 39.47*, 40.27, 46.90, 47.83, 48.07, 52.60*, 52.67*, 55.03*, 55.73*, 55.77*, 56.67*, 57.43*, 59.03*, 62.37, 62.40, 66.07*, 66.33*, 69.13* 73.93* (n=78, r=65)

Table 1: Survival times of NSCLC patients receiving two different therapies

Li (1994) and Qian et al. (1996) (see also Qian (1994)) assumed exponential and Weibull models, respectively, for analyzing the data corresponding to RT. For data corresponding to combined therapy CT+RT, they however assumed the same models but with a restrictive assumption on the scale parameters. The authors assumed that the logarithm of the ratio of the scale parameters for the models corresponding to CT+RT and those corresponding to RT is constant. This assumption makes sense with regard to the exponential model as the ratio becomes simply the hazard ratio but it is certainly not appealing with the assumption of Weibull model. The authors finally considered a comparison of two therapies in a Bayesian framework based on estimated survival functions and concluded that the combined therapy CT+RT does provide a significant improvement over RT in terms of the survival of the patients. We perhaps do not find any significant work where an unrestricted analysis of the data is given and also where an attempt has been made on model comparison.

In order to formalize our analysis, we first simulated the posterior (5) separately for RT and CT+RT data using the Gibbs sampler algorithm. The details of the implementation of the Gibbs sampler algorithm for censored data situations can be seen in Section 2. For the initial values of α and β, we considered maximum likelihood estimates using the corresponding data sets whereas for the initial values of unknown censored observations x_j, we used the corresponding truncation points c_j ($j= r+1,\ldots, n$). We next considered a single chain of long run through Gibbs algorithm and the convergence monitoring was done for both (α,β) and unknown

censored observations using the ergodic averages. Finally, samples of size 1000 corresponding to each variate were picked up from the generating chain using equidistant (every 10^{th}) outcomes. The gaps were chosen to make serial correlation negligibly small. These samples can be regarded as random samples from the distributions of the corresponding unknowns. Thus the sample-based inferences can be easily drawn once the samples for the corresponding unknowns are made available. Sample based estimates in the form of posterior modes are shown in Table 2. The table also provides the estimated modal values of the corresponding unknown censored observations. All these estimates are based on samples of size 1000 from each of the unknowns and will be used for further inferences. The estimated survival curve corresponding to each therapy is shown in Figure 1 when the underlying model is Weibull distribution. These estimated curves are based on the modal values of the corresponding estimates and have been drawn using R software.

Variate	Estimates corresponding to	
	RT	CT+RT
α	15.421	26.669
β	0.968	0.926
Censored observations (in an ordered form)	38.554, 44.401, 69.817, 70.358, 77.294, 106.001	46.851, 64.093, 70.201, 71.579, 72.375, 72.624, 72.807, 72.955, 74.980, 77.516, 84.064, 84.382, 90.600

Table 2: Estimates based on sample of size 1000 corresponding to RT and CT+RT data when the underlying modelling assumption is Weibull

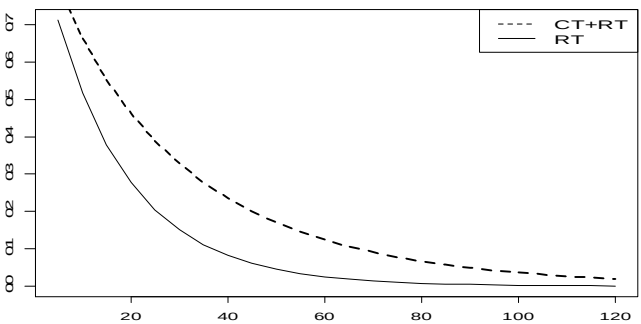

Figure 1: Estimated survival functions based on RT and CT+RT data when the underlying modelling assumption is Weibull.

A number of results can be reported likewise once the samples are made available but we shall concentrate on two important findings based on Table 2 and Figure 1. First, the estimated posterior mode of β is quite close to unity for both RT and CT+RT which, in turn, provides an impression that exponential model is a strong candidate for both the data sets. Second, the survival curve corresponding to CT+RT is, in general, higher than the corresponding curve for RT which shows that the combined therapy provides a better survival to the patients suffering from NSCLC (see Figure 1). A similar conclusion was drawn by earlier researchers but the simplicity of our procedure is certainly an added advantage.

Guided by our first conclusion above, we propose to consider the exponential modelling assumption (that is β=1 in the previous formulation) as well for the proposed

analysis and then intend to provide a comparison of the two models so that a better one can be recommended. The corresponding posterior is given in (6). We separately implemented our strategy for exponential distribution (see Section 2) on both RT and CT+RT data and generated a single chain of long run. It is to be noted that exponential distribution has a single parameter but several unknowns in the form of censored observations. We picked up samples of size 1000 from the corresponding distributions of each of the unknowns in a way similar to what has been discussed for Weibull modelling assumption. Sample based estimates of λ in the form of posterior modes are shown in Table 3 for both the data sets. The table also provides the estimated modal values of the corresponding unknown censored observations.

