BATHTUB HAZARD MODEL WITH COVARIATES IN THE PRESENCE OF RIGHT- AND INTERVAL-CENSORED DATA

Idari Ismail

MyStats 2022
9TH MALAYSIA STATISTICS CONFERENCE
Department of Statistics Malaysia
Dealing with Uncertainties: Unearthing Measures for Recovery
4TH OCT. 2022 (VIRTUAL) & 5TH OCT. 2022 (ILSM, SUNGKAI, PERAK)
Organised by: PRIME MINISTER'S DEPARTMENT DEPARTMENT OF STATISTICS MALAYSIA, BANK NEGARA MALAYSIA, MALAYSIA INSTITUTE OF STATISTICS

PRESENTATION OUTLINE

INTRODUCTION

INTRODUCTION

Reliability Analysis

Survival Analysis

- The specialized fields of mathematical statistics - developed to deal with the special type of time-to-event random variables.
- Reliability analysis - methods related to assessment and prediction of successful operation or performance of products.

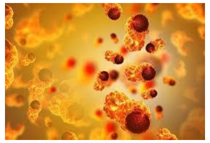
INTRODUCTION

- Survival analysis is the analysis of time-to-event data
- Deals with modelling and analysis of lifetime data(survival data or failure time)

EXPOSURE
cancer

survival time

EVENT
death



INTRODUCTION

CENSORING

- ♠ Occurs when we do not know the exact time-to-event for an included observation
- ♠ Due to incomplete observations

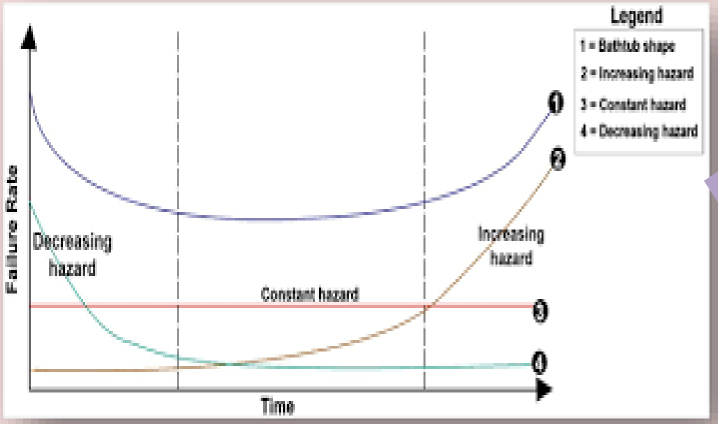
TYPES OF CENSORING

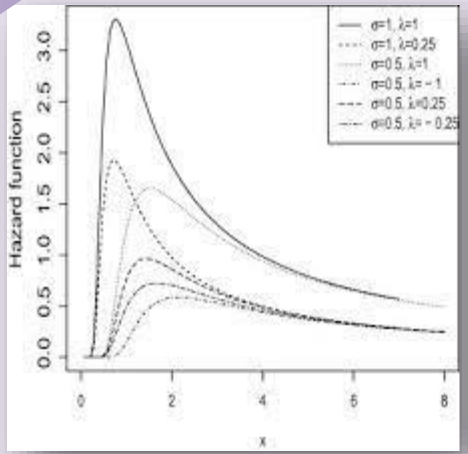
- ✓ Right censoring
- ✓ Left censoring
- ✓ Interval censoring

INTRODUCTION

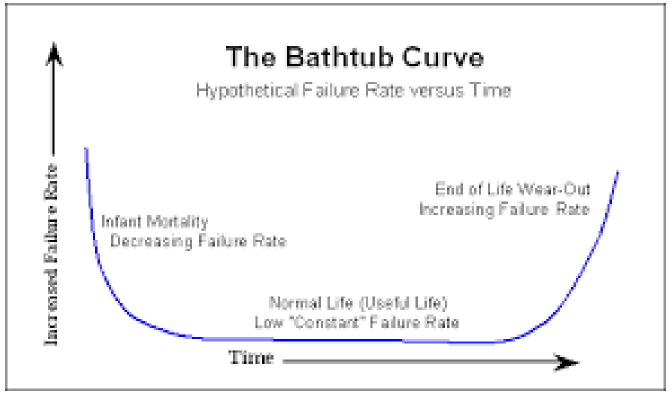
Failure Rate Function

- Increasing
- Decreasing
- Constant
- Bathtub-shaped
- Upside down bathtub-shaped





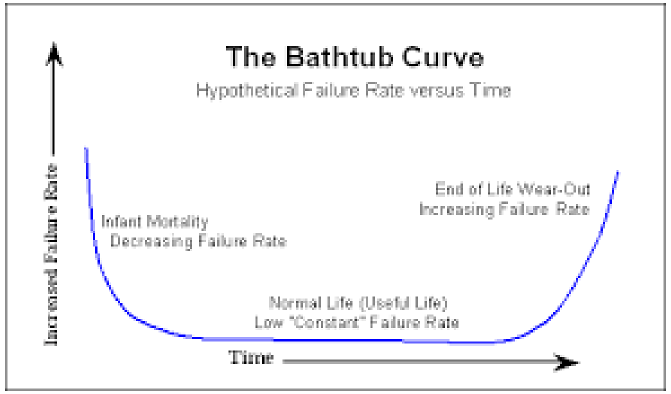
Bathtub-shaped failure rates



- Gamma or Weibull - accommodate monotone failure rates.
- failure rates firstly decrease, then stagnant at a constant level and eventually increase - resembles a bathtub, and thus, known as bathtub-shaped failure rate.

