RESEARCH COMMUNICATION

Applying Conventional and Saturated Generalized Gamma Distributions in Parametric Survival Analysis of Breast Cancer

Parvin Yavari[1]*, Alireza Abadi[1,], Farzaneh Amanpour[2], Chris Bajdik[3]

Abstract

   <u>Background</u>: The generalized gamma distribution statistics constitute an extensive family that contains nearly all of the most commonly used distributions including the exponential, Weibull and log normal. A saturated version of the model allows covariates having effects through all the parameters of survival time distribution. Accelerated failure-time models assume that only one parameter of the distribution depends on the covariates. <u>Methods</u>: We fitted both the conventional GG model and the saturated form for each of its members including the Weibull and lognormal distribution; and compared them using likelihood ratios. To compare the selected parameter distribution with log logistic distribution which is a famous distribution in survival analysis that is not included in generalized gamma family, we used the Akaike information criterion (AIC; $r=l(b)-2p$). All models were fitted using data for 369 women age 50 years or more, diagnosed with stage IV breast cancer in BC during 1990-1999 and followed to 2010. <u>Results</u>: In both conventional and saturated parametric models, the lognormal was the best candidate among the GG family members; also, the lognormal fitted better than log-logistic distribution. By the conventional GG model, the variables "surgery", "radiotherapy", "hormone therapy", "erposneg" and interaction between "hormone therapy" and "erposneg" are significant. In the AFT model, we estimated the relative time for these variables. By the saturated GG model, similar significant variables are selected. Estimating the relative times in different percentiles of extended model illustrate the pattern in which the relative survival time change during the time. <u>Conclusions</u>: The advantage of using the generalized gamma distribution is that it facilitates estimating a model with improved fit over the standard Weibull or log-normal distributions. Alternatively, the generalized F family of distributions might be considered, of which the generalized gamma distribution is a member and also includes the commonly used log-logistic distribution.

Keywords: Survival analysis - parametric regression - generalized gamma distribution - breast cancer

Introduction

   When a parametric survival analysis is considered, we assume that the survival time (or a function of it) follows a given theoretical distribution (or model) and has an explicit relationship with the covariates (Lee and Wang, 2003). Survival time models that can be linearized by taking logs are called 'accelerated failure time' (AFT) models. In AFT models the effect of covariates is multiplicative over time (Hosmer and Lemeshow, 1989). The acceleration factor is a ratio of survival times corresponding to any fixed value of S(t).The acceleration factor describes the "stretching out" or contraction of survival functions when comparing one group to another (Kleinbaum and Klein, 2005). The underlying AFT assumption, for comparing two levels of covariates, is that the ratio of times to any fixed value of S(t) = q is constant for any probability q (Kleinbaum and Klein, 2005).

   The application of a constant relative time implies the assumption that the treatment only has an effect on one characteristic of the survival distribution, while commonly used survival distributions, like the lognormal and Weibull distribution, have both a shape and a scale parameter. In this paper, we fitted the generalized gamma distribution, the extensive family that contains nearly all of the most commonly used distribution including the exponential, Weibull and lognormal; and we extend the analysis to covariates having effects through all the parameters of survival time distribution. We compared these models with the conventional AFT form of parametric models which assumes that only one parameter of distribution depends on the covariates. All models were fitted to population-based data from British Columbia, Canada (BC).

   The relationship of the generalized gamma distribution to the exponential, Weibull and lognormal allows us to evaluate the appropriateness of these distributions relative to each other and to a more general distribution. Fitting the extended distribution to the conventional AFT form shows how a variable can affect the survival time of breast cancer patients.

[1]Department of Health and Community Medicine, School of Medicine, [2]Department of Biostatistics, Paramedical sciences faculty, Shahid Beheshti University of Medical Sciences, Tehran, Iran, [3]Cancer Control Research Program, BC Cancer Agency, Vancouver, British Columbia, Canada *For correspondence: P.yavari-grc@sbmu.ac.ir

Materials and Methods

Study Design

All models were fitted using data for 369 women age 50 years or more, diagnosed with stage IV breast cancer in BC during 1990-1999 and followed to 2010. The data for each patient included information about their treatment, survival and tumor hormone receptors.

Treatment(s) assigned to patients included hormone therapy, chemotherapy, surgery, and radiotherapy. For each of the four treatment types, a corresponding variable had the value one if the subject received the treatment and zero otherwise.

Survival time was defined as the period between the diagnosis of disease and death or end of patient follow up. A censorship variable had the value one if the patient died of breast cancer and zero otherwise.

The presence of hormone receptors in a tumor has been proven to affect a patient's survival time, and was included in model as a binary variable, erposneg, with the value one if the tumor was hormone receptor positive and zero if it was hormone receptor negative.

