ANTÓNIO MIGUEL HENRIQUES DA CUNHA ABREU

# MALE BREAST CANCER: PROGNOSTIC AND PREDICTIVE FACTORS OF RESPONSE TO THERAPY

Thesis submitted for a Doctoral Degree in Medical Sciences to Instituto de Ciências Biomédicas Abel Salazar da Universidade do Porto.

**Supervised by**
Doutor Carlos Alberto da Silva Lopes.
Professor Catedrático Jubilado
Instituto de Ciências Biomédicas Abel Salazar da Universidade do Porto (ICBAS-UP).

**Co-supervised by**
Doutor Rui Manuel Ferreira Henrique.
Professor Catedrático Convidado
ICBAS-UP

**And**

Mestre Maria Deolinda Paulino Pereira Sousa Pereira.
Diretora do Serviço de Oncologia Médica
Instituto Português de Oncologia do Porto, Francisco Gentil (IPOPFG), EPE.

# Abstract

Male breast cancer (MBC) is an uncommon disease accounting for only 1% of all breast cancers. Despite its rarity, the mortality associated with MBC and the increase in its incidence in recent years, has led to a progressive interest on this subject. Retrospective studies have already demonstrated important differences between MBC and female breast cancer (FBC) but the knowledge about its biological behaviour remains insufficient to define a personalised treatment strategy to MBC patients.

The aim of the present work was to study MBC prognosis based on: the definition of better intrinsic subgroups based on immunoexpression, the analysis of recurrence and of factors that could predict response to adjuvant endocrine therapy.

From clinical records of 111 MBC patients treated in a tertiary cancer center (with a median follow-up of 5.5 years), information was selectively analysed to answer specific research questions. Central pathology revision was performed with additional immunohistochemical (IHC) analysis of the cases in which tumour blocks were available. Blood samples were collected from patients who were in active treatment or surveillance in the center for genetic analysis.

The independent prognostic factors associated with poor outcome were: tumour size larger than 2cm (HR: 1.8, 95% CI: 1.0-3.4 years, $p = 0.049$), the absence of expression of estrogen receptor (HR: 4.9, 95% CI: 1.7-14.3 years, $p = 0.004$) and stage IV disease (HR: 5.3, 95% CI: 2.2-3.1 years, $p < 0.001$). The presence of lymph node metastasis seems to be related with poor outcome in univariate analysis (HR: 1.8, 95% CI: 0.9-3.2 years, p=0.05). The two classifications used to define FBC subgroups (based on three and four IHC markers), could not be linearly translated to MBC patients as some groups are rarely represented among these patients (triple negative: 2.7%-3.2% and HER2 enriched non luminal $\leq$ 1%), and they lost their prognostic discriminatory capacity when analysed only the most frequent groups (luminals). Cluster analysis based on a six-IHC panel, identified two important prognostic subgroups with extreme outcomes: one with better prognosis, composed by patients with tumours that express estrogen and progesterone receptor without expression of androgen receptor and HER2 and low ki67/p53, and another with the worst survival, associated with patients that had tumours who did not express progesterone receptor.

Recurrence was specifically analysed to evaluate its impact in survival outcome, using two gender cohorts that were matched for age at diagnosis, histological grade, stage, type of tumour and adjuvant treatments performed. MBC patients relapse more often to lung ($p = 0.003$) and presented poorer survival than FBC and this result remains across all of the five groups defined by relapse pattern. The difference in outcome seems to be related with the absence of systemic palliative treatment in recurrence time that is more commonly registered in MBC patients (21.1% vs 4.4%, $p = 0.018$). Patients treated with adjuvant tamoxifen categorized as poor tamoxifen metabolizers based on CYP2D6*4 polymorphism had also a higher risk of recurrence ($p = 0.0034$) and this effect was still observed when controlled important prognostic factors like:

size $>$ 2cm ($p = 0.001$), nodal status (N0 vs N+, $p = 0.004$), and advanced disease (stage III versus others, $p < 0.001$). These patients were associated with worse survival when tumours were larger than 2cm.

In conclusion, MBC patients have poorer outcome than FBC patients, even when stratified by important prognostic factors. Clinicians must look with major concerns for MBC patients that present with tumours with more than 2cm, or have lymph nodes involved or do not express ER or PR or present with distant metastasis at diagnosis (stage IV). Poor tamoxifen metabolizers based on CYP2D6*4 polymorphism have higher risk for relapse and probably need a more aggressive adjuvant approach. Considering recurrence time, palliative systemic treatment had a favourable impact in MBC patients' survival.

# Resumo

O carcinoma da mama no homem (CMH) é uma doença rara, correspondendo a apenas 1% de todos os cancros da mama. Apesar da baixa incidência, o aumento progressivo do número de casos nos últimos anos associado à elevada taxa de mortalidade, tem despertado o interesse para esta patologia. Estudos retrospetivos demonstraram já diferenças importantes entre o CMH e o carcinoma da mama na mulher (CMM), mas o conhecimento sobre o seu comportamento biológico ainda é insuficiente para definir uma estratégia de tratamento personalizada para estes doentes.

O objetivo do presente trabalho foi estudar o prognóstico dos doentes do sexo masculino com carcinoma da mama baseado: na definição de melhores subgrupos de prognóstico a partir de marcadores da imunohistoquímica (IHQ), na análise da recorrência e de fatores preditivos de resposta à terapêutica endócrina adjuvante.

A partir dos registos clínicos de 111 doentes tratados em um centro oncológico (tempo de follow-up mediano de 5.5 anos), foi colhida e analisada a informação de forma a responder a perguntas de investigação específicas. Foi realizada revisão patológica central procedendo-se a análise IHQ adicional nos casos em que os blocos de tumor estavam disponíveis, sendo ainda colhidas amostras de sangue para análise genética nos doentes que se encontravam em tratamento ativo ou em vigilância no centro.

Os fatores independentes de prognóstico associados a pior sobrevivência foram: tamanho do tumor superior a 2cm (HR: 1.8, IC 95%: 1.0-3.4 anos, $p = 0.049$), a ausência de expressão do receptor de estrogénio (HR: 4.9, IC 95%: 1.7-14.3 anos, $p = 0.004$) e o estádio IV (HR: 5.3, IC 95%: 2.2-3.1 anos, $p < 0.001$). A metastização ganglionar locorregional parece estar relacionada com pior prognóstico na análise univariada (HR: 1.8, IC 95%: 0.9-3.2 anos, $p = 0.05$). As duas classificações utilizadas para definir subgrupos de prognóstico no CMM (com base em três ou quatro marcadores da IHQ), não podem ser linearmente transpostas para os homens com carcinoma da mama, já que alguns subgrupos são raramente descritos nestes doentes (triplos negativos: 2.7%-3.2% e *HER2 enriched* não luminais $\leq$1%), perdendo a sua capacidade discriminativa de prognóstico quando analisados apenas os subgrupos mais frequentes (luminais). A análise por *clusters*, utilizando um painel de seis marcadores da IHQ, identificou dois subgrupos de prognóstico com resultados extremos: um associado a melhor prognóstico, composto pelos doentes com tumores que expressavam receptores de estrogénio e de progesterona sem expressão do receptor de andrógenio nem HER2 e com ki67 / p53 baixo e outro que apresentava a pior sobrevivência, composto pelos doentes cujos tumores não expressavam o receptor de progesterona.

A análise da recorrência baseou-se na comparação de duas coortes definidas por sexo, que foram emparelhadas por idade ao diagnóstico, grau histológico, estádio, tipo de tumor e tratamentos adjuvantes realizados. Os homens com carcinoma da mama recorreram mais frequentemente para o pulmão ($p = 0.003$) e apresentaram pior sobrevivência do que as mulheres,

mantendo-se esta tendência, de forma transversal nos cinco grupos definidos com base no padrão de recorrência. Esta diferença parece relacionar-se com a ausência de tratamento paliativo sistémico no diagnóstico da recorrência que se verifica mais frequentemente nos doentes homens (21.1% vs 4.4%, $p = 0.018$). Aqueles categorizados como *poor metabolizers* do tamoxifeno, de acordo com o polimorfismo da CYP2D6*4, apresentaram, também, um maior risco para recorrência ($p = 0.0034$) mantendo-se este resultado quando controlados importantes fatores de prognóstico como: tamanho do tumor superior a 2 cm ($p = 0.001$), N *status* (N0 *vs* N+, $p = 0.004$) e estádio avançado (estádio III *vs* os outros, $p < 0.001$). Estes doentes apesentaram também pior sobrevivência quando os tumores eram maiores que 2 cm.

Em conclusão os homens com carcinoma da mama apresentam pior sobrevivência quando comparados com as suas congéneres femininas mesmo quando controlados os principais fatores de prognóstico. Na prática clinica, deverão ser considerados de maior risco para morte por carcinoma da mama os doentes que apresentam tumores com mais de 2cm, ou que tenham envolvimento ganglionar locorregional ou que não expressam receptor de estrogénio ou de progesterona ou que se diagnostiquem em estádio IV. Os categorizados como *poor metabolizers* relativamente ao tamoxifeno com base no polimorfismo da CYP2D6*4 têm maior risco de recorrência e provavelmente beneficiarão de uma abordagem adjuvante mais agressiva. A realização de tratamento paliativo sistémico na primeira recorrência tem um impacto positivo na sobrevivência destes doentes.

# Acknowledgments

This work was only possible with the contribution of many people, who helped me in several different ways during these years.

First of all, I would like to thank my Brother Pedro and my Parents who always believed in me and in my capacities to develop this work.

Secondly, I would like to express my gratitude to my supervisors Prof. Carlos Lopes, Prof. Rui Henrique and Dr. Deolinda Pereira, who supported this thesis at various levels and provided all the necessary conditions and protection for this project to grow.

I had also the fortune of having had the help of other excellent colleagues without which this work would not have been possible. I must thank Dr. Noémia Afonso, for her bright insights and profound knowledge in many fields that definitively shaped me as an oncologist. From pathology service, thanks to Dr. Francisco Menezes and Dr. Conceição Leal and from Molecular Oncology Group to Dr. Mónica Gomes for their specialized contributions to this work, guaranteeing its quality and bringing for this project the true spirit of the translational research.

# List of Publications and Communications (within the scope of this PhD)

## Publications in International peer-review journals

1. CYP2D6*4 Polymorphism: a new marker of response to hormonotherapy in male breast cancer?

   **Miguel Henriques Abreu**, Mónica Gomes, Francisco Menezes, Noémia Afonso, Pedro Henriques Abreu, Rui Medeiros, Deolinda Pereira, Carlos Lopes.

   The Breast, 2015; 24(4):481-6.

## Submitted Articles Waiting for Review

1. Patterns of recurrence and treatment in male breast cancer: a clue to prognosis?

   **Miguel Henriques Abreu**, Pedro Henriques Abreu, Noémia Afonso, Rui Henrique, Carlos Lopes.

2. Male breast cancer: looking for better prognostic subgroups.

   **Miguel Henriques Abreu**, Noémia Afonso, Pedro Henriques Abreu, Francisco Menezes, Paula Lopes, Rui Henrique, Deolinda Pereira, Carlos Lopes.

3. Predicting breast cancer recurrence using machine learning tecnhiques: a systematic review.

   Pedro Henriques Abreu, Miriam S. Santos, **Miguel Henriques Abreu**, Daniel Castro Silva, Bruno Andrade.

4. A context-free framework applied to healthcare environments.

   **Miguel Henriques Abreu**, Pedro Henriques Abreu, Daniel Castro Silva, Hugo Amaro.

## Abstracts Published in International peer-review journals

1. Male breast cancer: looking for prognostic subgroups.

   **Miguel Henriques Abreu**, Noémia Afonso, Pedro Henriques Abreu, Francisco Menezes, Paula Lopes, Rui Henrique, Deolinda Pereira, Carlos Lopes. Journal of Clinical Oncology, 2015; 33:5s (supplement; abst.e11562).

2. Male breast cancer: the experience of an oncological center.

   **Miguel Henriques Abreu**, Eduarda Matos, Noémia Afonso, Deolinda Pereira, Helena Rodrigues, Rui Henrique, Carlos Lopes. Annals of Oncology, 2012 (23), supplement 9.

## Oral Communications in Scientific Meetings

1. Carcinoma da mama no homem: um marcador preditivo de sensibilidade à terapêutica endócrina?

   **Miguel Henriques Abreu**, Noémia Afonso, Pedro Henriques Abreu, Rui Henrique, Deolinda Pereira, Carlos Lopes. Encontros da Primavera 2014. Évora (Portugal). $27^{th}$-$29^{th}$ March 2014.

## Poster Presentations in Scientific Meetings

1. Male breast cancer: looking for prognostic subgroups.

   **Miguel Henriques Abreu**, Noémia Afonso, Pedro Henriques Abreu, Francisco Menezes, Paula Lopes, Rui Henrique, Deolinda Pereira, Carlos Lopes. American Society of Clinical Oncology (ASCO) Congress. Chicago (USA). $29^{th}$ May-$2^{nd}$ June 2015.

2. Carcinoma da mama no homem: estudo retrospetivo multicêntrico.

   **Miguel Henriques Abreu**, Eduarda Matos, Francisco Menezes, Noémia Afonso, Ana Ferreira, Deolinda Pereira, Helena Rodrigues, Rui Henrique. Congresso Nacional de Senologia. Porto (Portugal). $17^{th}$-$19^{th}$ November 2012.

3. Male breast cancer: the experience of an oncological center.

   **Miguel Henriques Abreu**, Eduarda Matos, Noémia Afonso, Deolinda Pereira, Helena Rodrigues, Rui Henrique, Carlos Lopes. European Society of Medical Oncology (ESMO) Congress. Austria (Vienna). $28^{th}$ September-$2^{nd}$ October 2012.

## Award

1. Third award in Oral Communications, Encontros da Primavera 2014, Évora (Portugal).2014.

## Clinical Trial

1. Principal Investigator in Instituto Português de Oncologia do Porto, Francisco Gentil EPE (IPOPFG) in European Organisation for Research and Treatment of Cancer (EORTC) protocol 10085, including the QOL sub-study: Clinical and biological characterization of male breast cancer: an international EORTC, Breast International Group (BIG), Translational Breast Cancer Research Consortium (TBCRC) and North American Breast Cancer Group (NABCG) intergroup study.

# Contents

# List of Tables

# ACRONYMS

AI: Aromatase Inhibitor
AJCC: American Joint Committee on Cancer
AR: Androgen Receptor
BC: Breast Cancer
BCFI: Breast Cancer Free Interval
BCM: Breast Cancer Mortality
BCRec: Breast Cancer Recurrence
BCSS: Breast Cancer Specific-Survival
BIG: Breast International Group
CISUC: Center for Informatics and Systems of University of Coimbra
CM: Cancer Mortality
CT: Chemotherapy
CK: Cytokeratin
DDFS: Distant Disease-Free Survival
DFS: Disease-Free Survival
DRecFS: Distant Recurrence Free Survival
DSS: Disease Specific Survival
EORTC: European Organization for Research and Treatment of Cancer
EFS: Event-Free Survival
ER: Estrogen Receptor
ET: Endocrine Therapy
FBC: Female Breast Cancer
FU: Follow-Up
HE Slides: Hematoxylin Eosin Slides
HER2: Epidermal growth factor Receptor 2
HR status: Hormonal Receptors status
HR: Hazard Ratio
ICBAS-UP: Instituto de Ciências Biomédicas de Abel Salazar da Universidade do Porto
IHC: Immunohistochemistry
IPOPFG: Instituto Português de Oncologia do Porto, Francisco Gentil, EPE
M: Mortality
MBC: Male Breast Cancer
NABCG: North American Breast Cancer Group
N status: lymph Node status
OS: Overall Survival
PFS: Progression-Free Survival
PR: Progesterone Receptor
QOL: Quality Of Life

RecFS: Recurrence-Free Survival
Rec rates: Recurrence rates
RT: Radiotherapy
RFT: Recurrence-Free Time
SPSS: Statistical Package for Social Sciences
TBCRC: Translational Breast Cancer Research Consortium
TTP: Time To Progression
TTRec: Time To Recurrence
WHO: World Health Organization

# Chapter 1

# Introduction

Male breast cancer (MBC) is a rare disease, accounting for only 1% of all breast cancers (BC) in Portugal [1] and in the USA [2] and 0.5% in the Nordic countries [3]. Its incidence has risen over the past 25 years [4–7], from 1.0 per 100,000 in the late 1970s to about 1.2 per 100,000 in 2000-2004 [5], with an increasing interest on this subject, even though representing less than 0.5% of all cancer deaths in men annually [8].

Due to its rarity, no randomized trials have been conducted, with only one prospective study published [9] and most knowledge still come from retrospective studies that cover several decades [10], different geographic regions [11], and different patients' approaches. Because of that, treatment recommendations remain extrapolated from female breast cancer (FBC) trials, although there are already many differences between the two diseases [12, 13].

Male are often older at diagnosis (67 vs 62 years) [12–19], presented with a painless subareolar lump, mainly in left breast and rarely in the upper outer quadrant [13, 20] that may involve nipple retraction or bleeding [10, 21]. The majority of them have ductal carcinomas [22, 23] that express hormonal receptors (estrogen receptor, ER: 91-95% vs 76-78% and progesterone receptor, PR: 80-81% vs 67% in male and female respectively), with lobular tumours being less common in male [13, 24]. Black males seem to be more affected to this disease, compared to white (ratios from USA, 1 male: 100 female in whites and 1:70 in blacks), and black race is also associated with poor prognosis after adjust for clinical, demographic and treatment factors [12, 25].

About 15-20% of male with BC have a family history of the disease, compared to only 7% in general male population [13]. Genes involved in breast and ovarian cancer, BRCA1 and BRCA2, inherited in an autosomal dominant pattern, are though to account for 80% of multiple-case breast cancer families. Germline BRCA2 mutations have been reported in 4-14% of MBC patients, while BRCA1 mutations are less frequent, occurring in up to 4% of these patients [26–32]. The highest

prevalence of BRCA2 mutations in MBC patients is registered in Iceland where a founder mutation is present in 40% of them [33]. The lifetime risk of developing breast cancer associated with these mutations is inferior from that observed in female patients: 1-6% from BRCA1 and 7% from BRCA2 in female [34–36] and in the male population 0.1% [3, 35]. Male patients with BRCA2 mutations seem to have BC in younger age and may have poorer survival [37]. Other investigated genes that do not have a causal association with MBC disease are the androgen receptor gene, PTEN (Cowden's syndrome), mismatch repair genes (hMLH1), BRIP1 and RAD51C [38–43], with mixed data for PALB2, CHEK2 and CYP17 [29, 44–51]. Other risk factors related to MBC are conditions associated with imbalance between estrogen and androgen, such as the Klinefelter's syndrome (resulting in a 50-fold increased risk) [52, 53], testicular abnormalities that cause testosterone deficiency [52, 54], liver diseases [55], obesity [52, 56–58] and exogenous estrogen exposure [59, 60], previous exposure to chest wall irradiation, like in patients with past history of Hodgkin disease [25], history of breast trauma and nipple discharge [25, 54].

Having in mind these differences between MBC and FBC patients, studies in MBC setting are warranted to define the best approach for these patients in the era of personalized medicine.

## 1.1 Objectives and Organisation of the Thesis

Nowadays MBC treatments are still extrapolated from FBC standards although many studies had suggested that male patients present distinct aspects from females that could justify specific approaches.

The present work addresses prognostic and predictive factors of response to endocrine therapy, trying to define features with clinical relevance that could help in daily practice decisions.

The look for a better knowledge of this disease was assessed by specific objectives:

- The assessment of prognostic factors (related to patients, tumours and/or treatments performed);
- The evaluation of the relations between these prognostic factors;
- The identification of new patients' prognostic subgroups based on a routinely ImmunoHistoChemical (IHC)-panel;
- The characterization of recurrence as an important prognosis determinant;
- The assessment of the clinical relevance of CYP2D6*4 polymorphism in the efficacy of response to adjuvant endocrine treatment (tamoxifen).

Additionally, other topics were considered important for the present work and in the preparation of future developments:

- The construction of a flexible framework that allows clinicians to define the variables that are important in patients' characterization, enabling also basic data storage operations (insert, edit or remove data) and finally export these data to standard formats such as excel or SPSS;
- The identification of the better machine learning techniques to define patterns of BC recurrence.

According to the proposal and to integrate all the contents of this work, the thesis is organized in three main chapters that come out after an introduction chapter. Chapter two is divided in two sections dedicated to prognostic analysis and chapter three focuses on a potential predictive factor to response to adjuvant endocrine therapy (tamoxifen).

A brief summary of the contents by chapter, is presented below:

**Chapter 1** presents an overview about MBC disease, the description of PhD objectives, organization of the thesis and a brief dataset description.

**Chapter 2** is dedicated to prognostic analysis. In the first section, the principal prognostic factors in these patients are addressed and new BC subgroups based on a six- IHC panel are defined. In the second section, a recurrence analysis is described.

**Chapter 3** concerns to the role of CYP2D6*4 polymorphism as a potential clinical marker of response to tamoxifen.

These main chapters (2 and 3) start with a state of the art related to the analysed theme.

**Chapter 4** presents the final remarks and future perspectives, highlighting the principal findings of this work, and proposing future developments.

In order to perform a truly translational research that represents the holistic approach of cancer patients with the guarantee of high quality standards, this work was developed in partnership with Serviço de Anatomia Patológica do IPOPFG, Grupo de Oncologia Molecular IPOPFG and Laboratório CISUC, Departamento Engenharia Informática da Universidade de Coimbra.

## 1.2 Dataset: Brief Description

The work dataset comprised all MBC patients treated at Instituto Português de Oncologia do Porto, Francisco Gentil EPE (IPOPFG), during 1976-2014 (115 patients), with age over 18 years, histological confirmation of BC and minimum follow-up of one year.

The number of patients and the time periods analysed were fit according to the research objectives and will be explain in the following paragraphs.

In the first approach, clinical data were collected from files and presented in an international meeting (European Society of Medical Oncology congress, ESMO congress 2012). In that time, only patients treated until 2011 were considered, and there were no pathological revision of the cases. This first analysis allowed the planning of the future research namely the elaboration of a list of cases that did not underwent breast surgery in IPOPFG and so, after that many letters were send to other laboratories to ask the blocks for central pathological revision in IPOPFG. From the majority of subsequent works, time period considered were 1980-2014, for a total of 111 eligible patients with a median follow-up of 5.5 years. The patients' characteristics are presented in Table 1.1.

Patients presented a median age of 66 years, were previous or active smokers in 36.1%, alcohol consumers in 44.8%, being over-weight or obese in 66.7%. BRCA2 mutation was present in 8.3% and 12.5% developed a second malignancy, prostate cancer predominantly. Five-year overall survival was 68.8%.

It was possible to do additional IHC markers (total of six markers) in 95 patients that had tissue blocks available. With this information, unsupervised hierarchical clustering was performed to look for new MBC subgroups, described in chapter 2.

In the relapse analysis, we selected all recurrent cases (27 cases) from the population of 111 patients. Twenty-three were considered eligible for the study (had complete information about adjuvant and palliative treatments) and were matched with 69 FBC cases treated in the same Institution, in the same time-period, with the same stage and with controlled age ($\pm$ 5 years).

The CYP2D6*4 polymorphism study, chapter 3, was performed based on blood or tissue sample from patients treated with adjuvant tamoxifen (53 patients), between 1992-2012.

Table 1.1: Patients and tumours' characteristics

| | Characteristics | n | Frequency (%) |
|---|---|---|---|
| **Patients' age** | $\leq$ 60 years | 35 | 31.5 |
| *n=111* | > 60 years | 76 | 68.5 |
| **Histological type** | Ductal | 99 | 90.1 |
| *n=111* | Papillary | 3 | 2.7 |
| | Lobular | 1 | 0.9 |
| | Mixed | 7 | 6.3 |
| **Grade** | 1 | 14 | 13.7 |
| *n=102* | 2 | 62 | 60.8 |
| | 3 | 26 | 25.5 |
| **Tumour size** | $\leq$ 2cm | 48 | 43.2 |
| *n=111* | >2cm | 63 | 56.8 |
| **Nodal status** | N0 | 43 | 38.7 |
| *n=111* | N+ | 68 | 61.3 |
| **Stage*** | I | 25 | 22.6 |
| *n=111* | II | 29 | 26.1 |
| | III | 49 | 44.1 |
| | IV | 8 | 7.2 |
| **Estrogen Receptor (ER)** | Positive | 107 | 96.4 |
| *n=111* | Negative | 4 | 3.6 |
| **Progesterone Receptor (PR)** | Positive | 99 | 89.2 |
| *n=111* | Negative | 12 | 10.8 |
| **Androgen Receptor (AR)** | Positive | 72 | 63.7 |
| *n=95* | Negative | 23 | 20.4 |
| **HER2** | Positive | 9 | 8.1 |
| *n=111* | Negative | 102 | 91.9 |
| **Ki67** | Low | 65 | 57.5 |
| *n=95* | High | 30 | 26.5 |
| **p53** | Low | 75 | 66.4 |
| *n=98* | High | 23 | 20.4 |
| **Surgery** | Yes | 107 | 96.4 |
| *n=101* | No | 4 | 3.6 |
| **Adjuvant chemotherapy** | Yes | 44 | 42.7 |
| *n=103* | No | 59 | 57.3 |
| **Adjuvant endocrine therapy** | Yes | 86 | 83.5 |
| *n=103* | No | 17 | 16.5 |
| **Adjuvant radiotherapy** | Yes | 71 | 68.9 |
| *n=103* | No | 32 | 31.1 |

* According to $7^{th}$ edition of American Joint Committee on Cancer (AJCC) staging.

