RESEARCH COMMUNICATION

Cure Models for Estimating Hospital-Based Breast Cancer Survival

Ranganathan Rama[1], Rajaraman Swaminathan[1*], Perumal Venkatesan[2]

Abstract

Objective: Research on cancer survival is enriched by development and application of innovative analytical approaches in relation to standard methods. The aim of the present paper is to document the utility of a mixture model to estimate the cure fraction and compare it with other approaches. Methods: The data were for 1,107 patients with locally advanced breast cancer, who completed the neo-adjuvant treatment protocol during 1990-99 at the Cancer Institute (WIA), Chennai, India. Tumour stage, post-operative pathological node (PN) and tumour residue (TR) status were studied. Event free survival probability was estimated using the Kaplan-Meier method. Cure models under proportional and non-proportional hazard assumptions following log normal distribution for survival time were used to estimate both the cure fraction and the survival function for the uncured. Results: Event free survival at 5 and 10 years were 64.2% and 52.6% respectively and cure fraction was 47.5% for all cases together. Follow up ranged between 0-15 years and survival probabilities showed minimal changes after 7 years of follow up. TR and PN emerged as independent prognostic factors using Cox and proportional hazard (PH) cure models. Proportionality condition was violated when tumour stage was considered and it was statistically significant only under PH and not under non PH cure models. However, TR and PN continued to be independent prognostic factors after adjusting for tumour stage using the non PH cure model. A consistent ordering of cure fractions with respect to factors of PN and TR was forthcoming across tumour stages using PH and non PH cure models, but perceptible differences in survival were observed between the two. Conclusion: If PH conditions are violated, analysis using a non PH model is advocated and mixture cure models are useful in estimating the cure fraction and constructing survival curves for non-cures.

Keywords: Parametric - proportional hazard - mixture cure model - cure fraction - breast cancer

Asian Pacific J Cancer Prev, **11**, 387-391

Introduction

Research on cancer survival is enriched by development and application of innovative analytical approaches in comparison to standard methods. The widely used proportional hazard (PH) model (Cox, 1972) is based on the assumption that every individual is susceptible to the event of interest, say death due to disease studied. However, this assumption may not always be true as a large proportion of patients could be deemed as cured of the disease after sufficient follow-up. Research on many cancers (Tai et al., 2005) including breast (Tai et al., 2004), childhood lymphomas and leukaemias (Sposto, 2002) has shown that a significant proportion of patients with these cancers are cured after therapy. Hence, survival model incorporating a cure fraction is an important statistical tool for analyzing cancer survival data. Parametric cure model (PCM) was introduced by Boag in 1949 and later developed by Berkson and Gage in 1950. The objectives of the present paper are to demonstrate the utility of the PCM in eliciting the prognostic factors for time to event breast cancer data in relation to standard methods, to estimate the cure fraction and to construct the survival curve for the uncured using mixture models.

Materials and Methods

A consecutive series of 1,107 locally advanced breast cancer (LABC) patients who had completed the neo-adjuvant treatment protocol consisting of preoperative chemoradiotherapy followed by surgery between 1990 and 1999 at Cancer Institute (WIA), Chennai, formed the study material. The variables analyzed were tumour stage (3 levels with increasing severity: Stages 2B, 3A and 3B), postoperative pathology of pathologic tumour residue (TR: negative or positive) and pathologic nodal status (PN: negative or positive). These prognostic factors are described in detail elsewhere (Shanta et al., 2008). Event free survival (EFS) duration was defined as the minimum time elapsed to disease progression, disease recurrence, occurrence of second malignant neoplasm or death from any cause. Patients alive without disease were censored at

*[1]Department of Epidemiology, Biostatistics and Cancer Registry, Cancer Institute (WIA), [2]Department of Biostatistics, Tuberculosis Research Centre, India *For Correspondence: iarcsurvival@yahoo.co.uk*

the date of last follow-up. EFS probability was estimated using Kaplan-Meier method (Kaplan and Meier, 1958).

Statistical description of cure models

Cure models estimate both the cure fraction and the survival function for the uncured. The two most commonly used cure models are the mixture model and the non-mixture model. In mixture cure model, the population is divided into two sub-populations so that an individual is either cured with proportion π or uncured with proportion $(1 - \pi)$, will experience the event, in the absence of censoring. Let the distribution function of the uncured be $F(t)$. The mixture cure survival model can be written as

$$S(t) = \pi + (1 - \pi)(1 - F(t)) \quad (1)$$

Where F(0) and F(∞) = 1, so that S(0) = 1 and S(∞) = π. The hazard function for this model is

$$h(t) = \frac{(1 - \pi) f(t)}{S(t)} \quad (2)$$

Where $f(t)$ is the density corresponding to $F(t)$.