INTRODUCTION

Bathtub-shaped failure rates



Can be observed when studying the lifespan of an industrial product or the lifetime of a biological entity (Dimitrakopoulou et al., 2007).

Example

considering a high failure rate in infant mortality which decreases to a certain level, then remains constant for some time, and eventually increases (Gaver & Acar, 1979).

Previous studies

Gaver & Acar (1979), Mudholkar & Srivastava (1993), and Smith & Bain (1975) among others.

INTRODUCTION

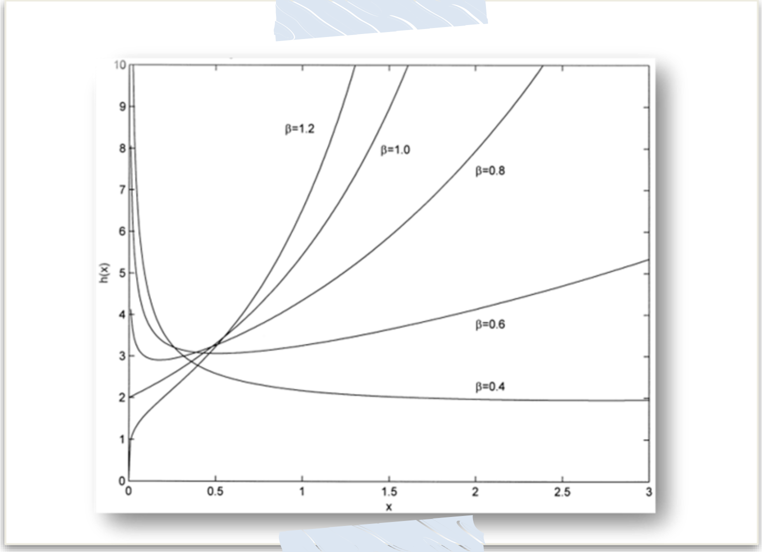
Bathtub hazard model

Chen (2000)

no two-parameter distribution that the failure rate exhibit bathtub-shaped.

Proposed by Chen (2000)

Bathtub-shaped and increasing-depending on its parameter.



LITERATURE REVIEW

LITERATURE REVIEW

Wu et al. (2004)

simple method for conducting statistical test with regards to the shape parameter where the method can be applied for a type-II right-censored data.

Wu (2008)

discussed exact confidence interval and exact join confidence region for the parameters under progressive type-II censored sample.

LITERATURE REVIEW

Wang et al. (2014)

Type II censored sample - discussed interval estimations for the parameters in bathtub hazard model.

Sarhan et al. (2012)

parameter estimation of the bathtub hazard model by using maximum likelihood and Bayes method.

LITERATURE REVIEW

Sarhan & Mustafa (2022)

- developed a new lifetime distribution based on bathtub hazard model and generalized exponential distribution.
- parameter estimation using maximum likelihood method and Bayesian procedures.

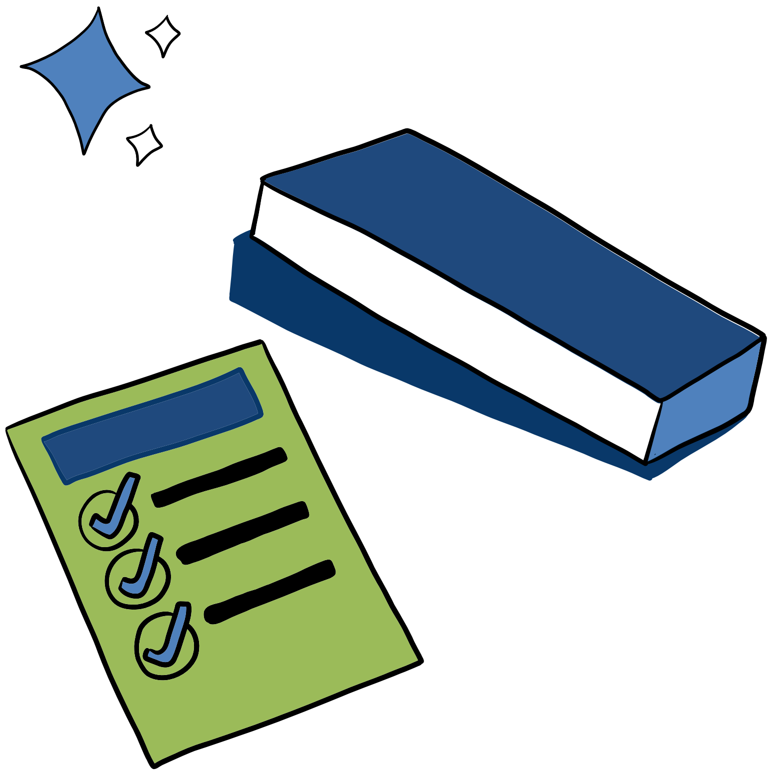
Chen & Gui (2020);

Zhang & Gui (2022)

- discussed parameter estimation of bathtub hazard model and presented confidence intervals for the model's parameters extensive research has been undertaken to study the bathtub hazard model.

the research to date has not focussed on investigating the bathtub hazard model in the presence of interval censored data

extend the bathtub hazard model by incorporating covariates in the presence of right- and interval-censored data.



