scientific reports

Check for updates

OPEN   # The comparison of censored quantile regression methods in prognosis factors of breast cancer survival

Akram Yazdani[1,2], Mehdi Yaseri[2], Shahpar Haghighat[3], Ahmad Kaviani[4] & Hojjat Zeraati[2]✉

The Cox proportional hazards model is a widely used statistical method for the censored data that model the hazard rate rather than survival time. To overcome complexity of interpreting hazard ratio, quantile regression was introduced for censored data with more straightforward interpretation. Different methods for analyzing censored data using quantile regression model, have been introduced. The quantile regression approach models the quantile function of failure time and investigates the covariate effects in different quantiles. In this model, the covariate effects can be changed for patients with different risk and is a flexible model for controlling the heterogeneity of covariate effects. We illustrated and compared five methods in quantile regression for right censored data included Portnoy, Wang and Wang, Bottai and Zhang, Yang and De Backer methods. The comparison was made through the use of these methods in modeling the survival time of breast cancer. According to the results of quantile regression models, tumor grade and stage of the disease were identified as significant factors affecting 20th percentile of survival time. In Bottai and Zhang method, 20th percentile of survival time for a case with higher unit of stage decreased about 14 months and 20th percentile of survival time for a case with higher grade decreased about 13 months. The quantile regression models acted the same to determine prognostic factors of breast cancer survival in most of the time. The estimated coefficients of five methods were close to each other for quantiles lower than 0.1 and they were different from quantiles upper than 0.1.

Abbreviations
AFT   Accelerated failure time
BCS   Breast conserving surgery
MRM   Modified radical mastectomy
QR   Quantile regression

In many medical studies, the outcome of interest is the time to event. For instance, in cancer, the event of interest is death or the relapse of illness, and in transplantation, the rejection of transplanted organ can be considered as the event. In case of the uncertainty of time for study inclusion and the incidence of an event in some studied units, time is regarded as the censored data[1]. The Cox proportional hazards (Cox) model is a widely used statistical method for the censored data. However, this model is limited by the assumption of a constant hazard ratio (HR) over time (i.e., proportionality), and models the hazard rate rather than the survival time directly[2]. Also, the complexity of the HR estimate interpretation was recognized as a problem in the Cox models. To overcome this limitations, other methods such as accelerated failure time (AFT) models and censored quantile regression (CQR) models were introduced for censored data with a more straightforward interpretation[3]. The AFT model is a model that assumes that a treatment or an exposure either extends or reduces the time to the development of event. Although the AFT model allows a direct interpretation of covariate effects on event time, it requires a strong assumption of homogeneous treatment effect[4]. The AFT model can only capture location shifts effects, and hence may fail to capture heterogeneity of covariate effects. The CQR model does not require the assumption

[1]Department of Biostatistics and Epidemiology, Faculty of Health, Kashan University of Medical Sciences, Kashan, Iran. [2]Department of Epidemiology and Biostatistics, School of Public Health, Tehran University of Medical Sciences, Tehran, Iran. [3]Breast Cancer Research Center, Motamed Cancer Institute, ACECR, Tehran, Iran. [4]Department of Surgery, Tehran University of Medical Sciences, Tehran, Iran. ✉email: zeraatih@tums.ac.ir

models	Assumptions	Advantages	Disadvantages
Cox method	Proportional hazard	No need to consider a specific probability distribution for the survival time; Can used in many types of survival model	The effect of the included covariates is multiplicative The complexity of the HR estimate interpretation
AFT method	The effect of a covariate is to accelerate or decelerate the life course of a disease by some constant Needs homogeneous covariates effect	direct interpretation of covariate effects on event time Can used in many types of survival model	Error term follow a specific probability distribution Failing to capture heterogeneity of covariate effects
Portnoy method	The model at lower quantiles are all linear (global-linearity)	The effect of covariates is not restricted to be constant No distributional assumptions about the regression error term	The 'global' linearity assumption
Bottai and Zhang method	The residuals follow a asymmetric Laplace distribution require -linearity assumption	The effect of covariates is not restricted to be constant Correct coverage and shorter computation time	Error term follow a Laplace distribution
Wang and Wang method	Require a locally linear quantile regression	Not require global-linearity assumption	Requires estimating the true distribution of the outcome variable
Yang method	Operates under the assumption that all the quantile functions are identifiable	Can handle different forms of censoring the estimator can achieve significant efficiency gains over the existing methods	It runs a risk of finding estimates even for non-identifiable quantile functions
De Backer method	Require a locally linear quantile regression	Consistency and asymptotic normality of estimator	Restrict to the estimation of the classical linear regression model

Table 1. Assumptions and features Cox, AFT, Portnoy, Wang and Wang, Bottai and Zhang, Yang and De Backer methods.

of a homogeneous covariate effect, while the effect estimates have the same straightforward interpretation as those in AFT model[3]. A censored quantile regression model is a model of the quantile function of failure time and investigates the covariate effects in different quantiles. In this model, the covariate effects can be changed for patients with different risks and is a flexible model for controlling the heterogeneity of covariate effects. Further, the failure time can be predicted by CQR model and the progression of a disease can be predicted[4].

For conditionally random right censoring, various approaches have been proposed that the most practical of them include: Portnoy[5], Peng and Huang[6], Wang and Wang[7], Bottai and Zhang[8], Yang et al.[9] and De Backer et al.[10] methods. Portnoy generalized the Kaplan–Meier method to estimate quantile of survival time with a recursively weighted estimation algorithm under the global linearity assumption of the conditional quantile functions[5]. Peng and Huang presented another method based on Nelson-Aalen estimator of cumulative hazard function requiring linearity assumption like Portnoy's method and present the result closed to Portnoy method[6]. To overcome the linearity assumption, Wang and Wang developed a method by non-parametrically estimating the conditional survival distribution via kernel smoothing[7]. In 2010, Laplace regression was introduced as a parametric method for modeling the conditional quantile of censored data by Bottai and Zhang[8]. They assumed error term follow asymmetric Laplace distribution and considered the Laplace regression model as a method for modeling the conditional quantiles of survival time. Nevertheless, this parametric assumption, shared by other methods in quantile regression (Liu and Botta[11], Farcomeni[12], Lee and Neocleous[13], Yuan and Yin[14]), has been shown assumption asymmetric Laplace distribution for error term, not to influence the results of the model under different data distributions. It was indicated in simulation studies that correct coverage and shorter computation time obtained with Laplace regression model compared with other alternative methods[8]. Yang et al. to estimate quantiles of survival time, employed a variation of the data augmentation algorithm[9]. Base on the general principle of data augmentation[15], in the algorithm, they first, impute censored values from the quantile functions then using the imputed values fit the quantile model. De Backer et al. investigate a new procedure for modeling right-censored data with a linear quantile regression[10]. They used "check" loss function that stems from the influential work of Koenker and Bassett[16], to circumvent the formulation of conditional quantiles. The assumptions and features of these models are summarized in Table 1.

The most of existing methods of estimation for censored quantiles are limited to right censored data. This paper has concentrated on the outcome data that were right censored, as this was the main feature of our motivating dataset. Subsequent work has greatly expanded the applicability of these methods to competing risks[17–19], recurrent events[20,21], various censoring types[17,22] and other settings. For example, Yang et al. proposed a new method for different forms of censoring including doubly censored and interval censored data[9]. Narisetty (2018) introduced a new approach for the cure rate quantile regression model[23]. Chen developed quantile regression estimators and proposed a quantile regression method with time-varying covariates[24].

