SURVIVAL ANALYSIS IN ACUTE MYELOID LEUKEMIA: A RETROSPECTIVE COHORT STUDY AND SIMULATION STUDY OF SMALL EVENTS PER INDEPENDENT VARIABLE

LAW KIAN BOON

DISSERTATION SUBMITTED IN FULFILMENT
OF THE REQUIREMENTS FOR THE DEGREE OF
MASTER OF SCIENCE

INSTITUTE OF MATHEMATICAL SCIENCES
FACULTY OF SCIENCE
UNIVERSITY OF MALAYA
KUALA LUMPUR

2017

UNIVERSITI MALAYA

ORIGINAL LITERARY WORK DECLARATION

Name of Candidate: Law Kian Boon

Registration/Matric No: SGP140003

Name of Degree: Master of Science

Title of Dissertation ("this Work"):

SURVIVAL ANALYSIS IN ACUTE MYELOID LEUKEMIA: A RETROSPECTIVE COHORT STUDY AND SIMULATION STUDY OF SMALL EVENTS PER INDEPENDENT VARIABLE

Field of Study: Statistics

I do solemnly and sincerely declare that:

(1) I am the sole author/writer of this Work;
(2) This Work is original;
(3) Any use of any work in which copyright exists was done by way of fair dealing and for permitted purposes and any except or extract from, or reference to or reproduction of any copyright work has been disclosed expressly and sufficiently and the title of the Work and its authorship have been acknowledged in this Work;
(4) I do not have any actual knowledge nor do I ought reasonably to know that the making of this Work constitutes an infringement of any copyright work;
(5) I hereby assign all and every rights in the copyright to this Work to the University of Malaya ("UM"), who henceforth shall be owner of the copyright in this Work and that any reproduction or use in any form or by any means whatsoever is prohibited without the written consent of UM having been first had and obtained;
(6) I am fully aware that if in the course of making this Work I have infringed any copyright whether intentionally or otherwise, I may be subject to legal action or any other action as may be determined by UM.

Candidate's Signature Date

Subscribed and solemnly declared before,

Witness's Signature Date

Name:
Designation:

ABSTRACT

Intensive chemotherapy for acute myeloid leukemia (AML) is used in both induction and consolidation treatments. The combination of fludarabine, high dose cytarabine and granulocyte colony-stimulating factor (FLAG) has been proven effective and safe as an induction treatment for refractory and poor risk AML, but not as a consolidation treatment especially in older AML patients. Hence, a retrospective cohort study was conducted to investigate the role of the FLAG regimen as consolidation treatment in older AML patients. Survival data of 41 eligible older patients were analyzed by using Kaplan-Meier method, log-rank test and Cox model. The results showed that patients consolidated with the FLAG regimen had a longer overall survival (OS) and disease free survival (DFS) when compared to non-FLAG regimens. The primary treatment variable (FLAG) was strongly associated with the survival outcomes with statistically different OS (log-rank, $p = 0.0025$) and DFS (log-rank, $p = 0.0026$). However, the regression analysis was performed at low events per independent variable (EPV) condition. The validity of Cox coefficient estimate requires at least 10 to 20 EPV, which can be difficult to achieve in a small study. Therefore, a simulation study was performed to assess the performance of Cox coefficient estimate for the primary treatment variable at low EPV spectrum. Our results showed that 3 and 4 EPV were associated with highest level of bias and disparity in accuracy, precision and statistical properties. At 5 to 6 EPV, the performance of Cox model started to gain stability. Above 6 EPV, increasing the number of events was less likely to improve the overall performance of the Cox model. The FLAG regimen should be used as part of consolidation for AML as the regimen improved both OS and DFS among older AML patients. The EPV rule has exceeded the number of outcome events required by a variable of strong association to the survival outcomes.

ABSTRAK

Kemoterapi intensif untuk leukemia myeloid akut (AML) telah digunakan dalam kedua-dua rawatan induksi dan konsolidasi. Gabungan fludarabine, cytarabine dos tinggi dan faktor perangsang koloni granulosit (FLAG) telah dibuktikan sebagai satu rawatan induksi yang berkesan dan selamat untuk AML refraktori dan AML berisiko buruk, tetapi bukan sebagai rawatan konsolidasi terutamanya bagi pesakit tua AML. Oleh itu, satu kajian kohort retrospektif telah dilakukan untuk menyiasat peranan regimen FLAG sebagai rawatan konsolidasi di kalangan pesakit tua AML. Data kehayatan daripada 41 pesakit tua yang layak telah dianalisis dengan menggunakan kaedah Kaplan-Meier, ujian log-rank dan model Cox. Keputusan menunjukkan bahawa pesakit yang dikonsolidasikan oleh regimen FLAG telah mencapai kehayatan keseluruhan (OS) dan kehayatan bebas penyakit (DFS) yang lebih panjang berbanding dengan regimen bukan FLAG. Pembolehubah rawatan utama (FLAG) didapati berkait secara kuat dengan hasil kehayatan dengan lengkung kehayatan yang berbeza secara statistik bagi OS (log-rank, p = 0.0025) dan DFS (log-rank, p = 0.0026). Namun, analisis regresi telah dilakukan dalam keadaan kejadian per pembolehubah tak bersandar (EPV) yang rendah. Kesahihan anggaran pekali koefisien Cox memerlukan sekurang-kurangnya 10 hingga 20 EPV, yang agar sukar dicapai dalam kajian kecil. Maka, satu kajian simulasi telah dijalankan untuk mengesahkan kelakuan anggaran pekali koefisien Cox bagi pembolehubah utama pada EPV spektrum rendah. Keputusan kami menunjukkan bahawa 3 dan 4 EPV telah dikaitkan dengan tahap pincang dan perbezaan yang paling tinggi dalam kejituan, kepersisan dan sifat berstatistik. Pada 5 hingga 6 EPV, prestasi model Cox mula menjadi stabil. Atas 6 EPV, peningkatan bilangan kejadian adalah kurang berkemungkinan untuk menambahbaikan prestasi keseluruhan model Cox tersebut. Regimen FLAG harus digunakan sebagai sebahagian rawatan konsolidasi bagi AML kerana regimen tersebut telah menambahbaikkan kedua-dua OS dan DFS di golongan pesakit tua AML. Peraturan EPV tersebut telah melebihi bilangan kejadian yang diperlukan oleh pembolehubah yang bersekutuan secara kuat dengan kesudahan kehayatan.

ACKNOWLEDGEMENTS

I would like to gratefully and sincerely thank Prof. Dr. Nor Aishah Hamzah, Dr. Ng Kok Haur and Dato Dr. Chang Kian Meng for all the guidance, understanding, patience, and most importantly, their mentorship during my postgraduate study at University of Malaya. The mentorship was paramount in providing a clear direction in completing my study goal. Their guidance and words are encouraging and inspiring, that led me to not only grow as a professional statistician, but also an independent thinker and problem solver. I am not sure many postgraduate students are given the opportunity to develop their own individuality and self-sufficiency by being allowed to work with such independence. For everything both of you have done for me, I sincerely thank you.

Finally, I would like to thank my wife Ying for her support, encouragement, patience, tolerance and love that were undeniably the most important part of my motivation to complete the postgraduate study.

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

LIST OF SYMBOLS

α	Shape parameter
β	Regression coefficient
$\hat{\beta}$	'true' regression coefficient
$\tilde{\beta}$	Simulated regression coefficient
$\bar{\tilde{\beta}}$	Average of simulated regression coefficient
d	Number of events at time *t*
e	exponential
E_i	Number of expected events
f	Number of patients alive just before time *t*
g	Total number of groups
γ	Constant hazard rate
$h(t\|\mathbf{X})$	Hazard rate at time t given $\mathbf{X}$
$h_0(t)$	Baseline hazard function
k	Number of independent variables
λ	Scale parameter
m	Total number of simulations
O_i	Number of observed events
P	Event prevalence
p	*p*-value
$S(t)$	Probability of being alive at time t
$S(t-1)$	Probability of being alive at time $t-1$
U	Random number
Uni	Uniform distribution
T	Survival time

T_{DFS}	Simulated survival time in DFS
T_{OS}	Simulated survival time in OS
T'_{DFS}	Original survival time in DFS
T'_{OS}	Original survival time in OS

LIST OF ABBREVIATIONS

AML	Acute myeloid leukemia
APML	Acute promyelocytic leukemia
Ara-C	Cytarabine
Ara-CTP	Ara-C-5'-triphosphate
ASV	Actual sample variance
CI	Confidence interval
CIF	Cumulative incident function
CR	Complete remission
DFS	Disease free survival
EPV	Events per independent variable
ESMO	European Society for Medical Oncology
FAB	French-American-British
FLAG	Fludarabine, high dose cytarabine and granulocyte colony-stimulating factor
FLAG-Ida	FLAG plus idarubicin
FLAMSA	FLAG plus amsacrine
G-CSF	Granulocyte colony-stimulating factor
HCT	Hematopoietic cell transplantation
HiDAC	High dose Ara-C
HiDAC-Ida	High dose Ara-C plus idarubicin

HR	Hazard ratio
IQR	Interquartile range
KM	Kaplan-Meier
MDS	Myelodysplastic syndrome
MiDAC	Mitoxantrone plus intermittent dose Ara-C
Mito-FLAG	FLAG plus mitoxantrone
MMV	Mean model variance
NCCN	National Comprehensive Cancer Network
OS	Overall survival
SE	Standard error
SEER	Surveillance, Epidemiology and End Results

CHAPTER 1: GENERAL INTRODUCTION

1.1 Background of Research

Treatment for acute myeloid leukemia (AML) can be divided into induction treatment and consolidation treatment. Many cytotoxic regimens have been studied and reported as being comparable to the gold standard therapy of cytarabine (Ara-C) plus daunorubicin. Among those regimens investigated, fludarabine, high dose cytarabine and granulocyte colony-stimulating factor (G-CSF) or FLAG, with or without other agents has received much attention due to comparable efficacy and relatively low cardiac toxicity. While most patients who receive an induction treatment will enter complete remission (CR), the most common cause of death in AML is related to relapse and subsequent complications (Krug et al., 2011; Ramos et al., 2015). In another word, the lack of superior post remission treatment leads to poor survival outcomes in patients. Despite of the fact that an effective consolidation treatment is a prerequisite for long term survival in AML, cytotoxic regimens with optimum treatment cycles have yet been optimized for consolidation treatment.

Acute myeloid leukemia is a rare and highly malignant form of leukemia. The disease has the lowest survival rate among all forms of leukemia (Deschler & Lübbert, 2006). It is also the most frequent form of leukemia that causes large number of cancer-related deaths every year in the western world. The Surveillance, Epidemiology and End Results (SEER) program of the National Cancer Institute in the United States reported a 5-year survival rate of 23% for those who are less than 55 years old, whereas the corresponding rate for those who are above 55 years was only 11% (Bethesda, 2015). Survival rates continue to improve for younger patients, but not for older patients in the past few decades. Survival rates differ between younger and older patients because of the treated patients are more likely to be younger and less likely to have poor performance

indicators and comorbidity score as compared with the untreated patients (Medeiros et al., 2015).

Acute myeloid leukemia is also a disease of late adulthood with reported median age at diagnosis from 65 to 70 years old (Thein et al., 2013; Bethesda, 2015). However, a real-world study has shown that up to 60% of elderly AML patients remain untreated following diagnosis (Medeiros et al., 2015). Elderly patients are generally regarded as not suitable for intensive chemotherapy due to many reasons; the toxicity of cytotoxic drugs is increased in the elderly; older patients are more prone to have particular risk factors for poor outcomes, for example, poor risk cytogenetics and the overexpression of genes associated with drug resistance; older patients also tend to have poor performance status, comorbidities and other medical contraindications to intensive chemotherapy. Therefore, nearly half of older AML patients are often given palliative treatment and low dose therapy straight. Research has shown that palliative treatment and low dose therapy are much more inferior than the intensive chemotherapy (Estey, 2007). For example, a study has reported that a low dose Ara-C treatment results in a CR rate of 7% when compared with 56% from intensive chemotherapy (Heiblig et al., 2016). The poor survival outcome in older AML patients worldwide motivates us to identify a better consolidation treatment for them.

The cohort study is conducted to address the unmet concerns as abovementioned. The framework of the cohort study therefore covers the unmet treatment need for older AML patients and the potential use of the FLAG regimen as part of consolidation treatment for them. The primary treatment variable is binary, consisting of a group of older patients who were consolidated with the FLAG regimen and a group of older patients who received non-FLAG regimens. In the study, we apply methods of survival analysis such as Kaplan-Meier (KM) product limit method, log-rank test and regression

analysis with Cox proportional hazards model to describe, quantify and compare the survival data from both FLAG and non-FLAG groups.

In clinical research, covariate effects are adjusted by using a regression model. The most common regression model used in survival analysis is the proportional hazards model developed by Cox (1972). The model is also called Cox proportional hazards model or Cox model. The Cox model can be used to analyze different type of distributions as long as the assumption of proportional hazards or constant hazard ratio (HR) is fulfilled. Like logistic regression model, the validity of the Cox model relies on the number of outcome events in relation to the number of predictor variables incorporated into the model. This property has been investigated in many simulation studies with aim to establish a definitive standard for trustworthy regression analysis with Cox model.

In survival analysis, it is generally recognized that 10 to 20 events per independent variable (EPV) are needed for a regression analysis with logistic model or Cox model (Concato et al., 1995; Peduzzi et al., 1995; Peduzzi et al., 1996). Inadequate EPV can cause biased coefficient estimates. But in analysis of causal influences for observational studies, control of confounders may require adjustment for more covariates than the rule of EPV allows (Greenland, 1989). Moreover, for rare disease like AML, it is very challenging to conduct clinical studies in specific patient cohorts that fulfill a particular treatment condition. This may explain the lack of investigations involving consolidation treatments in older AML patients.

It is noteworthy that the rule of EPV for logistic regression model and Cox model is established based on simulation studies with independent variables of moderate associations with the study outcomes. Hence, a low EPV condition may not necessarily affect the validity of regression analysis for variables of high association with study

outcomes. Besides, there is no single rule based on EPV that would guarantee an accurate estimation of logistic regression parameters (Courvoisier et al., 2011).

In our simulation study, the cohort dataset with 26 events for overall survival (OS) and 27 events for disease free survival (DFS) in AML was simulated to investigate the performance of Cox regression estimates for the primary treatment variable (FLAG versus non-FLAG). The primary treatment variable was found highly associated with the survival outcomes at less than 10 EPV condition. Our argument is that high variable association would requires less events, therefore would also require less EPV for accurate coefficient estimate.

1.2 Significance of Research

Our research addresses a few problems in AML and regression analysis with proportional hazards model. There is a serious lack of research in using the FLAG regimen as part of consolidation treatment in AML. Many physicians would continue the post remission treatment with similar cytotoxic agents used in the induction treatment or with high dose Ara-C (HiDAC) alone. The lack of EPV is also a very common problem in clinical research involving regression analysis with Cox model. Although many simulation studies have been conducted to evaluate problems associated with the lack of EPV, the outcomes are not consistent. As a rule of thumb, 10 to 20 EPV are usually required for regression analysis with logistic model or Cox model. Besides, variable association has never been sufficiently investigated in the past simulation studies. Therefore, our study is significant as it helps to answer the research problems below.

a) Is the FLAG regimen an effective consolidation treatment for AML?
b) Can the FLAG regimen help prolong the OS and DFS of older patients when being used as part of consolidation treatment for AML?

c) Does a variable showing high association to survival outcome require less EPV in regression analysis with Cox model?

d) What is the required EPV for variable showing high variable association to survival outcomes?

1.3 Objectives of Research

There are two primary objectives for this dissertation.

a) The objective of the cohort study is to assess the role and to quantify the effect of a treatment variable (FLAG versus non-FLAG) to the survival outcomes of older AML patients.

b) The objective of the simulation study is to investigate the performance of the Cox coefficient estimate of the treatment variable (FLAG versus non-FLAG) at low EPV condition and to justify the low EPV requirement for variable of high variable association with survival outcome.

The two objectives sound different, but are actually inter-connected. The first objective is designed to answer research problems 1.2(a) and 1.2(b), regarding to the use of the FLAG regimen as consolidation treatment in older AML patients. Whether the FLAG regimen is indeed a better choice of consolidation treatment for older AML patients as compared with other regimens is a clinical problem of interest. The first objective can only be achieved through either retrospective or prospective cohort study with real time data, sound research methodology, and more importantly with a valid statistical analysis. A retrospective cohort study is usually necessary before the conduct of a large confirmatory study as the later would require huge funding and longer duration. With sound methodology, a retrospective cohort study helps answer research questions in a fast and effective manner.

The conduct of the cohort study gives rise to research problems 1.2(c) and 1.2(d) as stated above. The second objective is designed to address those problems, which remain unsolved for regression analysis with proportional hazards model. The second objective thus supports the first objective, and at the same time helps redefine the requirement of EPV for high variable association in small research. This dissertation will provide a new perspective for regression analysis with Cox model when it comes to small research with high impact variable.

1.4 Outline of Dissertation

Chapter 1 discusses the general framework of the dissertation by providing background of research, objectives of research, significance of research and outlines of the dissertation.

Chapter 2 provides a comprehensive literature review on the use of FLAG regimens in AML, relevant simulation studies, a preliminary review on the concept of survival analysis and various methods used in the study. Approaches used to generate random survival times using proportional hazards model is also discussed in this preliminary review.

Chapter 3 describes in details the conduct of a retrospective cohort study to investigate the survival outcomes of older AML patients who were consolidated by the FLAG regimen versus non-FLAG regimens. KM product limit method, log-rank test, and regression analysis with Cox proportional hazards model are used to describe the role and to quantify the effect of the FLAG regimen to patients' survival outcomes.

Chapter 4 describes the design and the conduct of a simulation study to investigate the performance of the Cox coefficient estimate for the primary treatment variable (FLAG

versus non-FLAG) at low EPV condition. The primary treatment variable was found to have high variable association with the survival outcomes.

Chapter 5 provides concluding remarks to all the significant findings based on the cohort study in AML and the simulation study.

CHAPTER 2: LITERATURE REVIEW

Acute myeloid leukemia is a diagnosis for a wide range of myeloid malignancies. The heterogeneous genetic etiology of AML causes treatment outcome and survival rate to vary between young and old patients (Deschler & Lübbert, 2006; Thein et al., 2013; Bethesda, 2015). A study has shown that treated AML patients are more likely to be younger and less likely to have poor performance indicators and comorbidity score when compared with untreated AML patients (Medeiros et al., 2015). AML is also a disease of late adulthood with reported median ages ranging from 65 to 70 years (Appelbaum et al., 2006; Juliusson et al., 2009). The effect of age on both patient and disease related factors results in a higher incidence of early death during chemotherapy, lower rate of CR and a reduced chance for long-term survival. Majority of older AML patients probably do not receive specific treatment, and those who receive standard regimens have a CR rate of 40% to 60% and median survival time of less than a year, which are lower than younger AML patients with a CR rate of above 70% and median survival time of more than 12 months (Estey, 2007; Cheng et al., 2015).

Cytotoxic treatments for AML are well established, and are typically used in induction and consolidation treatments. The induction treatment aims to achieve a CR by reducing the leukemic cells in the bone marrow and blood to normal, while consolidation treatment is given after CR to eliminate the remaining small amount of circulating leukemic blasts in the body. A non-randomized study has shown that post-remission therapy is a prerequisite for long-term survival (Büchner et al., 1985), however the optimal post remission therapy remains unclear and is a question of active research. As there are no generally accepted consolidation protocols for older AML patients, regimens used in induction treatment such as HiDAC and mitoxantrone plus intermittent dose Ara-C (MiDAC) are usually adopted. Besides, older AML patients are often treated with low-

dose Ara-C and palliative therapy as an alternative to the intensive chemotherapy. But outcomes with those alternative treatment are generally inferior with a median survival time of 4 to 9 months, when compared with 12 to 16 months in patients receiving intensive chemotherapy (Tilly et al., 1990; Heiblig et al., 2016).

