PROGNOSTIC VALUE OF MOLECULAR SUBTYPES AND OTHER FACTORS ON DISEASE-FREE SURVIVAL IN WOMEN WITH BREAST CANCER AT UNIVERSITI SAINS MALAYSIA HOSPITAL

By

OH CHEN WEI

Thesis submitted in fulfilment of the requirements
for the degree of
Master of Science (Medical Statistic)

April 2018

ACKNOWLEDGEMENT

The author would like to specifically like to thank Professor Dr Norsa'adah Bachok, the supervisor for her advice and encouragements in the preparation of this dissertation; Dr Maya Mazuwin Yahya, the co-supervisor for her kind advice to the clinical aspect of the study; Dr. Najib Majdi Bin Yaacob, the co-supervisor for his support and advice on study sampling matter. Assoc. Prof. Dr Sarimah Abdullah, the coordinator and all the lecturers in Unit of Biostatistics and Research Methodology, Universiti Sain Malaysia (USM) for the fruitful lectures and advice; Professor Dr Than Winn, Assoc. Prof Dr Sarimah Abdullah and Dr. Siti Azrin Ab. Hamid, for their valuable comment as panel of viva; Mr Azri Nasruddin and Ms Nurzulaikha Mahd Ablah, the senior medical statisticians for thoughtful and practical view on gathering information; his course mates for their moral support and kind assistance; and most importantly, his beloved family for their endless love and countless supports on the drawbacks throughout the research.

The study has received financial support from the Tabung Insentif Pembangunan Pengajian Siswazah (TIPPS), School of Medical Sciences, USM. The author thanks Record Department of USM Hospital for providing the patient's data and Human Ethics Committee of USM for reviewing the study.

TABLE OF CONTENTS

CONTENTS

LIST OF TABLES

LIST OF FIGURES

LIST OF ABBREVIATION

CI	Confidence Interval
DFS	Disease-Free Survival
DF	Degree of Freedom
ER	Estrogen Receptor
EBCTCG	Early Breast Cancer Trialist's Collaborative Group
HER2	Human Epidermal Growth Factor Receptor 2
HR	Hazard Ratio
IDC	Invasive Ductal Carcinoma
ILC	Invasive Lobular Carcinoma
IQR	Interquartile Range
LML	Log Minus Log
LNs	Number of positive lymph node
LNR	Lymph Node Ratio
LR	Likelihood Ratio
NCR	National Cancer Registry
OR	Odds Ratio
OS	Overall Survival
PABC	Pregnancy-associated Breast Cancer
PR	Progesterone Receptor
PS	Power and Sample Size Calculation
RCT	Randomised Clinical Trial
SEER	Surveillance, Epidemiology, and End Results
SD	Standard Deviation
SPSS ®	Statistical Package for Social Science
STATA®	Data Analysis and Statistical Software
TNM	TNM Classification of Malignant Tumours

TNBC	Triple-Negative Breast Cancer
USM	Universiti Sains Malaysia

LIST OF SYMBOLS

<	Less than
>	More than
≤	Less than or equal to
≥	More than or equal to
α	Level of significance (alpha)
B	Regression coefficient
=	Equal
%	Percentage
1-β	Power
N	Number of samples

ABSTRAK

Tajuk: Nilai Prognostik Subjenis Molekul dan Faktor-faktor Lain Terhadap Kemandarian Tanpa Penyakit dalam Kalangan Wanita dengan Penyakit Kanser Payudara di Hospital USM.

Pengenalan: Dengan kejadian baru sebanyak 18,206 pesakit pada tahun 2007-2011, kanser payudara adalah kanser yang paling kerap di Malaysia. Kira-kira 2,572 pesakit mati akibat kanser payudara pada tahun 2012.

Objektif: Kajian ini mengakses kemandarian bebas penyakit 5 tahun, subjenis molekul dan faktor-faktor prognostik lain untuk kanser payudara.

Kaedah: Kajian secara retrospektif satu pusat dalam kalangan 208 pesakit dengan peyakit kanser payudara tanpa metastasis yang telah menerima pembedahan antara tahun 2007 dan 2015. Data yang dikumpul termasuk demografi pesakit, ciri-ciri tumor, profil histopatologi dan rawatan. Pesakit dibahagikan kepada empat kumpulan: Luminal A: ER + dan / atau PR + dan HER2- Luminal B: ER + dan / atau PR + dan HER2 +; HER-2 diperkaya: ER-, PR- dan HER2 +; Triple negatif: ER-, PR- dan HER2-. Nilai nisbah bilangan nodus positif dan nodus diperiksa dikelaskan kepada 0, ≤0.250,> 0.20 dan ≤ 0.65 serta > 0.65. Analisis Kaplan Meier dijalankan. Regresi *cox proportional hazard* digunakan untuk mengenalpasti faktor prognostic.

Keputusan: Kadar kemandirian tanpa penyakit antara 208 pesakit adalah 55% (95% SK:.46.6-62.7) Jenis kanser antara pesakit adalah: *Luminal A* (35.2%), *Luminal B* (29.0%), *HER2-enriched* (18.1%), dan *Triple negatif* (17.6%). Dalam analisis ringkas, subjenis molekul tidak dikaitkan dengan kemandirian tanpa penyakit. Namun begitu, bilangan nodus positif (LNs) serta LNR adalah faktor prognostik bagi kemandirian tanpa penyakit. Dalam analisis berganda, LNs menjadi tidak bermakna dan LNR kekal sebagai faktor prognostic, serta menunjukkan risiko pesakit mengalami sebarang kejadian semula atau kematian untuk adalah lebih tinggi (NB 1.17; 95% CI: 1.10 – 1.26) bagi setiap peningkatan 0.1 unit dalam LNR.

