THE ASSOCIATION OF MAMMOGRAPHIC DENSITY, AGE AND POSITIVE FAMILY HISTORY WITH MOLECULAR SUBTYPES AMONG BREAST CANCER PATIENT IN HOSPITAL UNIVERSITI SAINS MALAYSIA

DR SITI FITRIWATI KHAZIS BINTI ISMAIL

DISSERTATION SUBMITTED IN PARTIAL FULFILMENT OF THE REQUIREMENTS FOR THE DEGREE OF MASTER OF MEDICINE (RADIOLOGY)

UNIVERSITI SAINS MALAYSIA

2021

ACKNOWLEDGEMENT

First and foremost, I would like to express my gratitude to Almighty Allah SWT. The completion of this thesis would not have been possible without His blessings and grace. The completion of this thesis would also not have been possible without the participation, cooperation, and assistance from so many people whose names may not all be enumerated. Their contribution is sincerely appreciated and gratefully acknowledged. However, I would like to express my deepest and utmost appreciation particularly to the following:

1. Prof Madya. Dr. Juhara Haron, my supervisor, who provided continuous assistance and guidance in the completion of this thesis.
2. Dr. Chandran a/l Nadarajan and Dr. Wan Faiziah bt Wan Abdul Rahman (my co-supervisors) who always supported and guided me in completing the thesis.
3. Prof Madya. Dr. Mohd Shafie Abdullah, Prof Madya. Dr. Mohd Ezane Aziz, Dr. Khairil Amir Sayuti, Dr. Ahmad Tarmizi Musa, Dr. Wan Aireene Wan Ahmed, Dr. Ahmad Hadif Zaidin Samsudin, Dr. Nur Asma Sapiai, Dr. Nasibah Mohamad, and to all of whom directly or indirectly contributed their ideas, comments, and support in the completion of this study.
4. Colleagues and all the staff at the Department of Radiology and BestARi Hospital USM.
5. My beloved parents, Dr Ismail bin Daut, Puan Saripah Khazaimah and my family.
6. Mr. Samsudin and Puan Yazmin from the record office who helped in tracing all the data that I needed.

DISCLAIMER

I declare that this dissertation records the results of the study performed by me and that it is of my own composition.

……

(SITI FITRIWATI KHAZIS BINTI ISMAIL)

Date:

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

LIST OF SYMBOLS, ABBREVIATIONS, AND ACRONYMS

BRCA 1	Breast cancer gene 1
BRCA 2	Breast cancer gene 2
BestARi	Breast cancer awareness and research unit
Bi-RADS	Breast Imaging-Report and Data System
CC	Craniocaudally
ER	Oestrogen receptor
HER2	Human Epidermal Growth Factor Receptor 2
HPE	Histopathology Examination
USM	Universiti Sains Malaysia
LIS	Laboratory Information System
MLO	Mediolateral Oblique
PACS	Picture Archiving and Communication System
PR	Progesterone Receptor

ABSTRAK

Pengenalan: Kanser payuudara juga di kenali sebagai barah yang sering berlaku pada kumpulan wanita di Malaysia. Pelbagai faktor yang dikenali sebagai risiko untuk mendapat kanser payudara seperti umur, sejarah keluarga dengan kanser payudara, kepadatan tisu payudara di mammogram dan subjenis molekul kanser payudara. Pengimejan mammogram adalah piawai emas untuk kanser payudara.

Melalui pengimejan mammogram, kepadatan payudara dapat dikenalpasti. Pengenalpastian subjenis molekul ini akan membantu dalam pemilihan kemoterapi yang lebih tepat untuk pesakit. Subjenis molekul ditentukan berdasarkan ekspresi reseptor hormon dari tisu payudara yang terjejas. Ekspresi reseptor hormon dari tisu payudara yang terjejas mungkin dipengaruhi oleh faktor risiko yang serupa untuk kanser payudara. Peranan untuk menentukan subjenis molekul serta faktor risiko yang terlibat adalah penting untuk mencapai keberkesanan dalam menangani penyakit ini.

Metodologi: Kajian ini dilakukan berdasarkan kajian keratin rentas di kalangan pesakit yang telah di kenal pasti menhidapi barah payudara di Hospital USM. Untuk menjalankan kajian, ini kami telah memilih pesakit yang berusia 18 tahun ke atas, hasil mamogram and subjenis molekul. Mamogram terbahagi kepada padat dan tidak padat, sementara subjenis molekul dibahagikan kepada negatif bukan tiga dan negatif tiga. Kami juga mengambil kira parameter lain, seperti umur dan sejarah keluarga yang menghidapi barah payuudara. Bahagian payudara tidak padat dan padat di kalangan pesakit barah payudara digolongkan sebagai analisis deskriptif. Hubungan antara ketumpatan mamografi, usia, sejarah keluarga barah payudara dan subjenis molekul diuji menggunakan' multiple logistic regression'.

Keputusan: Seramai 280 peserta telah mengambil bahagian dalam kajian ini, hampir semua adalah orang Melayu (n = 248,98.6%) dan umur majoriti pesakit adalah lebih dari 50 tahun (n = 220,78.6%). Lebih separuh daripada 280 peserta mempunyai ketumpatan payudara yang tidak padat (n = 159,56.8%). Tiga perempat daripada 280 peserta diklasifikasikan dalam kes negatif tiga kali ganda (n = 247,87.5%). Ini, sama dengan sejarah keluarga barah payudara, iaitu tiga perempat daripada jumlah keseluruhan 280 peserta tidak mempunyai sejarah keluarga barah payudara (n = 246,87.9%). Regresi logistik berganda tidak menunjukkan hubungan yang signifikan antara subjenis molekul kanser payudara dan kepadatan payudara, usia dan sejarah keluarga barah payudara.