The non-mixture cure model defines an asymptote for the cumulative hazard and hence for the cure fraction. The non-mixture PCM takes the form

$$S(t) = \pi^{F(t)} \quad (3)$$

This model can be derived under a threshold model for tumour recurrence, where $F(t)$ refers to the distribution of division times of each cell in a homogenous clone of cells (Sposto, 2002). The hazard function for the non-mixture PCM is

$$h(t) = - \ln(\pi) f(t) \quad (4)$$

When $F(t)$ does not depend on covariates, non-mixture PCM satisfies the PH assumption. Hence, results using these models are analogous to using Cox PH models.

The function $F(t)$ takes the form of parametric or semi-parametric distributions, with support over (0, ∞). The choice of the parametric models for our study is lognormal distribution, whose distribution is given by

$$F(t) = \Phi(\ln[\lambda t])^{\gamma} \quad (5)$$

where Φ(.) is the standard normal distribution function with the scale parameter λ and the shape parameter γ.

The cure fraction π may vary by covariates and this dependence can be modelled with advantage using different link functions in different situations. If the covariates matrix is X, then some of the important link functions are as follows:

1. Identity link, $\pi = \beta' X$: Here, the covariate effects are in units of the cure fraction.

2. Logistic link, $\log(\pi_i/(1 - \pi_i) = \beta' X$: Here, covariate effects are expressed as (log) odds ratios, and interpreted in a similar way as in logistic regression.

3. Complementary log-log link, $\log(- \log(\pi_i) = \beta' X$: It is useful for the non-mixture model as the covariate effects expressed are analogous to those obtained from standard Cox regression analysis and are exactly analogous to the non-mixture model under PH.

Covariates can also influence $F(t)$ through the scale and shape parameters. In non-mixture PCM, if either λ or γ depend on covariates then the PH assumption does not apply to these covariates. This fact provides a natural way to obtain a likelihood ratio test of the PH assumption that are analogous to Cox regression based tests of covariate effects. Analysis of prognostic factors and estimate of cure fraction were carried out by non-mixture cure models using Stata version 10 software was used for analysis (Lambert, 2007).

Results

Table 1 describes the number, proportion and event free survival estimates for factors of pathologic node status, tumour residue and tumour stage categories of breast cancer cases. Five and ten-year event free survival of all stages together were 64.2% and 52.6% respectively. The proportion of cases by tumour stage ranged between 31-35% and a decreasing survival with increasing stage was forthcoming. Tumour residue negative cases constituted 45% while those positive accounted for 55% and their survival at 10 years were 63.3% and 42.8% respectively. Post-operative node negative cases comprised 59% and those positive were 41% with ten-year event free survival of 64.1% and 35.1% respectively.

The follow up duration ranged between 0 and 15 years with a median follow-up duration was 82 months among those without experiencing any event and 27 months among those experiencing any event. The number of events was the maximum in the second year and decreased gradually. Event free survival probabilities for all cases together showed minimal changes after 7 years of follow up (Figure 1): survival percentages were 64.2, 55.3, 53.7 and 52.6 for the years 5, 8, 9 and10 respectively. The same

Table 1. Number of Cases, Proportion (%) and Survival (%) of Factors of Tumour Stage, Tumour Residue (TR) and Pathologic Node (PN) Status of Breast Cancer, Cancer Institute (WIA), Chennai, India, 1990-99

Tumour Stage	Tumour Residue	Pathologic Node	No.	%	Survival % 5 years	Survival % 10 years
Stage 2B	TR-	PN-	162	14.6	78.0	70.0
		PN+	32	2.9	58.0	49.0
	TR+	PN-	96	8.7	81.3	60.8
		PN+	76	6.9	61.0	46.8
Stage 3A	TR-	PN-	122	11.0	84.1	78.8
		PN+	57	5.1	46.5	38.6
	TR+	PN-	98	8.9	67.0	57.5
		PN+	107	9.7	50.3	28.6
Stage 3B	TR-	PN-	89	8.0	64.5	57.4
		PN+	38	3.4	50.1	45.1
	TR+	PN-	85	7.7	63.0	47.4
		PN+	145	13.1	42.0	28.7
All Stages			1107	100.0	64.2	52.6