Statistical Analysis

To determine the appropriate parametric survival model, we began the analysis by fitting a model based on the generalized gamma (GG) distribution. This is a three-parameter family with location (β), scale ($\sigma > 0$) and shape (λ) parameters, which contains nearly all of the most commonly used distributions including the exponential ($\lambda = \sigma = 1$), Weibull ($\lambda = 1$) and log normal ($\lambda = 0$).

The classical AFT model only allows that covariate effects are modeled through the parameter β. If we extend the analysis to covariates having effects through the parameters σ and/or λ, these are no longer AFT models (Cox et al., 2007). We fitted both the conventional GG model and the saturated form for each of its members including the Weibull and lognormal distribution; and compared them using likelihood ratios. To compare the selected parameter distribution with log-logistic distribution that is not included in generalized gamma family, we used the Akaike information criterion (AIC; r=l(b)-2p). The candidate distribution with the largest r value is the distribution that fits the data the best (Lee and Wang, 2003).

In conventional models, we can estimate the relative time by exponentiating the coefficient of a variable, since in AFT models the ratio of survival time is constant during the time. However for saturated models in which all of the parameters depend on covariates, we need to calculate the relative time RT(p) for different times. The relative times are defined for 0< p<1 as the ratio of the corresponding quantile functions, RT(p) =t_1 (p)/t_0 (p). RT(p) is the time required for p individuals in the exposed or treated population to experience the event of interest. So RT(p)-fold the time is required for the same proportion of events to occur in the reference population. The links between the quantiles of the gamma and those of the GG facilitates the use of softwares such as SAS that provides the quantile function for the gamma distribution to obtain the percentiles of the GG (Cox et al., 2007).

Table 1. Comparison between Parametric Regression Models in Saturated and Conventional Forms

Distribution	Conventional		Saturated	
	X_L	AIC	X_L	AIC
GG	Ref	-552.28	Ref	-556.60
Weibull	53.34	-576.95	59.20	-576.22
Lognormal	1.24	-550.90	10.06	-541.65
Log-logistic	---	-551.67	---	-550.93

*X_L, Likelihood ratio statistic relative to generalized gamma; Ref, Reference distribution

Table 2. Relative Time Estimated in Conventional and Saturated Parametric Models

Variable	Conventional		Saturated		
	RT	95%CI	RT(0.25)	RT(0.5)	RT(0.75)
Surgery	1.49	(1.16,1.99)	1.3	1.55	1.86
Radiotherapy	1.65	(1.25,2.22)	2.09	1.68	1.35
Erposneg	3.02	(2.68,4.42)	2.84	3	3.17
Hormone therapy if erposneg=1					
	2.7	(1.72,3.81)	2.4	2.53	2.67
Hormone therapy if erposneg=0					
	1.33	(0.39,1.97)	1.14	1.19	1.24

$$S_{GG(\beta,\sigma,\lambda)} = S_{G(\beta\lambda/\sigma,|\lambda|)}(t^{\lambda/\sigma})$$
$$= 1 - cdf (\text{'gamma'}, t^{\lambda/\sigma}, \lambda^{-2}, \lambda^2 e^{\beta\lambda/\sigma}) \quad \text{if } \lambda > 0$$
$$= cdf (\text{'gamma'}, t^{\lambda/\sigma}, \lambda^{-2}, \lambda^2 e^{\beta\lambda/\sigma}) \quad \text{if } \lambda < 0$$

Results

The results of comparison between parametric models have been summarized in Table 1. In both conventional and saturated parametric models, the lognormal was the best candidate among the GG family members according to the likelihood ratio statistic; the lognormal fitted better than log-logistic according to the AIC statistic.

The relative times estimated by fitting conventional and saturated parametric model have been summarized in Table 2.

In the conventional model the variables surgery (p=0.002), radiotherapy (p=0.001), hormone therapy (p<0.001), and erposneg (p<0.001) were meaningful and the interaction between hormone therapy and erposneg (p=0.027) was significant. In the AFT model, we estimated the relative time for variables surgery, radiotherapy and erposneg by exponentiating the coefficient of the each variable, and because of interaction between hormone therapy and erposneg, the relative time of hormone therapy was estimated by exponentiating its coefficient in each category of erposneg separately. According to the conventional model, patients who receive surgery have longer survival times rather than those who don't; similarly, patients who receive radiotherapy have longer survival times than those who don't; in addition, hormone therapy increased the survival time of patients with both positive and negative hormone receptor status. In general, patients with positive hormone receptor status had higher survival times than patients with negative hormone receptor status; likewise, patients with positive hormone receptor status who received hormone therapy had higher survival time than patients with negative hormone receptor

status who received hormone therapy.