Kesimpulan: Kajian ini melaporkan jenis kanser payudara jenis *Luminal* adalah subjenis kanser payudara yang paling biasa di kalangan pesakit wanita. Kajian ini juga menunjukkan

bahawa LNR adalah faktor prognostik yang lebih baik dalam meramalkan kemandirian tanpa peyakit jika berbanding dengan LNs. Subjenis molekul tidak merupakan faktor prognostik.

ABSTRACT

Title: Prognostic Value of Molecular Subtypes and Other Factors on Disease-Free Survival in Women with Breast Cancer at Universiti Sains Malaysia Hospital.

Introduction: Breast cancer is the most common cancer in Malaysia with a 5-year incidence of 18,206 patients during 2007-2011. Approximately 2,572 patients died of breast cancer in 2012 in Malaysia.

Objective: The study assessed the five-year disease-free survival (5-y DFS), distribution of molecular subtypes and other prognostic factors of breast cancer.

Methods: A single-center retrospective review was conducted on 208 patients with non-metastatic, operable breast cancer treated with mastectomy or breast-conserving surgery between 2007 and 2015. Data collected included patients' demographics, tumour's characteristics, histopathological profiles, receptor status and treatment modalities. Molecular subtypes of breast cancer were subdivided into four groups: Luminal A: ER+ and/or PR+ and HER2- Luminal B: ER+ and/or PR+ and HER2+; HER-2 enriched: ER-, PR- and HER2+; Triple negative: ER-, PR- and HER2-. Ratio of positive lymph nodes (LNs) and dissected LNs was classified as 0, ≤0.20, >0.20 and ≤ 0.65 and > 0.65. Kaplan-Meier survival analysis was performed. Cox proportional hazards regression was used to determine prognosis factor.

Result: The distribution of molecular subtypes among 208 patients was: Luminal A (35.2%), Luminal B (29.0%), HER2-enriched (18.1%), and Triple-negative (17.6%). In univariable analysis, the molecular subtype was not associated with DFS. However, both absolute number of positive lymph nodes and the LNR were significant prognostic factors of DFS. In multivariable analysis, the positive LNs lost significance and the LNR remained as a prognostic factor of DFS, patients had increased risk (aHR: 1.17; 95% CI: 1.10 - 1.26) to experience any recurrence or death for every increasing of 0.1 unit in LNR when radiotherapy was adjusted for.

Conclusion: The present study reported the Luminal type breast cancer was the most common subtype of breast cancer among female patients. The study also demonstrates that LNR is a better prognostic factor in predicting DFS than the absolute number of positive LNs. Molecular subtype is not a significant prognostic factor.

CHAPTER 1 - INTRODUCTION

1.1 Global Epidemiology of Breast Cancer

Breast cancer is the second most common cancer in the world and the most frequent cancer among women by far, with an estimated of 1.67 million new cases diagnosed in 2012 (25% of all cancers and 12 % of new cancer) (Ferlay *et al*., 2015). In term of mortality, the breast cancer caused around 522,000 deaths in 2012 and ranked as the fifth cause of death from cancer overall. In less developed countries, the breast cancer is the most frequent cause of cancer-related death in women population (324,000 deaths, 14.3%). It is the second cause of cancer-related death after lung cancer in developed countries (15.4%) and most common cause of cancer death in developing nation (14.3%) (Ferlay *et al*., 2015).

In USA, the American Cancer Society estimates more than 250,000 women were diagnosed with invasive breast cancer in 2017. However, the death rate from breast cancer had decreased by 1.9% per year among female patients in the nation (Ryerson *et al*., 2016).

In Asia-Pacific region, 404,000 cases of female breast cancer were diagnosed during 2012, corresponding to a rate of 30 per 100,000. In Asia, the incidence of breast cancer was lower compared to Northern America and Europe (Ferlay *et al*., 2015). Consistent with the worldwide figure, breast cancer was the most common cancer among women in the Asia-Pacific region, accounting for 18% of all type of cancer. In term of mortality, 116,000 females were estimated to have died from breast cancer in 2012, representing 22% of total death in global (Youlden *et al*., 2014).

A research noted the incidence rates of breast cancer were relatively high in Southeastern Asia (Philippines, Singapore and Thailand) when compared to other region in Asia such as China, Japan, Korea and Taiwan.(Shin 2010).

1.2 Epidemiology of Breast Cancer in Malaysia

Breast cancer was the most common cancers, leading to death among women in Malaysia. A total of 18,206 cases of breast cancer were diagnosed during 2007-2011, accounting for 17.7% of all cancer incidence (Azizah *et al*. 2016). The GLOBALCAN 2012 study estimated the age-standardised mortality rate is 18.9 per 100000 population (Ferlay *et al*. 2015). Approximately 2,572 patients died of breast cancer in 2012. Larger increment in breast cancer mortality was observed in Malaysia as compared to other regions in Asia such as Hong Kong and Singapore (Youlden *et al*., 2014).

Among races in Malaysia, the overall cumulative risk of having breast cancer was 3.4 across the population and was highest among Chinese (4.5%) and lowest among Malays (2.9%). The incidence rate increase since the second decade of life and the peak incidence occurred in the 5th decade of life. The proportion of breast cancer detected at stage I, II, III and IV were 20%, 20%, 23% and 37%, respectively (Azizah *et al*., 2016).

1.3 Justification of The Study

Estrogen receptor, progesterone receptor and HER2 receptor are important in the management of breast cancer as it will further classify breast cancer into four distinct molecular subtypes, i.e., Luminal A, Luminal B, HER2 overexpressing and Triple-negative breast cancer, along with different prognostic implications (Carey *et al*., 2006). In Malaysia, studies on the topic regarding the prognostic value of breast cancer molecular subtype or hormonal receptor / HER2 receptor status are very few. To the best of current knowledge, only one study had been conducted in Sarawak determined the prevalence and distribution pattern of each molecular subtype in Malaysia (Devi *et al*., 2012), and none has been published with regards to the prognostic value of distinct subtype in the Malaysian population.

