Türkiye Klinikleri Biyoistatistik Dergisi
Turkiye Klinikleri Journal of Biostatistics Turkiye Klinikleri J Biostat. 2021;13(2):131-7

ORİJİNAL ARAŞTIRMA ORIGINAL RESEARCH DOI: 10.5336/biostatic.2020-79287

An Application of Cox Mixed Cure Model for Breast Cancer Patients

Meme Kanserli Hastalar için Cox Karma İyileşme Modelinin Uygulanması

Şirin ÇETİN[a], Özge PASİN[b], İsa DEDE[c], Ayşe ÜLGEN[d]

[a]Department of Biostatistics, Gaziosmanpaşa University Faculty of Medicine, Tokat, TURKEY
[b]Department of Biostatistics, İstanbul University İstanbul Faculty of Medicine, İstanbul, TURKEY
[c]Department of Medical Oncology, Mustafa Kemal University Faculty of Medicine, Hatay, TURKEY
[d]Department of Biostatistics, Girne American University Faculty of Health Sciences, Kyrenia, TRNC

ABSTRACT Objective: Survival analyses are used in cancer research in the selection of treatment methods that may be effective in survival time and in investigating prognostic factors. In the literature, Cox regression analysis was performed without investigating whether the assumptions were met in survival analysis. The aim of this study is; to give information about the features of Cox mixed cure models and to show its application for breast cancer patients. **Material and Methods:** In our study, it was observed that breast cancer data did not provide the proportional hazard assumption. In addition, since the survival curve has a flattened tail after 72 months, this study was analyzed with a mixed cure model. In the model, cured and uncured patients are modeled separately. **Results**: In our study, the median survival time of the patients was found to be 72 months (95% confidence interval; 51.88-92.12). Progesterone receptor, estrogen receptor, c-erbB-2, age and lymph node were found to be a significant risk factor for the survival of breast cancer patients in cured and uncured part (for all of them $p<0.05$). **Conclusion:** Nowadays, cancer patients are more likely to recover. For this reason, we think that it would be appropriate to evaluate the survival analyses of cured and uncured cancer patients separately with a mixed cure model instead of Cox regression analysis in modeling the survival analysis of cancer patients.

Keywords: Mixed cure model; breast cancer; estrogen receptor; progesterone receptor

ÖZET Amaç: Kanser araştırmalarında, sağkalım süresinde etkili olabilecek tedavi yöntemlerinin seçiminde ve prognostik faktörlerin bulunmasında, sağkalım analizleri kullanılmaktadır. Literatürde, sağkalım analizlerinde varsayımların sağlanıp sağlanmadığı araştırılmadan, Cox regresyon analizinin yapıldığı görülmektedir. Bu çalışmanın amacı; Cox karma iyileşme modellerinin özellikleri hakkında bilgi vermek ve meme kanserli hastalar için uygulamasını göstermektir. **Gereç ve Yöntemler:** Çalışmamızda, meme kanseri verilerinin orantılı hazard varsayımını sağlamadığı görülmüştür. Ayrıca yaşam süresi eğrisi, 72. aydan sonra düzleşmiş bir kuyruğa sahip olduğundan bu çalışmada, karma iyileşme modeli ile analiz yapılmıştır. Modelde, hastaların iyileşmiş ve iyileşmemiş olanları ayrı ayrı modellenmektedir. **Bulgular:** Çalışmamızda, meme kanseri hastalarının iyileşmiş ve iyileşmemiş olanları ayrı ayrı modellenerek hastaların medyan sağkalım süresi 72 ay (%95 güven aralığı; 51,88-92,12) olarak bulunmuştur. Progesteron reseptörü, östrojen reseptörü, c-erbB-2, yaş ve lenf nodunun iyileşen ve iyileşmeyen bölümlerinde de meme kanseri hastalarının sağkalımı üzerinde anlamlı bir risk faktörü olduğu görüldü (hepsi için $p<0,05$). **Sonuç:** Gün geçtikçe kanser hastalarının iyileşme olasılıkları artmaktadır. Bu sebeple kanser hastalarının sağkalım analizlerinin modellenmesinde, Cox regresyon analizi yerine karma iyileşme modellerinin ayrı ayrı değerlendirilmesinin uygun olacağını düşünmekteyiz.

