Published in final edited form as:

Stat Med. 2009 July 20; 28(16): 2077–2094. doi:10.1002/sim.3598.

A new generalization of Weibull distribution with application to a breast cancer data set

Abdus S. Wahed^{*}, The Minh Luong, and Jong-Hyeon Jeong

Department of Biostatistics, Graduate School of Public Health, University of Pittsburgh, Pittsburgh, PA 15261, USA.

SUMMARY

In this article, we propose a new generalization of the Weibull distribution, which incorporates the exponentiated Weibull distribution introduced by Mudholkar and Srivastava [1] as a special case. We refer to the new family of distributions as the beta-Weibull distribution. We investigate the potential usefulness of the beta-Weibull distribution for modeling censored survival data from biomedical studies. Several other generalizations of the standard two-parameter Weibull distribution are compared with regards to maximum likelihood inference of the cumulative incidence function, under the setting of competing risks. These Weibull-based parametric models are fit to a breast cancer dataset from the National Surgical Adjuvant Breast and Bowel Project (NSABP). In terms of statistical significance of the treatment effect and model adequacy, all generalized models lead to similar conclusions, suggesting that the beta-Weibull family is a reasonable candidate for modeling survival data.

Keywords

Competing risk; Cumulative incidence; Weibull distribution

1. INTRODUCTION

A suitable parametric model is often of interest in the analysis of survival data, as it provides insight into characteristics of failure times and hazard functions that may not be available with non-parametric methods. The Weibull distribution [2] is one of the most commonly used families for modeling such data. However, only monotonically increasing and decreasing hazard functions can be generated from the classic two-parameter Weibull distribution; as such this two-parameter model is inadequate when the true hazard shape is of unimodal or bathtub nature.

The widely used Kaplan-Meier product-limit estimator is a flexible method to model survival data but is noted to often be inefficient [3]. Other semi-parametric methods, such as proportional hazards modelling, require assumptions that may not be plausible in many situations [4]. Meanwhile, various parametric techniques have been developed to incorporate the wide variety of patterns in survival data. Some of the proposed parametric models have incorporated a shape parameter into the classic Weibull distribution to account for additional possible hazard shapes. One of the first models of this type was proposed by

Kalbfleisch [5], but this model may be impractical in the presence of censored data as it often requires the evaluation of an incomplete gamma integral or beta ratio.

Mudholkar and Srivastava [1] introduced an exponentiated version of the Weibull model that included an additional shape parameter. The distribution has a closed form of probability density, survival, and hazard functions, that are flexible and able to generate a wide variety of frequently observed hazard shapes, including unimodal and bathtub. It can be shown that this extension of the Weibull can be achieved through a simple application of the probability integral transform, using the densities of Beta and two-parameter Weibull distributions. Mudholkar, Srivastava, and Kollia [6] proposed another generalization of the Weibull model, which is able to generate similar types of hazard shapes as the exponentiated model; however irregularities may arise as the support of the distribution becomes dependent on the parameter space. A four-parameter generalization of the Weibull distribution was introduced by Jeong [7] based on stable distributions proposed by Hougaard [8], induced under the semiparametric frailty model. This model can also be viewed as a generalization of the model of Mudholkar et al. [6].

A decade has passed since the exponentiated Weibull was proposed, however a limited amount of literature is available on its application to survival data. This article introduces the family of beta-Weibull distributions encompassing regular two-parameter Weibull and the exponentiated Weibull distributions. It focuses on the potential use of the beta-Weibull distribution as a model for survival data in the presence of competing risks, in comparison with other generalizations of the Weibull model.

In the analysis of breast cancer data, it is often of interest to investigate only a subset of any first events. For example, investigators may be interested in investigating local or regional recurrence of the original breast cancer, considering other events such as distant recurrence, new primary cancers other than breast, or deaths prior to any disease [9]. In this case, a competing risks situation arises, i.e. local or regional events against other events. It is well-known that the cumulative incidence function estimates the proportion of local or regional recurrences correctly as a function of time in the presence of other events [5,10]. In this paper, the cumulative incidence function will be parameterized and compared by using various existing Weibull distributions along with the proposed beta-Weibull distribution. These parametric approaches will be illustrated in a data set from one of the phase III breast cancer clinical trials performed by National Surgical Adjuvant Breast and Bowel Project (NSABP).

The article is organized as follows. In Section 2, the beta-Weibull family of distributions is introduced along with an overview of existing Weibull-based families. In Section 3, maximum likelihood estimation and corresponding inferential procedures are described for the generalized Weibull models in the setting of competing risks. In Section 4, these models are fitted to a previously analyzed breast cancer data set, and performance of the Weibull-based approaches are evaluated by using conventional model assessment methods. We conclude with a few remarks in Section 5.

2. THE WEIBULL DISTRIBUTION AND ITS EXTENSIONS

The classic two-parameter Weibull distribution has the probability density function (pdf)

$$g_1(x; \gamma, \beta) = \frac{\gamma}{\beta} (x/\beta)^{\gamma-1} e^{-(x/\beta)^\gamma}, 0 < x < \infty, 0 < \gamma, \beta < \infty \quad (1)$$

with its cumulative distribution function being

$$G_1(x; \gamma, \beta) = 1 - e^{-(x/\beta)^\gamma}, 0 < x < \infty. \quad (2)$$

While this convenient distribution can be easily incorporated into practical tools for survival analysis such as proportional hazards and accelerated failure time models, only monotonically increasing or decreasing hazard shapes can be properly estimated. To overcome this limitation, several generalized versions of the classic two-parameter Weibull model have been proposed. In this section we describe two such models that require at most two extra parameters to be estimated.

