RESEARCH ARTICLE

Determination of a Change Point in the Age at Diagnosis of Breast Cancer Using a Survival Model

Mahbubeh Abdollahi1, Ebrahim Hajizadeh1*, Ahmad Reza Baghestani2, Shahpar Haghighat3

Abstract

Breast cancer, the second cause of cancer-related death after lung cancer and the most common cancer in women after skin cancer, is curable if detected in early stages of clinical presentation. Knowledge as to any age cut-off points which might have significance for prognostic groups is important in screening and treatment planning. Therefore, determining a change-point could improve resource allocation. This study aimed to determine if a change point for survival might exist in the age of breast cancer diagnosis. This study included 568 cases of breast cancer that were registered in Breast Cancer Research Center, Tehran, Iran, during the period 1986-2006 and were followed up to 2012. In the presence of curable cases of breast cancer, a change point in the age of breast cancer diagnosis was estimated using a mixture survival cure model. The data were analyzed using SPSS (versions 20) and R (version 2.15.0) software. The results revealed that a change point in the age of breast cancer diagnosis was at 50 years age. Based on our estimation, 35% of the patients diagnosed with breast cancer at age less than or equal to 50 years of age were cured while the figure was 57% for those diagnosed after 50 years of age. Those in the older age group had better survival compared to their younger counterparts during 12 years of follow up. Our results suggest that it is better to estimate change points in age for cancers which are curable in early stages using survival cure models, and that the cure rate would increase with timely screening for breast cancer.

Keywords: Breast cancer - survival analysis - survival cure model - change point in age of diagnosis - cut-off point

Introduction

Cancer is the disease of cells in which cancer cells form malignant neoplasm and they are unlimitedly and uncontrollably reproduced (Hajian et al., 2011). Carcinogenesis, the process of changing a natural cell to cancer cell, is a complex and multistage process (Hatami et al., 2004). Breast cancer is a multistage cancer which appears by a small mass in breast(s) (first stage). This tumor gradually grows and expands to areas near breast (auxiliary lymph nodes) (second stage). The mass might extend to sternum bone, chest wall, breast skin, or lymph nodes above and below the sternum (third stage). In the third stage of sickness, it might be operable or inoperable. If the sickness reaches the lymph nodes above the collarbone, it is inoperable. In the fourth stage, the disease penetrates other organs such as bones, lungs, brain, and liver. Therefore, this disease is curable if it is diagnosed in early stages (Parker, 2012). Breast cancer, after lung cancer, is the second leading cause of cancer mortality among women, and it is the most common cancer in women after skin cancer. In Iran, breast cancer is the third common cancer after cervical and skin cancers (Hatami et al., 2004). Since cancer is the second leading cause of human mortality (Hanahan and Weinberg, 2000) and health care plays an important role, the progress of medical science in the field of cancer is a major goal in health care programs (Goodman et al., 2006). Disease free survival is one of common criteria to evaluate cancer patients. Disease free survival is defined as the time from the onset of illness until recurrence or death (Lamont et al., 2006).

Multiple studies focus on prognostic group definition in oncology. Classifying quantitative prognostic variable, and determination of change point (cut point) for quantitative variable is important to comprehend the disease and plan the treatment. Prognostic homogenous groups can be created by classifying an important quantitative prognostic variable (Buettner et al., 1997). There are many different methods to calculate change point (cut point) such as using determined change point value in previous studies, using sample quantile like median. Other methods that are worth mentioning are optimized change point (Buettner et al., 1997), p-value minimization (Heinzl and Tempfer, 2001), and change

1Department of Biostatistics, Faculty of Medical Sciences, Tarbiat Modares University, 2Department of Biostatistics, Faculty of Paramedical Sciences, Shahid Beheshti University of Medical Sciences, 3Breast Cancer Research Center (BCRC), ACECR, Tehran, Iran *For correspondence: hajizadeh@modares.ac.ir

Asian Pac J Cancer Prev 17, Cancer Control in Western Asia Special Issue, 5-10
point models (Wang et al., 2007). Multiple studies were conducted to determine cut point using change point models. Goodman et al. (2006) conducted a study to determine change point in hazard function in survival analysis. In this paper, hazard function and change point were estimated in hazard function trend. The models were fitted on data for breast cancer patients in order to determine change point in hazard function. To this end, SEER data were used where 7,224 white and 682 black women attended the study. The results were as follows: change point was reported 1.3 after the diagnosis of cancer for white women who developed breast cancer as well the hazard experienced rising trend after three years of diagnosis and then falling until the end of study. For black women, three change points were obtained at 3, 20, and 22 years in hazard function as well hazard function rose until the first change point, then fell until the second point, after that rose to the third point and finally declined until the end of the study (Goodman et al., 2006).