Kesimpulan: Kajian kami menunjukkan bahawa faktor-faktor yang berkaitan seperti kepadatan payudara mamogram, usia pesakit dan sejarah keluarga barah payudara tidak mempunyai kaitan dengan subjenis molekul tisu payudara. Ini menunjukkan bahawa faktor-faktor ini tidak mempengaruhi patofisiologi subjenis molekul.

Kata kunci: Barah payudara, Mammogram, subjenis molekular, faktor risiko.

ABSTRACT

Introduction. Breast cancer is one of the commonest cancers among females. Multiple risk factors such as age, family history, mammogram density, and molecular subtypes are associated with the formation of breast carcinoma. Each factor causes the carcinogenesis of breast cancer in different mechanisms. Mammogram is one of the gold imaging modalities to diagnose breast cancer. Based on the imaging, a woman's breasts can be categorised as non-dense and dense parenchyma. There is a clinically proven association between mammogram parenchyma density and breast cancer development. In addition, the prognosis of breast cancer is linked to the molecular subtypes, namely the "non-triple-negative" and "triple-negative" subtype depending on type of hormonal receptor that detected. To date, the association between molecular subtypes and the major risk factors of breast cancer is still poorly defined, especially among the Malaysian population. Therefore, the aim of this study was to assess the association between molecular subtypes with the mammogram density and other risk factors.

Patients and methods. A cross-sectional study involving breast cancer patients of 18 years and above was conducted in the Hospital USM. Their mammogram and molecular subtype results were obtained. The mammogram was categorised as non-dense and dense while the molecular subtypes were divided into non-triple-negative and triple-negative. Other variables collected in this study were the patients' age and family history of breast cancer. The proportion of non-dense and dense breasts among the patients was presented as descriptive analysis. The association between mammographic density, age, family history of breast cancer, and molecular subtype was tested using multiple logistic regression.

Results. There were 280 participants in this study. The majority of them were Malays (n = 248, 98.6%) and aged 50 years and above (n = 220, 78.6%). Slightly more than half of them (n = 159, 56.8%) had non-dense breast. A high number of them reported no family history of breast cancer, (n = 246, 87.9%). Multiple logistic regression showed that the molecular subtypes of breast cancer was not significantly related to breast density, age, and family history of breast cancer.

Conclusions. Based on the findings, breast density, age, and family history were not directly related to the molecular subtypes of breast cancer. Nonetheless, a case-control study with a larger sample size is recommended to obtain a more comprehensive understanding of the relationship between breast density, age, and family history with the molecular subtype of breast tissue.

Keywords. Breast cancer, Mammogram, Molecular subtypes, risk factors

CHAPTER 1: BACKGROUND

1.1 Introduction.

Breast cancer is the commonest type of cancer affecting Malaysian women after cervical cancer. Based on the Malaysian National Cancer Registry in 2007-2011, breast cancer topped the list with a total of 18,343 newly diagnosed cases during the period that accounted for 17% of the total 103,507 newly diagnosed cancer patients nationwide (Ab, Saleha and Hashimah, 2007).

As one of the major health problems and the top leading cancer-related causes of death among the worldwide population, early detection and timely management are crucial ('WHO Breast cancer', 2018). Efficient data collection and proper analysis via a comprehensive cancer registry are important to provide clear epidemiological and clinical information regarding breast cancer (Shukla et al., 2017).

Multiple risk factors have been linked with the development of breast cancer. They included non-modifiable factors such as age, family history, age of menarche and menopause, parity status, duration of lactation, infertility or subfertility, dietary intake, anthropometry, and breast tissue component (Shukla et al., 2017). Apart from that, modifiable risk factors such as the lifestyle of the patients also predispose to breast cancer. Together, all these factors play a role in the carcinogenesis pathway of breast cancer (Sung et al., 2010).

Most common and accurately assessable history of non-modifiable risk factors such as age and family history were chosen. Besides being a risk factor for developing breast cancer, age also affects the prognosis of breast cancer (Lee et al., 2015). In addition, those with a background

of family members with breast cancer, regardless of primary or secondary relatives, are predisposed to breast cancer (Jiang et al., 2012). Age and family history were established as the most common risk factors of breast cancer (Yip, Pathy and Teo, 2014). Patients can provide relatively accurate history regarding age and family history (Kluttig and Schmidt-Pokrzywniak, 2009). A recent study by Paige et al. shows that patients with non-modifiable risk factors accounted for the majority of the preventable breast cancer cases (Maas *et al.*, 2016). However, these two risk factors were non-modifiable in nature. They were also not directly related to any external exposure that can be targeted in the effort to reduce breast cancer (Maas *et al.*, 2016).