Yip *et al*. reviewed 421 breast cancer research in Malaysia from 1996 to 2014 and suggested future research on the reason why the Malay ethnicity had a poorer survival may be warranted

by studying pharmacogenomics in connection with chemotherapy as a possible factor (Yip *et al*., 2014). Another author recommended the prognostic value of various biomarkers, e.g., HER-2, in local populations (Ong andYip, 2003) to be studied in future research. Even though a similar study by Leow *et al*. (2007) had been conducted to determining prognostic factor for overall survival of breast cancer, the study did not include HER-2 receptor status and the information on ER and PR receptor status were limited, hence the distribution of molecular subtype was not reported. In addition to that, there is limited published study to review the locoregional recurrence of breast cancer in Malaysia.

This study was conducted in Kelantan, a state located in northern region of Peninsular Malaysia. This research aimed to examine the distribution pattern of molecular subtype among patients, and the prognostics value of molecular subtypes and/or ER/PR/HER2 receptors in women with breast cancer with a recent population data (2007-2015). DFS was used to provide some information on the locoregional recurrence for Malaysian female diagnosed with breast cancer.

Along with molecular subtype as the primary interest, this study also aimed to investigate the other potential patient-related, clinical and pathological related and treatment-related factor toward breast cancer DFS. With the above interest in mind, this single centre study was expected to eventually provide additional value to improve the breast cancer management in patients living in Kelantan, and reasonably to be extrapolated to whole Malaysian breast cancer population.

1.4 Research Questions

- What were the overall locoregional recurrence and five-year disease-free survival (5-y DFS) for women diagnosed with non-metastatic, operable breast cancer treated with mastectomy or breast-conserving surgery at USM hospital?

- What was the distribution of molecular subtype among women diagnosed with non-metastatic, operable breast cancer treated with mastectomy or breast-conserving surgery at USM hospital?
- Was the molecular subtype the prognostic factor of the DFS for women diagnosed with non-metastatic, operable breast cancer treated with mastectomy or breast-conserving surgery at USM hospital?
- What were the other factors associated with DFS for women diagnosed with non-metastatic, operable breast cancer treated with mastectomy or breast-conserving surgery at USM hospital?

1.5 General Objectives

- To study the overall locoregional recurrence, 5-y DFS and prognostic factor among women diagnosed with non-metastatic, operable breast cancer treated with mastectomy or breast-conserving surgery at USM hospital

1.6 Specific Objectives

- To determine the overall locoregional recurrence and 5-y DFS among women diagnosed with non-metastatic, operable breast cancer treated with mastectomy or breast-conserving surgery at USM hospital
- To determine the distribution of molecular subtype survival among women diagnosed with non-metastatic, operable breast cancer treated with mastectomy or breast-conserving surgery at USM hospital
- To determine the association of molecular subtype, patient-related, clinical and pathological related and therapy-related factor toward 5-y DFS among women diagnosed with non-metastatic, operable breast cancer treated with mastectomy or breast-conserving surgery at USM hospital

1.7 Research Hypothesis

- The molecular subtype was associated with 5-y DFS among women diagnosed with non-metastatic, operable breast cancer treated with mastectomy or breast-conserving surgery at USM hospital.

- The patient-related factors, clinical and pathological-related factors, therapy-related factors were associated with 5-y DFS among women diagnosed with non-metastatic, operable breast cancer treated with mastectomy or breast-conserving surgery at USM hospital.

CHAPTER 2 - LITERATURE REVIEW

2.1 Literature Search Strategies

Literature search strategy includes using the phrase, using Boolean operator and keywords as well as searching by author and citation. Search engines were used to obtain the relevant articles on this topic. Searching databases housing relevant journals, for example, Science Direct, Scopus, Pubmed, Google Scholar, EBSCOhost, Springerlink, Web of Science was given preference over single-journal search engine except for journals, for example Online Access Catalogue (OPAC Krisalis) of Perpustakaan Hamdan Tahir, USM.

Keywords:

"breast cancer" or "breast carcinoma" or "breast tumor" AND

"Disease-free survival" or "DFS" AND

"molecular subtype" or "Luminal A" or "Luminal B" or "HER2 overexpress" or "Triple negative" or "basal-like" AND/OR

"Hormonal receptor" or "estrogen receptor" or "ER" or "progesterone receptor" or "PR" AND/OR

"HER2 receptor" or "neu" or "erbB-2" or "HER2/neu" AND

"prognostic factor", "associated factor", "predictive factor" AND

"International" or "Malaysia"

Table 2-1: The literature search results by using Google Scholar, PubMed, Scopus and Web of Science.

	Search Engine			
	Google Scholar	PubMed	Scopus	Web of Science
Using Phrase				
breast cancer disease-free survival	2,140,000	1443	1760	1794
breast cancer disease-free survival molecular subtype	110,000	32	52	85
breast cancer disease-free survival molecular subtype Malaysia	2500	0	0	0
Using Boolean Operators and keywords				
"Disease-free survival" AND "breast cancer" AND "prognostic factor"	4230	211	182	86
"Disease-free survival" AND "breast cancer" AND "molecular subtype"	3220	32	34	38
Citation Search	Using author's name or title of the article			

2.2 Survival Probability and DFS of Breast Cancer

A review of US National Cancer Institute's SEER database (1975-2013) revealed that the five-year relative survival for localized, regionalized and distant breast cancer is 98.6%, 84.3% and 24.1%, respectively (Howlader *et al.*, 2016). The most recent data indicated that the survival rates for female patients with advanced breast cancer (Stage III and IV) are lower than those for women with earlier stage cancer. The five-year survival rate for stage III breast cancer is approximately 72%, while metastatic (stage IV) breast cancer possess a much lower five-year relative survival rate at about 22% (American Cancer Society, 2017).