Anahtar kelimeler: Karma iyileşme modeli; meme kanseri; östrojen reseptörü; progesteron reseptörü

According to the data of the World Health Organization, the risk of getting cancer and the risk of patients dying from cancer is increasing day by day. Therefore, there is a great interest in studies on cancer and new methods are being developed every day. Breast cancer is a heterogeneous disease. Different continents,

Correspondence: Şirin ÇETİN
Department of Biostatistics, Gaziosmanpaşa University Faculty of Medicine, Tokat, TURKEY/TÜRKİYE
E-mail: cetinsirin55@gmail.com

Peer review under responsibility of Turkiye Klinikleri Journal of Biostatistics.

Received: 29 Sep 2020 ***Received in revised form:*** 24 Jan 2021 ***Accepted:*** 15 Feb 2021 ***Available online:*** 15 Mar 2021

counties, ethnicities, and individuals may have different cancer risks, stages, and survival probailities prognostic factors are the most important issues for cancer patients. These factors have an important impact on the treatment and management strategy. The term prognosis may refer to two different things: one on the recurrence of the disease, and another on the death of the patient. As a result, survival analysis on breast cancer can also be classified by the analysis of the treatment-to-reccurence time or treatment-to-death time. Survival analyses are used to examine the natural course of the disease in cancer research or to select the treatment methods that can be most effective on the patient's survival. Recently, special combined treatment methods have been used for cancer patients. The clinician must make the right decision when applying these personalized treatment methods. It is very important to obtain risk analyses of the patient's death risk or recurrence risk before starting personalized treatment methods. These analyses can be used as a tool to help the clinician make the right decision. Therefore, in order to determine the risk analysis, it is necessary to create the right models first.[1-7]

Studies on survival methods are constantly being updated. However, when the literature is examined, it is seen that the analyses for many cancer types are performed directly by Cox analysis methods without examining the distribution of survival rate, duration and characteristics of the patients. Cox regression analysis is the most widely used method for determining risk factors affecting survival time. But in Cox regression model, the proportional hazards assumption must be provided. The main assumption of the Cox regression method is that patients will definitely encounter the event of interest (death, metastasis or recurrence) from the study result. However, this condition cannot be provided in some situations. With the help of the current technologies, treatment conditions and the knowledge of clinicians, there have been many developments in the field of medicine. Considering today with past, with the improving technologies, knowledge and experiences with most of the cancer associated deaths became preventable and be delayed. Therefore, presence of many individuals who did not experience the event in recent years (death, recurrence, etc.) on cancer studies may be mentioned. However, here, the term recovery is not the clinical improvement of the patient, but the survival of the patient after being followed for a long period of time. The Cox regression analysis is based on the assumption that there has to be higher numbers of individuals who experienced the studied incidence, so the use of this model in cases with high numbers of patients may cause misleading results. The cure models are used as an alternative method in this case.

In conclusion, it is very important to select the appropriate method by examining the survival distribution of the data. In order to obtain unbiased results, appropriate survival methods should be selected depending on the characteristics of the data. The aim of this study is to analyze the data of breast cancer patients with a mixed cure model.

COX MIXED CURE MODEL (PROPORTIONAL HAZARDS MIXED CURE MODEL)

The developments in cancer therapies resulted with the development of cure models among the statistical research. The analysis of the most of the survival data is based on the overall survival or progression-free survival. However, most patients may not have experienced the investigated event in the long term. Therefore, the cure model is a method that may be used for investigating the heterogeneity between the cured and uncured patients in the long term. A simple way of identifying whether a specific data set will have a subset of survivors in the long term is to examine the survival curve. The cure model may be an appropriate and useful way to analyse the data if the survival curve has a stability after a specific point at the end of the study. Cure in the clinical sense, is the complete disappearance of the disease, and the symptoms. However, cure in statistical sense is enabled with the equivalency of the mortality rate to the expected mortality rate.[8-11]

Cure models generate the survival modellings considering the probability of cure when there are individuals who will not encounter the resultant incident in the data set. The mixed cure model was first recommended by Boag and Gage. Both the cure ratio of the individuals from disease, and the survival ratio of patients who did not cured could be estimated simultaneously in the model they developed.[8,9,12]

In the model, T is the failure time of the event, x and z were the observed covariants, 1-π(z) was the possibility of the cured patients associated to z, and S(t|x) is the possibility of the uncured patients associated to x, and the complex cured ratio is expressed as follows.[12]

$$S_{pop}(t|x,z) = \pi(z)S(t|x) + 1 - \pi(z) \quad (1)$$

If the proportional hazard model is used in S(t|x) estimation, the mixed cure model is described as the proportional hazards mixed model, and, if the accelerated failure model is used in S(t|x) estimation, the model is described as the accelerated failure mixed cure model.