# References


[1] Registo Oncológico Nacional (2008) Available at: http://www.roreno.com.pt/images/stories/pdfs/ro_nacional_2008.pdf (accession on 29/10/2015).

[2] Siegel R, Naishadham D, Jemal A. (2013) Cancer statistics, 2013. CA: A Cancer Journal for Clinicians, **63**, 11-30.

[3] Engholm G, Ferlay J, Christensen N, et al. (2013) NORDCAN: Cancer incidence, mortality, prevalence and survival in the Nordic countries, version 5.3. Association of the Nordic cancer registries. Danish Cancer Society.p1-2.

[4] Giordano SH, Cohen DS, Budzar AU, et al. (2004) Breast carcinoma in men:a population-based study. Cancer, **101(1)**, 51-57.

[5] Stang A, Thomssen C. (2008) Decline in breast cancer incidence in the United States: what about male breast cancer? Breast Cancer Research and Treatment, **112(3)**, 595-596.

[6] Speirs V, Shaaban AM. (2008) The rising incidence of male breast cancer. Breast Cancer Research and Treatment, **115(2)**, 429-30.

[7] White J, Kearins O, Dodwell D, et al. (2011) Male breast carcinoma: increased awareness needed. Breast Cancer Research, **13(5)**, 219.

[8] Jemal A, Siegel R, Xu J, Ward E. (2010) Cancer statistics, 2010. CA: A Cancer Journal for Clinicians, **60(5)**, 277-300.

[9] Bagley CS, Wesley MN, Young RC, et al. (1987) Adjuvant chemotherapy in males with cancer of the breast. American Journal of Clinical Oncology, **10(1)**, 55-60.

[10] Giordano SH. (2005) A review of the diagnosis and management of male breast cancer. The oncologist, **10(7)**, 471-479.

[11] Johansson I, Killander F, Linderholm B, Hedenfalk I. (2014) Molecular profiling of male breast cancer-Lost in translation? The International Journal of Biochemistry & Cell Biology, **53**, 526-35.

[12] Anderson WF, Althuis MD, Brinton LA, Devesa SS. (2004) Is male breast cancer similar or different than female breast cancer? Breast Cancer Research and Treatment, **83(1)**, 77-86.

[13] Giordano SH, Budzar AU, Hortobagyi GN. (2002) Breast cancer in men. Annals of Internal Medicine, **137(8)**, 678-687.

[14] Thomas DB. (1993) Breast cancer in men. Epidemiologic Reviews, **15(1)**, 220-231.

[15] O'Malley CD, Prehn AW, Shema SJ, Glaser SL. (2002) Racial/ethnic differences in survival rates in a population-based series of men with breast carcinoma. Cancer, **94(11)**, 2836-2843.

[16] Cutuli B, Lacroze M, Dilhuydy JM, et al. (1995) Male breast cancer: results of the treatments and prognostic factors in 397 cases. European Journal of Cancer, **31A(12)**, 1960-1964.

[17] Mabuchi K, Bross DS, Kessler II. (1985) Risk factors for male breast cancer. Journal of the National Cancer Institute, **74(2)**, 371-375.

[18] Nahleh ZA, Srikantiah R, Safa M, et al. (2007) Male breast cancer in the veterans affairs population: a comparative analysis. Cancer, **109(8)**, 1471-1477.

[19] Nilsson C, Koliadi A, Johansson I, et al. (2013) High proliferation is associated with inferior outcome in male breast cancer patients. Modern Pathology, **26(1)**, 87-94.

[20] Goss PE, Reid C, Pintilie M, et al. (1999) Male breast carcinoma:a review of 229 patients who presented to the Princess Margaret Hospital during 40 years:1955-1996. Cancer, **85(3)**, 629-639.

[21] Ruddy KJ, Winer EP. (2013) Male breast cancer: risk factors, biology, diagnosis, treatment, and survivorship. Annals of Oncology, **24**, 1434-43.

[22] Fentiman IS, Fourquet A, Hortobagyi GN. (2006) Male breast cancer. Lancet, **367(9510)**, 595-604.

[23] Korde LA, Zujewski JA, Kamin L, et al. (2010) Multidisciplinary meeting on male breast cancer. Summary and research recommendations. Journal of Clinical Oncology, **28(12)**, 2114-2122.

[24] Weigelt B, Geyer FC, Reis-Filho Js. (2010) Histological types of breast cancer: how special are they? Molecular Oncology, **4(3)**, 192-208.

[25] Sasco AJ, Lowenfels AB, Pasker-de-Jong P. (1993) Review article: epidemiology of male breast cancer. A meta-analysis of published case-control studies and discussion of selected aetiological factors. International Journal of Cancer, **53(4)**, 538-549.

[26] Basham VM, Lipscombe JM, Ward JM, et al. (2001) BRCA1 and BRCA2 mutations in a population-based study of male breast cancer. Breast Cancer Research, **4(1)**, R2.

[27] Chodick G, Struewing JP, Ron E, et al. (2008) Similar prevalence of founder BRCA1 and BRCA2 mutations among Ashkenazi and non-Ashkenazi men with breast cancer: evidence from 261 cases in Israel, 1976-1999. European Journal of Medical Genetics, **51(2)**, 141-147.

[28] Couch FJ, Farid LM, DeShano ML, et al. (1996) BRCA2 germline mutation in male breast cancer cases and breast cancer families. Nature Genetics, **13(1)**, 123-125.

[29] Ding YC, Steele L, Kuan CJ, et al. (2010) Mutations in BRCA2 and PALB2 in male breast cancer cases from the United States.Breast Cancer Research and Treatment, **126(3)**, 771-778.

[30] Friedman LS, Gayther SA, Kurosaki T, et al. (1997) Mutation analysis of BRCA1 and BRCA2 in a male breast cancer population. American Journal of Human Genetics, **60(2)**, 313-319.

[31] Ottini L, Rizzolo P, Zanna I, et al. (2008) BRCA1/BRCA2 mutation status and clinical-pathologic features of 108 male breast cancer cases from Tuscany: a population-based study in central Italy. Breast Cancer Research and Treatment, **116(3)**, 577-586.

[32] Struewing JP, Coriaty ZM, Ron E, et al. (1999) Founder BRCA1/2 mutations among male patients with breast cancer in Israel. American Journal of Human Genetics, **65(6)**, 1800-1802.

[33] Thorlacius S, Olafsdottir G, Tryggvadotir L, et al. (1996) A single BRCA 2 mutation in male and female breast cancer families from Iceland with varied cancer phenotypes. Nature Genetics, **13(1)**, 117-119.

[34] Levy-Lahad E, Friedman E. (2007) Cancer risks among BRCA1 and BRCA2 mutations carriers. British Journal of Cancer, **96**, 11-15.

[35] Liede A, Karlan BY, NArod SA. (2004) Cancer risks for male carriers of germline mutations in BRCA1 or BRCA2: a review of the literature. Journal of Clinical Oncology, **22(4)**, 735-742.

[36] Tai YC, Domchek S, Parmigiani G, Chen S. (2007) Breast cancer risk among male BRCA1 and BRCA 2 mutation carriers. Journal of National Cancer Institute, **99(23)**, 1811-1814.

[37] Kwiatkowska E, Teresiak M, Filas V, et al. (2003) BRCA 2 mutations and androgen receptor expression as independent predictors of outcome of male breast cancer patients. Clinical Cancer Research, **9(12)**, 4452-4459.

[38] Wooster R, Mangion J, Eeles R, et al. (1992) A germline mutation in the androgen receptor gene in two brothers with breast cancer and Reifenstein syndrome. Nature Genetics, **2(2)**, 132-134.

[39] Lobaccaro JM, Lumbroso S, Belon C, et al. (1993) Androgen receptor gene mutation in male breast cancer. Human Molecular Genetics, **2(11)**, 1799-1802.

[40] Syrjakoski K, Hyytinen ER, Kuukasjarvi T, et al. (2003) Androgen receptor gene alteration in Finnish male breast cancer. Breast Cancer Research and Treatment, **77(2)**, 167-170.

[41] Fackenthal JD, Marsh DJ, Richardson AL, et al. (2001) Male breast cancer in Cowden syndrome patients with germline PTEN mutations. Journal of Medical Genetics, **38(3)**, 159-164.

[42] Silvestri V, Rizzolo P, Falchetti M, et al. (2010) Mutation analysis of BRIP1 in male breast cancer cases: a population-based study in central Italy. Breast Cancer Research and Treatment, **126(2)**, 539-543.

[43] Silvestri V, Rizzolo P, Falchetti M, et al. (2011) Mutation screening of RAD51C in male breast cancer patients. Breast Cancer Research, **13(1)**, 404.

[44] Blanco A, de la Hoya M, Balmaña J, et al. (2011) Detection of a large rearrangement in PALB2 in Spanish breast cancer families with male breast cancer. Breast Cancer Research and Treatment, **132(1)**, 307-315.

[45] Falchetti M, Lupi R, Rizzolo P, et al. (2007) BRCA 1/BRCA2 rearrangements and CHECK2 common mutations are infrequent in Italian male breast cancer cases. Breast Cancer Research and Treatment, **110(1)**, 161-167.

[46] Sauty de Chalon A, Teo Z, Park DJ et al. (2009) Are PALB2 mutations associated with increased risk of male breast cancer? Breast Cancer Research and Treatment, **121(1)**, 253-255.

[47] Ohayon T, Gal I, Barruch RG, et al. (2004) CHEK2*1100delC and male cancer risk in Israel. International Journal of Cancer, **108(3)**, 479-480.

[48] Silvestri V, Rizzolo P, Zanna I, et al. (2010) PALB2 mutations in male breast cancer: a population-based study in central Italy. Breast Cancer Research and Treatment, **122(1)**, 299-301.

[49] Syrjakoski K, Kuukasjarvi T, Auvinen A, Kallioniemi O-P. (2003) CHEK2 1100C is not a risk factor for male breast cancer population. International Journal of Cancer, **108(3)**, 475-476.

[50] Wasielewski M, den Bakker MA, van den Ouweland A, et al. (2008) CHEK2 1100C and male breast cancer in the Netherlands. Breast Cancer Research and Treatment, **116(2)**, 397-400.

[51] Young IE, Kurian KM, Annink C, et al. (1999) A polymorphism in the CYP17 gene is associated with male breast cancer. British Journal of Cancer, **81(1)**, 141-143.

[52] Brinton LA, Carreon JD, Gierach GL, et al. (2009) Etiologic factors for male breast cancer in the US Veterans Affairs medical care system database. Breast Cancer Research and Treatment, **119(1)**, 185-192.

[53] Hultborn R, Hanson C, Kopf I, et al. (1997) A prevalence of Klinefelter's syndrome in male breast cancer patients. Anticancer Research, **17(6D)**, 4293-4297.

[54] Thomas DB, Jimenez LM, McTiernan A, et al. (1992) Breast cancer in men: risk factors with hormonal implications. American Journal of Epidemiology, **135(7)**, 734-748.

[55] Sorensen HT, Friis S, Olsen JH, et al. (1998) Risk of breast cancer in men with liver cirrhosis. American Journal of Gastroenterology, **93(2)**, 231-233.

[56] Brinton LA, Richesson DA, Gierach GL, et al. (2008) Prospective evaluation of risk factors for male breast cancer. Journal of the National Cancer Institute, **100(20)**, 1477-1481.

[57] Ewertz M, Holmberg L, Tretli S, et al. (2001) Risk factors for male breast cancer- a case control study from Scandinavia. Acta Oncologica, **40(4)**, 467-471.

[58] Hsing AW, McLaughlin JK, Cocco P, et al. (1998) Risk factors for male breast cancer (United States). Cancer causes & control, **9(3)**, 269-275.

[59] Medras M, Alicja F, Pawel J, et al. (2006) Breast cancer and long-term hormonal treatment of male hipogonadism. Breast Cancer Research and Treatment, **96(3)**, 263-265.

[60] Thellenberg C, Malmer B, Travelin B, Gronberg H. (2003) Second primary cancers in men with prostate cancer: an increased risk of male breast cancer. Journal of Urology, **169(4)**, 1345-1348.

## Related Articles

Article 1 - A context-free framework applied to healthcare environments (Unpublished) 1.3

# A Context-Free Framework Applied to Healthcare Environments

Miguel Henriques Abreu[a], Pedro Henriques Abreu[b], Daniel Castro Silva[c], Hugo Amaro[b]

[a] *Portuguese Institute of Oncology of Porto, Rua António Bernardino de Almeida s/n, 4200-465, Porto Portugal*
[b] *Department of Informatics Engineering, Faculty of Sciences and Technology of Coimbra University and CISUC, Pólo II - Pinhal de Marrocos 3030-290, Coimbra Portugal*
[c] *Faculty of Engineering of Porto University, Department of Informatics Engineering, Laboratory of Artificial Intelligence and Computer Science, Rua Dr. Roberto Frias, s/n 4200-465 Porto Portugal*

**Abstract**

Electronic Healthcare Systems have come to represent an important role in medical institutions, helping physicians and professionals in their activity. However, some of these systems present some drawbacks, such as context-dependency; inability to support knowledge extraction; and are sometimes developed without consulting the professionals that will use them, leading to low acceptance and use levels.

In this paper, we propose a flexible system that is context-independent, allowing any physician to define the system's working environment, thus being able to be used in any service or department of a healthcare unit. The personalization capabilities allow it to adapt to the needs of a team or person, promoting acceptance as well as collaborations through the system.

The developed framework was tested by a team of 7 oncologists interested in studying male breast cancer. The results are very promising, with high user satisfaction levels, as well as system usability and performance levels.

*Keywords:* Flexible Healthcare System, Personalizable Healthcare System, Personalized Queries, Digital Medical Records

## 1 Introduction

An Information System (IS) can be defined as the overall information processing in an organization, including the involved human players and the used information technology [1]. Recent advances in the technological field lead to a propitious context for the developing of IS in healthcare environments, reducing clinical error, supporting healthcare professionals, increasing the efficiency of care or/and improving the quality of patient care [2] [3].

Historically, and following Haux's work [4], the evolution of IS in healthcare environments can be synthesized in seven direction: (1) from paper-based records towards computer-based information storage and processing tools [5] (this migration was not universally followed as illustrated below); (2) from local to global information system architectures that include not only systems enclosing multiple hospital centers but also changing the paradigm of the information system itself to a more patient-centered approach rather than an institution-centered approach [6]; (3) from healthcare professionals to patients and consumers, which reflects the need to provide support to more actors than just healthcare professionals [7]; (4) from using data solely for patient care towards supporting healthcare planning and clinical research [8]; (5) from technical to strategic information management priorities emphasizing that the main problem today is not technical but it encompasses organizational, social issues among others; (6) towards including new types of data related to the appearance or improvement of new exams/analysis [9]; and (7) towards ubiquity and the inclusion of new devices for data collection and health monitoring.

In spite of such research works, many success and failures stories have composed the spectrum of IS in healthcare [10]. In this particular topic, one question can arise: how can the success of an IS be measured? Success, like failure, is a multidimensional and contested concept [11]. Following Delone and McLean (two of the most cited authors) [12], success can be classified in six dimensions: Information quality, system quality, information use, user satisfaction, individual impact and organisational impact. Ten years later, the same authors [13] updated their model presenting information quality, system quality, service quality, intention to use, user satisfaction and net benefits as new dimensions to evaluate success in IS (a full revision work about this topic can be

found in [14]).

One of the consequences of low IS acceptance levels is the use of ad-hoc databases by the physicians for their research works, either directed towards a scientific publication of certain medical findings or simply to understand the reality of a particular unit/service. However, this brings forward several drawbacks: since physicians are not technically qualified to use more advanced tools, these databases are usually built using simple tools, not adequate for the task at hand; it is very hard to maintain these databases up-to-date because they are frequently stored in a single computer, without distributed/online access and in many cases only few physicians know of its existence; patient characterization is based on the expertise of one physician (or a small group of physicians), which means that if the characterization changes, the clinical files will need to be revised again and a new file will be produced to conduct new studies.

To increase physician satisfaction, a flexible and fully configurable Clinical System (CS) is proposed in this paper. A CS (sometimes called Department Electronic Medical Record [15]) is a system that is a specialized service for a specific department [16]. As the system was developed in partnership with the department of medical oncology of the Portuguese Institute of Oncology of Porto (IPO-Porto), one case was used for the initial validation of the system: Male Breast Cancer. IPO-Porto is a tertiary cancer center that treats more than 10.000 new patients per year. In Portugal (a country with a population of approximately 10 million), there are more than 150 IS spread over 104 public healthcare institutions [17]. However, despite some institutions having more than one IS being used at the same time (85% in 2011), half the information produced in hospital units was in traditional paper format, and professionals usually prefer to maintain their patient's records in physical files, even if the digital information exists [18]. There are several possible explanations for that fact: systems do not cover all the needs of the clinician in his daily practice; in some cases the multiple systems in use are unable to communicate with each other; the computational demands of the systems are sometimes too high for the available computers in the hospitals, thus turning the use of the system into a time-consuming task; the rotation of different systems in a specific unit (typically this change occurs every 3 years); or the need of specific training to be able to proficiently use this type of systems (as these systems are not developed by clinicians,

most of the operations are not too intuitive).

The main goal of this project it to provide physicians with a framework that allows them to manage patient records in a simple and intuitive manner. This system presents the following functionalities (these functionalities and others are fully described in section 3:

- Automatic Database Model Creation. When using the developed tool, the database model is automatically created during a file import operation or by specifying variables and relationships between them;
- Creation of Personalized Queries. Users can create their own queries by picking the variables involved in the query and a set of logical operators to specify selection and presenation restrictions on the data. The framework also allows for these queries to be shared with other professionals.
- Data Import/Export from other sources. The developed tool is able to import/export data from/to Microsoft Excel (probably one of the most used software tools) and IBM SPSS software.

To the best of our knowledge, this constitutes the first effort to propose and develop a flexible CS.

All project phases followed the general recommendations proposed by Ammenwerth et al. [3] as well as the design-reality gap model [19] and the evaluation was inspired by the framework proposed by Delone and McLean [13] (for a full revision on the evaluation topic please consult [20] [16]).

The developed framework was tested by a team of seven oncologists for a period of three months, and was then evaluated in terms of user satisfaction, system usability and system performance.

The results show that the physicians were satisfied with the system, as it was rapidly ready to use in accordance with the specifications they came up with. The versatility and flexibility of the system are pointed out as some of the main features that contribute to its acceptance.

In the future, the goal is to expand this concept to an IS in a central hospital facility promoting its use by the physicians, also promoting data and knowledge sharing across departments.

The rest of this article is structured as follows. Section 2 describes some of the related work done in the context of IS in healthcare, and in particular those developed targeting a specific service. Section 3 describes the proposed system architecture and the modules that compose it. Section 4 presents the system evaluations, in terms of user satisfaction, system usability and performance. Finally, in section 5, the main conclusions from the developed work are drawn and some pointers for future work are provided.

## 2 Related work

There are several literature review works that clearly illustrate the past, present and future directions in this area (Gunasekaran et al. [21]; Murphy et al.[22]; Yusof et al. [16]; Fichman et al.[23]). In this section, a set of works englobing different pathologies will be described.

Jordan et al. [24] developed a system called MAGIC that helps clinicians in the detection of two classes of abnormal events related to hemodynamics (that includes episodes of hypotension, hypertension, bradycardia, and tachycardia) and by laboratory results (including acidosis, alkalosis, hypercardia, hypoxia, low saturation, hyponatremia, hypernatremia, hypokalemia, hyperkalemia, hypocalcemia, hypercalcemia, anemia, hypoglycemia, and hyperglycemia).

Monitoring patient information during surgery, such as vital signs, inhaled anesthetics or ventilation parameters, among others and using several well-known scores, such as Acute Physiologic and Chronic Health Status Evaluation, Multi-organ System Illness Score and the therapeutic Intervention Severity Score, this system was able to detect the two event classes.

At the end, the performance of MAGIC systems was compared with 46 clinicians using 24 patients as the test set. The system presented very good accuracy in the laboratory results (more than 95%); however, some differences were detected in comparison with the clinicians related to hemodynamics events, and should be a point to

revise in the future.

Slater et al. [25] constructed a system that supports four levels of coding (non-operative or post-procedural admission, diagnostic group, specific condition and for injury and infection the aetiological factor) related to the reasons for admitting children to intensive care. More than 19.000 children admitted in intensive care from the period between 1997-2000 in New Zealand and Australia were enrolled in this study, with two diagnoses detected in 61% of the cases, three diagnoses in 29%, four diagnoses in 13%, five diagnoses in 6% and finally, six diagnoses in 3% of the cases.

Cimino et al. [26] developed an ambulatory record application called DOP (Decision-supported Outpatient Practice System). In its application, they support some services like progress notes, review of reports from ancillary systems and health maintenance reminders. This system uses the Medical Entity Dictionary composed by more than 67.000 terms to help physicians find relationships between used terms and to maintain a controlled terminology in patient records. These terms were divided into five groups: patient problems, adverse reaction, new medication, existing medication and medication modification. To evaluate the system, 27 different users (8 attending physicians, 18 resident physicians, and 1 nurse) used the system, in order to identify how physicians deal with controlled terminology in a system. The users tried to add 238 terms, 151 related to medication. In almost 70% of the 151 cases, these terms are related to adding dose route and dose frequency information. At the end, and in spite of the fact that the authors detected some important lines of improvements (e.g sometimes the user makes mistakes due to some lack of terms), they claimed that this application constitutes an important contribution to the area.

Kinn et al. [27] presented an Electronic Medical Record (EMR) tool for cardiac patients. This tool stored medical information and allows physicians to produce some data queries. After some developments they decided to merge the EMR tool with a new one called Virtual Lipid Clinic, capable of monitoring cholesterol in enrolled patients. This tool extracts key elements from data stored in the EMR tool and, most importantly, it provides timely online alerts when the signals of the patients come outside the range of the pre-established guidelines. At the end, two main groups were constructed and divided: the first is the control group (the patients that are exclusively followed by pe-

riodical consultations and whose data is stored in paper – n=764) and the test group (the patients that have periodical consultation and whose data is stored in the EMR tool and monitored by the Virtual Lipid Clinic – n=1109). The final results showed that the cholesterol of the test group patients presented better values and statistical significance in comparison to the control group patients.

In conclusion, in spite of the fact that many EMR have been developed for many pathologies, some issues still remain. First, these systems are developed for a specific domain, which means that if a new physician wants to use it for a new context (a new pathology, for instance) it will not be possible, mainly because the system presents specificities inherent to the primary domain. Also, and even when the physicians help in the development phase (e.g. selecting the variables to be included in the system), after the development it is not common to be able to update the set of variables that is included in the system. These two issues were addressed in the solution proposed in this paper.

## 3 Project Architecture

This section describes the project architecture, and the modules that comprise the system, using the male breast cancer as the running example. The proposed system is organized in three layers, as shown in Fig. 1.

The system supports two types of users: system administrators, responsible for managing the system itself and the other users; and physicians, who use the system in support of their daily activities. The Interface layer contains two front-ends, one for each type of user, and each of which providing access to different modules on the Logical layer of the system. The Administrator front-end provides access to the system and user management modules, while the Medical front-end provides access to the Patient Management module, the communications module and the search engine module. All these modules require access to the system database.