2.1. A three-parameter extension

A three-parameter generalization [6] of the Weibull distribution is defined by the survival function

$$S(x; \gamma, \beta, \lambda) = \{1 + (x/\beta)^\gamma / \lambda\}^{-\lambda}, 0 < \gamma, \beta < \infty \quad (3)$$

and the corresponding hazard function

$$h(x; \gamma, \beta, \lambda) = \frac{\gamma}{\beta} \frac{(x/\beta)^{\gamma-1}}{1 + (x/\beta)^\gamma / \lambda}.$$

The regular case of this generalized Weibull distribution occurs when $\lambda \geq 0$ with the density having support $(0, \infty)$, generating monotonically decreasing or unimodal hazard functions. In particular, the distribution approaches the two-parameter Weibull as $\lambda \rightarrow \infty$. When $\lambda < 0$, monotonically increasing or bathtub hazard shapes are generated, but the support of the density becomes parameter-dependent, within the range $(0, \beta(-\lambda)^\alpha)$. Similar to other non-regular densities, in these situations the likelihood may become unbounded and thus maximum likelihood estimates may not exist, or alternately, although the maximum likelihood estimates exist, their asymptotic properties may not hold for the classical asymptotic theory [11].

2.2. A four-parameter extension

A four-parameter Weibull model was proposed by Jeong [7]. This generalization of the Weibull distribution incorporates an additional parameter τ to the three-parameter version proposed by Mudholkar [6] that was previously discussed. This generalization can be characterized by the survival function

$$S(x; \gamma, \beta, \lambda, \tau) = \exp \left[- \frac{\lambda^{1-\tau} \{(x/\beta)^\gamma + \lambda\}^\tau - \lambda}{\tau} \right], 0 < x < \infty, 0 < \gamma, \beta, \lambda < \infty, -\infty < \tau < \infty$$

with the corresponding hazard function

$$h(x; \gamma, \beta, \lambda, \tau) = \frac{\gamma (x/\beta)^\gamma \lambda^{1-\tau}}{x [(x/\beta)^\gamma + \lambda]^{1-\tau}}.$$

This distribution reduces to Mudholkar's generalized Weibull distribution as τ approaches to zero. As one would expect, it covers a wider range of shapes than the three-parameter extension and is regular, regardless of the parameter values.

2.3. The beta-Weibull distribution

It has been demonstrated by Wahed [12] and Ferreira and Steel [13] that any parametric family of distribution can be incorporated into larger families through an application of the probability integral transform. Specifically, given two valid probability density functions (pdf) $g_1(\cdot)$ and $g_2(\cdot)$, with the latter having support on the unit interval, a third valid density function $f(\cdot)$ may be obtained by applying the equation

$$f(x)=g_2(G_1(x))g_1(x), \quad (4)$$

where $G_1(\cdot)$ is the cumulative distribution function (cdf) corresponding to $g_1(\cdot)$. Using this technique, we obtain yet another generalization of the two-parameter Weibull distribution as described below.

To obtain the appropriate extension of the Weibull distribution, we considered $g_1(\cdot)$ as the two-parameter Weibull pdf given by (1) and $g_2(\cdot)$ as a two-parameter beta distribution with pdf

$$g_2(u;\alpha_1, \alpha_2) \propto u^{\alpha_1-1}(1-u)^{\alpha_2-1}, 0 < u < 1, 0 < \alpha_1, \alpha_2 < \infty. \quad (5)$$

Using (1) and (5) in (4) we obtain a new probability density function

$$f(x;\alpha_1, \alpha_2, \beta, \gamma) = \frac{1}{B(\alpha_1, \alpha_2)} \frac{\gamma}{\beta} (1 - e^{-(x/\beta)^\gamma})^{\alpha_1-1} e^{-\alpha_2(x/\beta)^\gamma} (x/\beta)^{\gamma-1}, \quad (6)$$

where $0 < x < \infty$, $0 < \alpha_1, \alpha_2, \gamma, \beta < \infty$ and $B(\alpha_1, \alpha_2) = \Gamma(\alpha_1)\Gamma(\alpha_2)/\Gamma(\alpha_1+\alpha_2)$. We will refer to this distribution as the beta-Weibull distribution throughout the article. The density given by (6) is a proper density and when $\alpha_1 = \alpha_2 = 1$, it reduces to the two-parameter Weibull density with scale parameter β and the shape parameter γ . This model is generated entirely based on two commonly used probability distributions, beta and Weibull, and therefore can be easily implemented using conventional statistical software packages. This process of using the beta distribution to generate a new class of distribution has been previously considered by Jones [14]. The survival function of the beta-Weibull distribution cannot be expressed in a closed form; however it may be estimated by evaluating the regularized beta function at the cumulative distribution function of the Weibull distribution. Specifically, the survival function of the beta-Weibull distribution (6) is given by

$$S(x;\alpha_1, \alpha_2, \gamma, \beta) = \frac{1}{B(\alpha_1, \alpha_2)} \int_{1-e^{-(x/\beta)^\gamma}}^1 u^{\alpha_1-1}(1-u)^{\alpha_2-1} du = 1 - I(1-e^{-(x/\beta)^\gamma}; \alpha_1, \alpha_2), \quad (7)$$

where

$$I(v; \alpha_1, \alpha_2) = \frac{\int_0^v u^{\alpha_1-1} (1-u)^{\alpha_2-1} du}{\int_0^1 u^{\alpha_1-1} (1-u)^{\alpha_2-1} du} \quad (8)$$

is the incomplete beta function. The corresponding hazard function is obtained as

$$h(x; \alpha_1, \alpha_2, \beta, \gamma) = \frac{\gamma (1 - e^{-(x/\beta)^\gamma})^{\alpha_1-1} e^{-\alpha_2(x/\beta)^\gamma} (x/\beta)^{\gamma-1}}{\beta (1 - I(1 - e^{-(x/\beta)^\gamma}; \alpha_1, \alpha_2))}. \quad (9)$$

The survival and hazard functions are plotted in Figures 1 and 2, respectively. The hazard plots show that this model can generate unimodal, v-shaped and bathtub shaped hazards. The parameters, α_1 and α_2 , are related to the beta-component of the distribution; as a result, the mean and variance of the distribution increases as α_1 increases, as for the beta distribution. Similarly, the survival probability is higher for larger α_1 . The shape of the hazard function is influenced by the parameter α_1 and γ , while β remains as a scale parameter. The second parameter from the beta distribution, α_2 , also primarily acts as a scale parameter as it is absorbed into the Weibull function through the exponent term $e^{-(x/\beta)^\gamma}$.