Recent years, use of quantile regression has increased in cancer research. Base on PubMed search, there are more than 200 publications on applications of quantile regression related to cancer research from 2015 to 2021. Breast cancer is the most common type of cancer after the lung cancer and the most common cause of death from cancer among women[25]. In 2000, there were 10 million new cases of breast cancer, i.e. 25% of all cancer cases around the world, and, it is expected to reach 15 million in 2025[26]. The five-year survival rate of breast cancer range from less than 40% to 80% in low-income to high-income countries around the world[27]. Knowledge of the survival-associated predictors in breast cancer has an important role in the process of treatment and patient care. Age at diagnosis, tumor size, grade, type of auxiliary treatment (radiotherapy, stage of disease, number of involved lymph nodes, chemotherapy, hormone therapy), type of surgery (Modified radical mastectomy (MRM)

and Breast conserving surgery (BCS), metastasis and recurrence have been identified as the most important risk factors of breast cancer survival[28,29].

The present study aimed to compare quantile regression methods included Portnoy, Wang and Wang, Bottai and Zhang, Yang and De Backer methods with Cox and AFT models. The comparison was made through the use of these methods in modeling the survival time of breast cancer using the data collected from Imam Khomeini hospital, Tehran, Iran.

Methods

Data description. The present study aimed to determine the factors affecting breast cancer survival using CQR models. The studied data included 800 females patients with breast cancer (based on breast cancer pathology diagnosis) referring to Imam Khomeini Hospital, Tehran, Iran during 1996–2005. The required data were extracted from patients' files. The latest condition of the patient was informed via phone contact.

This study was approved by the Ethics Committee of School of Public Health & Allied Medical Sciences-Tehran university of Medical Sciences (approval ID: IR.TUMS.SPH.REC.1397.212) and was carried out according to relevant guidelines and regulations. The informed consent was obtained from all participants.

The event of interest was death from breast cancer, and the survival time was defined as the duration (months) from diagnosis to death due to breast cancer. The prognostic factors included age at diagnosis (year), type of surgery (Modified radical mastectomy (MRM) and Breast conserving surgery (BCS), tumor grade (grade 1–3), and stage of disease (Stage 1–4) based on the seventh edition of the TNM classification. The p-values less than 0.05 were considered to be significant. The analysis was performed using STATAversion 12 and quantreg package in R.

Models. Quantile regression is a statistical technique intended to inference about conditional quantile functions. This method offer a mechanism for estimating models for the conditional median function, and the full range of other conditional quantile functions.

Let T denote the failure time and $\tilde{X} = (X_1, X_2, \ldots, X_p)^T$, denote a p × 1 vector of covariates. let $Q_Y(\theta|\tilde{X})$ = inf{t: $Pr(Y \le t|\tilde{X}) \ge \theta$}, denote the θth conditional quantiles of Y = log(T), (or another monotone transformation of T) given $\tilde{X}$, where $\theta \in (0, 1)$. For randomly censored data, let C denote time to censoring and let $\tilde{T} = \min(T, C)$ and $\delta = I(T \le C)$. The observed data consist of n i.i.d replicates of $(\tilde{T}, \delta, \mathbf{X})$, denoted by $(\tilde{T}_i, \delta_i, \mathbf{X}_i)$, i = 1,…, n. Define $\tilde{Y} = \log(\tilde{T})$, $\tilde{Y}_i = \log(\tilde{T}_i)$.

The linear QR model takes the form

$$Q_Y(\theta|\mathbf{X}) = \mathbf{X}^T\beta(\theta), \theta \in (\theta_L, \theta_U) \tag{1}$$

where $0 < \theta_L < \theta_U < 1$, and $\beta(\theta)$ is a vector of unknown regression coefficients that represents the change in the θth conditional quantile of Y given a one-unit change in the corresponding covariate[30]. When $\theta_L = \theta_U$, model (1) is referred to as a locally linear quantile regression model. When $\theta_L < \theta_U$, model (1) referred to as a globally linear quantile regression model.

The most applying methods during the recent years are Portnoy[5], Wang and Wang[7], Bottai and Zhang[8], Yang et al.[9] and De Backer et al.[10] methods. In the following, we present a brief overview of methodological framework for these models.

Portnoy method. Portnoy[5], using Efron's[31] interpretation of Kaplan–Meier as shifting mass of censored observations to the right, proposed an estimation algorithm under the standard random right censoring assumption to estimate $\beta(\theta)$.

The grid-based procedure presented in Neocleous et al.[32] defines a grid of θ -values, g_n, as, for $0 < \theta_1 < \theta_2 < \cdots < \theta_K = \theta_U$. Let $\|g_n\| = \max\{\theta_r - \theta_{r-1} : r = 1, 2, \ldots, R\}$. We will adopt the grid g_n throughout this section. $\beta(\theta_1)$ is estimated from applying uncensored QR when no censoring occurs below the θ_1th conditional quantile of T. Then, $\beta(\theta_{r+1})$ is a value of **b** (**b** is a vector of unknown regression coefficients) minimizing sequentially for r = 1, 2,…, R, by

$$\min\left\{\sum_{i \notin G} \rho_\theta\left(\tilde{Y}_i - \mathbf{X}_i^T\mathbf{b}\right) + \sum_{i \in G} [w_{r+1,i}\rho_\theta\left(\tilde{Y}_i - \mathbf{X}_i^T\mathbf{b}\right) + (1 - w_{r+1,i})\rho_\theta\left(Y^* - \mathbf{X}_i^T\mathbf{b}\right)]\right\} \tag{2}$$

where Y* is an extremely large value, $\rho_\theta(x) = x\{\theta - I(x < 0)\}$, and G is the set of indices of censored observations that have been previously crossed. The weight $w_{r+1,i} = (\theta_{r+1} - \theta_l)/(1 - \theta_l)$, approximates the conditional probability $Pr(C_i < T_i < \exp\{\mathbf{X}_i\beta(\theta_{r+1})\}|C_i < T_i, \mathbf{X}_i)$, based on the estimates for $\beta(\theta_1), \beta(\theta_2) \ldots, \beta(\theta_r)$[5].

Bottai and Zhang method. Bottai and Zhang[8], to estimator $\beta(\theta)$ considered a regression model where the error term is assumed to follow asymmetric Laplace distribution. They explored its use in the estimation of conditional quantiles of a continuous outcome variable given a set of covariates in the presence of random censoring.

They supposed that exists a fixed r-dimensional parameter vector $\beta(\theta)$ such that

$$T_i = \mathbf{X}^T\beta(\theta) + \varepsilon_i \tag{3}$$

where ε_i is an independent and identically distributed residual whose θth quantile equals zero $(P(\varepsilon_i \le 0|x_i) = \theta)$.

Let T_i conditionally on X_i, follows a form of asymmetric Laplace distribution with probability density function

$$f(T_i) = \exp\left\{\left(I(t_i \le \mu_i) - \theta\right)\frac{t_i - \mu_i}{\sigma(\theta)}\right\}\frac{\theta(1-\theta)}{\sigma(\theta)}$$

where $\sigma(\theta) \in (0, \infty), \mu_i = \mathbf{X}^T\beta(\theta)$.

In the presence of censored observations, the likelihood function is proportional to

$$l(\beta(\theta), \sigma(\theta)|T_i) = \prod_{i=1}^{N}\left((f(T_i))^{\delta_i}(1 - F(T_i))^{(1-\delta_i)}\right)$$

The maximum likelihood estimators for the parameters are defined as maximizes of $l(\beta(\theta), \sigma(\theta)|T_i)$. They used algorithm proposed by Nelder and Mead[33], to estimate parameters and inference on the parameters obtained by bootstrapping the point estimates for quantile of interest[8].