In addition, breast tissue composition also affects the risk of developing breast cancer. Breast tissues consist of adipose tissue and fibro glandular tissue (Javed and Lteif, 2013). Prolonged exposure to oestrogen hormone during the development of breast is associated with an elevated risk of developing breast cancer (Dall and Britt, 2017). The pathophysiology of the relationship between breast density and breast cancer is still poorly understood. However, a study by Lokate (Lokate et al., 2011) highlighted the positive association between breast density with carcinogenesis depending on the histological type of breast cancer cells. Breast density is based on the ratio of the amount between epithelial and stromal cells. Breast cancer is predominantly composed of epithelial cells. Thus, the larger the number of epithelial cells or the denser the breast, the greater the risk is to develop breast cancer (Pettersson et al., 2014). Furthermore, a higher breast density can decrease the sensitivity of the detection of a non-dominant lesion on mammogram (Nazari and Mukherjee, 2018).

Breast imaging can be performed via ultrasound or digital mammography or a combination of both. Ideally, mammography is the best modality to assess breast tissue composition. Currently,

there is a new technology that combines digital mammography with tomosynthesis. It helps to determine the pattern of breast density and provides important information to guide the decision-making for further management (Yang et al., 2013).

There are various ways to interpret breast density on mammography based on quantitative and qualitative measurements, such as those suggested by Wolfe, Boyd, and the TABAR or Bi-RADS classifications (Garrido-Estepa et al., 2010). Each of these classifications has it owns pros and cons. The Breast Imaging Reporting and Data System (Bi-RADS) classification developed by the American College of Radiology in 2016 was the main tool used in the Hospital Universiti Sains Malaysia (HUSM) to classify breast density. It is also the most widely used classification method domestically and internationally.

Based on the latest fifth edition of the Bi-RADS lexicon, the classification of breast density is categorised into four groups, namely Composition A, B, C, and D. Composition A is almost entirely fatty breast tissue while Composition B shows scattered areas of fibro glandular density. Compositions C and D refer to heterogeneously dense and extremely dense breast tissues (Rao et al., 2016). Composition A and B are categorised as non-dense breasts while composition C and D refer to dense breasts (Irshad et al., 2016). Most of the studies evaluated breast density based on the Volpara density grade category in the mammogram report by calculating the volume of fibro glandular breast tissue (Kato et al., 1995). However, in HUSM, the Bi-RADS classification was applied in the mammogram report.

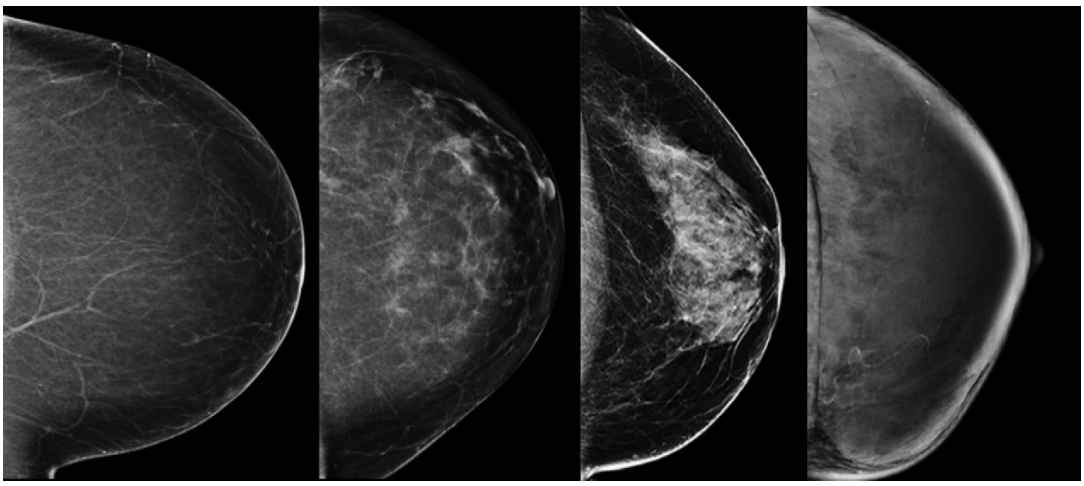

Figure 1 (a-d): Craniocaudal view of mammogram images show almost entirely fatty breast tissues, scattered area of fibro glandular breast tissues, heterogeneously dense, and extremely dense breast (from left to right) (Rao et al., 2016).

In the literature, the expression of hormone receptors such as oestrogen receptor (ER), progesterone receptor (PR), HER2 receptor, and basal subtype on breast cancer tissues have been widely assessed in histopathology examination. The expression of these hormone receptors is vital in the development, management, and prognosis of breast cancer (Fragomeni, Sciallis and Jeruss, 2018). As such, an in-depth understanding of their roles is important in planning an effective and targeted treatment for breast cancer (Pourzand et al., 2011). The affected breast tissues can be obtained either by ultrasound-guided biopsy or invasive surgical procedure (Nadji et al., 2005).

ER is a major player in normal breast development. On the contrary, PR is important in breast cancer development, especially in postmenopausal women. The growth and proliferation of breast cells are controlled by the HER2 gene. HER2 gene amplification leads to the overexpression of HER2. Therefore, the HER2 receptor gene test is compulsory in all cases of

