# PARAMETRIC SURVIVAL MODELS WITH INTERVAL CENSORED DATA IN DETERMINING PROGNOSTIC FACTORS OF PATIENTS OF LUNG CANCER

SITI AFIQAH BINTI MUHAMAD JAMIL

A thesis submitted in
fulfillment of the requirement for the award of the
Doctor of Philosophy in Sciences



Faculty of Applied Sciences and Technology
Universiti Tun Husein Onn Malaysia

OCTOBER 2020

Dedicated to my beloved father and mother,
Muhamad Jamil Bin Rosdi and Kalsom Binti Husin for their endless moral, financial support from the beginning until the end of my study.

To my siblings,
Siti Farahia Binti Muhamad Jamil, Siti Zaheera Binti Muhamad Jamil, Siti Halissa Binti Muhamad Jamil, Faris Hazmi Bin Muhamad Jamil, and Faris Hassiff Bin Muhamad Jamil for their greatest support all the way in completing this dissertation.

This work has been a sign of my love towards all of you.

## ACKNOWLEDGEMENT

Bismillahirrahmanirrahim. Praise to the Almighty, Allah S.W.T, for His wisdom and blessing giving me the opportunity to complete this PhD thesis. I would like to express my deepest gratitude and thank you to the following individuals who had supervised, guided and helped me during the preparation of this dissertation.

My parents, Muhamad Jamil Bin Rosdi and Kalsom Binti Husin, my sibling and family who gave great support and spared their time for the success of this dissertation.

My supervisor, Dr. Mohd Asrul Affendi Bin Abdullah, for his support, valued encouragement, advices, comments, suggestions and contribution throughout the completion of this dissertation.

My co-supervisor, Dr. Kek Sie Long, for his great support, continuous cooperation and guided me all the way in completing my dissertation.

My Dean, Deputy Deans of the Universiti Tun Hussein Onn Malaysia, UTHM, for the permission and full cooperation in the process of collecting the data.

Hospital Sultan Ismail (HSI), Johor Bahru, National Malaysian Registry Report (NMRR), Ministry of Health (MOH), National Institutes of Health (NIH) for the permission and full cooperation while collecting the data.

Last but not least, to all lecturers, staff and colleagues in Department of Mathematics and Statistics, Faculty of Applied Sciences and Technology (FAST) for their kind support.

**ABSTRACT**

In clinical trials, biological research and medical studies that involved periodically follow-ups, it is predominently to have censored data. The censored data can be either left, right or interval censored where it reflects on the uncertainty of survival time until an event occur. Survival analysis can accommodates both fixed and time varying covariates with the presence of censored data. The survival time of parametric distribution of Weibull, exponential and log-logistic were derived by using the inverse cumulative distribution function with the hazard and survival function. Standard estimation values such as, the mean square error (MSE), root mean square error (RMSE), mean absolute error (MAE), and mean asolute percentage error (MAPE) were employed in comparing different distributions and number of sample sizes. Besides, Akaike Information Criterion (AIC), Bayesian Information Criterion (BIC), Corrected Akaike Information Criterion ($AIC_C$) been evaluated in finding the best fit model towards the survival time of lung cancer. Thus, the exponential model was found to be the most reliable with censored and fixed covariate of simulation and lung cancer data while the log-logistic appeared to be practically more stable than Weibull in estimating the censored with varying effect covariate. Meanwhile, the prognostic factors that were significant involved the types of lung cancer, gender and some other interactions.  Somehow, increased number of sample either in simulation or bootstrap makes the results to be approximately more reliable as the biases decreased.

# ABSTRAK

Dalam ujian klinikal, penyelidikan biologi and kajian perubatan yang melibatkan tindakan lanjut secara berkala, ia mempunyai peluang yang lebih besar untuk mendapat data yang disensor. Sensor data sama ada kiri, kanan atau selang ditapis dimana ia mencerminkan ketidakpastian masa hidup sehingga berlaku kejadian. Analisis survival boleh menampung kovariat tetap dan masa yang berbeza dengan kehadiran data yang sensor atau data yang ditapis. Masa kelangsungan parametrik Weibull, exponential and log-logistik diperoleh dengan menggunakan kaedah fungsi pengedaran kumulatif songsang dengan fungsi bahaya dan kelangsungan hidup. Nilai-nilai seperti min kesilapan (MSE), min akar kesilapan (RMSE), kesilapan ralat mutlak (MAE), kesilapan peratusan absolute (MAPE) digunakan untuk membandingkan sensor data dengan bilangan saiz sampel yang berbeza. Selain itu, Kriteria Maklumat Akaike (AIC), Kriteria Maklumat Bayesian (BIC), Kriteria Maklumat Akaike (AICC) yang diperbetulkan telah diperiksa untuk mencari model yang sesuai untuk kanser paru-paru. Oleh itu, model eksponen didapati paling dipercayai dengan kovariat yang disensor, dan tetap bagi data simulasi dan data kanser paru-paru manakala log logistik kelihatan lebih stabil daripada Weibull dalam menganggarkan yang disensor dengan pelbagai kesan kovariat. Sementara itu, faktor prognostik yang penting melibatkan jenis paru-paru, jantina dan beberapa interaksi lain antara kovariat. Tambahan lagi, peningkatan saiz sampel dalam simulasi dan bootstrap menjadikan keputusan analisis lebih baik apabila bias atau anggaran kesilapan kesilapan semakin menurun.

## CONTENTS

# LIST OF TABLES

# LIST OF FIGURES

# LIST OF SYMBOLS

| | |
|---|---|
| $\alpha$ | - Alpha (Maximum rate) |
| $\beta$ | - Beta (minimum rate) |
| $\theta$ | - Theta |
| $a$ | - Shape parameter |
| $e$ | - Estimation error ($Y_i - Y_i$) |
| $b$ | - Scale parameter |
| $\lambda$ | - Lambda |
| $n$ | - Number of sample |
| $N$ | - Number of population |
| $NaN$ | - Not a number |
| $Inf$ | - Infinity |
| $\rho$ | - Correlation coefficient |
| $\phi$ | - Degrees of freedom |
| $\ell$ | - Log-likelihood function |
| $\delta$ | - Delta |
| $\Sigma$ | - Sum |
| $L_i$ | - Left censored |
| $R_i$ | - Right censored |
| $I_i$ | - Interval censored |
| $\delta_i$ | - Binary form for left and right censored |
| $\rho_i$ | - Binary form for interval censored |
| $t_i$ | - Time of survival |
| $T_i$ | - Survival time |
| $\propto$ | - Proportional to |
| $DL_s$ | - Detection limits |
| $h(t)$ | - Hazard function |

| | |
|---|---|
| $s(t)$ | - Survival function |
| $H(t)$ | - Cumulative hazard $(\Lambda(t))$ |
| $F(t)$ | - Probability density function |
| $L(a,b)$ | - Likelihood function |
| $HbA^{1c}$ | - Glycated haemoglobin |

# LIST OF ABBREVIATIONS

*AEL* - Adjusted empirical likelihood
*AFT* - Akaike Failure Time
*AIC* - Akaike Information Criterion
$AIC_C$ - Corrected Akaike Information Criterion
*AIDS* - Acquired Immune Deficiency Syndrome
*ASR* - Age Standardised Incidence Rates
*BIC* - Bayesian Information Criterion
*BMI* - Body Mass Index
*CCC* - Concordance correlation coefficient
*CCRT* - Chemoradiotherapy
*CMV* - Cytomegalovirus infection
*CMLE* - Consistent maximum likelihood estimator
*DGP* - Data generating process
*EM* - Expectation maximisation algorithm
*ESLD* - End-stage liver disease
*EL* - Empirical likelihood
*FC* - Fixed covariate
*GLM* - Generalised Linear Model
*HIV* - Human immunodeficiency virus
*HDI* - Human development index
*ICER* - In-cost effectiveness ratio
*IARC* - International Agency for Research on Cancer
*KM* - Kaplan-Meier
*LC* - Lung cancer
*LOD* - Limit of detection
*LRT* - Likelihood ratio test
*MAKNA* - National cancer society in Malaysia

*MLE* - Maximum Likelihood Estimation
*MAE* - Mean absolute error
*MAPE* - Mean absolute percentage error
*MBE* - Mean bias error
*MSE* - Mean square error
*MSMLE* - Modified semiparametric maximum likelihood estimator
*MICE* - Multiple imputation chained equation
*MIDMLE* - Midpoint maximum likelihood estimator
*MSV* - Model selection variance
*NA* - Nelson-aelan
*NCR* - National cancer registry
*NRIM* - Newton Raphson iterative method
*NSCLC* - Non-small cell lung cancer
*NPMLE* - Non-parametric maximum likelihood estimator
NBFH - Non-parametric Behrens- Fisher hypothesis
*ODS* - Outcome dependent sampling
*PH* - Proportional hazard
*PIC* - Partly-interval censored data
*PAH* - Polycyclic aromatic hydrocarbon
*PO* - Proportional odds
*QOL* - Quality of life
*RMSE* - Root mean square error
*RNA* - Ribonucleic acid
*RM* - Reconstruction method
*SA* - Sensitivity analysis
*ST* - Score test
*SD* - State dependent
*SCLC* - Small cell lung cancer
*SE* - Standard error
*TVC* - Time varying covariate
WLR - Weighted log rank
WKM - Weighted Kaplan-Meier
*VTSABD* - Virginia Twin Study of Adolescent Behaviour
*XRF* - X-ray fluorescence

