Far East Journal of Mathematical Sciences (FJMS)
© 2017 Pushpa Publishing House, Allahabad, India
http://www.pphmj.com
http://dx.doi.org/10.17654/MS102123165
Volume 102, Number 12, 2017, Pages 3165-3176 **ISSN: 0972-0871**

DETERMINING THE SURVIVAL RATES OF BREAST CANCER PATIENT IN MIXTURE MODEL

Nurkaromah Dwidayati and Zaenuri

Department of Mathematics
Faculty of Mathematics and Natural Sciences
Universitas Negeri Semarang
Indonesia

Abstract

This paper determines the survival rates of patient with breast cancer using mixture model, and estimates the proportions of survival probabilities of uncured patients until pre-determined times. Baseline survival function could not be fully eliminated on EM (expectation maximization) algorithm. For estimating the function of baseline survival, the assumption of PH (proportional hazard) model is employed, in line with Cox's PH model. Based on baseline survival functions, survival rates at pre-determined times in accordance with the existing certain characteristics of the patients could then be obtained.

1. Introduction

The mixture model was based on the function of probability distributions (both discretional and continuing distributions in the forms of mixtures):

$$f(\cdot) = \sum_{i=1}^{n} p_i f_i(\cdot), \tag{1}$$

Received: July 2, 2017; Accepted: September 13, 2017
2010 Mathematics Subject Classification: 62P10.
Keywords and phrases: survival rate, breast cancer, mixture model.

where $f(\cdot)$ constitutes the density function describing the mixture elements, p_i constitutes the mixture weight with $p > 0, \sum_{i=1}^{n} p_i = 1$ and n constitutes the numbers of the elements contained in the mixture.

The mixture model is called *parametric cure mixture model* when it employs standard probability distributions (as Weibull, Gompertz exponential distributions) and generalized *F*. Discussions on the parametric mixture model can be found in [1-13].

Mixture model not using standard probability distributions is called *non-parametric cure mixture model* used for estimating the cure rates, as introduced in [14], by applying the assumption of PH model for the failure time distributions of non-cured patients. The employed method was similar to the one of Cox's PH model, namely non-parametric model. However, the results from the two similar methods could not be separated from Monte Carlo approach which used the concept of likelihood function. The reference [15] used Kaplan-Meier's survival estimators (for estimating failure time distributions for uncured patient) and EM algorithm, but the model could not produce covariates for uncured patients.

Chance to be cured, which usually is known as cure rate or surviving fraction, is defined as asymptotic value of the survival function for unlimited time to come, written as $\lim_{t\to\infty} S(t)$ [16]. For this example, t states the observed survival time period. Statistical inferences on cure rates are based on any given survival rate functions $\{S(t) = P(T \geq t)\}$ and can be written in as:

$$S(t) = a + (1 - a)S_0(t), \tag{2}$$

where $a = P(T = \infty)$ or the chance to be cured, and $S_0(t) = P(T \geq t \mid X < \infty)$.

The likelihood function for the mixture model is given by

$$L = \prod_{i=1}^{n} | \pi(z_i) f_u(t_i | x_i) |^{\delta_i} [\pi(z_i) S_u(t_i | x_i) + 1 - \pi(z_i)]^{1-\delta_i}. \qquad (3)$$

For estimating the unknown parameters in the mixture model we used EM algorithm, which consists of *E*-step and *M*-step. *E*-step calculates the function of expectancy for the likelihood logs, in which the purposes were to estimate the density function, the survival function and the proportions of uncured patients. *M*-step concerns with the maximization of the likelihood functions, which relates to the estimations of the density function, the survival function and the proportion of the uncured patients.

EM algorithm constitutes an iterated approach for studying the model of data (data which have some lost values) through 4 steps [17]: (1) select an initial association for the parameters of the model, (2) determine the expected values of the lost data, (3) write the inducted parameters for a new model based on the combination between the expected and the original values in the data, and (4) when the parameters are not convergent, repeat step (2) by using a new model.

2. Methods

This research was designed using deductive-analytic approach. The population for this research consisted of patient who sought help from RSUP Dr Sardjito Yogyakarta (RSUP Sardjito) for breast cancer. The sample for this research consisted of breast cancer patients (BCP), taken from the same population based on a five-year survival.