2.4. Restricted beta-Weibull distribution

The skewing mechanism based on the beta distribution has also been considered in other settings. For example, Ferreira and Steel [15] used a restricted version of the beta distribution to obtain skewed versions of multivariate distributions where they imposed the restriction $\alpha_2 = 1/\alpha_1$ on the beta-parameters. We impose the same constraint on (6) to obtain what we will refer to as a restricted beta-Weibull distribution.

2.5. Exponentiated Weibull distribution

Notice that when $\alpha_2 = 1$, the beta-Weibull distribution defined by (6) reduces to

$$f(x; \alpha_1, \beta, \gamma) = \frac{\alpha_1 \gamma}{\beta} (1 - e^{-(x/\beta)^\gamma})^{\alpha_1-1} e^{-(x/\beta)^\gamma} (x/\beta)^{\gamma-1}, 0 < x < \infty, 0 < \gamma, \beta, \alpha_1 < \infty, \quad (10)$$

which is the so-called exponentiated Weibull distribution [1,16]. For this specific case, both survival function and hazard function are evaluable in closed forms and are given by

$$S(x; \alpha_1, \beta, \gamma) = 1 - \{1 - e^{-(x/\beta)^\gamma}\}^{\alpha_1} \quad (11)$$

and

$$h(x; \alpha_1, \beta, \gamma) = \frac{\alpha_1 \gamma}{\beta} \frac{(1 - e^{-(x/\beta)^\gamma})^{\alpha_1-1} e^{-(x/\beta)^\gamma} (x/\beta)^{\gamma-1}}{1 - \{1 - e^{-(x/\beta)^\gamma}\}} \quad (12)$$

respectively.

Setting $\alpha = \alpha_1$ for notational convenience, the derivative of the hazard function with respect to x can be expressed as

$$h'(x; \alpha, \beta, \gamma) = m(x; \alpha, \beta, \gamma) \left\{ \gamma (x/\beta)^\gamma \left[\alpha e^{-(x/\beta)^\gamma} - 1 + (1 - e^{-(x/\beta)^\gamma})^\alpha \right] + (\gamma - 1)(1 - e^{-(x/\beta)^\gamma}) [1 - (1 - e^{-(x/\beta)^\gamma})^\alpha] \right\}, \quad (13)$$

where

$$m(x; \alpha, \beta, \gamma) = \frac{\alpha \gamma (1 - e^{-(x/\beta)^\gamma})^{\alpha-2} e^{-(x/\beta)^\gamma} (x/\beta)^{\gamma-2}}{\beta^2 \{1 - [1 - e^{-(x/\beta)^\gamma}]^\alpha\}^2}. \quad (14)$$

Notice that $m(x; \alpha, \beta, \gamma)$ is non-negative regardless of any values of x or parameters.

As discussed in Mudholkar and Srivastava [1], for $\alpha < 1$, the hazard shapes are unimodal if $\alpha < 1/\gamma$ and monotonically decreasing otherwise; likewise when $\alpha > 1$ bathtub shapes are generated for $\alpha > 1/\gamma$, with monotonically increasing shapes otherwise. The hazard function is monotonically increasing when $\alpha \geq 1, \gamma > 1$, while $\alpha \leq 1, \gamma < 1$ implies a monotonically decreasing function; when both parameters are equal to 1, the distribution reduces to the exponential distribution (constant hazard). However, when one of the parameters is greater than 1 and the other is less than 1, more shapes can be generated. If $\alpha > 1$ and $\gamma < 1$, a unimodal hazard shape is generated if $\alpha > 1/\gamma$ and is monotonically decreasing otherwise; if $\alpha < 1$ and $\gamma > 1$, bathtub shapes are generated when $\alpha < 1/\gamma < 1$ and are monotonically increasing otherwise. Behavior of the hazard function with respect to the parameters is depicted in Figure 3 for values of α less than 1 (left) and greater than 1 (right). Similarly, the hazard shapes for different values of the shape parameter γ for $\alpha < 1$ (left) and $\alpha > 1$ (right) are plotted in Figure 4. Different types of hazard shapes may be obtained by altering this parameter while keeping the other two parameters fixed.

While both exponentiated Weibull distribution and three-parameter generalized Weibull (Section 2.1) reduce to the two-parameter Weibull distribution when λ approaches ∞ and λ approaches 1, there is no clear link between the parameters that comprise these distribution functions. This is apparent from the behavior of respective hazard functions; $\lambda < 0$ implies either a monotonically increasing or bathtub shape for the generalized Weibull, while $\alpha > 1$ may result in a monotone hazard function that is increasing or decreasing, or a unimodal function. Similarly, $\lambda > 0$ produces a monotonically decreasing or unimodal shape for the generalized Weibull, whereas if $0 < \alpha < 1$ for the exponentiated Weibull, a bathtub shape is possible, along with monotonically decreasing or increasing shapes. The exponentiated Weibull is a generalization of the Burr Type X distribution $S(x) = [1 - e^{-(x/\beta)^\gamma}]^\alpha$, while the family of the three-parameter generalized Weibull includes that of the Burr Type XII family of distributions [17] as a special case.

3. INFERENCE UNDER COMPETING RISKS

When multiple cause-specific events, or competing risks, are present, the cumulative incidence function correctly assesses the cumulative probability of a particular cause-specific event while taking into account other types of events [5]. This prevents the cumulative probability of a sub-distribution of cause specific events of interest from being overestimated, which would occur if failures due to other causes are considered as censored [10]. For data consisting of only two types of cause-specific events, the cumulative incidence function for cause-specific events of type 1, under the assumption of cumulative hazards, is defined as

$$F_1(t; \phi) = \int_0^t S_2(u; \phi_2) f_1(u; \phi_1) du, \quad (15)$$

where $S_2(\cdot, \phi_2)$ and $f_1(\cdot, \phi_1)$ are survival function for events of type 2 and probability density function for events of type 1, respectively. In this paper, we will focus on parameterizing $S_2(\cdot, \phi_2)$ and $f_1(\cdot, \phi_1)$ via Weibull-based models to estimate the cumulative incidence function.