# LIST OF APPENDICES

# LIST OF PUBLICATIONS

**Year 2016:**

1. Abdullah, M. A. A., Jamil, S. A. M., & Jalal, Y. T. M. T. (2016, October). Modelling count data: An application to a breast cancer data in Malaysia. In *AIP Conference Proceedings* (Vol. 1775, No. 1, p. 030003). AIP Publishing.
2. Abdullah, M. A. A., Naing, N. N., & Jamil, S. A. M. (2016, October). Mixture statistical modelling for predicting mortality human immunodeficiency virus (HIV) and tuberculosis (TB) infection patients. In *AIP Conference Proceedings* (Vol. 1775, No. 1, p. 030004). AIP Publishing.
3. Abdullah, M., Jamil, M., Afiqah, S., Wahid, A., Athirah, N., Zaimi, E. M., ... & Abd Ghaffar, U. F. (2016). Annual Wholesale Distributor Spending on Products. *Advanced Science Letters*, *22*(12), 4059-4064.

**Year 2017:**

4. Jamil, S. A. M., Abdullah, M. A. A., Kek, S. L., Nor, M. E., Mohamed, M., & Ismail, N. (2017, September). Detecting overdispersion in count data: A zero-inflated Poisson regression analysis. In *Journal of Physics: Conference Series* (Vol. 890, No. 1, p. 012170). IOP Publishing.
5. Jamil, S. A. M., Abdullah, M. A. A., Kek, S. L., Nor, M. E., Mohamed, M., & Ismail, N. (2017, September). Analysing count data of Butterflies communities in Jasin, Melaka: A Poisson regression analysis. In *Journal of Physics: Conference Series* (Vol. 890, No. 1, p. 012171). IOP Publishing.
6. Jamil, S. A. M., Abdullah, M. A. A., Kek, S. L., Olaniran, O. R., & Amran, S. E. (2017, September). Simulation of parametric model towards the fixed covariate of right censored lung cancer data. In *Journal of Physics: Conference Series* (Vol. 890, No. 1, p. 012172). IOP Publishing.

7. Amran, S. E., Abdullah, M. A. A., Kek, S. L., & Jamil, S. A. M. (2017, September). Analysis of survival in breast cancer patients by using different parametric models. In *Journal of Physics: Conference Series* (Vol. 890, No. 1, p. 012169). IOP Publishing.

**Year 2018:**

8. Abdullah, M. A. A., Jamil, S. A. M., Kormin, F., & Mamat, M. (2018). Comparing Generalized Linear Model of Count Data: Application towards Pteridophyta species. *International Journal of Engineering & Technology*, *7*(3.28), 58-63.
9. Mohamad, N., Che Him, N., Rusiman, M. S., Sufahani, S., & Jamil, S. A. M. (2018). Application FCM in Modelling DIR for Selangor Using Negative Binomial GAM. *International Journal of Engineering & Technology*, *7*(4.30), 1-4.
10. Afiqah Muhamad Jamil, S., Asrul Affendi Abdullah, M., Sie Long, K., Fazilla Mohd Jupri, N., & Mamat, M. (2018). A Stepwise Logistic Regression Analysis: An application toward Poultry Farm Data in Johor. *International Journal of Engineering & Technology*, *7*(3.28), 68-71. doi:http://dx.doi.org/10.14419/ijet.v7i3.28.20968

**Year 2020:**

11. Abdullah, M. A. A., Zaimi, E. M., & Jamil, S. A. M. Parametric Frailty Model with Time-Dependent Covariates. **(IJRTE)**

**Year 2020: (Work in progress)**

12. Analysis of Parametric Survival Models in right censored data using simulation procedure.
13. Estimating Pteridophyta species diversity of abundance and occupancy by modelling zero-inflated count data.
14. Parametric Weibull Time-Varying Covariate Model for HIV-TB Mortality.

# REFERENCES


Aboueissa, A. E.-M. A. and Stoline, M. R. (2006). Maximum likelihood estimators of population parameters from doubly left-censored samples. *Environmetrics*. 17(8), 811–826.

Akhtar-Danesh, N. and Finley, C. (2015). Temporal trends in the incidence and relative survival of non-small cell lung cancer in Canada: A population-based study. *Lung Cancer*. 90(1), 8–14.

Alharpy, A. M. and Ibrahim, N. A. (2013). Parametric tests for partly interval-censored failure time data under Weibull distribution via multiple imputation. *Journal of Applied Sciences.* 13(4), 621.

Alharpy, A. M. and Ibrahim, N. A. (2014). A two-sample parametric test for partly interval-censored data with nonproportional hazard. *Mathematical Problems in Engineering.*

Ali, Z., Hosseini, M., Mahmoodi, M., Mohammad, K., Zeraati, H. and Naieni, K. H. (2015). A Comparison between Accelerated Failure-time and Cox Proportional Hazard Models in Analyzing the Survival of Gastric Cancer Patients. *Iranian journal of public health.* 44(8), 1095.

Allison, P. D. (1995). Survival analysis using SAS: A practical guide. Cary, NC: SAS Institute.

Al Tamimi, T. M., Al-Bar, A., Al-Suhaimi, S., Ibrahim, E., lbrahim, A., Wosornu, L., & Gabriel, G. S. (1996). Lung cancer in the eastern region of Saudi Arabia: a population-based study. *Annals of Saudi medicine*, *16*, 3-11.

Alvarez, F., Lippi, F., & Passadore, J. (2017). Are State-and Time-Dependent Models Really Different?. *NBER Macroeconomics Annual*, *31*(1), 379-457.

Amirsehi, K. (1997). Testing for a time-dependent covariate effect in the linear risk model.

Aridgides, P. D., Janik, A., Bogart, J. A., Duffy, S., Rosenbaum, P. and Gajra, A. (2013). Radiotherapy for Stage III Non–Small-Cell Lung Carcinoma in the Elderly (Age 70 years). *Clinical lung cancer*. 14(6), 674–679.

Aryal, S., Bhaumik, D. K., Santra, S. and Gibbons, R. D. (2009). Confidence interval 103 for random-effects calibration curves with left-censored data. *Environmetrics*. 20(2), 181–189.

Atem, F. D., Qian, J., Maye, J. E., Johnson, K. A., & Betensky, R. A. (2017). Linear regression with a randomly censored covariate: application to an Alzheimer's study. *Journal of the Royal Statistical Society: Series C (Applied Statistics)*, *66*(2), 313-328.

Ata, N., & Sözer, M. T. (2007). Cox regression models with nonproportional hazards applied to lung cancer survival data. *Hacettepe Journal of Mathematics and Statistics*, *36*(2), 157-167.

Athreya, K. B. (1987). Bootstrap of the mean in the infinite variance case. *The Annals of Statistics*, 724-731.

Austin, P. C. (2012). Generating survival times to simulate Cox proportional hazards models with time-varying covariates. Statistics in medicine. 31(29), 3946–3958.

Ayabe, T., Tomita, M., Asada, T., Tsuchiya, K., Nemoto, M., & Nakamura, K. (2017). Patient-Reported Outcomes of Chemotherapy Involving Non-Small Cell Lung Cancer: Evaluation by Questionnaires of Quality of Life Regarding Anti-Aging and Anti-Cancer Drugs. *Advances in Lung Cancer*, *6*(02), 13.

Bach, P. B. (2009). Smoking as a factor in causing lung cancer. *JAMA*, *301*(5), 539-541.

Baghestani, A. R., Daneshva, T., Pourhoseingholi, M. A. and Asadzadeh, H. (2016). Survival of Colorectal Cancer in the Presence of Competing-Risks-Modeling by Weibull Distribution. *Asian Pacific journal of cancer prevention:* APJCP. 17(3), 1193–1196.

Baghestani, A. R., Gohari, M. R., Orooji, A., Pourhoseingholi, M. A. and Zali, M. R. (2015). Evaluation of parametric models by the prediction error in colorectal cancer survival analysis. *Gastroenterology and Hepatology from bed to bench.* 8(3), 183.

Balakrishnan, N. and Mitra, D. (2012). Left truncated and right censored Weibull data and likelihood inference with an illustration. *Computational Statistics & Data Analysis*. 56(12), 4011–4025.

Barton, M. B., Frommer, M., & Shafiq, J. (2006). Role of radiotherapy in cancer control in low-income and middle-income countries. *The lancet oncology*, *7*(7), 584-595.