The data for this research were secondary data collected from clinical studies and medical records of the BCP treated in the Hospital. The medical records consisted of registered numbers, the patient’ name, dates of births, dates of first visits, dates of check outs or dates of last re-visits, disease stadiums, types of treatments and post-treatment health statuses. The data were analyzed deductively based on reviews on the results of previous studies, clinical definitions, clinical assumptions and clinical theorems.

3. Results and Discussions

The data on the BCP based on respective ages, survivals, clinical stadiums and received treatments are illustrated in a histogram as shown in Figure 1.

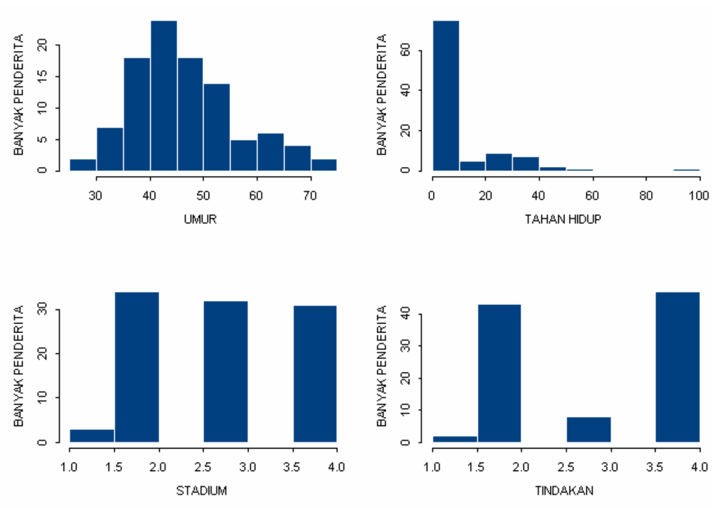


Figure 1. Histogram BCP RSUP Sardjito.

Based on the data shown in Figure 1, we can figure out the ages of the BCP seeking for treatments in RSUP Sardjito, as can be seen in Table 1.

Table 1. Statistical values of the ages of BCPs RSUP Sardjito

Statistical values	Value (years of ages)
Means	46.79
Standards of deviations	9.83
Minimum	27
Maximum	74

From Table 1, it can be seen that the average age of BCPs treated in RSUP Sardjito was 46.79. This is in accordance with the finding of previous review that in Indonesia, the highest incidence of breast cancer was found in females of productive ages (40-49 years old) [18]. As comparisons, the average age of BCPs in Jakarta was 46 years old while in Surabaya was 47. The life times of BCPs treated in RSUP Sardjito can be seen in Table 2.

Table 2. Life times of BCPs RSUP Sardjito

Life times	Value (in months)
Mean	9.21
Standard of deviation	15.63
Minimum	0.03
Maximum	48.5

From Table 2, it can be seen that the average life time of BCPs treated in RSUP Sardjito was 9.21 months (less than a year). This low average of life time was caused by the BCP’s age and stadium of the disease at the time when the individual first came to RSUD for seeking medical help (as illustrated in the histogram in Figure 1). The illustration of the life times of BCPs based on ages is shown in Figure 2.

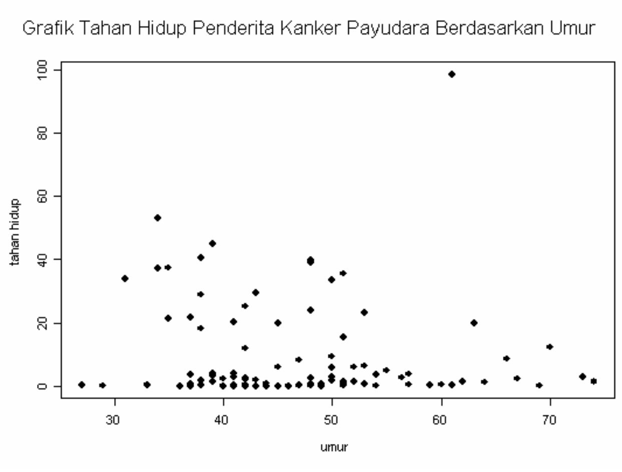


Figure 2. Life time BCP based ages.