Wang and Wang method. A locally weighted method was proposed by Wang and Wang[7] to estimate a locally linear quantile regression model, which assumes that $\theta_L = \theta_U$, in model (1), i.e.

$$Q_Y(\theta|\mathbf{X}) = \mathbf{X}^T\beta(\theta) \qquad (4)$$

Wang and Wang[7], for random censoring, by twisting the idea of the self-consistent Kaplan–Meier estimator[31], proposed to modify the standard quantile loss function. The fundamental idea of Wang and Wang[7] is to redistribute the probability mass $(\Pr(T_i > C_i|C_i, \mathbf{X}_i))$, of the censored cases to the right through a local weighting scheme[7]. An estimator of $\beta(\theta)$ can be obtained by minimizing the following objective function of β:

$$n^{-1}\sum_{i=1}^{n}[w_i(F_0)\rho_\theta\left(\tilde{Y}_i - \mathbf{X}_i^T\mathbf{b}\right) + \{1 - w_i(F_0)\}\rho_\theta\left(Y^* - \mathbf{X}_i^T\mathbf{b}\right)] \qquad (5)$$

where $F_0(t|\boldsymbol{x}) \equiv \Pr(T > t|\mathbf{X} = \boldsymbol{x})$ is known and

$$w_i(F_0) = \begin{cases} 1 & F_0(C_i|\mathbf{X}_i) > \theta \text{ or } \delta = 1 \\ \frac{\theta - F_0(C_i|\mathbf{X}_i)}{1 - F_0(C_i|\mathbf{X}_i)} & F_0(C_i|\mathbf{X}_i) < \theta \text{ and } \delta = 0 \end{cases}$$

Wang and Wang[7] proposed to minimize the objective function (5), when $F_0(t|\boldsymbol{x})$ is unknown, with $F_0()$ replaced by the Beran's local Kaplan–Meier estimator[34], $\widehat{F}(.)$,

$$\widehat{F}(t|\boldsymbol{x}) = 1 - \prod_{j=1}^{n}\left\{1 - \frac{B_{nj}(x)}{\sum_{k=1}^{n}I(\tilde{Y}_k > \tilde{Y}_j)B_{nk}(x)}\right\}^{N_j(t)}$$

where $N(t) = I(\tilde{Y} \le t, \delta = 1)$, and $B_{nk}(x)$ is a sequence of nonnegative weights adding up to 1, for example, Nadaraya Watson's type weight, $B_{nk}(x) = K(\frac{x-x_k}{h_n})/\sum_{i=1}^{n}K(\frac{x-x_i}{h_n})$, where $K()$ is a density kernel function and h_n is a positive bandwidth converging to 0 as $n \to \infty$[7].

De Backer method. De Backer et al.[10] proposed to estimate model (4) based on a minimum distance loss function, given by $\sum_{i=1}^{n}\{1 - \widehat{F}(X_i^\tau\beta(\theta)|X_i) - \theta\}^2$. They further suggested using a smooth double kernel version of $\widehat{F}(.|X_i)$. Let Y_i^u denote the i-th order statistic of the uncensored responses, $n^u = \sum_{i=1}^{n}\delta_i$, and let $H^*(t) = \int_{-\infty}^{t}K(u)du$, for some kernel density K. They propose to estimate $F(t|\boldsymbol{x})$ by $\widehat{F}^s(t|\boldsymbol{x})$, where.

$$\widehat{F}^s(t|\boldsymbol{x}) = \int H\left(\frac{\boldsymbol{t}-\boldsymbol{u}}{\boldsymbol{h_T}}\right)d\widehat{F}(u|\boldsymbol{x}) = \sum_{i=1}^{n^u}\left(\widehat{F}\left(Y_{(i)}^u|\boldsymbol{x}\right) - \widehat{F}\left(Y_{(i-1)}^u|\boldsymbol{x}\right)H\left(\frac{\boldsymbol{t}-Y_{(i)}^u}{\boldsymbol{h_T}}\right)\right)$$

Yang method. Yang et al.[9] proposed a new and unified approach, to estimate the quantile regression model (1) with $\theta_U = 1$. They used a variation of the data augmentation algorithm base on the general principle of data augmentation[15]. The algorithm starts with a set of initial values, $\beta^{(0)}(\theta_k), k = 1, \ldots, M_n$, obtained by parallel quantile regression estimators or existing quantile regression estimators. Draw $Y_i^{(u)}$, for $u = 1, \ldots, U$, from the quantile process approximated by $\mathbf{X}_i^T\widehat{\beta}^{(u-1)}(\theta_k)$, conditional on the set of possible values for Y_i. Then, obtain updated estimates $\widehat{\beta}^{(u)}(\theta_k)$, via a standard uncensored quantile regression, based on a pairwise bootstrapping sample of size n from $\{\mathbf{X}_i, Y_i^{(u)}\}_{i=1}^{n}$. The final estimates obtain from $\widehat{\beta}(\theta) = U^{-1}\sum_{u=1}^{U}\widehat{\beta}^{(u)}(\theta_k)$[9].

The proposed method adapts easily to different forms of censoring including doubly censored and interval censored data[9].

Ethics approval and consent to participate. This study was approved by the Ethics Committee of School of Public Health & Allied Medical Sciences-Tehran University of Medical Sciences. Approval ID: IR.TUMS.SPH.REC.1397.212.

Written informed consent for publication of their clinical details was obtained from the patient relative of the patient. A copy of the consent form is available for review by the Editor of this journal.

Variables	Cox model			AFT model		
	HR	p-value	95% Cl	Coef	p-value	95% Cl
Age	0.99	0.753	(0.98, 1.01)	0.002	0.749	(− 0.01, 0.01)
Surgical procedure						
MRM	1.00			1.00		
BCS	0.58	0.079	(0.32, 1.06)	0.40	0.104	(− 0.08, 0.88)
Stage	3.12	< 0.001	(2.55, 3.81)	− 0.94	< 0.001	(− 1.11, − 0.77)
Grade	1.71	< 0.001	(1.27, 2.30)	− 0.45	< 0.001	(− 0.69, − 0.21)

Table 2. Multivariate analysis of prognostic factors of breast cancer survival with Cox model and AFT model. *HR* Hazard Ratio, *Coef.* Estimated parameter, *CI* Confidence Interval, *AFT* Accelerated failure time.
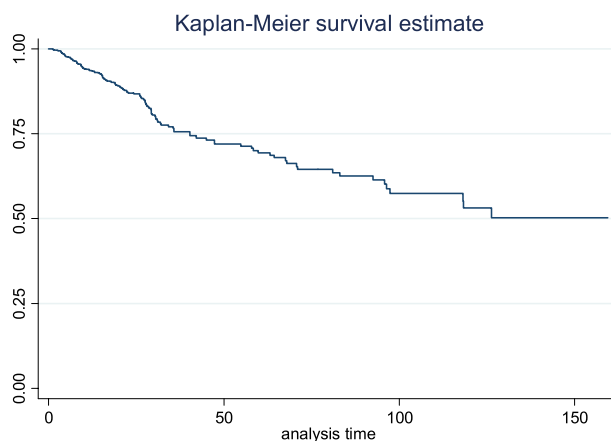
Figure 1. Kaplan–Meier plot of survival time of patient with breast cancer.

Results

The median follow-up time was 22.32 months with inter-quartile ranged from 13.10 to 30.51 months. During the follow-up, 143 (17.9%) patients died due to breast cancer and 657 (82.1%) survived or censored. Mean (SD) of age at diagnosis was 48.86 (13.63) years, and 106 (13.3%), 586 (73.3%), 65 (8.1%) and 43 (5.4%) patients were diagnosed in stages of disease 1 to 4, respectively. Further, 487 (60.9%) had tumor grade 2, 21 (27.3%) had tumor grade 3. 67 (84.3%) of patients had undergone MRM surgery. The proportional hazard assumption was confirmed at the significant level 0.05.