The set of parameter vector $\phi = (\phi_1, \phi_2)$ of a given Weibull-based model for a given data set can be estimated through the maximum likelihood method. For the cause-specific event of type k ($k = 1, 2$), the observed data for the i th individual is denoted by (x_i, δ_{ki}) , where x_i represents the event time (failure/censoring) and δ_{ki} is an indicator variable, taking value 1 when the k th cause-specific event occurs. The likelihood function for the k th cause-specific event can be expressed as

$$l_k(\phi_k) = \prod_{i=1}^n \{f_k(x_i, \phi_k)\}^{\delta_{ki}} \{S_k(x_i, \phi_k)\}^{(1-\delta_{ki})}. \quad (16)$$

For all the Weibull-based models described in the previous section, the likelihood function (16) cannot be maximized analytically. Thus, an iterative procedure such as the Newton-Raphson method needs to be used to obtain maximum likelihood estimates for the parameter vector ϕ_k for k th cause-specific event. We have used the S-Plus function *nlminb* to minimize $-l_k(\phi_k)$ to obtain the maximum likelihood estimates of ϕ_k . The initial values of the parameters for more complicated models were chosen from the results of fitting a classical two-parameter Weibull model to the data, while the initial values for both skewing beta parameters were set to 1 (equivalent to the classical Weibull). When the Newton-Raphson procedure is used to obtain the MLEs, in order to facilitate convergence it is suggested to set the options of the minimization procedure such that the iterative step length is of similar magnitude between each parameter [18]. Once the MLE of ϕ_k is obtained, the variance of these estimates are calculated by inverting the observed information matrix.

The corresponding estimators for the cumulative incidence and survival functions are obtained by substituting the parameter estimates into the respective functions. For instance, the cumulative incidence function (15) is estimated by

$$\widehat{F}_1(t; \widehat{\phi}) = \int_0^t S_2(u; \widehat{\phi}_2) f_1(u; \widehat{\phi}_1) du. \quad (17)$$

The standard errors of the estimated cumulative incidence function and the survival probability is then obtained through an application of the multivariate delta method. For instance, an estimate of the approximate variance of the estimated cumulative incidence function $F_1(\cdot)$ is given by (see [7] for derivation)

$$\widehat{var}(\widehat{F}_1(t; \widehat{\phi})) = \left(\frac{\partial F_1(t; \phi)}{\partial \phi} \right) \bigg|_{\phi = \widehat{\phi}} var(\widehat{\phi}) \left(\frac{\partial F_1(t; \phi)}{\partial \phi} \right)^T \bigg|_{\phi = \widehat{\phi}}. \quad (18)$$

Equality of survival or cumulative incidence functions between independent treatment groups can then be tested using the Wald test. For testing the equality of two parametric

functions $\theta^{(1)}(t)$ and $\theta^{(2)}(t)$ from two different treatments at a specific time point t , the appropriate test statistic is

$$Z = \frac{\widehat{\theta}^{(1)}(t) - \widehat{\theta}^{(2)}(t)}{\sqrt{\widehat{\text{var}}(\widehat{\theta}^{(1)}(t)) + \widehat{\text{var}}(\widehat{\theta}^{(2)}(t))}}, \quad (19)$$

where $Z \sim N(0, 1)$. For instance, the appropriate Wald test statistic to compare two cumulative incidence functions at time t will have the form

$$Z = \frac{\widehat{F}_1^{(2)}(t; \phi) - \widehat{F}_1^{(1)}(t; \phi)}{\sqrt{\widehat{\text{var}}(\widehat{F}_1^{(1)}(t; \phi)) + \widehat{\text{var}}(\widehat{F}_1^{(2)}(t; \phi))}}. \quad (20)$$

4. APPLICATION OF WEIBULL-BASED MODELS TO BREAST CANCER DATA

We fitted the models to a data set from a breast cancer study (Protocol B-20) performed by the National Surgical Adjuvant Breast and Bowel Project Study (NSABP) [19]. The study was designed to test if the addition of chemotherapy to tamoxifen would result in increased disease-free survival compared to tamoxifen alone. Here the disease-free survival events include breast cancer recurrences in local, regional, or distant sites, second primary cancers other than breast, or death prior to previously mentioned events, whichever occurs first. In this section, the cause-specific events of interest will be the local or regional recurrences, and the other competing events consist of distant recurrences, second primaries, and death without evidence of disease (see Figure 5). A total of 2,363 patients were randomized to tamoxifen, CMF (a chemotherapy containing the alkylating agent cyclophosphamide) or CMFT (CMF plus tamoxifen). The sample consists of 770 of the patients in the tamoxifen group and 766 in the CMFT group. We present our data only up to first 10 years as the risk set become sparse afterwards.

4.1. Maximum likelihood estimation

Maximum likelihood estimates for the parameters from different model fits are displayed in Table I for the cause-specific models, and Table II for the disease-free model. The Akaike Informational Criterion (AIC) is also provided. For the 2-parameter, 3-parameter, 4-parameter and exponentiated versions of the Weibull model, the standard errors were obtained from inverting the matrix of minus the second-derivatives of the log-likelihoods, evaluated at the MLEs. Bootstrapping, using 1000 replicates, was used to estimate the standard errors for the restricted and full beta-Weibull models, for which the second derivatives of the log-likelihood were not available in a closed form.

Within the tamoxifen group, the 4-parameter Weibull model had the best fit as assessed by the AIC followed by the restricted and unrestricted beta-Weibull model, although there were hardly any difference between the AIC's for the latter two models. The two-parameter Weibull model had the largest AIC. By contrast, in the CMFT group there were no significant differences among the fits, regardless of the number of parameters in the model.

The estimated overall hazard function from each of the six model fits is displayed in Figure 6 for each type of the competing events and in Figure 7 for all disease-free events. The non-parametric hazard estimates are life table estimates of the annual hazard rate at the midpoint of each year. Model fits seem generally better for the other events, as local or regional

recurrences were relatively infrequent and had more variable hazard shapes (see more on model adequacy in Section 4.2).