From Figure 2, it can be seen that the higher the age of the BCP, the lower would her life time be. The life times of BCPs treated in RSUP Sardjito were assumed as distributed in Weibull fashion with θ and β parameters. This assumption is supported by time interval in probability plots as shown below (Figure 3).

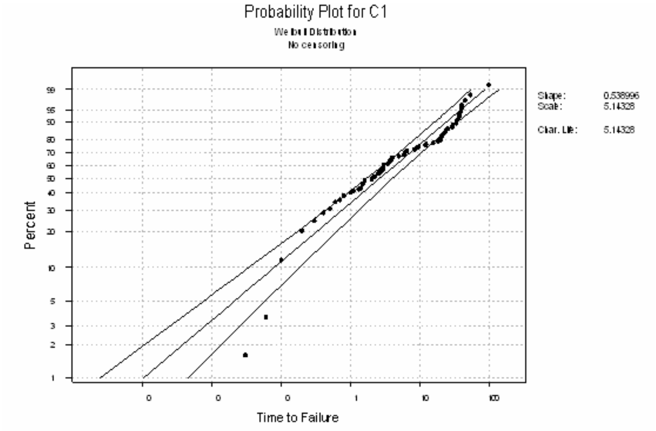


Figure 3. Probability plot.

From the above assumption we can then construct a plot for the hazard function (Figure 4) and another one for the survival function (Figure 5).

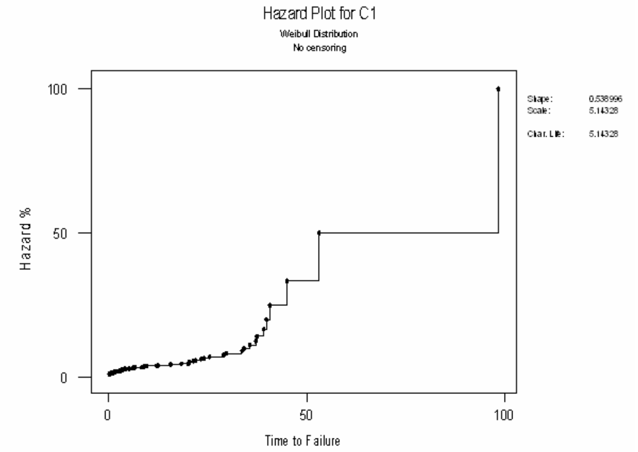


Figure 4. Hazard function.

Hazard function refers to conditional failure rate, which is the probability of mortality for a very short time interval, with an assumption that the individual would remain alive during the earliest time interval. In other words, it also means a limited probability stating that a BCP would die in a very short time interval of $[t, t + \Delta t]$ if it was known that the individual would remain alive until the time of t. This probability constitutes an inclination to die as a function of the individual's age, meaning that $h(t)$ is the proportion of the individual's ages during which she would die in a time interval of $[t, t + \Delta t]$. Hazard function implies a mortality risk per unit of time during the aging process.

In the cases of BCPs treated in RSUP Sardjito, based on Figure 4, it can be seen that the bigger the life time of *t* (the individual's age keeps increasing) the higher would the value of the hazard function be. It shows that the older BCPs have higher risk to die than the younger ones.

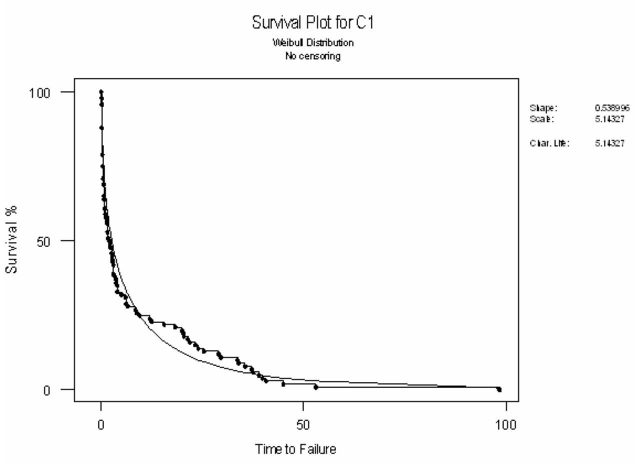


Figure 5. Survival function.