Real Data Application – breast cancer data

Midpoint imputation was applied – compared with no imputation

METHODOLOGY

METHODOLOGY

Cumulative Hazard Function

$$F(t) = 1 - e^{\lambda\left(1-e^{t^{\alpha}}\right)} \quad , t \geq 0$$

Probability Density Function

$$f(t;\lambda,\alpha) = \lambda\alpha t^{\alpha-1} e^{t\alpha} e^{\lambda\left(1-e^{t^{\alpha}}\right)}$$

Survival Function

$$S(t;\lambda,\alpha) = e^{\lambda\left(1-e^{t^{\alpha}}\right)}$$

Bathtub hazard model

Hazard Function

$$h(t;\lambda,\alpha) = \lambda\alpha t^{\alpha-1} e^{t^{\alpha}}$$

The failure rate function becomes bathtub-like as $\boldsymbol{\lambda} < \mathbf{1}$ and is increasing when $\boldsymbol{\lambda} \geq \mathbf{1}$

METHODOLOGY

Let the parameter λ be a function of the covariates

$$\lambda_i = e^{-\beta_0 - \beta_1 x_i}$$

The failure rate function for a data set with a fixed covariate x_i where $i = 1, 2, \ldots, n$

$$h(t) = e^{-\beta_0 - \beta_1 x_i} \alpha t_i^{\alpha - 1} e^{t_i^{\alpha}}$$

METHODOLOGY

The likelihood function for the full sample if there are no censored observation is :

$$L(\theta)=\prod_{i=1}^{n} f(t_i)=\prod_{i=1}^{n}\left\{ e^{(-\beta_0-\beta_1 x_i)}\,\alpha t_i^{\alpha-1} e^{t_i^{\alpha}}\, e^{e^{(-\beta_0-\beta_1 x_i)}\left(1-e^{t_i^{\alpha}}\right)} \right\}$$

Censoring Indicator

$\delta_{I_i} = 1$ if the subject is interval-censored, 0 otherwise

$\delta_{R_i} = 1$ if the subject is right censored, 0 otherwise

$\delta_{E_i} = 1$ if subject is not censored (exact survival time is observed), 0 otherwise

METHODOLOGY

The likelihood function for right-, interval-censored or uncensored **without any imputation** is,

$$L(\theta)=\prod_{i=1}^{n}\left[f(t_i)\right]^{\delta_{E_i}}\left[S(t_{R_i})\right]^{\delta_{R_i}}\left[S(t_{L_i})-S(t_{R_i})\right]^{\delta_{I_i}}$$

$$=\prod_{i=1}^{n}\left[e^{\left(-\beta_0-\beta_1 x_i\right)}\alpha t_i^{\alpha-1}e^{t_i^{\alpha}}e^{e^{\left(-\beta_0-\beta_1 x_i\right)}\left(1-e^{t_i^{\alpha}}\right)}\right]^{\delta_{E_i}}$$

$$\mathrm{X}\left[e^{e^{\left(-\beta_0-\beta_1 x_i\right)}\left(1-e^{t_{R_i}^{\alpha}}\right)}\right]^{\delta_{R_i}}\mathrm{X}\left[e^{e^{\left(-\beta_0-\beta_1 x_i\right)}\left(1-e^{t_{L_i}^{\alpha}}\right)}-e^{e^{\left(-\beta_0-\beta_1 x_i\right)}\left(1-e^{t_{R_i}^{\alpha}}\right)}\right]^{\delta_{I_i}}$$

METHODOLOGY

Log-likelihood function

$$\ell(\theta)=\sum_{i=1}^{n}\delta_{E_i}\left[\left(-\beta_0-\beta_1 x_i\right)+\ln\left(\alpha\right)+(\alpha-1)\ln(t_i)+t^{\alpha}+e^{\left(-\beta_0-\beta_1 x_i\right)}(1-e^{t_i^{\alpha}})\right]$$

$$+\delta_{R_i}\left[e^{\left(-\beta_0-\beta_1 x_i\right)}(1-e^{t_{R_i}^{\alpha}})\right]+\delta_{I_i}\ln\left[e^{e^{\left(-\beta_0-\beta_1 x_i\right)}\left(1-e^{t_{L_i}^{\alpha}}\right)}-e^{e^{\left(-\beta_0-\beta_1 x_i\right)}\left(1-e^{t_{R_i}^{\alpha}}\right)}\right]$$