The generalized Weibull models, including the beta-Weibull versions, were able to capture more complex unimodal hazard shapes (local or regional recurrences) as well as the basic monotone shapes (other events). Within the tamoxifen group, according to the non-parametric fit there was an apparent peak at 2–3 years which leveled off after 10 years. This peak was inadequately captured by the two-parameter model, which only produced monotonic shapes; the Weibull extensions produced unimodal shapes that had captured this peak. The curves from the exponentiated model were closest to those from the two-parameter model, tending to produce a more conservative peak at 2–3 years than those from the other four extensions of the Weibull. The hazards from the restricted version of beta-Weibull tended to follow more closely those from the full beta-Weibull model. It is interesting to observe that, as the number of parameters increases, a model tends to pick up the peak earlier in this dataset.

There did not appear to be a similar single peak for the observed events in the CMFT group. As such, monotonically increasing hazard shapes from the two-parameter Weibull model provided an adequate fit to the observed non-parametric curves. The three and four-parameter models very closely followed the two-parameter fit.

4.2. Model adequacy

Model adequacy was investigated using signed deviance residuals with censored observations [20] to assess goodness of fit (Tables III and IV). These are obtained by calculating the expected number of events e_y during each year y as $N_y h_y$, where N_y is the number of patients at risk at the beginning of each year, and h_y is the hazard function integrated over the time interval. The goodness-of-fit between the observed number of events S_y and expected events e_y for each of the models can be assessed through signed deviance residuals (R_y), the squared sum of which approximately follows a χ^2 distribution, where

$$R_y = \sqrt{2} \times \text{sign}(S_y - e_y) \left[S_y \log \frac{S_y}{e_y} + (N_y - S_y) \log \left(\frac{N_y - S_y}{N_y - e_y} \right) \right]^{1/2}. \quad (21)$$

The models tended to provide a relatively poor fit to the data for the endpoint of local or regional recurrences, in large part due to their relative scarcity and the irregular hazard shapes; the models provided better fits for the other events and when all events were grouped together to estimate the overall survival curve.

4.3. Cumulative incidence and survival

The cumulative incidence function was estimated using all six Weibull models for each type of competing events, i.e. local or regional recurrences and other events.

The non-parametric cumulative incidence was estimated through Gray's method [21], while the parametric cumulative incidence functions were obtained through the invariance property of the maximum likelihood estimates from all six Weibull models. To demonstrate how closely the parametric cumulative incidence curves followed the non-parametric estimates, we only presented the estimated cumulative incidence curves from the four-parameter Jeong model and the restricted beta-Weibull model in Figure 8 for each event type. We have chosen these two Weibull models as they had lower AIC values compared to other models. Figure 9 shows the corresponding disease-free survival curves. There were

hardly any differences in cumulative incidence or survival estimates between the two models. The fits from the Weibull models generally followed the non-parametric cumulative incidence curves closely. There is a significant reduction in local or regional recurrences in the CMFT group, but no apparent difference between the two treatment groups in other types of events (Table V). A formal Wald test also concluded that there was a significant improvement in disease-free survival in the CMFT group, which was mainly due to reduction in local or regional recurrences. Both Jeong model and the restricted beta-Weibull model resulted in the same conclusions.

5. CONCLUSION

In this article, we have presented a new generalization of Weibull distribution, the beta-Weibull family, by applying the beta skewing mechanism on the classical two-parameter Weibull distribution. This generalization is easily obtained by transforming the two-parameter Weibull model through a beta distribution and hence can be implemented easily using standard statistical software packages.

The beta-Weibull, two-parameter, exponentiated and other existing generalized models were used to model competing risks and disease-free survival in data from a previously published breast cancer study. The two-parameter Weibull distribution produced simple monotone hazard shapes, as expected, that did not reflect pattern of the hazard shape in the observed data. On the other hand, the extensions of the Weibull model were able to more accurately capture the observed hazard pattern. In terms of statistical significance of the treatment effect and model adequacy, all extended models lead to similar conclusions, suggesting that beta-Weibull family could play a reasonable role as a candidate for modeling survival data, including competing risks. One limitation of the beta-Weibull family is that the survival and the hazard functions cannot be expressed in closed form and hence numerical integration techniques were needed to evaluate them.

REFERENCES

1. Mudholkar GS, Srivastava DK. Exponentiated weibull family for analyzing bathtub failure-rate data. *IEEE Transactions on Reliability*. 1993; 42:299–302.
2. Weibull WA. A statistical distribution of wide applicability. *Journal of Applied Mechanics*. 1951; 18:293–297.
3. Miller RG. What price Kaplan-Meier? *Biometrics*. 1983; 39:1077–1081. [PubMed: 6671119]
4. Cox, DR.; Oakes, D. *Analysis of Survival Data*. New York: Chapman & Hall Ltd; 1984.
5. Kalbfleisch, JD.; Prentice, RL. *The Statistical Analysis of Failure Time Data*. New York: John Wiley & Sons; 1980. p. 321
6. Mudholkar GS, Srivastava DK, Kollia GD. A generalization of the Weibull distribution with application to the analysis of survival data. *Journal of the American Statistical Association*. 1996; 91:1575–1583.
7. Jeong JH. A new parametric family for modelling cumulative incidence function: application to breast cancer data. *J.R. Statistic. Soc. A*. 2006; 169:289–303.
8. Hougaard P. Survival models for heterogeneous populations derived from stable distributions. *Biometrika*. 1986; 73:387–396.
9. Taghian A, Jeong JH, Mamounas E, Anderson S, Bryant J, Deutsch M, Wolmark N. Patterns of locoregional failure in patients with operable breast cancer treated by mastectomy and adjuvant chemotherapy with or without tamoxifen and without radiotherapy: results from five National Surgical Adjuvant Breast and Bowel Project randomized clinical trials. *Journal of Clinical Oncology*. 2004; 22:4237–4239. [PubMed: 15452185]