Similarly, the data from UK (2002-2006) revealed that the five-year relative survival for stage I, II and III are 99%, 88% and 55% respectively. However, the five-year survival rate dramatically dropped to 15% for stage IV breast cancer (Cancer Research UK, 2017). About one-third of patients with early-stage breast cancer will subsequently progress to metastatic disease (O'Shaughnessy, 2005).

A population-based study based on data from 13,060 patient who were admitted to the hospital in Malaysia during 2000 – 2005 reported that the overall five-year survival (OS) rate was 49% with median survival time of 68.1 months (Abdullah *et al*., 2013). Other researchers in Malaysia reported a range from 43.5% – 69% of OS and a similar decreasing trend of OS across the stage I to stage IV breast cancer patient (Ibrahim *et al*., 2012; Nur Aishah *et al*., 2008; Saxena *et al*., 2012). The previous study in Kelantan revealed that the breast cancer patient suffered a poorer prognosis compared to another region in Malaysia, in particular, the OS was 25.84% (Leow *et al*., 2007). Table 2.2 summarized studies from local society that showed a relatively poorer overall survival compared to developed nations.

Table 2-2: The five-year overall survival based on whole study population and staging of cancer at presentation during diagnosis of disease.

Author	Institution	n	% (95% CI)				
			5-y OS	Stage I	Stage II	Stage III	Stage IV
Leow *et al*., 2007	USM Hospital 1987-2000	185	25.8	100	IIA 58.3 IIB 31.8	IIIA 31.6 IIIB 4.4 IIIC 11.1	3.6
Taib *et al.*, 2011	UMMC 1993-97	423	58.4	81.7	72.4	39.9	12.8
	UMMC 1998-2002	965	(54.0,63.0)				

Author	Institution	n	% (95% CI)				
			5-y OS	Stage I	Stage II	Stage III	Stage IV
Ibrahim *et al.*, 2012	HKL 2005-2009	868	43.5	58.0 (54.2, 61.8)	52.7 (50.2, 55.1)	39.0 (35.8, 42.6)	19.8 (17.0, 22.7)
Saxena *et al.*, 2012	UMMC 1993-2007	3,320	69.0 (67.0,71.1)	93.0 (91.9,94.1)	79.0 (77.8,80.3)	52.0 (49.4, 54.6)	12.0 (6.8, 17.1)
	NUH Singapore 1993-2007	2,141	80.0 (79.0,80.9)	98.0 (97.0,99.0)	85.0 (83.7,86.3)	66.0 (62.5,69.6)	230 (16.6,29.5)
Abdullah *et al.*, 2013	National Registration Dept 2000-2005	10,230	49.4	NA	NA	NA	NA

NA: Not applicable

The development of a variety of new treatment agent and the multiplication lines of treatment in breast cancer have significantly reduced mortality in certain contexts. This therapeutics advancement has resulted in the need for new surrogate endpoints to be used in breast cancer research. Therefore, DFS may be obtained in shorter time frame compared to overall survival (OS) and had been used as surrogate endpoints in non-metastatic breast cancer settings. The adoption of a surrogate endpoint such as disease-free survival has been strongly influenced by the need to reduce the subject's sample size and the cost of research especially in the randomized controlled trial (Gourgou-Bourgade *et al*., 2015).

The researcher reported women diagnosed with breast cancer had a 5-y DFS range from 68.5% - 88.5%. The DFS vary across studies based on the disease presentation, patients characteristics and treatment received by patients *(Kim et al*., 2005; Diniz *et al.*, 2016; Wen *et al.*, 2016). Locoregional recurrence (LRR) after surgery especially mastectomy without distant metastases was common in breast cancer management, although the proportion of patient suffered from LRR notably varies among published studies as a result of differences in disease characteristics, surgical and adjuvant treatments (Clemons *et al*., 2001) A local study indicated

the overall LRR in breast cancer patient are 16.4% after a median follow-up period of 67 months. Table 2.3 summarizes studies from global that showed overall survival and DFS of breast cancer patients.

Table 2-3: OS, DFS and LRR

Author, Date Country	Study Design, Number of subjects	Presentation /treatment	OS	DFS	LRR
Katz *et al.*, 2000 USA	Five clinical trial n= 1,031	Stage II, IIIa Mastectomy and chemotherapy	10-y: 65%	10-y: 55%	NA
Nielsen *et al.*, 2006 USA	RCT n=3,083	Stage II, III Mastectomy	5-y: 36%	NA	NA
Zhang *et al.*, 2015 USA	Retrospective cohort n=160	Neoadjuvant therapy surgery and adjuvant Radiotherapy	3-y: 85.4%	3-y: 68.5%	NA
Diniz *et al.*, 2016 Brazil	Retrospective cohort n=563	Non-metastatic breast cancer	NA	5-y: 72%	NA
Kim *et al.*, 2005 Korea	Retrospective cohort, 1994-2001 n=805	Stage I and II breast cancers received breast-conserving treatment and radiotherapy	NA	5-y: 88.5%	5-y: 2.8%
Wen *et al.*, 2016 Korea	Retrospective cohort n=701	Early breast cancer with N1 Without postmastectomy radiotherapy	5-y: 91.2%	5-y: 86.3% 10-y: 77.5%	NA
Li *et al.*, 2014 PR China	Retrospective cohort, 1998-2007 (n=439)	Postmastectomy radiotherapy with four and more positive axillary lymph node	5-y: 70.7%	NA	5-y: 12.2%
Choong *et al.*, 2010 Malaysia	Retrospective cohort, 1998-2002 (n=522)	Mastectomy	NA	NA	Overall: 16.4%

NA: Not applicable

2.3 Classification of Molecular Subtype of Breast Cancer and Its Prognostic Value

Biologic molecular and genetic knowledge has provided a new understanding of breast cancer that categorized into different subtype that possessed different clinical characteristic, treatment response and disease outcome (Prat and Perou, 2011). Although the modern microarray gene profiling is the best way to categorise the breast cancer subtype, the high cost and practicability limited the use in clinical setting hence it warranted a classification by routinely obtained surrogate marker (Parker *et al*., 2009).