invasive breast carcinoma (Fragomeni, Sciallis and Jeruss, 2018). Specifically, the molecular subtype of breast cancer can be divided into non-triple-negative, triple-negative, or basal-like subtypes. ER-positive, PR-positive, and HER2-positive breast cancer patients are classified as having the non-triple-negative subtype. These subtypes were classified into one category in view of the similar plan of management. To date, hormonal treatment for specific management targeting a single hormonal receptor expression has been established (Gonçalves Jr *et al.*, 2018a). On the other hand, the absence of these receptors indicates triple-negative subtype. Triple-negative breast cancer is a group of distinct tumours that display different clinical and biological factors (Bhatti *et al.*, 2014; Gonçalves Jr *et al.*, 2018a). The characteristics of these tumour subtypes are important in determining the treatment and prognosis of breast cancer. Each molecular subtype of the tumour calls for tailored management, either surgically or empirically (Gonçalves Jr *et al.*, 2018a). These subtypes are important in determining the type of treatment and the prognosis of breast cancer (Gonçalves Jr et al., 2018a). However, the majority of the studies that investigated the link between breast density, risk factors, and hormonal expression of breast cancer were conducted among the Caucasian population. There is a lack of studies among the Asian population. In view of the important roles of these molecular subtypes on breast tissue development, treatment, and prognosis of breast cancer, this study aimed to assess the relationship between the molecular subtypes and non-modifiable risk factors of breast cancer patients in HUSM.

1.2 Objectives

1.2.1 General Objective

To examine the association between mammographic breast density and non-modifiable risk factors of breast cancer such as age and family history among patients of different molecular subtypes of breast cancer.

1.2.2 Specific Objective

1 To determine the proportion of dense and non-dense breasts among breast cancer patients in HUSM.

2 To determine the association between mammographic density and molecular subtypes among breast cancer patients.

3 To determine the association between age and family history with molecular subtypes among breast cancer patients.

1.3 Hypothesis

There is a higher number of breast cancer patients who have dense breasts. An association between mammographic density and molecular breast cancer subtypes was recognised among breast cancer patients. Association between non-modifiable risk factors such as age and family history with the molecular subtypes also present among breast cancer patients.

1.4 Research question

1. What is the distribution of mammography density among breast cancer patients in HUSM?
2. Is there any association between mammographic density and molecular breast cancer subtypes among breast cancer patients in HUSM?
3. Is there any association between non-modifiable risk factors such as age and family history with the molecular subtypes among breast cancer patients in HUSM?

CHAPTER 2*:* LITERATURE REVIEW

2.1. Breast anatomy & physiology

The human breast is a modified sebaceous gland that arises from the ectodermal derivative. The left and right breasts are almost symmetrical bilaterally. The breast is located at the anterior chest wall, between the third and seventh ribs (De Benedetto et al., 2016). Breast tissues composed of parenchymal and stromal elements (Javed and Lteif, 2013). Exteriorly, it is covered by the skin, nipple, subcutaneous tissues at the anteroposterior aspect, and the pectoralis major muscle posteriorly (Irigo, Coscarelli and Rancati, 2017).

The mammary glands form the internal aspect of the breast. Mammary glands are made up of glandular, adipose, and fibrous tissues. The glandular tissue of the parenchyma consists of 15-20 lobes with alveolar ducts. The alveolar ducts converge into the lobular ducts before forming the main duct. Another important structure of the breast is the stroma that is made up of dense fibrous and adipose tissues. The stroma consists of three parts, namely the subcutaneous part (between the skin and the gland), intraparenchymal part (the area consisting of lobes and lobules), and the retro-mammary part (the area behind the gland) (De Benedetto et al., 2016).

The human breast undergoes different phases of development due to hormonal interaction. The development starts from in vitro until menopause (Gusterson and Stein, 2012). During these phases of development, the breast tissue composition undergoes remodelling as a result of the significant hormonal influence exerted during pregnancy and lactation (Macias and Hinck, 2012).

The evolution of breast tissue begins with the formation of the secretory gland in early puberty. The formation of multiple lobules and the growth of the glandular ducts are influenced by oestrogen, progesterone, and prolactin. The secretion of these hormones fluctuates especially during pregnancy and lactation. During puberty and early adulthood, the breast components are composed of predominantly fibro glandular tissues. However, later in life during the menopause phase, the elasticity of breast tissue reduces as a result of the decreased level of natural oestrogen. Eventually, the fibro glandular tissues will regress and be replaced by adipose tissues (Javed and Lteif, 2013). On the contrary, prolonged exposure to oestrogen such as hormonal therapy throughout the stages of breast development could lead to an elevated risk of breast cancer development (Dall and Britt, 2017).

The breast tissues that consist of adipose tissues and fibro glandular tissues can be examined via imaging modalities such as mammograms. The percentage of fibro glandular tissue in the breast will determine the percentage of breast density shown in mammograms. Multiple studies on breast density have been conducted to determine its influence on breast cancer pathophysiology (Zulfiqar, Rohazly and Rahmah, 2011; Sartor et al., 2016; Heller et al., 2018). Breast density can be influenced by sociodemographic factors such as age, parity, and lactation history as well as clinical factors such as body mass index, anovulatory cycle, history of taking hormone replacement therapy such as oestrogen and progesterone-based pill, history of breast surgery (Heller et al., 2018).