Survival function implies the chance of BCP to remain alive until the time period beyond the time of *t*. From the graph shown above, it can be seen that higher the time *t*, lower would the chance of the individual to remain alive beyond time *t*.

The illustrations of our assumptions on life time distributions for the BCPs treated in RSUP Sardjito is shown in Figure 6.

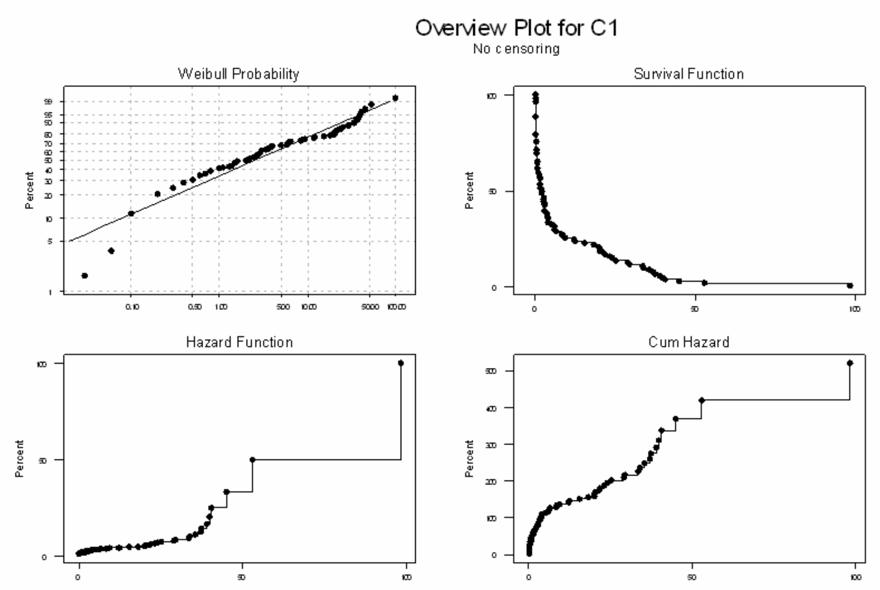


Figure 6. Life time distributions.

Most of the BCPs treated in RSUP Sardjito (64%) had high stadiums (stadium III and stadium IV). Those who came to RSUP Sardjito in early stadiums (stadium I and stadium II) were only 36%. For more detailed illustrations, see Table 3.

Table 3. Stadium of BCPs RSUP Sardjito

Stadium	Percentage (%)
I	3
II	33
III	32
IV	32

Relating to the disease stadiums of the BCPs seeking for medical help in RSUP Sardjito, the treatments were given in four categories: (1) medication in low dosages, (2) mastectomy, (3) radiation and (4) chemotherapy. From the histogram in Figure 1, it can be seen that the BCPs were mostly treated by means of chemotherapy and mastectomy. For more detailed information, see Table 4.

Table 4. Medical treatments on BCPs RSUP Sardjito

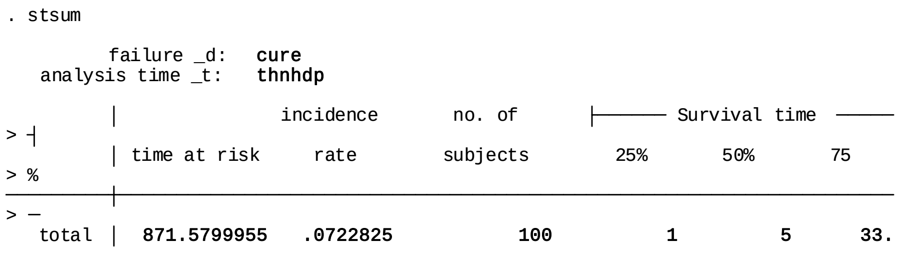
Treatments	Percentage (%)
Low dosage medications	2
Mastectomy	43
Radiation	8
Chemoterapy	47

From Table 4, it can be seen that the treatments given to the BCPs are mostly chemotherapy followed by mastectomy. This is in line with the facts that 64 BCPs seeking for medical treatments in RSUP Sardjito were in high stadiums (III and IV) and 33 were in stadium II. More detailed information on the statuses of BCPs treated in RSUP Sardjito can be seen in Table 5.