### *3.1 System Administration*

The administrator is responsible for defining the system model that will support all physician activities. He has access to the system definition module (for system config-

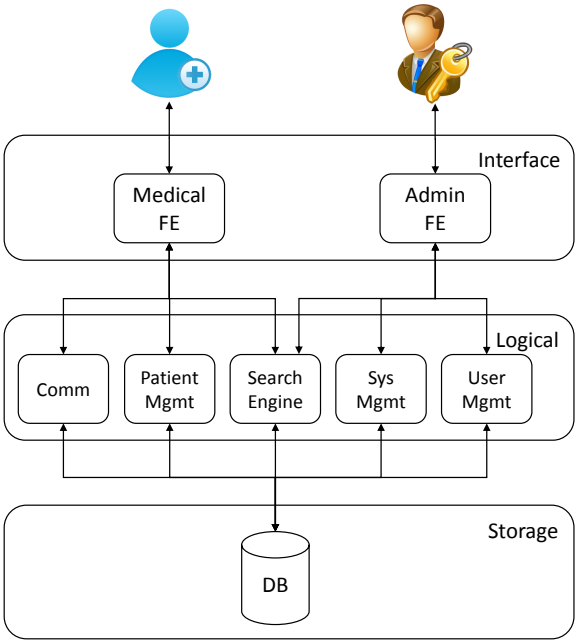



Figure 1: Global System Architecture

uration) and to the user management module, described below. The system interface is made so that these administrative responsibilities can be performed by a physician, and not necessarily by an informatics expert.

### *3.1.1 System Definition*

The proposed approach uses a dynamic database phylosophy, where tables and relashionships in a relational database are created dynamically as the model is defined by the system administrator, attempting to mimic the model a database engineer would come up with if modeling the specific reality.

The overall idea is based on a hierarhical variable definition. Variables can either be simple (in which case, we'll refer to them as fields) or compound, in which case they can contain other variables. This structure allows for everything to be interpreted as a variable, from patients to simple fields.

A patient, the main variable of the system, is defined as comprised by a set of variables, each of which corresponds to a report on the patient (such as surgery report, pathological report, and so on). Each of these reports is itself a variable, composed

by other variables, which can be fields or compound variables (which act as groups of variables).

From a more technical perspective, each compound variable is stored in the database as a table, while fields are stored as columns in the table corresponding to the variable containing the field. A set of relationships can be defined between variables, as to express how they relate to each other:

1. One-to-many relationships. A column is added to the table on the *many* side of the relationship that references the table on the *one* side of the relationship. For instance, a patient can have many blood pressure (BP) measurements (which are defined as compound variables having the test date, values, and other possible variables or fields), but each BP measurement belongs to one patient only.
2. Many-to-many relationships. A new intermediate table is created to store the relationship between the two variables, including the value and references to both variables. This kind of relationship can be useful for shared information. A typical example is geographical area information, such as the postal code. Since more than one patient can share the same postal code, and to avoid information redundancy, the intermediate table allows for the variable to be shared by several patients.

All these relationships, as well as information regarding each of the variables, are stored in one table, named the Master Variable Control Table (MVCT), which allows for a centralization of this information, along with a reduction of processing costs during database operations (including interface generation). Table 1 shows the structure of this table, and the information it contains.

Table 2 shows an example of the MVCT with a variable (Epidemiological Data) containing three other variables (Age, Family History and BRCA2 Mutation).

Figure 2 shows the interface used to define such information.

### *3.1.2 User Management*

User management is based on the concept of users and user groups on one hand, and variables on the other. User groups are used to group users with similar characteristics (for instance, all chemotherapists are grouped together), in order to facilitate

Table 1: Master Variable Control Table

| Field Name | Field Description |
|---|---|
| id | Variable unique ID |
| pid | Parent Variable ID (if any) |
| name | Name of Variable |
| type | Data Type |
| rel | Relationship with Parent |
| min | Minimum value (if applicable) |
| max | Maxiimum value (if applicable) |
| default | Default Value |
| format | Accepted Formats |
| order | Order during Form Generation |
| optional | Variable Mandatory or Optional |
| target | Variable alias target |
| notes | Human-friendly Notes |
| cond | Variable visibility in Form |

user management. The system creates a permission matrix from existing variables and existing users and groups, thus allowing the administrator to define which variables are editable by each user. Associated with user management appears the concept of view, which consists in pre-configured work desktops or dashboards associated with users or user groups. If a user has no views associated to him, a standard generic view is used, automatically generated with all variables he has access to. Customized views can be used to provide access to more variables than the ones the user has permissions to edit (in this case, the view will provide read-only access to those variables) or to restrict the variables the user can see.

### *3.2 Medical Perspective*

The system provides three main modules to be used by physicians: Patient Management module, Search Engine module and Communications module.

Table 2: Example content of the MVCT Table

| id | pid | name | type | rel | min | max | default | format | order | optional |
|---|---|---|---|---|---|---|---|---|---|---|
| 1 | -1 | Epidemiological Data | 99 | 0 | 0 | 0 | null | null | 0 | 1 |
| 2 | 1 | Age | 0 | 0 | 1 | 140 | null | null | 1 | 0 |
| 3 | 1 | Family History | 2 | 0 | 0 | 1024 | null | ((.—\n—\r)*) | 2 | 1 |
| 4 | 1 | BRCA2 Mutation | 4 | 0 | 0 | 0 | Yes | Yes+_+No | 3 | 1 |

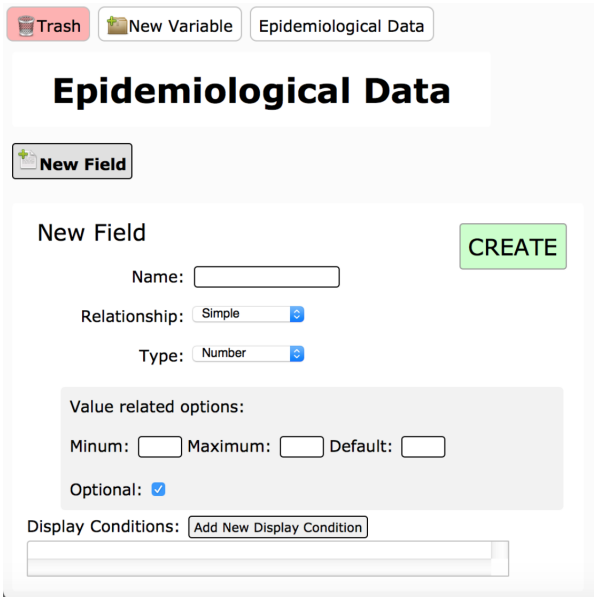



Figure 2: Example Interface

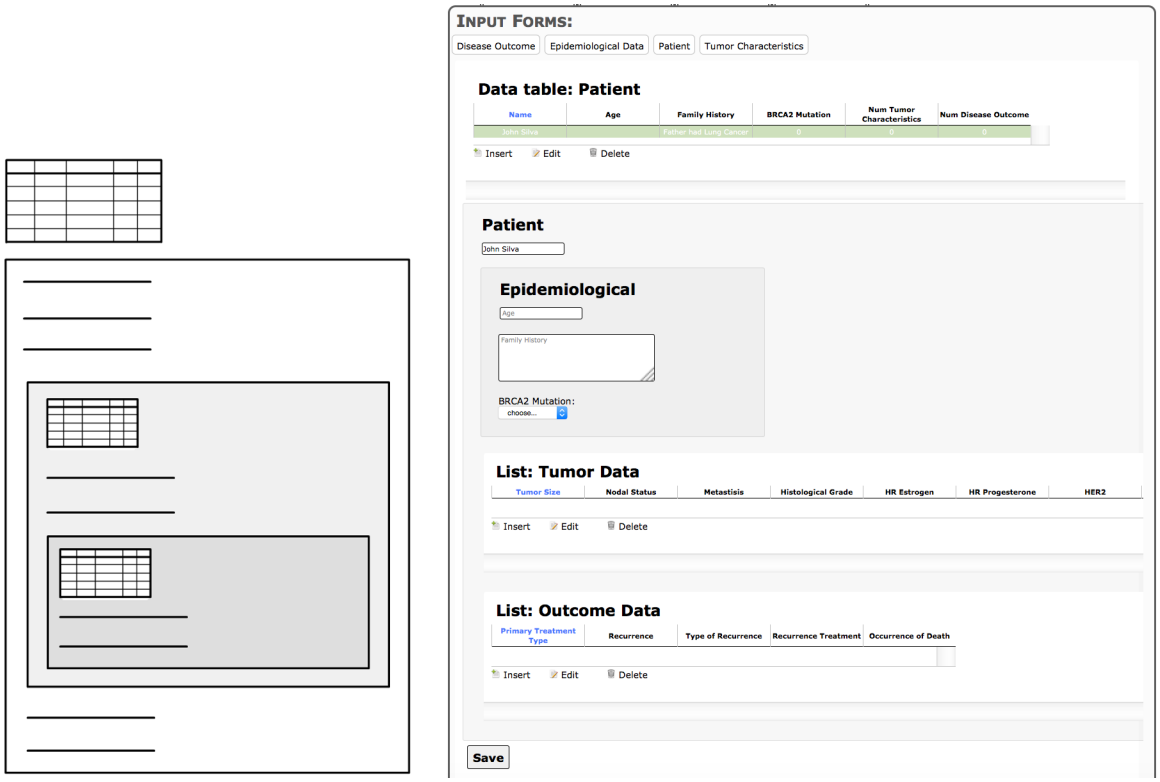
### *3.2.1 Patient Management*

This module provides physicians with functionalities to deal with patient records, allowing them to list, search, create, view and edit patient details.

Given the generic and dynamic nature of the system, the user interface cannot be predefined, but needs to be automatically generated from the information in the database. This interface is based on the concept of nested forms, where each form corresponds to one compound variable, as illustrated in Fig. 3a.

To facilitate user interaction, avoiding unnecessary long forms, nested forms are initially collapsed, and can be expanded or collapsed again as the user navigates the form.

Attending to the constraints on each variable and field, the form generator creates the appropriate form element (text area, multiple selection fields, and so on) to display. In the case of variables linked by one-to-many or many-to-many relationships, a table is added to the form with the existing values, so that one (or more) can be selected for the patient, or a new one can be created. Figure 3b shows an example of a form generated by the interface generator, including the interface for the example presented above for




(a) Nested Forms (b) Example Interface

Figure 3: Interface Generator

the Epidemiological Data variable.

### *3.2.2 Search Engine*

The system provides full support for creating and sharing personalized queries. In the developed system, the administrator can provide some pre-configured queries to be available to all users, and physicians can create their own personalized queries, as well as share them with other users. A personalized query is defined as a set of variables to select, associated with possible group options (such as 'show results grouped by gender'); a set of filters that define the patients whose records will be used to retrieve and compile the data (e.g. 'all patients above 65 years'); and possible grouping operations (such as averaging variables). The system transforms this configuration into a query to be made on the database, and stores the query configuration so it can be reused in the

future, or shared with other physicians. Figure 4 shows an example of a personalized query (one created by the administrator and shared with all users), in this case for the overall survival of male breast cancer patients.

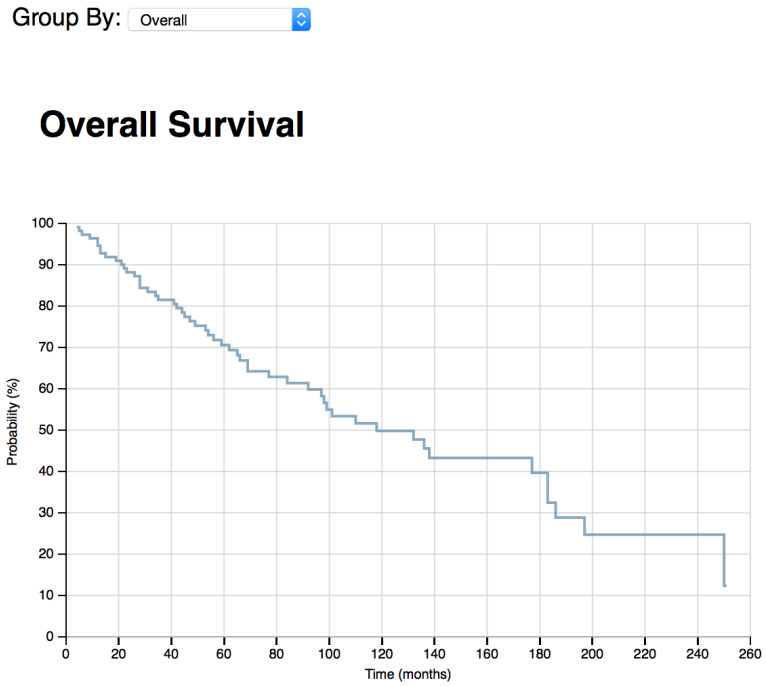



Figure 4: Overall Survival in Male Breast Cancer Patients

### *3.3 Communications*

The system provides two means of internal communications between users: private messages and forum. Private messages can be send to a single or multiple recipients. User groups are used as a basis to send messages to multiple recipients, and the system provides functionalities similar to those of a regular e-mail client (such as reply, reply to all, forward, delete). Selected information from user profiles can also be used to select multiple recipients (for instance, all users who belong to a certain department, or have a certain specialization), which provides extra flexibility to the module.

The forum provides regular functionalities, with the additional feature of providing group-based access control without the need of further configurations: user groups

created for user management are automatically mapped into sections of the forum accessible only by users in that group. Additionally, some selected information from user accounts (such as department and hospital unit) is also used to automatically create and grant access to specific sections of the forum. This avoids the need to manually create these groups, all the while still allowing new sections and sub-sections to be created by the users.

## 4 System Evaluation

The system was made available to the breast cancer unit in IPO-P where it was used by a team of physicians composed by 7 oncologists interested in studying the male breast cancer pathology.

In an initial stage, a team comprised by the 7 oncologist and the specialists who developed the system was assembled, and for a period of four weeks the group brainstormed, created and fine tuned a model for the system (this model is the set of variables and its structure, and is organized in the form of a clinical record, for familiarity reasons). After that, the system was configured for the specific pathology, and data from the first 12 weeks of use of the prototype system has been collected and analyzed.

At the end of this period, the system was evaluated, attempting to determine user satisfaction, system usability and system performance.

### *4.1 User Satisfaction*

To evaluate the user satisfaction, two different strategies have been used. The first strategy consists in identifying the number accesses and operations performed in the system. The second one consists in user satisfaction questionnaires.

The number of system operations was collected and divided into 5 categories, as shown in Fig. 5.

We can see that insertions happen only in the first 5 weeks, when the oncologists inserted the 100 patients into the system. The number of edits to patient records was also higher in the first few weeks (possibly made after the insertion) and then stabilized to approximately 2 edits per day. Patient record views has a similar behavior. Regarding the creation and viewing of personalized queries, 14 queries were created during the

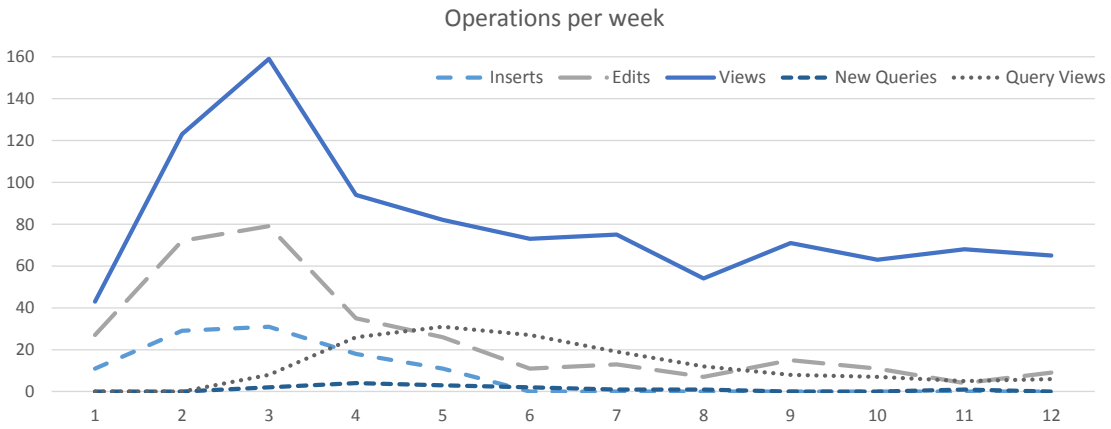



Figure 5: Number of system access

operational test phase, most of them between weeks 2 to 6. The number of query views was also higher during these weeks, stabilizing afterwards at about 7 query views per week in the last weeks.

These results seem to indicate that after an inicial period of records creation, system use stabilizes to 'normal' levels. It should be noted that in the test context, there are no new patients to handle, and thus patient insertions were effectively over at the end of the fifth week.

### *4.2 System Usability*

In order to evaluate system usability, users were once again faced with a questionnaire aimed at ascertaining the level of usability for the system. The Computer System Usability Questionnaire (CSUQ) [28] was used to evaluate system usability. The questions that comprise the CSUQ are shown in Appendix Appendix A, and Table 3 shows the obtained results for each question.

The overall CSUQ score was of 5.52, with a value of 5.61 for the System Usefulness sub-scale, 5.37 for the Information Quality sub-scale and 5.52 for the Interface Quality sub-scale.

Regarding the individual questions, most of them (16 of 19) obtained a median of 5 (out of 7) or higher, which constitutes good results. At this point it is import to highlight that 10 questions present a median of 6 or higher. However, some questions obtained

Table 3: CSUQ Results

| Question | Median | Semi-Interquartile Range |
|---|---|---|
| 1 | 6 | 0.5 |
| 2 | 6 | 0.5 |
| 3 | 4 | 0.5 |
| 4 | 5 | 0.25 |
| 5 | 5 | 0.5 |
| 6 | 6 | 0.75 |
| 7 | 6 | 0.5 |
| 8 | 5 | 0.5 |
| 9 | 3 | 0.75 |
| 10 | 3 | 0.75 |
| 11 | 6 | 0.75 |
| 12 | 6 | 0.5 |
| 13 | 7 | 0.25 |
| 14 | 5 | 0.75 |
| 15 | 6 | 0.5 |
| 16 | 7 | 0.5 |
| 17 | 5 | 0.75 |
| 18 | 5 | 0.75 |
| 19 | 6 | 0.25 |

lower classifications, and in particular 2 questions (questions 9 and 10) presented a median of 3. These two questions are related to the output that the system generates in an error situation, which attending to the achieved results should be a point to improve in the future.

The oncologists were also asked to point out positive and negative aspects regarding the system. The most cited positive aspects include the concept of views (this concept, previously explained, was considered one of the key features of the system, given the levels of personalization it provides); custom queries (the majority of the re-

spondents mentioned this feature also as very helpful); data export capabilities (the possibility to export the data to excel was also considered a key feature, as it allows the clinicians to use the data on further studies without much effort in compiling or preparing the data). As negative aspects, the oncologists refer design issues (such as color scheme, size of text and action buttons, and color of some icons); the several steps necessary to create a custom query (respondents suggest that this process should be made easier).

### *4.3 System Performance*

System performance was evaluated by measuring access times to the system, and operational demands.

Table 4 shows the results regarding load times for two pages (the Login Page (LP), which is the first page of the system; and the Authenticated Page (AP), which shows the authenticated user his View to the system) in two situations (First View, which means that all resources are loaded from the web; and Repeat View, which means that most resources are already cached and no download is necessary). These results were obtained using the WebPageTest tool (WebPageTest is an online website performance assessment tool, available at `http://www.webpagetest.org/`) and Mozilla Firefox Developer Tools (More information available from `https://developer.mozilla.org/en-US/docs/Tools`).

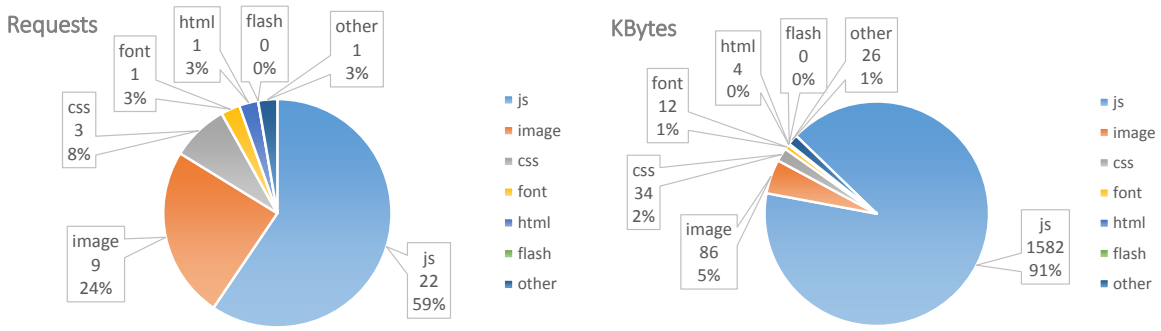
The results show that page load times are within normal webpage load times, and that, as expected, repeat views have much lower load times, since most resources are already cached. Figure 6 shows the number and total size (in KB) of the requests necessary to load the authenticated page of the system for the first time. As can be seen in Table 4, for the following page loads, the number of requests decreases from 36 to 2, and the total size decreases from 1.7 MB to 31 KB, which translates into lesser network demand during system use.

## 5 Conclusions and Future Work

In this work, a flexible framework to support different departments inside a healthcare institution was proposed. With this tool, a team of physicians can characterize

Table 4: System Load Performance (LP - Login Page; AP - Authenticated Page)

| | | | | | Document Complete | | | Fully Loaded | | |
|---|---|---|---|---|---|---|---|---|---|---|
| | **Time** | **First Byte** | **Start Render** | **DOM Elems.** | **Time** | **Req.** | **Bytes In** | **Time** | **Req.** | **Bytes In** |
| LP First View | 1.828s | 0.323s | 1.830s | 30 | 1.828s | 9 | 634 KB | 1.997s | 10 | 635 KB |
| LP Repeat View | 0.386s | 0.184s | 0.491s | 30 | 0.386s | 1 | 3 KB | 0.561s | 3 | 4 KB |
| AP First View | 2.376s | 0.06s | 1.641s | 1906 | 2.376s | 33 | 1524 KB | 2.53s | 36 | 1746 KB |
| AP Repeat View | 0.294s | 0.110s | 0.386s | 1906 | 0.294s | 2 | 4 KB | 1.637s | 2 | 31 KB |




(a) Number of Requests

(b) Bytes Transfered

Figure 6: Number of Requests and Bytes Transfered for the Authenticated Page (First View)

their patients and control all this information using personalized views of the system, thus allowing them to manage this data in a more suitable and personalized manner. The system is context-independent, which means that it can easily support every department, allowing each physician of each department to have customized views of the system and data. This tool also supports the creation of personalized queries over the data, providing physicians with a intuitive manner to create and share these queries, and possible new information that can be extracted from them, with other physicians. To the best of the author's knowledge, this is the first tool that allows for a full customization of both the system model and queries executed on the defined system model.

The implementation was tested with a team of 7 oncologists from a service interested in studying male breast cancer. After an initial setup, the oncologists' use of the system was monitored, as to extract some information. Results show that the use of the system is consistent with the expectations, and the users show high levels of satisfaction with it. Usability was evaluated with a standard questionnaire that showed the users to believe the system has high levels of usability. Also, system performance was evaluated and the results show that the system has standard performance levels for a website.

Some aspects have been identified as improvements to the system, to be implemented in the future. Most of these are related to the design and usability of the user interface, as detected by the physicians. Other points have also been identified, such

as the diversity of formats supported for data import/export.

## References


[1] Winter, A., Ammenwerth, E., Bott, O., Brigl, B., Buchauer, A., Gräber, S., Grant, A., Häber, A., Hasselbring, W., Haux, R., Heinrich, A., Janssen, H., Kock, I., Penger, O., Prokosch, H., Terstappen, A., and Winter, A. (2001) *International Journal of Medical Informatics* **64(2-3)**, 99–109.

[2] Bates, D. W., Cohen, M., Leape, L. L., Overhage, J. M., Shabot, M. M., and Sheridan, T. (2001) *Journal of the American Medical Informatics Association* **8(4)**, 299–308.

[3] Ammenwerth, E., Graber, S., Herrmann, G., Burkle, T., and Konig, J. (2003) *International Journal of Medical Informatics* **71**, 125–135.

[4] Haux, R. (2006) *International Journal of Medical Informatics* **75**, 268–281.

[5] Haux, R., Ammenwerth, E., Herzog, W., and Knaup, P. (2002) *International Journal of Medical Informatics* **66**, 3–120.

[6] Itkonen, P. (2002) *Methods of Information in Medicine* **41(5)**, 387–392.

[7] Shah, S. G. S., Fitton, R., Hannan, A., Fisher, B., Young, T., and Barnett, J. (2015) *International Journal of Medical Informatics* **85**, 111–118.