METHODOLOGY

The likelihood function for right-, interval- censored or uncensored **with imputation** method is,

$$L(\theta)=\prod_{i=1}^{n}\left[f(t_i)\right]^{\delta_{E_i}}\left[S(t_{R_i})\right]^{\delta_{R_i}}\left[f(\tilde{t}_i)\right]^{\delta_{I_i}}$$

$$=\prod_{i=1}^{n}\left[e^{\left(-\beta_0-\beta_1x_i\right)}\alpha t_i^{\alpha-1}e^{t_i^{\alpha}}e^{e^{\left(-\beta_0-\beta_1x_i\right)}\left(1-e^{t_i^{\alpha}}\right)}\right]^{\delta_{E_i}}\text{X}\left[e^{e^{\left(-\beta_0-\beta_1x_i\right)}\left(1-e^{t_{R_i}^{\alpha}}\right)}\right]^{\delta_{R_i}}$$

$$\text{X}\left[e^{\left(-\beta_0-\beta_1x_i\right)}\alpha\tilde{t}_i^{\alpha-1}e^{\tilde{t}_i^{\alpha}}e^{e^{\left(-\beta_0-\beta_1x_i\right)}\left(1-e^{\tilde{t}_i^{\alpha}}\right)}\right]^{\delta_{I_i}}$$

METHODOLOGY

Log-likelihood function

$$\ell(\theta)=\sum_{i=1}^{n}\delta_{E_i}\left[\left(-\beta_0-\beta_1 x_i\right)+\ln\left(\alpha\right)+(\alpha-1)\ln(t_i)+t_i^{\alpha}+e^{\left(-\beta_0-\beta_1 x_i\right)}(1-e^{t_i^{\alpha}})\right]$$

$$+\delta_{R_i}\left[e^{\left(-\beta_0-\beta_1 x_i\right)}(1-e^{t_{R_i}^{\alpha}})\right]$$

$$+\delta_{I_i}\left[\left(-\beta_0-\beta_1 x_i\right)+\ln\left(\alpha\right)+(\alpha-1)\ln(\tilde{t}_i)+\tilde{t}_i^{\alpha}+e^{\left(-\beta_0-\beta_1 x_i\right)}(1-e^{\tilde{t}_i^{\alpha}})\right]$$

METHODOLOGY

midpoint imputation - the intervals are replaced by the midpoint.

imputation technique - has an advantage over other methods due to its simplicity and ease of implementation (Chen & Sun, 2010).

RESULTS

Simulation Study

$N = 1000$

$n = 50, 100, 150, 250$

The values of the parameters α, β_0 and β_1 were particularly set at 0.4, 3.3 and 0.9 respectively

R programming language

Simulation study

Step 1	Generate covariate values $\boldsymbol{x_i}$ from a standard normal distribution
Step 2	Generate a sequence of random numbers $\boldsymbol{u_i}$ **from a standard uniform distribution** on the unit interval (0,1) to obtain the event time, $\boldsymbol{t_i}$ for $\boldsymbol{i = 1, 2, \ldots, n}$
Step 3	Generate censoring times, $\boldsymbol{c_i}$ from an exponential distribution with the value of $\boldsymbol{\mu}$ would be modified to obtain the desired censoring proportion (cp).
Step 4	Generate survival time $\boldsymbol{t_i}$

Simulation Study

Generate survival time $\boldsymbol{t_i}$

$$t_i = \left(\ln\left(1 - \frac{\ln(1-U_i)}{e^{(-\beta_0-\beta_1 x_i)}} \right) \right)^{1/\alpha}$$

RESULTS

Summary of standard error (SE) values for parameter estimates

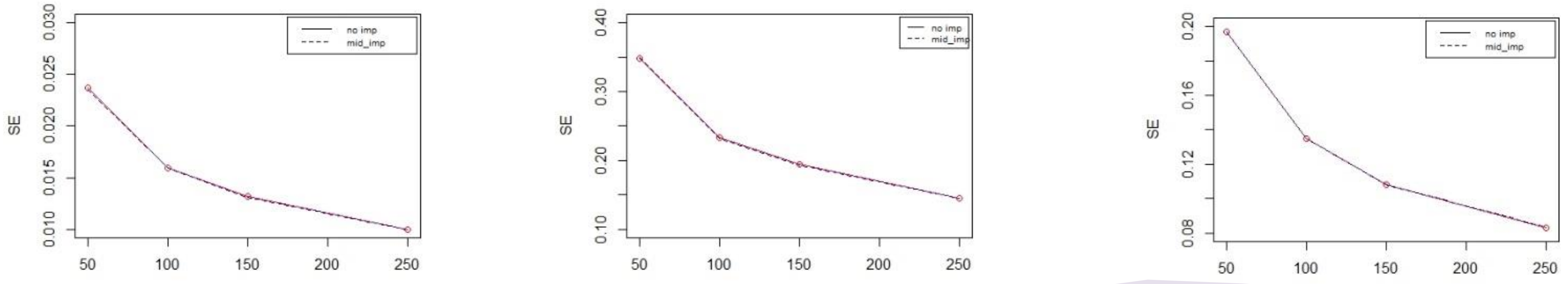
Estimates	n	No Imputation				Midpoint Imputation			
		cp = 0	cp = 10%	cp = 20%	cp = 30%	cp = 0	cp = 10%	cp = 20%	cp = 30%
$\hat{\alpha}$	50	0.02366	0.02622	0.02631	0.03059	0.02348	0.02413	0.02426	0.02598
	100	0.01598	0.01726	0.02090	0.02011	0.01594	0.01631	0.01786	0.01690
	150	0.01313	0.01431	0.01519	0.01642	0.01312	0.01346	0.01354	0.01364
	250	0.01000	0.01070	0.01157	0.01379	0.01000	0.01055	0.01038	0.01112
$\hat{\beta}_0$	50	0.34859	0.35192	0.36139	0.37577	0.34886	0.35055	0.35350	0.35769
	100	0.23250	0.24050	0.26225	0.25066	0.23210	0.23733	0.25004	0.24054
	150	0.19378	0.19813	0.20156	0.20563	0.19371	0.19476	0.19348	0.19698
	250	0.14554	0.15099	0.15458	0.16515	0.14549	0.15027	0.14669	0.15524
$\hat{\beta}_1$	50	0.19702	0.20701	0.20877	0.22334	0.19668	0.19990	0.20078	0.19993
	100	0.13490	0.13417	0.15676	0.15401	0.13498	0.14063	0.13735	0.13067
	150	0.10798	0.10984	0.11410	0.11779	0.10809	0.10852	0.10823	0.10558
	250	0.08325	0.08127	0.08743	0.09193	0.08327	0.08346	0.08300	0.08481