The gold standard of therapy for AML is known as the '3+7' regimen, with daunorubicin from 45 mg/m^2 to 60 mg/m^2 intravenously daily for 3 days plus Ara-C from 100 mg/m^2 to 200 mg/m^2 as a continuous infusion daily for 7 days (Sperr et al., 2003; Burnett, Wetzler, & Lowenberg, 2011). The '3+7' regimen results in CR rates of more than 40% and long-term survivals of more than 15%. The '3+7' regimen is always used as a comparison standard for a newly developed regimen. Understanding the mechanism of action of Ara-C in AML leads to studies of HiDAC as consolidation treatment in younger patients and those with diploid karyotypes (Kantarjian et al., 2008).

Studies have shown that daunorubicin can be replaced by other anthracyclines such as idarubicin, mitoxantrone, and amsacrine of equivalent dose for improved rates in CR, survival and remission duration. Those outcomes are found in younger AML patients, not older AML patients. The typical risks of induction treatment include infections due to neutropenia, which are more common in older patients; mucositis; skin rash induced by Ara-C; and cardiac side effects of anthracyclines such as cardiomyopathy and arrhythmia. Therefore, induction treatment for AML should only be performed in experienced hematological centers (Krug et al., 2011).

Combination of fludarabine and Ara-C has yielded comparable outcomes in several studies. The regimen has been proven effective for older patients, especially for those with cardiovascular events (Kantarjian et al., 2006). Fludarabine and Ara-C are sometimes given together with G-CSF for refractory and relapse AML. The regimen is known as FLAG, and its widespread use is motivated by an improved accumulation of

more Ara-C-5'-triphosphate (Ara-CTP) at the presence of fludarabine in the leukemic cells. Fludarabine is a purine analog which acts by inhibiting the ribonucleotide reductase that increases the formation of Ara-CTP, an active Ara-C metabolite in the leukemic cells. The action accelerates the destruction of more leukemic blasts. Granulocyte colony-stimulating factor is a glycoprotein which stimulates the bone marrow to produce granulocytes and release them into the bloodstream. The ability of G-CSF in recruiting quiescent cells to s-phase makes the leukemic blasts more sensitive to Ara-CTP attack (Gandhi et al., 1993; Gandhi et al., 1995).

Therapeutic effect of the FLAG regimen can be further intensified by adding other chemotherapeutic agents. The effectiveness and acceptable toxicity profile of FLAG regimen and its intensified versions such as FLAG plus mitoxantrone (Mito-FLAG), FLAG plus idarubicin (FLAG-Ida), and FLAG plus amsacrine (FLAMSA) have been frequently investigated and reported in patients with prior history of myelodysplastic syndrome (MDS), poor prognosis, secondary AML and patients with unfavorable risk factors such as old age and chromosomal abnormalities (Clavio et al., 1996; Estey et al., 1999; Carella et al., 2001; de la Rubia et al., 2002; Ossenkoppele et al., 2004; Ferrara et al., 2005). However, majority of investigations focus primarily on the role of the FLAG regimen as a first line treatment or salvage regimen, rather than as a consolidation treatment.

Cox model is very useful in survival analysis as it allows multiple independent variables to be regressed on survival times through hazard function. Like logistic regression, the performance of the Cox coefficient estimate is affected by the number of outcome events observed rather than the number of subjects followed up. Too few events in relation to the number of independent variables included into a regression model can causes biased coefficient estimate as a result of overfitting the model (Courvoisier et al., 2011). Performance of the Cox model is frequently investigated in simulation studies. In

1995, the impact of varying EPV in Cox model was investigated to establish a definitive EPV value for trustworthy estimates. The simulation study was done based on a cardiovascular cohort study with 36 EPV (252 deaths and 7 variables). With a constant daily hazard rate, the exponential distribution was fitted to generate the survival times to simulate the Cox model. The coefficient estimates of 7 independent variables were assessed from 2 to 25 EPV, and compared with the 'true' coefficient estimates from the cardiovascular cohort study. At the end, the authors concluded that at least 10 EPV are required for trustworthy estimation in regression analysis with Cox model (Concato et al., 1995; Peduzzi et al., 1995). However, the conclusion was rather conservative as numerous factors that could possibly affect the Cox coefficient estimates were not handled in the simulation study (Vittinghoff & McCulloh, 2007). Those factors are strength of association between a variable to the survival outcome, prevalence of the positive value in a binary variable, interaction between variables and sequential selection of variables in regression procedure.

Nonetheless, the simulation study helped reveal most of the problems encountered at very low EPV, like:-

- Low EPV increases bias and may produce both overestimation and underestimation of the true effect,
- Low EPV may cause the loss of normality in the distribution of regression coefficients and increases the chance of falsely extreme values,
- Large sample properties of proportional hazards model variance may not hold at low EPV,
- The power to detect significant effects also decreases at low EPV, causing problem such as "underfitting",
- Low EPV also tends to cause problems in significance testing,

- At very low EPV, the *z*-statistics does not have Gaussian distribution, and the narrow distribution lead to an overly conservative test in which the null hypothesis is rejected less often than the stated significance level, and
- At low EPV, model non-convergence increases.

Besides, the 7 independent variables evaluated in the simulation study had only moderate associations with the survival outcome. So, whether the strength of association modifies the impact of EPV was not ascertained. Lastly, the simulation study was conducted at fixed number of variables, therefore the impact of varying variables was also not known. As a result, the recommendation of 10 EPV becomes the rule of thumb for regression analysis with Cox model (Clark et al., 2003b).

Vittinghoff and McCulloch (2007) conducted a large factorial simulation study to re-evaluate the EPV requirement for both logistic and Cox models. The reason was that the control of confounding in some observational studies may require adjustment for more covariates than the rule of EPV allows. Factors that were not investigated in other simulation studies like sample size, types of variable, and predictor prevalence were examined. The study focuses primarily on confidence interval (CI) coverage for regression coefficient (β), and related type I error rate of the test of null hypothesis (H_0: $\beta = 0$), secondarily on bias in the regression coefficient, and only indirectly on variability and power. Finally, the authors suggested that the rule of 10 EPV could be reduced to 5 to 9 EPV especially for variable with high predictor prevalence.

The requirement of EPV has an implication in the calculation of sample size. Currently, sample size is estimated based on the rule of 10 EPV. To understand factors other than EPV that may affect the estimation of sample size, a simulation study based on a large database of over 2 million anonymized patients was conducted (Ogundimu, Altman, & Collins, 2016). The objective of the study was to investigate the requirement

of EPV for prediction models with a range of binary predictors with varying prevalence that reflect the clinical practice. The results showed that EPV should be data driven, and higher EPV (≥ 20) were required to eliminate bias when many low-prevalence binary predictors are added into an otherwise stable model. The authors concluded that it is difficult to have a definitive EPV value suitable for every situation. An EPV ≥ 20 should be considered when a dataset includes many low-prevalence binary predictors, if this cannot be guaranteed then the used of penalized likelihood approach such as Firth's estimator and Bayesian approaches is recommended as it has been shown to reduce bias in parameter estimates on data with rare events (Lin et al., 2013).

2.1 Preliminary

2.1.1 Survival analysis

Survival analysis aims to analyze time to event. Methods available in survival analysis can be categorized as non-parametric, semi-parametric and parametric methods. Non-parametric methods include life table, KM product limit method, cumulative incidence function (CIF), log-rank test and Gray's test. The life table and KM method are used to describe distribution of survival probability; the CIF is used to analyze competing risk; log-rank test is used to test equality of survival function; and the Gray's test is used to test equality of CIF between two or more groups (Zwiener, Blettner, & Hommel, 2011). Non-parametric methods are very useful in describing the survival experience between two or more groups. Method like KM estimator can be used to plot survival curve for two and more groups and calculate median survival duration with 95% confidence limit. Those methods are non-parametric because of taking no specific form for the distribution of the survival times during the counting process. The only drawback is that those methods do not incorporate the effect of covariates which may affect the survival times (Clark et al., 2003a).

In survival analysis, the effect of covariates can be controlled or adjusted by using a regression model. Covariate adjustment is required to obtain more accurate estimates with higher statistical power (Ford & Norrie, 2002). Regression model is available as semi-parametric model or parametric model in survival analysis. In this dissertation, we focus primarily on the application of non-parametric methods and semi-parametric regression model. In survival analysis, regression analysis primarily involves the use of Cox model because of its robustness. The Cox model is used to estimate the effect of a set of independent variables to patients' survival times. The regression analysis yields a HR that indicates the proportion of the hazard changes between two groups for a binary variable. For a continuous variable, the HR indicates the proportion of hazard changes from a one-unit increase in the continuous variable (George, Seals, & Aban, 2014). According to a simulation study, a regression analysis with Cox model will require at least 10 to 20 EPV for trustworthy estimate (Concato et al., 1995; Peduzzi et al., 1995; Peduzzi et al., 1996).

The hazard is defined as the probability that if an event of interest has not already occurred, it will occur in the next time interval, divided by the length of that interval (Spruance et al., 2004). When the time interval is made very short, the hazard becomes an instantaneous rate for an individual who has already survived up to a certain point in time. The Cox model is semi-parametric as the model does not require a particular probability distribution to represent survival times. However, it does assume that the effects of the independent variables upon survival are constant over time and are additive in one scale (Allison, 2010). For this reason, the Cox model is more robust than other regression models, but only valid for cases that meet the proportional hazards assumption. The proportional hazards assumption requires the HR between two variable groups to be constant over time.

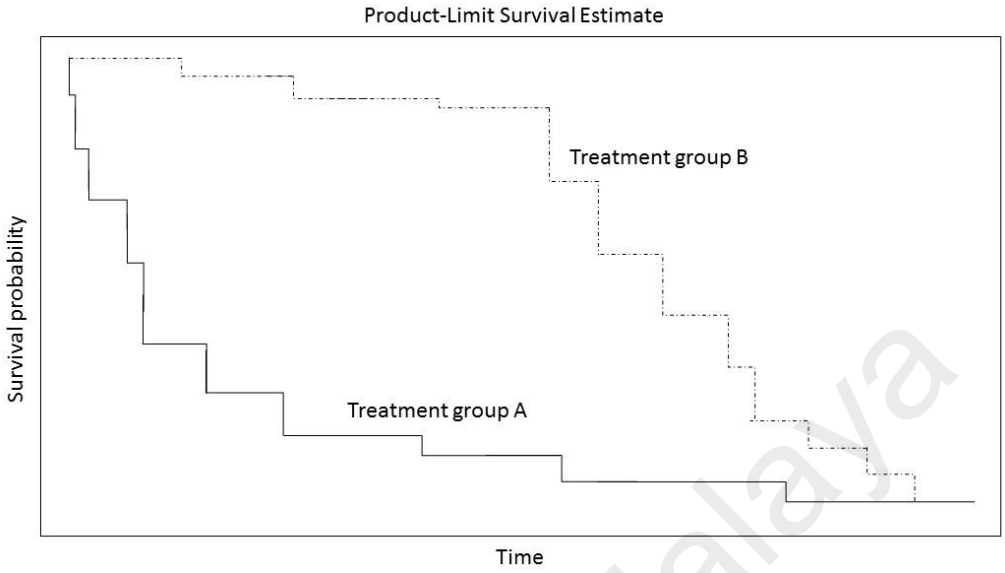

Figure 2.1: The survival curves for two treatment groups with different survival probabilities and similar proportions of outcome events. Treatment group A had a shorter median survival duration as compared with treatment group B.

In survival analysis, an event is the change of quality status in a person like death or occurrence of disease. When death becomes an event of interest in a cohort study, survival times of patients are also called overall survival. Survival time and outcome event are important measures in cancer research. Those measures help differentiate the impact of two interventions that are having identical proportion of outcome events with different time to events (George, Seals, & Aban, 2014). This situation is graphically presented in Figure 2.1. When treatment group A had most of the events happened shortly after the treatment, and treatment group B had less events observed until just before the end of the study, these two groups were considered clinically different. Treatment group A had median survival that was shorter than treatment group B. The survival probabilities and median survival times of these treatment groups can be plotted using KM product limit method.

2.1.2 Censoring in survival analysis

Besides, survival times present few distinct features that cannot be handled by conventional methods. For example, the distribution of survival times is rarely a normal distribution, which causes the ordinary least squares regression method to be inaccurate (Bewick, Cheek, & Ball, 2004). The skewness of the distribution of survival times is mainly caused by the occurrence of many early events and few late ones. This situation is often observed in cancer research. For example, when a disease free duration of cancer patients is measured, most of the recurrences for cancer can happen quite early within few months, but few patients may present prolonged remission (Figure 2.2).

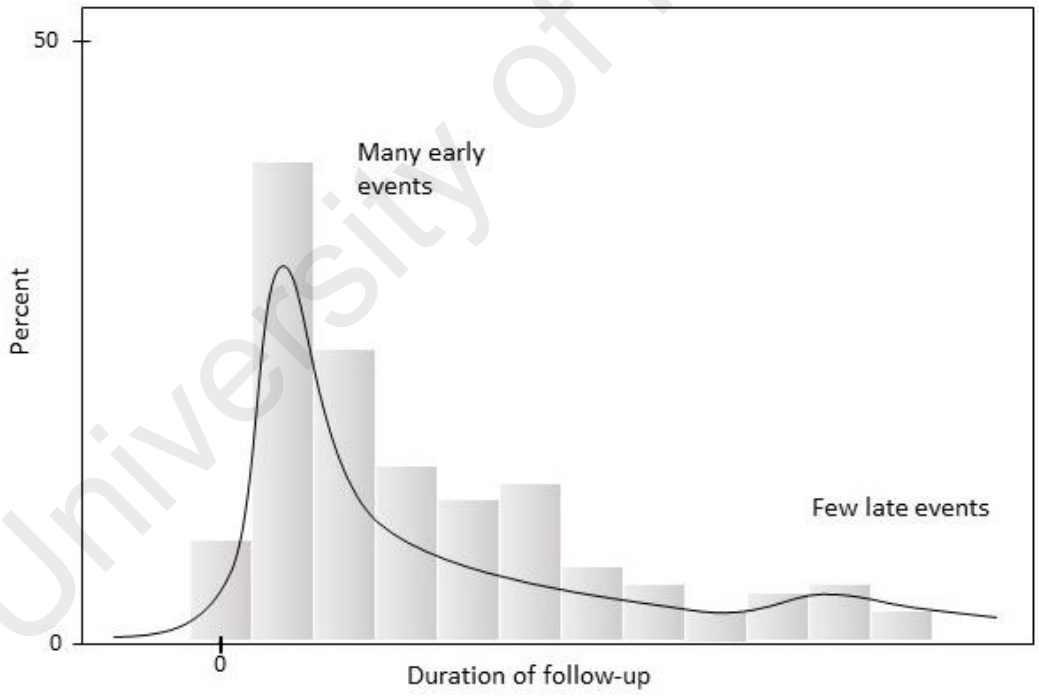

Figure 2.2: The distribution of survival times is often positively skewed due to the occurrence of many early events and relatively few late events.

On the other hand, survival time may consist of few observations with incomplete time to event. This phenomenon is called censoring in survival analysis. Censoring is caused by situation that a subject has not yet experienced an event at the end of study or

loss to follow-up during the study. There are three types of censoring: right censoring, left censoring and interval censoring. Survival analysis successfully deal with the censoring by incorporating the time to event for censored observations into the counting process, and allows comparison between the numbers of survivors in each group at multiple points in time (Spruance et al., 2004). This technique helps reduce considerable bias as a result of covering both complete and incomplete event times. When all event times are analyzed, survival analysis generates more powerful result then other analytical methods (George, Seals, & Aban, 2014).

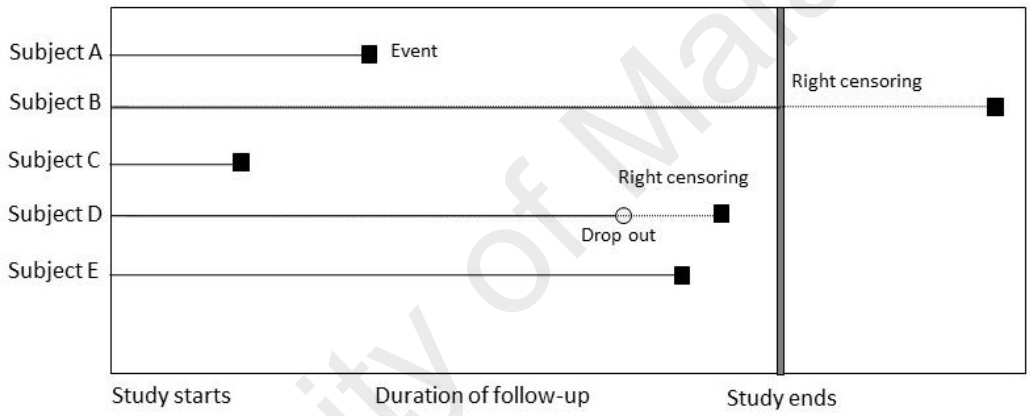

Figure 2.3: Examples of right censoring: The time to event was incomplete for subject B due to the end of study before occurrence of an event. For subject D, the time to event was incomplete due to early drop out.

Right censoring refers to observations that are terminated before the occurrence of events (Allison, 2010). Right censoring is far more common than left censoring and interval censoring in clinical research. Figure 2.3 illustrates some examples of right censoring. For example, subject B was right censored because of not experiencing a disease before the end of study follow-up. Subject D was right censored due to drop-out from the study. As the drop-out happened before the occurrence of the disease, the true event time was basically unknown. Therefore, both subjects B and D had follow-up times that were less than the 'true but unknown' event times. Although the true event times are

uncertain and longer than the observed times, the observed times are still valuable and informative as the fact that the patients went through a certain amount of times without experiencing diseases is itself informative (George, Seals, & Aban, 2014).

Conversely, an observation is left censored when a subject has a follow-up time that is longer than the true event time. Interval censoring happens when the true event time is somehow between a time interval. Although survival times of censored observations can be analyzed in survival analysis, censoring must be prevented as incomplete survival time may still cause either underestimation or overestimation of the 'true' effect. Besides, censoring should be non-informative and not related to the event of interest. Violation of this assumption can invalidate just about any sort of methods used in survival analysis (Clark et al., 2003a).

2.1.3 Kaplan-Meier product limit method

Many methods are available to analyze survival times. For example, the KM method is used to compute survival probability for both censored and uncensored events. The KM method is univariate and non-parametric as it assumes no specific form for the distribution of the survival times and necessarily ignores the effect of covariates (Kaplan & Meier, 1958). The KM survival curve can be generated when the probability of being alive at a particular time is plotted against time by using the equation below,

$$\mathrm{S}(t) = \mathrm{S}(t-1)\left(1 - \frac{\mathrm{d}}{\mathrm{f}}\right), \tag{2.1}$$

where $\mathrm{S}(t)$ is the probability of being alive at time t, $S(t-1)$ is the probability of being alive at time $t-1$, f is the number of patients alive just before time t, and d is the number of events at time t. The occurrence of an event every time changes the value of the $\mathrm{S}(t)$ and turns the KM estimator into a step function, in which the estimated survival probabilities are constant between adjacent event times and only decrease at each

occurrence of an event (Clark et al., 2003a). A KM survival curve describes the survival experience in an intuitive way. The KM method also allows the estimation of median survival time. Median survival time is preferred than mean survival time in survival analysis as the distribution of survival times is rarely normal.

2.1.4 The log-rank test

The comparison of survival functions for two or more groups can be performed with the log-rank test (Peto & Peto, 1972). The log-rank test computes the test statistics, X^2 by comparing the number of observed events, O_i for a particular variable group, i=1, 2, …, g groups to the number of expected events, E_i via the equation,

$$X^2 = \sum_{i=1}^{g} \frac{(O_i - E_i)^2}{E_i} . \quad (2.2)$$

The X^2 is then compared to the Chi-square distribution with (g-1) degrees of freedom. This method allows a *p*-value to be generated to assess the statistical significance of the difference between two survival curves.