[8] Zvarova, J., Preiss, J., and Sochorova, A. (1997) *International Journal of Medical Informatics* **45**, 59–64.

[9] Kulikowski, C. A. (2002) *Methods of Information in Medicine* **41(1)**, 20–24.

[10] Dahlen, R. W. (1997) *Bulletin of the Medical Library Association* **85(4)**, 443.

[11] Markus, M. and Tanis, C. (2000) The enterprise systems experience - from adoption to success In R. W. Zmud, (ed.), Framing the Domains of IT Research: Glimpsing the Future Through the Past, chapter 10, pp. 173–207 Pinnaflex Cincinnati, Ohio.

[12] Delone, W. H. and McLean, E. R. (1992) *The Institute of Management Sciences* **3(1)**, 60–95.

[13] Delone, W. H. and McLean, E. R. (2003) *Journal of Management Information Systems* **19(4)**, 9–30.

[14] Van derMeijden, M. J., Tange, H. J., Troost, J., and Hasman, A. (2003) *Journal of the American Medical Informatics Association* **10(3)**, 235–243.

[15] Haÿrinen, K., Saranto, K., and Nykänen, P. (2008) *International Journal of Medical Informatics* **77**, 291–304.

[16] Yusof, M. M., Papazafeiropoulou, A., Paul, R. J., and Stergioulas, L. K. (208) *International Journal of Medical Informatics* **77**, 377–385.

[17] Public Hospitals in Portugal (2013).

[18] Almeida, A. A Gestão Sistémica da Informação nos Hospitais Públicos Portugueses: uma Perspectiva Actual (2011) Master Thesis of Science of Documentation and Information at Faculty of Arts, University of Lisbon.

[19] Heeks, R. (2006) *International Journal of Medical Informatics* **75**, 125–137.

[20] Rahimi, B. and Vimarlund, V. (2007) *Journal of Medical Systems* **31(5)**, 397–432.

[21] Gunasekaran, A., Ngai, E. W. T., and McGaughey, R. E. (2006) *European Journal of Operational Research* **173**, 957–983.

[22] Murphy, E. C., Ferris, F. L., and O´Donnell, W. R. (2007) *Investigative Ophthalmology & Visual Science* **48(10)**, 4383–4389.

[23] Fichman, R. G., Kohli, R., and Krishnan, R. (2011) *Information Systems Research* **22(3)**, 419–428.

[24] Jordan, D. A., Mckeown, K. R., Concepcion, K. J., Feiner, S. K., and Hatzivassiloglou, V. (2001) *Journal of the American Medical Informatics Association* **8(3)**, 267–280.

[25] Slater, A., Shann, F., and McEniery, J. (2003) *Intensive Care Medicine* **29**, 271–277.

[26] Cimino, J. J., Patel, V. L., and Kushniruk, A. W. (2001) *Journal of the American Medical Informatics Association* **8(2)**, 163–173.

[27] Kinn, J. W., O'Toole, M. F., Rowley, S. M., Marek, J. C., Bufalino, V. J., and Brown, A. S. (2001) *The American Journal of Cardiology* **88**, 163–164.

[28] Lewis, J. R. (1995) *International Journal of Human-Computer Interaction* **7(1)**, 57–78.

## Appendix A CSUQ Questions

Table A.5: Questions of the CSUQ

| Number | Question |
|---|---|
| 1 | Overall, I am satisfied with how easy it is to use this system |
| 2 | It was simple to use this system |
| 3 | I can effectively complete my work using this system |
| 4 | I am able to complete my work quickly using this system |
| 5 | I am able to efficiently complete my work using this system |
| 6 | I feel comfortable using this system |
| 7 | It was easy to learn to use this system |
| 8 | I believe I became productive quickly using this system |
| 9 | The system gives error messages that clearly tell me how to fix problems |
| 10 | Whenever I make a mistake using the system, I recover easily and quickly |
| 11 | The information (such as online help, on-screen messages, and other information) provided with this system is clear |
| 12 | It is easy to find the information I needed |
| 13 | The information provided for the system is easy to understand |
| 14 | The information is effective in helping me complete the tasks and scenarios |
| 15 | The organization of information on the system screens is clear |
| 16 | The interface of this system is pleasant |
| 17 | I like using the interface of this system |
| 18 | This system has all the functions and capabilities I expect it to have |
| 19 | Overall, I am satisfied with this system |

# Chapter 2

# Prognosis

Classically, there is the misperception that MBC is an inherently aggressive disease, associated with a poor outcome [1–5]. However, the unfavourable prognosis of these patients has been attributed, in recent years, to important factors not controlled in previous publications [6–13], namely age and stage [14]. Other implicated factors in MBC patients' outcomes were: comorbid conditions [15], delay in diagnosis [16] or even differences in tumour biology [1]. For example, in a 1941 study [17] the average in diagnostic delay was 29 months, in 1995 the mean duration of symptoms was 21 months [18], and it remains in more recent series with a mean of 6-10 months [19, 20].

The principal large series that compared survival between the two genders are summarized in Table 2.1, and present conflicting results even when important prognostic factors were matched.

Having in mind that BC is a heterogeneous disease, it is important to understand which factors really affect prognosis and should be matched when comparing outcomes for MBC and FBC.

As in female, tumour size and lymph nodes involvement are the most consistent prognostic factors described in MBC patients [18, 25–33]. Males with tumours measuring between 2 and 5 cm have a 40% higher risk of death than those with smaller lesions and patients with nodal involvement have a 50% higher risk of death than dose with N0 disease [14]. Ten-year breast cancer specific-survival (BCSS) for N0 was described as 77-84%, for N1: 44-50% and for N2: 14-14%. N+ disease is more prevalent in MBC patients [33], and was also reported in small ($\leq$2 cm) and hormonal receptor positive tumours, which means that this disease cannot be considered an indolent one, *a priori* [35].

The prognostic relevance of histological grade is contradictory [14, 36, 37], with some studies reporting the association between high grade and poor prognosis [36] while others have not state it [14, 37]. Tumour grading is not always simple to assess and inter-observer discrepancies are a major concern [35].

Table 2.1: Comparison of survival outcomes between genders

| Reference | n, male | n, female | Matched factors | Results |
|---|---|---|---|---|
| Scott-Conner CE, et al. 1999 [21] | 4755 | 624174 | Age, stage. | Similar relative OS, trend toward worst survival in male with stage III/IV. |
| Giordano SH, et al. 2004 [14] | 2524 | 380856 | Age, stage. | Similar relative OS. |
| Miao H, et al. 2011 [3] | 2665 | 459846 | Age, year of diagnosis, FU time, stage, treatments. | Better relative BCSS for male. Inferior survival and BCSS for men when patients were not matched. |
| Nilsson C, et al. 2011 [4] | 99 | 369 | Age, year of diagnosis. | Inferior relative and OS for male. |
| Foerster R, et al. 2011 [22] | 108 | 108 | Age, year of diagnosis, N status, grade, stage, HR expression, HER2 expression. | Similar OS. |
| Shaaban AM, et al. 2012 [23] | 251 | 263 | Age, grade, N status | Similar OS. |
| Gnerlich JL, et al. 2012 [2] | 1541 | 244518 | Controlling for confounders | Inferior BCSS for men with stage I disease. Inferior OS for men with stage I-III |
| Greif JM, et al. 2012 [5] | 13457 | 1439866 | NA | Inferior OS for male. Better 5-year OS for women stage I-II. |
| Chen X, et al. 2013 [1]* | 150 | 300 | Age, year of diagnosis and stage | Inferior DFS and OS for male. |
| Kwong A, et al. 2014 [24]* | 132 | 8118 | Year of diagnosis, stage, size, N status | Inferior OS for male but better BCSS for male. |

* Chinese population;
BCSS: breast cancer specific-survival; DFS: disease free survival; FU: follow-up; HR: hormonal receptors; NA: not applicable; N status: lymph node status; OS: overall survival.

In the majority of series grade 2 is the most often observed [14, 15, 38], perhaps because these difficulties could be bigger in male since there is lack of normal breast tissue that is a prerequisite to assessing nuclear atypia correctly. Other factors related to prognosis namely breast cancer subgroups and relapse will be analysed in next sections.

## 2.1 Breast cancer subgroups

From the publication of Perou et al. [39] in 2000, we understand that BC is a heterogeneous disease that could be divided into comprehensive subgroups associated with differences in treatment response and outcome [40]. This first work was done using gene expression profiles to define the intrinsic BC subtypes, that were translated thereafter into clinical practice using surrogate IHC markers [41–43], based on expression of a small number of proteins. This expression correlates with the transcriptional subtypes in 75-90% [44].

Over the years, IHC-based subtypes changed, with the addition of new markers (namely ki67) and redefinition of IHC markers cut-offs. Now we have already three definitions [44–46], summarized in Table 2.2, with proved prognostic implications in FBC patients [39, 47, 48]. Basal-like and HER2 subgroups display the worst prognosis and luminal A the best one, while luminal B tumours have an intermediate prognosis probably related to less sensibility to endocrine therapy compared to luminal A and higher proliferation [49–52].

Table 2.2: Definitions of FBC subtypes based in IHC markers

| |
|---|
| **Luminal A-like tumours:** |
| Definition I: ER and/or PR positive, HER2 negative;<br>Definition II: ER and/or PR positive, HER2 negative, low ki67;<br>Definition III: ER positive, PR positive*, HER2 negative, low ki67. |
| **Luminal B-like tumours:** |
| Definition I: ER and/or PR positive with/out HER2 positive;<br>Definition II: ER and/or PR positive with/out HER positive and/or high ki67;<br>Definition III: ER positive, HER2 positive and/or high ki67 and/or PR negative*. |
| **Triple negative (basal-like) tumours:** ER, PR and HER2 negatives, and sometimes also CK5/6 and/or EGFR and/or CK14 positive. |
| **HER2-enriched:** ER and PR negatives, HER positive. |

*Based on a 20% cut-off for PR;
ER: estrogen receptor, PR: progesterone receptor, CK: cytokeratin.

Contrarily to female, only a few studies were conducted in male that addressed BC subtypes and showed conflicting results, probably related to different IHC definitions used, without an uniform distribution of patients by all subtypes [53, 54]. The most important studies are listed

in the following table (Table 2.3). There was no study that has used the new definition of BC subgroups in MBC patients, that was proposed by Prat et al. [46] with the redefinition of PR cutt-off.

Most MBC patients have luminal-like tumours, and HER2 positive and triple negatives are rare, indicating that FBC subgroups could not be linearly translated for male and could have a limited relevance in these patients.

Many explanations were stated to justify the disparities between frequencies of classical FBC subtypes in male. Luminal tumours in female patients are associated with older age and postmenopausal status [48]. Like in postmenopausal women, there are only low levels of circulating estrogen in males. Most of the estrogen is synthesized in peripheral tissue and has local effects in a paracrine or autocrine manner, which is important for the development of hormone-dependent breast cancers and probably explains the high incidence of this type of tumours in males [14, 54, 59, 60]. However, and even in luminal tumours, there were already documented differences in molecular basis (DNA aberrations and gene expression patterns), implying that MBC may be a distinct disease [37, 61]. Weber- Chapuis et al. [62], described that although a larger fraction of male tumours were ER positive compared to female ones, they were only weakly associated with antigens under estrogen control and more often positive for antigens under androgen control, while the opposite was true for FBC. This support the rationale that not all ER positive MBC tumours behave in the same way as ER positive in women, but rather seem to share features with both ER positive and ER negative FBC. In the same line, Johansson et al. [61] compared the global gene expression in MBC and FBC patients and identified two new subgroups in male, with half of them remain unclassified by the previous definitions: luminal M1 and luminal M2. The luminal M1 consists in a more aggressive phenotype with a poorer prognosis. Even though MBC tumours were mainly ER positive by IHC, there was a significant difference in ER-related gene signalling between the groups, with luminal M1 displaying an inferior correlation to ER-signalling. This is different from female, where only ER negative tumours are identified with similarly reduced ER-signalling scores [35, 37].

Triple negative tumours, are associated with young women [48, 63] and BRCA1 mutations [64, 65] and the low frequency of these tumours in men could be explained by higher age at diagnosis and low frequency of BRCA1 [14, 59, 66–69]. HER2 amplified tumours were observed in no more than 11% of male patients and the prognostic impact of this factor was not evaluated [70, 71]. Cecilia et al. [35] showed that the majority of HER2 amplified cases in their male series were IHC 1+ or 2+ and not 3+. It has been hypothesized that tumours, which are negative for HER2 by IHC, but amplified by in situ hybridization, could be falsely IHC-negative due to epitope loss during fixation [72]. Another alternative explanation could be that protein expression is not as highly correlated with gene amplification in MBC [35].

Table 2.3: MBC studies addressing BC subgroups

| Reference | n | Markers | Analysis performed | Definition, according Table 2.2 | Distribution by subgroups |
|---|---|---|---|---|---|
| Ge Y, et al. 2009 [54] | 42 | ER, PR<br>HER2<br>CK 5/6 | IHC | I | **. Luminal A-like:** 83%<br>**. Luminal B-like:** 17%<br>(0% HER2 positive)<br>**. Triple negative (basal-like):** 0%<br>**. HER2-enriched:** 0% |
| Kornegoor R, et al. 2012 [55] | 134 | ER, PR<br>HER2<br>Ki67<br>CK 5/6, CK14 | IHC | II | **. Luminal A-like:** 75%<br>**. Luminal B-like:** 21%<br>(3% HER2 positive)<br>**. Triple negative (basal-like):** 4%<br>**. HER2-enriched:** 0% |
| Sánchez-Muñoz A, et al. 2012 [56] | 43 | ER, PR<br>HER2<br>Ki67<br>CK 5/6 | IHC | II | **. Luminal A-like:** 44%<br>**. Luminal B-like:** 51%<br>(0% HER2 positive)<br>**. Triple negative (basal-like):** 5%<br>**. HER2-enriched:** 0% |
| Shaaban AM, et al. 2012 [23] | 251 | ER, PR<br>HER2<br>CK5/6,14,18,19 | Tissue microarrays | I | **. Luminal A:** 98%<br>**. Luminal B:** 0%<br>(0% HER2 positive)<br>**. Triple negative (basal-like):** 2%<br>**. HER2-enriched:** 0% |
| Nilsson C, et al. 2013 [35] | 197 | ER, PR<br>HER2<br>Ki67<br>CK 5/6 | IHC | I | **. Luminal A-like:** 81%<br>**. Luminal B-like:** 11%<br>**. Triple negative (basal-like):** 1%<br>**. HER2-enriched:** 0% |
| | | | | II | **. Luminal A-like:** 41%<br>**. Luminal B-like:** 51%<br>(9.6% HER2 positive)<br>**. Triple negative (basal-like):** 1%<br>**. HER2-enriched:** 0% |
| Yu XF, et al. 2013 [57] | 68 | ER, PR, HER2 | IHC | I | **. Luminal A-like:** 60.29%<br>**. Luminal B-like:** 25.00%<br>**. Triple negative:** 5.88%<br>**. HER2-enriched:** 8.82% |

ER: estrogen receptor; PR: progesterone receptor; IHC: immunohistochemical analysis; CK: cytokeratin.

All of these findings could suggest that female subgroups do not adequately identify the aggressive forms of MBC [23, 35, 54, 57], and accurate subtyping of MBC is essential to developing an appropriate therapeutic strategy, ideally based on a IHC-panel clinical feasibly and reproducible in routine practice.

### Related Articles

Article 2 - Male breast cancer: looking for better prognostic subgroups (Unpublished) 2.1.1.

# Male breast cancer: looking for better prognostic subgroups

Miguel Henriques Abreu[a], Noémia Afonso[a], Pedro Henriques Abreu[b], Francisco Menezes[c], Paula Lopes[c], Rui Henrique[c,d], Deolinda Pereira[a], Carlos Lopes[d]

[a]*Department of Medical Oncology, Portuguese Institute of Oncology of Porto, Rua António Bernardino de Almeida s/n, 4200-465, Porto Portugal*
[b]*Department of Informatics Engineering, Faculty of Sciences and Technology of Coimbra University and CISUC, Pólo II - Pinhal de Marrocos 3030-290, Coimbra Portugal*
[c]*Department of Pathology, Portuguese Institute of Oncology of Porto, Rua António Bernardino de Almeida s/n, 4200-465, Porto Portugal*
[d]*Department of Pathology and Molecular Immunology, Biomedical Sciences Institute of Abel Salazar, University of Porto, Rua Jorge Viterbo Ferreira, 228, 4050-313 Porto, Portugal*

---

**Abstract**

**Purpose:** Male Breast Cancer (MBC) remains a poor understood disease. Prognostic factors are not well established and specific prognostic subgroups are warranted.

**Patients/Methods:** Retrospective revision of 111 cases treated in the same Cancer Center. Blinded-central pathological revision with immunohistochemical (IHQ) analysis for estrogen (ER), progesterone (PR) and androgen (AR) receptors, HER2, ki67 and p53 was done. Cox regression model was used for uni/multivariate survival analysis. Two classifications of Female Breast Cancer (FBC) subgroups (based in ER, PR, HER2, 2000 classification, and in ER, PR, HER2, ki67, 2013 classification) were used to achieve their prognostic value in MBC patients. Hierarchical clustering was performed to define subgroups based on the six-IHQ panel.

**Results:** According to FBC classifications, the majority of tumours were luminal: A (89.2%; 60.0%) and B (7.2%; 35.8%). Triple negative phenotype was infrequent (2.7%; 3.2%) and HER2 enriched, non-luminal, was rare ($\leq$1% in both). In multivariate analysis the poor prognostic factors were: size$>$2cm (HR: 1.8; 95% CI: 1.0-3.4years, $p$=0.049), absence of ER (HR: 4.9; 95% CI: 1.7-14.3 years, $p$=0.004) and presence of distant metastasis (HR: 5.3; 95% CI: 2.2-3.1 years, $p<0.001$). FBC subtypes were independent prognostic factors ($p$=0.009, $p$=0.046), but when analyzed only luminal groups, prognosis did not differ regardless the classification used ($p>0.20$). Clustering defined

different subgroups, that have prognostic value in multivariate analysis ($p$=0.005), with better survival in ER/PR+, AR-, HER2- and ki67/p53 low group (median:11.5 years; 95% CI: 6.2-16.8 years) and worst in PR- group (median: 4.5 years; 95% CI: 1.6-7.8 years).

**Conclusion:** FBC subtypes do not give the same prognostic information in MBC even in luminal groups. Two subgroups with distinct prognosis were identified in a common six-IHQ panel. Future studies must achieve their real prognostic value in these patients.

*Keywords:* Male breast cancer, prognostic subgroups, survival

---

## 1 Introduction

Male breast cancer(MBC) accounts for only 1% of all breast cancers [1] with an incidence of 63/100.000 in 2008 in Portugal [2]. Despite its rareness, the morbi/mortality associated to this disease [3] has led to a progressive interest on it over the recent years although its etiology and tumor behavior remain poor understood [4]. Female breast cancer(FBC) is a heterogeneous disease and since its first comprehensive classification into four subtypes proposed by Perou et al. [5] and translated thereafter to clinical feasible and reproductive immunohistochemistry(IHC) markers, much has been learned about tumours behavior and patients outcomes and how different patients should be managed with a special attention for poor prognostic subgroups (triple negative and HER2 enriched tumours). The transposition of these subgroups to MBC reality indicates that they do not adequately identify the aggressive forms of MBC [6, 7, 8, 9] as the classical triple negative and HER 2 positive groups are rarely described in male setting. Even in the luminal group, there is an actual understanding that some of these tumours in male, might not have an active estrogen receptor(ER) pathway although they express ER in IHC analysis, leading to the question whether these patients respond to endocrine treatment in the same way as female with luminal tumours [10]. For these reasons, the relation between MBC subtypes and prognosis is not yet been stablished [6] and new subgroups, based on the conjunction of a limited biomarkers panel clinical feasible and reproductive, are warranted to better understand this disease and to provide the basis for optimal patient management [4]. The purpose of this study is to define

the subgroups that better describe the prognosis of MBC patients using the definitions already described to FBC and proposing a new one.

## 2 Patients and Methods

Study population comprised 111 cases of MBC treated in the same Cancer Center, the Portuguese Institute of Oncology of Porto, during 1980 to 2012 (maximum follow-up: 23 years).

Patients' information was retrospectively collected from clinical files and active follow-up was used to ascertain prognosis in patients that have been discharged the Institution. To reduce the effect of time, time-dependent variables were adjusted or grouped in main categories (stage was adjusted to $7^{th}$ edition of AJCC Staging, and treatment was grouped in categories).

Patients, or in case of death their substitutes, given written consent to participate on the study. This study was approved for local ethic committee.

### *2.1 Histologic evaluation*

One breast Pathologist centrally reviewed the HE slides. The histologic classification was based on WHO criteria and histologic grade in the Nottingham system. IHQ study for estrogen (ER) and progesterone (PR) receptors and HER2 was performed in all cases. Additional study for ki67, androgen receptor (AR) and p53 was performed in the cases with tissue blocks available. ER was considered positive if$\geq$1% cells showed nuclear staining [11], and for PR, two cut offs of positivity were used ($\geq$1% [11] or $>$20% [12]). Ki67 was interpreted as low or high by use a 14% threshold [13], p53 was considered high if$\geq$5% cells showed accumulation [14] and AR was considered positive when at least 10% of nuclei were stained [15]. Cases were considered HER2 positive when they are IHC-3+ according to the Dako score or FISH-amplified defined according ASCO/CAPO guidelines [16]. This analysis was blind for clinical outcomes.

### *2.2 IHQ subtyping*

Molecular subtypes based on two surrogate IHQ definitions were evaluated. For all cases (111) and according to the classical definition initially proposed by Perou et al.

[5] and adapted to clinic using surrogate IHC markers [13, 17, 18], we set the typing standard groups: ER positive and/or PR positive($\geq$1%) and HER2 negative for luminal A; ER positive and/or PR positive($\geq$1%) and HER2 positive for luminal B; ER and PR negative(0%) and HER2 positive for HER2 enriched; ER/PR negative(0%) and HER2 negative for basal-like(triple negative). This classification is referred on the text as 2000 classification, see Figure 1. In the 95 cases where it was possible to validate the ki67,

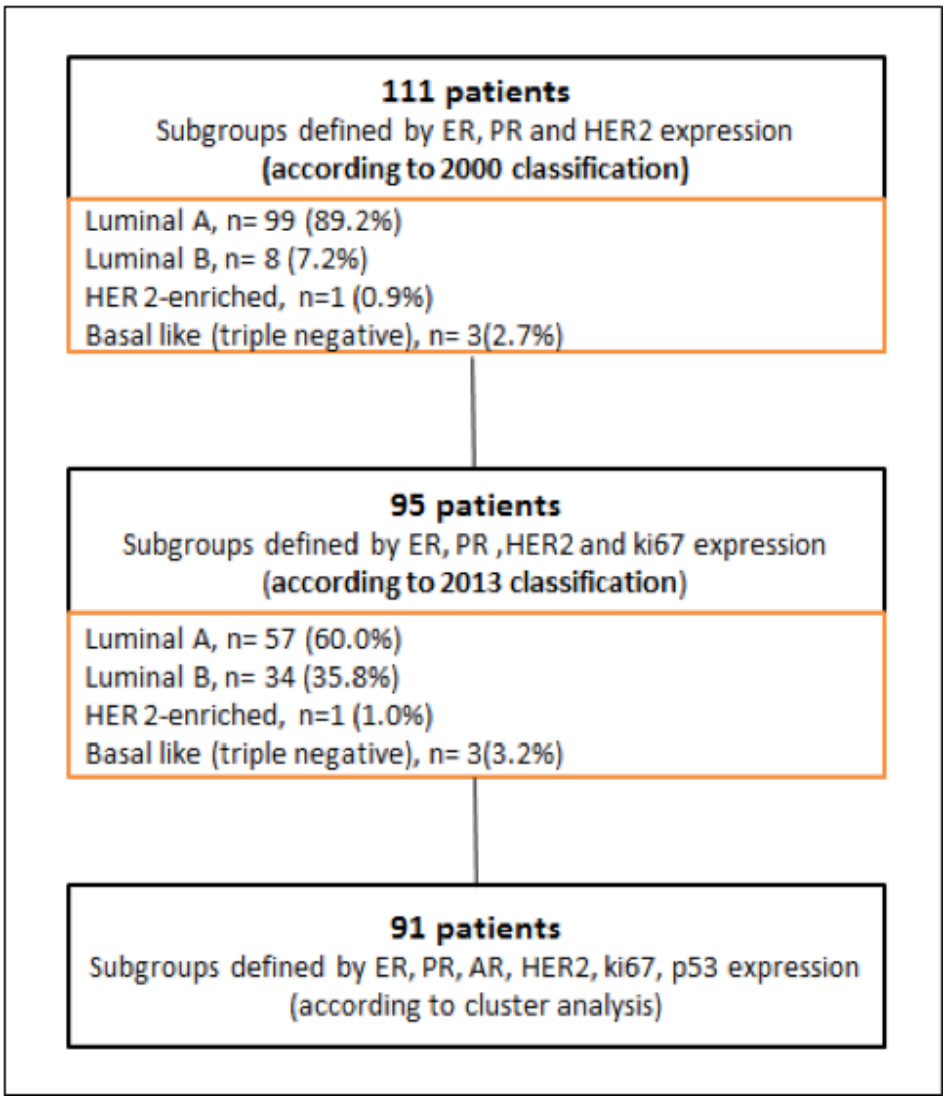



Figure 1: Subgroups analyzed

we compare the previous classification to the most recent one proposed by Prat et al. [12] reclassifying the luminal tumours in luminal A when they are ER positive($\geq$1%), PR positive($>$20%), HER2 negative and low ki67 and luminal B when they are ER positive($\geq$1%), HER2 positive and/or high ki67 and/or PR negative($\leq$20%). This classification is referred on the text as 2013 classification.