RESULTS

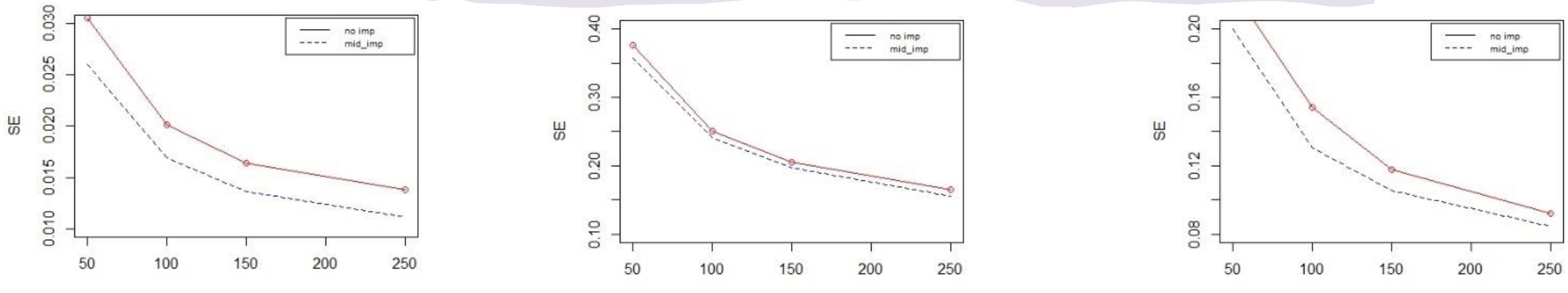
Summary of root mean square error (RMSE) values for parameter estimates

Estimates	n	No Imputation				Midpoint Imputation			
		cp = 0	cp = 10%	cp = 20%	cp = 30%	cp = 0	cp = 10%	cp = 20%	cp = 30%
$\hat{\alpha}$	50	0.02469	0.02692	0.02752	0.04108	0.02481	0.02448	0.02456	0.02610
	100	0.01651	0.01790	0.03225	0.03607	0.01649	0.01632	0.01864	0.01767
	150	0.01329	0.01634	0.02426	0.03574	0.01329	0.01366	0.01414	0.01524
	250	0.01010	0.01391	0.02391	0.04283	0.01010	0.01092	0.01103	0.01437
$\hat{\beta}_0$	50	0.36208	0.35199	0.36146	0.38923	0.36461	0.35911	0.36435	0.36099
	100	0.23819	0.24060	0.28328	0.28088	0.23791	0.23926	0.25028	0.24104
	150	0.19541	0.20160	0.22416	0.25231	0.19546	0.19498	0.19390	0.19699
	250	0.14630	0.15929	0.19029	0.24768	0.14626	0.15030	0.14673	0.15556
$\hat{\beta}_1$	50	0.20190	0.21165	0.21389	0.22505	0.20183	0.20307	0.20789	0.20208
	100	0.13617	0.13527	0.15815	0.15488	0.13623	0.14236	0.13924	0.13238
	150	0.10897	0.11018	0.11420	0.11811	0.10905	0.11000	0.10974	0.10673
	250	0.08368	0.08127	0.08766	0.09195	0.08370	0.08428	0.08358	0.08490