2.1.5 The Cox model

Both KM method and log-rank test do not allow the adjustment for the effect of covariates. Adjustment for the effect of co-variates effect helps improve the estimation and can only be achieved with regression modelling. The most popular regression model used in survival analysis is the Cox model. In the Cox model, survival times are modelled through hazard function, which indicates the probability of having an event given that the subjects have survived up to a given point of time (Bewick, Cheek, & Ball, 2004). The Cox model is expressed as

$$h(t|\mathbf{X}) = h_0(t)\exp(\mathbf{X\beta}), \quad (2.3)$$

where $h(t|\mathbf{X})$ is the hazard rate at time t given $\mathbf{X}$, t is the time to event for each individual, $\mathbf{X}$ is a vector of one or more independent variables $X_1, \ldots, X_p$. For the j-th individual let the values of $\mathbf{X}$ be $\mathbf{X_j} = (X_{1j}, \ldots, X_{pj})$ for $j = 1, 2, \ldots, n$, $\boldsymbol{\beta}$ is a $p \times 1$ vector of unknown parameters and $h_0(t)$ is the baseline hazard function for a standard set of conditions $\mathbf{X} = \mathbf{0}$.

Through the Cox model, we can easily quantify the relationship of a variable of interest (e.g. new treatment) to the survival time in control of several explanatory covariates such as age, gender and race. The baseline hazard, $h_0(t)$ is an unspecific function. For this reason, the Cox model becomes a semi-parametric model and robust. The Cox model assumes that the variables act multiplicatively on the hazard at any point in time and the effect is constant over time (George, Seals, & Aban, 2014). The assumption is called proportional hazards assumption and is one of the most important assumptions for regression analysis with Cox model.

2.1.6 Generating survival times to simulate the Cox model

The performance of Cox coefficient estimate under certain pre-specified conditions can be evaluated through a simulation study. However, the simulation study is not straight forward as the Cox model is formulated through a hazard function. The simulation study requires random survival times to be generated based on a parametric distribution. Exponential and Weibull distributions are commonly used in simulation study of Cox model because of sharing the assumption of proportional hazards (Bender, Augustin, & Blettner, 2005). The steps required in generating random survival times to simulate Cox model using the Exponential or Weibull distribution have been outlined by Bender, Augustin, & Blettner (2005).

The survival function, $S(t|\mathbf{X}_j)$ of the Cox model (Equation (2.3)) is given by

$$S(t|\mathbf{X}_j) = \exp\left[-h_0(t) \times \exp(\mathbf{X}_j\boldsymbol{\beta})\right], \quad (2.4)$$

where $h_0(t) = \int_0^t h_0(v)\, dv$ is the cumulative baseline hazard function.

The distribution function of the Cox model is given by

$$F(t|\mathbf{X}_j) = 1 - \exp\left[-h_0(t) \times \exp(\mathbf{X}_j\boldsymbol{\beta})\right]. \quad (2.5)$$

Let Y be a random variable with distribution function F, then U=F(Y) follows a uniform distribution on the interval from 0 to 1, abbreviated as Uni [0,1]. Moreover, if U~Uni[0,1], then (1–U)~Uni[0,1] too. Let T be the survival time of Cox model, then it follows from Equation (2.5) that

$$U = \exp\left[-h_0(T) \times \exp(\mathbf{X}_j\boldsymbol{\beta})\right] \sim \text{Uni}[0,1]. \quad (2.6)$$

If $h_0(t) > 0$ for all t, then h_0 can be inverted and the survival time T of the Cox model can be expressed as

$$T = h_0^{-1}\left[-\log(U) \times \exp(-\mathbf{X}_j\boldsymbol{\beta})\right]. \quad (2.7)$$

The inverse of the cumulative hazard function for the exponential distribution with constant hazard rate, γ is given by

$$h_0^{-1}(t) = \gamma^{-1}t\,. \quad (2.8)$$

By inserting Equation (2.8) into Equation (2.7), survival times of a Cox model with constant baseline hazard can be expressed by

$$T = \gamma^{-1}\left[-\log(U) \times \exp(-\mathbf{X}_j\boldsymbol{\beta})\right]. \quad (2.9)$$

The inverse of the cumulative hazard function for the Weibull distribution with scale parameter, λ and shape parameter, α is given by

$$h_0^{-1}(t) = (\lambda^{-1}t)^{\frac{1}{\alpha}}. \tag{2.10}$$

By inserting Equation (2.10) into Equation (2.7), survival time of a Cox model with baseline hazard of the Weibull distribution can be generated by

$$T = \left(\lambda^{-1}\left[-\log(U) \times \exp(-\mathbf{X}_j\boldsymbol{\beta})\right]\right)^{\frac{1}{\alpha}}. \tag{2.11}$$

In practice, the assumption of a constant hazard rate is rarely practical, therefore the exponential distribution is of limited use. With a shape parameter, the Weibull distribution becomes more flexible and can fit more types of data. This attributes to the popularity of Weibull distribution in many applications (Nelson, 1982).

CHAPTER 3: FLUDARABINE, HIGH DOSE CYTARABINE AND GRANULOCYTE COLONY-STIMULATING FACTOR (FLAG) AS CONSOLIDATION CHEMOTHERAPY IN OLDER PATIENTS WITH ACUTE MYELOID LEUKEMIA: A RETROSPECTIVE COHORT STUDY

3.1 Introduction

Chemotherapy based on high dose infusional Ara-C and anthracyclines is the primary treatment option for AML. This combination is therapeutically intensive and can only be given to a clinically fit patient. Most of the AML patients need to undergo multiple cycles of chemotherapy from induction to consolidation. The induction chemotherapy aims to achieve CR by reducing the total leukemic cell population in the body from around 10^{12} cells to below the cytologically detectable level of about 10^{9} cells. The consolidation chemotherapy is given to sustain the disease remission by eliminating the remained leukemic blasts that are undetected by the current diagnostic tests. Consolidation chemotherapy is the pre-requisite for long term survival in patients. However, the optimal regimen for consolidation especially in older AML patients remains unclear (Krug et al., 2011).

The combination of fludarabine and Ara-C has been advocated as effective and safe for most refractory and relapsed AML because of higher remission rate and low toxicity profile (Pastore et al., 2003; Lee et al., 2009; Luo et al., 2013). The regimen is sometimes given simultaneously with G-CSF for better efficacy. The combination is called FLAG. Therapeutic effect of the FLAG regimen can be further intensified by adding other agents into the regimen, for example Mito-FLAG, FLAG-Ida, FLAMSA and all-trans retinoic acid. The potential benefits of the FLAG regimen and its intensified versions have been evaluated in patients with refractory AML and patients with poor prognosis or unfavorable risk factors such as old age, chromosomal abnormalities and history of myelodysplastic syndrome (MDS). The study outcomes were highly

appreciated (Clavio et al., 1996; Estey et al., 1999; Carella et al., 2001; de la Rubia et al., 2002; Alwan et al., 2014).

A synergistic anti-leukemic mechanism is believed to exist among the member agents in FLAG regimen. The presence of fludarabine increases the formation of Ara-CTP, an active metabolite in the circulating leukemic blasts and this accelerates the destruction of more leukemic blasts. The presence of G-CSF stimulates the bone marrow to produce more granulocytes and stem cells. This ability in recruiting quiescent cells to s-phase could make the leukemic blasts more sensitive to Ara-CTP attack (Gandhi et al., 1993; Gandhi et al., 1995).

Most patients who undergo induction chemotherapy will enter CR. However, the most common cause of AML death is subsequent relapse, or in another context the lack of superior chemotherapy regimens for post remission treatment (Ramos et al., 2015). Most guidelines including European Society for Medical Oncology (ESMO) and National Comprehensive Cancer Network (NCCN) recommend standard induction chemotherapy with 3 days anthracycline and 7 days Ara-C, with or without additional agent (Fey & Buske, 2013; “NCCN Clinical Guidelines in Oncology, Acute Myeloid Leukemia," version 2.2014), but there are no recommendations about the use of the FLAG regimen as consolidation chemotherapy. This may be caused by the lack of prospective trials with the use of the FLAG regimen, as well as the number of cycles in post remission treatment. Therefore, a cohort study is urgently needed to assess and compare the survival outcomes with regards to the use of FLAG as the consolidation chemotherapy in older AML patients.

3.2 Materials and Methods

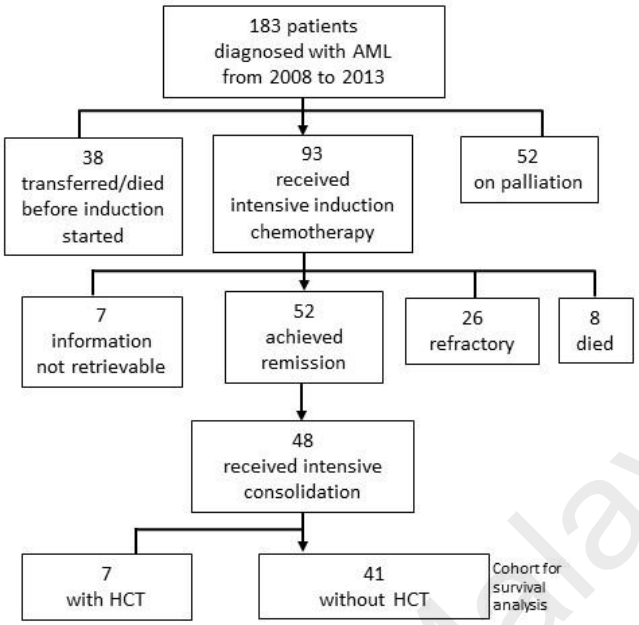

Figure 3.1: The CONSORT diagram for the retrospective cohort study. A total of 183 patients were evaluated. At the end, only 41 patients were eligible and included in the survival analysis.

The retrospective cohort study was conducted at Ampang Hospital, the national referral center for hemato-malignancies and treatments in Malaysia. Figure 3.1 summarizes the selection of eligible patients from a total of 183 patients diagnosed with AML (excluding acute promyelocytic leukemia, APML) from 2008 to 2013. The percentage of older AML patients (≥ 54 years old) who were treated with intensive chemotherapy was only 50.8%. From the 52 patients who achieved CR, 48 patients were further consolidated by chemotherapy. Finally, a total of 41 patients who had not undergone hematopoietic cell transplantation (HCT) were enrolled for the cohort study. Data of interest were retrieved and analyzed after the study protocol had been approved by the Medical Research and Ethics Committee and the Hospital. Characteristics of patients such as age, race, comorbidities found at diagnosis, types of induction chemotherapy, French-American-British (FAB) sub-types and survival outcomes were summarized according to the types of consolidation regimens (FLAG and non-FLAG).

The characteristics of patients are presented in Table 3.1. The cut-off age in AML is arbitrary. In this study, age of 54 years was used based on the suitability of patients to receive intensive chemotherapy according to institutional practice.

Table 3.1: The characteristics of the older AML patients consolidated by the FLAG and non-FLAG regimens.

Characteristics	Total (41 patients)	FLAG (34 patients)	non-FLAG (7 patients)
	n (%)	n (%)	n (%)
1. Age at diagnosis, years			
• Median (Min, Max)	61.2 (54.1, 75.3)	61.7 (54.3, 75.3)	58.7 (54.1, 64.3)
2. Gender			
• Male	17 (41.5)	11 (32.4)	6 (85.7)
• Female	24 (58.5)	23 (67.6)	1 (14.3)
3. Race			
• Malay	20 (48.8)	17 (50.0)	3 (42.8)
• Chinese	15 (36.6)	13 (38.2)	2 (28.6)
• Indian	6 (14.6)	4 (11.8)	2 (28.6)
4. FAB subtypes			
• M2	6 (14.6)	5 (14.7)	1 (14.3)
• M4	9 (22.0)	8 (23.5)	1 (14.3)
• M5	6 (14.6)	5 (14.7)	1 (14.3)
• M6	1 (2.4)	1 (2.9)	0 (0)
• Not specified	19 (46.3)	15 (44.1)	4 (57.1)
5. Number of comorbidities at diagnosis			
• 4	2 (4.9)	2 (5.9)	0 (0)
• 3	8 (19.5)	7 (20.6)	1 (14.3)
• 2	12 (29.3)	10 (29.4)	2 (28.6)
• 1	7 (17.0)	7 (20.6)	0 (0)
• 0	12 (29.3)	8 (23.5)	4 (57.1)
6. Regimen for induction treatment			
• Anthracyclines + Ara-C	39 (95.2)	32 (94.2)	7 (100.0)
• Anthracyclines + Ara-C + etoposide	1 (2.4)	1 (2.9)	0 (0)
• Anthracyclines + etoposide + cyclophosphamide	1 (2.4)	1 (2.9)	0 (0)
7. Survival outcome			
• Alive	12 (29.3)	12 (35.3)	0 (0)
• Loss to follow-up	3 (7.3)	3 (8.8)	0 (0)
• Death	26 (63.4)	19 (55.9)	7 (100.0)
- Chemotherapy related death	6 (14.6)	4 (11.8)	2 (28.6)

The result was presented in absolute count (n) and percentage (%), unless otherwise specified. FAB: French-American-British; Anthracyclines include mitoxantrone and daunorubicin. Min: minimum; Max: maximum.

The eligibility for study was determined based on history of remission and chemotherapy. First, a patient must undergo both induction and consolidation treatments before relapse or death. A CR must be achieved before the start of a consolidation treatment. A chemotherapy is classified as a consolidation treatment only if it is used after the attainment of a CR. A CR is defined as clearance of leukemic blasts in the bone marrow to less than 5 percent of all nucleated cells, morphologically normal hematopoiesis and return of peripheral blood cells count to normal level (Fey & Buske, 2013). In this study, we excluded patients who had received hematopoietic cell transplantation (HCT) as transplantation is a totally different treatment modality for AML and transplantation associated mortality could be a competing risk to the outcome event.

The post induction treatment response was confirmed by bone marrow aspiration and/or biopsy and complete blood counts. The OS is measured from the date of diagnosis until death of any causes, with observations censored for patients last known alive. The DFS is measured from the date of attainment of a CR to recurrence of AML or death of any causes (Delgado et al., 2014). The date of diagnosis, CR and event are based on date documented in the patient's electronic medical record. We defined chemotherapy related mortality as death within 30 days after chemotherapy.

The study cohort included patients with a median age at diagnosis of 61.2 years old (range, 54.1 – 75.3 years), majority female (58.5%) and Malay (48.8%) patients. Majority of patients (29 patients, 70.7%) were found to have at least one comorbidity at diagnosis. Standard induction regimen with anthracycline plus Ara-C was used in 39 out of 41 patients (95.1%). There were about 26 deaths (event for OS) and 27 relapses/deaths (event for DFS) found during the study period. Fifteen patients were right censored including 3 cases of loss to follow-up. The minimum follow-up duration were 24 months.

3.2.1 Consolidation chemotherapy

The types of chemotherapy and the number of cycles used as consolidation are summarized and presented in Table 3.2.

Table 3.2: The types of chemotherapy and the number of cycles used as consolidation in the study cohort.

Consolidation chemotherapy	**Number of Patients**	**Cycles of induction chemotherapy (mean)**	**Cycles of consolidation chemotherapy (mean)**	**Total cycles of chemotherapy (mean)**
FLAG group	**34**	**57 (1.4)**	**86 (2.1)**	**143 (3.5)**
A. FLAG only	19	28 (1.5)	32 (1.7)	60 (3.2)
a. FLAG	11			
b. FLAG-gemtuzumab	6			
c. FLAG + FLAG-gemtuzumab	2			
B. Combination	15	20 (1.3)	40 (2.7)	60 (4.0)
a. FLAG + MiDAC	4			
b. FLAG + MiDAC + HiDAC	4			
c. FLAG + maintenance	3			
d. FLAG + HiDAC	2			
e. FLAG-IDA + HiDAC + maintenance	1			
f. FLAG- gemtuzumab + HiDAC	1			
Non-FLAG group	**7**	**9 (1.3)**	**14 (2.0)**	**23 (3.3)**
a. HiDAC + MiDAC	3			
b. HiDAC + MiDAC + maintenance	1			
c. HiDAC	1			
d. DA + azacitidine	1			
e. HiDAC-Ida	1			

DA: daunorubicin plus Ara-C; Ida: Idarubicin; Maintenance includes Ara-C + etoposide, Ara-C + thioguanine or decitabine

Of the 41 eligible patients, 34 (82.9%) patients were consolidated with at least a cycle of FLAG and were designated as 'FLAG' group; the remaining 7 (17.1%) patients receiving other types of chemotherapy for consolidation were designated as 'non-FLAG'

group. Among those patients in the FLAG group, 19 (46.3%) patients were further designated as 'FLAG only' group as they received only FLAG for consolidation and 15 patients (36.6%) were further designated as 'combination' group as they were consolidated with combination of FLAG and other regimens. Other regimens used as consolidation treatment were HiDAC and MiDAC. Maintenance chemotherapy with low dose Ara-C, thioguanine, decitabine and etoposide were given to patients who were considered unfit for more intensive treatments. The total number (mean) of cycles of chemotherapy delivered to patients were 143 (3.5) cycles including 60 (3.2) cycles for the FLAG only group, 60 (4.0) cycles for the combination group and 23 (3.3) cycles for the non-FLAG group. The number (mean) of cycles of chemotherapy used as consolidation were 32 (1.7) cycles for the FLAG only group, 40 (2.7) for the combination group and 14 (2.0) for the non-FLAG group.

There was not specific institutional standard that determined the type of chemotherapy a patient should receive for consolidation. The choice of chemotherapy was primarily physician oriented, while considering other factors such as patient's motivation and family supports, cytogenetics, type of comorbidities and so on. Patient's motivation and family supports were very crucial as intensive chemotherapy like the FLAG regimen was more likely to cause longer hospitalization and more complications.

The FLAG regimen was once used with gemtuzumab ozogamicin, a drug-linked monoclonal antibody which was marketed by Wyeth (now Pfizer) as Mylotarg from 2000 to 2010. The monoclonal antibody was withdrawn from the market in June 2010 when a clinical trial failed to demonstrate additional clinical benefit (survival time) in AML patients and observed a greater number of deaths occurred in the test group (Jefferson, 2010). In this study, a subset of patients who received FLAG plus gemtuzumab ozogamicin for consolidation was also designated as the 'FLAG only' group.

3.2.2 Statistical analysis

The survival data was analyzed by using SAS software, version 9.4. The characteristics of patients were summarized using descriptive statistics by types of consolidation treatment (FLAG and non-FLAG). In the analysis, survival outcomes were compared as follows: FLAG versus non-FLAG, FLAG only versus non-FLAG, and combination versus non-FLAG. The KM method was used to plot survival curves and to estimate median survival duration. Equality of survival curves between groups was tested by using the log-rank test with significance level set at 0.05. Bonferroni adjustment was used in multiple strata comparison for log-rank test to maintain a familywise type I error rate of 0.05.

A regression analysis with Cox model were performed to estimate the effect of the consolidation treatment (FLAG and non-FLAG) to OS and DFS. The effect of consolidation was quantified by using HR and adjusted for other explanatory variables such as age, race and sex. To ensure valid results, the proportional hazards assumption was checked by using cumulative sums of martingale-based residuals and tested with supremum test[1]. A *p*-value, $p > 0.05$ would indicate that the assumption holds for the Cox model (Lin, Wei, & Ying, 1993).

A HR less than one (negative regression coefficient) indicates that the use of FLAG as consolidation treatment was associated with a protective effect. If this is true, the FLAG group should experience longer OS and DFS when compared to the non-FLAG group. As the HR does not translate directly into information about the duration of time until events, a ratio of median survival durations between groups should be reported (Spruance et al., 2004). For this reason, we derived the ratio of median survival durations

[1] Cumulative sums of martingale-based residuals is a plot of standardized score residuals over time for checking the adequacy of the Cox model. If the residuals get unusually large at any time point, the method suggests the proportional hazards assumption may not hold for a particular Cox model.

for the FLAG, FLAG only and combination versus the non-FLAG group. If consolidation with FLAG regimen was truly effective, the ratio of median survival times should be greater than one.