Table 2 displays the results of analysis of Cox and AFT models. Based on the results of Cox model, tumor grade and stage of disease had a significant effect on breast cancer survival. In other words, the hazard of death was 3.12 times for a higher stage of the disease, while the hazard ratio of death equaled to 1.71 for grade of tumor. In the AFT model with Weibull distribution, a shape parameter equaled to 1.24 that was significantly different from 1. Based on AFT model, the stage of disease and grade of tumor were the factors affecting survival time. Thus, increasing the stage of disease and the grade of tumor, decreased the median (or other quantiles) of survival time by a factor of 0.38 (exp(−0.94)) and a factor of 0.63 (exp(−0.45)) respectively.

As shown in the Kaplan–Meier plot in Fig. 1, at the end of the follow-up the minimum percentile of survival was 51%. Thus, we considered the10th, 20th and 40th percentiles of survival time in CQR model and considered the bandwidth 0.05 that was used in Wang and Wang and De Backer methods. According to the results of all quantile regression models in 20th percentile, tumor grade and stage of the disease were identified as significant factors affecting the survival time. However, the effectiveness of those factors varies in each model. In this regard, in Bottai and Zhang method 20th percentile of survival time for a case with higher unit of stage decreased 14.28 months and for a case with higher grade decreased12.53 months. In Yang method it was 16.64 and 14.53 months less and in Portnoy method 18.06 and 18.31 months, respectively. The result for 10th, 20th and 40th percentiles of survival time were shown in Table 3.

The CQR coefficients estimated and the 95% confidence intervals (CI) with Portnoy, Bottai and Zhang, Yang, Wang and Wang and De Backer methods and conditional quantile effects estimated by Cox model for θ ∈ (0.01, 0.10, …, 0.40) were displayed in Figs. 2, 3,4,5 and 6. Figure 7 shows the estimated coefficients of five methods. In the Cox model, the estimated quantile measure for each covariate was computed using Eq. (9) of Portnoy[5]. The effects of the Cox model were almost the same in different quantiles while they changed in quantile regression models as the quantiles vary.

5

	Coef. (95% Cl)		
Quantiles	0.10	0.20	0.40
Bottai and Zhang method			
Intercept	48.96 (22.65, 75.26)*	79.97 (44.82, 115.12)*	133.74 (87.57, 179.92)*
Age	0.02 (− 0.23, 0.26)	0.03 (− 0.30, 0.36)	0.09 (− 0.36, 0.53)
Surgical procedure (BCS)	11.28 (− 2.97, 25.54)	14.34 (− 1.07, 29.75)	24.07 (3.03, 45.14)*
Stage	− 10.59 (− 14.23, − 6.95)*	− 14.28 (− 19.01, − 9.56)*	− 28.87 (− 35.001, − 22.74)*
Grade	− 6.85 (− 13.14, − 0.57)*	− 12.53 (− 20.88, − 4.18)*	− 12.39 (− 23.60, − 1.19)*
Portnoy method			
Intercept	56.94 (37.86,76.65)*	102.05 (81.86,145.50)*	199.77 (87.30,304.03)*
Age	0.08 (− 0.23,0.37)	0.19 (− 0.99,0.62)	0.06 (− 0.60,0.45)
Surgical procedure (BCS)	14.34 (5.49,27.91)*	10.36 (1.35,22.13)*	22.31 (− 26.34,116.88)
Stage	− 12.61 (− 16.96, − 8.56)*	− 18.06 (− 24.26, − 6.85)*	− 43.58 (− 49.06, − 39.17)*
Grade	− 10.24 (− 13.48, − 3.36)*	− 18.31(− 25.81,13.60)*	− 13.92 (− 32.98, 20.41)
Wang and Wang method			
Intercept	17.57 (0.51, 45.50)*	42.05 (5.51, 69.10)*	111.39 (34.85, 160.72)*
Age	0.02 (− 0.19, 0.19)	− 0.02 (− 0.22, 0.25)	− 0.08 (− 0.48, 0.56)
Surgical procedure (BCS)	11.95 (3.42, 31.14)*	17.68 (− 1.85, 35.57)	8.96 (− 8.36, 60.73)
Stage	− 5.26 (− 8.41, − 2.69)*	− 8.26 (− 11.37, − 5.19)*	− 18.59 (− 24.51, − 10.65)*
Grade	− 2.33 (− 8.11, 0.32)	− 6.51 (− 10.14, − 0.45)*	− 11.57 (− 25.53, − 3.30)*
De Backer method			
Intercept	52.79 (25.79, 82.24)*	73.91 (40.12, 126.08)*	123.42 (74.45, 171.28)*
Age	0.02 (− 0.31, 0.24)	− 0.07 (− 0.36, 0.41)	− 0.03 (− 0.46, 0.58)
Surgical procedure (BCS)	3.62 (− 5.37, 25.95)	9.24 (− 5.37, 27.11)	10.98 (− 15.25, 41.63)
Stage	− 10.05 (− 13.62, − 5.88)*	− 12.56 (− 20.15, − 9.09)*	− 26.81 (− 35.98, − 15.86)*
Grade	− 5.91 (− 14.88, − 0.81)*	− 8.76 (− 19.49, − 3.01)*	− 5.30 (− 21.46, 2.24)
Yang method			
Intercept	53.11 (23.09, 93.38)*	87.20 (44.17, 133.21)*	198.89 (44.49, 227.96)*
Age	0.06 (− 0.26, 0.32)	0.12 (− 0.23, 0.51)	0.03 (− 0.29, 0.49)
Surgical procedure (BCS)	15.63 (− 1.59, 31.70)	14.71 (− 0.92, 37.20)	11.05 (− 16.74, 30.51)
Stage	− 12.04 (− 17.59, − 7.07)*	− 16.64 (− 23.77, − 9.74)*	− 40.23 (− 41.28, − 9.74)*
Grade	− 9.68 (− 16.89, − 2.70)*	− 14.53 (− 26.03, − 5.47)*	− 14.20 (− 25.80, − 3.51)*

Table 3. Multivariate analysis of prognostic factors of breast cancer survival with Portnoy, Wang and Wang, Bottai and Zhang, Yang and De Backer methods. *Coef.* Estimated parameter, *CI* Confidence Interval. *P-value < 0.05.

Discussion

In the present study, CQR methods were compared in modeling the prognosis factors of breast cancer survival based on the breast cancer data collected from Imam Khomeini Hospital. Tehran, Iran.

The analyses of accelerated failure-time models and Cox model showed that the stage of disease and grade of tumor were identified as the prognostic factors of breast cancer survival. According to the Cox model, the stage of disease and tumor grade increased the risk of death. In the AFT model, the median of survival time decreased as the stage of disease and tumor grade increased. The analyses of CQR models showed that type of surgery, stage and grade were the effective factors on the survival of patients with breast cancer. These results are consistent with most findings in the existing literature, although a direct comparison of the effect size is difficult due to the fact that the majority of works report hazard ratios or odds ratios. Among the significant factors in this study, the effect of surgical method on the survival of breast cancer patients is one of the most addressed issues in recent studies[35–37]. Our study showed that 20th percentile survival time increased for women with BCS surgery, based on Portnoy methods. In the study of Hofvind, by controlling other factors, the hazard of death in MRM is 1.7 times higher than BCS[38]. Meanwhile, there has been no significant difference between two surgical methods in the study of Quan[39]. In most of quantiles stage and grade have significant effect on survival. In recent study, Saadatmand describe overall survival of female patients with breast cancer from two time cohorts (1999–2005 and 2006–2012) in a nationwide population based study[40]. Their results emphasize the importance of tumor stage at diagnosis of breast cancer, as it still greatly affects overall survival. Rottenberg examine differences in survival among older women diagnosed with breast cancer, according to age and disease stage at time of diagnosis. And showed that stage disease among older women became a less powerful predictor of mortality with rising age[41].