2.2 Breast cancer molecular subtypes

Breast carcinoma is a heterogeneous disease composed of many biologically distinct entities. The molecular subtypes of breast cancer are categorised based on the gene expression that is validated via an immunohistochemical (IHC) surrogate panel. The IHC surrogate panel includes oestrogen receptor (ER), progesterone receptor (PR), and human epidermal growth factor receptor 2 (HER2). There are four well-established molecular subtypes, namely Luminal A (ER +, PR+/-, HER2 -), luminal B (ER +, PR +/-, HER +/-), and HER2 overexpression and basal-like (Cell, F., & To et al., 2013; El Chediak et al., 2017; Lian et al., 2017). However, a study by Goncalves et al it divides further molecular subtypes into two categories. ER-positive, PR-positive, and HER2 overexpression are classified as "non-triple-negative" group. These subtypes were classified into one category in view of the similar plan of management. To date, no specific management targeting a single hormonal receptor expression has been established (Gonçalves Jr *et al.*, 2018a). On the other hand, the absence of these receptors indicates triple-negative subtype. Any breast cancer tissues with negative ER, PR, and HER2 receptors would be categorised under the triple-negative group. It is a group of distinct tumours that display different clinical and biological factors (Bhatti *et al.*, 2014; Gonçalves Jr *et al.*, 2018a). The affected breast tissues can be obtained either by ultrasound-guided biopsy or invasive surgical procedure. These tissues should be kept in formalin and embedded with paraffin before being cut into a smaller size. The small size tissue must be heated up overnight before undergoing immunohistochemistry (IHC) staining the next day. Monoclonal antibody 1D5 is used for oestrogen receptor (ER) whereas monoclonal anti-PR antibody 636 is used for progesterone receptor (PR) (Nadji et al., 2005). As for human epidermal growth factor receptor 2 (HER2) expression, monoclonal antibody 4B5 is used as the staining solution (Yim et al., 2019).

The expression of each of the main hormonal receptor has a different role to play in breast development. For example, ER is involved in normal breast development. Indirectly, ER is also involved in breast cancer pathophysiology as it induces cancer cell growth (Han et al., 2018). On the other hand, PR is important in breast cancer development, especially in postmenopausal women. Meanwhile, HER2 (c-erbB-2) is an oncogene under the family of epidermal growth factor receptors. The growth and proliferation of breast cells are controlled by the HER2 gene. Overexpression of HER2 is a result of HER2 gene amplification. Invasive breast cancer with HER2 positive is considered as high-grade tumours but they showed good response with HER2 targeted therapy such as Trastuzumab (Rosai and Ackerman's Surgical Pathology - 2 Volume Set - 11th Edition). Therefore, the HER2 receptor gene test is compulsory in all cases of invasive breast carcinoma (Fragomeni, Sciallis, and Jeruss, 2018).

In addition, the relationship between these hormonal receptors and mammographic density during different phases of breast development has been widely researched upon. Majority of the studies have shown positive relationship between hormonal receptors and mammographic density especially those with positive estrogen and progesterone hormones(Conroy *et al.*, 2011a; Yaghjyan *et al.*, 2015). However study by Colditz in 2004, negative association between hormonal receptor and mammographic density was observed This could be possibly attributed to environmental factors such as body mass index and age that can affect breast density indirectly (Vachon *et al.*, 2000; Colditz *et al.*, 2004). The advancement in the hormonal receptor markers of ER and PR has greatly facilitated breast cancer treatment. The ER and PR positivity refers to at least 1% of nuclear positivity on IHC staining (Rosai and Ackerman's Surgical Pathology - 2 Volume Set - 11th Edition, no date).

2.3 Breast imaging & Mammographic density

Breast cancer screening is a proven method that is effective in providing valuable data for the diagnosis and prognosis of the condition. Routine breast screening can lower the mortality risk of breast cancer (Godavarty et al., 2015). There are three main methods of breast cancer screening, which includes breast self-examination (BSE), clinical breast examination (CBE), and mammography. BSE and CBE can help to detect breast cancer at the earliest symptomatic stage whereas mammography helps to detect early lesions at the curable stage (Teh et al., 2015). Besides mammography, other breast imaging modalities include ultrasonography, computed tomography, magnetic resonance imaging, positron emission tomography, and single-photon emission computed tomography (Jafari et al., 2018).

2.4. Breast cancer risk

There are many risk factors of breast cancer, one of it being breast density (Freer, 2015). A study showed an elevated risk of developing breast cancer when more than 75% area in a mammogram was composed of dense tissue (Byrne, Schairer and Wolfe, 1995). This could be attributed to the fact that breast cancer predominantly arises from epithelial cells. As previously outlined, breast tissue is formed by epithelial and stromal cells whereas breast density is quantified by the ratio between the epithelial and stromal cells in the breast tissue. The higher the number of epithelial cells, the higher the risk of breast cancer. In other words, the percentage of breast density is a good risk predictor of breast cancer development (Pettersson and Tamimi, 2012). An earlier study reported that women with dense breasts recorded a four-fold higher risk of breast cancer (N F Boyd et al., 1998). The same study also concluded that mammographic breast density is among the strongest risk factors of breast cancer in the study

population. Another study established a linear association between quantitative measures of breast density and risk of breast cancer (McCormack and dos Santos Silva, 2006). 'Triple test' refers to the combination of mammography and ultrasonography in breast cancer screening as well as the fine needle aspiration cytology (FNAC) of the tissues in the breast lump. Triple test is the standard diagnostic modality of breast cancer and it has been associated with 100% sensitivity (Kharkwal, Sameer, and Mukherjee, 2014).