The survival models had 6.5 EPV for OS (26 deaths divided by 4 variables) and 6.75 EPV for DFS (27 relapses/deaths divided by 4 variables). As both survival models had EPV values less than the 10 to 20 EPV rule recommended (Peduzzi et al., 1995), a simulation study was conducted to assess the performance of the coefficient estimates of the primary treatment variable (FLAG and non-FLAG) at low EPV condition. Further details of the simulation study can be found in the Chapter 4.

3.3 Results

3.3.1 Kaplan-Meier survival curves and median survival durations

Relative to the non-FLAG group, the FLAG group consisted of patients with median age at diagnosis of 3 years older (61.7 years versus 58.7 years) and majority female (67.6% versus 14.3%). Majority of patients in the FLAG group had at least 2 extra comorbidities (55.9%) while majority of patients in the non-FLAG group were free of comorbidity at diagnosis (57.1%). Chemotherapy regimens used for induction treatment were similar in both groups. Standard induction protocol with anthracyclines and high-dose Ara-C was used in more than 94% of patients in the FLAG group and all patients in the non-FLAG group (Table 3.1). All patients had entered CR before undergoing consolidation treatment. Among the 26 deaths in the cohort, 19 (55.9%) cases of death were from the FLAG group and 7 (100.0%) cases of death were from the non-FLAG group. The chemotherapy related mortality were 6 (14.6%) cases in total, with 4 (11.8%) cases from the FLAG group and 2 (28.6%) cases from the non-FLAG group. Within the FLAG group, 10 (52.6%) cases of death with 3 (15.8%) cases of chemotherapy related

death were from the FLAG only group; 9 (60.0%) cases of death with only 1 (6.7%) case of chemotherapy related death were from the combination group.

Figure 3.2 shows the OS curves of the FLAG group versus non-FLAG group. The survival curves were significantly different (log-rank, p = 0.0025). The median OS was longer for the FLAG group when compared to the non-FLAG group (18.70 vs 8.09 months). The ratio of median OS (FLAG: non-FLAG) was 2.31. Figure 3.3 shows the DFS curves of the FLAG group versus non-FLAG group. The survival curves were significantly different (log-rank, p = 0.0026). The median DFS was longer for the FLAG group when compared to the non-FLAG group (13.84 vs 4.44 months). The ratio of median DFS (FLAG: non-FLAG) was 3.12.

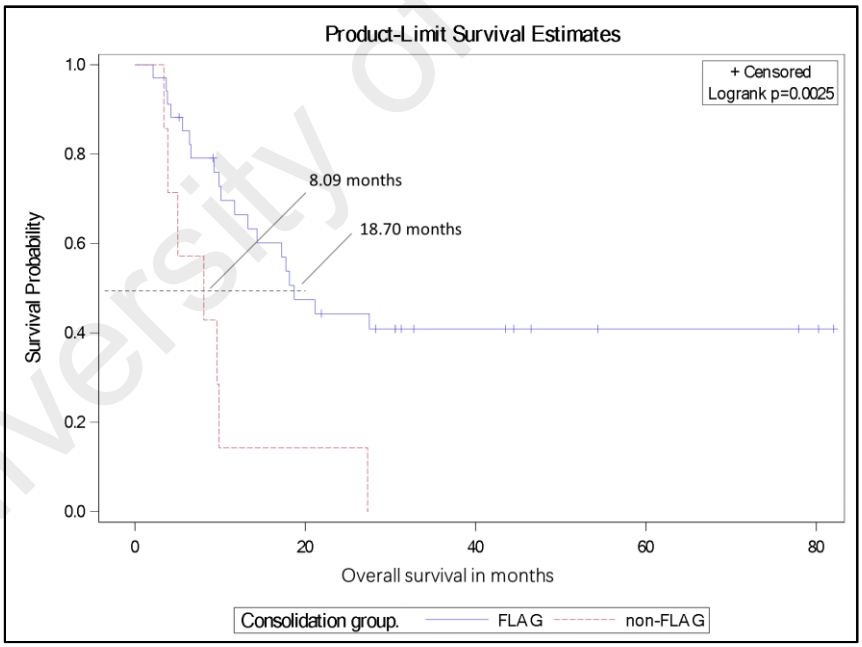

Figure 3.2: The KM survival curves and median OS for older AML patients consolidated with the FLAG and non-FLAG regimens.

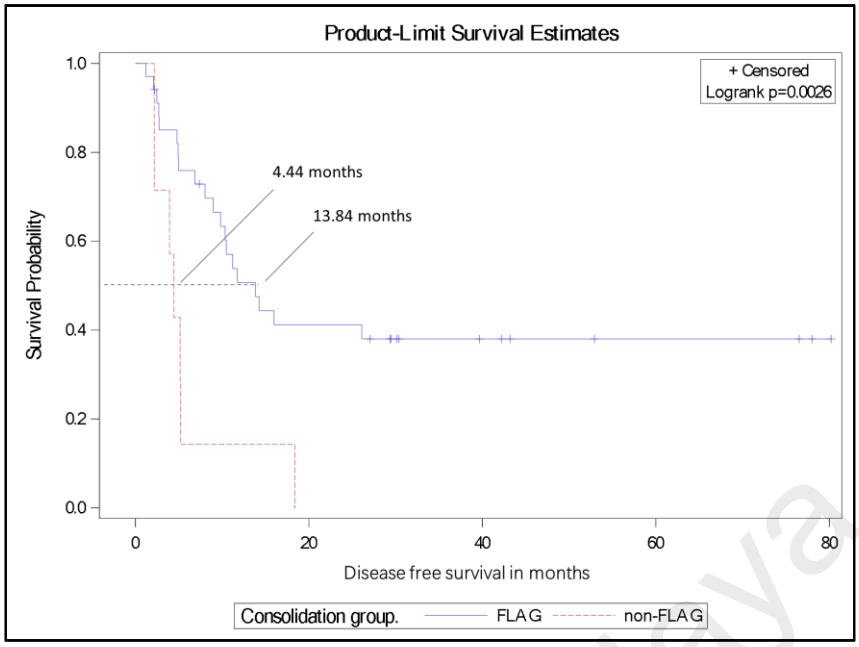

Figure 3.3: The KM survival curves and median DFS for older AML patients consolidated with the FLAG and non-FLAG regimens.

Figure 3.4 shows the OS curves for the FLAG only, combination and non-FLAG groups. The survival curves were significantly different (log-rank, p = 0.0089). The combination group had the longest median OS of 24.32 months, followed by the FLAG only group of 17.72 months when compared with the non-FLAG group of 8.09 months. When Bonferroni correction was applied, only the combination group had the OS curve significantly different from the non-FLAG group (Table 3.3). The ratios of median OS were 2.19 for the FLAG only group and 3.01 for the combination group, when compared with the non-FLAG group.

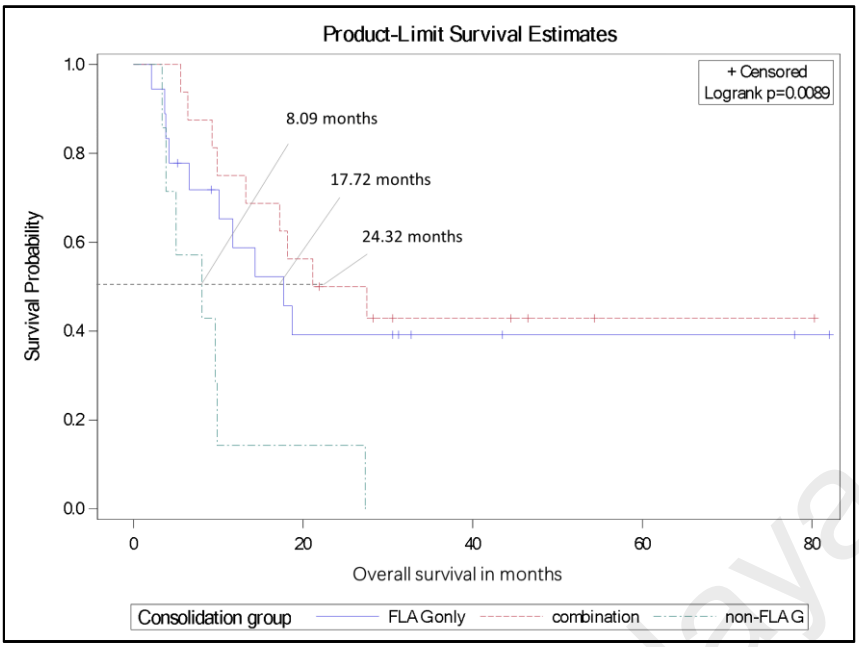

Figure 3.4: The KM survival curves and median OS for older AML patients consolidated with the FLAG only, combination and non-FLAG regimens.

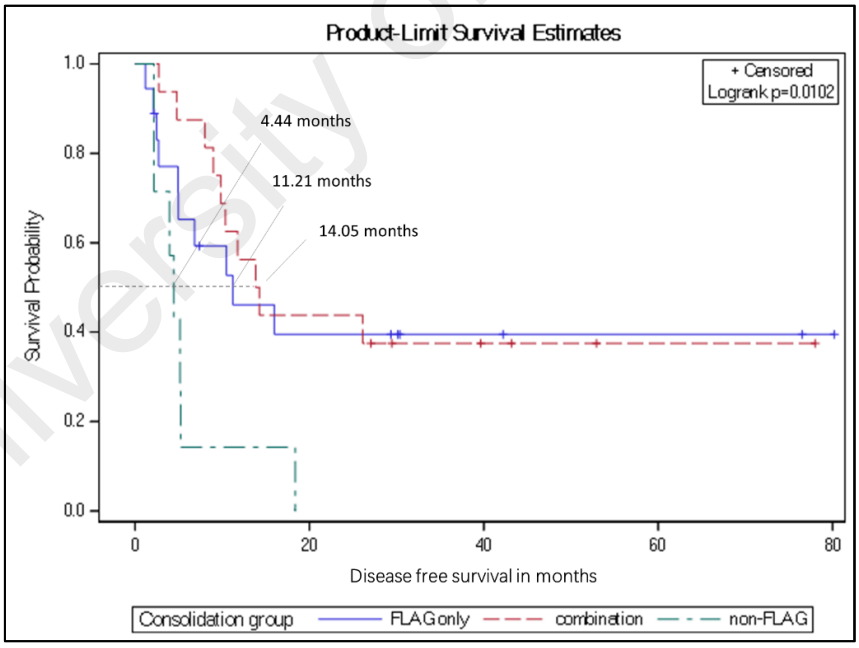

Figure 3.5: The KM survival curves and median DFS for older AML patients consolidated with the FLAG only, combination and non-FLAG regimens.

Figure 3.5 shows the DFS curves for the FLAG only, combination and non-FLAG groups. The survival curves were significantly different (log-rank, p = 0.0102). The

combination group had the longest median DFS of 14.05 months, followed by the FLAG only group of 11.21 months when compared with the non-FLAG of 4.44 months. When Bonferroni correction was applied with significance level set at 0.0167, no significant differences were found among the three groups (Table 3.3). The ratios of median DFS were 2.52 for the FLAG only group and 3.16 for the combination group, when compared with the non-FLAG group.

Table 3.3: The multiple strata comparisons for the log-rank test for both OS and DFS.

Strata Comparison	**OS**		**DFS**	
	Chi-Square	***p*-value**	**Chi-Square**	***p*-value**
FLAG only vs non-FLAG	2.6736	0.1020	3.5491	0.0823
Combination vs non-FLAG	6.0716	0.0137	5.0084	0.0214
FLAG only vs combination	0.3465	0.5561	0.0640	0.8003

a. With Bonferroni correction, the significance level is set at 0.05/3=0.0167.

3.3.2 Regression analysis with the Cox model

Table 3.4 summarizes the results for both univariable and multivariable regression analysis. Univariable regression analysis refers to the construction of a Cox model with an independent variable only. The coefficient estimate of the independent variable generated is not adjusted for other variable effects. Multivariable regression analysis refers to the construction of a Cox model with more than one independent variable. In this study, four independent variables were used. The assumption of proportional hazards was satisfied for all variables as presented by Table 3.5.

Table 3.4: The unadjusted and adjusted HR of the treatment variable for both OS and DFS.

Comparisons	**Univariable (unadjusted)**		**Multivariable (adjusted)**	
	HR (95% CI)	***p*-value**	**HR (95% CI)**	***p*-value**
OS				
FLAG versus non-FLAG	0.274 (0.112-0.671)	0.0046	0.245 (0.085-0.708)	0.0094
a. FLAG only versus non-FLAG	0.313 (0.116-0.843)	0.0216	0.389 (0.097-1.568)	0.1845
b. Combination versus non-FLAG	0.239 (0.086-0.661)	0.0058	0.214 (0.070-0.650)	0.0065
DFS				
FLAG versus non-FLAG	0.277 (0.113-0.676)	0.0048	0.217 (0.072-0.656)	0.0068
a. FLAG only versus non-FLAG	0.302 (0.112-0.814)	0.0180	0.279 (0.069-1.125)	0.0727
b. Combination versus non-FLAG	0.254 (0.094-0.691)	0.0072	0.201 (0.064-0.632)	0.0060

a. Information of the full regression models can be found in APPENDIX C.
b. The HR is obtained by taking exponentiation of the coefficient estimates as shown in APPENDIX C.

In the univariable regression analysis for OS, statistically significant difference in the hazard rate was obtained in the FLAG group versus non-FLAG group (HR = 0.274, $p = 0.0046$), in the FLAG only group versus non-FLAG group (HR = 0.313, $p = 0.0216$) and in the combination group versus non-FLAG group (HR = 0.239, $p = 0.0058$). When the outcomes were adjusted for covariate differences, statistically significant difference in the hazard rate was obtained in the FLAG group versus non-FLAG group (HR = 0.245, $p = 0.0094$) and in the combination group versus non-FLAG group (HR = 0.214, $p = 0.0065$).

In the univariable regression analysis for DFS, statistically significant difference in the hazard rate was obtained in the FLAG group vs non-FLAG group (HR = 0.277, $p = 0.0048$), in the FLAG only group versus non-FLAG group (HR = 0.302, $p = 0.0180$) and in the combination group versus non-FLAG group (HR = 0.254, $p = 0.0072$). When the outcomes were adjusted for covariate differences, statistically significant difference in the hazard rate was obtained in the FLAG group versus non-FLAG group (HR = 0.217, $p = 0.0068$) and in the combination group versus non-FLAG group (HR = 0.201, $p = 0.0060$).

Table 3.5: The results of supremum test for checking the proportional hazards assumption.

Model	**Variables**	**Maximum Absolute Value**	***p*-value**
OS	FLAG versus non-FLAG	0.6819	0.6420
	Race- Chinese	1.1201	0.1500
	Race - Indian	0.8993	0.2920
	Sex - female	0.5404	0.8210
	Age at diagnosis	0.4981	0.7780
OS	FLAG only versus non-FLAG	1.4396	0.3210
	Combination versus non-FLAG	0.8930	0.4850
	Race- Chinese	1.1106	0.1650
	Race - Indian	0.8822	0.2960
	Sex - female	0.7597	0.7740
	Age at diagnosis	0.4861	0.8780
DFS	FLAG versus non-FLAG	0.8677	0.4210
	Race- Chinese	0.7906	0.4790
	Race - Indian	1.0157	0.2040
	Sex - female	0.5966	0.6920
	Age at diagnosis	0.8851	0.2200
DFS	FLAG only versus non-FLAG	1.6610	0.2060
	Combination versus non-FLAG	1.1189	0.2790
	Race- Chinese	0.8193	0.4540
	Race - Indian	1.0066	0.2090
	Sex - female	0.6997	0.7750
	Age at diagnosis	0.8987	0.2700

3.3.3 Performance of regression coefficients at low EPV condition

The number of outcome events in relation to the number of independent variables used in the regression analysis with Cox model for both OS and DFS do not meet the 10 to 20 EPV rule. Therefore, the performance of the coefficient estimate of the primary

treatment variable was checked in a simulation study. Based on the result of the simulation study, we confirmed that the Cox model required only 5 to 6 EPV for reliable estimation of the 'true' effect of the treatment variable for both OS and DFS. More works and discussions follow in Chapter 4.

3.4 Discussion

The consolidation treatment variable (FLAG and non-FLAG) had a strong association with both OS and DFS as shown in Figures 3.2 to 3.5. Patients who were consolidated with at least a cycle of the FLAG regimen had a significant longer median OS and median DFS than patients who received only non-FLAG regimens. Similar trend was observed for patients who received the FLAG only regimen and the combination regimens.

It is noteworthy that although the FLAG only group had a lower mean cycle of chemotherapy than the non-FLAG group, patients in the FLAG only group had experienced longer median OS and median DFS than the non-FLAG group. Patients with the combination regimens recorded the longest median OS and median DFS when compared to the FLAG only group and non-FLAG group. It is observed that the combination group also recorded the highest number and mean cycles of consolidation chemotherapy, followed by the non-FLAG group and FLAG only group. More cycles of consolidation chemotherapy during consolidation might delay disease progression and relapse, causing longest median OS and median DFS in the combination group. However, we could not determine the causal relationship between the number of consolidation cycles and patient's survival due to limitations in the study design and sample sizes.

Lowest mortality rate was observed in the FLAG only group when compared with the combination group and non-FLAG group. However, the FLAG only group was having higher chemotherapy related mortality when compared with the combination group, but

still lower than the non-FLAG group. Majority of patients in the non-FLAG group even though did not have comorbidities, but were not considered for more intensive chemotherapy like FLAG. This indicated that the number of comorbidity a patient presented at diagnosis was not the primary consideration for intensive chemotherapy.

A protecting effect was associated with the use FLAG regimen as consolidation treatment in older patients. This protecting effect caused significant reduction in hazards by approximately 60% to 80% and led to longer median survival in both OS and DFS. The strong variable association to the survival outcome allowed the regression analysis with Cox model to reliably estimate the true effect of the FLAG regimen even at low EPV condition. Our simulation study showed that the number of outcome events was sufficient. The EPV threshold was identified at 5 to 6 EPV. Above the threshold, the performance of the coefficient estimate of the treatment variable was stable and less likely to be affected by increasing EPV. More discussion about the EPV requirement for variable with strong association to survival outcome can be found in Chapter 4.

Post induction remission have improved over years, but AML relapse remains the main cause of treatment failure. Instead of treating the patients after the relapse, the FLAG regimen and its intensified version seem somewhat protective before the occurrence of a relapse. The effectiveness of FLAG and its intensified versions relies on the enhanced intracellular conversion of Ara-C to its active metabolite Ara-CTP in the presence of fludarabine and G-CSF (Ossenkoppele et al., 2004). The enhanced killing mechanism at the intracellular level might prevent the mutational adaptation in the remaining small number of leukemic blasts which could eventually lead to relapse.

In older patients, AML is highly heterogeneous and the treatment is often complicated by problems such as comorbidities, performance status, tolerance, risk of early death, pre-existing MDS and lack of motivation. Even majority of older patients up

to 80 years of age can tolerate and benefit from intensive chemotherapy, less than half would eventually be treated intensively for longer survival (Juliusson et al., 2009; Oran & Weisdorf, 2012). The effectiveness and toxicity profile of FLAG and its intensified versions have been investigated and reported in some clinical studies involving older patients in relapse and refractory AML, secondary AML, as well as those with poor prognosis and poor performance status. The highly appreciated outcomes warrant the use of the FLAG regimen as part of consolidation treatment for longer survival in patients. Finally, the use of FLAG was not limited by cardiotoxicity found in anthracyclines, which could be detrimental to some older patients with concurrent cardiac problems. This would allow more patients to receive the consolidation treatment for longer remission and survival.

3.5 Conclusion

A significantly longer remission and survival were found when the FLAG regimen was used as part of consolidation treatment in older AML patients. By using regression analysis with Cox model, we adjusted the effect of the FLAG for covariate differences. The use of FLAG as consolidation contributed to substantial reduction in hazard rates. Therefore, the regimen should be recommended for consolidation treatment, especially for older patients who are still fit for intensive chemotherapy.