### *2.3 Statistical Analysis*

In the 91 patients with complete information for the six IHQ markers, hierarchical clustering was performed to define subgroups, as this tool has been already used in MBC setting and was described as a potent one for subdividing breast cancer patients into novel, clinically relevant groups [8, 10, 19]. For this purpose the linkage algorithm was used and, as all markers were categorical variables(negative/positive), similarities between patients were achieved by the jaccard distance measure. This analysis was performed using statistical program R (`http://www.r-project.org`).

Differences between patients and tumours' characteristics were evaluated with Pearson's x2 for categorical variables. Survival analysis was performed using Kaplan-Meier survival curves and Cox-regression model was used to achieve prognostic rule of different factors in uni/multivariate analysis. SPSS for windows version 15.0 (SPSS Inc., Chicago, IL, USA) was used for statistical calculations. P-values $\leq$0.05 were regarded as significant.

## 3 Results

All patients were Caucasians, with a median age at diagnosis of 66 years-old (range from 33 to 95 years-old), see Table 1.

The tumours less differentiated were more frequently associated with N+ disease, $p$=0.005 (in grade 3 tumours, N0: 19.2% vs N+: 80.8%) and with p53 accumulation, $p$=0.033 (p53 accumulation in grade 3 tumours was 47.6%).

The expression of ER was correlated with expression of PR, $p<0.001$(ER+ tumours were also PR+ in 92.5%) and the absence of expression of hormonal receptors(ER and/or PR), were associated with advanced stages $p$=0.006 and 0.024 respectively (in ER-: stage III: 50%, stage IV: 50% and in PR-: stage III: 41.7%, stage IV: 25%). PR-

Table 1: Clinical and pathological characteristics

| | Characteristics | *n* | Frequency (%) |
|---|---|---|---|
| **Patients age** | ≤60 years | 35 | 31.5 |
| *n=111* | >60 years | 76 | 68.5 |
| **Histological type** | Ductal | 99 | 90.1 |
| n=111 | Papillary | 3 | 2.7 |
| | Lobular | 1 | 0.9 |
| | Mixed | 7 | 6.3 |
| **Grade** | 1 | 14 | 13.7 |
| *n=102* | 2 | 62 | 60.8 |
| | 3 | 26 | 25.5 |
| **Tumor size** | ≤2 cm | 48 | 43.2 |
| *n=111* | >2 cm | 63 | 56.8 |
| **Nodal status** | N0 | 43 | 38.7 |
| *n=111* | N+ | 68 | 61.3 |
| **Stage*** | I | 25 | 22.6 |
| *n=111* | II | 29 | 26.1 |
| | III | 49 | 44.1 |
| | IV | 8 | 7.2 |
| **ER** | positive | 107 | 96.4 |
| *n=111* | negative | 4 | 3.6 |
| **PR** | positive | 99 | 89.2 |
| *n=111* | negative | 12 | 10.8 |
| **AR** | positive | 72 | 63.7 |
| *n=95* | negative | 23 | 20.4 |
| **HER2** | positive | 9 | 8.1 |
| *n=111* | negative | 102 | 91.9 |
| **ki67** | low | 65 | 57.5 |
| *n=95* | high | 30 | 26.5 |
| **p53** | low | 75 | 66.4 |
| *n=98* | high | 23 | 20.4 |
| **Surgery** | yes | 107 | 96.4 |
| *n=101* | no | 4 | 3.6 |
| **Adjuvant chemotherapy** | yes | 44 | 42.7 |
| n=103 | no | 59 | 57.3 |
| **Adjuvant hormonotherapy** | yes | 86 | 83.5 |
| *n=103* | no | 17 | 16.5 |
| **Adjuvant radiotherapy** | yes | 71 | 68.9 |
| *n=103* | no | 32 | 31.1 |

*According to $7^{th}$ edition of AJCC Staging;

ER- estrogen receptor, PR-progesterone receptor, AR-androgen receptor.

tumours were also associated with N+ disease(91.7%), $p$=0.024. The expression of AR was related to PR expression, $p$=0.002 (AR- tumours were also PR- in 71.4%) but not with ER expression, $p$=0.39. Patients with expression of AR frequently present low values of ki67, $p$=0.006, and low accumulation of p53, $p$=0.046. Low ki67 was also related to low accumulation of p53 ($p$=0.04).

Apart from the classification used (2000 or 2013), most tumours were luminal: A (89.2%; 60.0%) or B (7.2%; 35.8%), respectively. Triple negative phenotype was infrequent (2.7%; 3.2%) and HER2 enriched(non-luminal) tumours were rare ($\leq$1% in both). The introduction of ki67 determination and PR negativity in the distinction of luminal groups in 2013 classification, changed 26 patients (32.1%) from the luminal A to luminal B group.

### *3.1 Cluster Analysis*

Cluster analysis divided patients in three distinct groups (A, B, C), Figure 2. Clinicopathological features and IHQ profiles of the principle ones(A and B) could be seen in Table 2. Positivity of ER and PR were considered if$\geq$ 1% cells showed nuclear staining. The C group was composed only for one patient that had a large tumour ($>$2cm), with nodal involvement (N+) and no expression of any marker, and was not considered for comparisons with the others, in Table 2. The patients included in the A group were characterized by the absence of PR. Half of them did also no expression of ER or HER2, which means that this group contained the other triple negative patient that differ from that included in the C group by the positivity of AR.

In B group, all patients had luminal tumours (expressed ER and PR) and are distinguished in 2 subgroups by HER 2 expression: B1(HER2-) and B2 (HER2+). The B1 subgroup (HER2-) could also be subdivided in B1.1 (AR- and ki67/p53 low) and B1.2 (others). According to 2013 classification, in B1.1 there was only luminal A patients and in B2 there was only luminal B. In B1.2 there was also luminal A (62%) and B (38%). The expression of AR differed between these patients: in B1.1, all were AR-, in B1.2 97% expressed it and in B.2 all were AR+.

Patients in A group differed from those in B group only in IHQ profile, see Table 2.

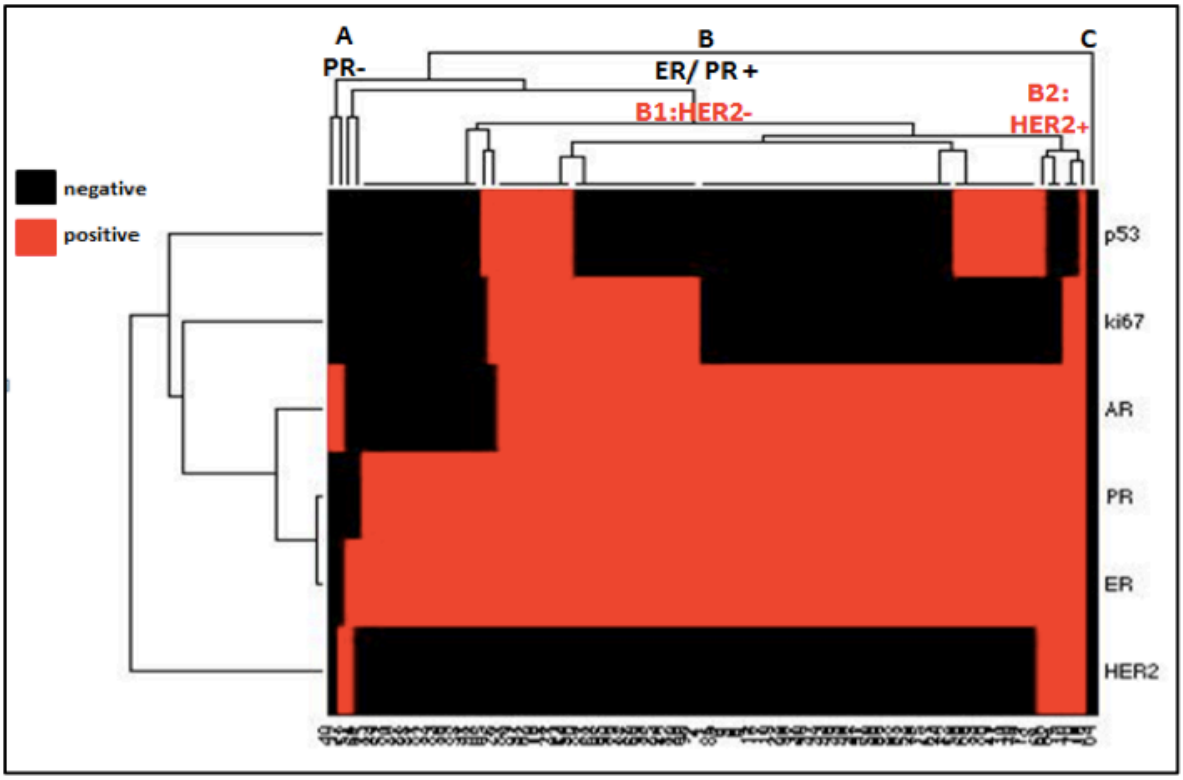



Figure 2: Hierarchical clustering of 6 immunohistochemical markers in 91 patients. The clustergram and corresponding dendogram indicate relationships between patients and immunohistochemical markers. ER- estrogen receptor, PR-progesterone receptor, AR-androgen receptor.

Table 2: Clinicopathological and immunohistochemical phenotype according to identified clusters of male breast cancer patients

| Characteristics | | A (n=4), | B1 (n=80) | | B2 (n=6), | *p value* |
|---|---|---|---|---|---|---|
| | | n (%) | B1.1 (n=14) n (%) | B1.2 (n=66), n (%) | n (%) | |
| **Patients age** | ≤60 years | 1 (25) | 5 (36) | 19 (29) | 2 (33) | 0.95 |
| | >60 years | 3 (75) | 9 (64) | 47 (71) | 4 (67) | |
| **Grade** | 1-2 | 2 (50) | 8 (57) | 46 (75) | 4 (67) | 0.14 |
| | 3 | 2 (50) | 6 (43) | 15 (25) | 2 (33) | |
| **Tumor size** | ≤2 cm | 2 (50) | 5 (38) | 29 (44) | 2 (333) | 0.89 |
| | >2 cm | 2 (50) | 9 (64) | 37 (56) | 4 (67) | |
| **Nodal status** | N0 | 0 | 5 (38) | 30 (46) | 0 | 0.05 |
| | N+ | 4 (100) | 8 (62) | 36 (54) | 6 (100) | |
| **Molecular type*** | Luminal A | 1 (25) | 14 (100) | 41 (62) | 0 | <0.001 |
| | Luminal B | 1 (25) | 0 | 25 (38) | 6 (100) | |
| | HER-2 enriched | 1 (25) | 0 | 0 | 0 | |
| | Triple negative | 1 (25) | 0 | 0 | 0 | |
| **ER** | positive | 2 (50) | 14 (100) | 66 (100) | 6 (100) | <0.001 |
| | negative | 2 (50) | 0 | 0 | 0 | |
| **PR** | positive | 0 | 14 (100) | 66 (100) | 6 (100) | <0.001 |
| | negative | 4 (100) | 0 | 0 | 0 | |
| **AR** | positive | 2 (50) | 0 | 64 (97) | 6 (100) | <0.001 |
| | negative | 2 (50) | 14 (100) | 2 (3) | 0 | |
| **HER2** | positive | 2 (50) | 0 | 0 | 6 (100) | <0.001 |
| | negative | 2 (50) | 14 (100) | 66 (100) | 0 | |
| **ki67** | low | 4 (100) | 14 (100) | 41 (62) | 3 (50) | 0.01 |
| | high | 0 | 0 | 25 (38) | 3 (50) | |
| **p53** | low | 4 (100) | 14 (100) | 45 (68) | 4 (67) | 0.05 |
| | high | 0 | 0 | 21 (32) | 2 (33) | |

*According to 2013 classification;

ER- estrogen receptor, PR-progesterone receptor, AR-androgen receptor.

### *3.2 Survival Analysis*

Fifty-five patients(49.5%) died, with a 5-year overall survival of 68.8%.

In univariate analysis, the patients with better survival had tumours$\leq$2cm (HR: 2.3, 95%CI: 1.3-4.2 years, $p$=0,005), with ER expression (HR: 5.1, 95%CI: 1.8-14.3 years, $p$=0.002), and did not have distant metastasis at diagnosis(HR: 6.9, 95%CI: 6.9-16.4 years, $p<0.001$), see Table 3.

When analyzed only the group of non-metastatic patients, survival was also different ($p$=0.028): stage I median: 20.0 years 10.9-29.1 years; stage II median: 9.8 years, 8.0-11.6 years; stage III median: 7 years, 4.7-9.3 years, Figure 3.

N status and PR expression seem to define different prognostic groups, although not statistically significant (N0 vs N+, HR: 1.8, 95%CI: 0.9-3.2 years, $p$=0.05 and PR positive vs negative, HR: 1.9, 95%CI: 0.9-4.0 years, $p$=0.07). In multivariate analysis tumour size, ER expression and presence of distant metastasis at diagnosis remained independent prognostic factors.

The two classifications used defined four patients' subgroups with different survivals. In 2000 classification (n=111) median survivals were: luminal A (n=99): 11.1 years (5.9-16.1 years), luminal B (n=8): 9.6 years (4.6-14.5 years), HER2 enriched, non-luminal (n=1): 6 years and for triple negative (n=3): 1.3 years (0-3.9 years), $p<0.001$.

In 2013 classification (n=95) median survivals were: luminal A (n=58): 10.5 years (8.1-12.9 years), luminal B (n=33): 11.8 years (8.7-14.9 years), HER2 enriched, non-luminal (n=1): 6 years and for triple negative (n=3): 1.3 years (0-3.9 years), $p<0.001$. However, when analyzed only the luminal groups, prognosis did not differ regardless the classification used ($p>0.20$).

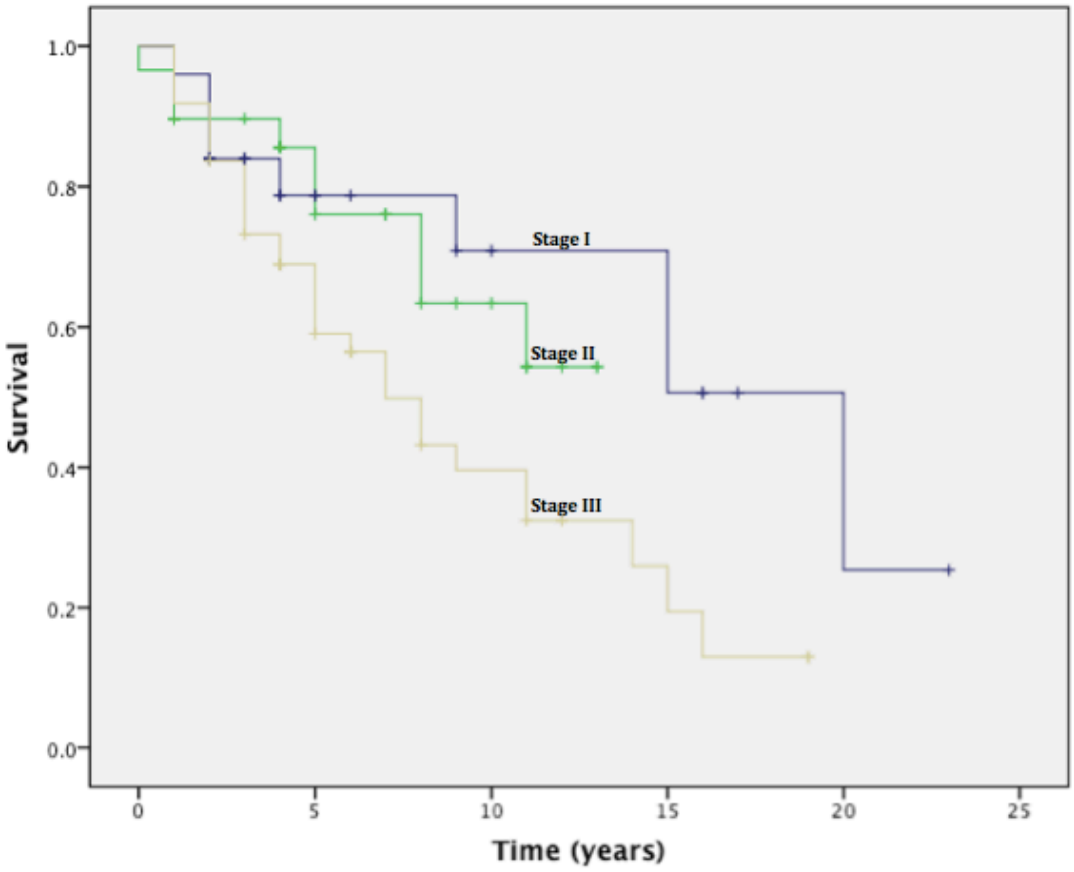
According to cluster groups, the median survival for A group (n=4) was 4,5 years (1.6-7.8), for B1.1 (n=66) was 11.5 years (6.2-16.8), for B1.2 (n=14) was 10.3 years (8.4-12.2), for B.2 (n=6) was 4.9 years (3.6-6.2) and for C (n=1) was 0,5 year (0.5-0.5), $p<0.001$.

When compared each subgroups and the prognostic factors not mutually exclusive (tumour size and presence of metastasis), in a multivariate analysis, molecular-like and cluster subgroups remained independent prognostic factors ($p$=0.009 for 2000 classification, $p$=0.046 for 2013 classification and $p$=0.045 for cluster groups).

Table 3: Survival analysis by Cox regression model

| Characteristics | Univariate analysis | | | Multivariate analysis | | |
|---|---|---|---|---|---|---|
| | **HR** | 95% CI, years | ***p value*** | **HR** | 95% CI, years | ***p value*** |
| **Age at diagnosis, years** | | | | | | |
| ≤60 vs >60 | 1.3 | 0.7-2.3 | 0.43 | ——- | ——- | ——- |
| **Grade** | | | | | | |
| 1/2 vs 3 | 0.9 | 0.4-1.7 | 0.69 | ——- | ——- | ——- |
| **Tumor size, cm** | | | | | | |
| ≤2 vs >2 | 2.3 | 1.3-4.2 | **0.005** | 1.8 | 1.0-3.4 | **0.049** |
| **Nodal status** | | | | | | |
| N0 vs N+ | 1.8 | 0.9-3.2 | 0.05 | ——- | ——- | ——- |
| **M status** | | | | | | |
| M0 vs M1 | 6.9 | 6.9-16.4 | <**0.001** | 5.3 | 2.2-3.1 | <**0.001** |
| **ER** | | | | | | |
| Positive vs negative | 5.1 | 1.8-14.3 | **0.002** | 4.9 | 1.7-14.3 | **0.004** |
| **PR** | | | | | | |
| Positive vs negative | 1.9 | 0.9-4.0 | 0.07 | ——- | ——- | ——- |
| **AR** | | | | | | |
| Positive vs negative | 1.4 | 0.8-2.6 | 0.26 | ——- | ——- | ——- |
| **HER2** | | | | | | |
| Negative vs positive | 0.9 | 0.3-2.5 | 0.85 | ——- | ——- | ——- |
| **ki67** | | | | | | |
| Low vs high | 0.6 | 0.3-1.2 | 0.14 | ——- | ——- | ——- |
| **p53** | | | | | | |
| Low vs high accumulation | 0.7 | 0.3-1.4 | 0.27 | ——- | ——- | ——- |

AR-androgen receptor; CI- confidence interval; ER- estrogen receptor; HR- hazard ratio; M status- metastasis presented at diagnosis; PR-progesterone receptor.




| Patients at risk | | | | | | |
|---|---|---|---|---|---|---|
| Stage I | 25 | 12 | 7 | 6 | 1 | 0 |
| Stage II | 29 | 17 | 7 | 0 | 0 | 0 |
| Stage III | 49 | 27 | 11 | 3 | 0 | 0 |

Figure 3: Overall survival by stage, in non-metastatic patients

## 4 Discussion

MBC is a rare disease and due to this, tumour biology remains poor understood [3, 8, 10, 20]. Most published data derived from retrospective studies that cover long time periods, with small samples and/or include patients from different hospitals and sometimes countries [4]. Many authors used also data from cancer registries [21, 22] with limitations regarding the absence of a clinical and pathological central revision and lacking of important information not usually recorded like treatments performed.

From some years ago, we know that FBC is a heterogeneous disease and can be divided into subgroups that define patients with different treatments response and outcomes [5, 23, 24]. Compared to women, some differences in DNA aberrations and gene expression patterns have been demonstrated, implying that MBC may be a different disease and hence, the molecular subtyping may also differ [7, 10]. However only a few studies were underwent in this field in male patients with different IHQ definitions and non-standardized classifications used [3].

Our series was based in a retrospectively analysis of MBC patients treated in the same Center, and clinical and pathological information were active collected and reviewed. Our sample characteristics show that we have a robust series. The majority of patients had $\geq$60 years old [22], presented with ductal carcinomas (followed by papillary histology and lobular is rare) [8, 25], moderately differentiated or undifferentiated [8], with nodal involvement at diagnosis [7], that often expressed ER/PR [17, 25] and AR [19]. The expression of HER2, high values of ki67 or p53 was not frequent [22]. The expression of AR relates to PR and low values of ki67 and p53 [8].

Like Nilsson et al. [7], tumour size and ER expression were independent prognostic factors but contrarily to them nodal status presented a borderline value in univariate analysis ($p$=0.05) with a tendency for worst prognosis to N+ disease. Similarly to us, N status was also not described as an independent prognostic factor by Yu et al. [6]. Differences in sample size could explain these findings, namely the differences in the ratio N+/N0 with a few N0 cases in our population. However the high prevalence of N+ disease in these patients compared to females [7], in our series 61.3%, even in small and ER positives tumours, determined that ER positive disease in men can not be always considered an indolent one and it is important to take other factors into account to

deal with them [7]. The notion of joint many markers to define distinct disease behaviors was first described in 2000, with the statement of the four intrinsic breast cancer subgroups. In MBC patients and using the classification based only in three markers (ER, PR and HER2), studies showed, like in our series, a high prevalence of luminal A tumours (75-98%) followed by luminal B (0-21%) and few cases of basal-like (0-4%) and HER2-enriched, non luminal (0%) [3, 8, 9]. Luminal patients showed better prognosis comparing to other subgroups but the low frequency of some groups (basal like and HER2-enriched) limits these comparisons and some authors opted to analyze luminal vs non-luminal groups [6]. Some hypotheses were stated to justify the low prevalence of HER2/basal like tumours in male patients namely related to age at diagnosis, BRCA1 frequency and IHQ technique and so, prognostic value of these subgroups were not conclusive [3, 6, 7].