The CQR models allow covariate effects to change in people with different risk. Thus, it is a flexible model for controlling heterogeneity due to covariates. In our studies, coefficients of prognostic factor were different in quantiles that showed different effect of prognostic factor survival time in each quantile. Base on Portnoy method
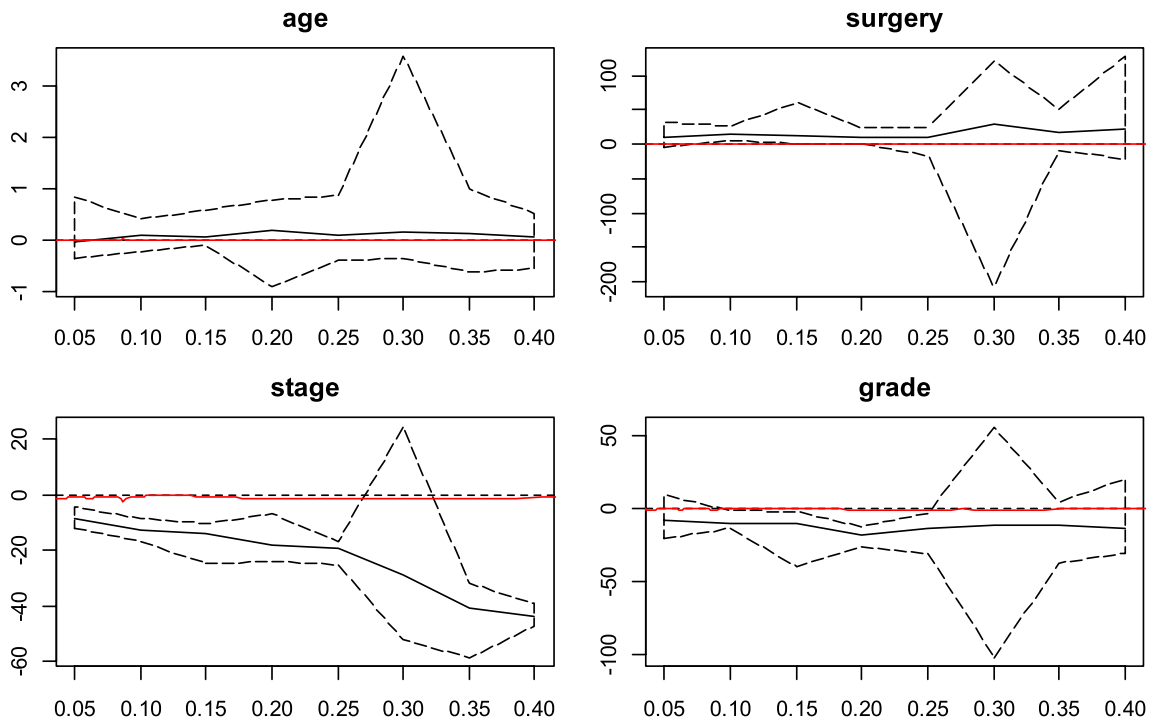
Figure 2. Censored quantile regression coefficients plots and their confidence intervals (dashed line) for Portnoy method and conditional quantile effects estimated by Cox model (red line).
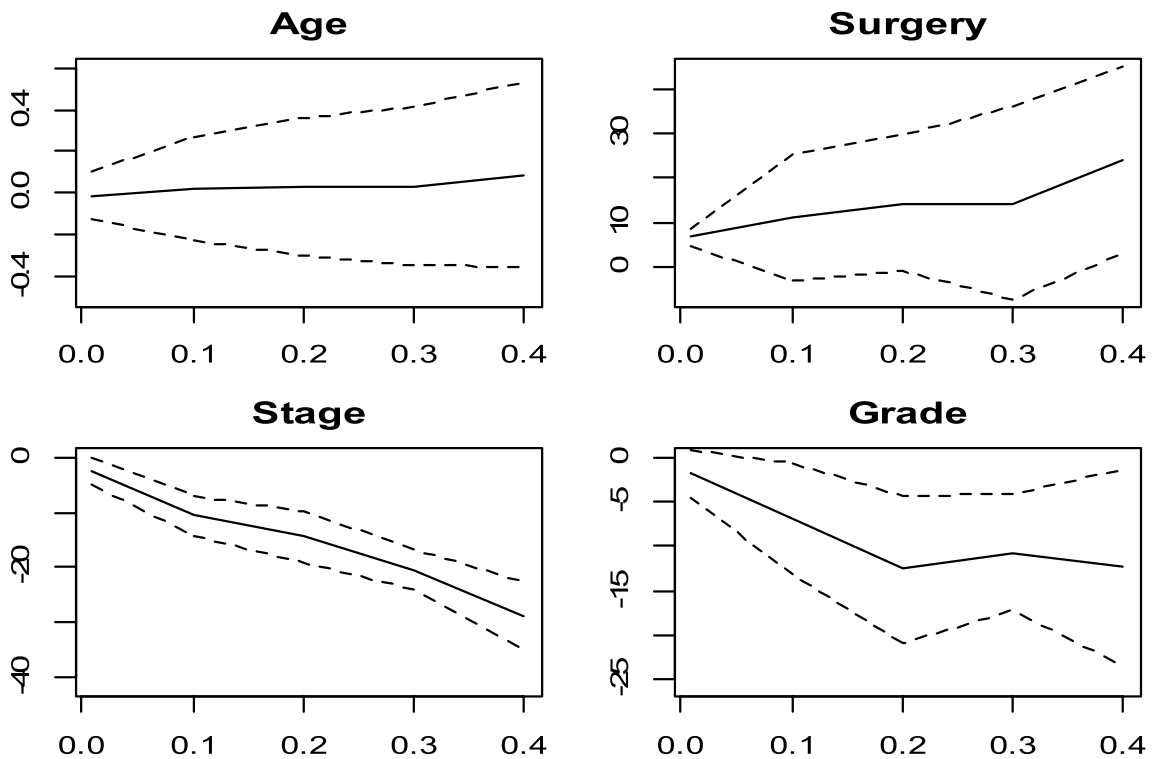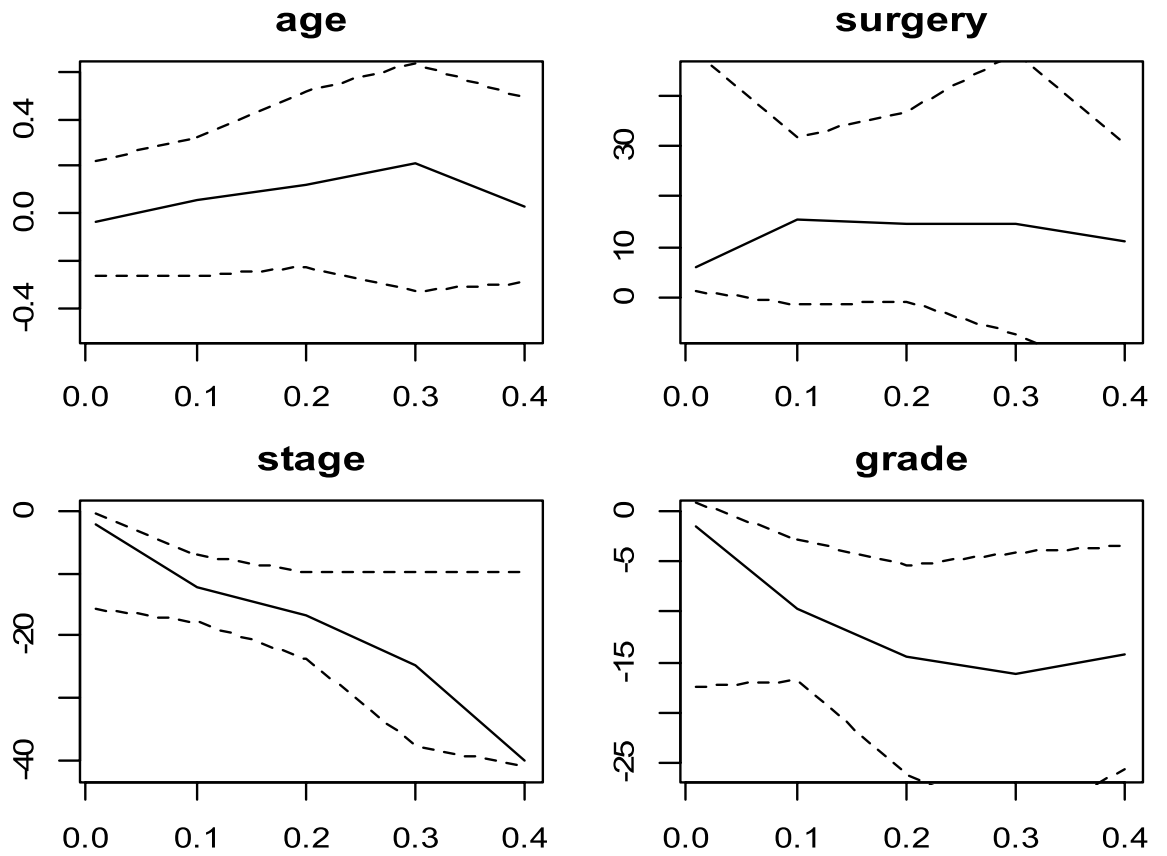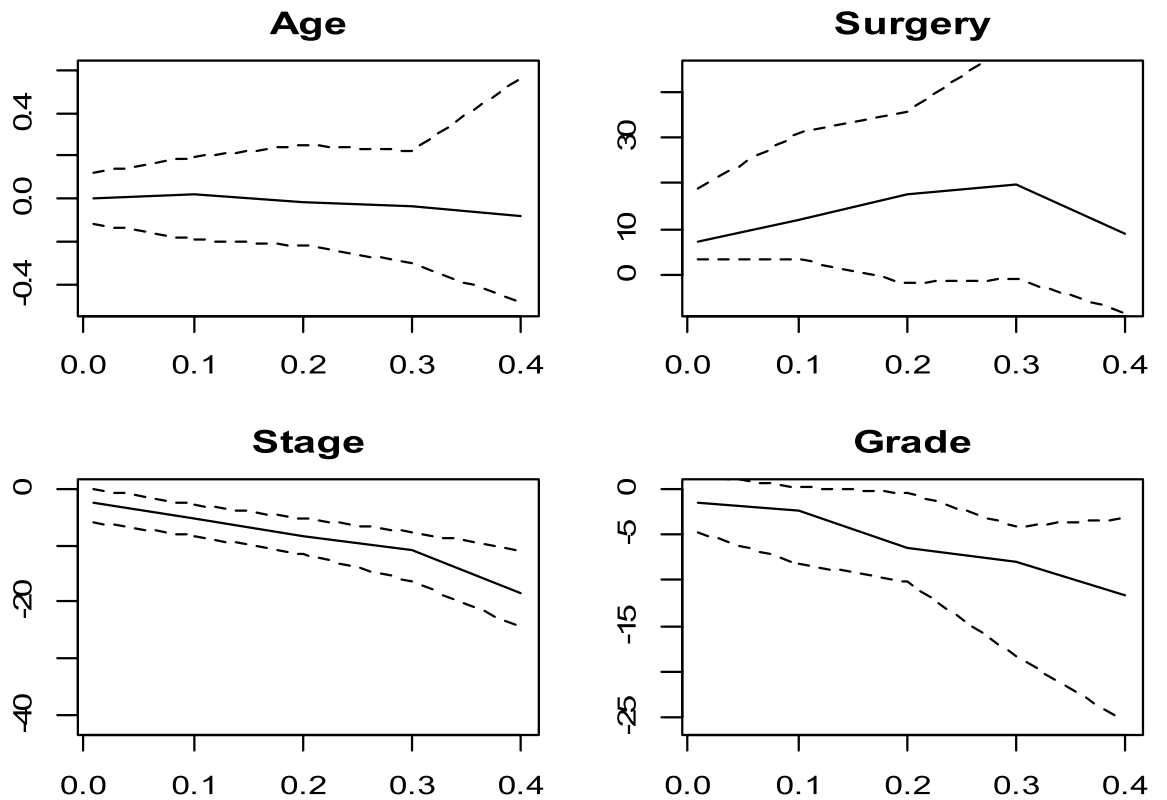