Table 2. Comparison of Regression Coefficients Obtained through Multifactorial Analysis of Prognostic Factors by using PH and Non PH Models, Breast Cancers Treated at Cancer Institute (WIA), Chennai, India, 1990-99

Model	Coef.	SE	p value	Coef.	SE	p value
	Model 1			Model 2		
Cox model						
Tumour Residue	0.223	0.103	0.031	0.193	0.104	0.063
Path. Node	0.783	0.101	<0.001	0.730	0.102	<0.001
Stage 3A				0.213	0.124	0.086
Stage 3B				0.482	0.121	<0.001
Cure PH model						
Cure fraction						
Tumour Residue	0.217	0.104	0.036	0.186	0.104	0.073
Path. Node	0.790	0.101	<0.001	0.736	0.102	<0.001
Stage 3A				0.221	0.124	0.076
Stage 3B				0.489	0.121	<0.001
Intercept	-0.733	0.111	<0.001	-0.933	0.132	<0.001
Scale						
Intercept	3.888	0.116	<0.001	3.893	0.116	<0.001
Shape						
Intercept	-0.018	0.062	0.778	-0.019	0.062	0.759
Cure Non-PH model						
Cure fraction						
Tumour Residue	0.462	0.170	0.007	0.344	0.160	0.031
Path. Node	0.531	0.185	0.004	0.625	0.204	0.002
Stage 3A				-0.536	0.655	0.413
Stage 3B				-0.352	0.615	0.568
Intercept	-0.731	0.160	<0.001	-0.295	0.595	0.620
Scale parameter						
Tumour Residue	0.393	0.210	0.062	0.242	0.215	0.260
Path. Node	-0.396	0.244	0.105	-0.178	0.278	0.521
Stage 3A				-1.125	0.954	0.238
Stage 3B				-1.264	0.884	0.153
Intercept	3.871	0.228	<0.001	4.838	0.899	<0.001
Shape parameter						
Tumour Residue	0.069	0.121	0.567	-0.045	0.124	0.718
Path. Node	-0.208	0.125	0.097	-0.099	0.149	0.508
Stage 3A				-0.425	0.290	0.143
Stage 3B				-0.395	0.249	0.113
Intercept	0.049	0.116	0.675	0.366	0.247	0.138

Model 1 includes factors of tumour residue (TR) and pathologic node (PN); Model 2 includes factors of TR, PN and tumour stage; PH: proportional hazard; TR negative, PN negative and tumour stage 2B served as reference categories

was observed for factors of tumour residue, pathologic node and tumour stage (Figure 1). The univariate survival differences in the factors of variables by log rank test were significant (p<0.001) for tumour residue, pathologic node and tumour stage.

Analysis of prognostic factors for event free survival was done using non-mixture cure model with PH assumptions and compared with Cox model (Table 2). The same was compared using cure model with non PH assumption. Lognormal kernel was used for the survival distribution since the deviance value -2Log L was the least for the lognormal as compared to exponential and Weibull distributions. Test for PH assumption was done using Schoenfeld residuals. Multifactorial modelling was performed under twin settings: In Model 1, the variables analyzed were tumour residue and pathologic node status and in model 2, tumour stage was the additional variable. Model 1 revealed that the impact of both the variables was

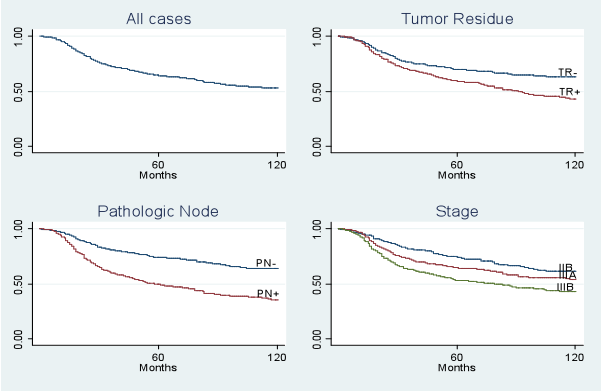

Figure 1. Kaplan Meier Survival for All Cases Together and Factors of Prognostic Variables, Breast Cancer, Cancer Institute (WIA), Chennai, India, 1990-99