The recent breast cancer subgroups proposed by Prat et al. [12], redefined the luminal groups using a new cut-off for PR. From the three markers classification to the new one, in our series 26 patients (32.1%) changed from the luminal A to luminal B group mainly because of ki67 discrimination. For our knowledge this new classification was not previously used in male setting. The most cited MBC studies that already defined luminal B patients using ki67 value [3, 26, 27] stated conflicting results. Similar to us, Kornegoor et al. [3] showed that these patients are less common than luminal A (21% vs 75%), but Snchez-Muoz et al. [27], described that they are more frequent than luminal A (51% vs 44%) and had also a better prognosis. The difficulties associated with the evaluation and standardization of ki67, namely the different cut-offs used, limit the ability to draw any firm conclusions [28], and even in male patients there are studies that found that it has no prognostic value [25, 29] while others found that it could be prognostic [30]. Although all the considerations, our series show that both classifications (2000 and 2013) defined different prognostic groups but when we look only for luminal groups, prognosis did not differ regardless the classification used ($p>0.20$), which probably means that their survival prediction capacity was related to the extremely low values of the rare groups, remained weak for the most frequent ones (luminals). Cause of that we defined new groups based in a six-IHQ panel. With cluster analysis, two principle groups were found, the biggest one with ER/PR positive tumours(distinguished by

HER2 expression) and a small one that contained all PR negative tumours. In the group with expression of hormonal receptors and absence of HER2, patients with low ki67/p53 and AR negative constituted the group that had better prognosis. This finding launches the debate about the role of routinely perform AR in male patients as it is related to hormonal receptors and is rarely negative. However as it was described as positive in some basal-like tumours there may be a subgroup of these patients that benefit from the determination of AR, namely to distinguish some of them with better prognosis. In this line, our cluster analysis separated the two triple negative patients according to AR expression. In the literature there are three publications in MBC setting that used hierarchical clustering analysis to define subgroups. Johansson et al. [10] using gene expression analysis to demonstrate that male and female breast cancer are different. Even in luminal patients, defined by expression of ER by IHQ, they reported a group with low ER signaling by gene expression leading the notion that luminal group is not homogeneous and there are some patients with a more aggressive behavior that could probably respond not so well to endocrine therapy. Shaaban et al. [8] did a comparison between male and female patients with a focus on hormone receptor profile. They described also a predominance of patients with expression of ER but they did not use ki67 in the definition of luminal B patients which difficult the direct comparison of the results. The publication that could be directly compared to our, with similar results, was performed by Kornegoor et al. [19] that performed a cluster analysis in MBC patients using a nine IHQ panel to define four groups. The first one, called hormone receptor-negative, contained tumours that do not express ER or PR. We remember that our A group was only constituted by patients with PR negative and contained also patients that are ER negative. The prognostic value of PR in male patients was already stated in a large retrospective study [31] and Kornegoor et al. [19] correlated it with other adverse prognostic factors as high mitotic count. Like us, these authors [19] also described three other clusters, all of them with tumours that express ER, one that contained the HER2 positive tumours and the other two that did not differ in ki67, p53 and AR but were distinguished by the histological grade (in low and intermediate grade). However in the article there is not described the distribution of patients subgroups by histological grade, and the prognostic value of this factor in MBC was not confirmed in

many studies [6, 7, 32] possibly related to the difficulty of obtain a consistent histological grade in male patients.

In conclusion this study show that tumour size, the presence of distant metastasis at diagnosis and the expression of ER are independent prognostic factors in MBC. FBC subgroups do not seem to provide similar prognostic information even with the most recent classification. Two important prognostic groups were defined by the cluster analysis: one with bad prognosis(the group with absence of PR) and one with good prognosis (the group with expression of ER and PR, and absence of HER2, AR, and low Ki67 and p53). More research is needed to confirm our findings and to clearly define the MBC prognostic subgroups to optimize treatment strategies and improve survival in these patients.

## References


[1] Siegel R, Naishadham D, Jemal A. (2013) Cancer statistics, 2013. CA: A Cancer Journal for Clinicians **63**, 11-30.

[2] Registo Oncologico Nacional (2008) `www.roreno.pt`, accessed in 9.1.2015.

[3] Kornegoor R, Verschuur-Maes AHJ, Buerger H, et al. (2012) Molecular subtyping of male breast cancer by immunohistochemistry. Modern Pathology, **25(3)**, 298-404.

[4] Johansson I, Killander F, Linderholm B, Hedenfalk I. (2014) Molecular profiling of male breast cancer- lost in translation? The International Journal of Biochemistry & Cell Biology, **53**, 526-535.

[5] Perou CM, Sorlie T, Eisen MB et al. (2000) Molecular portraits of human breast tumours. Nature **406(6797)**, 747-752.

[6] Yu X-F, Feng W-L, Miao LL, et al. (2013) The prognostic significance of molecular subtype for male breast cancer: a 10-year retrospective study. Breast, **22(5)**, 824-827.

[7] Nilsson C, Johansson I, Ahlin C, et al. (2013) Molecular subtyping of male breast cancer using alternative definitions and its prognostic impact. Acta Oncologica **52(1)**, 102-109.

[8] Shaaban AM, Ball GR, Brannan RA, et al. (2012) A comparative biomarker study of 514 matched cases of male and female breast cancer reveals gender-specific biological differences. Breast Cancer Research and Treatment **133(3)**, 949-958.

[9] Ge Y, Sneige N, Eltorky MA, et al. (2009) Immunohistochemical characterization of subtypes of male breast carcinoma. Breast Cancer Research **11**, R28.

[10] Johansson I, Nilsson C, Berglund P, et al. (2012) Gene expression profiling of primary male breast cancers reveals two unique subgroups and identifies N-acetyltransferase-1 (NAT1)as a novel prognostic biomarker. Breast Cancer Research **14(1)**, R31.

[11] Hammond ME, Hayes DF, Dowsett M, et al. (2010) American Society of Clinical Oncology/College of American Pathologists guidelines recommendations for immunohistochemical testing of estrogen and progesterone receptors in breast cancer (unabridged version). Archives of Pathology & Laboratory Medicine **134(7)**, e48-e72.

[12] Prat A, Cheang MCU, Martin M, et al. (2013) Prognostic significance of progesterone receptor-positive tumor cells within immunohistochemically defined luminal A breast cancer. Journal of Clinical Oncology **31(2)**, 203-209.

[13] Cheang MC, Chia SK, Voduc D, et al. (2009) ki67 index, HER2 status and prognosis of patients with luminal B breast cancer. Journal of the National Cancer Institute **101(10)**, 736-750.

[14] de Jong JS, van Diest PJ, van der Valk P, Baak JP. (2001) Expression of growth factors, growth factor receptors and apoptosis related proteins in invasive breast cancer: relation to apoptotic rate. Breast Cancer Research and Treatment **66(3)**, 201-208.

[15] Andre S, Pinto AE, Laranjeira C, et al. (2007) Male and female breast cancer-differences in DNA ploidy, p21 and p53 expression reinforce the possibility of distinct pathways of oncogenesis. Pathobiology **74(6)**, 323-327.

[16] Wolff AC, Hammond ME, Hicks DG, et al. (2014) Recommendations for human epidermal growth factor receptor 2 testing in breast cancer: American Society of Clinical Oncology/College of American Pathologists clinical practice guideline update. Journal of Clinical Oncology **31(31)**, 3997-4013.

[17] Cheang MC, Voduc D, Bajdik C, et al. (2008) Basal-like breast cancer defined by five biomarkers has superior prognostic value than triple-negative phenotype. Clinical Cancer Research **14(5)**, 1368-1376.

[18] Hugh J, Hanson J, Cheang MC, et al. (2009) Breast cancer subtypes and response to docetaxel in node-positive breast cancer: use of an immunohistochemical definition in the BCIRG 001 trial. Journal of Clinical Oncology **27(8)**, 1168-1176.

[19] Kornegoor R, Verschuur-Maes A, Buerger H, et al. (2012) Immunophenotyping of male breast cancer. Histopathology **61(6)**, 1145-1155.

[20] Abreu MH, Gomes M, Menezes F, et al. (2015) CYP2D6*4 polymorphism: a new marker of response to hormonotherapy in male breast cancer? Breast **24(4)**, 481-486.

[21] Anderson WF, Jatoi I, Tse J, Rosenberg PS. (2010) Male breast cancer: a population-based comparison with female breast cancer. Journal of Clinical Oncology **28(2)**, 232-239.

[22] Giordano SH, Cohen DS, Budzar AU, et al. (2004) Breast carcinoma in men: a population-based study. Cancer **101(1)**, 51-57.

[23] Sorlie T, Perou CM, Tibshirani R, et al. (2001) Gene expression patterns of breast carcinomas distinguish tumor subclasses with clinical implications. Proceedings of the Natlonal Academy of Sciences of the USA **98(19)**, 10869-10874.

[24] van't Veer LJ, Dai H, van de Vijver MJ, et al. (2002) Gene expression profiling predicts clinical outcome of breast cancer. Nature **415(6871)**, 530-536.

[25] Weigelt B, Geyer FC, Reis-Filho JS. (2010) Histological types of breast cancer: how special are they? Molecular Oncology **4(3)**, 192-208.

[26] Nilsson C, Koliadi A, Johansson I, et al. (2013) High proliferation is associated with inferior outcome in male breast cancer patients. Modern Pathology **26(1)**, 87-94.

[27] Sánchez-Muñoz A, Román-Jobacho A, Pérez-Villa L, et al. (2012) Male breast cancer: immunohistochemical subtypes and clinical outcome characterization. Oncology **83(4)**, 228-233.

[28] Polley MY, Leung SC, McShane LM, et al. (2013) An international Ki67 reproducibility study. Journal of National Cancer Institute **105(24)**, 1897-906.

[29] Wang-Rodriguez J, Cross K, Gallagher S, et al. (2002) Male breast carcinoma: correlation of ER, PR, ki-67, HER2- Neu and p53 with treatment and survival, a study of 65 cases. Modern Pathology **15(8)**, 855-861.

[30] Rayson D, Erlichman C, Suman VJ, et al. (1998) Molecular markers in male breast carcinoma. Cancer **83(9)**, 1947-1955.

[31] Foerster R, Foerster FG, Wulff V et al. (2011) Matched-pair analysis of patients with female ad male breast cancer: a comparative analysis. BMC Cancer **11**, 335.

[32] Cutili B, Le-Nir CC, Serin D, et al. (2010) Male breast cancer. Evolution of treatment and prognostic factors. Analysis of 489 cases. Critical Reviews in Oncology Hematology **73(3)**, 246-254.

## 2.2 Recurrence

Cancer recurrence is an important but missing variable in national cancer registries and the majority of hospitals report incomplete information for more than half of their patients [73].

In MBC patients, there is no study that specifically addressed recurrence, and the majority of data come from retrospective epidemiologic series, that covered prolonged time periods, summarized in Table 2.4.

As we can see, relapse was documented in a large range of values, between 7.8 and 60.9%, mainly in distant sites. In FBC patients, and for many years now there is a clear perception that patterns of relapse influence prognosis and correlate with type of tumour and adjuvant treatments performed [85].

In luminal tumours (HR positive), more than half of all recurrences occur 6 years or more after diagnosis, particularly following 5 years of adjuvant endocrine therapy, with an annual distant recurrence risk after this therapy of 1-4%, depending on the extent of initial disease [86]. Late relapses in HR positive BC represent a significant clinical challenge as there is limited understanding about the underlying mechanisms of hormone-resistance and late relapse. In HER2 negative luminal tumours recurrence risk is partitioned by tumour grade. Low grade cases have very low early risk, but 20% of patients fall-off in 10 or more years after diagnosis. Higher-grade cases have risk over more than 20 years [87]. The risk of locoregional relapse was not related with type of breast surgery (conservative vs mastectomy) [88].

The luminal/HER2+ and HER2 enriched tumours (HR negative) are associated with higher rates of brain, liver and lung metastasis [85]. The HER2 disease displays also different patterns of relapse and metastatic spread depending on HR status, with a median relapse free survival of 19.5 months after surgery in HR negative patients compared to 32.0 months in HR positive patients. Younger age, stage III and no expression of ER were independent risk factors for relapse in these patients [89]. HER2+ tumours seem to recur more locally than triple negatives after conservative surgery, with no differences in local recurrence after mastectomy between these BC subtypes [88, 90]. The effect of HER2 status on distant recurrence in early stage BC differs according to recurrence site, ER expression and time [89]. For brain and pleura recurrences, the effect of HER2 depends on ER status in ways that significantly changed over time. For bone recurrences, the effect of HER2 does not depend on ER status but changes significantly over time. For liver and distant lymph node recurrences, there is a significant effect of HER2 status that does not change with time or ER status. For lung recurrences, rates do not significantly vary with HER2 status [89]. When they occur, severe bone metastasis and massive hepatic metastasis are independent risks for early relapse [91].

Table 2.4: MBC studies with recurrence information described

| Reference | n | Event (recurrence) | 5-year DFS | Recurrence type |
|---|---|---|---|---|
| Yoney A, et al. 2009 [74] | 39 | 16 (41%) | 65.8% | Local (5.1%)<br>Distant (35.9%) |
| Cutuli B, et al. 2010 [36] | 489 | 142 (29%) | 89% | Local (6.6%)<br>Axillar (5.3%)<br>Distant (22.5%): bone (29%), lung (24%) |
| Liukkonen S, et al. 2010 [75] | 58 | 14 (24.1%) | ——— | Local (40%)<br>Distant (60%) |
| Liu T, et al. 2011 [76] | 87 | 41 (47.1%) | 66.3% | Local (8%)<br>Distant (39.1%) |
| Selcukbiricik F, et al. 2012 [77] | 86 | 39 (45.3%) | 72.4% | Locoregional (17.4%)<br>Distant (34.1%) |
| de Ieso PB, et al. 2012 [78] | 63 | 18 (29%) | ———- | Local (1.3%)<br>Locoregional (4%)<br>Distant (22%) |
| Eryilmaz MA, et al. 2012 [79] | 25 | 5 (20%) | 49% | Local (20%)<br>Distant (80%) |
| Arslan UY, et al. 2012 [80] | 118 | 38 (32.2%) | 60% | Multiple (28.9%)<br>Bone (28.9%), Locoregional (21.1%),<br>Lung(15.8%), Liver (2.6%), Brain (2.6%) |
| Chen X, et al. 2013 [1] | 150 | 87 (58%) | 65.6% | ———*  |
| Kwong A, et al. 2014 [24] | 132 | ——— | 91.3% | ——— |
| Foerster R, et al. 2014 [81] | 41 | 25 (60.9%) | ———- | Lung (51.2%)**<br>Bone (56.1%)**<br>Liver (17.1%)** |
| Aggarwal A, et al. 2014 [82] | 51 | 4 (7.8%) | ———- | ——— |
| Iorfida M, et al. 2014 [83] | 99 | 43 (43.4%) | 75.5% | ——— |
| Masci G, et al. 2015 [84] | 87 | 26 (29.8%) | 50% | Local (7.7%)<br>Distant (92.3%) |

*Distant recurrence more frequent than in women contrarily than local recurrence;
**Analyses included metastatic patients at diagnosis and recurrent patients;
———-Not described; DFS - Disease-Free Survival.

Patients with HER2+/HR- tumours are more likely to present with high histologic grade and higher stages than those with HER2+/HR+. The first ones are less likely to experience first recurrence in bone but more in brain, and this lower risk of bone involvement persists when adjusted for age, stage and adjuvant trastuzumab and when first and subsequent sites of recurrence were both considered [92]. Patients with HER2+/HR- disease have a significantly increase hazard of early (0-2 and 2-5 years), but not late death (>5 years) compared to HER2+/HR+ [92].

Patients with triple negative tumours have an increased likelihood of distant recurrence and death within 5 years of diagnosis. Risk of distant relapse peaks at 3 years with a rapidly declined thereafter [93]. An 8-years follow-up study revealed that locoregional and distant metastases were present in 25% of the patients, with a specific BC-mortality higher than 75% [94]. These tumours are also associated with a significantly higher rate of brain, lung and distant nodal metastasis but a significantly lower rate of liver and bone metastasis comparing with other BC-subtypes [85]. Independent risk factors for recurrence included increased tumour size, positive nodal status, advanced stage and type of chemotherapy (adjuvant vs neoadjuvant) [94]. Tumour size was responsible for recurrence despite lack of involvement of lymph nodes [95]. Konigsberg et al. [96] used a cut-off of age (65 years) to analyse the patterns of recurrence in triple negative patients. Distant visceral metastases occurred significantly more than bone metastases in both age groups. However, local recurrences, bone and secondary lymph node metastases were more frequent in younger patients [96]. The elderly received significantly less chemotherapy than younger patients.

According to previous section, female BC subgroups are not linearly replicated in male and even concerning adjuvant treatments, comparative studies showed inconsistent results between genders, even with some prognostic factors controlled, as see in Table 2.5.

From these studies we could state that MBC patients seem to underwent less lumpectomy, with distinct axillar approaches, and radiotherapy. However, breast-conservative surgery is acceptable in male because they have, commonly, unicentric disease and are amenable to adjuvant radiotherapy [100, 101] and sentinel lymph node biopsy is also a valid option for nodal evaluation in male with clinically node-negative disease [13, 102]. The use of radiotherapy in MBC patients, with regional disease changed over time with an increase tendency to its use [103].

Table 2.5: Comparison of BC first treatments between genders

| Reference | n, male | Matched factors | Surgery type | CT | ET | RT |
|---|---|---|---|---|---|---|
| Scott-Conner CE, et al. 1999 [21] | 4755 | Age diagnosis; Ethnicity; Stage. | ——— more mastectomies | Less | ——- | More after mastectomy Less after lumpectomy |
| Wang J, et al. 2009 [97] | 151 | NA | ——— * more mastectomies | Less | Less | Less |
| Nilsson C, et al. 2011 [4] | 99 | Age diagnosis; Year of diagnosis. | Equal more mastectomies | Equal | Equal | Less After mastectomy was equal |
| Foerster R, et al. 2011 [22] | 108 | Age; Grade; N status; Stage; HR status; HER2 status; Year of diagnosis. | Equal more mastectomies | Less | Equal | Less |
| Greif JM, et al. 2012 [5] | 13457 | NA | ——— more mastectomies | Equal | Equal | Less |
| Cloyd JM, et al. 2013 [98] | 6039 | Age; Race; Stage; Grade; Year of diagnosis. | ——— ** more mastectomies | ——— | ——- | Less after lumpectomy |
| Iorfida M, et al. 2014 [83] | 99 | Age; Tumour size; N status; Grade; HR status; HER2 status; ki67 status; Year of surgery. | ——— more mastectomies | Equal | Equal | Less |
| Rushton M, et al. 2014 [99] | 72 | Age; Stage; Year of diagnosis. | ——— more mastectomies | Equal | Equal | Less |

**NOTE:** Tendencies described are related to MBC patients;
*11% of male did not perform surgery; male were less likely to perform sentinel node biopsy and more likely to undergo level 2 axillary surgery;
** Male were less likely to receive lymph node sampling;
CT: chemotherapy; ET: endocrine therapy; HR status: hormonal receptors status; N status: nodal status; RT: radiotherapy;
NA: not applied; ——— not described.

# References


[1] Chen X, Liu X, Zhang L, et al. (2013) Poorer survival of male breast cancer compared with female breast cancer patients may be due to biological differences. Japanese Journal of Clinical Oncology, **43(10)**, 954-963.

[2] Gnerlich JL, Deshpande AD, Jeffe DB, et al. (2012) Poorer survival outcomes for male breast cancer compared with female breast cancer may be attributable to in-stage migration. Annals of Surgical Oncology, **18(7)**, 1837-1844.

[3] Miao H, Verkooijen HM, Chia KS, et al. (2011) Incidence and outcome of male breast cancer: an international population-based study. Journal of Clinical Oncology, **29(33)**, 4381-4386.

[4] Nilsson C, Holmqvist M, Bergkvist L, et al. (2011) Similarities and differences in the characteristics and primary treatment of breast cancer in men and women- a population based study (Sweden). Acta Oncologica, **50(7)**, 1083-1088.

[5] Greif JM, Pezzi CM, Kimberg VS, et al. (2012) Gender differences in breast cancer: analysis of 13000 breast cancer in men from the national cancer data base. Annals of Surgical Oncology, **19(10)**, 3199-3204.

[6] Nahleh ZA, Srikantiah R, Safa M, et al. (2007) Male breast cancer in the Veterans Affairs population: a comparative analysis. Cancer, **109(8)**, 141-147.

[7] El-Tamer MB, Komenaka IK, Troxel A, et al. (2004) Men with breast cancer have better disease specific survival than women. Archives of Surgery, **139(10)**, 1079-1082.

[8] Marchal F, Salou M, Marchal C, et al. (2009) Men with breast cancer have same disease-specific and event-free survival as women. Annals of Surgical Oncology, **16(4)**, 972-978.

[9] Anderson WF, Althuis MD, Brinton LA, Devesa SS. (2004) Is male breast cancer similar or different tan female breast cancer? Breast cancer Research and Treatment, **83(1)**, 77-86.

[10] Hill TD, Khamis HJ, Tyczynski JE, Berkel HJ. (2005) Comparison of male and female breast cancer incidence trends, tumor characteristics, and survival. Annals of Epidemiology, **15(10)**, 773-780.

[11] Macdonald G, Paltiel C, Olivotto IA, Tydesley S. (2005) A comparative analysis of radiotherapy use and patient outcome in males and females with breast cancer. Annals of Oncology, **16(9)**, 1442-1448.

[12] Meijer-van Gelder ME, Look MP, Bolt-de Vries J, et al. (2001) Clinical relevance of biologic factors in male breast cancer. Breast Cancer Research and Treatment, **68(3)**, 249-260.

[13] Boughey JC, Bedrosian I, Meric-Bernstam F, et al. (2006) Comparative analysis of sentinel lymph node operation in male and female breast cancer patients. Journal of American College of Surgeons, **203(4)**, 475-480.

[14] Giordano SH, Cohen DS, Budzar AU, et al. (2004) Breast carcinoma in men: a population based study. Cancer, **101(1)**, 51-57.

[15] Janssen-Heijnen ML, Houterman S, Lemmens VE, et al. (2005) Prognosis impact of increasing age and co-morbidity in cancer patients: a population-based approach. Critical Reviews in Oncology Hematology, **55(3)**, 231-240.

[16] Fentiman IS, Fourquet A, Hortobagyi GN. (2006) Male breast cancer. Lancet, **367(9510)**, 595-604.

[17] Sachs MD. (1941) Carcinoma of the male breast. Radiology, **37(4)**, 458-467.

[18] Cutuli B, Lacroze M, Dilhuydy JM, et al. (1995) Male breast cancer: results of the treatments and prognostic factors in 397 cases. European Journal of Cancer, **31A(12)**, 1960-1964.

[19] Ribeiro GG, Swindell R, Harris M, et al. (1996) A review of the management of the male breast carcinoma based on an analysis of 420 treated cases. Breast, **5(3)**, 141-146.

[20] Joshi MG, Lee AK, Loda M, et al. (1996) Male breast carcinoma: an evaluation of prognostic factors contributing to a poorer outcome. Cancer, **77(3)**, 490-498.

[21] Scott-Conner CEH, Jochimsen PR, Menck HR, Winchester DJ. (1999) An analysis of male and female breast cancer treatment and survival among demographically identical pairs of patients. Surgery, **126(4)**, 775-780.

[22] Foerster R, Foerster FG, Wulff V, et al. (2011) Matched-pair analysis of patients with female and male breast cancer: a comparative analysis. BMC Cancer, **11**, 335.

[23] Shaaban AM, Ball GR, Brannan RA, et al. (2012) A comparative biomarker study of 514 matched cases of male and female breast cancer reveals gender-specific biological differences. Breast Cancer Research and Treatment, **133(3)**, 949-958.

[24] Kwong A, Chau WW, Mang OWK, et al. (2014) Male breast cancer: a population-based comparison with female breast cancer in Hong Kong, Southern China:1997-2006. Annals of Surgical Oncology, **21(4)**, 1246-1253.

[25] Donegan WL, Redlich PN, Lang PJ, Gall MT. (1998) Carcinoma of the breast in males: a multiinstitutional survey. Cancer, **83(3)**, 498-509.

[26] Joshi MG, Lee AK, Loda M, et al. (1996) Male breast cancer: an evaluation of prognostic factors contributing to a poorer outcome. Cancer, **77(3)**, 490-498.

[27] Borgen PI, Wong GY, Vlamis V, et al. (1992) Current management of male breast cancer. A review of 104 cases. Annals of Surgery, **215(5)**, 451-457.

[28] Yap HY, Tashima CK, Blumenschein GR, Eckles NE. (1979) Male breast cancer: a natural history study. Cancer, **44(2)**, 748-754.

[29] Salvadori B, Saccozzi R, Manzari A, et al. (1994) Prognosis of breast cancer in males: an analysis of 170 cases. European Journal of Cancer, **30A(7)**, 930-935.

[30] Vetto J, Jun SY, Paduch D, et al. (1999) Stages at presentation, prognostic factors, and outcome of breast cancer in males. American Journal of Surgery, **177(5)**, 379-383.

[31] Williams WL Jr, Powers M, Wagman LD. (1996) Cancer of the male breast: a review. Journal of the National Medical Association, **88(7)**, 439-443.