Barker, N. (2005). A practical introduction to the bootstrap using the SAS system. In *SAS Conference Proceedings: Phuse 2005: October 10-12 2005; Heidelberg, Germany SAS*.

Behera, M., Fowler, E. E., Owonikoko, T. K., Land, W. H., Mayfield, W., Chen, Z., ... & Heine, J. J. (2011). Statistical learning methods as a preprocessing step for survival analysis: evaluation of concept using lung cancer data. *Biomedical engineering online*, *10*(1), 97.

Bewick, V., Cheek, L. and Ball, J. (2004). Statistics review 12: survival analysis. *Critical care*. 8(5), 389.

Betensky, R. A., Lindsey, J. C., Ryan, L. M., & Wand, M. P. (1999). Local EM estimation of the hazard function for interval-censored data. *Biometrics*, *55*(1), 238-245.

Beyaztas, U., & Alin, A. (2013). Jackknife-after-bootstrap method for detection of influential observations in linear regression models. *Communications in Statistics-Simulation and Computation*, *42*(6), 1256-1267.

Beyaztas, U., & Alin, A. (2014). Sufficient jackknife-after-bootstrap method for detection of influential observations in linear regression models. *Statistical Papers*, *55*(4), 1001-1018.

Bharathi, H., & Arulananth, T. S. (2017). A review of lung cancer prediction system using data mining techniques and self-organizing map (SOM). *International Journal of Applied Engineering Research*, *12*(10), 2190-2195.

Boruvka, A., & Cook, R. J. (2015). A Cox-Aalen Model for Interval-censored Data. *Scandinavian Journal of Statistics*, *42*(2), 414-426.

Biau, D. J., Kernéis, S., & Porcher, R. (2008). Statistics in brief: the importance of sample size in the planning and interpretation of medical research. *Clinical orthopaedics and related research*, *466*(9), 2282-2288.

Bray, F., & Piñeros, M. (2016). Cancer patterns, trends and projections in Latin America and the Caribbean: a global context. *salud pública de méxico*, *58*(2), 104-117.

Brendel, M., Janssen, A., Mayer, C.-D. and Pauly, M. (2014). Weighted log rank permutation tests for randomly right censored life science data. *Scandinavian Journal of Statistics*. 41(3), 742–761.

Briggs, A. H., Weinstein, M. C., Fenwick, E. A., Karnon, J., Sculpher, M. J., Paltiel, A. D., Force, I.-S. M. G. R. P. T. et al. (2012). Model parameter estimation and uncertainty: a report of the ISPOR-SMDM Modeling Good Research Practices Task Force-6. *Value in Health.* 15(6), 835–842.104

Brown, M. L., Goldie, S. J., Draisma, G., Harford, J., & Lipscomb, J. (2006). Health service interventions for cancer control in developing countries. *Disease control priorities in developing countries*, *2*, 569-589.

Burnham, K. P. and Anderson, D. R. (2003). Model selection and multimodel inference: a practical information-theoretic approach. *Springer Science & Business Media*.

Boos, D. D., & Brownie, C. (1989). Bootstrap methods for testing homogeneity of variances. *Technometrics*, *31*(1), 69-82.

Bose, A., & Chatterjee, S. (2002). Comparison of bootstrap and jackknife variance estimators in linear regression: second order results.

Burr, D. (1994). A comparison of certain bootstrap confidence intervals in the Cox model. *Journal of the American Statistical Association*, *89*(428), 1290-1302.

Carrasco, J. M., Ortega, E. M., & Paula, G. A. (2008). Log-modified Weibull regression models with censored data: Sensitivity and residual analysis. *Computational Statistics & Data Analysis*, *52*(8), 4021-4039.

C¸ etiny¨urek Yavuz, A. and Lambert, P. (2011). Smooth estimation of survival functions and hazard ratios from interval-censored data using Bayesian penalized Bsplines. *Statistics in medicine.* 30(1), 75–90.

Chen, C.-M. and Shen, P.-s. (2017). Semiparametric regression analysis of failure time data with dependent interval censoring. *Statistics in medicine*. 36(21), 3398–3411.

Chen, Q., May, R. C., Ibrahim, J. G., Chu, H. and Cole, S. R. (2014). Joint modelling of longitudinal and survival data with missing and left-censored time-varying covariates. *Statistics in medicine*. 33(26), 4560–4576.

Cheng, T. Y. D., Cramb, S. M., Baade, P. D., Youlden, D. R., Nwogu, C., & Reid, M. E. (2016). The international epidemiology of lung cancer: latest trends, disparities, and tumor characteristics. *Journal of Thoracic Oncology*, *11*(10), 1653-1671.

Chong-Kin, L. (2014). Lung Cancer in Malaysia–Epidemiology and Management Practices.

Chang, M. N., & Yang, G. L. (1987). Strong consistency of a nonparametric estimator of the survival function with doubly censored data. *The Annals of Statistics*, *15*(4), 1536-1547.

Chang, C. C., & Keisler, H. J. (1990). *Model theory* (Vol. 73). Elsevier.

Chakraborti, S., & Van der Laan, P. (1997). An overview of precedence-type tests for censored data. *Biometrical Journal*, *39*(1), 99-116.

Chen, C. H., & George, S. L. (1985). The bootstrap and identification of prognostic factors via Cox's proportional hazards regression model. *Statistics in medicine*, *4*(1), 39-46.

Claeskens, G., Hjort, N. L. et al. (2008). Model selection and model averaging. Cambridge Books.

Cohen, A. C. (1965). Maximum likelihood estimation in the Weibull distribution based on complete and on censored samples. *Technometrics*. 7(4), 579–588.

Collett, D. (2015). Modelling survival data in medical research. CRC press. Cox, D. R. and Oakes, D. (1984). *Analysis of survival data*. vol. 21. CRC Press.

Cook, R. J., Zeng, L., & Lee, K. A. (2008). A multistate model for bivariate interval-censored failure time data. *Biometrics*, *64*(4), 1100-1109.

Cramer III, R. D., Bunce, J. D., Patterson, D. E., & Frank, I. E. (1988). Crossvalidation, bootstrapping, and partial least squares compared with multiple regression in conventional QSAR studies. *Quantitative Structure‐Activity Relationships*, *7*(1), 18-25.

Cufari, M. E., Proli, C., Phull, M., Raubenheimer, H., Al Sahaf, M., Asadi, N., Perikleous, P., Allan, A., Shedden, L., Chavan, H. et al. (2015). Increasing Incidence of Non-Smoking Lung Cancer: Presentation of Patients with Early Disease to a Tertiary Institution in the UK. In *JOURNAL OF THORACIC ONCOLOGY*, vol. 10. ELSEVIER SCIENCE INC 360 PARK AVE SOUTH, NEW YORK, NY 10010-1710 USA, S793–S794.

Curado, M. P., & de Souza, D. L. B. (2014). Cancer burden in Latin America and the Caribbean. *Annals of global health*, *80*(5), 370-377.

Danaei, G., Vander Hoorn, S., Lopez, A. D., Murray, C. J., Ezzati, M., & Comparative Risk Assessment collaborating group (Cancers. (2005). Causes of cancer in the world: comparative risk assessment of nine behavioural and environmental risk factors. *The Lancet*, *366*(9499), 1784-1793.

Daniel, R. M., Cousens, S. N., De Stavola, B. L., Kenward, M. G., & Sterne, J. A. C. (2013). Methods for dealing with time-dependent confounding. *Statistics in medicine*, *32*(9), 1584-1618.

Denne, J. S., Stone, A. M., Bailey-Iacona, R., & Chen, T. T. (2013). Missing data and censoring in the analysis of progression-free survival in oncology clinical trials. *Journal of biopharmaceutical statistics*, *23*(5), 951-970.

De Gruttola, V., & Lagakos, S. W. (1989). Analysis of doubly-censored survival data, with application to AIDS. *Biometrics*, 1-11.

Didkowska, J., Wojciechowska, U., Mańczuk, M., & Łobaszewski, J. (2016). Lung cancer epidemiology: contemporary and future challenges worldwide. *Annals of translational medicine*, *4*(8).

Dudani, S., Leighl, N. B., Ho, C., Pantarotto, J. R., Zhu, X., Zhang, T. and Wheatley-Price, P. (2016). Approach to the non-operative management of patients with stage II non-small cell lung cancer (NSCLC): A survey of Canadian medical and radiation oncologists. *Lung Cancer*. 94, 74–80. 105.

Dugué, A. E., Pulido, M., Chabaud, S., Belin, L., & Gal, J. (2016). How to deal with interval-censored data practically while assessing the progression-free survival: a step-by-step guide using SAS and R software. *Clinical Cancer Research*, *22*(23), 5629-5635.

e Castro, A. T., Clemente, J., Carvalho, L., Freitas, S. and Cemlyn-Jones, J. (2016). Small-cell lung cancer in never-smokers: A case series. *Lung Cancer*. 93, 82–87.