CHAPTER 4: ASSESSING THE PERFORMANCE OF THE COX REGRESSION COEFFICIENT WITH STRONG ASSOCIATION TO THE SURVIVAL OUTCOMES AT LOW EPV CONDITION

4.1 Introduction

Cox proportional hazards model (Cox model) is widely used in cancer research to describe and estimate the association of multiple variables towards patients' survival times through hazard function (George, Seals, & Aban, 2014). The performance of Cox model and its coefficient estimates rely primarily on the independent variables and the outcome events, and are frequently investigated in simulation study (Bender, Augustin, & Blettner, 2005; Burton et al., 2006). The relationship of the outcome events and the independent variables can be expressed as

$$\text{EPV} = \frac{\text{Number of outcome events}}{\text{Number of independent variables}}. \tag{4.1}$$

The value of EPV is determined by the number of outcome events and independent variables, therefore it increases with either increasing the outcome events or decreasing the independent variables. Several simulation studies have shown that a higher EPV value contributes to a more reliable regression analysis with Cox model, and 10 to 20 EPV are generally required for the Cox model (Concato et al., 1995; Peduzzi et al., 1995; Steyerberg, Eijkemans, & Habbema, 1999; Vittinghoff & McCulloh, 2007; Courvoisier et al., 2011; Ogundimu, Altman, & Collins, 2016). Though the EPV has provided a useful indication for assessing the performance of a proportional hazards model and sample size calculation, in fact the value does not reflect anything about the independent variables, other than just the quantity. The use of EPV rule in Cox model necessitated further investigations into other factors that may affect the performance of the Cox model and its estimates, and the EPV associated with those factors.

The EPV rule for Cox model is established based on several large simulation studies. As the performance of Cox model relies primarily on the outcome events and independent variables, not on sample size, simulation studies using large data set may overlook problems encountered in small scale studies. In real situation, not every clinical study is conducted with large sample size. Some single center clinical studies especially those on rare disease conditions and unique cohort of patients, small sample size and outcomes events are almost expected. Therefore, it is not uncommon to encounter studies with Cox models that do not meet the EPV rule (Concato et al., 1995). This problem is especially true in some cancer studies involving rare malignancies such as acute leukemia (Robak & Robak, 2013).

The requirement of EPV somewhat depends on the simulated models and the characteristics of independent variables examined. According to a large factorial simulation study with a primary predictor variable and other variables as adjustment covariates, the authors were able to show that 10 to 20 EPV exceeded the minimum EPV required for certain conditions. In fact, 5 to 9 EPV were sufficient for Cox models with high prevalence variables (Vittinghoff & McCulloh, 2007). For Cox model with many low prevalence variables, more than 20 EPV were generally needed to eliminate bias of the Cox coefficient estimates (Ogundimu, Altman, & Collins, 2016). For Cox models with less than 6 EPV, partial likelihood methods may be too conservative, alternative estimators such as Firth's estimator and Bayesian approaches can be used (Lin, Chang, & Liao, 2013).

As the performance of a Cox model relies on some factors other than EPV, assessing the performance and output of the Cox model based on a definitive EPV value seems not convincing. To date, it is still uncertain as how variable association will affect the performance of the Cox coefficient estimate and what is the EPV for strong variable association. A strong variable association may require less events for the same precision

and accuracy in Cox coefficient estimate as compared to a weak variable association. This further suggests the EPV requirement may change according to the degree of association, and that a variable may respond to a different EPV value than others. Therefore, the rule of 10 to 20 EPV needs to be re-assessed for variable association.

Here, we conduct a Monte Carlo simulation study based on the AML retrospective cohort study reported in Chapter 3. The primary objective of this simulation study is to assess the accuracy, precision and statistical properties of the Cox coefficient estimate of a primary treatment variable with strong variable association to the survival outcomes. In addition, an EPV threshold will be determined for the primary treatment variable. The EPV threshold is the minimum EPV required for stable coefficient estimate. The threshold value can be used to estimate the minimum number of events required or the effective sample size for a large confirmatory study.

4.2 Design of the Simulation Study

4.2.1 Background of the cohort study and data source

Acute myeloid leukemia is a rare and aggressive form of leukemia. The disease is characterized by rapid growth and accumulation of abnormal white cells in bone marrow that interferes with the production of normal blood cells (Robinson & Broadfield, 2005; Deschler & Lübbert, 2006). Treatment for AML includes two phases of intensive chemotherapy, which are induction and consolidation treatments, and hematopoietic stem cell transplantation (Kantarjian et al., 2010; Burnett, Wetzler, & Lowenberg, 2011).

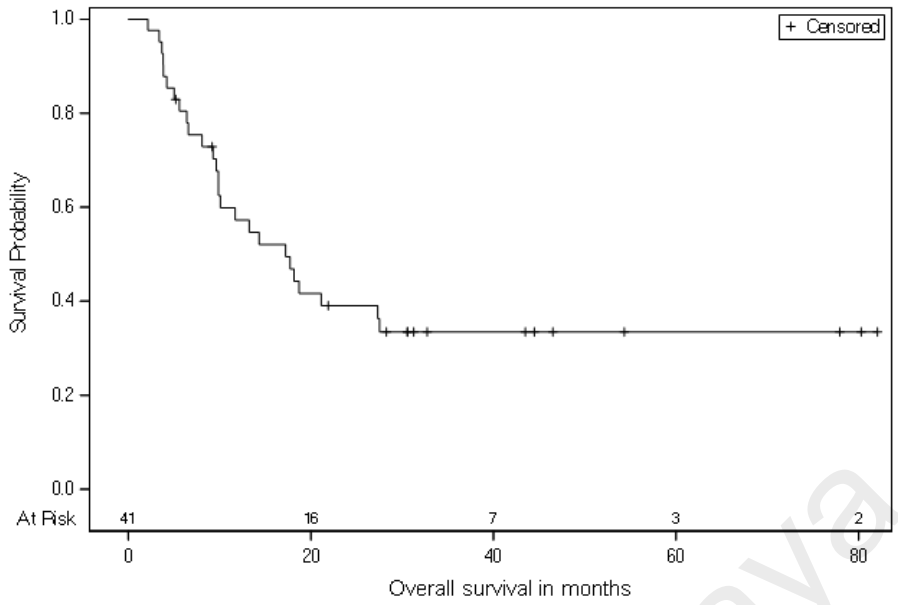

Figure 4.1: The KM survival curve for OS of older AML patients with a median OS duration of 17.23 months (95% CI: 9.86 – 27.51).

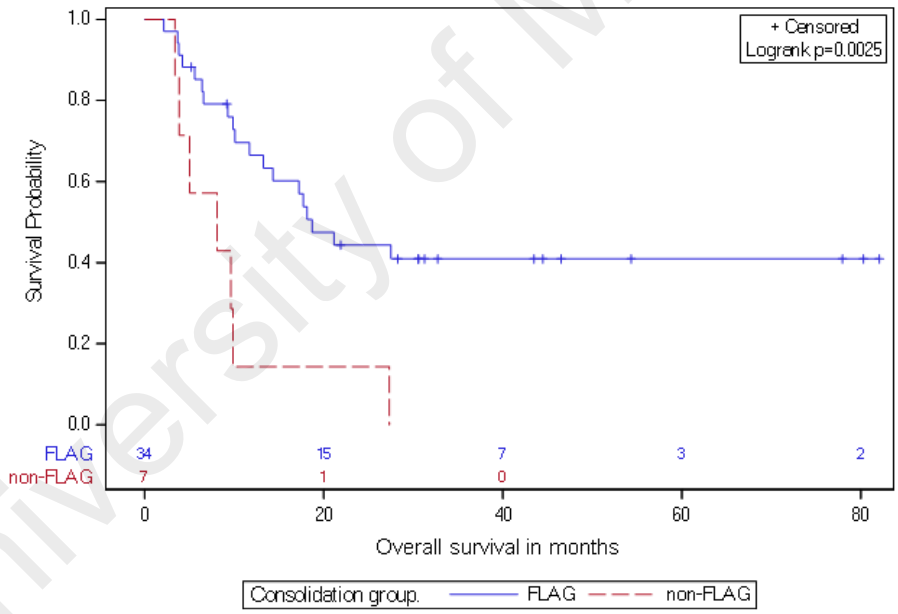

Figure 4.2: The KM survival curves for OS of the older AML patients consolidated with the FLAG regimen and non-FLAG regimens. The FLAG group experienced a longer OS than the non-FLAG group.

The cohort study consisted of 41 older patients who were diagnosed with AML from 2008 to 2013. The aim of the cohort study was to compare the OS and DFS of older patients who had been consolidated with different types of chemotherapy regimen. The

primary treatment variable was binary; it consisted of a group of patients who were consolidated with FLAG regimen (coded as '1') and a group of patients who were consolidated with non-FLAG regimen (coded as '0'). The effect of the primary treatment variable on patient's OS and DFS was adjusted for three other variables (sex, race, and age at diagnosis) in the Cox model.

The EPV was calculated based on Equation (4.1) with 6.5 for OS (26 deaths divided by 4 variables) and 6.75 for DFS (27 relapses/deaths divided by 4 variables). According to the KM OS curve as shown in Figure 4.1, the primary treatment variable was strongly associated with the OS as it caused the OS to differ significantly between those who received the FLAG regimen and those who received non-FLAG regimens (Figure 4.2). We defined a strong variable association to survival outcome as a variable that causes a statistically significant difference, for example at significance level of 0.05, in the survival probabilities at small number of outcome events. The survival curve of the DFS for the treatment variable is similar of OS as presented by Figure 3.3.

The effect of the primary treatment variable was estimated by using a Cox model as expressed by Equation (2.3). Table 4.1 summarizes the results of regression analysis with the Cox model. The primary treatment variable had a prevalence of 0.83 for the FLAG group and 0.17 for the non-FLAG group. Let $\hat{\beta}$ be an estimate of the 'true' parameter of interest, β of the Cox model in the cohort study. The regression coefficient of the primary treatment variable was reported as $\hat{\beta}_{OS} = -1.4073$ ($p = 0.0094$) for OS and $\hat{\beta}_{DFS} = -1.5289$ ($p = 0.0068$) for DFS. Taking the exponentiation of the regression coefficient ($e^{\hat{\beta}}$) yielded a HR of 0.245 for OS and 0.217 for DFS. The results suggested that the use of FLAG regimen as the consolidation chemotherapy was associated with improved OS and DFS in the patients. The assumption of proportional hazards was tested

by using cumulative sums of martingale residuals with supremum test and was satisfied for all variables for the Cox model as shown in Table 3.5.

Table 4.1: The summary of statistics of all variables used in the regression analysis with Cox model for the cohort study.

Variables (coding)	Prevalence	OS				DFS			
		$\hat{\boldsymbol{\beta}}$	SE	*p*	HR	$\hat{\boldsymbol{\beta}}$	SE	*p*	HR
1. Treatment - FLAG (=1) - non-FLAG (=0)	 0.83 0.17	-1.4073	0.542	0.0094	0.245	-1.5289	0.5650	0.0068	0.217
2. Gender - Female (=1) - Male (=0)	 0.59 0.41	-0.3403	0.432	0.4305	0.712	-0.2894	0.4378	0.5088	0.749
3. Race - Chinese (=1) - Indian (=1) - Malay (=0)	 0.36 0.15 0.49	 0.3969 0.1209	 0.397 0.121	 0.3651 0.8419	 1.487 1.128	 0.4268 0.5613	 0.4380 0.5643	 0.3299 0.3199	 1.532 1.753
4. Age	-	0.0293	0.045	0.5137	1.030	0.0637	0.0456	0.1626	1.066

SE = standard error.

4.2.2 Simulation study

The simulation study was performed by using SAS software (version 9.4). Survival times were approximated by a parametric distribution that shares the assumption of proportional hazards with Cox model (Bender, Augustin, & Blettner, 2005). The survival times for OS were first fitted by the Weibull distribution with a scale parameter, λ and a shape parameter, α. A Goodness-of-fit test showed that the survival times for OS fitted well with the Weibull distribution with $\lambda = 2.864 \times 10^{-2}$ and $\alpha = 1.129$ ($p > 0.05$). By using the inverse cumulative hazard function for the Weibull distribution (see Equation (2.11)), the random survival time, T_{os}, of each patient in OS can be simulated by using

$$T_{os} = \left(\lambda^{-1}\left[-\log(U) \times \exp\left(-\mathbf{X}_j\hat{\boldsymbol{\beta}}\right)\right]\right)^{\frac{1}{\alpha}}, \tag{4.2}$$

where the random number, U is uniformly distributed, as U~Uni[0,1], $\widehat{\boldsymbol{\beta}}$ is the set of 'true' regression coefficients, λ is the scale parameter and α is the shape parameter of the Weibull distribution as given above. Survival times for DFS were generated by using the following equation,

$$T_{DFS} = T_{OS} \times \frac{T'_{DFS}}{T'_{OS}}, \tag{4.3}$$

where T'_{DFS} was the original survival time for an individual patient in DFS and T'_{OS} was the original survival time for an individual patient in OS. The unique ratio $\left(\frac{T'_{DFS}}{T'_{OS}}\right)$ between OS and DFS available for each patient was used to generate the survival times in DFS. This approach helped avoid situations whereby an individual patient would have DFS longer than OS. As the DFS was a subset to OS, the inherent relationship between OS and DFS should be maintained throughout the simulation process.

The simulated survival times were then sorted from shortest to longest and combined with a preset outcome status (APPENDIX A) based on the number of outcome events required to achieve the desired EPV value. Events are coded as '1' and non-events are coded as '0' in the preset outcome status dataset. Non-event refers to censored observations either by default censoring or drop-out censoring. For EPV spectrum ranging from 3 to 9, the number of outcome events required were 12, 16, 20, 24, 28, 32 and 36, respectively. The simulated survival times, preset outcome status and variable values of each individual patient were then analyzed by using PROC PHREG[2] in SAS to estimate the simulated regression coefficient, $\widetilde{\boldsymbol{\beta}}$ and the corresponding variance, $\mathbf{s^2}$. The simulation process was repeated for m=1000 times at each EPV value. This allowed the assessment of the performance of the Cox model and its estimates at different EPV values.

[2] PROC PHREG is a SAS command that implements the Cox regression model developed by Sir David Cox in 1972 by using new estimation method called partial likelihood. Partial likelihood estimates still have two of the three standard properties of maximum likelihood: they are consistent and asymptotically normal (Allison, 2010).

In this simulation study, only the performance of the primary treatment variable is investigated in the following section.

4.2.3 Performance of the simulated Cox coefficient estimates

Table 4.2 summarizes the indexes used to assess the accuracy, precision and statistical properties of the simulated Cox coefficient estimates for the primary treatment variable. Accuracy was evaluated by using relative bias and proportion of simulations with percentage absolute relative bias greater than 100%, in accordance to the index used in establishment of EPV rule (Concato et al., 1995; Peduzzi et al., 1995). The absolute relative bias was calculated by taking the value of $(\tilde{\beta} - \hat{\beta})/\hat{\beta}$ in percentage. The absolute relative bias indicates disparity of the simulated regression coefficients ($\tilde{\beta}$) from the 'true' regression coefficient ($\hat{\beta}$). A proportion of simulations with absolute relative bias greater than 100%, which is closer to zero indicates a lower degree of bias. The mean relative bias was calculated by taking the value of $(\bar{\tilde{\beta}} - \hat{\beta})/\hat{\beta}$ in percentage, where $\bar{\tilde{\beta}}$ is the average value of $\tilde{\beta}$ found in m=1000 simulations. A mean relative bias of more than 15% is regarded as problematic (Vittinghoff & McCulloh, 2007).

Table 4.2: The indexes used to assess the accuracy, precision and statistical properties of the simulated Cox coefficient estimates.

Parameters	**Indexes**
A. Accuracy	• Relative bias • Proportion of simulations with absolute relative bias greater than 100%
B. Precision	• Mean model variance (MMV) • Actual sample variance (ASV) • Ratio of MMV:ASV
C. Statistical properties	• 95% CI coverage • Statistical power • Minimum and maximum values of HR

Precision of coefficient estimate was evaluated by using mean model variance (MMV), actual sample variance (ASV) and the ratio of MMV:ASV. The MMV is the average of the corresponding variances for the $\tilde{\beta}$. The ASV is calculated from the simulated regression coefficients by using $\sum_{i=1}^{m}(\tilde{\beta}_i - \bar{\tilde{\beta}})^2/(m-1)$. A departure between the MMV and ASV would cause the ratio of MMV:ASV to deviate from the value of 1. Large sample properties of the Cox model might be lost if the ratio of MMV:ASV was substantially deviated from one (Concato et al., 1995).

Statistical properties of the regression coefficient were checked by using 95% CI coverage, which was the proportion of the simulated models in which 95% CI of a simulated coefficient estimate to include the 'true' regression coefficient. For assessing the power, the proportion of simulated models in which the *z*-statistics exceeded the value of –1.28 for 90% power was calculated. Besides, we also evaluated the minimum and maximum HR for the primary treatment variable at each EPV value.

4.3 Results

Figures 4.3 and 4.4 present the distributions of the simulated Cox regression coefficients for the primary treatment variable at different numbers of EPV in both OS and DFS. All the distributions of the regression coefficients are asymptotically normal. But the distributions became more dispersed when EPV decreased.

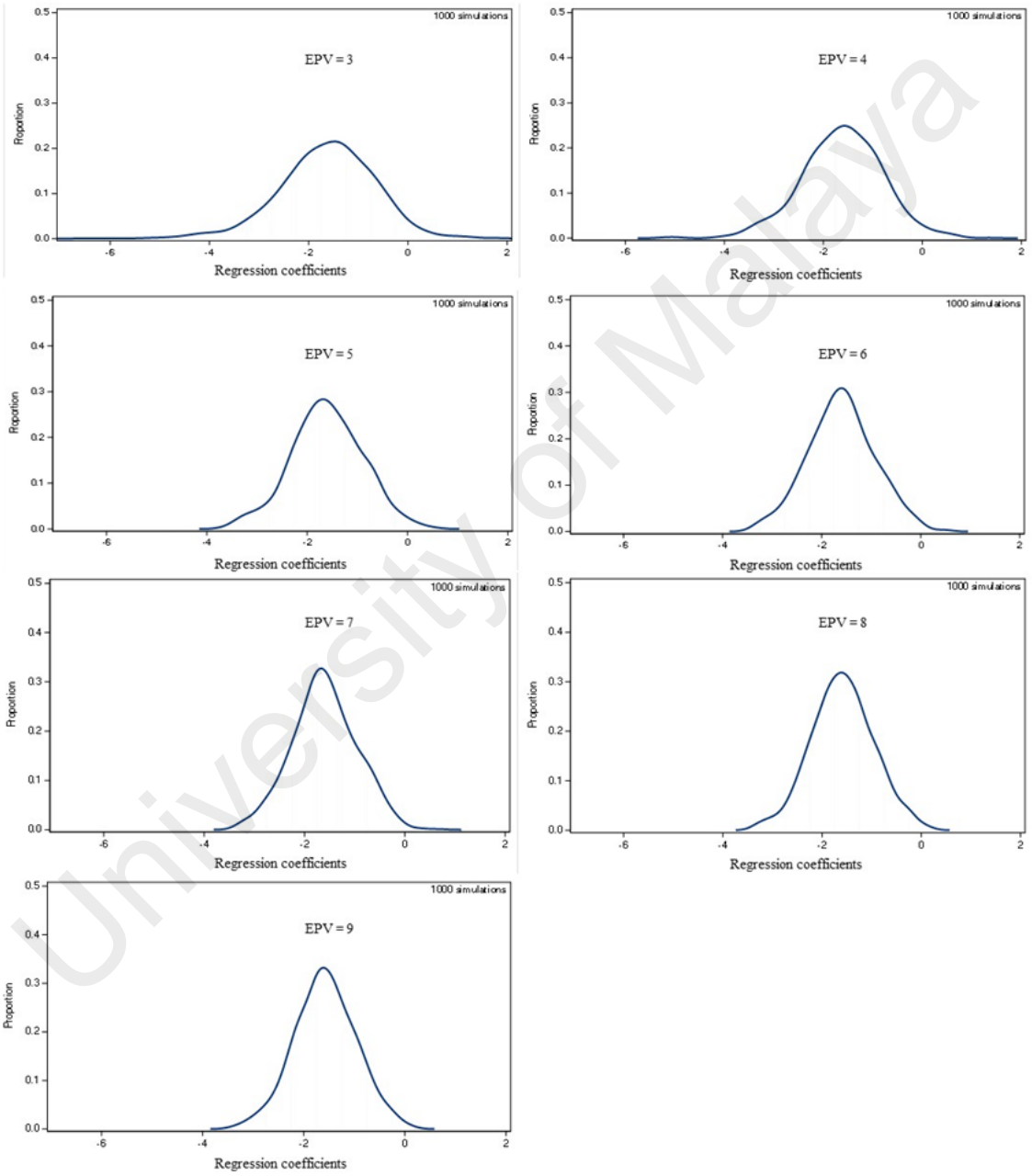

Figure 4.3: The distributions of the simulated Cox regression coefficients for the treatment variable in OS at different numbers of EPV.