[32] Guinee VF, Olsson H, Moller T, et al. (1993) The prognosis of breast cancer in males. A report of 335 cases. Cancer, **71(1)**, 154-161.

[33] Herman K, Lobaziewicz W, Skotnicki P, et al. (2000) Male breast cancer. Does the prognosis differ compared to female? Neoplasma, **47(3)**, 191-195.

[34] Uppsala-Orebro. (2010) Regional Oncologic Center U, Sweden. Register for breast cancer.

[35] Nilsson C, Johansson I, Ahlin C, et al. (2013) Molecular subtyping of male breast cancer using alternative definitions and its prognostic impact. Acta Oncologica, **52(1)**, 102-109.

[36] Cutuli B, Le-Nir CC, Serin D, et al. (2010) Male breast cancer. Evolution of treatment and prognostic factors. Analysis of 489 cases. Critical Reviews in Oncology Hematology, **73(3)**, 246-254.

[37] Johansson I, Nilsson C, Berglund P, et al. (2012) Gene expression profiling of primary male breast cancers reveals two unique subgroups and identifies N-acetyltransferase-1 (NAT1) as a novel prognostic biomarker. Breast Cancer Research, **14(1)**, R31.

[38] Chavez-MacGregor M, Clarke CA, Lichtensztajn D, et al. (2013) Male breast cancer according to tumor subtype and race: a population based study. Cancer, **119(9)**, 1611-1617.

[39] Perou CM, Sorlie T, Eisen MB et al. (2000) Molecular portraits of human breast tumors. Nature, **406(6797)**, 747-752.

[40] Johansson I, Killander F, Linderholm B, Hedenfalk I. (2014) Molecular profiling of male breast cancer- lost in translation? The International Journal of Biochemistry & Cell Biologyl, **53**, 526-535.

[41] Cheang MC, Voduc D, Bajdik C, et al. (2008) Basal-like breast cancer defined by five biomarkers has superior prognostic value than triple-negative phenotype. Clinical Cancer Research, **14(5)**, 1368-1376.

[42] Hugh J, Hanson J, Cheang MC, et al. (2009) Breast cancer subtypes and response to docetaxel in node-positive breast cancer: use of an immunohistochemical definition in the BCIRG 001 trial. Journal of Clinical Oncology, **27(8)**, 1168-1176.

[43] Cheang MC, Chia SK, Voduc D, et al. (2009) ki67 index, HER2 status and prognosis of patients with luminal B breast cancer. Journal of the National Cancer Institute, **101(10)**, 736-750.

[44] Kaufmann M, Pusztai L, Biedenkopf Expert Panel Members. (2011) Use of standard markers and incorporation of molecular markers into breast cancer therapy: consensus recommendations from an international expert panel. Cancer, **117(8)**, 1575-1582.

[45] Goldhirsch A, Winer EP, Coates AS, et al. (2013) Personalizing treatment of women with early breast cancer: highlights of the St Gallen international expert consensus on the primary therapy of early breast cancer. Annals of Oncology, **24(9)**, 2206-2223.

[46] Prat A, Cheang MCU, Martin M, et al. (2013) Prognostic significance of progesterone receptor-positive tumor cell within immunohistochemically defined luminal A breast cancer. Journal of Clinical Oncology, **31(2)**, 203-209.

[47] Nielsen TO, Hsu FD, Jensen K, et al. (2004) Immunohistochemical and clinical characterization of the basal-like subtype of invasive breast carcinoma. Clinical Cancer Research, **10(16)**, 5367-5374.

[48] Carey LA, Perou CM, Livasy CA, et al. (2006) Race, breast cancer subtypes and survival in the Carolina Breast Cancer Study. Journal of the American Medical Association, **295(21)**, 2492-2502.

[49] Hu Z, Fan C, Oh DS, et al. (2006) The molecular portraits of breast tumors are conserved across microarrays platforms. BMC Genomics, **7**, 96.

[50] Parker JS, Mullins M, Cheang MCU, et al. (2009) Supervised risk predictor of breast cancer based on intrinsic subtypes. Journal of Clinical Oncology, **27(8)**, 1160-1167.

[51] Sorlie T, Perou CM, Tibshirani R, et al. (2001) Gene expression patterns of breast carcinomas distinguish tumor subclasses with clinical implications. Proceedings of the National Academy of Sciences of the USA, **98(19)**, 10869-10874.

[52] Sorlie T, Tibshirani R, Parker J, et al. (2003) Repeated observation of breast tumor subtypes in independent gene expression data sets. Proceedings of the National Academy of Sciences of the USA, **100(14)**, 8418-8423.

[53] Ciocca V, Bombonati A, Gatalica Z, et al. (2006) Cytokeratin profiles of male breast cancers. Histopathology, **49(4)**, 365-370.

[54] Ge Y, Sneige N, Eltorky MA, et al. 2009. Immunohistochemical characterization of subtypes of male breast carcinoma. Breast Cancer Research, **11**, R28.

[55] Kornegoor R, Verschuur-Maes AH, Buerger H, et al. (2012) Molecular subtyping of male breast cancer by immunohistochemistry. Modern Pathology, **25(3)**, 398-404.

[56] Sanchez-Muñoz A, Román-Jobacho A, Pérez-Villa L, et al. (2012) Male breast cancer: immunohistochemical subtypes and clinical outcome characterization. Oncology, **83(4)**, 228-233.

[57] Yu XF, Feng WL, Miao LL, et al. (2013) The prognostic significance of molecular subtype for male breast cancer: a 10-year retrospective study. The Breast, **22(5)**, 824-827.

[58] Cardoso F, Barlett J, Slaets L, et al. (2015) Characterization of male breast cancer: first results of the EORTC10085/TBCRC/BIG/NABCG international male BC program. Cancer Res, **75** ,S6-05.

[59] Anderson WF, Jatoi I, Tse J, Rosemberg PS. (2010) Male breast cancer: a population-based comparison with female breast cancer. Journal of Clinical Oncology, **28(2)**, 232-239.

[60] Muir D, Kanthan R, Kanthan SC. (2003) Male versus female breast cancers. A population-based comparative immunohistochemical analysis. Archives of Pathology & Laboratory Medicine, **127(1)**, 36-41.

[61] Johansson I, Nilsson C, Berglund P, et al. (2011) High-resolution genomic profiling of male breast cancer reveals differences hidden behind the similarities with female breast cancer. Breast Cancer Research and Treatment, **129(3)**, 747-760.

[62] Weber- Chappuis K, Bieri-Burger S, Hurlimann J. (1996) Comparison of prognostic markers detected by immunohistochemistry in male and female breast carcinomas. European Journal of Cancer, **32A(10)**, 1686-1692.

[63] Rakha EA, El-Sayed ME, Green AR, et al. (2007) Breast carcinoma with basal differentiation: a proposal for pathology definition based on basal cytokeratin expression. Histopathology, **50(4)**, 434-438.

[64] Eisinger F, Jacquemier J, Charpin C, et al. (1998) Mutations at BRCA1: the medullary breast carcinoma revisited. Cancer Research, **58(8)**, 1588-1592.

[65] Foulkes WD, Stefansson IM, Chappuis PO, et al. (2003) Germline BRCA1 mutations and a basal epithelial phenotype in breast cancer. Journal of the National Cancer Institute, **95(19)**, 1482-1485.

[66] Friedman LS, Gayther SA, Kurosaki T, et al. (1997) Mutation analysis of BRCA1 and BRCA2 in a male breast cancer population. American Journal of Human Genetics, **60(2)**, 313-319.

[67] Stratton MR, Ford D, Neuhasen S, et al. (1994) Familial male breast cancer is not linked to the BRCA1 locus on chromosome 17q. Nature Genetics, **7(1)**, 103-107.

[68] Tai YC, Domchek S, Parmigiani G, Chen, S. (2007) Breast cancer risk among male BRCA1 and BRCA2 mutation carriers. Journal of the National Cancer Institute, **99(23)**, 1811-1814.

[69] Wolpert N, Warner E, Seminsky MF, et al. (2000) Prevalence of BRCA1 and BRCA2 mutations in male breast cancer patients in Canada. Clinical Breast Cancer, **1(1)**, 57-63.

[70] Bloom KJ, Govil H, Gattuso P, et al. (2001) Status of HER-2 in male and female breast carcinoma. American Journal of Surgery, **182(4)**, 389-392.

[71] Rudlowski C, Friedrichs N, Faridi A, et al. (2004) HER-2/neu gene amplification and protein expression in primary male breast cancer. Breast Cancer Research and Treatment, **84(3)**, 215-223.

[72] Dowsett M, Brtlett J, Ellis IO, et al. (2003) Correlation between immunohistochemistry (HercepTest) and fluorescence in situ hybridization (FISH) for HER-2 in 426 breast carcinomas from 37 centres. Journal of Pathology, **199(4)**, 418-423.

[73] In H, Bilimoria KY, Stewart AK, et al. (2014) Cancer recurrence: an important but missing variable in national cancer registries. Annals of Surgical Oncology, **21(5)**, 1520-1529.

[74] Yoney A, Kucuk A, Unsal M. (2009) Male breast cancer: a retrospective analysis. Cancer Radiothérapie, **13(2)**, 103-107.

[75] Liukkonen S, Saarto T, Maenpaa H, Sjostrom.mattson J. (2010) Male breast cancer: a survey at the Helsinki University Central Hospital during 1981-2006. Acta Oncologica, **49(3)**, 322-327.

[76] Liu T, Tong Z, He L, Zhang L. (2011) Clinicopathological characteristics and survival analysis of 87 male breast cancer cases. Breast Care, **6(6)**, 446-451.

[77] Selcukbiricik F, Tural D, Aydogan F, et al. (2012) Male breast cancer: 37-year data study at a single experience center in Turkey. Journal of Breast cancer, **16(1)**, 60-65.

[78] de Ieso PB, Potter AE, Le H, et al. (2012) Male breast cancer: a 30-year experience in South Australia. Asia-Pacific Journal of Clinical Oncology, **8(2)**, 187-193.

[79] Eryilmaz MA, Igci A, Muslumanoglu M, et al. (2012) Male breast cancer: a retrospective study of 15 years. Journal of the Balkan Union of Oncology, **17(1)**, 51-56.

[80] Arslan UY, Oksuzogiu B, Ozdemir N, et al. (2012) Outcome of non-metastatic male breast cancer: 118 patients. Medical Oncology, **29(2)**, 554-560.

[81] Foerster R, Schroeder L, Foerster F, et al. (2014) Metastatic male breast cancer. a retrospective cohort analysis. Breast care, **9(4)**, 267-271.

[82] Aggarwal A, Liu ML, Krasnow SH. (2014) Breast cancer in male veteran population: an analysis from VA cancer registry. Journal of Community and Supportive Oncology, **12(8)**, 293-297.

[83] Iorfida M, Bagnardi V, Rotmensz N, et al. (2014) Outcome of male breast cancer: a matched single-Institution Series. Clinical Breast Cancer, **14(5)**, 371-377.

[84] Masci G, Caruso M, Caruso F, et al. (2015) Clinicopathological and immunohistochemial characteristics in male breast cancer: a retrospective case series. Oncologist, **20(6)**, 586-592.

[85] Kennecke H, Yerushalmi R, Woods R, et al. (2010) Metastatic behavior of breast cancer subtypes. Journal of Clinical Oncology, **28(20)**, 3271-3277.

[86] Lim E, Metzger-Filho O, Winer EP. (2012) The natural history of hormone receptor-positive breast cancer. Oncology Journal, **26(8)**, 688-694.

[87] Esserman L. (2011) Biologic markers determine both the risk and the timing of recurrence in breast cancer. Breast Cancer Research and Treatment, **129(2)**, 607-616.

[88] Lowery A, Kell MR, Glynn RW et al. (2012) Locoregional recurrence after breast cancer surgery: a systematic review by receptor phenotype. Breast Cancer Research and Treatment, **133(3)**, 831-841.

[89] Hess KR, Esteva FJ. (2013) Effect of HER2 status on distant recurrence in early stage breast cancer. Breast Cancer Research and Treatment. **137(2)**, 449-455.

[90] Voduc K, Cheang MC, Tyldesley S. (2010) Breast cancer subtypes and the risk of local and regional relapse. Journal of Clinical Oncology, **28(10)**, 1684-1691.

[91] Park YH, Lee S, Cho EY, et al. (2010) Patterns of relapse and metastatic spread in HER2-overexpressing breast cancer according to estrogen receptor status. Cancer Chemotherapy and Pharmacology, **66(3)**, 507-516.

[92] Vaz-Luis I, Ottesen RA, Hughes ME, et al. (2012) Impact of hormone receptor status on patterns of recurrence and clinical outcomes among patients with human epidermal growth factor-2-positive breast cancer in the national comprehensive cancer network: a prospective cohort study. Breast Cancer Research, **14(5)**, R129.

[93] Dent R, Trudeau M, Pritchard KI, et al. (2007) Triple-negative breast cancer: clinical features and patterns of recurrence. Clinical Cancer Research, **13**, 4429-4434.

[94] Stewart L, Conant L, Gao F, Margenthaler JA. (2014) Predictive factors and patterns of recurrence in patients with triple negative breast cancer. Annals of Surgical Oncology, **21(7)**, 2165-2171.

[95] Pogoda K, Niwinska A, Murawska M, Pienkowski T. (2013) Analysis of pattern, time and risk factors influencing recurrence in triple-negative breast cancer patients. Medical Oncology, **30(1)**, 388.

[96] Konigsberg R, Pfeiler G, Klement T, et al. (2012) Tumor characteristics and recurrence patterns in triple negative breast cancer: a comparison between younger (<65) and elderly (>=65) patients. European Journal of Cancer, **48(16)**, 2962-2968.

[97] Wang J, Kollias J, Marsh C, Maddern G. (2009) Are males with early breast cancer treated differently from females with early breast cancer in Australia and New Zeland? Breast, **18(6)**, 378-381.

[98] Cloyd JM, Hernandez-Boussard T, Wapnir I. (2013) Poor compliance with breast cancer treatment guidelines in men undergoing breast-conserving surgery. Breast Cancer Research and Treatment, **139(1)**, 177-182.

[99] Rushton M, Kwong A, Visram H, et al. (2014) Treatment outcomes for male breast cancer: a single-center retrospective case-control study. Current Oncology, **21(3)**, 400-407.

[100] Golshan M, Rusby J, Dominguez F, Smith BL. (2007) Breast conservation for male breast carcinoma. Breast, **16(6)**, 653-656.

[101] Dietz JR, Partridge AH, Gemignani ML, et al. (2015) Breast cancer management updates. Young and older, pregnant, or male. Annals of Surgical Oncology, **22(10)**, 3219-3224.

[102] Flynn LW, Park J, Patil SM, Cody HS, Port ER. (2008) Sentinel lymph node biopsy is successful and accurate in male breast carcinoma. Journal of the American College of Surgeons, **206(4)**, 616-621.

[103] Fields EC, DeWitt P, Fisher CM, Rabinovitch R. (2013) Management of male breast cancer in the United States: a surveillance, epidemiology and end results analysis. 2013. International Journal of Radiation Oncology Biology and Physics, **87(4)**, 747-752.

## Related Articles

Article 3 - Patterns of Recurrence and Treatment in Male Breast Cancer: A Clue to Prognosis? (Unpublished) 2.2.1.
Article 4 - Predicting Breast Cancer Recurrence using Machine Learning Techniques: A Systematic Review (Unpublished) 2.2.2

# Patterns of Recurrence and Treatment in Male Breast Cancer: A Clue to Prognosis?

Miguel Henriques Abreu[a], Pedro Henriques Abreu[b], Noémia Afonso[a], Deolinda Pereira[a], Rui Henrique[c,d], Carlos Lopes[d]

[a]*Department of Medical Oncology, Portuguese Institute of Oncology of Porto, Rua António Bernardino de Almeida s/n, 4200-465, Porto Portugal*
[b]*Department of Informatics Engineering, Faculty of Sciences and Technology of Coimbra University and CISUC, Pólo II - Pinhal de Marrocos 3030-290, Coimbra Portugal*
[c]*Department of Pathology, Portuguese Institute of Oncology of Porto, Rua António Bernardino de Almeida s/n, 4200-465, Porto Portugal*
[d]*Department of Pathology and Molecular Immunology, Biomedical Sciences Institute of Abel Salazar, University of Porto, Rua Jorge Viterbo Ferreira, 228, 4050-313 Porto, Portugal*

---

**Abstract**

**Background:** Male breast cancer (MBC) patients seem to have inferior survival compared to female (FBC) ones, which is not fully explained by usual prognostic factors. Recurrence analysis could show differences in relapse patterns and/or in patients' approaches that justify these outcomes.

**Material/Methods:** Retrospective analysis of 111 MBC patients treated in a cancer center between 1990-2014. For each patient, three matched recurrent FBC patients were selected by: diagnosis' year, age (within 5 years), stage and tumors' type (only luminal-like were considered). Differences between cohorts were assessed by $X^2$ test and hierarchical clustering was performed to define subgroups according to local of relapse. Survival curves were calculated by Kaplan-Meier and compared using log-rank test. Statistical significance was defined as $p<0.05$.

**Results:** Groups were balanced according to age, histological grade, stage, expression of hormonal receptors and adjuvant treatments performed. Median time to recurrence was equivalent, 43 vs 48 months, $p=0.72$, with the majority of patients presented with distant metastases, $p=0.69$, mainly in bone, with more lung involvement in male, $p=0.003$. Male patients were more often proposed to symptomatic treatment (21.1% vs 4.4%, $p=0.02$). Survival was poorer for male, median: 5 years (95%CI:4.1-5.9 years)

vs 10 years (95%CI:7.8-12.2 years), $p<0.001$, and this tendency remained in the five cluster subgroups, that identified five patterns of relapse, $p=0.003$.

**Conclusions:** MBC patients had the worst survival, even after controlling important factors, namely the local of relapse. Palliative systemic treatment had favorable impact in prognosis and its frequently avoidance in male could justify the outcomes differences.

*Keywords:* MBC, FBC, recurrence, luminal tumors, prognosis

---

## 1 Introduction

Male breast cancer (MBC) is a rare disease, accounting for less than 1% of all breast cancer cases [1]. Recently and due to the increase in its incidence [2], some studies were published suggesting a worst outcome for male [1, 2, 3, 4, 5] compared to female (FBC) patients, justified by: comorbid conditions [6] and delay in the diagnosis in male patients [1], different age and stage at diagnosis [2, 4, 7] or even differences in tumors biology [5]. Despite the remarkable improvements in breast cancer (BC) characterization, accurate prediction of breast cancer behavior is often still difficult to achieve [8], particularly in luminal disease where the mechanisms of treatment resistance, late relapse and dormancy are still not well-understood [9, 10]. Recurrence plays an important rule in BC prognosis and although new therapies and approaches might decrease the incidence of relapse, compared to past records [11], the hazard of dying after metastization was equal for patients diagnosed between 1978-1984 and 1995-2003 in a population-based study [11] demonstrate that a special focus to advanced disease is still required.

In the absence of studies in MBC patients specifically designed to characterized recurrence the aim of this study was to compare patterns of relapse and patients approaches in two genders groups, trying to estimate its real effect in patients' outcomes.

## 2 Patients and Methods

Clinical records from MBC patients treated at Portuguese Institute of Oncology of Porto (IPO-Porto) between 1990-2014, maximum follow-up of 30 years (median of 5.5 years), were retrospectively reviewed looking for cancer relapse. From a total of 111

patients, recurrence was documented in 27 patients (24.3%). Four patients were lost in follow-up, considering 23 patients for the final analysis. For each patient, three matched recurrent female patients treated in the same Institution were selected. The variables used to randomly assigned matches were: year of diagnosis, age at diagnosis (within 5 years), stage and biological subtype (only luminal-like tumors were considered in order to obtain the necessary sample size and to compare more homogeneous groups). Luminal-like tumors were defined by immunohistochemistry classification, IHQ, as tumors that express estrogen receptor in more than 1% of their cells with or without expression of progesterone receptor using the same cutoff and without co-expression of HER2). The local ethics committee approved this study.

### *2.1 Statistical analysis*

Differences between matched cohorts were assessed by $X^2$ test to compare categorical variables. Hierarchical clustering was performed to define subgroups, according to local of relapse. For this purpose, the linkage algorithm with the Unweighted average distance (UPGMA) to compute the distance between cluster was used and, as all markers were categorical variables (present/absent), similarities between patients were achieved by the mahalanobis distance measure. This analysis was performed using statistical program R (`http://www.r-project.org`). At the end, the cophenetic correlation coefficient was calculated for the hierarchical cluster tree and a value of 0.96 was obtained (for a maximum of 1.0). Survival curves were calculated by the Kaplan-Meier method and compared using the log- rank test, considering survival from relapse, the time from the diagnosis of recurrence to death and overall survival as the time from diagnosis to death (from any cause). Statistical significance was defined as $p<0.05$.

SPSS for Windows version 19.0 (SPSS Inc., Chicago, IL, USA) was used for statistical calculations.

## 3 Results

The analyzed population comprised 23 MBC patients and 69 FBC patients. Characteristics of both cohort groups could be seen in Table 1. Groups were balanced accord-

ing to age, histological grade, stage, expression of hormonal receptors and adjuvant treatments performed.

Histological type presented some differences between groups, $p$=0.047, with no lobular tumors in MBC patients and no exclusive papillary in FBC patients. Tumor size showed also a slightly difference, $p$=0.06, with a tendency to more T3 tumors in FBC patients (18.8% vs 4.3%) and more T4 tumors in MBC patients (30.4% vs 7.2%). The type of surgery and adjuvant treatments performed, were equal ($p>0.05$). The regimens of chemotherapy were the same, $p$=0.30, composed by taxanes and antracyclines in 50% (male) vs 38.1% (female). On the other hand, the type of endocrine therapy was different, $p$=0.01, with 84.2% of male underwent only tamoxifen vs 45% female, with no cases of use of aromatase inhibitors in consecutive regimen after tamoxifen (switch regimen) in male and with more women that did only aromatase inhibitors (43.3% vs 15.8%).

## 4 Recurrence

Median time from diagnosis to recurrence were equal: male 43 months (2-144 months) and 48 months in female (3-323 months), $p$=0.72. The majority of patients were symptomatic (male: 17 (73.9%); female: 50 (72.5%), $p$= 0.89), and presented with elevation of Ca15.3 marker (male: 8 (88.9%); female: 57 (89.1%), $p$=0.98), Histological and/or cytological confirmation of metastasis were performed in 7 male (30.4%) and 36 female (52.2%), $p$=0.007 and showed IHQ changes related to primary tumor in 3 male patients, with 2 new negative progesterone receptor (40% of change) and 1 new negative estrogen receptor (20% of change) and in female 5 new cases of negative estrogen receptor (12.8% of change), and 11 new cases of negative progesterone receptor (36.7% of change).

Patients' recurrence was illustrated in Table 2. The locals described are not mutually exclusive. The majority of patients relapse at distant sites (male: 21 (91.3%) and in female: 61 (88.4%), $p$=0.69), mainly in bone, male: 5 (21.7%) and female: 30 (43.5%), with more lung metastasis registered in male, $p$=0.03. There was no brain or lymph node metastases in MBC patients at time of first relapse. Two female patients recurred for peritoneum and one for eye, described in Table 2 as other sites.