RESULTS

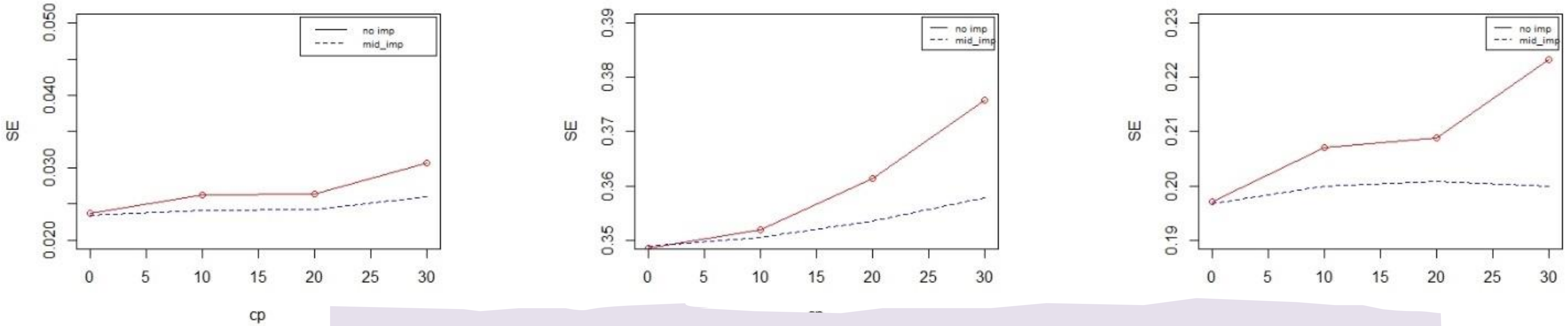


Line plot of SE for each parameter estimate at cp=0

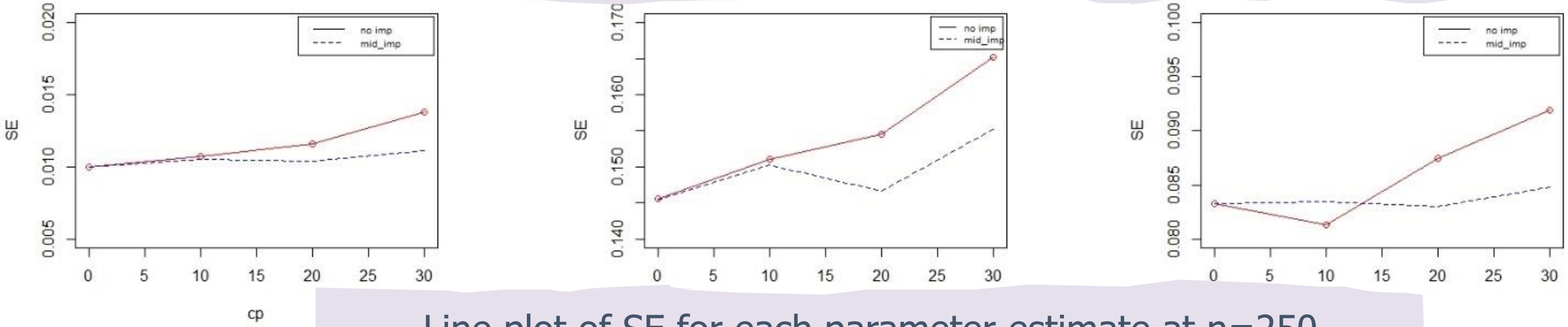


Line plot of SE for each parameter estimate at cp=30

RESULTS

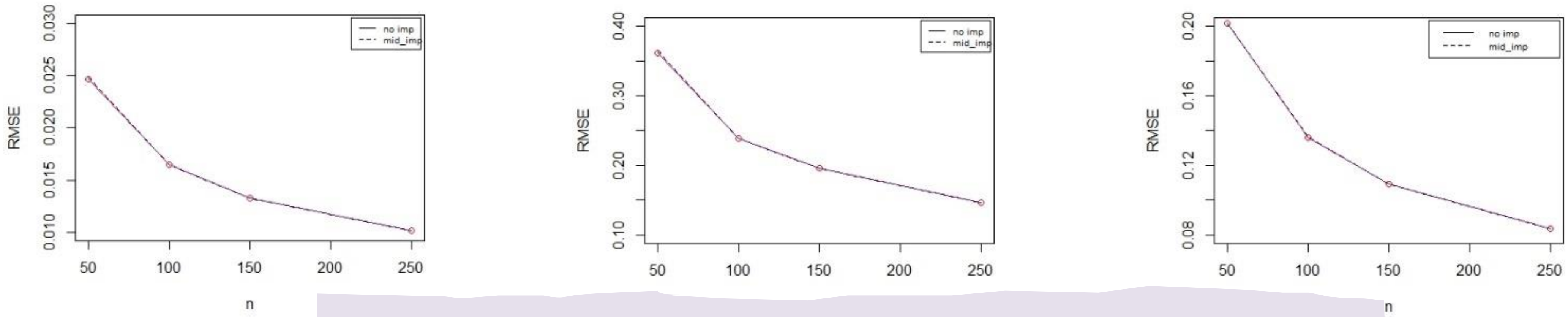