10. Gooley T, Leisenring W, Crowley J, Storer B. Estimation of failure probabilities in the presence of competing risks: new representations of old estimators. *Statistics in Medicine*. 1999; 18:695–706. [PubMed: 10204198]
11. Smith RL. Maximum likelihood estimation in a class of nonregular cases. *Biometrika*. 1985; 72:67–90.
12. Wahed AS. A general method of constructing extended families of distributions from an existing continuous class. *Journal of Probability and Statistical Science*. 2006; 4:165–177.
13. Ferreira JT, Steel M. A constructive representation of univariate skewed distributions. *Journal of the American Statistical Association, Theory and methods*. 2006; 101:823–829.
14. Jones MC. Families of Distributions arising from Distributions of Order Statistics. *Test*. 2004; 13:1–43.
15. Ferreira JT, Steel M. Model comparison of coordinate-free multivariate skewed distributions with an application to stochastic frontiers. *Journal of Econometrics*. 2007; 137:641–673.
16. Mudholkar GS, Srivastava DK, Freimer M. The exponentiated Weibull family: a reanalysis of the bus-motor-failure data. *Technometrics*. 1995; 37:436–445.
17. Rodriguez RN. A guide to the Burr type XII distributions. *Biometrika*. 1977; 64:129–134.
18. Thisted, RA. *Elements of Statistical Computing*. New York: Chapman & Hall; 1988.
19. Fisher B, Dignam J, Wolmark N, DeCillis A, Emir B, Wickerham DL, Bryant J, Dimitrov N, Abramson N, Atkins J, Shibata H, Deschenes L, Margoless R. Tamoxifen and Chemotherapy for Lymph Node-Negative, Estrogen Receptor-Positive Breast Cancer. *Journal of the National Cancer Institute*. 1997; 89.22:1673–1682. [PubMed: 9390536]
20. Efron B. Logistic regression, survival analysis, and the Kaplan-Meier curve. *Journal of the American Statistical Association*. 1988; 83:414–425.
21. Gray R. A class of k-sample tests for comparing the cumulative incidence of a competing risk. *Annals of Statistics*. 1988; 16:1141–1154.

Figure 1.
Survival function of the beta-Weibull distribution with varying α_1 , for fixed $\gamma = 1$ and $\beta = 2$.

Figure 2.

Unimodal (left; $\alpha_1 = 4$, $\gamma = 0.5$, $\beta = 2.0$) and v-shaped (right; $\alpha_1 = 0.4$, $\gamma = 2.0$, $\beta = 2.0$) hazard shapes with constant α_1 and varying α_2 (left), and α_2 and varying α_1 (right), for the beta-Weibull distribution.

Figure 3.

Hazard shapes for exponentiated Weibull, with varying α and fixed $\beta = 2.0$, for $\gamma = 0.5$ (left), and $\gamma = 2.0$ (right).

Figure 4. Hazard functions of exponentiated Weibull distribution for varying γ and fixed $\beta = 2.0$, for $\alpha = 0.2$ (left), and $\alpha = 2.0$ (right). Three types of hazard shapes are generated in each graph, with respect to γ only.

*In cases of multiple events, first event is used to group the competing events

Figure 5.
Competing risks in breast cancer.

Figure 6.

Estimated event-specific hazard functions; a non-parametric fit (straight lines) and fitted curves from each of the 6 Weibull models, for local or regional recurrences (left) and other events (right), and the tamoxifen (top) and CMFT (bottom) groups.

Figure 7.
Estimated hazard functions for all disease-free events, from a non-parametric fit (straight lines) and fitted curves from each of the six Weibull models, for the tamoxifen (left) and CMFT (right) groups.

Figure 8. Cumulative incidence function estimates and 95% confidence intervals, between tamoxifen (solid) and CMFT (dashed) groups, for local or regional recurrences (top) and other events (bottom) using the restricted beta-Weibull model (left panels) and 4-parameter Jeong model (right panel).

Figure 9.

Estimated disease-free probability in tamoxifen (solid) and CMFT (dashed) groups using the restricted beta-Weibull model (left) and 4-parameter Jeong model (right), along with the corresponding 95% confidence intervals. Non-parametric estimates (non-smooth curve) of disease-free survival was computed using the Kaplan-Meier method.

Table 1
Maximum likelihood estimates of Weibull parameters from breast cancer data (competing events)

Distribution	tamoxifen group			
	specific		other	
	par	mle(se)	mle(se)	AIC
2-parameter Weibull	γ	0.983 (0.130)	1.099 (0.084)	2088.06
	β	126.53 (47.39)	35.84 (4.57)	
3-parameter Weibull	γ	2.434 (0.819)	1.656 (0.305)	2076.84
	β	8.039 (5.217)	12.19 (4.76)	
	λ	0.017 (0.009)	0.128 (0.063)	
4-parameter Weibull	γ	4.763 (4.114)	1.187 (0.271)	2061.16
	β	3.083 (1.464)	10.50 (4.43)	
	λ	0.007 (0.007)	0.003 (0.001)	
	τ	0.133 (0.102)	0.070 (0.032)	
Exponentiated Weibull (<i>beta-Weibull with fixed $\alpha_2 = 1$</i>)	γ	0.130 (0.022)	0.130 (0.015)	2076.66
	β	0.012 (0.027)	0.0003 (0.0006)	
	α	26.41 (10.95)	71.23 (29.97)	
Restricted beta-Weibull (<i>beta-Weibull with fixed $\alpha_2 = 1/\alpha_1$</i>)	γ	0.349 (0.069)	0.422 (0.041)	2070.50
	β	0.012 (0.366)	0.096 (0.547)	
	α	70.02 (44.79)	15.67 (9.18)	
Beta-Weibull	γ	0.448 (0.134)	0.352 (0.147)	2072.12
	β	0.040 (0.096)	0.027 (0.092)	
	α_1	33.80 (174.65)	38.25 (53.24)	
	α_2	0.010 (0.206)	0.062 (0.862)	
CMFT				
			specific	other