Variate	Estimates corresponding to	
	RT	CT+RT
α	14.703	26.371
Censored observations (in an ordered form)	37.974, 44.312, 69.438, 70.155, 76.840, 85.800	51.221, 62.705, 66.262, 68.203, 68.514, 69.725, 69.771, 70.025, 73.482, 76.764, 77.194, 77.668, 85.597

Table 3: Estimates based on sample of size 1000 corresponding to RT and CT+RT data when the underlying modelling assumption is exponential

Figure 2 presents the estimated survival curves for the two therapies using the estimated modal values given in Table 2. These curves are more or less similar to those shown in Figure 1 and provide exactly the same conclusion that was drawn using the Weibull model for the two data sets. That is, combined therapy CT+RT provides significant improvement in the survival of patients in comparison to those who are treated with RT alone. Another important finding is based on the estimates reported for censored observations. It is to be noted that these observations correspond to the patients who failed to report during follow up and, as such, their actual survival times could not be recorded. Based on the estimated modal values of these censored observations, one can at least get an idea of actual survival times for these patients. It is to be noted that exponential model, in general, provides smaller estimated values for the highest ordered censored observations than those based on Weibull modelling assumption. This may not be a striking finding but an underestimated value is certainly a good indicative for deciding improved therapy. Moreover, we should not expect enough survival for such category of patients who left the study (or treatment) after surviving for a longer duration of time (see Tables 1-3).

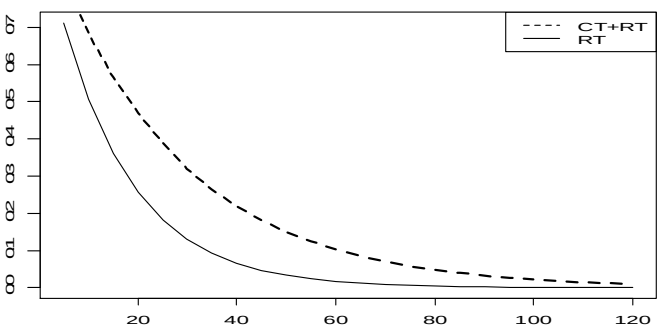

Figure 2: Estimated survival functions based on RT and CT+RT data when the underlying modelling assumption is exponential.

The conclusion in favour of combined therapy CT+RT was noted by earlier authors as well who considered the two data sets and the above two modelling assumptions though their approaches were slightly different (see, for example, Li (1994) and Qian et al. (1996)). Moreover, none of the authors tried comparing the models or testing β against unity. They simply considered either exponential or Weibull model without giving any justification of the fact that why they are using these models. Besides, they took some unrealistic assumptions especially when they used Weibull model for the reported data.

In the above study we advocated in favour of the exponential model simply because the estimated β, when Weibull model was considered to be a true model, was found to be close to unity. This was obviously a vague criterion and, therefore, we propose to consider a comparison of the two models based on BIC. Before we begin, we shall however study compatibility of the models using posterior predictive p-value obtained by considering the Bayesian version of chi-square discrepancy measure. In order to obtain the same, we first considered the sample-based output, each of size 1000, of the unknown parameters involved in both exponential and Weibull models separately for the two data sets. We replaced the censored observations with the corresponding estimated modal values as mentioned before. Using each observation of the sample-based output in the first step; we then generated 1000 predictive samples with sizes equal to those of observed data and correspondingly obtained D_2 and D_1 based on predictive and observed data sets, respectively. Posterior predictive p-values based on chi-square discrepancy measure were then obtained using (7) as proportion of times D_2 exceeds D_1 (see also Section 3). The values were found to be 0.458(0.711) and 0.248(0.649), respectively, when Weibull and exponential models were considered to be the true models. The bracketed values correspond to CT+RT data.

It is obvious from the results that both the models are compatible for the two data sets and none can be rejected. It is, however, important to mention here a few things before we close our discussion. First, we are aware with the fact that model compatibility study based on posterior predictive p-values has invited a few criticisms (see, for example, Bayarri and Berger (1998)) especially the fact that it incorporates double use of data, once in simulating posteriors and second, in obtaining the p-value. This can be an important demerit but in either case it can be used as a preliminary tool as mentioned earlier (see also Upadhyay and Mukherjee (2008)). Second, once exponential model is justified for the data in hand, the Weibull model being a more complex generalization is certainly justified. Therefore, we do not need to consider the compatibility of latter but we have done simply for the sake of completeness of our study. A word of remark: the model compatibility study or the p-values should not be taken as model selection tool so we are not recommending any particular model at this stage. The parsimony principle, however, suggests that since both the models are compatible, we should go with the simple exponential model.

To complete the study for recommending a model, we evaluated BIC for the two models. These values were found to be 576.36(658.53) and 583.57(668.38), respectively, for exponential and Weibull models where the bracketed values correspond to those based on CT+RT data. Since the values corresponding to exponential model are, in general, smaller to those corresponding to Weibull model, we may safely recommend exponential model for both RT and CT+RT data although the values corresponding to the two models are not wide apart from each other. The same conclusion was drawn by parsimony principle as well, which we advocated earlier when both the models were found compatible with the data but making a conclusion after comparing the two models provides an added safety.