Fernando Alarid et al. (2013) conducted a study to determine cut point for disease survival time among colorectal cancer patients in stage 3. In this paper, spline model was used instead of Bayesian Markov chain to determine change point of survival time. The estimations of these models were far more accurate than those of ordinary models (Alarid, 2014).

Minh Luong et al. (2013) conducted a study in order to determine change point using hidden Markov Model. Then they used this model to find mutations of chromosome 10 in colorectal cancer patients. In this study which did not have 261,563 SNP factor, mutation points of chromosome 10 were obtained in colorectal cancer patients (Luong et al., 2013). Age change point is the age when a process begins to alter (Assareh and Mengersen, 2012). There is strong evidence that breast cancer is preventable and clinical practice guidelines-based screening is advised (Parker, 2012). Also, knowing where change point occurs is important for screening (Goodman et al., 2006). Therefore, change point cure model is used to estimate change point or cut point of age variable. The advantage of this model lies in the fact that some percentages of breast cancer patients have long survival time and they are considered cured. In other words, if this disease is diagnosed in early stages, some patients might have longer survival time compared to others and even they might cure. In such cases, using standard survival models such as semi parametric model and parametric survival models are not appropriate because these models assume that all participants in the study under consideration will experience consequences such as death or recurrence (Farewell, 1982; Maller and Zhou, 1996). In this state in which there is patients with long survival time or cured, survival cure models will be used to analyze survival data (Machin et al., 2006; Taweab et al., 2015).

In this paper, survival cure model is used to determine change point in diagnosis of breast cancer age. We will reach change point of breast cancer diagnosis age and cure ratio for the breast cancer patients in Tehran using mixture cure model with exponential distribution which has one change point in age covariate (Othus et al., 2012). It is noteworthy that likelihood function falls into two sections in this model: the first section of likelihood function is related to data which are smaller than change point parameter; the second section is related to larger data than change point parameter. Then cure ratio parameter, exponential distribution rate, and age covariate change point are simultaneously estimated using a parametric model and smoothing methods. Since determining change point in cancer diagnosis age helps planning the treatment substantially, one of objectives in this paper is to obtain change point in breast cancer diagnosis age.

Materials and Methods

In this longitudinal study, as many as 568 breast cancer patients who referred to Breast Cancer Research Center, Tehran, from 1986 to 2006 were studied. The beginning of the study was based on pathological diagnosis of breast cancer. Patients were contacted according to their telephone numbers in their files. Their latest disease progression, demographic and clinical information were completed after 6 years of follow up in 2012. The result of this study was disease free survival time which is the time from the breast cancer diagnosis, the first pathology report about breast cancer development, until the first recurrence. R software version 2.15.0 and SPSS software version 16 were used to analyze data. After appropriate detection of cure model fitness to data, the exponential cure model likelihood function was written along with change point in age covariate. Accordingly, maximum likelihood estimation was obtained for model parameters using numerical search methods.

Cure model

there are two groups of cure models: Mixture cure model and Non-Mixture cure model (Li et al., 2007; Taweab et al., 2015). In mixture cure model, patients fall into two groups: cured and uncured.

In mixture cure model, if p is cure ratio, survival function for cure model is as follows:

\[ S(t) = p + (1-p) S_U(t) \]

Where \( S_U(t) \) is survival function for uncured patients which used Gamma, lognormal, Weibull, etc. distributions (Maller and Zhou, 1996; Li et al., 2007). For this model, when \( t=0 \), then \( S(0) = p + (1-p) S_U(0) = p + (1-p) = 1 \); it means that all individuals are alive at \( t=0 \). When \( t = \infty \), then \( S(\infty) = p + (1-p) S_U(\infty) = p + (1-p) \times 0 = p \). It means that a proportion of people (p) remain alive (Machin et al., 2006).

The main objective in mixture cure models is to estimate cure ratio and survival function for the uncured (Corbieri et al., 2009). If cure ratio estimation is close to zero, then the cure model might not be appropriate (Othus et al., 2012). Censoring data have random patterns. For this reason, safe individuals cannot be detected from censoring (Maller and Zhou, 1996) and only cure probability in a certain community can be estimated through maximum likelihood methods (Maller and Zhou, 1996; Corbieri et al., 2009).

Two assumptions need to be studied prior to cure model usage. The first pre-assumption is testing the presence of cured fraction in community (Do a significant
number of people have long-term survival or not) and the second pre-assumption is studying long enough duration of follow-up. However, clinical experiences and biological evidence prove long enough duration of follow-up in some cases. In the case of presence of cured patients in the study, Kaplan-Meier survival curve does not reach zero on the far right and it remains at a horizontal straight line. Parametric tests can be used to study the cured patients in data (Maller and Zhou, 1996).