Many investigators had developed an alternative method by using immunochemical tests for estrogen receptor (ER), progesterone receptor (PR) and HER2 receptor as surrogate approaches for molecular subtypes of breast cancer (Goldhirsch *et al*., 2011). In particular, five intrinsic breast cancer subtypes surrogated by immunohistochemical (IHC) analysis that recommended by the St. Gallen International Breast Cancer Conference (2011) are: luminal A [ER+ and/or PR+, Ki67 low and HER2−], luminal B [ER+ and/or PR+, Ki67 high and HER2−], luminal HER2+ [ER+ and/or PR+, any Ki67 and HER2+], HER2 [ER−, PR−, and HER2+], and triple negative (TN) [ER−, PR−, and HER2−](Goldhirsch *et al*., 2011). However, the classification remained controversial. Some investigators considered any HER2+ tumor that is ER+ and/or PR+ to be luminal B (Carey *et al*., 2006; Tamimi *et al*., 2008). On the other hand, some researcher suggested to use another biomarker Ki67 to define Luminal B into 2 types further, one is Luminal B(HER2−) with a high proliferation index Ki67, and the other one is Luminal B (HER2+) (Goldhirsch *et al*., 2011; Zafrani *et al*., 1994). However, some researchers against Ki67 usage and had suggested replacing with tumor grade to helping define the luminal B phenotypes (Brouckaert *et al*., 2013). To date, the debate on the cutoff point of Ki 67 remained. Some researchers had suggested a lower value of the Ki67 index to be used for better differentiation between ‘Luminal A’ and ‘Luminal B (HER2 negative)’(Bustreo *et al*., 2016). In contrast, some study reported low reproducibility of Ki67 result among the laboratory (Varga *et al*., 2012; Polley *et al*., 2013). The recent Saint Gallen Breast Cancer Conference has recommended the use of the median Ki67 value of local laboratory in

classifying breast cancer molecular subtype (Coates *et al*., 2015). In the circumstances, if Ki67 value was unavailable, some researcher identified four categories of breast cancer: luminal A, luminal B, HER2 overexpressing, and basal-like or triple negative based on the status of ER, PR and HER2 (Perou *et al*., 2000; Sørlie *et al*., 2001). The Ki67 test was however not available in USM hospital.

Most of the breast cancers (up to 75%) expressed either estrogen receptor or progesterone receptor (hormone-receptor [HR]–positive) (Setiawan *et al*., 2009). In general, Luminal A type cancers tend to be slow-growing and less aggressive compared to other subtypes. These tumors were associated with the better prognosis because expression of hormone receptors is a predictive factor of hormonal therapy's response (Anderson *et al*., 2014; Blows *et al*., 2010). Most of the patients possessed Luminal A type breast cancer (44.7%) (Hennigs *et al*., 2016).

Luminal B was similar to luminal A breast cancers in term of ER+ and/or PR+ and are further defined by being highly positive for Ki67 (an indicator of a large proportion of actively dividing cells) or HER2. About 10% of breast cancers are ER+ and/or PR+ and HER2+. Luminal B breast cancers tend to be higher grade and more aggressive than luminal A breast cancers (Parise and Caggiano, 2014).

In US, about 12% of breast cancers are triple negative, so called because all three receptors were expressing negative status (ER-, PR-, and HER2-); however, the ethnicity is an associated factor in US population: Blacks are at about two times higher risk of getting triple-negative breast cancer than white women. Triple-negative breast cancers tend to have a poorer prognosis than other breast cancer types, as these subtypes are not warranted for hormonal or neoadjuvant therapy (Blows *et al*., 2010; Parise and Caggiano, 2014).

About 4% of breast cancers produce excess HER2 and do not express hormone receptors. These cancers tend to grow and spread more aggressively than other breast cancers. In general, this breast cancer subtype is associated with poorer short-term prognosis compared to ER+ breast cancers (Blows *et al*., 2010).

As for Asia, a study of 346 Chinese and Japanese in New York City had a higher proportion of Luminal A breast cancer (66.7% and 80.0 %, respectively) when compared to Filipinos and Koreans. On the other hand, Filipinos had a higher proportion of HER2 positive cancers (45.6%) while Koreans had a higher proportion of triple-negative cancers (23.55%) (Chuang *et al*., 2012). A study reviewed 471 Japanese and noted that 65% were Luminal A, 8.7% were Luminal B, 12.5% were HER2 enriched, 7.9% were basal-like 5.9% were unclassified (Tamaki *et al*., 2013). Choi *et al*. (2007) found there was a difference in HER2 expression between Korean and white breast cancer patients (47.5% and 15.8%, respectively). A multi-region gene profiling study analysed 78 samples from Chinese subjects and indicated 27% were Luminal A, 29% were Luminal B, 22% were HER2 enriched, 13% were Basal-like and 9% of them are undetermined. The prevalence of intrinsic subtypes in Chinese was significantly different from Caucasian (Huang *et al*., 2015).