5. Conclusion

Advanced stage cancer patients are usually treated with RT or combined CT+RT. It has been a long and continuous debate among the medical practitioners that which therapy actually provides a better survival. A number of studies are performed earlier but most of these studies do not provide any convincing way for dealing with censored data although the studies

have shown that CT+RT does provide better survival. The present study provides a similar conclusion based on Weibull modelling assumption, deals systematically with censored data, and successfully obtains the estimated survival times for such censored data situations.

Weibull distribution is quite flexible and perhaps because of the same reason it was used earlier by Qian et al. (1996) for the data in hand although he did not provide any convincing argument for considering this model. Our proposed study not only examines the compatibility of the Weibull model with the data but also examines the suitability of exponential model so that the resulting inferences become easy to draw if the same is recommended. It has been successfully shown after comparing the two models that unnecessary complication by assuming Weibull model can be avoided. This is what parsimony principle also suggests after getting compatibility of the two models.

References

1. Bayarri, M.J. and Berger, J.O. (1998). Quantifying surprise in the data and model verification, Bayesian Statistics 6, Bernardo, J.M., Berger, J.O., David, A.P., and Smith, A.F.M., Eds, Oxford University Press, p.53-82.
2. Cox, D.R. (1972). Regression models and life tables, J. Roy. Statist. Soc. B. 34, p. 187-220.
3. Gamerman, D. and Lopes, H.F. (2006). Markov Chain Monte Carlo: Stochastic Simulation for Bayesian Inference, 2nd ed., Chapman & Hall / CRC.
4. Gelman, A. (2004). Parameterization and Bayesian modeling, J. Amer. Statist. Assoc., 99, p. 537-545.
5. Gelman, A., Carlin, J.B., Stern, H.S., Rubin, D.B. (2003). Bayesian Data Analysis. 2nd ed., Chapman & Hall, London.
6. Gelman, A., Meng, X.L. and Stern, H.S. (1996). Posterior predictive assessment of model fitness via realized discrepancies, Statistica Sinica, 6, 733-807.
7. Ibrahim, J.G., Ming-Hui Chen and Sinha, D. (2001). Bayesian Survival Analysis. Springer-Verlag, New York.
8. Ibrahim, J.G., Ming-Hui Chen and Sinha, D. (2004). Bayesian methods for joint modeling of longitudinal and survival data with applications to cancer vaccine trials, Statistica Sinica, 14, p. 863-883.
9. Kalbfleisch, J.D. and Prentice, R.L. (2002). The Statistical Analysis of Failure Time Data. Wiley, NewYork.
10. Kaplan, E.L., and Meier, P. (1958). Non parametric estimation from incomplete observations, J. Amer. Statist. Assoc., 53, p. 457-81.
11. Lawless, J.F. (2002). Statistical Models and Methods for Lifetime Data. 2nd ed., Wiley, New York.
12. Lee, E.T. and Wang, J.W. (2003). Statistical Mehods for Survival Data Analysis, Wiley, New York.
13. Li, C. (1994). Comparing Survival Data for Two Therapies: Nonhierarchical and Hierarchical Bayesian Approaches, Unpublished Ph.D. thesis, Department of Statistics, Duke University, USA.
14. Mantel, N. (1966). Evaluation of survival data and two new rank order statistics arising ill its consideration, Cancer Chemotherapy Reports, 50, p.163-70.
15. Martz, H.F. and Waller, R.A. (1982). Bayesian Reliability Analysis. Wiley, New York.
16. Qian, J. (1994). A Bayesian Weibull Survival Model, Unpublished Ph.D. thesis, Department of Statistics, Duke University, USA.
17. Qian, J., Stangl, D., and George, S. (1996). A Weibull model for survival data: using predictions to decide when to stop a clinical trial. Bayesian Biostatistics, Berry, D. and Stangl, D., Eds., Marcel Dekker, p.187-205.
18. Schwartz, G. (1978). Estimating the dimension of a model. Ann. Statist., 6, p.461-68.
19. Singpurwalla, N.D. (2006). Reliability and Risk: A Bayesian Perspective. Wiley, New York.

20. Sinha, D. and Dey, D.K. (1997). Semiparametric Bayesian Analysis of Survival Data, J. Amer. Statist. Assoc., 92, p. 1195-1212.
21. Smith, A.F.M. and Roberts, G.O. (1993). Bayesian computation via the Gibbs sampler and related Markov chain Monte Carlo Methods, J. Roy. Statist. Soc. B, 55, p. 2-23.
22. Upadhyay, S.K., Vasistha, N. and Smith, A.F.M. (2001). Bayes inference in life testing and reliability via Markov chain Monte Carlo simulation, Sankhya A, 63, p. 15-40.
23. Upadhyay, S.K. and Mukherjee, B. (2008). Assessing the value of the threshold parameters in the Weibull distribution using Bayes paradigm, IEEE Trans. Reliab., 57. p. 489-497.
24. Upadhyay, S.K. and Peshwani, M. (2008). Posterior analysis of lognormal regression models using the Gibbs sampler, Statistical Papers, 49, p. 59-85.