Table 5. Cure status on BCPs RSUP Sardjito PKPD

Cure status	Percentage (%)
Cure (1)	63
Uncured (0)	37

From Table 5, it can be seen that the BCPs who either died or just lost from the records at the end of this research were 37% while those who were cured or alive were 63%. The median survival time can be seen in Figure 7.



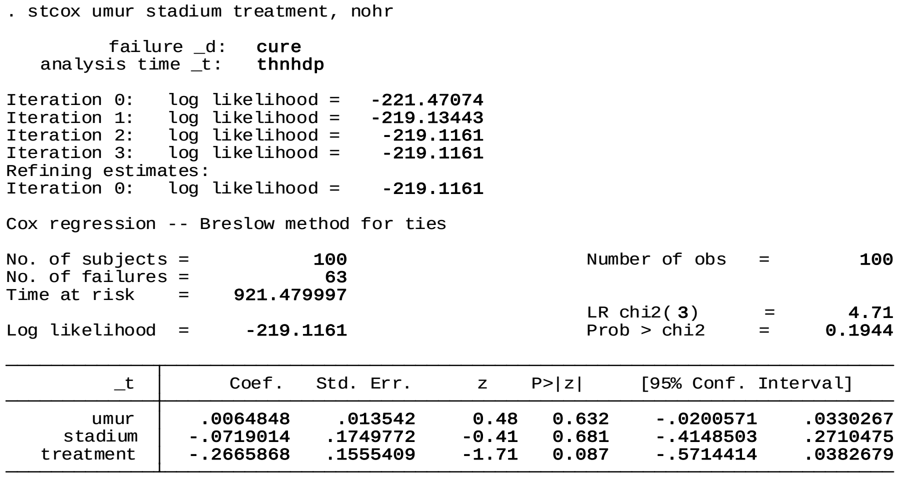
```
. stsum

         failure _d:  cure
   analysis time _t:  thnhdp

         |               incidence       no. of      ├────── Survival time  ─────
> ┤
         | time at risk     rate        subjects        25%       50%       75
> %
─────────┼──────────────────────────────────────────────────────────────────────
> ─
   total |  871.5799955   .0722825          100          1         5       33.
```

Figure 7. Survival time median.

The central inclination illustrated in Figure 7 is the median but not average due to the fact that with at least one individual having too short or to long (or both) life times, the average survival time would not be proportional (would be either too big or too small or both). Based on previous analyses, ages and disease stadiums of the BCPs and treatments influence the life times. The relationships among these variables are expressed in Cox's regression model for PH shown in Figure 8.



```
. stcox umur stadium treatment, nohr

         failure _d:  cure
   analysis time _t:  thnhdp

Iteration 0:   log likelihood =  -221.47074
Iteration 1:   log likelihood =  -219.13443
Iteration 2:   log likelihood =   -219.1161
Iteration 3:   log likelihood =   -219.1161
Refining estimates:
Iteration 0:   log likelihood =   -219.1161

Cox regression -- Breslow method for ties

No. of subjects =          100                     Number of obs   =       100
No. of failures =           63
Time at risk    =   921.479997
                                                   LR chi2(3)      =      4.71
Log likelihood  =    -219.1161                     Prob > chi2     =    0.1944

------------------------------------------------------------------------------
          _t |      Coef.   Std. Err.      z    P>|z|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
        umur |   .0064848    .013542     0.48   0.632    -.0200571    .0330267
     stadium |  -.0719014   .1749772    -0.41   0.681    -.4148503    .2710475
   treatment |  -.2665868   .1555409    -1.71   0.087    -.5714414    .0382679
------------------------------------------------------------------------------
```

Figure 8. Relation inter variable.