Apart from breast density, many other modifiable and non-modifiable risk factors can contribute indirectly to the risk of developing breast cancer. Exposure to hormonal therapy, high BMI, alcohol usage, and sedentary lifestyle are modifiable risk factors. On the other hand, non-modifiable risk factors include age, marital status, family history of breast cancer, parity, menopause and menarche status, and history of lactation. Among all the non-modifiable risk factors of breast cancer, age and family history are globally established as the most common risk factors (Maas et al., 2016). Patients can often provide relatively accurate history regarding age and family history (Kluttig and Schmidt-Pokrzywniak, 2009). Having a first-degree relative with a history of breast cancer is a well-established risk factor for breast cancer (Brewer et al., 2017). A recent study by Paige et al. shows that patients with non-modifiable risk factors accounted for the majority of the preventable breast cancer cases (Maas *et al.*, 2016). However, these two risk factors were non-modifiable in nature. They were also not directly related to any external exposure that can be targeted in the effort to reduce breast cancer (Maas *et al.*, 2016).

Age is also another prominent independent risk factor for breast cancer. Generally, women of older age have a higher risk of developing breast cancer. However, in a recent study, some young patients were also shown to develop breast cancer (Pourzand et al., 2011), depending on the type of hormone receptors detected within the cancer cells. Another study highlighted that

breast cancer patients of younger age often showed a positive hormone receptor expression that was associated with the type of breast cancer with unfavourable prognosis (Lee et al., 2015). Furthermore, evidence showed that elderly women with breast cancer often suffer from poorer outcome due to a delay in diagnosis and treatment (Tesarova, 2012). In general, the pathogenesis of breast cancer is highly complex. Therefore, it is important to understand the epidemiology of breast cancer to facilitate early detection and better prognosis (Assi et al., 2013).

Among the various modalities, ultrasonography can be used to identify and measure the size of the breast lesion. It can also be used for further intervention such as ultrasound-guided biopsy for the collection of breast tissue samples for further analysis. However, ultrasonography is unable to measure the density of the breast tissues precisely (Madjar, 2010).

In comparison, the mammogram can provide a better representation of breast density. The breast density is calculated based on the amount of radio-opaque epithelial and stromal tissues compared with the radiolucent adipose tissue. Therefore, a mammogram is the imaging modality of choice to assess breast density (Assi et al., 2011). Leborgne first published a study on the measurement of breast density with a mammogram in 1953. Further classification of breast density was subsequently proposed by Wolfe in 1976 (Freer, 2015).

Currently, the measurement of breast density follows the fifth edition of the Breast Imaging Reporting and Data System (Bi-RADS) lexicon (Rao et al., 2016). In this edition, the breast density is categorised into four groups. Composition A is almost entirely fatty breast tissue while Composition B shows scattered areas of fibro glandular density. Compositions C and D refer to heterogeneously dense and extremely dense breast tissues (Rao et al., 2016).

Furthermore, breast parenchyma with composition A and B are categorised as non-dense breasts while composition C and D refer to dense breasts. In short, Bi-RADS provides an accurate measurement of mammography density. Therefore, it is commonly used by clinicians in the determination of the breast cancer risk to assist decision-making (Irshad et al., 2016).

2.5 Rationale of the study

In view of the alarming incidence of breast cancer worldwide, in-depth study on the risk factors of breast cancer is warranted. Among the various risk factors, breast density is one of the established risk factors of breast cancer development. In addition, molecular subtypes, age, and family history should also be considered in research regarding risk factors of breast cancer development.

This study is proposed to analyse the pattern of breast cancer patient's profile and risk factors at HUSM which not clearly studied before. Furthermore, we want to determine any association between mammogram breast density and non-modifiable risk factors (i.e., age, family history) with molecular subtypes in breast cancer patient. Among all non-modifiable risk factors, this study focused on age and family history. These factors known to be the most common causes with accurately assessable history from the patient. Hypotheses suggest that a positive association existed between mammographic breast density, molecular subtypes, socio-demographic factors (i.e., age, family history), and the risk of developing breast cancer. By determining specific risk factors-subtype of breast cancer, this study also highlights the importance of identifying these factors to facilitate the early prevention of breast cancer by raise awareness among patient and practitioner. Furthermore, for better understanding the risk factors-subtype of breast cancer eventually develop effective preventive strategies.

