

TITLE PAGE

Type of manuscript (Original Article)

Title: Survival Analysis in the presence of a Cure Fraction Using Data of Dialysis Patients: A Bayesian Approach

Authors:

Shideh Rafati^{1,2}. PhD in Biostatistics. ORCID ID: 0000-0001-7108-414X

¹ Social Determinants in Health Promotion Research Center, Hormozgan Health Institute, Hormozgan University of Medical Sciences, Bandar Abbas, Iran.

² Department of Biostatistics and Epidemiology, Kerman University of Medical Sciences, Kerman, Iran. Email: shiderafati@yahoo.com

Mohammad Reza Baneshi^{2,3}. Professor of Biostatistics. ORCID ID: 0000-0002-6405-8688

² Department of Biostatistics and Epidemiology, Kerman University of Medical Sciences, Kerman, Iran.

³ Modeling in Health Research Center, Institute for Futures Studies in Health, Kerman University of Medical Sciences, Kerman, Iran. Email: rbaneshi2@gmail.com

Laleh Hassani⁴. Assistant Professor of health education and health promotion.

ORCID ID: 0000-0001-8446-0992

⁴ Mother and Child Welfare Research Center, Hormozgan University of Medical Sciences, Bandar Abbas, Iran. Email: 7hassani1969@gmail.com

Corresponding author: Abbas Bahrapour^{2,3}. Professor of Biostatistics. ORCID ID: 0000-0002-6343-9243

² Department of Biostatistics and Epidemiology, Kerman University of Medical Sciences, Kerman, Iran.

³ Modeling in Health Research Center, Institute for Futures Studies in Health, Kerman University of Medical Sciences, Kerman, Iran.

Tel: +98(0)9131404512 Fax: +983431325127

Email: abahrampour@yahoo.com a.bahrampour@griffith.edu.au

a_bahrampour@kmu.ac.ir

Survival Analysis in the presence of a Cure Fraction Using Data of Dialysis Patients: A Bayesian Approach

Abstract

Background: The aim of this study was to evaluate the goodness of fit of Bayesian mixture and non-mixture cure models to find the factors affecting dialysis patient's survival time where a significant proportion of the population has a long-term survival.

Study Design: A retrospective cohort study.

Methods: The data of 252 dialysis patients were used among whom 35 cases died. Since in this study a part of the population had long-term survival, Bayesian cure models were used and evaluated using DIC index. The data were analyzed by R and Openbugs Softwares.

Results: Of the 252 dialysis patients, 136(54%) were males and the mean (SD) age was 53.39 (18.09) years. The patient's follow-up time mean (SD) was 10.93(7.82) years. The 10 and 20-year survival rate of these patients were 87% and 73%, respectively. The findings show that the best fitting belonged to the Bayesian Non-mixture Cure Model (BNCM) with Dagum distribution. The variables of age, Body Mass Index, dialysis duration, frequency of dialysis, age of onset of dialysis, and occupation affected patients' survival based on BNCM with Dagum distribution.

Conclusions: The results demonstrated that the BNCM with Dagum distribution can be a good selection model to analyze survival data, where there is the possibility of a fraction of cure.

Keywords: Survival; Bayesian Approach; Cure Fraction; Dagum Distribution

Introduction

The advanced phase of chronic renal failure, in which life relies on transplantation or dialysis, is called end-stage renal disease (ESRD) [1]. Due to its chronic and disabling nature, this disorder may have a detrimental impact on patients' quality of life and contributes to decreased social interactions, depression, anger, decreased capacity of an individual to perform independent everyday life activities, and eventually increased mortality. It is necessary to determine the factors that influence the survival of these patients, with regard to the challenges and issues that dialysis patients face [2].

With significant advances in all areas of medical science, a significant proportion of patients with various types of diseases have long-term survival [3]. Thus, in many cases we have survival data that the number of patients experiencing death from a disease is much lower than that of censored cases. In such a situation, if all the events occurred at the beginning of the study and no event happened at the end of the Kaplan-Meier (K-M) curve, cure models can be used to provide a more accurate interpretation of survival data [4].

Cure models are specific types of survival models in which the population is a combination of susceptible cases (those who may experience the event, ie, patients with short-term survival) and cured/non-susceptible individuals who have long-term survival (those who never experience the event in the follow-up period) [5]. There are 2 main methods of mixture and non-mixture to model the survival data with a cure fraction.