In Malaysia, the only study on distribution of population molecular subtype reported that 48% luminal A (ER+ PR+ HER2-) breast cancer, 12% Luminal B with HER2 positive (ER+PR+HER2+), 29% TNBC and 11% HER2 overexpressing subtypes among the 1034 subjects (ER-PR-HER2+) (Devi *et al*., 2012) .

2.4 Socio-Demographic Prognostic Factors of Breast Cancer

An informative prognostic factor in should possess the following characteristic:

- Provides significant and independent prognostic value
- Feasible, reproducible, and widely available

- The conclusion is readily interpretable by the healthcare worker (Rosner an dLane, 1991)

2.4.1 Age

A study in Sweden with 57,068 women included indicated that both younger or older age at diagnosis was associated with a poorer relative survival compared to those whose age were amongst the middle (Adami *et al*., 1986). In a registry-based cohort of 22,017 women, patients age 20 to 35 years showed a worse five-year survival (74.7% compared to 83.8 – 88.3% for women ages 35 to 69 years) (Fredholm *et al*., 2009). Another research found younger age (<35) was an independent risk factor for disease relapse in operable breast cancer female (Han and Kang, 2010). Similarly, a study found young patients with breast cancer who did not receive adjuvant treatment were at higher risk of death from the disease (Kroman *et al*., 2000). The St. Gallen Consensus conferences back in 1998 and 2001 also indicated that age below 35 was a risk factor for disease relapse in women with node-negative breast cancer patients (Zujewski and Liu, 1998; Goldhirsch *et al*., 2001). In a study of 17,575 women with stage I to III breast cancer, patients with ≤ 40 years of age at diagnosis showed higher breast cancer mortality as compared to older patients (HR: 1.4; 95%CI: 1.2,1.7) (Partridge *et al*., 2016).

2.4.2 Ethnicity

Surveillance, Epidemiology and End Results (SEER) study in 2013 reported that black women had lower rates of breast cancer screening and longer delay to the initiation of the treatment, thus having 12.9% poorer five-year survival compared to whites (blacks, 55.9%; whites, 68.8%) (Silber *et al*., 2013). A recent review of breast cancer research in Malaysia by Yip *et al*., (2014) showed that Malays was associated with higher risk of all-cause mortality. However, the author also pointed out that the factors accounted for the ethnic differences were unclear. A study combining data from two hospitals located in Malaysia and Singapore demonstrated Malay ethnicity was significantly associated with larger tumour and later stage at presentation

compared to other races (Bhoo-Pathy *et al*., 2012). Ong and Yip (2000) observed the race is a notable predictor of recurrence-free survival but not significant forward overall survival.

2.4.3 Obesity

Protani *et al*. concluded in their recent systematic review and meta-analysis on some observational studies that obese women at the time of diagnosis had a 33% higher chance of disease recurrence or mortality when compared with normal-weight patient (Protani *et al*., 2010). Another single centre study analysing DFS also supported the prognostic value of weight, where it showed among never-smoking women, those who gained between 0.5 and 2.0 kg/m^2 (relative risk: 1.35; 95% CI, 0.93 to 1.95) or more than 2.0 kg/m^2 (RR: 1.64; 95% CI, 1.07 to 2.51) after diagnosis were at higher risk of mortality during 9 years of follow-up, compared with patients who had maintained their body weight (Kroenke *et al*., 2005).

2.4.4 Smoking

Cigarette smoking both before and after breast cancer diagnosis had been shown to relate to increasing of breast cancer mortality. Patients who smoked actively during the year before the cancer diagnosis were at higher hazard (HR 1.25, 95% CI: 1.13, 1.37) to die compared to never-smokers. On top of that, women who continue smoke lifestyle after diagnosis were even more likely than never-smokers to die of breast cancer (HR 1.72, 95% CI 1.13, 2.60) (Passarelli *et al*., 2016). Several studies also confirm that smoker or former smoker worsens the prognosis after surgery among women with breast cancer (Hellmann *et al*., 2010; Braithwaite *et al.*, 2012; Saquib *et al.*, 2013; Sollie and Bille, 2017).

2.4.5 Marital status

A recent study indicated unmarried patients suffered a 28% higher chance of death after diagnosed with breast cancer when compared to the married patient (Martínez *et al.*, 2017).

2.4.5 Parity status

A study highlighted nulliparous status was an independent predictor of poorer breast cancer survival after adjusting for age (Anderson *et al.*, 2014). However, the study also showed that along with no child, high parity patient (four or more children) also had significantly higher mortality from breast cancer when compared to women with a child (RR 1.49, 95% CI 1.20-1.85) (Butt *et al*., 2009).

2.4.6 Age at first childbirth and age at last childbirth

Study showed those who had breast cancer within 5-10 years of delivery have a 62% higher chance to carry a poorer outcome as compared to the other (Strasser-Weippl *et al*., 2015). A study in Danish women demonstrated breast cancer patients with first childbirth between 20 and 29 years had a significantly lower risk of death compared with women with first childbirth below the age of 20 years (20-24 years: RR = 0.88, 95% CI: 0.78-0.99; 25-29 years: RR = 0.80, 95% CI : 0.70-0.91) (Kroman *et al*., 1998)

2.5 Histopathological Prognostic Factors of Breast Cancer

2.5.1 Multifocal or multicentric tumours

According to a recent study, whether multifocal (i.e., invasive tumors identified within the same breast quadrant) or multicentric (i.e., invasive tumours identified in separate breast quadrants) influence prognosis was still controversial. A study with 288 pairs of breast cancer female using Cox regression analysis for multivariate analyses demonstrated multifocal or multicentric cancer were significant predictors for overall survival (HR = 1.57; 95% CI: NA) and overall-recurrence (HR = 1.74; 95% CI: NA). However, another study showed multifocal or multicentric breast cancer did not have predictive value on recurrence-free survival, BCSS, or OS (Weissenbacher *et al*., 2010; Lynch *et al*., 2012).