Line plot of SE for each parameter estimate at n=50

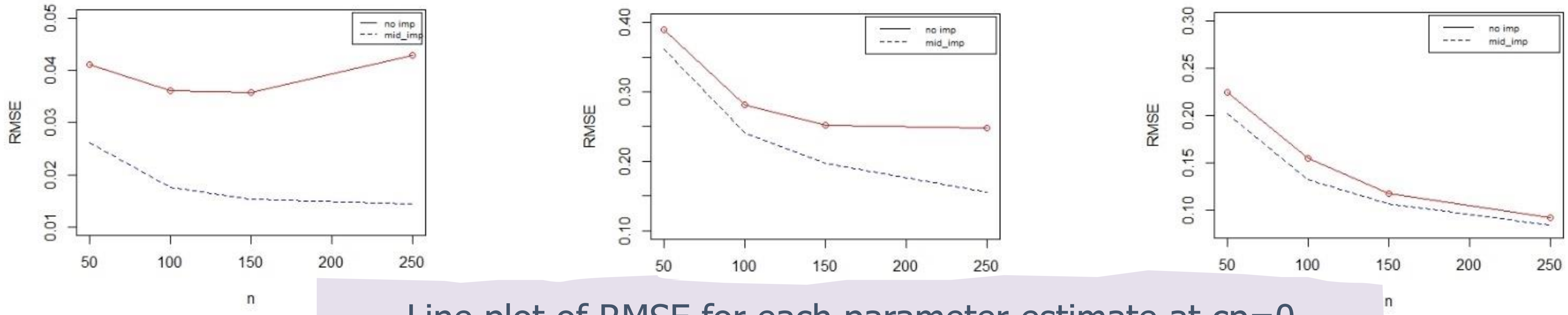


Line plot of SE for each parameter estimate at n=250

RESULTS

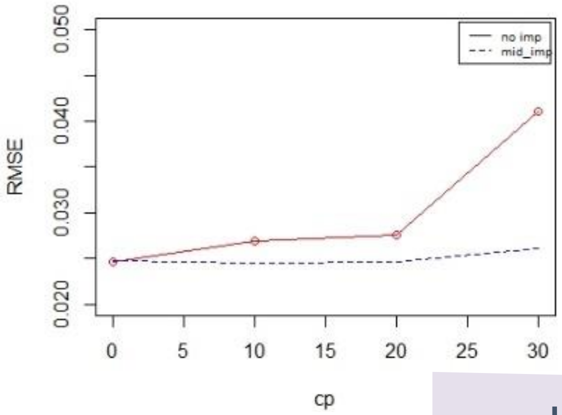


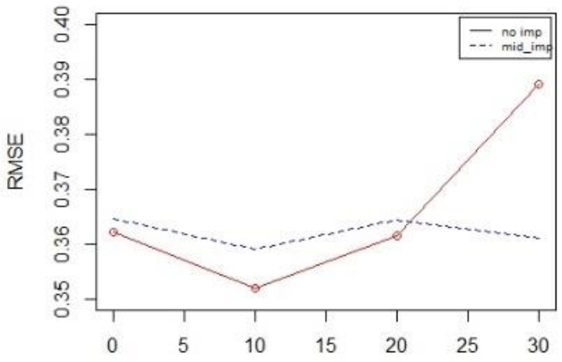
Line plot of RMSE for each parameter estimate at cp=0