The advantage of the mixed cure model is that it enables the separate modelling of the cure distributions of both the cured and uncured individuals. Generally using the logit function, the effects of the z covariates is modelled with

$$\pi(z) = \frac{\exp(bz)}{1+\exp(bz)} \quad (2)$$

The b in the formula is a vector of unknown parameters.[12]

The effects of z covariates is modelled with log(-log(1-π(z)))=bz by using log-log. When probit function is used, it is modelled as $\Phi^{-1}(\pi(z)) = bz$. The terms Φ(.) is the cumulative probability function of the standard normal distribution.

$O = (t_i, \delta_i, z_i, x_i)$ equivalence is the observation data of ith individual; $t_i$expresses the observed survival time, δ_i=1 is the absence of censor, $\delta_i = 0$ is the presence of censor, and, z_i, x_idescribe the covariates. When $\Theta(b, \beta, S_0(t)$are identified as the unknown parameters, the parameter values in the model is estimated with the help of EM algorithm. Y=1 describing the condition when the result occurred, Y=0 describing the condition when the result did not occur, the complete probability function is expressed as follows:[12]

$$\prod_{i=1}^{n}[1-\pi(z_i)]^{1-y_i}\pi(z_i)^{y_i}h(t_i|Y=1,x_i)^{\delta_i y_i}S(t_i|Y=1,x_i)^{y_i} \quad (3)$$

h(.), is the hazard function corresponding to S(.). The complete probability function logarithm may be written as: $I_c(b,\beta;0,y) = I_{c1}(b;0,y) + I_{c2}(\beta;0,y)$. Here it is described as:[12]

$$I_{c1}(b;0,y)=\sum_{i=1}^{n} y_i \log[\pi(z_i)] + (1-y_i)\log[1-\pi(z_i)]$$
$$I_{c2}(\beta;0,y)=\sum_{i=1}^{n} y_i\delta_i \log[h(t_i|Y=1,x_i)] + y_i \log[S(t_i|Y=1,x_i)] \quad (4)$$

The conditioned expected value of the complete probability is calculated according to$y_i's$ in the E step of the EM algorithm. $\Theta^{(m)} = (b^{(m)}, \beta^{(m)}, {S_0}^{(m)}(t))$being the present estimation, the expected value of $E(y_i|O,\Theta^{(m)})$ is described as,

$$w_i^{(m)} = E(y_i|O,\Theta^{(m)}) = \delta_i + (1-\delta_i)\frac{\pi(z_i)S(t_i|Y=1,x_i)}{1-\pi(z_i)+\pi(z_i)S(t_i|Y=1,x_i)}\Big|_{(0,\Theta^{(m)})} \quad (5)$$

If δ_i=1, $w_i^{(m)} = 1$ andδ_i=0$w_i^{(m)}$ will be the ratio of the uncured patients.[10-12] The second part of $E(y_i|O, \Theta^{(m)})$, i. is the conditioned probability of the patient being uncured. Because $\delta_i log w_i^{(m)}$= 0 and$\delta_i w_i^{(m)}$ is $= \delta_i$, is defined as:[12]

$$E(l_{c1}) = \sum_{i=1}^{n} w_i^{(m)} \log [\pi(z_i)] + (1 - w_i^{(m)}) \log [1 - \pi(z_i)]$$

$$E(l_{c2}) = \sum_{i=1}^{n} \delta_i \log [w_i^{(m)} h(t_i|Y = 1, x_i)] + w_i^{(m)} \log [S(t_i|Y = 1, x_i)] \quad (6)$$

The maximization of the unknown parameters is performed on M algorithm.