Table 3. Comparison of Cure Fractions of Factors of Prognostic Variables using PH and non PH Cure Models, Cancer Institute (WIA), Breast cancer treated at Cancer Institute (WIA), Chennai, India, 1990-99

Stage	PN	PH Model		Non PH Model	
		TR-	TR+	TR-	TR+
All stages	PN-	62.9	46.6	62.5	54.8
	PN+	43.7	27.4	34.7	27.1
Stage 2B	PN-	68.2	61.8	52.9	40.1
	PN+	42.5	36.1	30.1	17.3
Stage 3A	PN-	62.1	55.7	68.3	55.5
	PN+	36.5	30.0	45.5	32.7
Stage 3B	PN-	52.4	46.0	59.7	46.9
	PN+	26.7	20.3	36.9	24.1

TR: tumour residue; PN: pathologic node; PH: proportional hazard

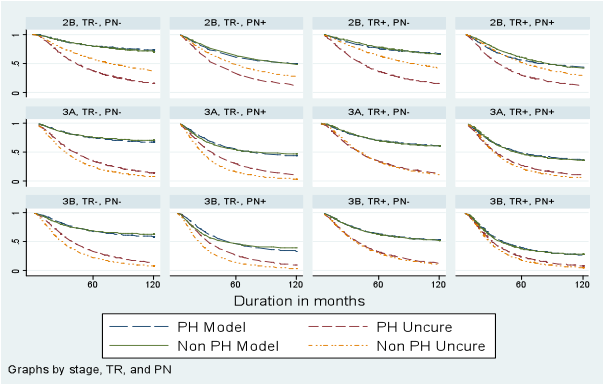

Figure 2. Survival Comparison of PH and Non-PH Cure Models of Factors of Prognostic Variables by Tumor Stage, Breast Cancer, Cancer Institute (WIA), Chennai, India, 1990-99

statistically significant and they emerged as independent prognostic factors under PH and non PH assumptions. When tumour stage was added to these variables in model 2, proportionality condition was violated. The deviance of the non PH cure model was the least of all the models in this study. Under non PH cure model, tumour stage was not statistically significant. The covariates introduced in the scale and shape parameters of non PH assumption were not statistically significant. Irrespective of tumour stage, both tumour residue and pathologic nodal status emerged as independent prognostic factors under appropriate cure model settings.

In Table 3, the above results are summarized as cure fractions for all stages together as well as stratified on tumour stage and compared for each category of combination of tumour residue (TR) and pathologic nodal (PN) status under PH and non PH assumptions based on cure models 1 and 2 as defined before. The cure fractions were decreasing with increasing tumour stage for all categories of tumour residue and pathological node status under PH assumption but not under non PH assumption. In both instances, a consistent ordering of cure fractions was forthcoming for all tumour stages: it was the highest for PN-TR- followed by PN-TR+, PN+TR- and PN+TR+. The cure fraction for all cases together was estimated as 47.5%.

The comparison of survival probabilities given by parametric models incorporating cure and uncure aspects under PH and non PH assumptions for factors of tumour residue and pathologic node for every tumour stage are shown in Figure 2. The event free survival curves based on cure models with PH and non PH assumptions showed minimal differences for each factor of PN, TR and tumour stage while there were perceptible differences in survival of the uncured.

Discussion

Cox model is the most widely used in the analysis of prognostic factors for cancer survival. This preference often leads to results being focused on hazard ratios or on survival curves by Kaplan-Meier method rather than on survival time or ratios of survival time and their distributions. This may tend to exaggerate the apparent survival differences among the groups of patients for whom the survival distributions actually overlap considerably. This is because the PH assumption, which is the basic assumption for application of the model, is not always tested (Valsecchi et al., 1996). In our study, stringent test for PH revealed some departure, which justified the use of parametric modelling of survival time under both PH and non PH assumptions.

The efficiency of any parametric model relies on the selection of correct parametric form of the survival distribution. A good discrimination among parametric models requires the censoring percentage as minimum (Narid et al., 2003). We therefore assessed a variety of different distributions and found that lognormal form fit the survival time data well. Tai et al., (2005) had demonstrated the applicability of log normal models in the survival of various cancers from Surveillance Epidemiology and End Results (SEER) data and compared lognormal and KM survival estimates. Pearlman (1976) estimated the growth rates of breast cancer that recurred in the scar by assuming that the recurrence started from a single cell were log normally distributed. Tai et al., (2005) also suggests that certain threshold year is required to wait before the statistical cure rate may be estimated for any cancer and thus highlighted the need for long-term follow-up. Owing to long follow up available in our study, proportion of long-term survivors was estimated through cure fraction by employing cure models under PH and non PH assumptions. The similarity of survival estimates by using cure model under PH assumption and Kaplan-Meier methods was striking.