In saturated model the variables surgery (p=0001), radiotherapy (p=0.004), hormone therapy (p<0.001), and erposneg (p<0.001) were meaningful in the shape parameter, and the interaction between hormone therapy and erposneg (p=0.021) was significant. In the scale parameter the variables surgery (p=0.021) and radiotherapy (p=0.002) were meaningful. Relative time was calculated in the 25th, 50th and 75th percentiles of the lognormal distribution for variables surgery, radiotherapy, and erposneg; the interaction between hormone therapy and erposneg, the relative time for hormone therapy was estimated in each category of erposneg separately. Estimating the relative times in different percentiles of extended model illustrate the pattern in which the relative survival time change during the time. The ratio of survival time for patients who received surgery increased over time; however, for patients who received radiotherapy the relative survival time decreased during the time. Patients who received hormone therapy, in both categories of positive and negative hormone receptor status, show an increasing pattern for relative survival time. In general, the ratio of survival time for patients with positive hormone receptor status and those with negative hormone receptor status has increased during the time.

Discussion

Most of the researches on parametric survival assume that the accelerated failure time assumption is true; that is, only one parameter of the distribution is related to covariates. However most parametric distributions have more than one parameter that may be depend on covariates and consequently cause the relative survival time not to be constant for every fixed survival function (Kleinbaum and Klein, 2005).

Our aim was to fit both conventional and extended forms of parametric models and compare their effects. Priority of the extended model to the AFT conventional model is that it can illustrate the pattern in which the intervention affect the survival time, in that it shows the change of relative survival time during the time. To our knowledge, no other study of this type in cancer survival analysis has considered the ratio of survival time using an extended form of the GG distribution. Cox (2007) illustrated in his paper how the relative time is not constant when all the parameters of distribution depend on covariates and as a result the saturated model works better than conventional AFT form (Cox et al., 2007). Meta-analysis of treatment effects based on the shape and scale parameters of parametric survival curves showed that saturated model offers a more flexible alternative to conventional model; in which only one parameter depends on covariates, in proportional hazards parametric models (Ouwens et al., 2011).

In our analysis, the lognormal distribution was the best fit for data in both conventional and saturated models. In saturated models, covariates were meaningful in both shape and scale parameters, and calculating the relative times for different percentiles of lognormal distribution showed the patterns by which relative times would change.

The lognormal distribution has a long history of usage in cancer survival analysis (Boag, 1949; Wang et al., 2010), and in many settings including breast cancer analysis, the lognormal model has been shown to be an appropriate survival model (Gamel et al., 1994; Royston, 2001; Chapman et al., 2006). In conventional models, the AIC was almost the same for lognormal and log-logistic distributions. Lognormal and log-logistic distributions have very similar shapes and yield similar results (Kleinbaum and Klein, 2005)

In conclusion, the generalized gamma distribution includes most of the commonly used parametric survival distributions: Weibull, exponential, lognormal, and Gamma. The advantage of using the generalized gamma distribution is that it facilitates estimating a model with improved model fit over the standard Weibull or lognormal distribution (Cox et al., 2007). Alternatively, the generalized F family of distributions might be considered, of which the generalized gamma distribution is a member and also includes the commonly used log-logistic distribution (Ciampi et al., 1986; Cox et al., 2007).

Acknowledgements

References

Boag JW (1949). Maximum likelihood estimates of the proportion of patients cured by cancer therapy. *J R Stat Soc B*, **11**, 15-44.

Chapman JA, Lickley HL, Trudeau ME, et al (2006). Ascertaining prognosis for breast cancer in node-negative patients with innovative survival analysis. *Breast J*, **12**, 37-40.

Cox C (2008). The generalized F distribution: An umbrella for parametric survival analysis. *Stat Med*, **27**, 4301-12.

Cox C, Chu H, Schneider MF, Munoz A (2007). parametric survival analysis and taxonomy of hazard functions for the generalized gamma distribution. *Stat Med*, **26**, 4352-74.

Gamel JW, Vogel RL, Valagussa P, et al (1994). Parametric survival analysis of adjuvant therapy for stage II breast cancer. *Cancer*, **74**, 2483-90.

Hosmer D, Lemeshow S (1989). Applied Survival Analysis. New York, Wiley.

Kalbfleisch J, Prentice R (1980). The Statistical Analysis of Failure Time Data. New York, Wiley.

Kleinbaum D, Klein M (2005). Survival Analysis. New York, Springer.

Lee E, Wang J (2003). Statistical Methods for Survival Data Analysis. New Jersey, Wiley.

Ouwens M, Philips Z, Jansen J (2011). Network meta-analysis of parametric survival curves. wileyonlinelibrary.

Royston P (2001). The lognormal distribution as a model for survival time in cancer, with an emphasis on prognostic factors. *Statistica Neerlandica*, **55**, 89-104.

Wang S, Kalpathy-Cramer J, Kim J, Fuller D, Thomas C (2010). Parametric survival models for predicting the benefit of adjuvant chemoradiotherapy in gallbladder cancer. *AMIA Annu Symp Proc*, **13**, 847-51.