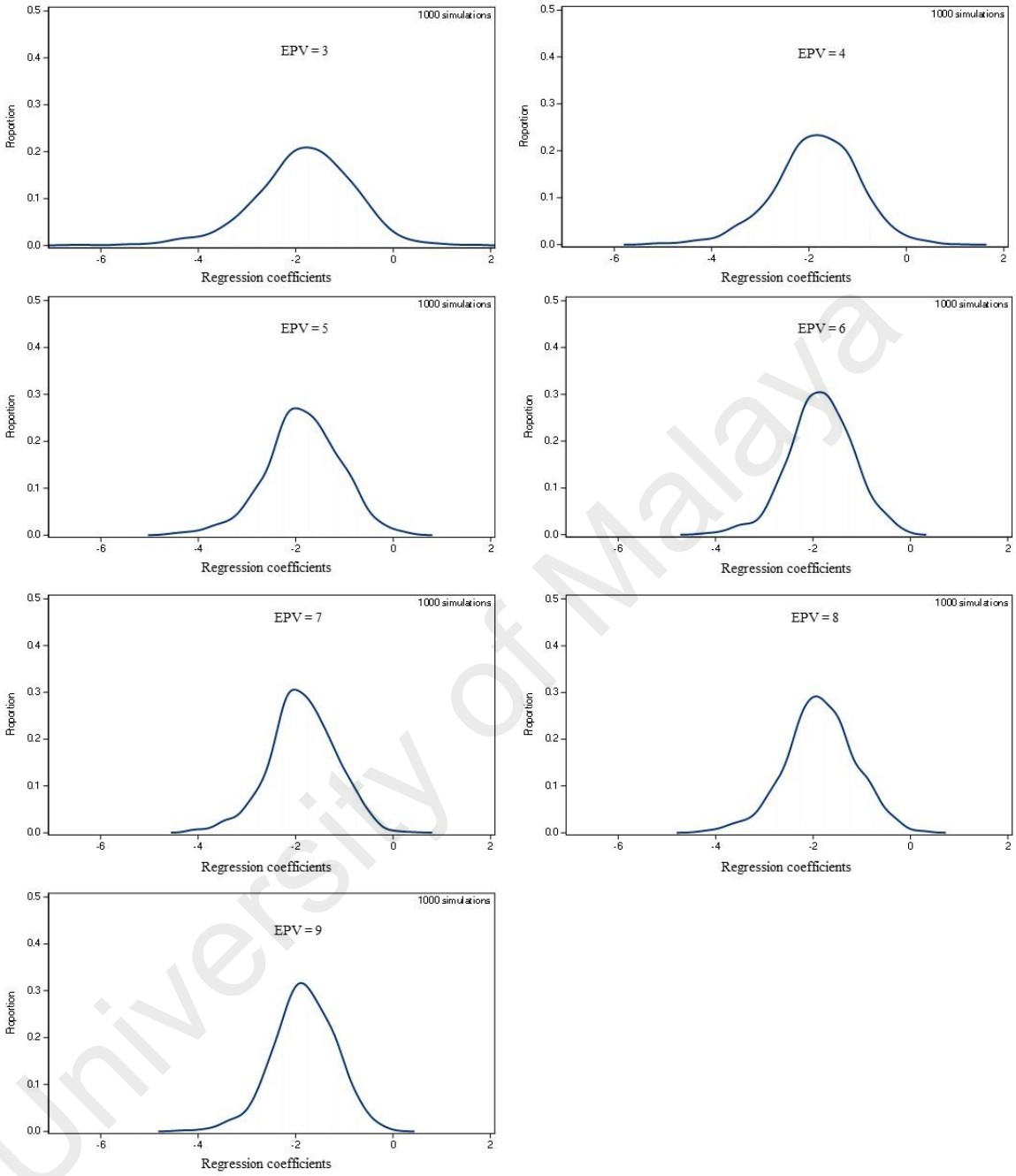

Figure 4.4: The distributions of the simulated Cox regression coefficients for the treatment variable in DFS at different number of EPV.

4.3.1 Accuracy of the simulated regression coefficients

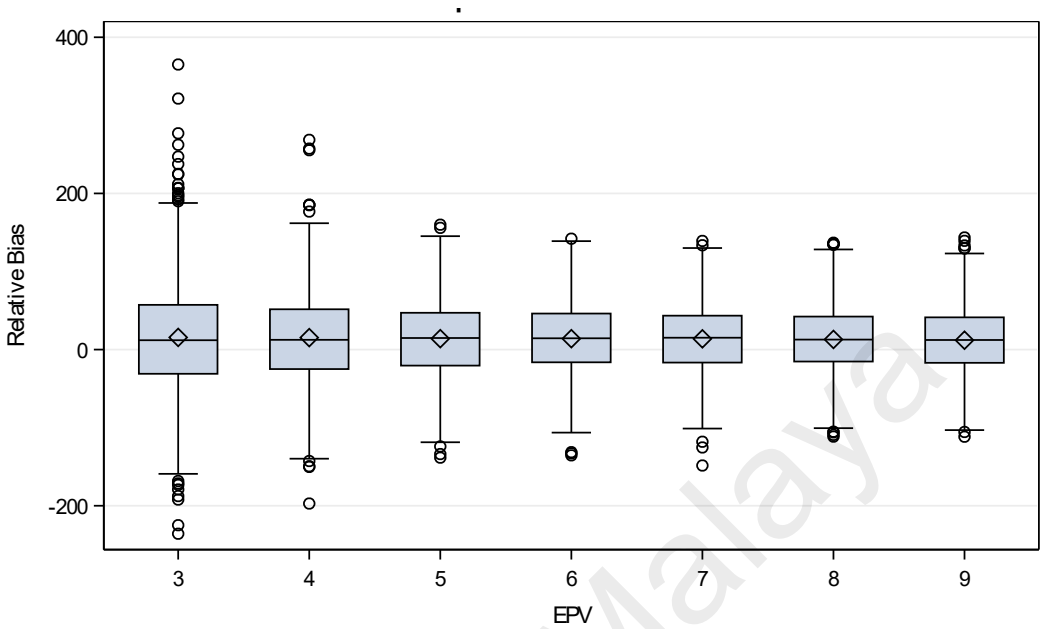

Figure 4.5: The box-plots of the individual relative bias of the simulated regression coefficients for OS.

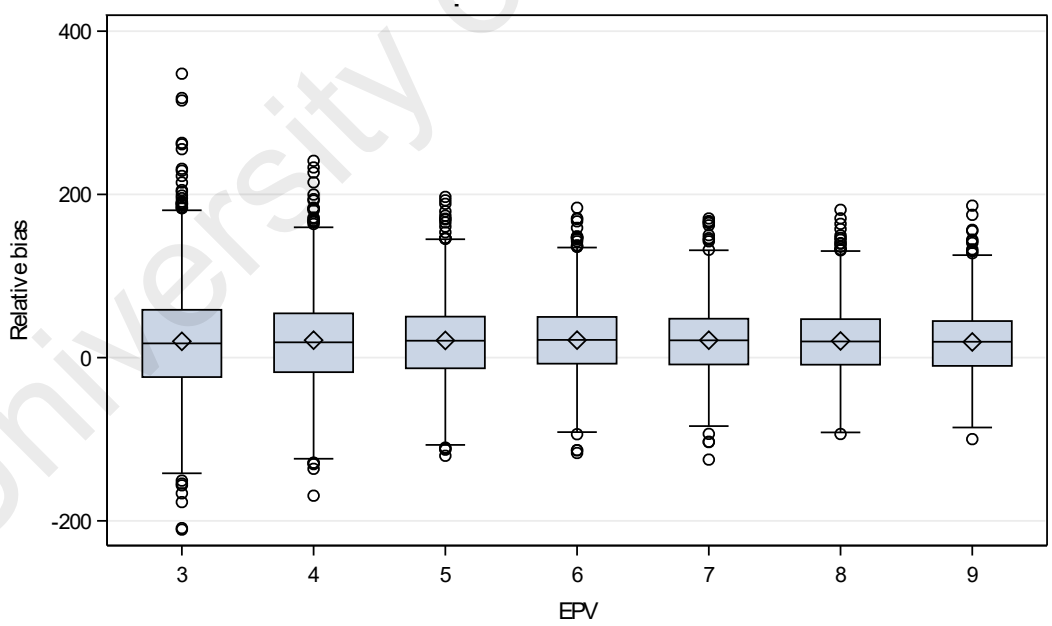

Figure 4.6: The box-plots of the individual relative bias of the simulated regression coefficients for DFS.

Figures 4.5 and 4.6 show the boxplots[3] of the individual relative bias of the simulated regression coefficients for both OS and DFS. According to the boxplots, bias greatly reduced when EPV increased. The bias appeared less varied from 5 to 9 EPV for both OS and DFS. The median and mean of the individual relative bias for both OS and DFS appeared unaffected by EPV change. Both mean and median were consistent and close to each other for all EPV values. Largest interquartile range (IQR) and most outliers were found at 3 EPV. The number of outliers greatly reduced from 5 to 9 EPV. More positive outliers were detected than negative outliers for both OS and DFS. Table 4.3 and Figure 4.7 show the mean relative bias for both OS and DFS across EPV. The mean relative bias was unaffected across EPV spectrum for both OS and DFS due to consistent mean values even at very low EPV. The simulated regression coefficients for OS had a mean relative bias of around 15% at 3 and 4 EPV and less than 15% from 5 to 9 EPV. The mean relative bias for DFS was consistently around 20% across the EPV spectrum.

Table 4.3: The mean relative bias in percentage of the simulated Cox regression coefficients for both OS and DFS.

EPV	OS	DFS
	Percentage (%)	**Percentage (%)**
3	15.57	19.98
4	15.32	21.37
5	13.98	21.08
6	13.96	21.43
7	13.78	21.32
8	12.98	20.42
9	11.98	19.18

[3] Upper and lower edges of box indicate third quartile (75th percentile) and first quartile (25th percentile). The length of the box indicates the interquartile range (IQR). The endpoints of the upper and lower whiskers indicate maximum and minimum data points still within 1.5 IQR from the edges of the box. The line inside the box indicates median and the symbol marker indicates mean.

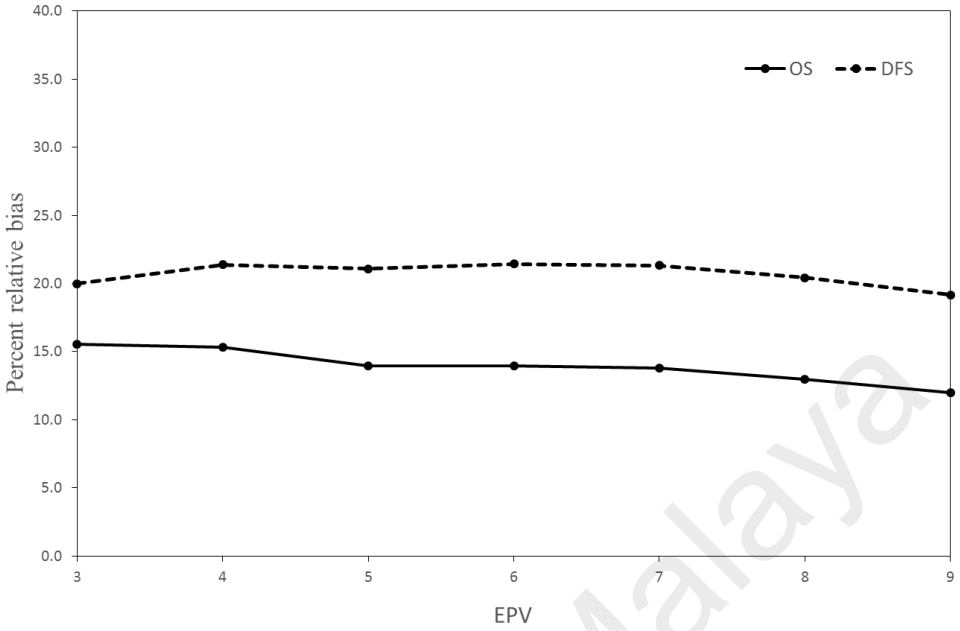

Figure 4.7: The mean relative bias of the simulated regression coefficients for both OS and DFS.

The proportion of simulations in which the absolute relative bias exceeded 100% of the 'true' regression coefficient ($\hat{\beta}$) decreased across EPV spectrum for both OS and DFS (Table 4.4 and Figure 4.8). At 6 EPV and above, less than 5% of the simulated models were found with absolute relative bias exceeded 100% of the 'true' regression coefficient. The result shows that bias increased substantially when the EPV was lower than 6. From 6 to 9 EPV, 50% increase (24 to 36 events) in the number of events only contributed to less than 2% improvement in the relative bias for both OS and DFS. This suggested that a strong variable association helped reduce bias in Cox coefficient estimate. Hence, less EPV are required for variable with strong variable association to the survival outcome.

Table 4.4: The proportion of the simulated Cox models with absolute relative bias exceeded 100% of the 'true' regression coefficients for both OS and DFS.

EPV	OS		DFS	
	Total	**Proportion**	**Total**	**Proportion**
3	134	0.134	124	0.124
4	98	0.098	102	0.102
5	64	0.064	64	0.064
6	47	0.047	47	0.047
7	34	0.034	44	0.044
8	33	0.033	33	0.033
9	29	0.029	36	0.036

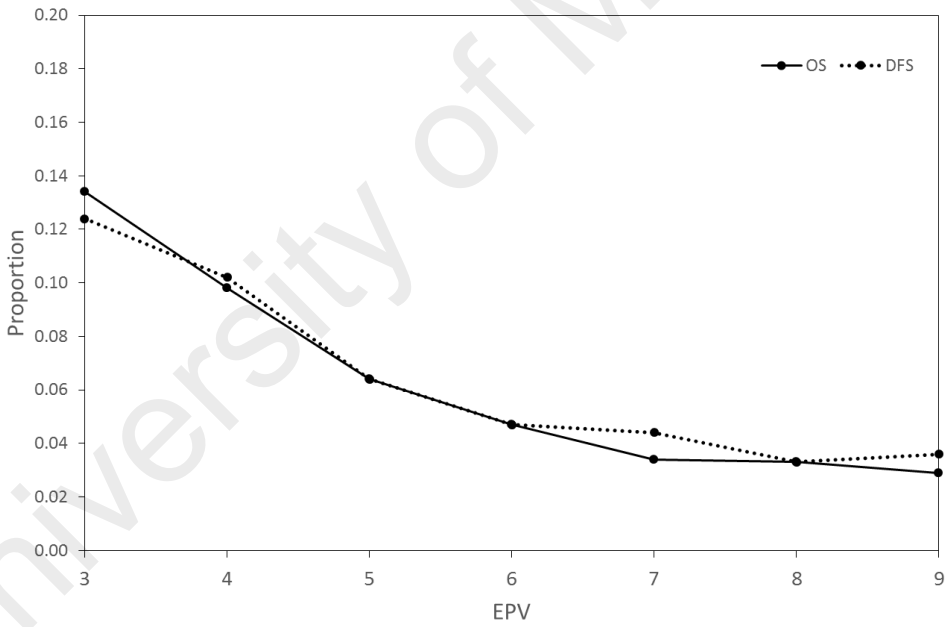

Figure 4.8: The proportion of the simulated Cox models with absolute relative bias exceeded 100% of the 'true' regression coefficients for both OS and DFS.

4.3.2 Precision of the simulated regression coefficients

Table 4.5 summarizes the MMV, ASV and the ratio of MMV:ASV for both OS and DFS. The MMV and ASV were relatively high at 3 and 4 EPV. Both MMV and ASV decreased with increasing EPV.

Table 4.5: The MMV, ASV and ratio of MMV:ASV of the simulated Cox regression coefficients for both OS and DFS.

EPV	OS			DFS		
	MMV	**ASV**	$\frac{\textbf{MMV}}{\textbf{ASV}}$	**MMV**	**ASV**	$\frac{\textbf{MMV}}{\textbf{ASV}}$
3	0.7920	0.9640	0.8216	0.8390	1.0680	0.7856
4	0.5860	0.6720	0.8720	0.6250	0.7540	0.8289
5	0.4780	0.5050	0.9465	0.5110	0.5760	0.8872
6	0.4210	0.4540	0.9273	0.4500	0.5080	0.8858
7	0.3860	0.4110	0.9392	0.4120	0.4590	0.8976
8	0.3640	0.3910	0.9309	0.3890	0.4320	0.9005
9	0.3500	0.3720	0.9409	0.3730	0.4070	0.9165

Departure between the MMV and ASV was found at 3 and 4 EPV for both OS (Figure 4.9) and DFS (Figure 4.10). The departure between the MMV and ASV was substantial at 3 and 4 EPV, but appeared consistently small from 5 to 9 EPV. Figure 4.11 presents the change of the ratio of MMV:ASV across the EPV spectrum. According to Table 4.5, the MMV of the simulated regression coefficients was underestimated by 18% at 3 EPV and 13% at 4 EPV for OS and by 21% at 3 EPV and 17% at 4 EPV for DFS. The change of the ratio of MMV:ASV over the EPV spectrum showed that Cox coefficient estimates had a considerably small departure between variances from 5 to 9 EPV. This implied that the precision of the Cox coefficient estimate would only be affected when the EPV was less than 5. Therefore, a strong variable association improved precision of Cox coefficient estimate at low EPV.

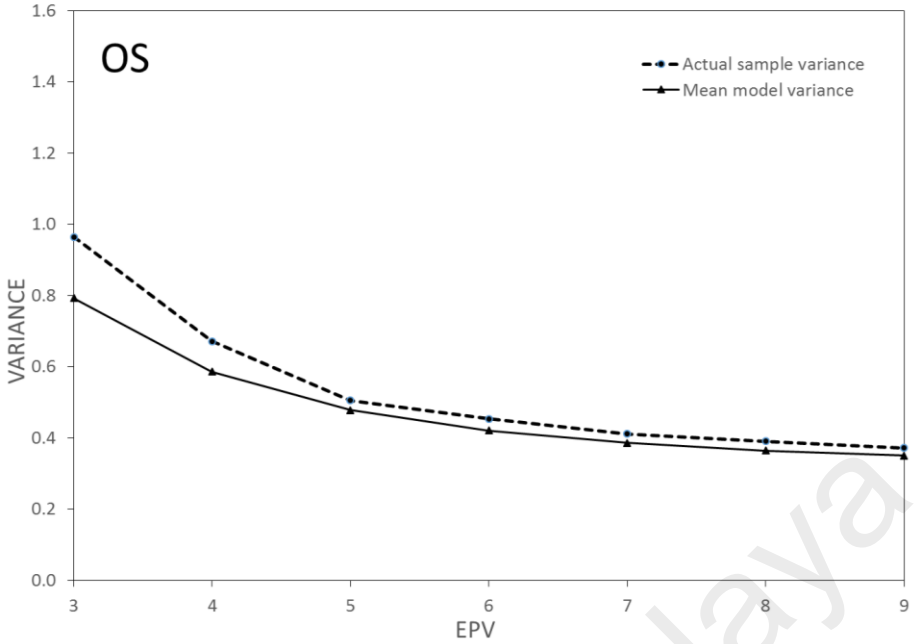

Figure 4.9: The MMV and ASV of the simulated regression coefficients of the treatment variable in OS.