Eberle, A., Jansen, L., Castro, F., Krilaviciute, A., Luttmann, S., Emrich, K., Holleczek, B., Nennecke, A., Katalinic, A., Brenner, H. et al. (2015). Lung cancer survival in Germany: A population-based analysis of 132,612 lung cancer patients. *Lung Cancer*. 90(3), 528–533.

Efron, B. (1979). Bootstrap methods: another look at the jackknife annals of statistics 7: 1–26. *View Article PubMed/NCBI Google Scholar*, *24*.

Efron, B. (1981). Nonparametric estimates of standard error: the jackknife, the bootstrap and other methods. *Biometrika*, *68*(3), 589-599.

Efron, B., & Gong, G. (1983). A leisurely look at the bootstrap, the jackknife, and cross-validation. *The American Statistician*, *37*(1), 36-48.

Efron, B. (1992). Bootstrap methods: another look at the jackknife. In *Breakthroughs in statistics* (pp. 569-593). Springer, New York, NY.

Efron, B. (2003). Second thoughts on the bootstrap. *Statistical Science*, *18*(2), 135-140.

Elfaki, F. A. M., Abobakar, A., Azram, M. and Usman, M. (2013). Survival model for partly interval-censored data with application to anti D in rhesus D negative studies. *World Academy of Science, Engineering and Technology*. 7, 899–902.

Elfaki, F. A. M., Azram, M. and Usman, M. (2012). Parametric Cox's Model for Partly Interval-Censored Data with Application to AIDS Studies. *International Journal of Applied Physics and Mathematics*. 2(5), 352.

Engel, J. (2010, July). On teaching bootstrap confidence intervals. In *Proceedings of the Eighth International Conference on Teaching Statistics. Voorburg, The Netherlands: International Statistical Institute*.

Ezzati, M., Lopez, A. D., Rodgers, A., Vander Hoorn, S., Murray, C. J., & Comparative Risk Assessment Collaborating Group. (2002). Selected major risk factors and global and regional burden of disease. *The Lancet*, *360*(9343), 1347-1360.

Farrington, C. P. (1996). Interval censored survival data: A generalized linear modelling approach. *Statistics in medicine*, *15*(3), 283-292.

Falcaro, M., & Pickles, A. (2007). A flexible model for multivariate interval-censored survival times with complex correlation structure. *Statistics in medicine*, *26*(3), 663-680.

Fazeli, Z., Pourhoseingholi, M. A., Vahedi, M., Abadi, A., Reza, A. and Baghestani, M. R. Z. (2015). Volume: 4: Issue-5: May-2015 ISSN: 2278-0246 Coden: IJAPBS www. ijapbs. com *TREND ANALYSIS OF LUNG CANCER MORTALITY IN IRAN*, 1995-2004.

Ferlay, J., Soerjomataram, I., Ervik, M., Dikshit, R., Eser, S., Mathers, C., Rebelo, M., Parkin, D., Forman, D. and Bray, F. (2018). GLOBOCAN 2012 v1. 0, Cancer Incidence and Mortality Worldwide: IARC Cancer Base No. 11. 2014. Available from: globocan. iarc. fr.

Feng, Y., Duan, R., & Sun, J. (2017). Nonparametric comparison of survival functions based on interval-censored data with unequal censoring. *Statistics in medicine*, *36*(12), 1895-1906.

Feng, Y., & Sun, J. (2014). Hazard ratio estimation for two-sample case under interval-censored failure time data. *Biometrical Journal*, *56*(2), 219-229.

Frydman, H., Gerds, T., Grøn, R. and Keiding, N. (2013). Nonparametric estimation in an illness-death model when all transition times are interval censored. *Biometrical Journal*. 55(6), 823–843. 106

Fradkin, D., Schneider, D., & Muchnik, I. (2006). Machine learning methods in the analysis of lung cancer survival data. *DIMACS Technical Report 2005–35*.

Fisher, L. D., & Lin, D. Y. (1999). Time-dependent covariates in the Cox proportional-hazards regression model. *Annual review of public health*, *20*(1), 145-157.

Filipovic, N., Sustersic, T., Vulovic, A., & Tsuda, A. (2019). Big Data and machine learning: new frontier in lung cancer care.

Foucher, Y., Giral, M., Soulillou, J. P., & Daures, J. P. (2007). A semi-Markov model for multistate and interval-censored data with multiple terminal events. Application in renal transplantation. *Statistics in medicine*, *26*(30), 5381-5393.

Fu, W., & Simonoff, J. S. (2017). Survival trees for interval-censored survival data. *Statistics in medicine*, *36*(30), 4831-4842.

Furrukh, M. (2013). Tobacco smoking and lung cancer: perception-changing facts. *Sultan Qaboos University Medical Journal*, *13*(3), 345.

Fukuoka, M., Wu, Y. L., Thongprasert, S., Sunpaweravong, P., Leong, S. S., Sriuranpong, V., ... & Duffield, E. L. (2011). Biomarker analyses and final overall survival results from a phase III, randomized, open-label, first-line study of gefitinib versus carboplatin/paclitaxel in clinically selected patients with advanced non–small-cell lung cancer in Asia (IPASS). *Journal of clinical oncology*, *29*(21), 2866-2874.

Fujimoto, D., Shimizu, R., Morimoto, T., Kato, R., Sato, Y., Kogo, M., ... & Nakagawa, A. (2015). Analysis of advanced lung cancer patients diagnosed following emergency admission. *European Respiratory Journal*, *45*(4), 1098-1107.

Gao, F., Zeng, D. and Lin, D.-Y. (2017). Semiparametric estimation of the accelerated failure time model with partly interval-censored data. *Biometrics*. 73(4), 1161–1168.

George, B., Seals, S. and Aban, I. (2014). Survival analysis and regression models. *Journal of Nuclear Cardiology*. 21(4), 686–694.

Geskus, R. B. (2011). Cause-specific cumulative incidence estimation and the fine and gray model under both left truncation and right censoring. *Biometrics*, *67*(1), 39-49.

Gerds, T. A., Kattan, M. W., Schumacher, M., & Yu, C. (2013). Estimating a time-dependent concordance index for survival prediction models with covariate dependent censoring. *Statistics in Medicine*, *32*(13), 2173-2184.

Gijbels, I. (2010). Censored data. *Wiley Interdisciplinary Reviews: Computational Statistics*. 2(2), 178–188.

Glynn, R. J., & Rosner, B. (2004). Multiple imputation to estimate the association between eyes in disease progression with interval-censored data. *Statistics in medicine*, *23*(21), 3307-3318.

Gu, M. G., & Zhang, C. H. (1993). Asymptotic properties of self-consistent estimators based on doubly censored data. *The Annals of Statistics*, 611-624.

Goetghebeur, E., & Ryan, L. (2000). Semiparametric regression analysis of interval-censored data. *Biometrics*, *56*(4), 1139-1144.

Goggins, W. B., Finkelstein, D. M. and Zaslavsky, A. M. (1999). Applying the Cox Proportional Hazards Model When the Change Time of a Binary Time-Varying Covariate Is Interval Censored. *Biometrics*. 55(2), 445–451.

Goodall, R. L., Dunn, D. T., & Babiker, A. G. (2004). Interval-censored survival time data: confidence intervals for the non-parametric survivor function. *Statistics in Medicine*, *23*(7), 1131-1145.

Gómez, G., Calle, M. L., Oller, R., & Langohr, K. (2009). Tutorial on methods for interval-censored data and their implementation in R. *Statistical Modelling*, *9*(4), 259-297.

Gong, Q., & Schaubel, D. E. (2013). Partly conditional estimation of the effect of a time-dependent factor in the presence of dependent censoring. *Biometrics*, *69*(2), 338-347.

Hajjar, R. R., Atli, T., Al-Mandhari, Z., Oudrhiri, M., Balducci, L., & Silbermann, M. (2013). Prevalence of aging population in the Middle East and its implications on cancer incidence and care. *Annals of oncology*, *24*(suppl_7), vii11-vii24.

Hanna, N., Shepherd, F. A., Fossella, F. V., Pereira, J. R., De Marinis, F., Von Pawel, J., ... & Lim, H. L. (2004). Randomized phase III trial of pemetrexed versus docetaxel in patients with non-small-cell lung cancer previously treated with chemotherapy. *Journal of clinical oncology*, *22*(9), 1589-1597.

Hamada, C. (1995). Bootstrap Cox regression using SAS procedures. *SAS Institute Japan Ltd*, *211*, 1258-62.

Harpole, D. H., Herndon, J. E., Young, W. G., Wolfe, W. G. and Sabiston, D. C. (1995). Stage I nonsmall cell lung cancer. A multivariate analysis of treatment methods and patterns of recurrence. *Cancer*. 76(5), 787–796.

Hashemian, A., Garshasbi, M., Pourhoseingholi, M. and Eskandari, S. (2017). A Comparative Study of Cox Regression vs. Log-Logistic Regression (with and without its frailty) in Estimating Survival Time of Patients with Colorectal Cancer. *Journal of Medical and Biomedical Sciences*. 6(1), 35–43.

