CHAPTER 2
THE ANALYSIS OF LUNG CANCER PATIENTS

Anamila Ariffin and Noraslinda Mohamed Ismail

2.1 INTRODUCTION

Survival analysis originated in seventeenth century when John Graunt presented the first weekly bill of mortality and Healey published the first lifetable (Kreager, 1988). The interest in durability or the "lifetime" of industrial devices started during World War II when the reliability of military equipment became an issue. After the war, the method was further developed and applied to the study of the survival time of cancer patients. Cancer researchers changed the term "lifetime analysis" to "survival analysis".

Survival analysis has been widely used in many fields of research such as medicine, economics, and political science (Gerard & Van, 2021). Depending on the field of research, survival analysis might also be called as event history analysis in sociology, failure time analysis in engineering and in reliability theory, transition analysis or duration analysis in economics. Generally, survival analysis is a collection of statistical methods for analysing the data which the outcome variable is time until an event occurs (González & Rossello, 2022; Song, 2021).

Survival analysis took place in era of 1950's where Kaplan and Meier (1958), proposed their famous estimator of the survival curve. Later in 1972, Cox introduced the proportional hazards model which

includes covariates. Most survival analysis must consider their analytical problem called censoring (González & Rossello, 2022). Theoretically, censoring happens when the information about individual survival time is incomplete, and the most survival data is right-censored. Right censoring is when the survival time becomes incomplete at the right side of the study period. However, the data can also be left-censored or interval-censored.

There are three models in survival analysis. They are parametric, semiparametric, and non-parametric. Parametric survival models were well-established in actuarial science and demography but have never dominated medical uses of survival analysis. This is because they need to take serious on the hazard baseline. Popular ones include the exponential, Weibull, and lognormal distributions. Devi (2015) discussed the use of survival analysis model using Kaplan-Meier method to model time until return to drug use for a set of drug addicted patients. Besides that, Cox regression model is also being used to investigate the effect of covariates in drug relapsing for a particular set of drug addicted patients. The result was that the probability of drug relapse decreases with time if the treatment is done for more time, chances of drug relapse are less.

The proportional hazards model that was proposed by Cox in 1972 has been widely used by many researchers for survival analysis. Abdulkabir et al. (2015) compared the gradient curve of the Cox proportional hazard and Weibull models using the data on tuberculosis diseases in Nigeria for the year 2011. It was observed that the shape parameter of the Weibull model does not depend or effect on the behaviour of the Cox proportional hazard model.

Caraviello et al. (2004) showed a comparison between a Weibull proportional hazards model and a conventional linear model for predicting the genetic merit of US Jersey Sires for daughter longevity. Based on the analysis in this study, the proportional hazards model generally yielded more accurate predictions according to the mentioned criteria, but differences in predictive ability between methods were smaller when using a Kullback-Leibler distance than with other

approaches. The results suggest that survival analysis methodology may provide more accurate predictions of genetic merit for longevity than conventional linear models.

Liu et al. (2012) estimated the sensitivity and specificity of tests for screening antibodies to human immunodeficiency virus (HIV) by using frequentist and Bayesian approaches. After comparing the results for 100% of sensitivity and, or specificity, they found that Bayesian approach acquired reasonable interval estimates of antibodies screening test to HIV. Meanwhile, frequentist approach expresses objectively the accuracy of each assay.

The proportional hazards model used in this chapter has been used extensively since 1972. In a fully parametric model, the lifetime distribution has been assumed to belong to a family of parametric distributions and reducing the regression problem to estimating the parameters from the data.

Recently, the proportional hazards model has been the most common from a Bayesian perspective (González & Hassan, 2021). It has been widely used in survival analysis. Gibbs Sampling is one of the new numerical algorithms which allow obtaining samples from posterior of interest and this new development has motivated the use of Bayesian methods in survival analysis. Gibbs Sampler is one of the best-known markov chain monte carlo (MCMC) samplings in computational literature. Specialized software packages called BUGS (Spiegelhalter, 1996) are created for implementing MCMC-based analysis of full probability models. These packages will treat all unknowns as random variables. This chapter describes the use of freely available software for the analysis of complex statistical models using MCMC techniques, called WinBUGS (Spiegelhalter, 2003) and R programming, in the context of population health management (PHM).

2.2 EARLIER STUDIES ON LUNG CANCER ANALYSIS

Lung cancer is the most prevalent type of cancer death, and 5-year survival varies according on stage of diagnosis. Although Stage I lung