Sposto (2002) had considered two classes of parametric cure models in the analysis of prognostic factors from the Children's Cancer Group and discussed the results in relation to Cox regression analysis under PH and non PH assumptions. Similar analysis was carried out to establish long-term survival and cure in young patients with Ewing's sarcoma (Weston et al., 2004). Lambert et al., (2007) investigated the importance of modelling the ancillary parameters in the selected parametric distribution. In our study, the results by Cox model and by cure model under PH assumption were analogous and

had elicited tumour residue, pathological node and tumour stage as independent prognostic factors for survival. However, owing to violation of PH condition, analysis by cure model under non PH assumption revealed that tumour residue and pathologic node were statistically significant but tumour stage was not. Further analysis also confirmed that both scale and shape parameters using lognormal distribution were dependent on the prognostic factors analysed. This amply justified the use of a parametric form of distribution for modelling survival time under non PH assumption in our study. Also, the results using cure fractions with respect to factors of tumour residue and pathologic node were consistent for all tumour stages together and across different tumour stages under PH and non PH assumptions. This demonstrated the utility of cure models in estimating long term survival by cure fraction using statistical cure models.

The models presented in this work could be extended in a number of ways. The choice of the link function is important in that it leads to different assumptions regarding the joint effect of covariates. The boundary problems associated with low or high cure fractions need more study.

In conclusion, it is imperative that in analysis of prognostic factors for survival, tests for PH assumption are routinely done. If any violation is detected, analysis under non PH condition is advocated and mixture cure models are useful in estimating the cure fraction and constructing survival curves for the non-cures.

Acknowledgments

The authors are grateful to the registry staff for their diligent work in data collection at registration and follow up without which this study would not have been possible. The registry is partly funded by Indian Council of Medical Research, New Delhi.

References

Berkson J, Gage RP (1950). Calculation of survival rates for cancer. *Proceedings of the Staff Meet Mayo Clinic*, **25**, 270-86.

Boag JW (1949). Maximum likelihood estimates of the proportion of patients cured by cancer therapy. *J Royal Statistical Society*, **11**, 15-44.

Cox DR (1972). Regression models and life tables (with discussion). *J Royal Statistical Society B*, **34**, 187-220.

Kaplan EL, Meier P (1958). Non-parametric estimation from incomplete observations. *J Am Statistical Assoc*, **53**, 457-81.

Lambert PC (2007). Modelling the cure fraction in survival studies. *Stata J*, **7**, 1-25.

Lambert PC, Thompson JR, Weston CL, et al (2007). Estimating and modelling the cure fraction in population-based cancer survival analysis. *Biostatistics*, **8**, 576-94.

Narid A, Schemper M (2003). Comparing Cox and parametric models in clinical studies. *Stat Med*, **22**, 3597-610.

Pearlman AW (1976). Breast cancer: Influence of growth rate on prognosis and treatment evaluation. A study based on mastectomy scar recurrences. *Cancer*, **38**, 1826-33.

Shanta V, Swaminathan R, Rama R, et al (2008). A retrospective analysis of locally advanced non-inflammatory breast cancer from Chennai, South India, 1990-99. *Int J Radiat Oncol Biol Phys*, **70**, 51-8.

Sposto R (2002). Cure model analysis in cancer: An application to data from the children's cancer group. *Stat Med*, **21**, 293-312.

Tai P, Yu E, Cserni G, et al (2005). Minimum follow-up time required for the estimation of statistical cure of cancer patients: verification using data from 42 cancer sites in the SEER database. *BMC Cancer*, **17**, 48-57.

Tai P, Yu E, Vinh-Hung V, et al (2004). Survival of patients with metastatic breast cancer: twenty-year data from two SEER registries. *BMC Cancer*, **2**, 60-8.

Valsecchi, MG, Silverstri D, Panieni PD (1996). Evaluation of long term survival: use of diagnostics and robust estimators with Cox's proportional hazards models. *Stat Med*, **15**, 2763-3780.

Weston CL, Douglas C, Craft AW, et al (2004). Establishing long-term survival and cure in young patients with Ewing's sarcoma. *Br J Cancer*, **91**, 225-32.