MATERIAL AND METHODS

Breast cancer data: In this study, 170 breast cancer patients from the Medical Oncology Clinic of Antakya State Hospital were collected. The death from the disease was defined as an end point. The study was approved by the Ethics Committee of Mustafa Kemal University Medical Faculty (2018/90) (decision number: 02; date: 24.05.2018). The following information is available on these patients: estrogen receptor (ER) status, progesterone receptor (PR) status, human epidermal growth factor receptor 2 (HER2) status (HER2 positive is also calledc-erbB-2-positive), age at the time of diagnosis, number of lymph nodes, tumor size, grade, status of receiving neoadjuvant chemotherapy and presence of metastasis.

We used data from breast cancer patients. Breast cancer is a complex disorder with heterogeneity, the mechanisms that cause this disease are not fully explained to date. Because of this heterogeneity, the natural course of breast cancer varies among patients. With advancing technology and treatment options, significant improvements have been made in the survival rates and recovery rates of breast cancer patients. In this case, the data analysis of breast cancer patients should be different from the classical analysis of life we know.

In the study, ER, PR, c-erbB-2 and age variables (50 years and over) were taken as risk factors that may affect the life span.

The Cox regression model used in classical life analysis has the assumption that every patient in the will experience the event of interest but this assumption is not always true because in some types of cancer, for example breast cancer, many cancer patients recover after treatment. In this case, the Cox regression model is not preferred since the survival of many patients followed up at the end of the study is unknown. In such cases, it is more appropriate to use the mixed cure model. So we use mixed cure model to breast cancer data. In the analysis, SPSS (version 21.0) and R (version 3.3) programs were used. Nominal alpha value is taken as 0.05.

In the cure model, for the cured part, coefficients of variation and levels can be used in order to obtain the cured rates. When calculating the ratio of the cured part, the link function used in the parameter estimation is taken into consideration and the calculations are made with a probit function. The survival curve of the patients was obtained by the Kaplan-Meier method.

RESULTS

The censoring rate was observed in 73.5% (125 people). The survival time was taken as months until the death of the patients. The median duration of follow-up is 72 months. The median survival time of the patients was found to be 72 months (95% confidence interval; 51.88-92.12). As can be seen from Figure 1, it was determined that the survival times of the patients were constant after a certain period of time. In this

case, it shows that not all patients encountered the event of interest, and some patients recovered. In addition, again, as expressed in the graph, while significant failures were observed in the first 72 months, there was no significant change in survival rates after the 72nd month. According to these findings, it can be stated that patients have a high probability of long-term survival.

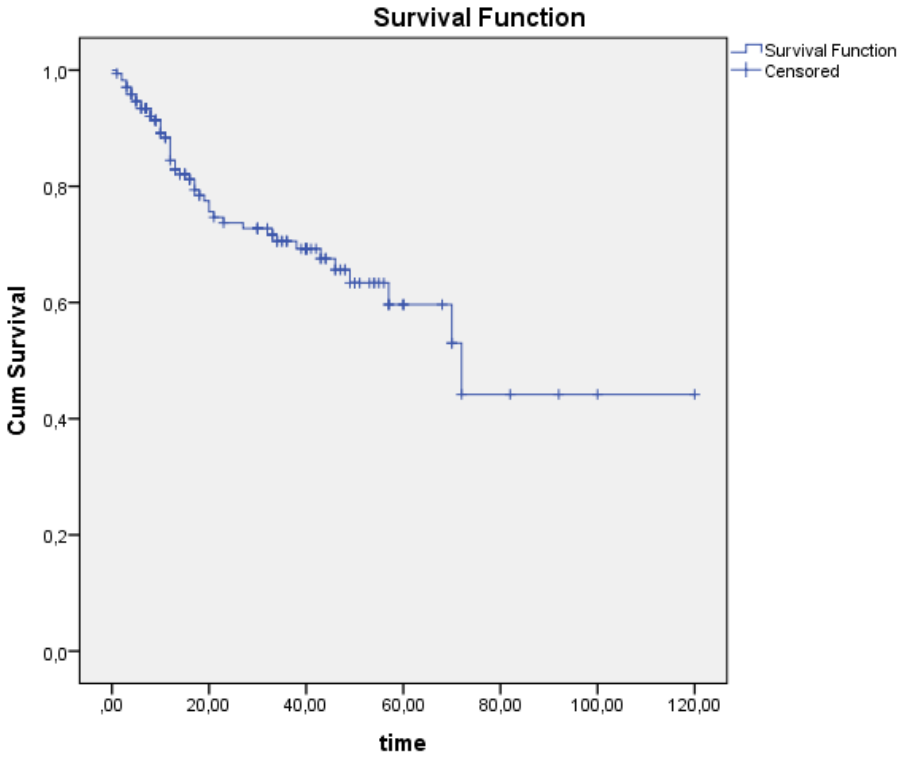

FIGURE 1: Kaplan-Meier curve.