Hachesu, P. R., Moftian, N., Dehghani, M., & Soltani, T. S. (2017). Analyzing a Lung Cancer Patient Dataset with the Focus on Predicting Survival Rate One Year after Thoracic Surgery. *Asian Pacific journal of cancer prevention: APJCP*, *18*(6), 1531.

Hawkins, N., Scott, D. A., Woods, B. S. and Thatcher, N. (2009). No Study Left Behind: A Network Meta-Analysis in Non–Small-Cell Lung Cancer Demonstrating the Importance of Considering All Relevant Data. *Value in Health*. 12(6), 996–1003.

Henschke, C. I., Yankelevitz, D. F., Libby, D. M., Pasmantier, M. W., Smith, J. P. and Miettinen, O. S. (2006). Survival of patients with stage I lung cancer detected on CT screening. *The New England journal of medicine*. 355(17), 1763–1771.

Hesterberg, T. (1996). Control variates and importance sampling for efficient bootstrap simulations. *Statistics and Computing*, *6*(2), 147-157.

Hesterberg, T. C. (1998). Simulation and bootstrapping for teaching statistics. In *American Statistical Association Proceedings of the Section on Statistical Education* (pp. 44-52). Alexandria, VA: American Statistical Association.

Hesterberg, T. C. (1999). Bootstrap tilting confidence intervals and hypothesis tests. *Hypothesis*, *6*(8), 4-5.

Hesterberg, T. (2011). Bootstrap. *Wiley Interdisciplinary Reviews: Computational Statistics*, *3*(6), 497-526.

Hesterberg, T. C. (2015). What teachers should know about the bootstrap: Resampling in the undergraduate statistics curriculum. *The American Statistician*, *69*(4), 371-386.

Hein, S. E., & Westfall, P. (2004). Improving tests of abnormal returns by bootstrapping the multivariate regression model with event parameters. *Journal of Financial Econometrics*, *2*(3), 451-471.

Higgins, E., Rees, T., Lester, J. and Davies, R. (2016). 48 Do stage I lung cancer patients have lung cancer symptoms? *Lung Cancer*. (91), S17.

Hirsch, F. R., Zaric, B., Rabea, A., Thongprasert, S., Lertprasertsuke, N., Dalurzo, M. L., & Varella-Garcia, M. (2017). Biomarker Testing for Personalized Therapy in Lung Cancer in Low-and Middle-Income Countries. *American Society of Clinical Oncology Educational Book*, *37*, 403-408.

Hosmer Jr, D. W., Lemeshow, S., & May, S. (2008). *Applied survival analysis: regression modeling of time-to-event data* (Vol. 618). Wiley-Interscience.

Hudgens, M. G., Li, C. and Fine, J. P. (2014). Parametric likelihood inference for interval censored competing risks data. *Biometrics*. 70(1), 1–9.

Hutson, A. D. (2002). Analytical bootstrap methods for censored data. *Advances in Decision Sciences*, *6*(2), 129-141.

I Civit, F. U. (1992). Some applications of the bootstrap to survival analysis. *Anuario de psicología/The UB Journal of psychology*, (55), 155-168.

Jemal, A., Bray, F., Center, M. M., Ferlay, J., Ward, E., & Forman, D. (2011). Global cancer statistics. *CA: a cancer journal for clinicians*, *61*(2), 69-90.

Jung, K. J., Jeon, C., & Jee, S. H. (2016). The effect of smoking on lung cancer: ethnic differences and the smoking paradox. *Epidemiology and health*, *38*.

Ji, S., Peng, L., Cheng, Y., & Lai, H. (2012). Quantile regression for doubly censored data. *Biometrics*, *68*(1), 101-112.

Kan, C. and Chan, K. (2016). A Review of Lung Cancer Research in Malaysia. *The Medical journal of Malaysia*. 71(Suppl 1), 70–78. 107

Kasymjanova, G., Kreisman, H., Correa, J. A., Dajczman, E. and Small, D. (2006). Does Granulocyte Colony–Stimulating Factor Affect Survival in Patients with Advanced Non-small Cell Lung Cancer? *Journal of Thoracic Oncology*. 1(6), 564–570.

Kiani, K., Arasan, J. and Midi, H. (2012). Interval estimations for parameters of Gompertz model with time-dependent covariate and right censored data. *Sains Malaysiana*. 41(4), 471–480.

Kim, M. Y., & Xue, X. (2002). The analysis of multivariate interval-censored survival data. *Statistics in Medicine*, *21*(23), 3715-3726.

Kim, J. S. (2003). Maximum likelihood estimation for the proportional hazards model with partly interval-censored data. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*. 65(2), 489–502.

Kim, Y. J., & Jhun, M. (2008). Cure rate model with interval censored data. *Statistics in medicine*, *27*(1), 3-14.

Kim, Y.-J., Cho, H., Kim, J. and Jhun, M. (2010). Median regression model with interval censored data. *Biometrical Journal*. 52(2), 201–208.

Kim, Y., Lim, J., & Park, D. (2015). Testing independence of bivariate interval-censored data using modified Kendall's tau statistic. *Biometrical Journal*, *57*(6), 1131-1145.

Kim, H.-Y., Williamson, J. M. and Lin, H.-M. (2016). Power and sample size calculations for interval-censored survival analysis. *Statistics in medicine*. 35(8), 1390–1400.

Kim, Y., Lee, J., & Kim, J. (2005). Bayesian bootstrap analysis of doubly censored data using Gibbs sampler. *Statistica Sinica*, 969-980.

Kim, M. Y., De Gruttola, V. G., & Lagakos, S. W. (1993). Analyzing doubly censored data with covariates, with application to AIDS. *Biometrics*, 13-22.

Kim, Y., Kim, J., & Jang, W. (2013). An EM algorithm for the proportional hazards model with doubly censored data. *Computational Statistics & Data Analysis*, *57*(1), 41-51.

Kimman, M., Norman, R., Jan, S., Kingston, D., & Woodward, M. (2012). The burden of cancer in member countries of the Association of Southeast Asian Nations (ASEAN). *Asian Pacific journal of cancer prevention*, *13*(2), 411-420.

Ko, Y. C., Cheng, L. S. C., Lee, C. H., Huang, J. J., Huang, M. S., Kao, E. L., ... & Lin, H. J. (2000). Chinese food cooking and lung cancer in women nonsmokers. *American journal of epidemiology*, *151*(2), 140-147.

Kundu, D. and Mitra, D. (2016). Bayesian inference of Weibull distribution based on left truncated and right censored data. *Computational Statistics & Data Analysis*. 99, 38–50.

Kuan, G. L. K., Papapreponis, P., & Hin, Y. M. (2015). State level variation of breast cancer cases in 2007 among Malaysian women. *Malaysian Journal of Public Health Medicine*, *15*(1), 62-68.

Kubota, T., Okano, Y., Sakai, M., Yamane, T., Shiota, N., Ohnishi, H., ... & Ogushi, F. (2014). A Phase II Study of Erlotinib in Patients with Previously Treated Non-Small Cell Lung Cancer. *Advances in Lung Cancer*, *3*(01), 10.

Lagakos, S. W. (1979). General Right Censoring and Its Impact on the Analysis of Survival Data. Biometrics. 35(1), 139–156. ISSN 0006341X, 15410420. Retrievable at http://www.jstor.org/stable/2529941.

Lamb, K. E., Williamson, E. J., Coory, M. and Carlin, J. B. (2015). Bias and precision of measures of survival gain from right-censored data. *Pharmaceutical statistics*. 14(5), 409–417.

Lam, W. K. (2005). Lung cancer in Asian women—the environment and genes. *Respirology*, *10*(4), 408-417.

Lan, Q., He, X., Costa, D. J., Tian, L., Rothman, N., Hu, G., & Mumford, J. L. (2000). Indoor coal combustion emissions, GSTM1 and GSTT1 genotypes, and lung cancer risk: a case-control study in Xuan Wei, China. *Cancer Epidemiology and Prevention Biomarkers*, *9*(6), 605-608.

Lapidus, N., Chevret, S. and Resche-Rigon, M. (2014). Assessing assay agreement estimation for multiple left-censored data: a multiple imputation approach. *Statistics in medicine*. 33(30), 5298–5309. 108.

Lawless, J. F. (2011). *Statistical models and methods for lifetime data* (Vol. 362). John Wiley & Sons.

Lee, E. T., & Wang, J. (2003). *Statistical methods for survival data analysis* (Vol. 476). John Wiley & Sons.

Lee, C. C., & Houghton, P. (2005). Cytotoxicity of plants from Malaysia and Thailand used traditionally to treat cancer. *Journal of ethnopharmacology*, *100*(3), 237-243.

Lee, M., Kong, L. and Weissfeld, L. (2012). Multiple imputation for left-censored biomarker data based on Gibbs sampling method. *Statistics in medicine*. 31(17), 1838–1848.