Figure 3. Censored quantile regression coefficients plots and their confidence intervals (dashed line) for Bottai and Zhang method.

high stage of disease decreased 10th, 20th and 40th percentiles survival time 12.61, 18.06 and 43.58 months, respectively.

The difference in the interpretation of the parameters is the biggest one among these models. The Cox model examines the covariate effects on the hazard function. In addition, the Cox model shows the hazard of death per unit of increase in covariate when the event is death. In this study, the hazard of death was 3.12 times for a

Figure 4. Censored quantile regression coefficients plots and their confidence intervals (dashed line) for Yang method.

Figure 5. Censored quantile regression coefficients plots and their confidence intervals (dashed line) for Wang and Wang method.
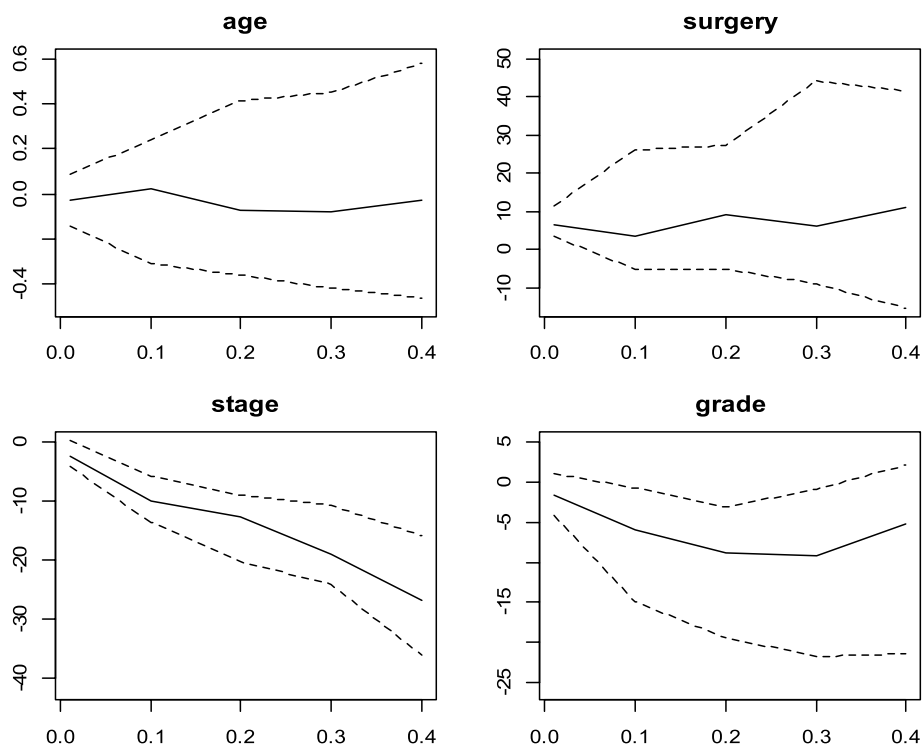
Figure 6. Censored quantile regression coefficients plots and their confidence intervals (dashed line) for De Backer method.
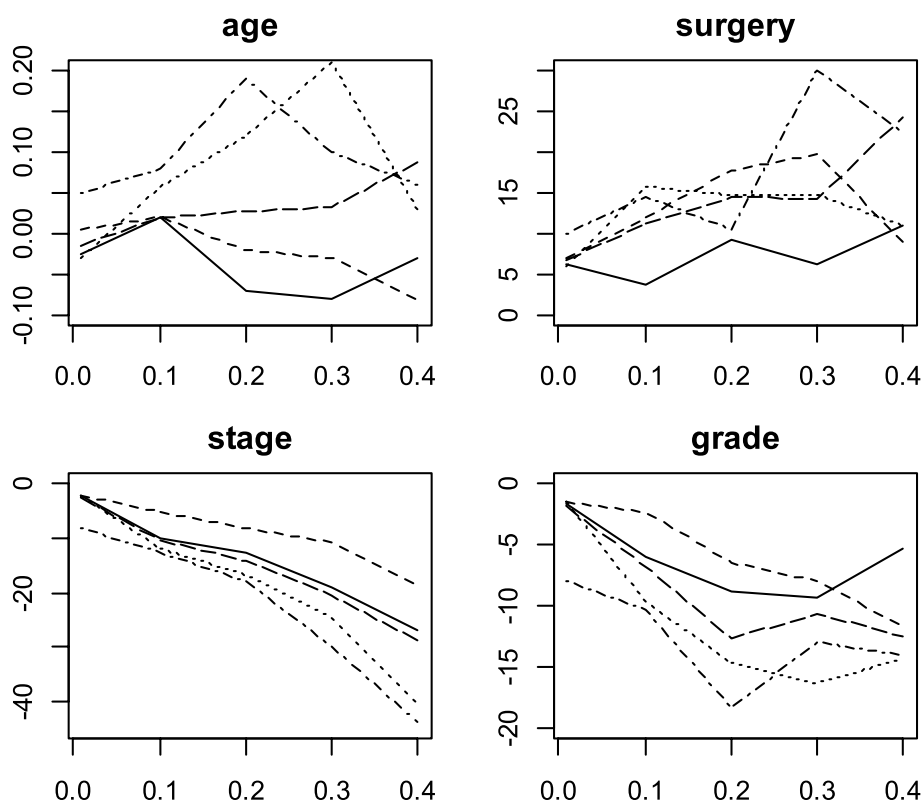