2.6 Conceptual framework

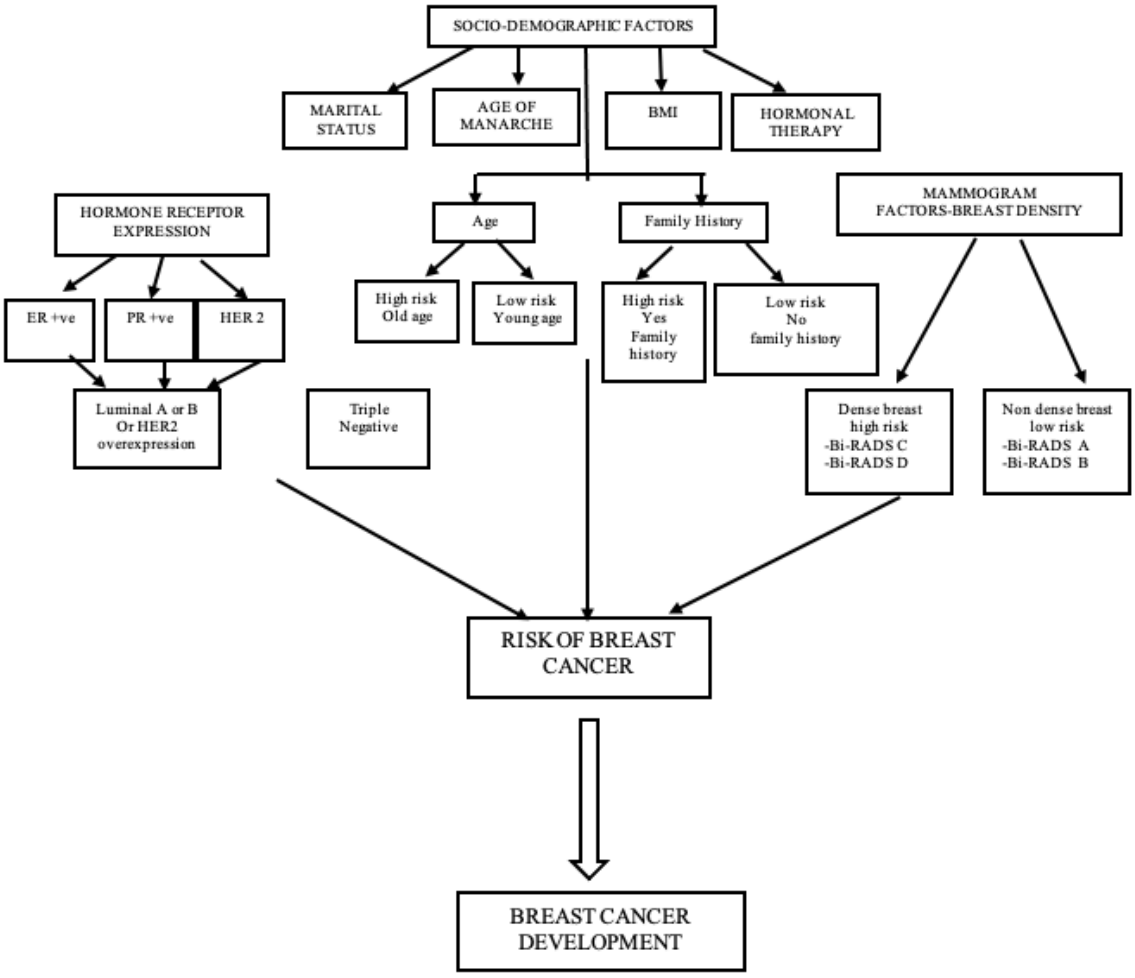

Figure 2: Conceptual framework of the risk factors of breast cancer.

The conceptual framework shows that multiple risk factors such as hormone receptor expression, breast density, and age of the patient are associated with breast cancer development.

CHAPTER 3: METHODOLOGY

3.1 Study Design

This is a retrospective cross-sectional study using secondary data from 1st January 2010 until 31st of March 2021.

3.2 Study Location and Duration

The study was conducted in Hospital USM from 1st January 2020 until 31st of March 2021.

3.3 Study population and sample

All patients with confirmed breast cancer who attended Hospital USM.

3.4 Sampling Technique

No sampling method was applied. All eligible patients that fulfilled the inclusion criteria were included in the study.

3.5 Inclusion Criteria

1. Female adults aged 18 years and above.
2. Underwent Mammogram.
3. HPE-confirmed breast cancer with known molecular subtypes from IHC staining for ER, PR, and HER 2.

3.6 Exclusion Criteria

1. Patient with a breast implant.
2. Recurrent breast cancer.
3. Case without ER, PR, and HER 2 results in previous HPE.

3.7 Sample Size Calculation

The sample sizes for Objectives 1 to 3 were calculated using the Samples size Calculator by Dr Wan Ariffin software.

For Objective 1, the sample size was calculated based on a previous study (Zulfiqar, Rohazly and Rahmah, 2011). Based on a 0.54 proportion of non-dense breast, a precision of 0.06, a confidence level of 95%, and an anticipated dropout rate of 10%, the calculated sample size was 288 (Appendix C).

For Objective 2, the sample size was calculated based on a recent study (Arora et al., 2010). The proportion of dense breasts in the triple-negative group was taken as control (P0) and the proportion of dense breasts in the Luminal A group was taken as the case (P1). For P0, the value was 0.14 whereas, for P1, the value was 0.50. Based on a significance level of 0.06, power of 85%, and the total dropout rate of 10%, the sample size needed was 64 (Appendix C).

For Objective 3, two sample size calculations were performed based on age and family history. Based on a 2017 study (El Chediak et al., 2017), the proportion of exposed (age) in the triple-negative group was taken as control (P0) and the proportion of exposed (age) in Luminal A

was taken as the case (P1). For P0, the value was 0.16 and for P1, the value was 0.40. With a significance level of 0.05, power of 85%, and the total dropout rate of 10%, the sample size needed was 138.

For family history, based on study in 2011 (Gaudet et al., 2011), the proportion of exposed (family history) in the triple-negative group taken as control (P0) and the proportion of exposed (family history) in the Luminal A group was taken as the case (P1). For P0, the value was 0.35 and for P1, the value was 0.60. With a significance level of 0.05, power of 85%, and total dropout rate of 10%, the sample size needed was 158.

From all the sample size calculations, the largest sample size obtained was for Objective 1. Therefore, the total number of patients that needed to be recruited in this study was 288 patients. However, after excluded the exclusion criteria, only 280 patients are fit for analysed.

3.8 Research tools

1. Mammogram machine in the Women Health Imaging (WISH) Radiology Department of Hospital USM(HUSM).

Brand: Hologic Selenia Dimensions, United State of America

Serial Number: 81001143041.

Started using in 2014.

Workstation: HOLOGIC SecurView Workstation.