tamoxifen group					
Distribution	par	specific		other	
		mle(se)	mle(se)	mle(se)	AIC
Distribution	par	mle(se)	mle(se)	mle(se)	AIC
2-parameter Weibull					
	γ	1.738 (0.445)	1.239 (0.100)	1.239 (0.100)	1585.49
	β	89.43 (53.21)	33.80 (4.18)	33.80 (4.18)	
3-parameter Weibull					
	γ	1.545 (0.775)	1.295 (0.213)	1.295 (0.213)	1587.32
	β	137.61 (245.71)	29.86 (12.47)	29.86 (12.47)	
	λ	-0.043 (0.147)	1.179 (3.947)	1.179 (3.947)	
4-parameter Weibull					
	γ	120.70 (206.07)	20.06 (21.76)	20.06 (21.76)	1589.18
	β	1.602 (0.776)	1.442 (0.439)	1.442 (0.439)	
	λ	2.77 (137.42)	0.011 (0.053)	0.011 (0.053)	
	τ	54.95 (2672.23)	0.804 (0.314)	0.804 (0.314)	
Exponentiated Weibull					
	γ	0.748 (1.668)	0.756 (0.720)	0.756 (0.720)	1587.30
	β	64.88 (133.31)	23.00 (26.53)	23.00 (26.53)	
	α	2.513 (6.700)	1.833 (2.308)	1.833 (2.308)	
Restricted beta-Weibull beta-Weibull with fixed $\alpha_2 = 1/\alpha_1$					
	γ	1.724 (1.903)	0.800 (0.576)	0.800 (0.576)	1587.30
	β	88.50 (53.00)	14.08 (15.27)	14.08 (15.27)	
	α	1.01 (120.20)	1.718 (2.231)	1.718 (2.231)	
Beta-Weibull					
	γ	4.320 (0.562)	0.725 (0.892)	0.725 (0.892)	1589.29
	β	88.54 (3.655)	41.68 (1.50)	41.68 (1.50)	
	α_1	0.400 (0.268)	1.923 (10.57)	1.923 (10.57)	
	α_2	0.915 (3.377)	1.763 (6.420)	1.763 (6.420)	

Table II
Maximum likelihood estimates of Weibull parameters from breast cancer data (all disease-free events)

Distribution	tamoxifen group			CMFT group		
	par	mle(se)	AIC	mle(se)	AIC	
2-parameter Weibull	γ	1.067 (0.070)	1851.92	1.272 (0.098)	1492.35	
	β	28.39 (2.82)		30.40 (3.35)		
3-parameter Weibull	γ	1.831 (0.325)	1842.22	1.299 (0.205)	1494.13	
	β	8.009 (2.566)		28.71 (11.18)		
	λ	0.126 (0.048)		2.70 (17.92)		
4-parameter Weibull	γ	1.550 (0.390)	1828.80	20.23 (20.89)	1496.13	
	β	6.046 (1.921)		1.440 (0.433)		
	λ	0.007 (0.003)		0.014 (0.077)		
	τ	0.114 (0.040)		0.830 (0.321)		
Exponentiated Weibull (beta-Weibull with fixed $a_2 = 1$)						
	γ	0.145 (0.018)	1838.5	0.875 (0.839)	1494.24	
	β	0.001 (0.002)		23.53 (21.92)		
	α	57.67 (24.87)		1.573 (1.930)		
Restricted beta-Weibull (beta-Weibull with fixed $a_2 = 1/\alpha_1$)						
	γ	0.438 (0.035)	1838.82	0.908 (0.734)	1494.24	
	β	0.315 (0.691)		16.48 (15.87)		
	α	6.803 (4.152)		1.508 (1.684)		
Beta-Weibull						
	γ	0.237 (0.104)	1836.65	0.871 (0.823)	1496.24	
	β	0.002 (0.033)		23.52 (30.49)		
	α_1	126.15 (62.70)		1.583 (14.55)		
	a_2	0.125 (0.232)		1.005 (8.64)		

Table III

Goodness-of-fit for breast cancer data for competing risks

tamoxifen group															
local or regional										other events					
Years	at risk	obs	2^a	3^b	4^c	$\exp(e_y)$	rest ^e	b-w ^f	obs	2	3	4	exp (e_y)	rest	b-w
(N_y)	(s_y)								(s_y)						
[0–1]	770	4	6.6	4.1	4.3	5.7	4.6	4.6	14	15.1	11.5	9.9	11.5	10.2	9.8
(1–2]	752	8	6.3	10.1	10.0	7.1	9.7	9.5	22	16.8	20.5	25.7	20.2	23.3	24.7
(2–3]	722	12	6.0	9.4	8.4	6.8	8.4	8.1	19	17.0	22.0	21.5	20.6	23.3	24.1
(3–4]	689	8	5.7	7.4	6.7	6.3	6.8	6.7	17	16.8	20.7	18.8	19.4	20.8	20.9
(4–5]	663	1	5.5	5.9	5.6	5.8	5.7	5.7	14	16.5	18.7	16.9	18.0	18.3	18.1
(5–6]	641	7	5.3	4.8	4.8	5.4	4.9	4.9	16	16.3	16.7	15.5	16.7	16.3	15.8
(6–7]	615	2	5.0	3.9	4.2	5.0	4.2	4.3	12	15.9	14.7	14.3	15.3	14.5	13.8
(7–8]	597	1	4.9	3.3	3.8	4.6	3.7	3.9	16	15.7	13.1	13.3	14.2	13.1	12.4
(8–9]	571	4	4.6	2.8	3.3	4.3	3.3	3.5	15	15.2	11.5	12.4	13.0	11.7	11.0
(9–10]	539	6	4.4	2.4	3.0	3.9	2.9	3.1	11	14.5	10.1	11.3	11.8	10.4	9.7
Total	6549	53							156						
$\sum R_y^2$															
p -value															
CMFT group										other events					
Years	at risk	obs	2	3	4	$\exp(e_y)$	rest	b-w	obs	2	3	4	exp (e_y)	rest	b-w
(N_y)	(s_y)								(s_y)						
[0–1]	766	0	0.3	0.4	0.4	0.3	0.3	0.3	8	9.8	9.4	9.3	9.2	9.2	9.2
(1–2]	755	2	0.7	0.7	0.7	0.7	0.7	0.7	14	13.1	13.3	13.7	13.5	13.5	13.5
(2–3]	738	1	1.0	1.0	1.0	1.0	1.0	1.0	15	14.5	14.9	15.0	15.1	15.1	15.1