2.5.2 Tumour, Nodes, Metastasis (TNM) system

Generally, tumour stage was a well-reported prognostic factor. The breast cancer is staged according to tumour size, number of lymph nodes involved, and presence of metastatic disease. (Tumour, Nodes, Metastasis [TNM] system). The SEER data in US revealed that five-year relative survival rates are 95, 85, 70, 52, 48 and 18 percent for stage I, IIA, IIB, IIIA, IIIB, and IV breast cancer, respectively (Newman, 2009). The data from UK (2002-2006) revealed that the five-year relative survival for stage I, II and III are 99%, 88% and 55% respectively (Cancer Research UK, 2017).

2.5.3 Tumour Size

Tumour Size (T) is defined as the greatest diameter of the primary breast tumour was an important prognostic factor in breast cancer (Fisher *et al*., 1969). The SEER database includes 13,464 women with node-negative breast cancer. Patients with tumors <1 cm had a 5-year OS of close to 99% compared with 89% for tumors between 1 cm and 3 cm and 86% for tumors between 3 cm and 5 cm (Carter et al., 1989). This association persists with longer follow-up. Rosen et al. examined the relationship between tumor size and 20-year recurrence-free survival and found a significant association, with a 20-year recurrence-free survival of 88% for tumors ≤1 cm, 72% for tumors 1.1 cm to 3 cm, and 59% for tumors between 3.1 cm and 5 cm (Rosen *et al.*, 1993).

2.5.4 Nodal Involvement

The number of axillary nodes with metastatic tumour growth was a well-reported independent prognostic factor for disease recurrence or mortality. Among women without metastatic progression at diagnoses, the five-year survival rate for those with localized and regional disease was 99% and 85%, respectively (Siegel *et al*., 2017). Another study showed the survival was related to number of nodes where the five-year survival was 96, 86, and 66 % for

patients who were node-negative, had one to three nodes involved or had more than four nodes involved, respectively (Carter *et al*., 1989).

2.5.5 Tumour morphology

Invasive ductal carcinoma (IDC) is the most common morphology type of breast cancer which account for more than 70% of all breast cancer new cases, followed by invasive lobular carcinoma (ILC) with a frequency of about 10 % as shown in a study reviewing US SEER record from 1992 – 2001 (Li *et al*., 2005). ILC had 16% lower risk of experiencing disease recurrence compared to IDC during the first six years of follow-up in a study; However, ILC had a 54 % higher risk of disease recurrence after six years (Pestalozzi *et al*., 2008).

2.5.6 Histologic grade

A study with 2200 patients found a very strong correlation between long-term survival and Nottingham combined histologic grade (Elston and Ellis, 1991). However, The Breast Task Force of The American Joint Committee on Cancer had decided to remove grading from the current TNM staging system for breast cancer due to the grading system was limited by low interobserver agreement and the lack of prognostic information in grade 2 tumours (Rosner and Lane, 1991; Singletary *et al*., 2002).

2.5.7 Lymphovascular Invasion

The presence of lymphovascular invasion appeared to be a factor indicating poor disease prognosis. This was shown in a cohort study of 1704 patients that did not receive any systemic therapy, in which peritumoral lymphovascular invasion was a prognostic factor for local recurrence and survival (Pinder *et al*., 1994). In a US study comprising 166 women with operable invasive BC who underwent adriamycin- and taxane-based NAC between 2000 and 2013 multivariate models adjusting for breast cancer subtype, LVI was significantly associated

with a decrease in progression-free survival (HR 3.76, 95 % CI 2.07-6.83, $p < 0.01$) and overall survival (HR 5.70, 95 % CI 2.08-15.64, $p < 0.01$) (Liu et al., 2016).

2.6 Treatment-Related Prognostic Factors of Breast Cancer

2.6.1 Adjuvant Chemotherapy

The consideration of systemic adjuvant therapy after the surgical operation based on the predicted response and the possible drug adverse reaction to a treatment. The final decision on treatment also takes patients' age, general health status, comorbidities and preference into account. Adjuvant systemic treatment started within 2 – 6 weeks after the surgery (Senkus *et al*., 2015). It was a general practice to administer systemic therapy to breast cancer patient with lymph node-positive breast cancer (Cianfrocca and Goldstein, 2004).

Administration of anthracycline-based polychemotherapy such as FAC or FEC for about 6 months would decrease the annual breast cancer mortality rate by 38% for women age below 50 years and by 20% for patients of age 50–69 years when diagnosed, regardless of the usage of endocrine therapy, stage, histopathological grade and ER status (EBCTCG, 2005).

2.6.2 Radiotherapy

A study reviewed 42,000 women with breast cancer in randomized treatment comparisons and found that those who received and completed radiotherapy treatment had much lower risk toward local recurrence (7% vs. 26%) and significant lower five-year overall survival (30.5% vs 35.9%; p=0.0002). (Clarke *et al*., 2005). A later meta-analysis of 10,801 operated women in 17 randomised trials comparing receipt of radiotherapy versus no radiotherapy confirmed the findings above. In this study, it showed radiotherapy reduced both 10-year risk of any local or distant recurrence from 35.0% to 19.3% (absolute reduction 15.7%, 95% CI 13.7–17.7, $p<0{\cdot}001$) and the five-year risk of mortality from 25.2% to 21.4% (absolute reduction 3.8%,

95% CI: 1.6–6.0, p<0·001) (Early Breast Cancer Trialists' Collaborative Group (EBCTCG) *et al*., 2011)

2.6.3 Hormonal Therapy

Hormone therapy or endocrine therapy played a main role in adjuvant therapy especially in patients with estrogen receptor positive breast cancer. American Society of Clinical Oncology (ASCO) suggested women with stage I to III breast cancer disease considered having tamoxifen for 10 years in its clinical practice guideline on adjuvant endocrine therapy for women with hormone receptor–positive breast cancer that published in 2014 (Burstein *et al*., 2016). A study initiated by the Early Breast Cancer Trialists' Collaborative Group noted that 5 years administration of adjuvant tamoxifen therapy successfully reduced the risk of recurrence or death by a third (Davies *et al*., 2011). Table 2.4 summarized result from those studies that exploring prognostic factor of breast cancer among female patient underwent surgery using DFS.