The existence of a long, stable flatness with high censoring rate at the tail of the K-M curve and the adequacy of follow-up time, indicate that the data are suitable for applying cure models [6]. To identify cured individuals, statistical tests can be used to detect patients with long-term survival. The null hypothesis of this test indicates that all people may experience death and there is no cure fraction. To reject or accept the stated hypothesis, critical values that have been set by Maller and Zhou can be used [7, 8].

Cure models do not work well when the sample size is small, and the high censoring rate at the end of the K-M curve is one of the favorable features required to use cure models. However, in the secondary data used in the present study, all events occurred at the beginning of the follow-up, and the K-M curve tail was stable. Still, the number of samples at the end of the study was low (Censoring rate at the end of the survival curve was not high.); thus, Bayesian analysis of cure models was used instead of classical analysis.

Bayesian inference of cure models were introduced by some authors, including Martine [9, 10], Swain [11], Chen [12], Ibrahim [13], and Castro [14].

However, to date, no study has compared the prediction performance of the Bayesian mixture and non-mixture cure models based on Dagum, Weibull, and log-logistic distributions. Therefore, in this study, the mentioned models were used and compared. The results of this study may be helpful for future studies and analysis of medical and health-related data, where there is the possibility of a fraction of cure.

Methods

Secondary Data

Of the total patients admitted from 2010 to 2016, the data of 252 patients were recorded in the dialysis ward of Bandar Abbas Hospitals, Iran. Mortality was considered as the event of interest and censored cases included those who were alive at the end of the study, excluded cases and those treated with kidney transplant. The survival time of the patients was calculated by years from the onset of dialysis to the end of the study in 2016.

Data were obtained from a given checklist including age, gender, body mass index (BMI), education, age of onset of dialysis, job, blood type, marital status, smoking, history of diabetes, hypertension, renal stones and obstruction, renal cysts and congenital diseases, dialysis

duration (hour per session), number of dialysis sessions per week, history of cardiac-respiratory diseases, history of anemia and familial history of chronic renal failure.

In the end, 19 variables, including gender, job (five indicator variables), blood type (three indicator variables), marital status, history of smoking, diabetes and hypertension (all binary), age of censoring or death time, education, dialysis duration, number of dialysis sessions, BMI, age of diagnosis (all continuous), are used in Bayesian cure models with regard to the low frequency in some categories of independent variables.

Ethical considerations

This manuscript has been approved by the Ethics Committee of Kerman University of Medical Sciences at No. IR.KMU.REC.1397.599. Written informed consent was obtained from all the participants.

Models

1. Mixture cure models

The population is separated into two parts in a mixture cure model. Cured or long-term survivors and survivors that are uncured or short-term. Let the probability of being cured be p ($0 < p < 1$) and therefore $(1 - p)$ is the probability of being susceptible to an individual[9]. The corresponding t -time survival function is as follows:

$$S(t) = p + (1 - p)S_0(t)$$

Where $S(t)$ is the total population's survival function. $S_0(t)$ is the baseline survival function for the susceptible individuals for which type I Dagum, Weibull and Log-logistic distributions are assumed in this study. The distributions described below have been introduced.

2. Non-mixture cure models

The survival function is defined in this case as

$$S(t) = p^{F_0(t)} = \exp[(\ln p)F_0(t)]$$

Where $F_0(t) = 1 - S_0(t)$ is the baseline cumulative distribution function for the susceptible people.

To model the cure probability (p) under both mixture and non-mixture cure models, we applied the logistic function.

3. The type I Dagum distribution

Suppose that survival time for the uncured individuals has the Dagum distribution with three parameters. The This distribution's cumulative distribution function is defined through (for $t > 0$)

$$F_0(t) = \left[1 + \frac{1}{a}t^{-b}\right]^{-c} = \left(\frac{1 + at^b}{at^b}\right)^{-c}$$

Where b and c are positive parameters of shape and the scale parameter is a ($a = \exp(X_i'\theta)$ the covariates can be included in the model through a). Notice that the $c=1$ case leads to the log-logistics distribution[10].

The third distribution is also assumed to be the Weibull distribution with two parameters $\alpha, \gamma > 0$ ($S_0(t) = \exp(-\gamma t^\alpha)$), which are the shape and scale parameters respectively.

Priors

The normal prior distributions $N(0, 100)$ were considered in the Bayesian analysis for the vector of the parameter θ . Also, in Dagum, Weibull, and log-logistic distributions, the uniform prior distribution was assumed for the shape parameters. Prior independence of the parameters was presumed in the model for all scenarios.