Lee, Y. C. A., & Hashibe, M. (2014). Tobacco, alcohol, and cancer in low and high income countries. *Annals of global health*, *80*(5), 378-383.

Lee, M. L. T., Whitmore, G. A., & Rosner, B. A. (2010). Threshold regression for survival data with time-varying covariates. *Statistics in medicine*, *29*(7-8), 896-905.

Li, L., Watkins, T., & Yu, Q. (1997). An EM algorithm for smoothing the self-consistent estimator of survival functions with interval-censored data. *Scandinavian Journal of Statistics*, *24*(4), 531-542.

Liao, X., Zucker, D. M., Li, Y. and Spiegelman, D. (2011). Survival Analysis with Error-Prone Time-Varying Covariates: A Risk Set Calibration Approach. *Biometrics*. 67(1), 50–58.

Lim, G. C. C. (2002). Overview of cancer in Malaysia. *Japanese Journal of Clinical Oncology*. 32(suppl 1), S37–S42.

Lim, H. J., & Sun, J. (2003). Nonparametric tests for interval-censored failure time data. *Biometrical Journal: Journal of Mathematical Methods in Biosciences*, *45*(3), 263-276.

Lin, X. and Wang, L. (2010). A semiparametric probit model for case 2 interval censored failure time data. *Statistics in medicine*. 29(9), 972–981.

Lin, G., He, X., & Portnoy, S. (2012). Quantile regression with doubly censored data. *Computational Statistics & Data Analysis*, *56*(4), 797-812.

Lindsey, J. C., & Ryan, L. M. (1998). Methods for interval-censored data. *Statistics in medicine*, *17*(2), 219-238.

Li, Y. M., Ohno, Y., Funaguchi, N., Gomyo, T., Sasaki, Y., Toyoshi, S., ... & Kawasaki, M. (2017). Granulocyte Colony-Stimulating Factor Enhances the Anticancer Effects of Cisplatin against Lung Cancer by Promoting Angiogenesis. *Advances in Lung Cancer*, *6*(01), 1.

Lynn, H. S. (2001). Maximum likelihood inference for left-censored HIV RNA data. *Statistics in medicine*. 20(1), 33–45.

Lynch, C. M., van Berkel, V. H., & Frieboes, H. B. (2017). Application of unsupervised analysis techniques to lung cancer patient data. *PloS one*, *12*(9).

Ma, X., & Xiang, L. (2013). Testing for the Presence of a Cure Fraction in Clustered Interval-Censored Survival Data. *Australian & New Zealand Journal of Statistics*, *55*(2), 173-190.

MacKenzie, G. and Peng, D. (2013). Interval-censored parametric regression survival models and the analysis of longitudinal trials. *Statistics in medicine*. 32(16), 2804–2822.

Magel, R. C., & Degges, R. (1998). Tests for ordered alternatives with right censored data. *Biometrical Journal: Journal of Mathematical Methods in Biosciences*, *40*(4), 495-518.

Malhotra, J., Malvezzi, M., Negri, E., La Vecchia, C., & Boffetta, P. (2016). Risk factors for lung cancer worldwide. *European Respiratory Journal*, *48*(3), 889-902.

Maathuis, M. H., & Wellner, J. A. (2008). Inconsistency of the MLE for the Joint Distribution of Interval‐Censored Survival Times and Continuous Marks. *Scandinavian journal of statistics*, *35*(1), 83-103.

Menon, M. and Saw, H. (1979). Lung cancer in Malaysia. *Thorax*. 34(2), 269–273.

Moghimi-Dehkordi, B., Safaee, A., Pourhoseingholi, M. A., Fatemi, R., Tabeie, Z. and Zali, M. R. (2008). Statistical comparison of survival models for analysis of cancer data. *Asian Pac J Cancer Prev*. 9(3), 417–20.

Miller, D. P. (2004, May). Bootstrap 101: Obtain robust confidence intervals for any statistic. In *SUGI* (Vol. 29, pp. 193-29).

Nadpara, P., Madhavan, S. S. and Tworek, C. (2015). Guideline-concordant timely lung cancer care and prognosis among elderly patients in the United States: a population-based study. *Cancer epidemiology*. 39(6), 1136–1144.

Nawumbeni, N. D., Luguterah, A. and Adampah, T. (2014). Performance of Cox Proportional Hazard and Accelerated Failure Time Models in the Analysis of HIV/TB Co-infection Survival Data. *Research on Humanities and Social Sciences*. 4(21), 94–102. 109

Neal, R. D., Hamilton, W. and Rogers, T. K. (2014). EASILY MISSED? bmj. 349, g6075.

Ning, J., Qin, J., & Shen, Y. (2010). Non‐parametric tests for right‐censored data with biased sampling. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, *72*(5), 609-630

Ning, J., Qin, J., & Shen, Y. (2011). Buckley–James‐type estimator with right‐censored and length‐biased data. *Biometrics*, *67*(4), 1369-1378.

Ng, M. P. (2002). A modification of Peto's nonparametric estimation of survival curves for interval‐censored data. *Biometrics*, *58*(2), 439-442.

Nguyen, E. K., Nicholas, G. A., & Song, X. (2018). Role of Radiotherapy-Induced Malignancies in Patients with Both Breast and Lung Cancer Diagnoses. *Advances in Breast Cancer Research*, *7*(03), 231.

Oral, F., Ekmekc¸i, ˙I. and Onat, N. (2015). Weibull distribution for determination of wind analysis and energy production. *World Journal of Engineering*. 12(3), 215–220. Organization, W. H. et al. (2018). Latest global cancer data: Cancer burden rises to 18.1 million new cases and 9.6 million cancer deaths in 2018.

O'Keeffe, L. M., Taylor, G., Huxley, R. R., Mitchell, P., Woodward, M., & Peters, S. A. (2018). Smoking as a risk factor for lung cancer in women and men: a systematic review and meta-analysis. *BMJ open*, *8*(10), e021611.

Odell, P. M., Anderson, K. M., & D'Agostino, R. B. (1992). Maximum likelihood estimation for interval-censored data using a Weibull-based accelerated failure time model. *Biometrics*, 951-959.

Oller, R., & Gómez, G. (2012). A generalized Fleming and Harrington's class of tests for interval-censored data. *Canadian Journal of Statistics*, *40*(3), 501-516.

Omar, Y., Tasleem, A., Pasquier, M., & Sagahyroon, A. (2018). Lung Cancer Prognosis System using Data Mining Techniques. In *HEALTHINF* (pp. 361-368).

Pajek, M., Kubala-Kuku´s, A., Bana´s, D., Braziewicz, J. and Majewska, U. (2004). Random left-censoring: a statistical approach accounting for detection limits in x-ray fluorescence analysis. *X-Ray Spectrometry*. 33(4), 306–311.

Parsons, A., Daley, A., Begh, R. and Aveyard, P. (2010). Influence of smoking cessation after diagnosis of early stage lung cancer on prognosis: systematic review of observational studies with meta-analysis. *Bmj*. 340, b5569.

Pan, W. (2000). A multiple imputation approach to Cox regression with interval-censored data. *Biometrics*, *56*(1), 199-203.

Parkin, D. M., Bray, F., Ferlay, J., & Pisani, P. (2005). Global cancer statistics, 2002. *CA: a cancer journal for clinicians*, *55*(2), 74-108.

Peto, R., & Peto, J. (1972). Asymptotically efficient rank invariant test procedures. *Journal of the Royal Statistical Society: Series A (General)*, *135*(2), 185-198.

Philip C Hoffman, E. E. V., Ann M Mauer (2000). Lung Cancer. *The Lancet*. 355(9202), 479–485.

Pope III, C. A., Burnett, R. T., Thun, M. J., Calle, E. E., Krewski, D., Ito, K. and Thurston, G. D. (2002). Lung cancer, cardiopulmonary mortality, and long term exposure to fine particulate air pollution. *Jama*. 287(9), 1132–1141.

Pourhoseingholi, M., Moghimi-Dehkordi, B., Safaee, A., Hajizadeh, E., Solhpour, A., Zali, M. et al. (2009). Prognostic factors in gastric cancer using log-normal censored regression model. *Indian Journal of Medical Research*. 129(3), 262.

Pourhoseingholi, M. A., Safaee, A., Moghimi-Dehkordi, B., Zeighami, B., Faghihzadeh, S., Tabatabaee, H. R. and Pourhoseingholi, A. (2008). Quality of life in breast cancer patients-a quantile regression analysis. *Asian Pacific Journal of Cancer Prevention*. 9, 487–490.

POURHOSSEIN, G. M., POURHOSSEIN, G. A., Vahedi, M., Moghimi, D. B., SAFAEI, A., Ashtari, S. and Zali, M. (2011). Alternative for the Cox Regression model: using Parametric Models to Analyze the Survival of Cancer Patients. 110

Powell, T. M. and Bagnell, M. E. (2012). Your survival guide to using time-dependent covariates. In Proceedings of the SAS Global Forum. *Citeseer*, 22–25.