Whether the proportional hazard assumption was met was investigated using the correlation test between the Schoenfeld residues and the rank of lifetimes. Our null hypothesis for test statistics is H_0: "proportional hazard assumption is satisfied". We concluded that the proportional hazard assumption was not provided for prognostic factors (p=0.007). Therefore, the proportional hazard assumption could not be provided for cancer patients data. The analysis of these data was done by cure mixed model analysis because both the proportional hazard assumption could not be satisfied and the survival rates remained constant after a certain period of time.

The results obtained as a result of applying the mixed cure model to the data set are shown in Table 1. In the cure model, the cured part and the uncured part are modeled separately. This feature of mixed models allows the effect of explanatory variables on the cured and uncured patients to be different When the p values for explanatory variables were examined, it was seen that PR, ER and c-erbB-2 were a significant risk factor on the survival of breast cancer patients (p=0.022; p=0.007; p<0.001). Also age and lymph node had a significant effect on survival rate of breast cancer patients in the cured part. For uncured part, it was ssen that PR, ER and c-erbB-2, age and lymph node had a significant effect on survival rate of breast cancer patients (respectively p=0.038, p=0.009, p=0.028, p=0.007, p=0.004). According to the results obtained from the mixed cure model analysis, the risk of failure of the PR negative patient is 1.7 times higher than the positive patient and ER negative patient is 2 times higher than the positive patient.

TABLE 1: Cure model results.

Proportional hazards mixed cure model for the cured part	$\widehat{\beta}$	Standard error	p value
ER	0.68	0.22	0.007
PR	-0.29	0.23	0.022
c-erbB-2	-0.91	0.24	p<0.001
Age	0.28	0.17	0.001
Lymph node	0.21	0.18	0.038
Grade	-0.98	0.76	0.125
Tumor size	0.02	0.01	0.062
Adjuvant	0.89	0.29	0.439
Proportional hazards mixed cure model for the uncured part	$\widehat{\beta}$	Standard error	p value
ER	1.37	0.66	0.038
PR	1.43	0.54	0.009
c-erbB-2	-1.79	0.81	0.028
Age	0.04	0.01	0.007
Lymph node	0.12	0.04	0.004
Grade	-0.75	0.40	0.064
Tumor size	0.04	0.01	0.092
Adjuvant	0.63	0.53	0.255

ER: Estrogen receptor; PR: Progesterone receptor.

By using variable coefficients and levels for the cured part, the desired cured rates can be obtained. While calculating the ratio of the cured part, the link function used in the parameter estimation is taken into consideration and calculations are made accordingly. According to these calculations, the recovery rate of breast cancer patients when younger than 50 years old, when hormone receptors (ER and PR) are negative and c-erbB-2 positive, is 60%; it has been observed that the recovery rate of breast cancer patients older than 50 years of age, who are hormone receptor (ER and PR) negative and c-erbB-2 positive, has decreased to 40%.

DISCUSSION

In this study, we suggested the use of a mixed cure model when the proportional hazard assumption is not satisfied in breast cancer patient data and the survival curve is stable after a certain period of time. When the literature is examined, when the survival times of the patients are constant after a certain period of time, many studies recommend cure model analysis in survival analysis.[13-15] Similarly, Rama et al. explained in a cancer data application that when proportional hazard assumptions are not satisfied, as in our study, the use of mixed models would be appropriate.[16]

CONCLUSION

The major advantage of mixed cure models is that, they can be modeled separately for the cured and uncured part, thereby allowing the interpretation of the effects of the explanatory variables on the cured and uncured patients separately. Nowadays, cancer patients are more likely to recover. For this reason, we think that it would be appropriate to evaluate the survival analysis of cured and uncured cancer patients separately with the mixed cure model.

Source of Finance

During this study, no financial or spiritual support was received neither from any pharmaceutical company that has a direct connection with the research subject, nor from a company that provides or produces medical instruments and materials which may negatively affect the evaluation process of this study.