Bayesian Inference

Here, the Bayesian mixture and nonmixture cure models were used based on Dagum, Weibull and log-logistic distributions. By combining the joint prior distribution and the likelihood function, the joint posterior distribution was obtained for the parameters of the model.

Posterior summaries of interest are obtained from simulated samples for the joint posterior distribution using standard Markov Chain Monte Carlo (MCMC) procedures. Also, 1,010,000 samples were generated for each parameter of interest. The first 10000 simulated samples were discarded as a burn-in period, which is usually used to minimize the effect of the initial values. The posterior summaries of interest were based on 20,000 samples, taking every 50 sample to have approximately uncorrelated values[9].

The Bayes estimates of the parameters were the mean of Gibbs samples, which were drawn from the joint posterior distribution. Convergence of the MCMC algorithm was monitored by history and autocorrelation plots for the simulated samples[9, 15]. Inferences were obtained using R and OpenBUGS Softwares.

In addition to the history plot, the Monte Carlo error (MC_error) were used to evaluate the accuracy of posterior estimates for each parameter, which is an estimate of the difference between the actual posterior mean and the estimated posterior mean for each parameter. If the Monte Carlo error value for each parameter is less than 5% of its standard deviation, then, convergence is obtained for that parameter, indicating no need for further simulation samples.

Model Selection

Deviance Information Criteria (DIC), as a measure of the goodness-of-fit, used to compare between the mixture and non-mixture models, where a lower DIC value indicates a better model fit.

Results

Overall, 252 hemodialysis patients were studied, of them, 35 (13.9%) cases faced the event of death and 217 (86.1%) cases were censored. The follow-up time mean (SD) was 10.93 (7.82) years. The 10 and 20-year survival rate of these patients were 87% and 73%, respectively. Table1 shows the patients characteristics.

Figure 1 displays a K-M plot for overall survival function, which shows "flatness" near 0.60 in the survival curve, and thus a cure model appears to be suitable for this data. Based on this diagram, flatness occurs after about 20 years (fig1).

In table2, the DIC values of Bayesian cure models based on Weibull, Log-logistic and type I Dagum distributions and in the presence all of covariates have been reported.

Based on the DIC index in Table 2, the goodness of fit of the non-mixture models is better than the mixture ones, and belongs to the dagum distributions. The worst performance is related to the Bayesian Weibull cure model (mixture and non-mixture) because the most DICs belong to these two models. The posterior summaries of parameters of non-mixture cure models with type I Dagum distribution have been presented in table 3.

In section short term survival in table3, the 95% credible interval for age of death or censoring time, BMI, duration of each dialysis, frequency of dialysis per week, age of onset of dialysis, and occupation does not include zero. Therefore, these variables have a significant effect on the short-term survival.

Based on this table, by adjusting other factors, for one unit of increase in age, the failure odds is increased by 0.11 ($OR=\exp(0.105)$). for one unit of increase in BMI and duration of each dialysis, the failure odds is reduced by 0.17 ($OR=\exp(-0.181)$) and 0.31 ($OR=\exp(-0.364)$), respectively. For one unit of increase in frequency of dialysis and age of onset of dialysis, the failure odds is increased by 0.28 ($OR=\exp(0.252)$) and 0.09($OR=\exp(0.084)$), respectively. The death odds of housewives is 0.17 less than employees ($OR=\exp(-0.192)$).

Also, in section long term survival, the 95% credible interval for age of death or censoring time, BMI and age of onset of dialysis does not include zero suggesting that these variables have significant effect on the long-term survival.

Discussion

A total of 252 dialysis patients with 35 (13.9%) death events were used for this study. Due to the low-event data, use of classical methods, such as Cox, was not possible. Also, all 35 death events (target event) occurred at the beginning of the study, and all cases were censored at the end of the study. However, the number of samples at the end of the study was low (The censoring rate was not high at the tail of the survival curve). Thus, Bayesian cure models were used to analyze the data.

Given the nature of the data of dialysis patients, this study aimed to evaluate and compare the performance of Bayesian cure models with Weibull, log-logistic, and Dagum distribution to analyze these data.

Based on the findings of the present study, the variables of age, Body Mass Index, dialysis duration, frequency of dialysis, age of onset of dialysis, and occupation affected dialysis patients' survival.