Mixture cure model along with change point in a covariate

This model was introduced by Megan Othus et al. in 2012 in order to determine change point at a quantitative variable. In this model, dividing mixture cure model likelihood function into two sections (less than change point parameter and greater than change point parameter), one is able to estimate quantitative variable change point. In addition, it is possible to estimate cured ratio and hazard rate in both sections of the model at the same time. In mixture cure model along with change point in a covariate, likelihood function is written as follows for each individual:

$$l=(f(t)^{\delta} S(t)^{1-\delta})^{(I(X<\tau))} (f(t)^{\delta} S(t)^{1-\delta})^{(I(X>\tau))}$$

Where $f(t)$ is density function of time variable till the event, $S(t)$ is survival function of time variable until the event, and $\tau$ is change point parameter for quantitative variable of $X$. Also, $\delta$ shows event condition for each individual so that it is considered one in case of event occurrence and zero in case of censoring. Then maximum likelihood method and numerical search methods are used to estimate change point model parameters.

Results

Of total number of 568 patients in the study, survival time of 180 patients (31.7%) was accurately recorded and 388 patients (68.3%) were considered right censoring because accurate recording of survival time is not possible for all participants. The mean age of cancer diagnosis is 46.3 with standard deviation of 11.1. Median of the diagnosis age was 45. Follow-up median was calculated 68.5 months. As many as 66.7% were 50 or younger and 33.3% were older than 50.

Figure 1 shows Kaplan-Meier curve to study the patients with long survival time. 25-year sufficient follow-up time is simply clear. As it is clear from the curve, Kaplan-Meier curve becomes flat after almost 23 years and remains unchanged until the end.

As it is clear from Kaplan-Meier curve, almost 40% of individuals have lived for 23 years after the follow up until the end of study. In other words, the majority of patients are not likely to die as a result of cancer. Also, after conducting parametric test to study cured individuals in the study, the hypothesis of cured patients was not rejected. Therefore, two pre-assumptions of cure model (the presence of cured patients in the study and sufficient follow-up time in the study) are true.

In order to reach change point in diagnosis age, exponential cure model along with change point to data was fit. Table 1 lists the results of model fitness. The results of table 1 indicate the fact that change point in age of colorectal cancer patients is nearly 50 years. Also, cure ratio is 35% among the patients who are 50 or younger (younger individuals) and cure ratio is 57% among the patients who are older than 50 (older individuals). The levels of cure are 0.01 and 0.02 among young and old patients, respectively.

Since change point in age was 50, Kaplan-Meier curves were drawn for 50 or younger and older-than-50 individuals (Figure 2). As it can be seen, after almost 12 years from the follow-up, individual survival of older people is longer than younger ones. In other words, survival time of the older is more than that of the younger;

| Parameter                      | Estimate | 95% CI    |
|-------------------------------|----------|-----------|
| Change-point estimate ($\tau$) | 49.5     | (46.5, 54.4) |
| Cure estimate for age < $\tau$| 0.4      | (0.5, 0.9)     |
| Hazard estimate for age < $\tau$| 0.0   | (0.0, 0.3)   |
| Cure estimate for age ≥ $\tau$| 0.6      | (0.5, 0.7)     |
| Hazard estimate for age ≥ $\tau$| 0.0   | (0.0, 0.2)   |

Figure 1. Estimated Kaplan-Meier Graph from Breast Cancer in Tehran

Figure 2. Kaplan-Meier Graph Based on Age of Breast Cancer Patients in Tehra
However, the graphs are close to each other in the first 20 years. Kaplan-Meier curve becomes flat after 9 years among younger individuals, and almost 22 years among the older individuals. In other words, the young and old accounted for 35% and 57% cured.

Discussion

This study was conducted to determine change point (cut point) of age covariate in breast cancer patients. It was obtained nearly 50 years. In 1999, Contal et al. used a change point method in semi-parametric Cox survival model in order to determine change point in breast cancer diagnosis age. In this method, assuming non-linear impact of diagnosis age on cancer survival time, change point of breast cancer diagnosis age was obtained 41 years. Also, this method reached 37 years using another method of cut point of cancer survival time which is inconsistent with the result of this paper (Contal and O’Quigley, 1999).