Based on the above analyses, a regression model for Cox's PH can be formulated as follows:

$$h(t) = h_0(t)\exp(0{,}0064848x_1 - 0{,}0719014x_2 - 0{,}2665868x_3),$$

where $x_1 :=$ age of the BCP, $x_2 :=$ disease stadium; $x_3 :=$ type of treatment

The function for survival rate can be written as follows:

$$S(t) = S_0(t)^{\exp(0{,}0064848x_1 - 0{,}0719014x_2 - 0{,}2665868x_3)}.$$

Examples. Assume that a 43-year old female individual gets breast cancer of stadium III and received chemotherapy (=4) as the necessary medical treatment for her. After treatment for 12 months, the survival rate and the hazard rate for this BCP are calculated. Based on baseline survival and baseline hazard rates, the survival rate of this BCP comes out to be

$$S_0(12) = 0{,}15410455 \quad \text{and} \quad h_0(12) = 1{,}8298352.$$

Thus,

$$S(12) = 0{,}15410455^{\exp(0{,}0064848*43 - 0{,}\,0719014*3 - 0{,}\,2665868*4)} = 0{,}503698,$$

and hence the chance of the BCP to survive after a 12-month treatment is 50,37%. The hazard rate can be calculated as below:

$$h(12) = 1{,}8298352 * \exp(0{,}0064848 * 43 - 0{,}0719014 * 3 - 0{,}2665868 * 4)$$

$$= 0{,}671005.$$

Hence, the chance of the BCP to die after the 12-month treatment is 67,1%.

4. Conclusions

Based on the above analyses, it can be concluded that to determine the survival rate of a BCP using the mixture model, the calculations should be based on BSF (baseline survival function). In the mixture model, the BSF could not be fully eliminated by using EM algorithm. Thus, BSF was estimated using the assumption of Cox's PH from which we can obtain the survival rates based upon pre-determined times in accordance with the given characteristics of the BCPs.

References

[1] J. W. Boag, Maximum likelihood estimates of the proportion of patients cured by cancer therapy, J. Royal Statist. Soc. 11 (1949), 15-53.

[2] D. R. Jones, R. L. Powles, D. Machin and R. J. Sylvester, On estimating the proportion of cured patients in clinical studies, Biometrie-Praximetrie 21 (1981), 1-11.

[3] V. T. Farewell, The use of mixture models for analysis of survival data with long-term survivors, Biometrics 38 (1982), 1041-1046.

[4] V. T. Farewell, Mixture models in survival analysis, are they worth the risk? Canad. J. Statist. 14 (1986), 257-262.

[5] A. B. Cantor and J. J. Shuster, Parametric versus non-parametric methods for cure rates based on censored survival data, Stat. Med. 11 (1992), 931-937.

[6] M. E. Ghitany, R. A. Maller and S. Zhou, Exponential mixture models with long-term survivor and covariates, J. Multivariate Anal. 49 (1994), 218-241.

[7] R. A. Maller and S. Zhou, Estimating the proportion of immunes in a censored sample, Biometrika 79 (1992), 731-739.

[8] J. W. Denham, E. Denham, K. B. Dear and G. V. Hudson, The follicular non- Hodgkin's Lymphomas-I. The possibility of cure, Eur. J. Cancer 32 (1996), 470-479.

[9] Y. Peng and J. W. Denham, Generalized *F* mixture model for cure rate estimation, Stat. Med. 17 (1998), 813-830.

[10] Y. Peng and K. G. B. Dear, A nonparametric mixture model for cure rate estimation, Biometric 56 (2000), 237-243.

[11] B. Muthen and K. Masyn, Discrete-time survival mixture, J. Edu. Behavioral Statist. 30 (2005), 27-58.

[12] F. Picard, An introduction to mixture models, Statistics for Systems Biology Group, Research Report No. 7, 2007.

[13] N. Dwidayati, S. H. Kartiko and Subanar, Estimation of the parameters of a mixture Weibull model for analyze cure rate, Appl. Math. Sci. 7 (2013), 5767-5778.

[14] A. Y. Kuk and C. Chen, A mixture model combining logistic regression and life model, Biometrika 79 (1992), 531-541.

[15] J. M. G. Taylor, Semi-parametric estimation in failure-time mixture models, Biometrics 51 (1995), 899-907.

[16] L. B. Klebanov and A. Y. Yakovlev, A new approach to testing for sufficient follow-up in cure-rate analysis, J. Statist. Plann. Inference 137 (2007), 3557-3569.

[17] W. Bridewell, P. Langley, S. Racunas and S. Borrett, Learning process models with missing data, Computational Leaning Laboratory, CSLI Stanford University, 2005.

[18] M. Ramli, Epidemiological Review of Breast Cancer in Indonesia, Book of Proceedings Jakarta International Cancer Conference'95, Jakarta, 1995.