The findings of present study revealed that older age was related to higher mortality in dialysis patients. This result is consistent with that of the previous studies[16, 17]. Also, similar to the

present study, Tsur's study confirmed that higher BMI, was significantly associated with longer survival[17], hence, high BMI is protective in these patients. Further, based on the present study, occupation is one of the most important factors affecting dialysis patients' survival that its effect is confirmed in past studies[18].

Based on the present study, for one unit of increase in the duration of dialysis, the odds of death was decreased by 0.31; the significance of this variable was verified in other studies[19]. Also, increasing one unit in the number of dialysis sessions per week increases the odds of death that is inconsistent with some other studies[20].

The results of this study indicated that the Bayesian non-mixture cure models are superior in compared to the Bayesian mixture cure models, and the results of some previous studies are consistent with those of the present study [15, 21]. However, The Swain study showed that the mixture cure model has a better fit than the nonmixture cure model based on the generalized Gompertz distribution [11].

Another study has compared mixture and non-mixture cure models based on the generalized modified Weibull distribution by Bayesian approach. Based on DIC values, it was found that the mixture and nonmixture cure models provide very close results [9]. Also, a study aimed to assess mixture and nonmixture cure models that revealed both classes fit the data well [6].

The results of Jafari Koushaki study showed the performance of the non-mixture cure model with log-logistic distribution is better than the Weibull non-mixture cure model, as the value of the DIC is lower for the log-logistic non-mixture cure model [22] that present study confirms it.

the implementation of the Bayesian cure models with the less used distribution of Dagum is the strength of the present study and the lack of more important and more effective variables on

patient survival such as adequacy of dialysis, creatinine level, urea, albumin, and hemoglobin is its weakness.

Conclusion

The results of the present study demonstrated that the Bayesian non-mixture cure model with type I Dagum distribution can be a good choice for analyzing survival data, where there is the possibility of a fraction of cure.

Abbreviations

DIC: Deviance Information Criteria; MC_error: Monte Carlo error; BMI: Body Mass Index; K-M curve: Kaplan-Meier curve; SD: Standard Deviation; BNCM: Bayesian Non-mixture Cure Model; ESRD: end-stage renal disease

Ethics approval and consent to participate

This manuscript has been approved by the Ethics Committee of Kerman University of Medical Sciences at No. IR.KMU.REC.1397.599. Written informed consent was obtained from all the participants. In addition to this, all methods were performed in accordance with relevant guidelines and regulations.

Consent for publication

Not applicable: individual information has not been published.

Availability of data and materials

The data sets used and/or analyzed during the current study are not publicly available due to confidentiality of data and subsequent research, but are available from the corresponding author on reasonable request.

Competing interests

The authors declare that they have no competing interests.

Funding

No funding was received for the study.

Authors' contributions

SR proposed the study. MB and LH were involved in the design of study with the supervision of AB. LH collected the data. SR performed the statistical analysis. SR, MB and AB prepared the first draft of the manuscript and the authors read, revised and approved the final manuscript.

Acknowledgements

The authors thank staff of Dialysis Ward of Shahid Mohammadi and Persian Gulf Hospitals located on Bandar Abbas (Iran) for assisting and allowing for data collecting. Also, the authors wish to acknowledge Martinez and colleagues[9] for their study, as their method is the original source of described method in this study.

References

1. Rafati F, Mashayekhi F, Dastyar N. Caregiver Burden and Spiritual Well-being in Caregivers of Hemodialysis Patients. *Journal of Religion and Health*. 2019;1-13.
2. Rafati S, Baneshi MR, Hassani L, Bahrampour A. Comparison of penalized cox regression methods in low-dimensional data with few-events: An application to dialysis patients' data. *Journal of research in health sciences*. 2019;19(3):e00452.