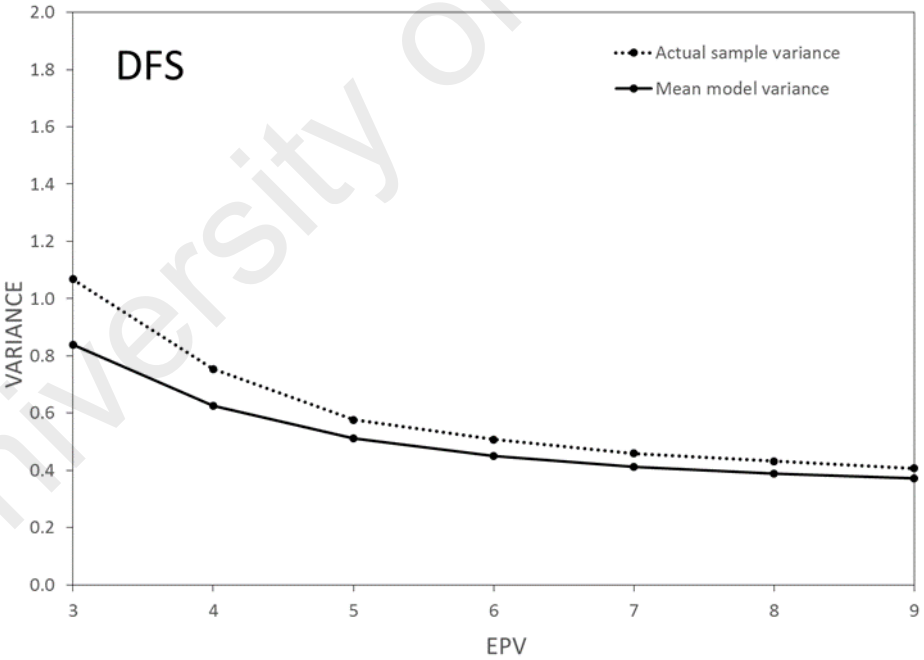

Figure 4.10: The MMV and ASV of the simulated regression coefficients of the treatment variable for DFS.

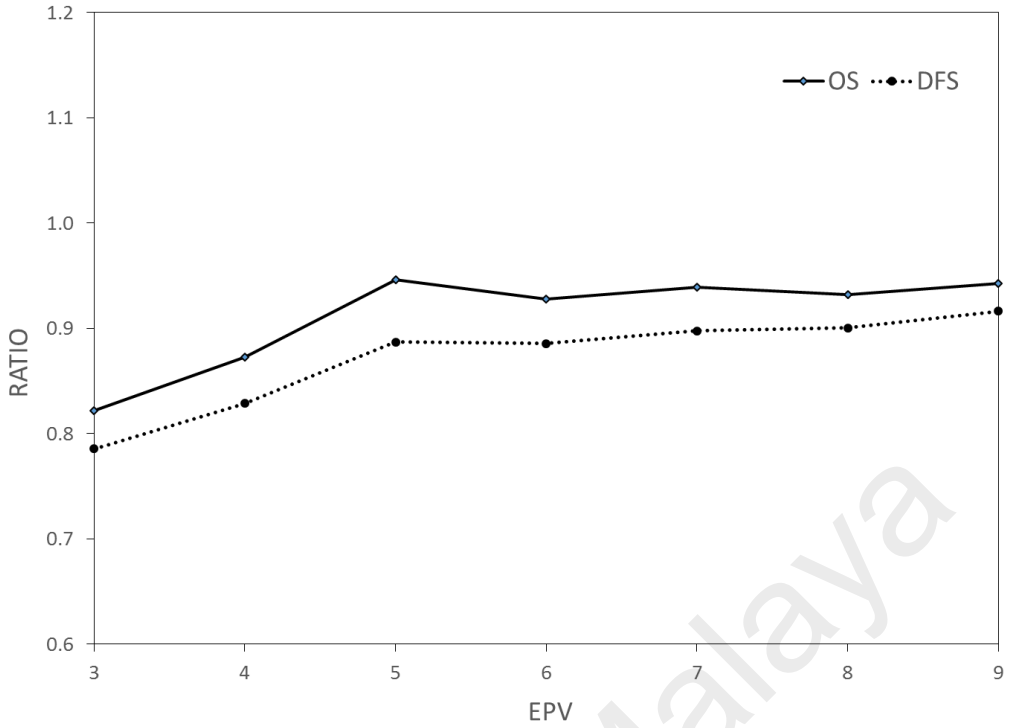

Figure 4.11: The ratio of MMV:ASV for the simulated regression coefficients of the treatment variable for both OS and DFS.

4.3.3 Statistical significance of the simulated regression coefficients

Statistical properties of the simulated regression coefficients were assessed by using three indexes. Table 4.6 shows the maximum and minimum HR of the treatment variable for both OS and DFS. The change of the maximum and minimum HR over the EPV spectrum is shown in Figure 4.12. The range of HR was widest at 3 and 4 EPV, and appeared less varied from 5 to 9 EPV. At 3 and 4 EPV, the protective effect of the primary treatment variable tended to be wrongly estimated. The 95% CI coverage also appeared stable and unaffected by EPV (Table 4.7 and Figure 4.13). The results suggest that a strong variable association may provide more stable CI coverage for the Cox regression coefficient.

Table 4.6: The maximum and minimum simulated HR at 95% level (two-tailed) of the treatment variable for both OS and DFS.

EPV	OS	DFS
3	6.763, 0.001	5.434, 0.001
4	3.926, 0.006	2.880, 0.005
5	1.717, 0.016	1.363, 0.011
6	1.647, 0.023	1.294, 0.013
7	1.429, 0.030	1.464, 0.016
8	1.075, 0.034	0.907, 0.014
9	0.963, 0.034	0.998, 0.013

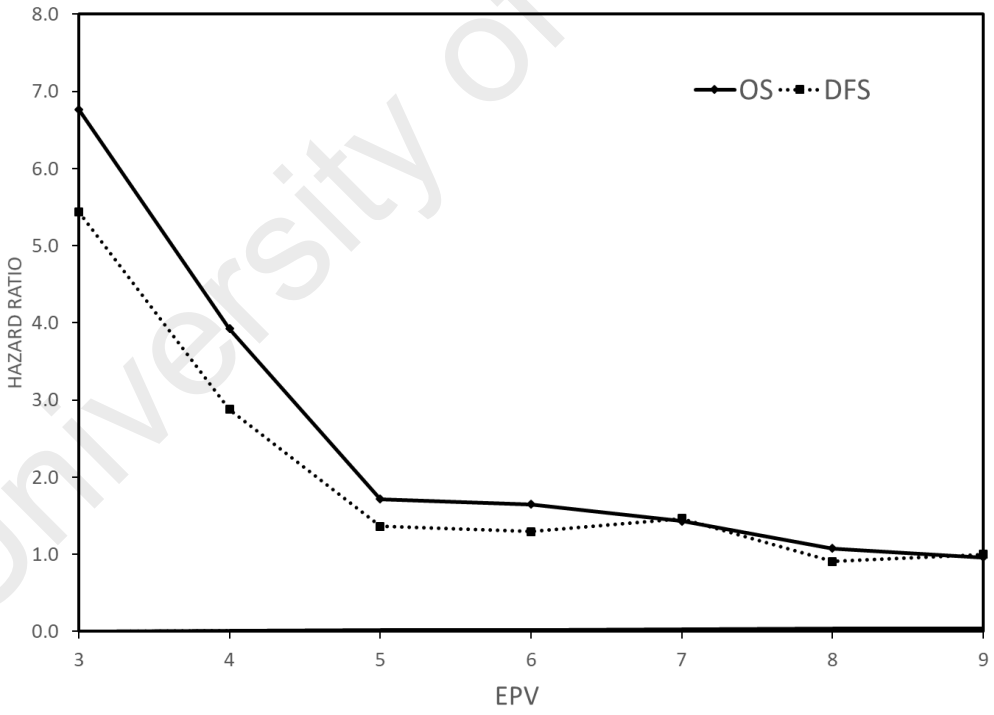

Figure 4.12: The maximum and minimum simulated HR at 95% level (two-tailed) of the treatment variable for both OS and DFS.

Table 4.7: The proportion of the simulated Cox models in which 95% CI of the simulated Cox regression coefficients included the 'true' regression coefficient for both OS and DFS.

EPV	OS	DFS
3	0.945	0.952
4	0.940	0.933
5	0.941	0.932
6	0.944	0.931
7	0.948	0.935
8	0.943	0.937
9	0.946	0.943

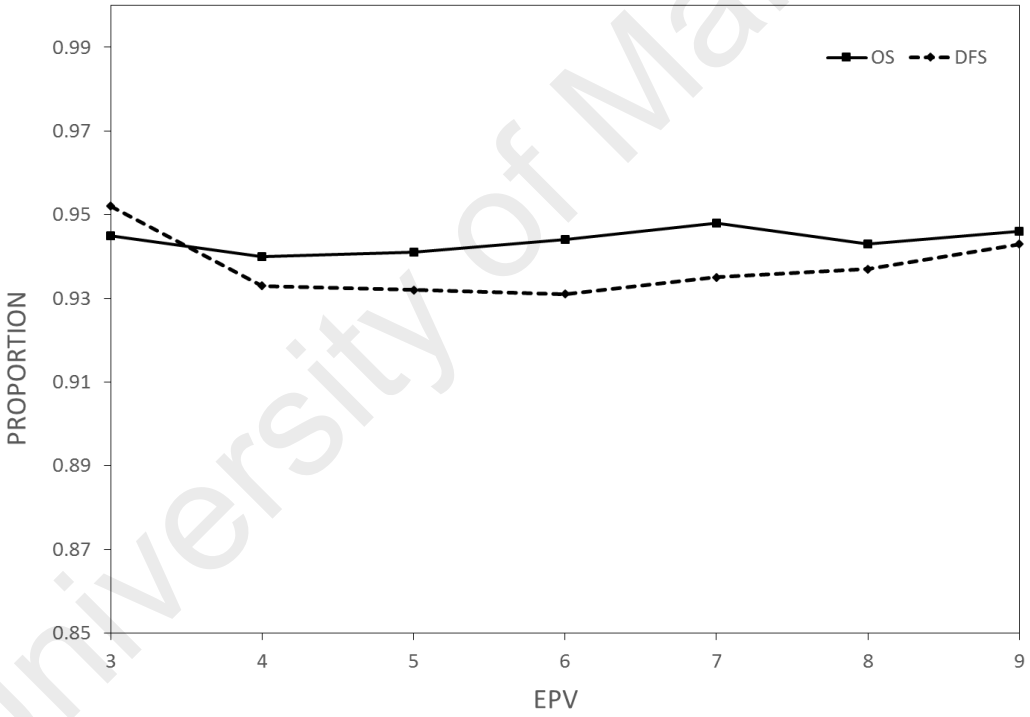

Figure 4.13: The proportion of the simulated Cox models in which 95% CI of the simulated Cox regression coefficients included the 'true' regression coefficient.

Table 4.8 and Figure 4.14 show the proportion of simulated models in which the *z*-statistics exceeded the value of –1.28 (a counterpart of 90% power). According to the figure, the power increased with increasing EPV. Ninety percent power was achieved at 8 EPV for OS and 5 EPV for DFS.

Table 4.8: The proportion of simulated Cox models with *z*-statistics more than –1.28 for 90% power for both OS and DFS.

EPV	OS	DFS
3	0.732	0.786
4	0.820	0.873
5	0.853	0.909
6	0.885	0.933
7	0.893	0.951
8	0.912	0.969
9	0.920	0.967

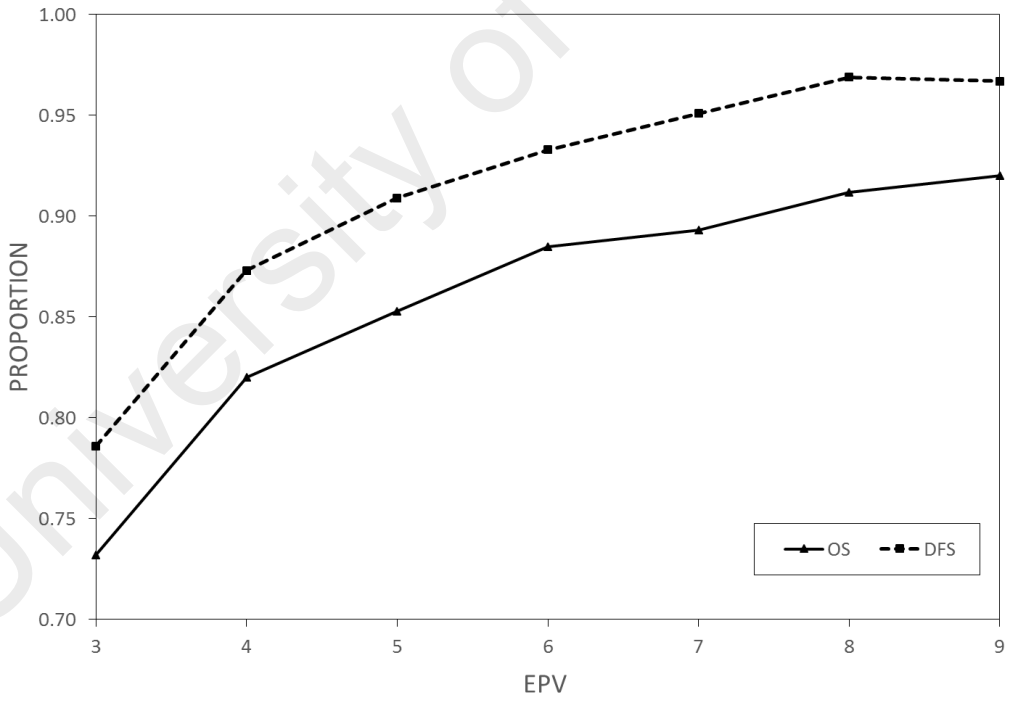

Figure 4.14: The proportion of simulated Cox models with *z*-statistics more than –1.28 for 90% power for both OS and DFS.

4.4 Discussion

Other simulation studies investigate the performance of Cox estimators and logistic estimators with varying EPV under conditions such as different correlation between variables, strength of estimator, number of variables, type of variables, and prevalence of binary variables (Peduzzi et al., 1996; Vittinghoff & McCulloh, 2007; Courvoisier et al, 2011; Ogundimu, Altman, & Collins, 2016). Even with very large dataset, conclusions from those simulation studies still vary. Therefore, it is arguable to have one definitive EPV value that applies to all conditions.

Our simulation study utilizes a small dataset as the validity of Cox model depends primarily on the outcome events and independent variables, not the sample size. Our study has shown that a variable with strong association to survival outcomes requires less EPV for reliable estimates, making it possible with less than 10 EPV. Our results show that the regression coefficient of the primary treatment variable had sufficient power, acceptable accuracy and precision at 6 EPV. Besides, our results are also consistent with other simulation studies that bias increases and statistical power decreases at very low EPV.

Interestingly, an EPV threshold for the primary treatment variable was found at 6 EPV. Above this value, variation and deviation appeared consistent and unaffected with increasing EPV. The EPV threshold may be variable-specific. It may also change according to the strength of association. However, further research is required to evaluate those assumptions. More importantly, the EPV threshold may be used to estimate the number of outcome events required to determine the sample size for a large confirmatory study.

It is possible that the relative bias could have been affected during the simulation process. An ideal approximation should bring all the simulated survival times close to the original survival times and when a comparison was made, zero relative bias should be

obtained at the EPV threshold. However, that was not achieved at the end. This indicates that the results may be affected during the simulation of survival times. The Weibull distribution has provided a good fit to the survival times, but is still far from a perfect fit. Nevertheless, the trend of how the indexes change over the EPV spectrum has shown that the regression analysis with Cox model should have provided a good estimation to the 'true' effect for the primary treatment variable at the given EPV value.

4.4.1 Implications of EPV in sample size estimation

A valid regression analysis with Cox model depends on sufficient number of events. Then, a sample size can be accurately estimated by knowing the number of events or EPV required. To do this, an investigator will need to consider the event prevalence (P), number of independent variables (k) and the EPV value. For example, let us assume that the three years prevalence of death in breast cancer was 60%. An investigator would like to assess the HR associated with a new chemotherapy in a group of terminal breast cancer patients, compared to the conventional treatment. Let us assume that the estimation of HR was done with adjustment of 5 other independent variables. Then, the sample size for the study $= \frac{\mathrm{k} \times \mathrm{EPV}}{\mathrm{P}} = \frac{6 \times \mathrm{EPV}}{0.60} = 10 \times \mathrm{EPV}$ patients over three years. As a rule of thumb, the sample size can be estimated by using 10 to 20 EPV, which gives the investigator an estimated 100 to 200 patients. However, if the new chemotherapy was effective and strongly associated to the survival outcome, 5 to 6 EPV may be sufficient, and the number of subjects could be reduced to 50 to 60 patients.

4.4.2 Strength of association

The requirement of 10 to 20 EPV was established by using variables with moderate association to the survival outcome (Peduzzi et al., 1995), therefore does not necessarily apply for variable with strong association to the survival outcome. Our results show that strong variable association requires less than 10 EPV. An investigator can

gauge the degree of association by studying KM survival curve. A variable with strong association should cause a significant difference in the KM survival curves with small number of events. To date, method for calculating an effective EPV for a particular variable is not available. Nevertheless, a simulation study can be conducted to identify an effective EPV threshold with helps from a statistician, provided the survival times can be approximated by a parametric distribution.

4.5 Conclusion

The performance of the Cox model not only relies on the EPV. The requirement of EPV may vary according to different conditions and characteristics of the independent variables, therefore it is arbitrary to have a definitive EPV value for all variables. Our simulation has shown that 3 and 4 EPV were associated with highest level of bias and disparity in accuracy, precision and statistical properties. At 5 to 6 EPV, the performance of the Cox coefficient estimate started to gain stability. Above 6 EPV, increasing the number of events was less likely to improve the overall performance of the Cox coefficient estimate. We were able to demonstrate that the EPV rule is uncertain and should not apply to a variable with strong association to the survival outcomes. An EPV threshold for the primary treatment variable was identified at 6, which was much lower than the EPV rule published in other literatures.

Our finding is useful for research of rare disease conditions with limited outcome events and variables of high association to survival outcome. This identified EPV threshold helps improve the estimation of sample size that can be applied for a large confirmatory study in the same scenario, as well as the interpretation of regression output by using the Cox model.

CHAPTER 5: CONCLUDING REMARKS

The retrospective cohort study is small considering the number of patients enrolled and amount of data analyzed. However, a good study should not be only judged by its size, other parts of the study such as novelty of research question, study design, analytical methodology, variables and outcomes measured are equally important.

The retrospective cohort study has successfully brought out an important treatment concept of intensifying the post induction treatment particularly for those with advanced age with the FLAG regimen. Not only that, further analysis has been done to differentiate the survival outcomes for those who had received the FLAG only regimen and those who had combination of regimens. Through a series of methods from KM to regression analysis with Cox model, we are able to analyze the complete survival and disease free experience of a group of older AML patients treated by local hematologists in a meaningful manner for the country. A clear difference in OS and DFS amongst those who took FLAG only, combination and non-FLAG chemotherapy evidences the need to intensify the post induction chemotherapy for those who are still clinically fit. The findings of the cohort study are also valuable for a large confirmatory study involving the FLAG regimen as consolidation later.