tamoxifen group														
local or regional										other events				
Years	at risk	obs	exp(e_y)			obs	exp(e_y)			rest	exp	4	rest	b-w
	(N_y)	(s_y)	2a	3b	4c	exp ^d	rest ^e	b-w ^f	(s_y)	2	3	4	exp	b-w
(3-4]	718	0	1.3	1.2	1.2	1.3	1.3	1.3	19	15.3	15.7	15.7	15.8	15.9
(4-5]	697	4	1.5	1.4	1.4	1.5	1.5	1.5	17	15.8	16.1	16.0	16.2	16.2
(5-6]	673	0	1.7	1.6	1.6	1.7	1.7	1.7	15	16.0	16.2	16.0	16.1	16.2
(6-7]	644	0	1.8	1.7	1.8	1.8	1.8	1.8	10	15.9	15.9	15.7	15.8	15.8
(7-8]	629	1	2.0	2.0	2.0	2.0	2.0	2.0	14	16.1	15.9	15.8	15.8	15.8
(8-9]	604	4	2.1	2.2	2.1	2.0	2.1	2.1	14	15.9	15.5	15.5	15.4	15.4
(9-10]	564	2	2.1	2.4	2.4	2.0	2.1	2.1	18	15.3	14.6	14.7	14.5	14.5
Total	6788	14	144											
$\sum R_y^2$			16.57	16.36	16.44	16.62	16.58	16.56	5.04	4.80	4.53	4.60	4.62	4.59
p -value			0.035	0.022	0.012	0.020	0.020	0.011	0.753	0.684	0.605	0.708	0.706	0.597
df	8	7	6	7	6	7	7	6	8	7	6	7	7	6
^a Two-parameter (classic) Weibull ^b Three-parameter Weibull ^c Four-parameter Weibull ^d Exponentiated Weibull ^e Restricted beta-Weibull ^f beta-Weibull														

Table IV

Goodness-of-fit for breast cancer data, for any event

tamoxifen group										
all events										
Years	at risk	obs	exp (e_y)							
	(N_y)	(s_y)	2^a	3^b	4^c	$\exp^d$	rest ^e	b-w ^f		
[0-1]	770	18	21.7	15.7	14.7	16.5	16.9	14.3		
(1-2]	752	30	23.2	30.7	35.4	28.2	29.1	34.0		
(2-3]	722	31	23.0	32.1	29.4	28.2	29.2	32.5		
(3-4]	689	25	22.5	28.8	25.3	26.1	26.9	27.9		
(4-5]	663	15	22.0	24.9	22.5	24.0	24.5	24.0		
(5-6]	641	23	21.6	21.5	20.4	22.0	22.4	20.8		
(6-7]	615	14	20.9	18.4	18.6	20.1	20.3	18.1		
(7-8]	597	17	20.5	16.0	17.3	18.5	18.7	16.1		
(8-9]	571	19	19.8	13.9	15.9	16.9	17.0	14.2		
(9-10]	539	17	18.8	12.0	14.5	15.2	15.3	12.4		
Total	6549	209								
$\sum R_y^2$										
			11.69	10.67	7.24	7.33	7.69	10.17		
p -value			0.166	0.154	0.299	0.396	0.361	0.118		
CMFT group										
all events										
Years	at risk	obs	exp (e_y)							
	(N_y)	(s_y)	2	3	4	exp	rest	b-w		
[0-1]	766	8	10.0	9.8	9.5	9.5	9.5	9.5		
(1-2]	755	16	13.9	14.0	14.4	14.2	14.2	14.2		
(2-3]	738	16	15.6	15.8	16.2	16.1	16.1	16.1		

tamoxifen group										
all events										
Years	at risk	obs								
	(N_y)	(s_y)	2 a	3 b	4 c	exp d	rest e	b-w f		
(3-4]	718	19	16.7	16.9	17.1	17.2	17.1	17.1	17.2	17.2
(4-5]	697	21	17.3	17.5	17.6	17.7	17.7	17.7	17.7	17.7
(5-6]	673	15	17.7	17.8	17.7	17.9	17.9	17.9	17.9	17.9
(6-7]	644	10	17.7	17.7	17.6	17.7	17.7	17.7	17.7	17.7
(7-8]	629	15	18.0	17.9	17.7	17.7	17.7	17.7	17.7	17.7
(8-9]	604	18	17.9	17.6	17.4	17.4	17.4	17.4	17.4	17.4
(9-10]	564	20	17.2	16.8	16.6	16.5	16.6	16.6	16.5	16.5
Total	6788	158								
$\sum R_y^2$										
			7.31	7.22	6.81	7.00	7.01	7.01	7.00	7.00
p -value										
			0.503	0.406	0.338	0.429	0.427	0.427	0.321	0.321
df										
			8	7	6	7	7	7	6	6
a Two-parameter (classic) Weibull										
b Three-parameter Weibull										
c Four-parameter Weibull										
d Exponentiated Weibull										
e Restricted beta-Weibull										
f beta-Weibull										

Treatment group comparisons of estimates of the cumulative incidence function under competing risks and survival probabilities for all disease-free events based on two Weibull distributions. The p-values are calculated using the Wald test.

Table V

Years	Treatment group	Cumulative incidence function				Survival function	
		Local or regional		Other events		All events	
		$\hat{F}_d$	$\widehat{se}(F_{ci})$	p-value	$\hat{F}_{ci}$	$\widehat{se}(F_{ci})$	p-value
Four-parameter Weibull							
2	tamoxifen	0.018	0.004	0.0001	0.046	0.006	0.059
	CMFT	0.001	0.001		0.030	0.005	0.969
5	tamoxifen	0.045	0.007	< 0.0001	0.119	0.011	0.060
	CMFT	0.006	0.003		0.091	0.010	0.903
10	tamoxifen	0.070	0.010	< 0.0001	0.206	0.017	0.736
	CMFT	0.019	0.006		0.196	0.025	0.785
Three-parameter restricted beta-Weibull							
2	tamoxifen	0.018	0.004	< 0.0001	0.043	0.006	0.093
	CMFT	0.001	0.001		0.029	0.005	0.969
5	tamoxifen	0.045	0.006	< 0.0001	0.123	0.010	0.027
	CMFT	0.006	0.002		0.091	0.010	0.903
10	tamoxifen	0.070	0.008	< 0.0001	0.208	0.015	0.562
	CMFT	0.019	0.005		0.196	0.015	0.785