Prentice, R. L., Kalbfleisch, J. D., Peterson Jr, A. V., Flournoy, N., Farewell, V. and Breslow, N. (1978). The analysis of failure times in the presence of competing risks. *Biometrics*, 541–554.

Qin, J., & Shen, Y. (2010). Statistical methods for analyzing right-censored length-biased data under Cox model. *Biometrics*, *66*(2), 382-392.

Rahimzadeh, M., Pourhoseingholi, M. A. and Kavehie, B. (2016). Survival rates for breast cancer in Iranian patients: A meta-analysis. *Asian Pacific Journal of Cancer Prevention*. 17(4), 2223–2227.

Ramchandani, R., Finkelstein, D. M. and Schoenfeld, D. A. (2015). A model informed rank test for right-censored data with intermediate states. *Statistics in medicine*. 34(9), 1454–1466.

Ranganathan, P., Pramesh, C. et al. (2012). Censoring in survival analysis: potential for bias. *Perspect Clin Res*. 3(1), 40.

Ramlau-Hansen, H., Chr. Bang Jespersen, N., Kragh Andersen, P., Borch-Johnsen, K., & Deckerf, T. (1987). Life insurance for insulin-dependent diabetics. *Scandinavian Actuarial Journal*, *1987*(1-2), 19-36.

Ridge, C. A., McErlean, A. M., & Ginsberg, M. S. (2013, June). Epidemiology of lung cancer. In *Seminars in interventional radiology* (Vol. 30, No. 02, pp. 093-098). Thieme Medical Publishers.

Rochowicz, J. A. (2011). Bootstrapping analysis, inferential statistics and EXCEL. *Spreadsheets Edu.(eJSiE)*, *4*, 1-23.

Safaee, A., Moghimi-Dehkordi, B., Zeighami, B., Tabatabaee, H., Pourhoseingholi, M. et al. (2008). Predictors of quality of life in breast cancer patients under chemotherapy. *Indian Journal of Cancer*. 45(3), 107.

Saffari, S. E., Adnan, R. and Greene, W. (2013). Investigating the impact of excess zeros on hurdle-generalized Poisson regression model with right censored count data. *Statistica Neerlandica*. 67(1), 67–80.

Sahinler, S., & Topuz, D. (2007). Bootstrap and jackknife resampling algorithms for estimation of regression parameters. *Journal of Applied Quantitative Methods*, *2*(2), 188-199.

Samar Ali, S. and Kannan, S. (2011). A diagnostic approach to Weibull-Weibull stress-strength model and its generalization. *International Journal of Quality & Reliability Management*. 28(4), 451–463. 111

Salim, E. I., Jazieh, A. R., & Moore, M. A. (2011). Lung cancer incidence in the Arab league countries: risk factors and control. *Asian Pac J Cancer Prev*, *12*(1), 17-34.

Samartzis, L. (2006). Survival and censored data. *Technical report*.

Sankaranarayanan, R., Ramadas, K., & Qiao, Y. L. (2014). Managing the changing burden of cancer in Asia. *BMC medicine*, *12*(1), 3.

Sattar, A., Sinha, S. K., Wang, X.-F. and Li, Y. (2015). Frailty models for pneumonia to death with a left-censored covariate. *Statistics in medicine*. 34(14), 2266–2280.

Sauerbrei, W., & Schumacher, M. (1992). A bootstrap resampling procedure for model building: application to the Cox regression model. *Statistics in medicine*, *11*(16), 2093-2109.

Severiano, A., Carriço, J. A., Robinson, D. A., Ramirez, M., & Pinto, F. R. (2011). Evaluation of jackknife and bootstrap for defining confidence intervals for pairwise agreement measures. *PLoS One*, *6*(5).

Schoop, R., Graf, E. and Schumacher, M. (2008). Quantifying the Predictive Performance of Prognostic Models for Censored Survival Data with Time-Dependent Covariates. *Biometrics*. 64(2), 603–610.

Sciacca, L. M. (2014). *Risk factors for attempting suicide among transgender individuals: Survival analysis with time-dependent covariates*. Palo Alto University.

Sekaran, U. and Bougie, R. (2011). Research method for business: A skill building approach.

Shao, F., Li, J., Ma, S. and Lee, M.-L. T. (2014). Semiparametric varying coefficient model for interval censored data with a cured proportion. *Statistics in medicine*. 33(10), 1700–1712.

Sharifian, A., Pourhoseingholi, M. A., Norouzinia, M. and Vahedi, M. (2014). Ovarian cancer in Iranian women, a trend analysis of mortality and incidence. *Asian Pac J Cancer Prev*. 15(24), 10787–10790.

Shelburne, R. C. (2013). European Press Conference for the Launch of the UN World Economic Situation and Prospects 2013.

Shen, H., & Cook, R. J. (2015). Analysis of interval-censored recurrent event processes subject to resolution. *Biometrical Journal*, *57*(5), 725-742.

Shin, H. R., Carlos, M. C., & Varghese, C. (2012). Cancer control in the Asia Pacific region: current status and concerns. *Japanese journal of clinical oncology*, *42*(10), 867-881.

Shin, H. R., Jung, K. W., Won, Y. J., & Park, J. G. (2004). 2002 annual report of the Korea Central Cancer Registry: based on registered data from 139 hospitals. *Cancer research and treatment: official journal of Korean Cancer Association*, *36*(2), 103.

Sheridan, J., & Collins, A. (2013). Adult lung cancer in southern Africa: epidemiology and aetiology. *Afr J Respir Med*, *8*.

Shepherd, F. A., Rodrigues Pereira, J., Ciuleanu, T., Tan, E. H., Hirsh, V., Thongprasert, S., ... & van Kooten, M. (2005). Erlotinib in previously treated non–small-cell lung cancer. *New England Journal of Medicine*, *353*(2), 123-132.

Singh, B. (2002). Inferences concerning exponential distributions in the presence of randomly right censored data with missing censored values. *Lifetime data analysis*, *8*(1), 69-88.

Singh, R. S. and Totawattage, D. P. (2013). The statistical analysis of interval-censored failure time data with applications.

Siannis, F. (2011). Sensitivity analysis for multiple right censoring processes: investigating mortality in psoriatic arthritis. *Statistics in medicine*, *30*(4), 356-367.

Smith, K. R. (2000). National burden of disease in India from indoor air pollution. *Proceedings of the National Academy of Sciences*, *97*(24), 13286-13293.

Sobue, T., Suzuki, T., Fujimoto, I., Matsuda, M., Doi, O., Mori, T., ... & Kono, K. (1994). Case-control study for lung cancer and cigarette smoking in Osaka, Japan: comparison with the results from Western Europe. *Japanese journal of cancer research*, *85*(5), 464-473.

Sparling, Y. H., Younes, N., Lachin, J. M. and Bautista, O. M. (2006). Parametric survival models for interval-censored data with time-dependent covariates. *Biostatistics*. 7(4), 599–614.

Stewart, B., Wild, C. P. et al. (2015). World cancer report 2014. World.

Stewart, B., Wild, C. P. et al. (2016). World cancer report 2014. World.

Sun, J. and Wei, L. (2000). Regression analysis of panel count data with covariate dependent observation and censoring times. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*. 62(2), 293–302.

Sun, L., Wang, L., & Sun, J. (2006). Estimation of the association for bivariate interval-censored failure time data. *Scandinavian Journal of Statistics*, *33*(4), 637-649.

Sun, S., Schiller, J. H., & Gazdar, A. F. (2007). Lung cancer in never smokers—a different disease. *Nature Reviews Cancer*, *7*(10), 778.

Sun, L., Zhu, L., & Sun, J. (2009). Regression analysis of multivariate recurrent event data with time-varying covariate effects. *Journal of Multivariate Analysis*, *100*(10), 2214-2223.

Snapinn, S. M., Jiang, Q. I., & Iglewicz, B. (2005). Illustrating the impact of a time-varying covariate with an extended Kaplan-Meier estimator. *The American Statistician*, *59*(4), 301-307.

Talamakrouni, M., Ghouch, A. E. and Van Keilegom, I. (2015). Guided censored regression. *Scandinavian Journal of Statistics*. 42(1), 214–233.

Temprosa, M. G. M. (2012). An Imputation-Estimation Algorithm Using Time-Varying Auxiliary Covariates for a Longitudinal Model When Outcome is missing by Design. Ph.D. Thesis. *The George Washington University*. 112.

Ten Haaf, K., Bastani, M., Cao, P., Jeon, J., Toumazis, I., Han, S. S., ... & Feuer, E. J. (2019). A comparative modelling analysis of risk-based lung cancer screening strategies. *JNCI: Journal of the National Cancer Institute*.

Thompson, R., Voit, E. and Scott, G. (2000). Statistical modeling of sediment and oyster PAH contamination data collected at a South Carolina estuary (complete and left-censored samples). *Environmetrics*. 11(1), 99–119.