Figure 7. Censored quantile regression coefficients plots for Portnoy (dotdash line), Wang and Wang (dashed line), Bottai and Zhang (longdash line), Yang (dotted line), and De Backer (solid line) methods.

higher stage of the disease, and the hazard of death was 1.71 times for a higher degree of tumor grade. Contrary to the proportional hazards model which describes how predictors influence the hazard function, the AFT model assumes a direct association between predictors and survival time, which makes interpretation easier. If the median survival time to event is considered and the accelerate factor is greater than one, the median survival time increases by the accelerate factor with increasing one unit covariate while the median survival time to event decreases if it is less than one. In the present study, based on AFT model, the median survival time decreased by a factor of 0.38 and 0.63 with increasing stage of disease and grade of tumor, respectively. The results of this model can be expressed as a proportional hazard, in which the interpretation is similar to the Cox model. The interpretation of coefficients in CQR model is considered as the changing rate quantile of dependent variable per one unit change in independent variable, like other linear models. If the event is death, it can be expressed as the covariate effects on the patients' lifetime.

Modeling the breast cancer data with CQR models indicated that most of the time all models acted the same to determine prognostic factors of breast cancer survival but sometime, significant factors and their coefficients were different. All models considered the stage of disease and grade of tumor as prognosis factors. With regard to the coefficients of covariates in different quantiles, the coefficients of Portnoy and Yang method were close to each other and the coefficients of Bottai and Zhang and De Backer methods were close to each other and they were different from Wang and Wang method. Peng and Huang compared their method with Portnoy's method and showed that both methods could represent very similar results[6]. Bottai and Zhang compared Laplace regression method with Peng-Huang and Portnoy methods by using simulation and indicated that the advantages of their method include giving the same results and accurate convergence, while two other methods sometimes failed to converge, and involve fast calculations[8]. Wang and Wang showed that the new approach adopts a preliminary local Kaplan–Meier estimator and results a weighted quantile regression. They established, utilizing results in modern empirical process theory, the consistency and asymptotic normality of the resulted estimator[7]. Base on simulation studies and the analysis of real data, the proposed method has shorter interval estimates than Portnoy's procedure[7]. De backer et al. in their study indicated in an extensive simulation study that the resulted quantile regression estimator respect to established check-based formulations have less variance results. From a theoretical prospect, both consistency and asymptotic normality of the proposed estimator for linear regression are obtained under classical regularity conditions[10]. Yang et al. indicated that the Yang's method presents an estimator is able to achieve significant efficiency gains in comparisons with Portnoy's estimator[9].

The assumptions required in each of the models should be considered while using these models. The proportional hazards assumption is the most important assumption of the Cox model and what it means is that the ratio of the hazards for any two individuals is constant over time. In this model, no assumptions are made about form of the baseline hazard. However, the distribution of survival time is sometimes specific or assuming a parametric form is logical. In these cases, parametric methods are used. Common parametric distributions in survival models include Weibull, Generalized Gamma, Log-Normal, Log-Logistic. In Bottai and Zhang method, it is assumed that the error terms follow asymmetric Laplace distribution. However, Yu and Moyeed (2001) showed that the model performs well when the error terms follow other distributions[42]. Portnoy and Yang methods require just global linearity assumption[5,9]. Wang and Wang and De Backer methods have local linearity assumption[7,10].

Computational time is another important issue in comparing these models. According to our data, the computational time of Portnoy, Bottai and Zhang and Yang methods is shorter than other methods.

It is necessary task to measure the goodness of survival models. Although for the model diagnostics of quantile regression with complete data some tools, such as the worm plot, have been proposed, for censored quantile regression is still greatly underdeveloped[43,44]. Designing effective model diagnostic tools for censored quantile regression warrants more in-depth research.

The high percentage of right censoring is regarded as one of the limitations of this study. By this way, modeling 50th percentile of survival time requires more follow-up time to increase the percentage of event. Thus, the comparison of models in higher quantiles was not possible.

Conclusions

For the CQR models, various approaches have been proposed that the most practical of them include: Portnoy[5], Wang and Wang[7], Bottai and Zhang[8], Yang et al.[9] and De Backer et al.[10] methods. Portnoy[5] generalized the Kaplan–Meier method with a recursively weighted estimation algorithm under the global linearity assumption of the conditional quantile functions. To overcome the linearity assumption, Wang and Wang[7] developed a method by non-parametrically estimating the conditional survival distribution via kernel smoothing. In 2010, Laplace regression was introduced as a parametric method for modeling the conditional quantile of censored data by Bottai and Zhang. Yang et al.[9] employed a variation of the data augmentation algorithm base on the general principle of data augmentation[15]. De Backer et al.[10] investigate a new procedure that used "check" loss function. The CQR methods acted the same to determine prognostic factors of breast cancer survival in most of the time. The estimated coefficients of five methods were close to each other for quantiles lower than 0.1 and they were different from quantiles upper than 0.1.

Data availability

The datasets used and/or analyzed during the current study are available from the corresponding author on reasonable request.