Various cut points were used in different breast cancer studies; for instance, GayosoDiz et al. in Spain (2013), Anderson et al. (2012), DeSanctis et al. in the US (2011) used 50 cut point for age which are consistent with the results of this study. However, Alieldin et al. (2014) used 40 cut point for age. Some cut points were used as age variable in some studies (Li et al., 2003; Linden et al., 2012; Minicozzi et al., 2013). The cut point used in these studies is not consistent with this study. This difference is associated with differences in statistical population and different cut points in previous studies.

In this study, the cured ratio is 57% among old individuals (50 and older), which is higher by 22% compared to younger ages (at or younger than 50). In other words, if cancer occurs in older ages, higher proportion of people are likely to heal up. According to studies conducted by Lee in 2010, this issue is associated with the fact that breast cancers diagnosed before menopause are often associated with adverse pathologic features.

The results of this paper show that survival of young patient is more than that of old patients in the first 12 years of follow-up. This relationships has an adverse condition after 12 years so that old patient survival is longer than that of young patient survival as time passes. The results in this paper are consistent with those of Rosenberg research in 2005. Rosenberg (2005) indicated that younger patients experience a more aggressive form of cancer. Also, the study conducted by Zahl and Tryril in 1997 pointed out the interaction between age and time which is consistent with this study, describing various impacts of age in patients’ survival using the interaction between age and time.

The results of some studies show that age was not found to be associated with patients’ survival (Khodabakhshi et al.; Akbari et al., 2012; Alieldin et al., 2014). The results of other studies show that survival was found to have an inverse association with age so that the young’s survival is more than that of the old (Berrino et al., 2009; Heydari et al., 2012). Also, the results of some other studies showed that the survival is more in the old than the young (Wingo et al., 1998; Maggard et al., 2003).

In this paper, mean and median of cancer diagnosis age are 46.25 and 45, respectively, consistent with the results of other studies in Iran which reported the mean age 45 to 50. This age is younger than European and American countries (Zafarghandi et al., 1998; Vahdaninia and Montazeri, 2004; Gohari et al., 2006) so that the mean age is as follows: mean age of 56.7 in the study conducted by Carlo et al (2005), age median of 60 based on the study conducted by Andres in Philadelphia (2013), age median of 49 based on the study conducted by Alieldin et al. in Cairo (2014), and age median of 61 in the US (Howlader et al., 2013).

In this paper, follow-up median was obtained 68.45 months. It was 46.8 month in the study conducted by Esserman et al (2012) in Germany, and 96 months in the study conducted by Goldhirsch in the US in 2013.

In this paper, age split was as follows: 33.3% in older-than-50 age group and 66.7% in 50 or younger age group. In other words, the occurrence is more in the young than the old which is consistent with the results of some studies (Sertkaya and Sözer, 2003; Mokhtari Hesari et al., 2014). On the contrary, the results of some studies show that the cancer occurrence percentage is higher among the old which is inconsistent with the results of this paper (Yaghmaei et al., 2008; DeSanctis et al., 2011).

In this paper, we found change point of age (cut point) using a cure model along with change point in age covariate. The advantage of this model is consideration of the cured with cure model, change point of age using a parametric model, simultaneous estimation of cure ratio, and rate of exponential distribution in a model simultaneously for both data sections (data prior to age change point and data after change age point).

The results of studies show that cancer family history is one of important occurrence factors of breast cancer (Parker, 2012). This issue is considered limitation due to lack of record. Another study in this field is proposed. Cure model along with a change point in a quantitative covariate can be used to calculate cut point at multistage cancers diagnosis age. In this study, in addition to determining cut-point in cancer diagnosis age, hazard rate and percentage of the cured were obtained before and after the diagnosis age. Assessment to cut point of disease diagnosis age and analysis of created prognostic groups are highly important in screening and treatment planning.

Acknowledgment

A debt of gratitude is owed to Breast Cancer Research Center, ACECR, Tehran to provide the data for this paper.

References

Akbari ME, Khayamzadeh M, Khoshnevis S, et al (2012). Five and ten years survival in breastcancer patients mastectomies vs. breast conserving surgeries personal experience. Iran J Cancer Prev, 1, 53-6.

Alarid F (2014). Calibration of piecewise markow models using a bayeslam change-point analysis through an iterative convex optimization algorithm. The 36th Annual Meeting of the Society for Medical Decision Making, Smdm.

Alieldin NH, Abo-Elazm OM, Bilal D, et al (2014). Age at diagnosis in women with non-metastatic breastcancer: Is it related to prognosis. J Egypt Natl Canc Instr, 26, 23-30.

Andres SA, Wittliff JL, Cheng A (2013). Protein tyrosine
Mahboubeh Abdollahi et al.

(1991-2002). *Koomesh*, 9, 111-6.