3. Kim S, Chen M-H, Dey DK, Gamerman D. Bayesian dynamic models for survival data with a cure fraction. *Lifetime Data Anal.* 2007;13(1):17-35.
4. Müller UU, Van Keilegom I. Goodness-of-fit tests for the cure rate in a mixture cure model. *Biometrika.* 2019;106(1):211-27.
5. Coelho-Barros EA, Achcar JA, Mazucheli J. Cure Rate Models Considering The Burr XII Distribution in Presence of Covariate. *Journal of Statistical Theory and Applications.* 2017;16(2):150-64.
6. 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.
7. Maller RA, Zhou X. *Survival analysis with long-term survivors*: John Wiley & Sons; 1996.
8. Hoseini M, Bahrampour A, Mirzaee M. Comparison of weibull and lognormal cure models with cox in the survival analysis of breast cancer patients in Rafsanjan. *Journal of research in health sciences.* 2017;17(1).
9. Martinez EZ, Achcar JA, Jácome AA, Santos JS. Mixture and non-mixture cure fraction models based on the generalized modified Weibull distribution with an application to gastric cancer data. *Comput Methods Programs Biomed.* 2013;112(3):343-55.
10. Martinez EZ, Achcar JA. A new straightforward defective distribution for survival analysis in the presence of a cure fraction. *Journal of Statistical Theory and Practice.* 2018;12(4):688-703.
11. Swain PK, Grover G, Goel K. Mixture and Non-Mixture Cure Fraction Models Based on Generalized Gompertz Distribution under Bayesian Approach. *Tatra Mountains Mathematical Publications.* 2016;66(1):121-35.
12. Chen M-H, Ibrahim JG, Sinha D. A new Bayesian model for survival data with a surviving fraction. *Journal of the American Statistical Association.* 1999;94(447):909-19.
13. Ibrahim NA, Taweab F, Arasan J. A parametric non-mixture cure survival model with censored data. *Computational Problems in Engineering*: Springer; 2014. p. 231-8.
14. Castro Md, Cancho VG, Rodrigues J. A Bayesian long-term survival model parametrized in the cured fraction. *Biometrical Journal: Journal of Mathematical Methods in Biosciences.* 2009;51(3):443-55.
15. Rafati S, Baneshi MR, Bahrampour A. Factors Affecting Long-Survival of Patients with Breast Cancer by Non-Mixture and Mixture Cure Models Using the Weibull, Log-logistic and Dagum Distributions: A Bayesian Approach. *Asian Pac J Cancer Prev.* 2020;21(2):485-90.
16. Tuğcu M, Kasapoğlu U, Şahin G, Apaydın S. The factors affecting survival in geriatric hemodialysis patients. *International journal of nephrology.* 2018;2018.

17. Tsur N, Menashe I, Haviv YS. Risk Factors Before Dialysis Predominate as Mortality Predictors in Diabetic Maintenance Dialysis patients. *Sci Rep.* 2019;9(1):1-8.
18. Tao S, Zeng X, Liu J, Fu P. Socioeconomic status and mortality among dialysis patients: a systematic review and meta-analysis. *Int Urol Nephrol.* 2019;51(3):509-18.
19. Goh A, Vathsala A. Native renal cysts and dialysis duration are risk factors for renal cell carcinoma in renal transplant recipients. *Am J Transplant.* 2011;11(1):86-92.
20. Chandrashekar A, Ramakrishnan S, Rangarajan D. Survival analysis of patients on maintenance hemodialysis. *Indian J Nephrol.* 2014;24(4):206.
21. Achcar JA, Coelho-Barros EA, Mazucheli J. Cure fraction models using mixture and non-mixture models. *Tatra Mountains Mathematical Publications.* 2012;51(1):1-9.
22. Jafari-Koshki T, Mansourian M, Mokarian F. Exploring factors related to metastasis free survival in breast cancer patients using Bayesian cure models. *Asian Pac J Cancer Prev.* 2014;15(22):9673-8.

Tables

Table1-the patients characteristics

Continuous Variables	Died (n=35)		Censored (n=217)		Pvalue
	Mean	SD	Mean	SD	
Age (year)	54.14	22.38	53.27	17.36	0.69
Age starting dialysis	42.77	19.38	42.90	16.72	0.04
Body Mass Index	21.51	3.61	23.10	4.30	0.76
dialysis duration (hours)	3.66	0.48	3.79	3.40	0.06
Categorical Variables	Number	Percent	Number	Percent	
Blood group O	16	45.7	92	42.4	0.71
Blood group A	5	14.3	57	26.3	0.12
Blood group B	14	40.0	58	26.7	0.10
Blood group AB	0	0.0	10	4.6	0.19
Employee	2	5.7	17	7.8	0.32
Jobless	10	28.6	36	16.6	0.08
Housewife	13	37.1	88	40.6	0.69
Farmer	2	5.7	6	2.8	0.36
Retired	4	11.4	38	17.5	0.36
Other occupation	4	11.4	32	14.7	0.60
Illiterate	16	45.7	70	32.3	0.21
Low literacy	15	42.9	102	47.0	0.65
Diploma	2	5.7	35	16.1	0.11
Collegiate	2	5.7	10	4.6	0.77
Males	18	51.4	118	54.4	0.74
Married	27	77.2	180	82.9	0.41
Tobacco use	15	42.9	76	35.0	0.36
Diabetes	24	68.6	110	50.7	0.04
Hypertension	18	51.4	134	61.8	0.24