Conflict of Interest

No conflicts of interest between the authors and/or family members of the scientific and medical committee members or members of the potential conflicts of interest, counseling, expertise, working conditions, share holding and similar situations in any firm.

Authorship Contributions

Idea/Concept: *Şirin Çetin;* ***Design:*** *Şirin Çetin;* ***Control/Supervision:*** *Şirin Çetin;* ***Data Collection and/or Processing:*** *İsa Dede;* ***Analysis and/or Interpretation:*** *Özge Pasin;* ***Literature Review:*** *Özge Pasin;* ***Writing The Article:*** *Şirin Çetin, Özge Pasin;* ***Critical Review:*** *Ayşe Ülgen.*

REFERENCES

1. Turashvili G, Brogi E. Tumor heterogeneity in breast cancer. Front Med (Lausanne). 2017;4:227. [Crossref] [PubMed] [PMC]
2. Leong SP, Shen ZZ, Liu TJ, Agarwal G, Tajima T, Paik NS, et al. Is breast cancer the same disease in Asian and Western countries? World J Surg. 2010;34(10):2308-24. [Crossref] [PubMed] [PMC]
3. Youlden DR, Cramb SM, Dunn NA, Muller JM, Pyke CM, Baade PD. The descriptive epidemiology of female breast cancer: an international comparison of screening, incidence, survival and mortality. Cancer Epidemiol. 2012;36(3):237-48. [Crossref] [PubMed]
4. Li CI, Malone KE, Daling JR. Differences in breast cancer stage, treatment, and survival by race and ethnicity. Arch Intern Med. 2003;163(1):49-56. [Crossref] [PubMed]
5. Chlebowski RT, Chen Z, Anderson GL, Rohan T, Aragaki A, Lane D, et al. Ethnicity and breast cancer: factors influencing differences in incidence and outcome. J Natl Cancer Inst. 2005;97(6):439-48. [Crossref] [PubMed]
6. Tao Z, Shi A, Lu C, Song T, Zhang Z, Zhao J. Breast cancer: epidemiology and etiology. Cell Biochem Biophys. 2015;72(2):333-8. [Crossref] [PubMed]
7. Ellsworth RE, Decewicz DJ, Shriver CD, Ellsworth DL. Breast cancer in the personal genomics era. Curr Genomics. 2010;11(3):146-61. [Crossref] [PubMed] [PMC]
8. Berkson J, Gage RP. Survival curve for cancer patients following treatment. Journal of the American Statistical Association. 1952;47(259):501-15. [Crossref]
9. Boag JW. Maximum likelihood estimates of the proportion of patients cured by cancer therapy. Journal of the Royal Statistical Society Series B (Methodological). 1949;11(1):15-53. [Crossref]
10. Kara P, Ata Tutkun N. Cox karma iyileşme modeli ve glioma veri kümesi üzerine bir uygulama [Cox mixture cure model and its application to glioma data set]. Turkiye Klinikleri J Biostat. 2017;9(3):241-56. [Crossref]
11. Othus M, Barlogie B, Leblanc ML, Crowley JJ. Cure models as a useful statistical tool for analyzing survival. Clin Cancer Res. 2012;18(14):3731-6. [Crossref] [PubMed] [PMC]
12. Cai C, Zou Y, Peng Y, Zhang J. smcure: an R-package for estimating semiparametric mixture cure models. Comput Methods Programs Biomed. 2012;108(3):1255-60. [Crossref] [PubMed] [PMC]
13. Achcar JA, Coelho-Barros EA, Mazucheli J. Cure fraction models using mixture and non-mixture models. Tatra Mt Math Publ. 2012;51(1):1-9. [Crossref]
14. Akhlaghi AA, Najafi I, Mahmoodi M, Shojaee A, Yousefifard M, Hosseini M. Survival analysis of Iranian patients undergoing continuous ambulatory peritoneal dialysis using cure model. J Res Health Sci. 2013;13(1):32-6. [PubMed]
15. Yu XQ, De Angelis R, Andersson TM, Lambert PC, O'Connell DL, Dickman PW. Estimating the proportion cured of cancer: some practical advice for users. Cancer Epidemiol. 2013;37(6):836-42. [Crossref] [PubMed]
16. Rama R, Swaminathan R, Venkatesan P. Cure models for estimating hospital-based breast cancer survival. Asian Pac J Cancer Prev. 2010;11(2):387-91. [PubMed]