Torre, L. A., Bray, F., Siegel, R. L., Ferlay, J., Lortet-Tieulent, J. and Jemal, A. (2015). Global cancer statistics, 2012. CA*: a cancer journal for clinicians*. 65(2), 87–108.

Topp, R., & Gómez, G. (2004). Residual analysis in linear regression models with an interval-censored covariate. *Statistics in medicine*, *23*(21), 3377-3391.

Troendle, J. F., & Yu, K. F. (2006). Likelihood approaches to the non-parametric two-sample problem for right-censored data. *Statistics in medicine*, *25*(13), 2284-2298.

Turnbull, B. W. (1976). The empirical distribution function with arbitrarily grouped, censored and truncated data. *Journal of the Royal Statistical Society: Series B (Methodological)*, *38*(3), 290-295.

Unkle, R. and Venkataraman, R. (2002). Relationship between Weibull and AMSAA models in reliability analysis: A case study. *International Journal of Quality & Reliability Management.* 19(8/9), 986–997.

Urman, A., Josyula, S., Rosenberg, A., Lounsbury, D., Rohan, T., & Hosgood, H. (2016). Burden of lung cancer and associated risk factors in Africa by region. *J Pulm Respir Med*, *6*(340), 2.

Ushijima, C., Yamazaki, K., Kouso, H., Katsura, M., Mori, R., Takeo, S., & Ishikawa, K. (2016). Surgical Treatment of Lung Cancer in Patients with Psychiatric Disorders: A Retrospective Study. *Journal of Cancer Therapy*, *7*(08), 553.

Ward, E., Jemal, A., Cokkinides, V., Singh, G. K., Cardinez, C., Ghafoor, A., & Thun, M. (2004). Cancer disparities by race/ethnicity and socioeconomic status. *CA: a cancer journal for clinicians*, *54*(2), 78-93.

Wang, L., Sun, L., & Sun, J. (2006). A Goodness-of-fit Test for the Marginal Cox Model for Correlated Interval-censored Failure Time Data. *Biometrical Journal: Journal of Mathematical Methods in Biosciences*, *48*(6), 1020-1028.

Wang, L., McMahan, C. S., Hudgens, M. G. and Qureshi, Z. P. (2016). A flexible, computationally efficient method for fitting the proportional hazards model to interval-censored data. *Biometrics*. 72(1), 222–231.

Wang, B. Y., Huang, J. Y., Cheng, C. Y., Lin, C. H., Ko, J. L., & Liaw, Y. P. (2013). Lung cancer and prognosis in Taiwan: a population-based cancer registry. *Journal of Thoracic Oncology*, *8*(9), 1128-1135.

Walser, T., Cui, X., Yanagawa, J., Lee, J. M., Heinrich, E., Lee, G., ... & Dubinett, S. M. (2008). Smoking and lung cancer: the role of inflammation. *Proceedings of the American Thoracic Society*, *5*(8), 811-815.

Wannemuehler, K. A., Lyles, R. H., Manatunga, A. K., Terrell, M. L., & Marcus, M. (2010). Likelihood-based methods for estimating the association between a health outcome and left-or interval-censored longitudinal exposure data. *Statistics in medicine*, *29*(16), 1661-1672.

Warschkow, R., Steffen, T., Thierbach, J., Bruckner, T., Lange, J., & Tarantino, I. (2011). Risk factors for anastomotic leakage after rectal cancer resection and reconstruction with colorectostomy. A retrospective study with bootstrap analysis. *Annals of surgical oncology*, *18*(10), 2772.

Waitman, L. R., Fisher, D. H., & King, P. H. (2006). Bootstrapping rule induction to achieve rule stability and reduction. *Journal of Intelligent Information Systems*, *27*(1), 49-77.

Weingaertner, V., Scheve, C., Gerdes, V., Schwarz-Eywill, M., Prenzel, R., Bausewein, C., Higginson, I. J., Voltz, R., Herich, L., Simon, S. T. et al. (2014). Breathlessness, functional status, distress, and palliative care needs over time in patients with advanced chronic obstructive pulmonary disease or lung cancer: a cohort study. *Journal of pain and symptom management*. 48(4), 569–581.

Winnett, A. and Sasieni, P. (2003). Iterated residuals and time-varying covariate effects in Cox regression. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*. 65(2), 473–488.

Wu, C. F. J. (1986). Jackknife, bootstrap and other resampling methods in regression analysis. *The Annals of Statistics*, *14*(4), 1261-1295.

Wu, Y., & Cook, R. J. (2015). Penalized regression for interval-censored times of disease progression: Selection of HLA markers in psoriatic arthritis. *Biometrics*, *71*(3), 782-791.

Wong, M., Lam, K. and Lo, E. (2011). Analysis of multilevel grouped survival data with time-varying regression coefficients. *Statistics in medicine*. 30(3), 250–259.

Xu, G., Sen, B., & Ying, Z. (2014). Bootstrapping a change-point Cox model for survival data. *Electronic journal of statistics*, *8*(1), 1345.

Yang, J. and Scott, D. W. (2013). Robust fitting of a Weibull model with optional censoring. *Computational statistics & data analysis*. 67, 149–161. 113

Yang, B. H., Parkin, D. M., Cai, L., & Zhang, Z. F. (2004). Cancer burden and trends in the Asian Pacific Rim region. *Annals of Epidemiology*, *14*(8), 596.

Yu, B. (2010). A Bayesian MCMC approach to survival analysis with doubly-censored data. *Computational statistics & data analysis*, *54*(8), 1921-1929.

Yu, X. Q., Luo, Q., Hughes, S., Wade, S., Caruana, M., Canfell, K., & O'Connell, D. L. (2019). Statistical projection methods for lung cancer incidence and mortality: a systematic review. *BMJ open*, *9*(8), e028497.

Yin, Z., Cui, Z., Ren, Y., Xia, L., Wang, Q., Zhang, Y., He, Q. and Zhou, B. (2016). Association between polymorphisms in pre-miRNA genes and risk of lung cancer in a Chinese non-smoking female population. *Lung Cancer*. 94, 15–21.

Yu, Q., Li, L., & Wong, G. Y. (2000). On consistency of the self-consistent estimator of survival functions with interval-censored data. *Scandinavian Journal of Statistics*, *27*(1), 35-44.

Yu, Q., & Wong, G. Y. (2005). Modified semiparametric maximum likelihood estimator in linear regression analysis with complete data or right-censored data. *Technometrics*, *47*(1), 34-42.

Yu, L., Liu, L. and Chen, D.-G. D. (2013). Weighted Least-Squares Method for Right-Censored Data in Accelerated Failure Time Model. *Biometrics*. 69(2), 358–365.

Yuan, J. M., Stram, D. O., Arakawa, K., Lee, H. P., & Mimi, C. Y. (2003). Dietary cryptoxanthin and reduced risk of lung cancer: the Singapore Chinese Health Study. *Cancer Epidemiology and Prevention Biomarkers*, *12*(9), 890-898.

Zainal Ariffin, O., Nor Saleha, I. et al. (2011). National cancer registry report 2007. Malaysia: Ministry of Health.

Zainudin, N., Elfaki, F. and Ali, M. Y. (2015). Confidence Interval for Survival Model based on Partly Interval-Censored data.

Zhao, Q., & Sun, J. (2004). Generalized log-rank test for mixed interval-censored failure time data. *Statistics in medicine*, *23*(10), 1621-1629.

Zhao, X., Zhao, Q., Sun, J. and Kim, J. S. (2008). Generalized Log-Rank Tests for Partly Interval-Censored Failure Time Data. *Biometrical Journal*. 50(3), 375–385.

Zhao, L. and Hu, X. J. (2013). Estimation with right-censored observations under a semi-Markov model. *Canadian Journal of Statistics*. 41(2), 237–256.

Zhou, Q., Cai, J., & Zhou, H. (2018). Outcome-dependent sampling with interval-censored failure time data. *Biometrics*, *74*(1), 58-67.

Zhang, Z., Sun, L., Zhao, X., & Sun, J. (2005). Regression analysis of interval-censored failure time data with linear transformation models. *Canadian Journal of Statistics*, *33*(1), 61-70.

Zhang, Z., Reinikainen, J., Adeleke, K. A., Pieterse, M. E., & Groothuis-Oudshoorn, C. G. (2018). Time-varying covariates and coefficients in Cox regression models. *Annals of translational medicine*, 6(7).

Zhang, C. H., & Li, X. (1996). Linear regression with doubly censored data. The Annals of Statistics, 24(6), 2720-2743.

Zhao, G. (2008). Nonparametric and parametric survival analysis of censored data with possible violation of method assumptions. The University of North Carolina at Greensboro.

Zhu, L., Tong, X., & Sun, J. (2008). A transformation approach for the analysis of interval-censored failure time data. *Lifetime data analysis*, *14*(2), 167-178.

Zucker, D. M. and Karr, A. F. (1990). Nonparametric survival analysis with time dependent covariate effects: a penalized partial likelihood approach. *The Annals of Statistics*, 329–353.