Table2- DIC values of Bayesian cure models

Models	DIC
	Mean
Mixture type I Dagum	280.2
Non-mixture type I Dagum	110.7
Mixture Log-logistic	280.6

Non-mixture Log-logistic	249.3
Mixture Weibull	523.2
Non-mixture Weibull	490.9

Table 3: Posterior summaries of the non-mixture type I Dagum cure models.

Parameter	Posterior mean	Posterior SD*	MC_error	95% Credible Interval	Odds Ratio
-----------	----------------	---------------	----------	-----------------------	------------

Short-term survival					
age of death or censoring time	0.105	0.039	<0.001	(0.048, 0.169)**	1.110
body mass index	-0.181	0.050	<0.001	(-0.292,-0.085)**	0.834
dialysis duration	-0.364	0.070	<0.001	(-0.825,-0.097)**	0.694
frequency of dialysis per week	0.252	0.100	<0.001	(0.242, 0.700)**	1.286
age of onset of dialysis	0.084	0.041	<0.001	(0.044, 0.171)**	1.087
Occupation(Housewives) Reference: Employees	-0.192	0.080	<0.001	(-0.721,-0.337)**	0.825
Long-term survival					
age of death or censoring time	-0.357	0.009	<0.001	(-1.889,-0.831)**	0.699
body mass index	-0.666	0.007	<0.001	(-1.253,-0.072)**	0.513
age of onset of dialysis	-0.620	0.067	<0.001	(-1.629,-0.500)**	0.537

* Standard Deviation ** significant at 95% level

Figure

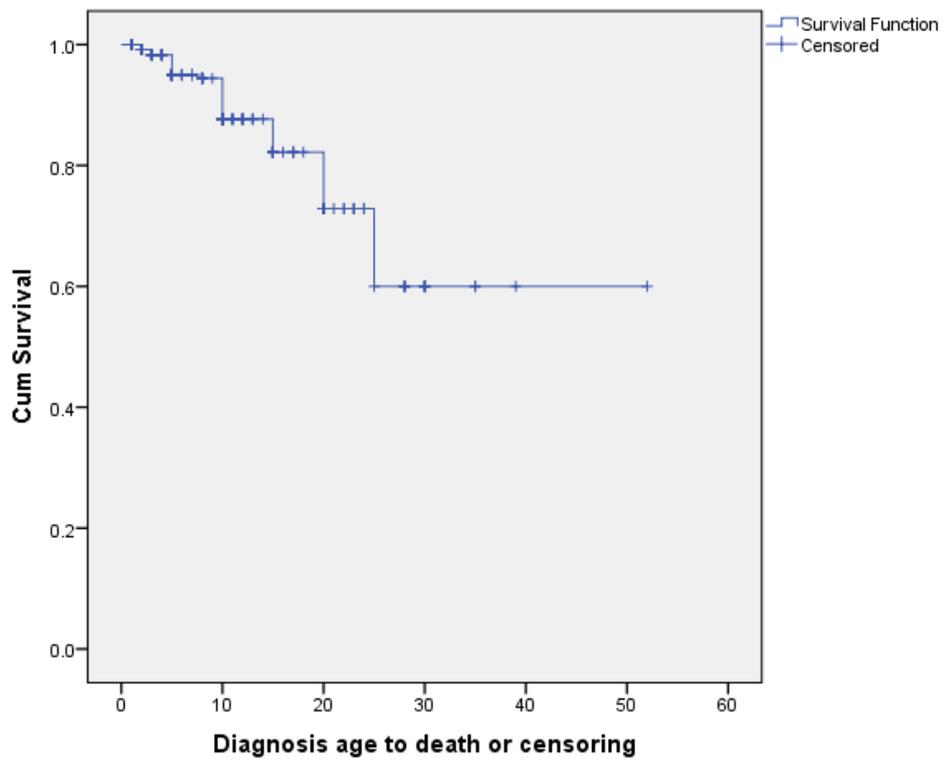


Fig1. Kaplan–Meier estimate of the overall survival function for the dialysis patient data