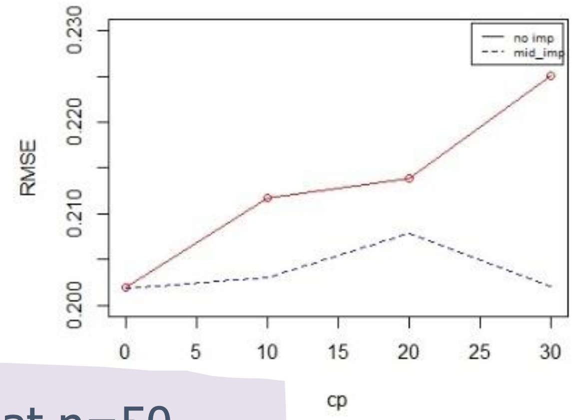


Line plot of RMSE for each parameter estimate at cp=0

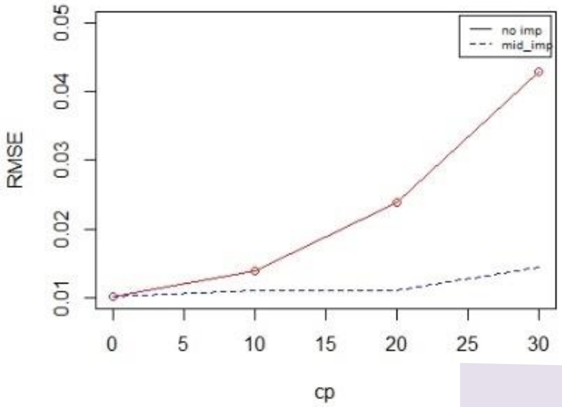
RESULTS

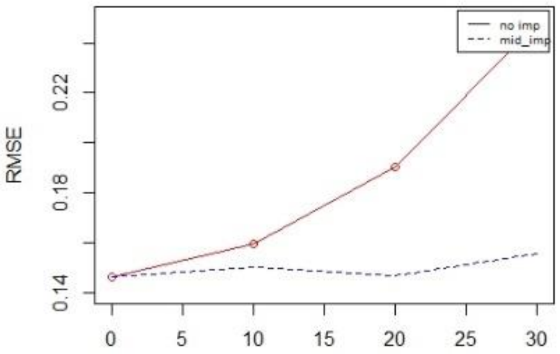


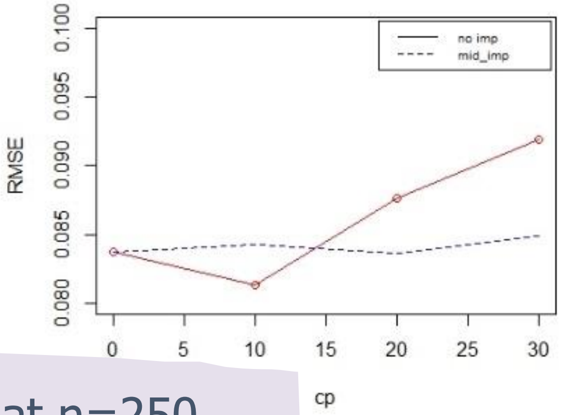




Line plot of RMSE for each parameter estimate at n=50







Line plot of RMSE for each parameter estimate at n=250

REAL DATA ANALYSIS

Breast Cancer Data

Study design	retrospective study by Finkelstein & Wolfe (1985).
Objective	To investigate the comparison between radiation therapy (RT) alone and radiation therapy combined with adjuvant chemotherapy (RCT)
Sample	94 breast cancer patients - Joint Centre for Radiation Therapy in Boston between 1976 and 1980.
Event of interest	time to cosmetic deterioration

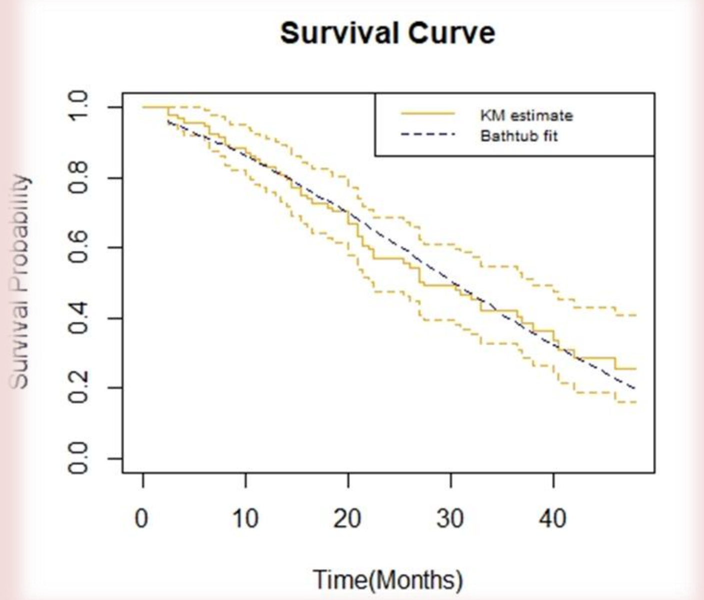
REAL DATA ANALYSIS

59.6%

Interval-censored

40.4%

Right-censored

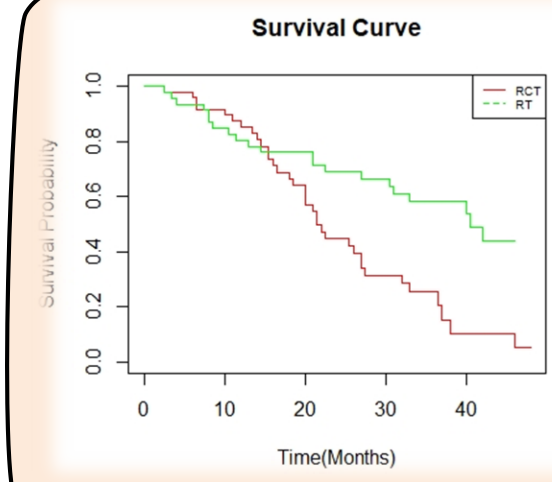


$\hat{\beta}_0$

REAL DATA ANALYSIS

Summary of maximum likelihood estimates for bathtub hazard model with and without covariate

	Parameters	Estimate (Standard error)	t value	Pr(>t)
Without covariate	$\hat{\alpha}$	0.4049(0.0216)	18.921	0.000
	$\hat{\lambda}$	0.0123(0.0043)	2.857	0.00427
With covariate	$\hat{\alpha}$	0.4255(0.0218)	19.565	0.000
	$\hat{\beta}_0$	4.1756(0.3548)	11.769	0.000
	$\hat{\beta}_1$	0.9585 (0.2821)	3.397	0.00068



$$H_0 : \beta_1 = 0$$
$$H_1 : \beta_1 \neq 0$$

Descriptive Statistics of time to cosmetic deterioration by treatment

Treatment	n	Mean	Standard error	Standard deviation	Median
RT	46	28.11	2.10	14.23	31.50
RCT	48	21.66	1.53	10.58	20.50

CONCLUSION

CONCLUSION

Simulation study

midpoint imputation outperforms since it obtains smaller SE and RMSE values for most cp levels and sample sizes.

Real data application

- bathtub hazard model shows a good fit to the data
- the results indicate that the treatment received by breast cancer patients has a significant effect on time to development of cosmetic deterioration.
- patients receiving RCT have a higher risk of developing deterioration
- Beadle et al. (1984) -adjuvant chemotherapy tends to increase development of breast retraction

CONCLUSION

Future Work

- Include other imputation approach

Conclusion

- This study provides support with the notion that the lifetime model with bathtub-shaped failure rates has an important contribution in medical study
- Demonstrates how data modelling could be undertaken considering the presence of interval-censored data.

THANK YOU

idarismail@gmail.com

StatsMalaysia
www.dosm.gov.my