The simulation study, as a statistical tool to assess the adequacy of the Cox models in the cohort study presents another important concept of low EPV requirement for variable with high association to the survival outcome. As a popular survival analysis technique in clinical study, it is prudent to understand how a coefficient estimate would behave under low EPV condition. The findings of the simulation study have a practical contribution towards study planning, particularly in the estimation of sample size by using EPV. However, more work should be carried out to investigate how the degree of association between a variable and a survival outcome could affect the EPV requirement.

REFERENCES

Allison, P. D. (2010). *Survival Analysis Using SAS: A Practical Guide*, Second Edition. Cary, NC: SAS Institute Inc.

Alwan, A. F., Matti, B. F., Naji, A. S., & Jawad, A. M. (2014). The efficacy of fludarabine, high dose cytosine arabinoside with granulocyte colony stimulating factor (FLAG) protocol as salvage therapy for refractory/relapsed acute leukemias in adult Iraqi patients. *Indian Journal of Hematology and Blood Transfusion*, *30*(4), 231-235.

Appelbaum, F. R., Gundacker, H., Head, D. R., Slovak, M. L., Willman, C. L., Godwin, J. E., …, Petersdorf, S. H. (2006). Age and acute myeloid leukemia. *Blood*, *107*(9), 3481-3485.

Bender, R., Augustin, T., & Blettner, M. (2005). Generating survival times to simulate Cox proportional hazards models. *Statistics in Medicine*, *24*(11), 1713 - 1723.

Bethesda. (2015). *SEER Cancer Statistics Facts: Acute Myeloid Leukemia (AML). National Cancer Institute*. Retrieved from https://seer.cancer.gov/statfacts/html/amyl.html

Bewick, V., Cheek, L., & Ball, J. (2004). Statistics review 12: Survival analysis. *Critical Care*, *8*(5), 389-394.

Büchner, T., Urbanitz, D., Hiddemann, W., Rühl, H., Ludwig, W. D., Fischer, J., . . . Zeile, G. (1985). Intensified induction and consolidation with or without maintenance chemotherapy for acute myeloid leukemia (AML): Two multicenter studies of the German AML Cooperative Group. *Journal of Clinical Oncology*, *3*(12), 1583-1589.

Burnett, A., Wetzler, M., & Löwenberg, B. (2011). Therapeutic advances in acute myeloid leukemia. *Journal of Clinical Oncology*, *29*(5), 487-494.

Burton, A., Altman, D. G., Royston, P., & Holder, R. L. (2006). The design of simulation studies in medical statistics. *Statistics in Medicine*, *25*(24), 4279-4292.

Carella, A. M., Cascavilla, N., Greco, M. M., Melillo, L., Sajeva, M. R., Ladogana, S., . . . Carotenuto, M. (2001). Treatment of "poor risk" acute myeloid leukemia with fludarabine, cytarabine and G-CSF (flag regimen): A single center study. *Leukemia & Lymphoma*, *40*(3-4), 295-303.

Cheng, H. W. B., Li, C. W., Chan, K. Y., Au, H. Y., Chan, P. F., Sin, Y.-C., . . . Sham, M. K. (2015). End-of-life characteristics and palliative care provision for elderly patients suffering from acute myeloid leukemia. *Support Care Cancer*, *23*(1), 111-116.

Clark, T. G., Bradburn, M. J., Love, S. B., & Altman, D. G. (2003a). Survival analysis part I: Basic concepts and first analyses. *British Journal of Cancer*, *89*(2), 232-238.

Clark, T. G., Bradburn, M. J., Love, S. B., & Altman, D. G. (2003b). Survival analysis part IV: Further concepts and methods in survival analysis. *British Journal of Cancer*, *89*(5), 781-786.

Clavio, M., Carrara, P., Miglino, M., Pierri, I., Canepa, L., Balleari, E., . . . Gobbi, M. (1996). High efficacy of fludarabine-containing therapy (FLAG-FLANG) in poor risk acute myeloid leukemia. *Haematologica*, *81*(6), 513-520.

Concato, J., Peduzzi, P., Holford, T. R., & Feinstein, A. R. (1995). Importance of events per independent variable in proportional hazards analysis I. background, goals, and general strategy. *Journal of Clinical Epidemiology*, *48*(12), 1495-1501.

Courvoisier, D. S., Combescure, C., Agoritsas, T., Gayet-Ageron, A., & Perneger, T. V. (2011). Performance of logistic regression modelling: Beyond the number of events per variable, the role of data structure. *Journal of Clinical Epidemiology*, *64*(9), 993-1000.

Cox, D. R. (1972). Regression models and life-tables. *Journal of Royal Statistical Society, Series B (Methodological)*, *34*(2), 187-220.

de la Rubia, J., Regadera, A., Martin, G., Cervera, J., Sanz, G., Martinez, J., . . . Sanz, M. (2002). FLAG-IDA regimen (fludarabine, cytarabine, idarubicin and G-CSF) in the treatment of patients with high-risk myeloid malignancies. *Leukemia Research*, *26*(8), 725-730.

Delgado, J., Pereira, A., Villamor, N., López-Guillermo, A., & Rozman, C. (2014). Survival analysis in hematologic malignancies: Recommendations for clinicians. *Haematologica*, *99*(9), 1410-1420.

Deschler, B., & Lübbert, M. (2006). Acute myeloid leukemia: Epidemiology and etiology. *Cancer*, *107*(9), 2099-2107.

Estey, E. (2007). Acute myeloid leukemia and myelodysplastic syndromes in older patients. *Journal of Clinical Oncology*, *25*(14), 1908 - 1915.

Estey, E. H., Thall, P. F., Pierce, S., Cortes, J., Beran, M., Kantarjian, H., . . . Freireich, E. (1999). Randomized phase II study of fludarabine + cytosine arabinoside + idarubicin ± all-trans retinoic acid ± granulocyte colony-stimulating factor in poor prognosis newly diagnosed acute myeloid leukemia and myelodysplastic syndrome. *Blood*, *93*(8), 2478-2484.

Ferrara, F., D'Arco, A. M., De Simone, M., Mele, G., Califano, C., Pocali, B., . . . Palmieri, S. (2005). Fludarabine and cytarabine as continuous sequential infusion for elderly patients with acute myeloid leukemia. *Haematologica*, *90*(6), 776-784.

Fey, M. F., & Buske, C. (2013). Acute myeloblastic leukaemias in adult patients: ESMO clinical practice guidelines for diagnosis, treatment and follow-up. *Annals of Oncology*, *24*(Suppl_6), 138-143.

Ford, I., & Norrie, J. (2002). The role of covariates in estimating treatment effects and risk in long-term clinical trials. *Statistics in Medicine*, *21*(19), 2899-2908.

Gandhi, V., Estey, E., Du, M., Nowak, B., Keating, M. J., & Plunkett, W. (1995). Modulation of the cellular metabolism of cytarabine and fludarabine by granulocyte-colony-stimulating factor during therapy of acute myelogenous leukemia. *Clinical Cancer Research*, *1*(2), 169-178.

Gandhi, V., Estey, E., Keating, M. J., & Plunkett, W. (1993). Fludarabine potentiates metabolism of cytarabine in patients with acute myelogenous leukemia during therapy. *Journal of Clinical Oncology*, *11*(1), 116-124.

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

Greenland, S. (1989). Modeling and variable selection in epidemiologic analysis. *American Journal of Public Health*, *79*(3), 340-349.

Heiblig, M., Elhamri, M., Tigaud, I., Plesa, A., Barraco, F., Labussière, H., . . . Thomas, X. (2016). Treatment wiht low-dose cytarabine in elderly patients (age 70 years or older) with acute myeloid leukemia: A single institution experience. *Mediterranean Journal of Hematology and Infectious Diseases*, *8*(1), e2016009.

Jefferson, E. (2010). *FDA: Pfizer voluntarily withdraws cancer treatment Mylotarg from U.S. market [Press release]*. Retrieved from https://wayback.archive-it.org/7993/20170114034251/http://www.fda.gov/NewsEvents/Newsroom/PressAnnouncements/ucm216448.htm

Juliusson, G., Antunovic, P., Derolf, A., Lehmann, S., Möllgård, L., Stockelberg, D., . . . Höglund, M. (2009). Age and acute myeloid leukemia: Real world data on decision to treat and outcomes from Swedish Acute Leukemia Registry. *Blood*, *113*(18), 4179-4187.

Kantarjian, H., Beran, M., Cortes, J., O'Brien, S., Giles, F., Pierce, S., . . . Estey, E. (2006). Long-term follow-up results of the combination of topotecan and cytarabine and other intensive chemotherapy regimens in myelodysplastic syndrome. *Cancer*, *106*(5), 1099-1109.

Kantarjian, H., O'Brien, S., Cortes, J., Wierda, W., Faderl, S., Garcia-Manero, G., . . . Freireich, E. J. (2008). Therapeutic advances in leukemia and myelodysplastic syndrome over the past 40 years. *Cancer*, *113*(S7), 1933-1952.

Kantarjian, H., Ravandi, F., O'Brien, S., Cortes, J., Faderi, S., Garcia-Manero, G., . . . Freireich, E. J. (2010). Intensive chemotherapy does not benefit most older patients (age 70 years or older) with acute myeloid leukemia. *Blood*, *116*(22), 4422-4429.

Kaplan, E. L., & Meier, P. (1958). Nonparametric estimation from incomplete observations. *Journal of the American Statistical Association*, *53*(282), 457-481.

Krug, U., Büchner, T., Beldel, W. E., & Müller-Tidow, C. (2011). The treatment of elderly patients with acute myeloid leukemia. *Deutsches Ärzteblatt International*, *108*(51-52), 863-870.

Lee, S. R., Yang, D. H., Ahn, J. S., Kim, Y. K., Lee, J. J., Choi, Y. J., . . . Kim, H. J. (2009). The clinical outcome of FLAG chemotherapy without idarubicin in patients with relapsed or refractory acute myeloid leukemia. *Journal of Korean Medical Science, 24*(3), 498-503.

Lin, D. Y., Wei, L. J., & Ying, Z. (1993). Checking the Cox model with cumulative sums of martingale-based residuals. *Biometrika, 80*(3), 557-572.

Lin, I. F., Chang, W. P., & Liao, Y. N. (2013). Shrinkage methods enhanced the accuracy of parameter estimation using Cox models with small number of events. *Journal of Clinical Epidemiology, 66*(7), 743-751.

Luo, S., Cai, F., Jiang, L., Zhang, S., Shen, Z., Sun, L., & Gao, S. (2013). Clinical study of Mito-FLAG regimen in treatment of relapsed acute myeloid leukemia. *Experimental and Therapeutic Medicine, 5*(3), 982-986.

Medeiros, B. C., Satram-Hoang, S., Hurst, D., Hoang, K. Q., Momin, F., & Reyes, C. (2015). Big data analysis of treatment patterns and outcomes among elderly acute myeloid leukemia patients in the United States. *Annals of Hematology, 94*(7), 1127 - 1138.

NCCN Clinical Guidelines in Oncology, Acute Myeloid Leukemia. (version 2.2014). National Comprehensive Cancer Network.

Nelson, W. (1982). *Applied life data analysis*. New York: John Wiley and Sons, Inc.

Ogundimu, E. O., Altman, D. G., & Collins, G. S. (2016). Adequate sample size for developing prediction models is not simply related to events per variable. *Journal of Clinical Epidemiology, 76*, 175 - 182.

Oran, B., & Weisdorf, D. J. (2012). Survival for older patients with acute myeloid leukemia: A population-based study. *Haematologica, 97*(12), 1916-1924.

Ossenkoppele, G. J., Graveland, W. J., Sonneveld, P., Daenen, S. M., Biesma, D. H., Verdonck, L. F., . . . Verhoef, G. E. (2004). The value of fludarabine in addition to ARA-C and G-CSF in the treatment of patients with high-risk

myelodysplastic syndromes and AML in elderly patients. *Blood*, *103*(8), 2908-2913.

Pastore, D., Specchia, G., Carluccio, P., Liso, A., Mestice, A., Rizzi, R., . . . Liso, V. (2003). FLAG-IDA in the treatment of refractory/relapsed acute myeloid leukemia: Single-center experience. *Annals of Hematology*, *82*(4), 231-235.

Peduzzi, P., Concato, J., Feinstein, A. R., & Holford, T. R. (1995). Importance of events per independent variable in proportional hazards regression analysis II. Accuracy and precision of regression estimates. *Journal of Clinical Epidemiology*, *48*(12), 1503-1510.

Peduzzi, P., Concato, J., Kemper, E., Holford, T. R., & Feinstein, A. R. (1996). A simulation study of the number of events per variable in logistic regression analysis. *Journal of Clinical Epidemiology*, *49*(12), 1373-1379.

Peto, R., & Peto, J. (1972). Asymptotically efficient rank invariant test procedures. *Journal of the Royal Statistical Society, Series A (General)*, *135*(2), 185-207.

Ramos, N. R., Mo, C. C., Karp, J. E., & Hourigan, C. S. (2015). Current approaches in the treatment of relapsed and refractory acute myeloid leukemia. *Journal of Clinical Medicine*, *4*(4), 665 - 695.

Robak, P., & Robak, T. (2013). Older and new purine nucleoside analogs for patients with acute leukemias. *Cancer Treatment Reviews*, *39*(8), 851-861.

Robinson, S. K., & Broadfield, L. (2005). Guidelines for the management of acute myelogenous leukemia in adults. Hematology Cancer Site Team. *Cancer Care Nova Scotia*.

Sperr, W. R., Hauswirth, A. W., Wimazal, F., Knöbl, P., Geissler, K., & Valent, P. (2003). Treatment concepts for elderly patients with acute myeloid leukemia. *Wiener klinische Wochenschrift*, *115*(13-14), 505-514.

Spruance, S. L., Reid, J. E., Grace, M., & Samore, M. (2004). Hazard ratio in clinical trials. *Antimicrobial Agents and Chemotherapy*, *48*(8), 2787-2792.

Steyerberg, E. W., Eijkemans, M. J. C., & Habbema, J. D. F. (1999). Stepwise selection in small data sets: A simulation study of bias in logistic regression analysis. *Journal of Clinical Epidemiology*, *52*(10), 935 - 942.

Thein, M. S., Ershler, W. B., Jemal, A., Yates, J. W., & Baer, M. R. (2013). Outcome of older patients with acute myeloid leukemia. *Cancer*, *119*(15), 2720-2727.

Tilly, H., Castaigne, S., Bordessoule, D., Casassus, P., Le Prisé, P. Y., Tertian, G., . . . Degos, L. (1990). Low-dose cytarabine versus intensive chemotherapy in the treatment of acute nonlymphocytic leukemia in the elderly. *Journal of Clinical Oncology*, *8*(2), 272-279.

Vittinghoff, E., & McCulloh, C. E. (2007). Relaxing the rule of ten events per variable in logistic and Cox regression. *American Journal of Epidemiology*, *165*(6), 710-718.

Zwiener, I., Blettner, M., & Hommel, G. (2011). Survival analysis: Part 15 of a series on evaluation of scientific publications. *Deutsches Arzteblatt International*, *108*(10), 163-169.

LIST OF PUBLICATIONS AND PAPERS PRESENTED

List of Publications

1. Fludarabine, High Dose Cytarabine and Granulocyte Colony-Stimulating Factor (FLAG) as Consolidation Chemotherapy in Older Patients with Acute Myeloid Leukemia: A Retrospective Cohort Study. *Indian Journal of Hematology and Blood Transfusion.* https://doi.org/10.1007/s12288-017-0790-3

   ISSN:0971-4502 (Print version), ISSN: 0974-0449 (Electronic version)

   Accepted: 6 February 2017, Published online: 14 February 2017

Indian J Hematol Blood Transfus
DOI 10.1007/s12288-017-0790-3

ORIGINAL ARTICLE

Fludarabine, High Dose Cytarabine and Granulocyte Colony-Stimulating Factor (FLAG) as Consolidation Chemotherapy in Older Patients with Acute Myeloid Leukemia: A Retrospective Cohort Study

Kian Boon Law[1,2] · Kian Meng Chang[3] · Nor Aishah Hamzah[2] · Kok Haur Ng[2] · Tee Chuan Ong[3]

Received: 2 November 2016 / Accepted: 6 February 2017

Abstract The study aimed to investigate the effect of consolidation treatment with fludarabine, high-dose cytarabine and granulocyte colony-stimulating factor or FLAG in older AML patients. The study included 41 eligible patients above 54 years old, who received both induction and consolidation chemotherapy for AML from 2008 to 2013. The study cohort had a minimum 24 months follow-up period. Survival analysis was carried out to assess patients' overall survival and disease free survival based on types of consolidation regimens. The consolidation treatment with FLAG exerted a protective effect to both overall survival and disease free survival in older patients. Patients who were consolidated with FLAG regimen had a significant longer overall survival (log-rank, $p = 0.0025$) and disease free survival (log-rank, $p = 0.0026$). The median overall survival was longer (18.70 months) with the use of FLAG when compared to non-FLAG group (8.09 months). The median disease free survival was also longer (13.84 months) with use of FLAG when compared to the non-FLAG group (4.44 months). Regression analysis with Cox model yielded hazard ratio of 0.245 ($p = 0.0094$) in overall survival and 0.217 ($p = 0.0068$) in disease free survival. The use of FLAG as consolidation treatment was associated with approximately 60–80% reduction in hazard rates. The result was adjusted for age, race and gender in regression analysis. Older AML patients had longer remission and survival when consolidated with FLAG regimen after the induction chemotherapy.

Keywords Acute myeloid leukemia · Consolidation treatment · Cytarabine · Fludarabine · Overall survival · Disease free survival

✉ Kok Haur Ng
kokhaur@um.edu.my

Kian Boon Law
kblaw@crc.gov.my

Kian Meng Chang
drchangkm@gmail.com

Nor Aishah Hamzah
aishah_hamzah@um.edu.my

Tee Chuan Ong
ongteechuan@gmail.com

[1] Clinical Trial Unit, Level 7, Hospital Ampang, Jalan Mewah Utara, Pandan Mewah, 68000 Ampang, Selangor, Malaysia

[2] Institute of Mathematical Sciences, Faculty of Science, University of Malaya, Kuala Lumpur 50603, Malaysia

[3] Department of Hematology, Hospital Ampang, Jalan Mewah Utara, Pandan Mewah, 68000 Ampang, Selangor, Malaysia

Introduction

Chemotherapy based on high dose infusional cytarabine (Ara-C) and anthracyclines is the primary treatment option for acute myeloid leukemia (AML). The combination is therapeutically intensive and can only be given to a clinically fit patient. Most of the AML patients need to undergo multiple cycles of chemotherapy from induction to consolidation. The induction chemotherapy aims to achieve complete remission (CR) by reducing the total leukemic cell population in the body from around 10^{12} cells to below the cytologically detectable level of about 10^9 cells. The consolidation chemotherapy is given to sustain the disease remission by eliminating the remained leukemic blasts that are undetected by the current diagnostic tests. Consolidation chemotherapy is the pre-requisite for long term survival in patients. However, the optimal regimen for

Published online: 14 February 2017

List of Papers Presented

No.	Title	Events and Dates
1.	A retrospective study to assess the overall survival of elderly AML treated with FLAG regimen as consolidation therapy	Research proposal presentation for Master program 28 July 2015
2.	Fludarabine, High Dose Cytarabine and Granulocyte Colony Stimulating Factor (FLAG) might improve survival outcomes of elderly patients with acute myeloid leukemia	Journal Club of Hematology Department, Hospital Ampang 16 October 2015
3.	Fludarabine, High Dose Cytarabine and Granulocyte Colony Stimulating Factor (FLAG) might improve survival outcomes of elderly patients with acute myeloid leukemia	12th MPS Pharmacy Scientific Conference 2015, 13 November 2015
4.	A retrospective study to assess the overall survival of elderly AML treated with FLAG regimen as consolidation therapy	Candidature defence for Master program 28 January 2016
5.	Assessing the performance of regression estimators with strong association to survival outcome at low events per independent variable: A simulation study	The 3rd International Statistical Conference 2016 (ISM-III) 9 August 2016