Table 2-4: Five-year overall survival based on whole study population and staging of cancer at presentation during diagnosis of disease.

Author, Date Country	Participants number	Presentation/treatment	Study design	Significant factor	Not significant factor
Overgaard *et al*., 1997 Netherland	N=1708	Stage II, III, high risk premenopausal Mastectomy + radiotherapy	RCT	Radiotherapy after mastectomy (HR=0.59; 95% CI: 0.51-0.67)	
Choi *et al*., 2017 South Korea	N = 2441 2009 - 2010	Underwent surgery; Stage I-III	cohort	Age; Tumour size; ER; PR; HER2; Molecular Subtype; histology Grade; Pathological grade LVI; BCT	
Kim *et al.*, 2016 South Korea	N = 3151 2006 - 2010	Stage I-III; node(-); no neoadjuvant therapy.	Medical records review	T stage; harvested lymph node	Histology Grade; ER

ER: Estrogen Receptor; PR: Progesterone Receptor; LVI: Lymphovascular invasion

2.7 Survival Analysis

Survival data study the time elapsed from some particular starting point to the occurrence of the event. One of the distinctive characteristics of survival data compared to other types of analysis is the presence of censoring. In survival analysis, subjects are followed over the specific period until the main event of interest occurs. Only some individuals experienced the event in interest while the others' are remained unknown. This phenomenon is known as censoring and it may occur in the following ways:

- A subject has not yet experienced the event by the time of the closure of the study.
- A subject is lost to follow up during the accrual period
- A subject experienced a different event that makes further follow up impossible

All of these situations were defined as right censoring (Bradburn *et al*., 2003a; Clark *et al*., 2003). Most survival data include right-censored observation (Hosmer *et al*., 2008). Other types of censored data available are interval and left censored (Clark *et al*., 2003).

The time-to-event data are generally described as survival and hazard. The survival probability also known as survival function of which denoted as *S(t)* represent the probability of an individual survives from the time origin toward a future time of endpoint. On the other hand, the hazard function of T, denoted as h(t) is the probability of an individual under observation has the event of interest occurred at a specific time point *t* in future. In sum, the survival indicates the cumulative not having an outcome in question, while the hazard indicates the chance of having an outcome in question (Clark *et al.*, 2003).

Statistical methods that are generally applied in survival analysis are Kaplan-Meier for univariable analysis and Cox proportional hazard model for multivariable analysis (Clark *et al*., 2003). The Kaplan-Meier survival estimator provides the probability of survive over the observed time frame including censored and uncensored, using the product-limit method (Kaplan and Meier, 1958).

Survival in two or more groups of subjects can be compared using log-rank test which is the most widely used method in medical research (Peto *et al*., 1977). The log-rank test calculates the number of expected events at each event time and for each group. These values are then summed up for all event times to obtain the total expected events in each group, denoted as E_i for group i, then the log-rank test compared the observed events, denoted as O_i for group I to the expected event E_i by following test statistic:

$$X^2 = \sum_{i=1}^{g} \frac{(Oi - Ei)^2}{Ei}$$

The calculated value can be compared to a χ^2 distribution with (g-1) degrees of freedom when g is the number of groups. A *P*-value can be calculated to determine the statistical significance of the differences between the survival curves (Peto *et al*., 1977).

Length of follow-up can be determined by the median follow-up time. Patient who does not experience an event can be included in a survival analysis, but the completeness of follow-up among patients is important. Loss to follow-up may be determined by demonstrating the total number of subjects lost during follow-up period of the study (Clark *et al*., 2003).

The Cox or proportional hazards model (Cox, 1972) is the most widely used multivariate analytical method for survival data in medical research. The survival regression model describes the relation between event and a number of covariates, as presented by the hazard function or hazards ratio. The hazard is defined as the instantaneous probability of having event at a given time. Clark *et* al. (2003) used the following formula to mathematically explains the Cox Model:

$$h(t) = h_o(t)x \exp(b_1x_1 + b_2x_2 + \cdots + b_ix_i)$$

where the hazard function h(t) is determined by a number of *i* variables, and the impact of variable on hazard was computed as respective coefficients (*b*). The h_o term also known as baseline hazard.

The objective of the study affected the choice of covariates, there are three possible scenarios to use a multivariable Cox model:

1. A single variable is under investigation for its association with survival along with another known variable
2. A set of variables are under investigations for their ability in predicting survival
3. A set of variables are under investigations for their ability in predicting survival, along with additional known variables

Stepwise selection is common choice for automated model building, but other approaches exist. Henderson and Vellemen (1981) suggested that the choices should be verified by a degree of munual modelling, where terms can be included or removed based on logical order rather than solely followed to statistical significance. Clark *et al.* (2003) demonstrated the rationale and limitation of semi-automated stepwise selection method and provided advice on hands-on modelling.

Residuals form models are a useful method for checking the fit of a regression model. There are three types of residual proposed in accessing model adequacy: Martingale residual, deviance residual and scaled Schoenfeld residuals. In general, the residual is skewed, it needs to have smoothing function for easier interpretation. The scatter graph of one of the residual and variable or survival time shall display an evenly scattered horizontal band and no obvious trend shall be observed (Bradburn *et al*., 2003b).