References

1. Klein, J. P. & Moeschberger, M. L. *Survival Analysis: Techniques for Censored and Truncated Data* Vol. 1230 (Springer, 2003).
2. Kalbfleisch, J. D. & Prentice, R. L. *The Statistical Analysis of Failure Time Data* Vol. 360 (Wiley, 2011).
3. Xue, X., Xie, X. & Strickler, H. D. A censored quantile regression approach for the analysis of time to event data. *Stat. Methods Med. Res.* **27**, 955–965 (2018).
4. Koenker, R. & Hallock, K. F. Quantile regression. *J. Econ. Perspect.* **15**, 143–156 (2001).
5. Portnoy, S. Censored regression quantiles. *J. Am. Stat. Assoc.* **98**, 1001–1012 (2003).
6. Peng, L. & Huang, Y. Survival analysis with quantile regression models. *J. Am. Stat. Assoc.* **103**, 637–649 (2008).
7. Wang, H. J. & Wang, L. Locally weighted censored quantile regression. *J. Am. Stat. Assoc.* **104**, 1117–1128 (2009).
8. Bottai, M. & Zhang, J. Laplace regression with censored data. *Biom. J.* **52**, 487–503 (2010).
9. Yang, X., Narisetty, N. N. & He, X. A new approach to censored quantile regression estimation. *J. Comput. Graph. Stat.* **27**, 417–425 (2018).
10. De Backer, M., El Ghouch, A. & Van Keilegom, I. Linear censored quantile regression: A novel minimum-distance approach. *Scand. J. Stat.* **47**, 1275–1306 (2020).
11. Liu, Y. & Bottai, M. Mixed-effects models for conditional quantiles with longitudinal data. *Int. J. Biostat.* **5**, 1 (2009).
12. Farcomeni, A. Quantile regression for longitudinal data based on latent Markov subject-specific parameters. *Stat. Comput.* **22**, 141–152 (2012).
13. Lee, D. & Neocleous, T. Bayesian quantile regression for count data with application to environmental epidemiology. *J. R. Stat. Soc. C* **59**, 905–920 (2010).
14. Yuan, Y. & Yin, G. Bayesian quantile regression for longitudinal studies with nonignorable missing data. *Biometrics* **66**, 105–114 (2010).
15. Tanner, M. A. & Wong, W. H. The calculation of posterior distributions by data augmentation. *J. Am. Stat. Assoc.* **82**, 528–540 (1987).
16. Koenker, R. & Bassett, G. Jr. Regression quantiles. *Econometrica* **46**, 33–50 (1978).
17. Li, R. & Peng, L. Quantile regression for left-truncated semicompeting risks data. *Biometrics* **67**, 701–710 (2011).
18. Peng, L. & Fine, J. P. Competing risks quantile regression. *J. Am. Stat. Assoc.* **104**, 1440–1453 (2009).
19. Fan, C., Ma, H. & Zhou, Y. Quantile regression for competing risks analysis under case-cohort design. *J. Stat. Comput. Simul.* **88**, 1060–1080 (2018).
20. Luo, X., Huang, C. Y. & Wang, L. Quantile regression for recurrent gap time data. *Biometrics* **69**, 375–385 (2013).
21. Sun, X., Peng, L., Huang, Y. & Lai, H. J. Generalizing quantile regression for counting processes with applications to recurrent events. *J. Am. Stat. Assoc.* **111**, 145–156 (2016).
22. Frumento, P. & Bottai, M. An estimating equation for censored and truncated quantile regression. *Comput. Stat. Data Anal.* **113**, 53–63 (2017).
23. Narisetty, N. & Koenker, R. Censored quantile regression survival models with a cure proportion. *J. Econ.* https://doi.org/10.1016/j.jeconom.2020.12.005 (2021).
24. Chen, S. Quantile regression for duration models with time-varying regressors. *J. Econ.* **209**, 1–17 (2019).
25. Moghadami, F. Z., Abolghasemi, J., Asgari, D. A. & Gohari, M. Survival analysis of patients with breast cancer using the Aalen's additive hazard model. (2011).
26. Saki, A., Hajizadeh, E. & Tehranian, N. Evaluating the risk factors of breast cancer using the analysis of tree models. *Horizon Med. Sci.* **17**, 60–68 (2011).
27. Akarolo-Anthony, S. N., Ogundiran, T. O. & Adebamowo, C. A. Emerging breast cancer epidemic: Evidence from Africa. *Breast Cancer Res.* **12**, 1–4 (2010).
28. Khodabakhshi, R., RezaGohari, M., Moghadamifard, Z., Foadzi, H. & Vahabi, N. Disease-free survival of breast cancer patients and identification of related factors. *Razi J. Med. Sci.* **18**, 27–33 (2011).
29. Roué, T. *et al.* Predictive factors of the survival of women with invasive breast cancer in French Guiana: The burden of health inequalities. *Clin. Breast Cancer* **16**, e113–e118 (2016).
30. Davino, C., Furno, M. & Vistocco, D. *Quantile Regression: Theory and Applications* Vol. 988 (Wiley, 2013).
31. Efron, B. *Proceedings of the fifth Berkeley symposium on mathematical statistics and probability* 831–853 (1967).
32. Neocleous, T., Branden, K. V. & Portnoy, S. Correction to censored regression quantiles by S. Portnoy, 98 (2003), 1001–1012. *J. Am. Stat. Assoc.* **101**, 860–861 (2006).
33. Nelder, J. A. & Mead, R. A simplex method for function minimization. *Comput. J.* **7**, 308–313 (1965).
34. Beran, R. *Nonparametric Regression with Randomly Censored Survival Data* (Springer, 1981).
35. Hartmann-Johnsen, O. J., Kåresen, R., Schlichting, E. & Nygård, J. F. Better survival after breast-conserving therapy compared to mastectomy when axillary node status is positive in early-stage breast cancer: A registry-based follow-up study of 6387 Norwegian women participating in screening, primarily operated between 1998 and 2009. *World J. Surg. Oncol.* **15**, 1–10 (2017).
36. Litière, S. *et al.* Breast conserving therapy versus mastectomy for stage I-II breast cancer: 20 year follow-up of the EORTC 10801 phase 3 randomised trial. *Lancet Oncol.* **13**, 412–419 (2012).
37. Hartmann-Johnsen, O. J., Kåresen, R., Schlichting, E. & Nygård, J. F. Survival is better after breast conserving therapy than mastectomy for early stage breast cancer: A registry-based follow-up study of Norwegian women primary operated between 1998 and 2008. *Ann. Surg. Oncol.* **22**, 3836–3845 (2015).
38. Hofvind, S. *et al.* Women treated with breast conserving surgery do better than those with mastectomy independent of detection mode, prognostic and predictive tumor characteristics. *Eur. J. Surg. Oncol.* **41**, 1417–1422 (2015).
39. Quan, M. L. *et al.* The effect of surgery type on survival and recurrence in very young women with breast cancer. *J. Surg. Oncol.* **115**, 122–130 (2017).
40. Saadatmand, S., Bretveld, R., Siesling, S. & Tilanus-Linthorst, M. M. Influence of tumour stage at breast cancer detection on survival in modern times: Population based study in 173 797 patients. *BMJ* **351**, 4901 (2015).
41. Rottenberg, Y., Naeim, A., Uziely, B., Peretz, T. & Jacobs, J. M. Breast cancer among older women: The influence of age and cancer stage on survival. *Arch. Gerontol. Geriatr.* **76**, 60–64 (2018).
42. Yu, K. & Moyeed, R. A. Bayesian quantile regression. *Stat. Probab. Lett.* **54**, 437–447 (2001).
43. Alhamzawi, R. & Yu, K. Bayesian Tobit quantile regression using g-prior distribution with ridge parameter. *J. Stat. Comput. Simul.* **85**, 2903–2918 (2015).
44. Alhamzawi, R. & Ali, H. T. M. Bayesian tobit quantile regression with penalty. *Commun. Stat. Simul. Comput.* **47**, 1739–1750 (2018).

Author contributions

Competing interests

The authors declare no competing interests.

Additional information

Correspondence and requests for materials should be addressed to H.Z.

Reprints and permissions information is available at www.nature.com/reprints.

Publisher's note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.