2. Picture Archive Communication System (PACS) in Hospital USM: Centricity PACS Version 6.0 SP GE Healthcare.

3.9 Variable definitions

Mammography density: The classification of breast composition density was based on the American College of Radiology Bi-RADS Lexicon 5th edition (Rao *et al.*, 2016) as explained in the introduction. The subjects were categorised as having non-dense and dense breast. Bi-RADS type A and B were grouped as 'non-dense' while Bi-RADS type C and D were grouped as 'dense' (Irshad *et al.*, 2016).

Age at first diagnosed with breast cancer: This variable was divided into two groups, i.e. Those less than or equal to 50 years old and those above 50 years (Yip, Pathy, and Teo, 2014).

Family history: This variable was divided as YES or NO, regardless of whether the family members were first or second-degree relative (El Chediak *et al.*, 2017) First-degree relatives included were parents, brothers, sisters, and children. Meanwhile, second-degree relatives referred to aunts, uncles, nieces, nephews, and grandparents (Jiang *et al.*, 2012)

Molecular subtype: The molecular subtypes of breast cancer are categorised as non-triple-negative (Luminal A, Luminal B, and HER2 overexpression) and triple-negative (negative ER, PR, and HER2) (Gonçalves Jr *et al.*, 2018b).

3.10 Data Collection

Patient cohort

This is retrospective study using secondary data. It is conducted in Hospital USM using data from 1st January 2010 until 31st March 2021 from Breast Cancer Awareness and Research Unit (BestARI) Hospital USM. We included breast cancer female patients who were 18 years old and above, underwent mammogram with available HPE breast tissue result. Exclusion criteria of those who had breast implant, recurrent breast cancer and absent data about hormonal receptor expression from their HPE result.

The study was approved by the Human Research Ethics Committee of Universiti Sains Malaysia (JEPeM code: USM/JEPeM/20010042) and complied with the Declaration of Helsinki (see APPENDIX B).

Imaging data

The data of all the patients who underwent a mammogram from 2010 to 2021 were collected from the PACS (Picture Archiving and Communication System). For mammograms, the images taken in MLO (mediolateral oblique) and CC (craniocaudal) by Hologic Selenia Dimensions. Based on the retrieved mammograms, breast density was determined based on the Bi-RADS classification. All the breast mammography findings of patients in HUSM were reviewed, verified, and reported by credited senior breast radiologists. The density of breast composition was classified based on the Fifth Edition of the American College of Radiology Bi-RADS lexicon (Rao *et al.*, 2016).

Validation

A senior breast radiologist with 10 years' experience in breast radiology and researcher independently classified a random sample of 15% of the mammographic density based on latest BI-RADS classification Lexicon fifth edition. According to the two radiological methods and the inter-rater reliability were assessed using Kappa test.

Interrater reliability using Cohen's kappa analysis shows a good agreement between two-raters using BI-RADS classification of mammographic breast density with Cohen's kappa coefficient of 1.00 and 0.87 ($p < 0.001$) respectively. Significant p-value is ≤ 0.05. The data are shown in Table 4.

Patient Profile

All data such as sociodemographic factors were retrieved from the patient's medical folders.

Sociodemographic Factors

The age of the patient at first diagnosis of breast cancer and family history of breast cancer were obtained from the medical record of the patients. Family history of breast cancer were later divided into 'Yes' or 'No'.

Pathology data

Breast tissue sample was obtained by surgical procedure or ultrasound guided breast biopsy. The specimen was place in 10% neutral buffered formalin. Then it underwent immunohistochemistry stained. Based on the hormone receptor expression in the histopathology examination (HPE) of breast lesions, patients with confirmed breast cancer were divided into the molecular subtypes of Luminal A, Luminal B, or HER2 overexpression. These subtypes were classified as "non-triple-negative". On the contrary, breast tissues with negative ER, PR, and HER2 were classified as the "triple-negative" molecular subtype.

3.11 Image Analysis

The mammogram taken by Hologic Selenia dimension were reviewed in Hologic Securview Workstation. The mammogram density was classified base on fifth edition of the American College of Radiology Bi-RADS lexicon (Rao *et al.*, 2016) as mentioned in the introduction.

3.12 Statistical Analysis & Hypothesis

Statistical Analysis

Data was entered and analysed using SPSS version 24. Demographic data was described using descriptively with frequency and percentage. For Objectives 2 and 3, the data were analysed with Chi-square Test followed by logistic regression analysis. Firstly, simple logistic regression was performed, and the results of each independent variable were presented as crude OR. Then, multiple logistic regression analysis was performed by auto-selection of both

forward and backward methods. Only variables with p-value <0.25 or any clinically important factors were selected for multiple logistic regression. Collinearity and interaction between the significant variables as well as the goodness of fit of the model were checked. The result was presented as the adjusted OR.

Hypothesis

Breast cancer patients have denser breasts. There is an association between mammographic density and molecular breast cancer subtypes among breast cancer patients. There is an association between non-modifiable risk factors such as age and family history with the molecular subtypes among breast cancer patients.

3.13 Confidentiality and Privacy

All subjects were identified by serial number and the available data were kept anonymous. The SPSS data were stored in password protected computer which only can be access by researchers. All the data were stored in CDs upon completion of study. Later, the database on the computer were erased once completed study. The results were presented as grouped data and thus it would not reveal any individual identity of the patients. The data were retained by the researchers for knowledge purposes only. Neither the name nor any identifying information was used in any publication or presentation resulting from this study.