Table 1: Patients' characteristics

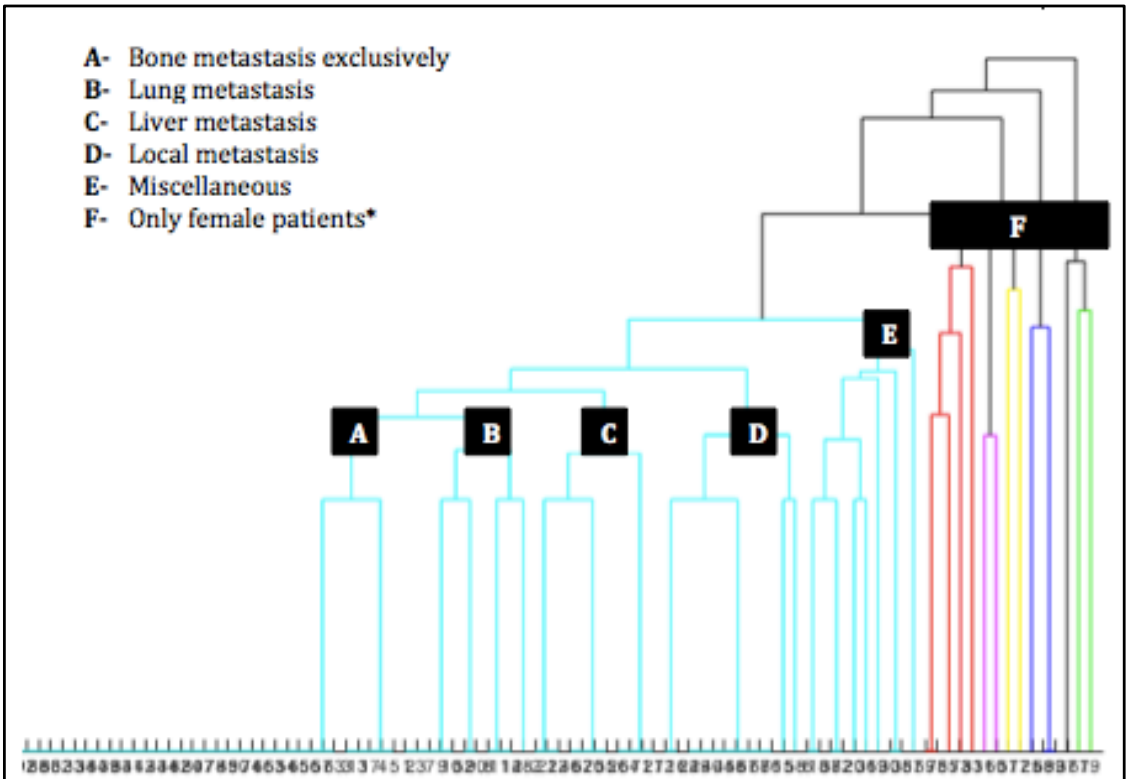
| **Characteristics** | **MBC patients** n=23 | **FBC patients** n=69 | ***p*** |
|---|---|---|---|
| **Median age at diagnosis, range** | 60 years (34-83 years) | 63 years (30-84 years) | 0.50 |
| **Post-menopausal status, n (%)** | ——— | 48 (52.2) | —- |
| **Histological type, n (%)** | | | |
| Ductal | 20 (87) | 47 (68.1) | 0.04 |
| Mix | 2 (8.7) | 13 (18.8) | |
| Other | 1 (4.3) | 9 (13.0) | |
| **Histological grade, n (%)** | | | |
| G1 | 2 (8.7) | 7 (10.1) | 0.51 |
| G2 | 16 (69.6) | 47 (68.1) | |
| G3 | 5 (21.7) | 15 (21.7) | |
| **T (size), n (%)** | | | |
| $>$2 cm | 16 (69.5) | 50 (72.4) | 0.06 |
| $\leq$2 cm | 7 (30.4) | 19 (27.5) | |
| **N (nodal status), n (%)** | | | |
| N0 | 6 (26.1) | 18 (26.5) | 0.40 |
| N+ | 17 (73.9) | 51 (73.5) | |
| **Stage, n (%)** | | | |
| I | 2 (8.7) | 6 (8.7) | 1.00 |
| II | 7 (30.4) | 21 (30.4) | |
| III | 14 (60.9) | 42 (60.9) | |
| **ER, n (%)** | | | |
| Positive | 23 (100) | 69 (100) | 1.00 |
| **PR, n (%)** | | | |
| Positive | 20 (87) | 62 (89.9) | 0.40 |
| Negative | 3 (13) | 7 (10.1) | |
| **HER2, n (%)** | | | |
| Negative | 23 (100) | 69 (100) | 1.00 |
| **Neoadjuvant treatment, n (%)** | | | |
| Yes | 3 (13.1) | 16 (23.2) | 0.14 |
| No | 20 (86.9) | 53 (76.8) | |
| **Surgery, n (%)** | | | |
| Mastectomy | 22 (95.7) | 56 (81.2) | 0.11 |
| Lumpectomy | 1 (4.3) | 13 (18.8) | |
| **Adjuvant chemotherapy, n (%)** | | | |
| Yes | 9 (39.1) | 42 (60.9) | 0.06 |
| No | 14 (60.9) | 27 (39.1) | |
| **Adjuvant radiotherapy, n (%)** | | | |
| Yes | 20 (87.0) | 53 (76.8) | 0.54 |
| No | 3 (13.0) | 16 (23.2) | |
| **Adjuvant endocrine therapy, n (%)** | | | |
| Yes | 19 (82.6) | 63 (91.3) | 0.32 |
| No | 4 (17.4) | 6 (8.7) | |

MBC: male breast cancer; FBC: female breast cancer;
ER: estrogen receptor; PR: progesterone receptor;
—-: not applicable.

Table 2: Type of recurrence by patients' cohort group

| **Recurrence** | **MBC patients** n=23 | **FBC patients** n=69 | ***p*** |
|---|---|---|---|
| **Local, n (%)** | 3 (13.0) | 10 (14.4) | 0.86 |
| **Regional Lymph nodes, n (%)** | 0 (0) | 4 (52.2) | —- |
| **Distant Lymph nodes, n (%)** | 0 (0) | 7 (10.1) | —- |
| **Bone, n (%)** | 12 (52.2) | 46 (66.7) | 0.21 |
| . Number of bones involved | | | |
| 1-5 | 6 (50) | 26 (56.5) | 0.81 |
| 5-10 | 2 (16.7) | 9 (19.6) | |
| $>$10 | 4 (33.3) | 11 (23.9) | |
| . Spine involvement | 11 (91.7) | 35 (77.8) | 0.28 |
| . Surgery | 2 (16.7) | 9 (19.6) | 0.82 |
| . Bisphosphonates | 5 (41.7) | 34 (73.9) | 0.03 |
| . Radiotherapy | 5 (41.7) | 28 (60.9) | 0.23 |
| **Lung, n (%)** | 9 (39.1) | 12 (17.4) | 0.03 |
| . Bilateral disease | 5 (50) | 7 (43.8) | 0.76 |
| . Pleural involvement | 4 (17.4) | 10 (14.5) | 0.70 |
| . Number of lesions | | | |
| $\leq$5 | 4 (57.1) | 5 (55.6) | 0.90 |
| $>$5 | 3 (42.9) | 4 (44.4) | |
| . Local treatment | 1 (4.3) | 0 (0) | 0.38 |
| **Liver, n (%)** | 3 (13.0) | 14 (20.3) | 0.44 |
| . Number of lesions | | | |
| $\leq$5 | 1 (50) | 3 (23.1) | 0.42 |
| $>$5 | 1 (50) | 10 (76.9) | |
| . Hepatic insufficiency | 1 (33.3) | 2 (15.4) | 0.47 |
| . Local treatment | 0 (0) | 0 (0) | 1.00 |
| **Brain, n (%)** | 0 (0) | 3 (4.3) | —- |
| . Parenchymal involvement | 0 (0) | 3 (100) | —- |
| . Leptomeningeal involvement | 0 (0) | 1 (33.3) | —- |
| . Number of lesions | | | |
| $=$1 | 0 (0) | 1 (33.3) | —- |
| $>$1 | 0 (0) | 2 (66.7) | |
| . Local treatment | 0 (0) | 3 (100) | —- |
| **Other sites, n (%)** | 0 (0) | 3 (4.3) | —- |
| **First line treatment** | | | |
| . Endocrine therapy | 11 (57.9) | 45 (66.2) | 0.04 |
| . Chemotherapy | 4 (21.1) | 20 (29.4) | |
| . Symptomatic treatment | 4 (21.1) | 3 (4.4) | |

—-: not possible to compare because there were no cases in one cohort group.




| Cluster Subgroup | Gender Group | Median overall survival, years | 95% CI, years |
|---|---|---|---|
| A | MBC patients | 10.0 | 3.4-16.6 |
| | FBC patients | 12.2 | 6.4-18.1 |
| B | MBC patients | 5.6 | 3.9-7.3 |
| | FBC patients | 14.0 | 14.0-14.0 |
| C | MBC patients | 2.0 | 0.9-3.2 |
| | FBC patients | 5.3 | 3.6-6.9 |
| D | MBC patients | 6.0 | 0.8-11.2 |
| | FBC patients | 9.8 | 4.8-14.8 |
| E | MBC patients | 6.3 | 4.6-8.1 |
| | FBC patients | 12.5 | 7.6-17.5 |

MBC: male breast cancer; FBC: female breast cancer; CI: confidence interval;
*Group F: not considered in the analysis.

Figure 1: Cluster subgroups according to local of relapse

Cluster analysis defined six different subgroups, according to local of relapse, please see Figure 1. One of them (Subgroup F), included the patients with lymph node metas-

tasis, brain metastasis and metastasis in other sites, that were all female patients, and because of that, it was not considered for analysis. The other five subgroups were composed by patients of both genders with, at least, one predominant site of metastasis: Subgroup A (5 male and 28 female), the major one, that included only patients with bone metastasis exclusively; Subgroup B (8 male and 3 female) that contained patients with lung metastasis predominantly; Group C (3 male and 8 female) with patients with liver metastasis; Group D (3 male and 9 female) with local recurrences and Group E (4 male and 6 female) that included a miscellaneous of patients, that included that ones with more than 2 sites of relapse.

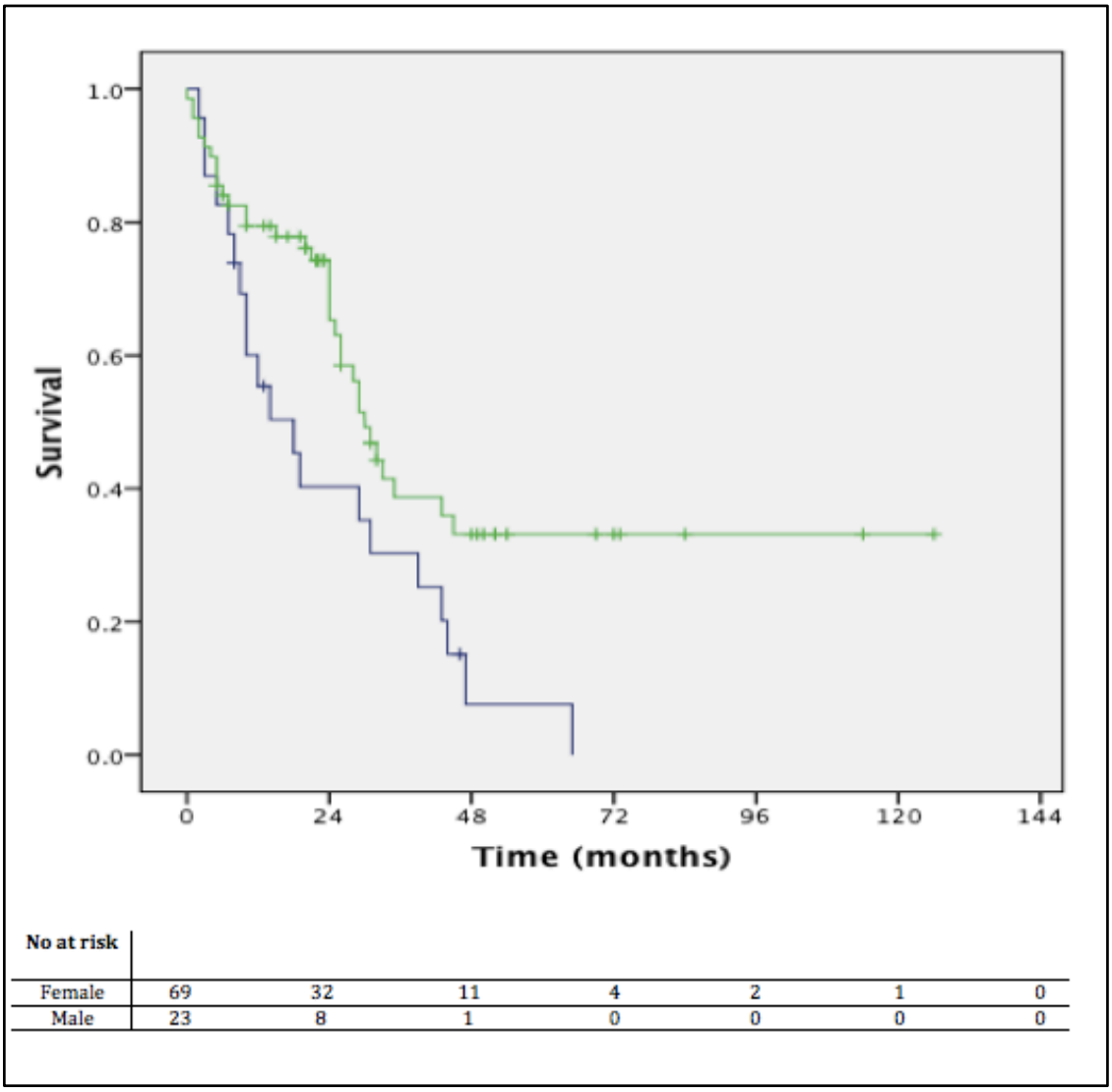
The good performance status in recurrence (ECOG: 1 in male: 10 (43.5%) and in female: 31 (44.9%), $p$=0.64, allowed palliative systemic treatments in 15 (79%) of male and 65 (95.6%) in female, with a median of lines of one in male and two in female, $p$=0.60. Male patients were more often proposed to symptomatic treatment at first diagnosis of recurrence than female (21.1% vs 4.4%,$p$=0.018).

## 5 Survival Analysis

Survival from recurrence diagnosis was worst in male: median 1 year (95% CI: 0-2.1 years), and for female median 2 years (95%CI: 1.6-2.4 years), $p$=0.004; for a HR of 1.9; as well as overall survival: for male median 5 years (95% CI: 4.1-5.9 years), and for female median 10 years (95% CI: 7.8-12.2 years), $p<0.001$, for a HR of 2.5, Figure 2 and 3. In first line, palliative systemic treatment instead of symptomatic one had a positive impact in overall survival in both groups (in male: 7 years vs 2 years and in female 10 years vs 3 years, $p$=0.013). Cluster analysis defined 5 subgroups with different prognosis, $p$=0.003, with male patients having worst prognosis in all subgroups than female patients, please see Figure 1.

## 6 Discussion

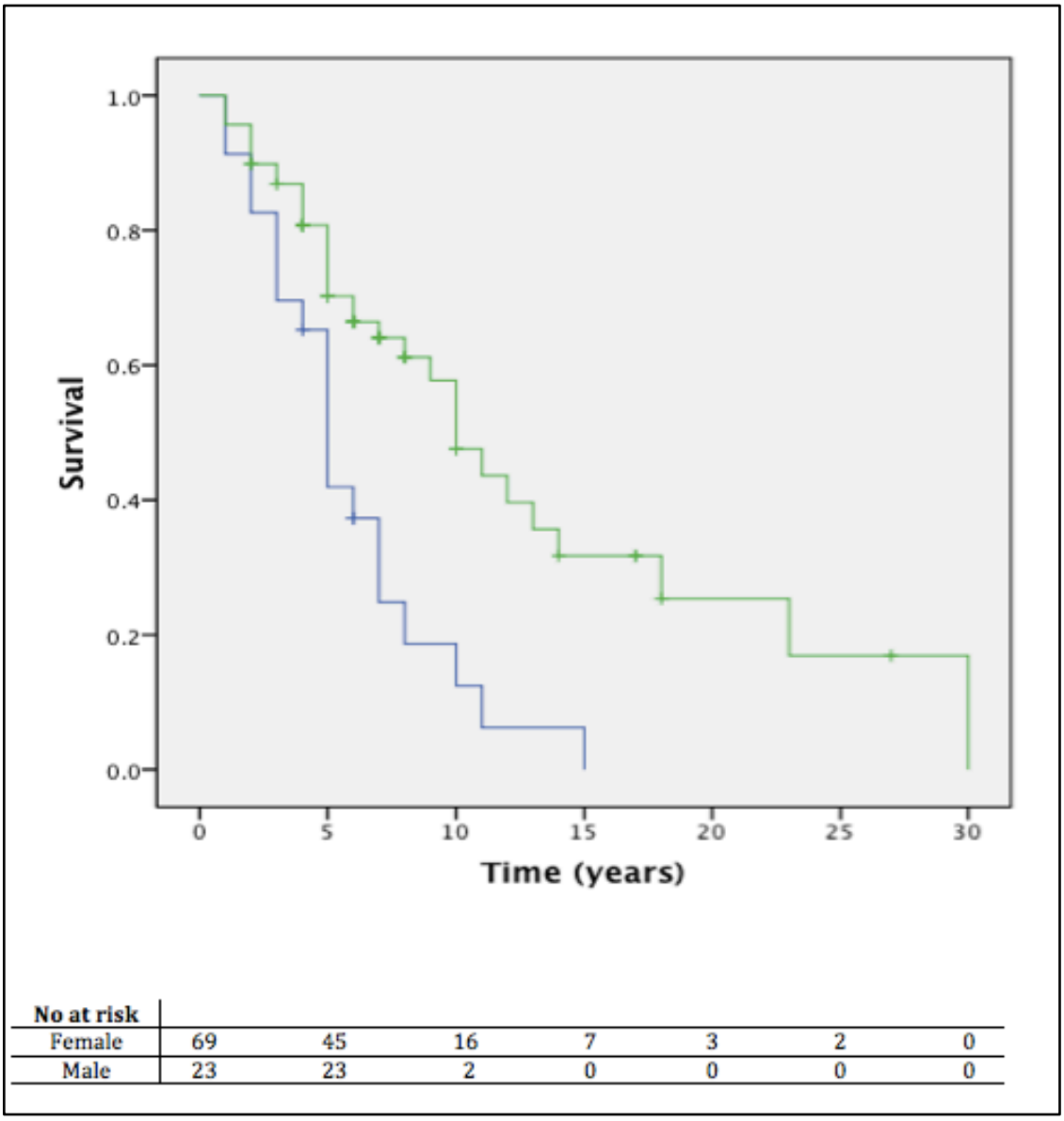
To our knowledge, this is the first study specifically designed to characterize relapse in MBC patients, with a comparison between two well-balanced matched gender cohorts.




| No at risk | | | | | | | |
|---|---|---|---|---|---|---|---|
| Female | 69 | 32 | 11 | 4 | 2 | 1 | 0 |
| Male | 23 | 8 | 1 | 0 | 0 | 0 | 0 |

Figure 2: Survival from recurrence diagnosis by cohort groups (Green Line – Female, Blue Line –Male)

With important prognostic factors controlled, namely the adjuvant treatments performed, we showed that MBC patients had a worse prognosis compared to female ones. Although recurred more often to lung, the main cause for poor prognosis seems to be the less aggressive treatment at recurrence diagnosis. Like in female, MBC patients that underwent palliative systemic treatments had better survival compared to that proposed only for symptomatic treatment.




| No at risk | | | | | | | |
|---|---|---|---|---|---|---|---|
| Female | 69 | 45 | 16 | 7 | 3 | 2 | 0 |
| Male | 23 | 23 | 2 | 0 | 0 | 0 | 0 |

Figure 3: Overall survival by cohort groups (Green Line – Female, Blue Line –Male)

Cancer recurrence is a critically important but missing variable in national cancer registries and the majority of hospitals report incomplete information for more than half of their patients [12]. In MBC, data from epidemiological series had a large range for recurrence rate, varying from 7.8% [13] to 60.9% [14] reflecting also a huge variation

in sample sizes, from 25 [14] to 489 [15] patients. Two of the largest published series [15, 16] reported, similar to us, a recurrence rate of 29.0% and 29.8% (142 events in 489 patients and 16 events in 91 patients). In FBC relapse, changes in IHQ markers, namely the non-expression of ER and/or PR [17] had a negative impact in prognosis, but in MBC patients this is not previous stated. In our series, cytological/histological confirmation of metastases in MBC patients was documented only in 7 cases, with PR variation in most of them. The real impact of these changes in male patients survival and the need to evaluate this in recurrence time must be evaluated in large series, having in mind that these patients have luminal-like tumors predominantly [2].

In the majority of cases, distant recurrence was the most prevalent [18, 19] documented mainly in bone [20, 21] as in our study. Like FBC patients, visceral metastases, in lung or liver, have a poor prognosis when compared to bone (exclusively) [22]. Brain metastasis at the time of first recurrence seems rare in MBC patients, described in 2 patients (7.32%) in Forester et al.[14] series. However, in this study, there were 15 patients (39%) with synchronous metastasis at diagnosis (stage IV) and 5 patients (16.7%) with HER2 positive disease, which could bias these results.

The positive survival impact of chemotherapy [14] and endocrine therapy [23, 24] in advanced disease was previously reported in male and was also corroborated in our study. Palliative chemotherapy strategies should follow female treatment guidelines [14] persisting some doubts about palliative endocrine therapy, namely about the need to combine or not the aromatase inhibitors with GnRH [23] or the rule of fulvestrant in these patients [24].

After controlling for important prognostic factors, we described a worse survival to MBC patients that are similar to other studies [1, 5], present however, a lower 5-year overall survival (65.2%) compared to some authors that reported more than 80% [15, 25]. This could be probably explained by the selection of our population that only contains patients that recurred and because of that have a poor prognosis.

In conclusion, this is the fist study in MBC setting that addressed a very important prognostic determinant: recurrence. However, even with all methodological precautions to avoid bias, there are also identifiable limitations in the study, strictly related to the rarity of this disease that requires a long period of time to compare patients outcomes,

with unavoidable changes in patients' approaches over the years. Because of that our findings need future confirmation, ideally in a large and prospective multicenter study focus also in comorbidities and causes of death.

## References


[1] Iorfida M, Bagnardi V, Rotmensz N, et al. (2014) Outcome of male breast cancer: a matched single-institution series. Clinical Breast Cancer **14(5)**, 371-377.

[2] Giordano SH, Cohen DS, Budzar AU, et al. (2004) Breast carcinoma in men: a population-based study. Cancer **101**, 51-57.

[3] Ribeiro GG, Swindell R, Harris M, et al. (1996) A review of the management of the male brest carcinoma based on an analysis of 420 treated patients. Breast **5(3)**, 141-146.

[4] Miao H, Verkooijen HM, Chia KS, et al. (2011) Incidence and outcome of male breast cancer: an international population-based study. Journal of Clinical Oncology **29(33)**, 4381-4386.

[5] Chen X, Liu X, Zhang L, et al. (2013) Poorer survival of male breast cancer compared with female breast cancer patients may be due to biological differences. Japanese Journal of Clinical Oncology **43(10)**, 954-963.

[6] Janssen-Heijnen ML, Houterman S, Lemmens VE, et al. (2005) Prognosis impact of increasing age and co-morbidity in cancer patients: a population-based approach. Critical Reviews in Oncology Hematology **55(3)**, 231-240.

[7] Goss PE, Reid C, Pintilie M, et al. (1999) Male breast carcinoma: a review of 229 patients who presented to the Princess Margaret Hospital during 40 years:1955-1996. Cancer **85(3)**, 629-639.

[8] Beca F, Santos R, Vieira D, et al. (2014) Primary relapse site pattern in women with triple-negative breast cancer. Pathology Research and Practice **210(9)**, 571-575.

[9] Esserman LJ, Moore, DH, Tsing PJ, et al. (2011) Biologic markers determine both the risk and the timing of recurrence in breast cancer. Breast Cancer Research and Treatment **129(2)**, 607-616.

[10] Lim E, Metzger-Filho O, Winer EP. (2012) The natural history of hormone receptor-positive breast cancer. Oncology **26(8)**, 688-694.

[11] Van den Hurk CJ, Eckel R, van de Poll-Franse LV et al. (2011) Unfavourable pattern of metastases in M0 breast cancer patients during 1978-2008: a population-based analysis of the Munich cancer registry. Breast Cancer Research and Treatment **128(3)**, 795-805.

[12] In H, Bilimoria KY, Stewart AK, et al. (2014) Cancer recurrence: an important but missing variable in national cancer registries. Annals of Surgical Oncology **21(5)**, 1520-1529.

[13] Aggarwal A, Liu ML, Krasnow SH. (2014) Breast cancer in male veteran population: an analysis from VA cancer registry. Journal of Community and Supportive Oncology **12(8)**, 293-297.

[14] Foerster R, Schroeder L, Foerster F, et al. (2014) Metastatic male breast cancer: a retrospective cohort analysis. Breast Care **9(4)**, 267-271.

[15] Cutuli B, Le-Nir CC, Serin D, et al. (2010) Male breast cancer. Evolution of treatment and prognostic factors. Analysis of 489 cases. Critical Reviews in Oncology Hematology **73(3)**, 246-254.

[16] Masci G, Caruso M, Caruso F, et al. (2015) Clinicopathological and immunohistochemical characteristics in male breast cancer: a retrospective case series. The oncologist **20(6)**, 586-592.

[17] Nishimura R, Osako T, Okumura Y, et al. (2011) Changes in the ER, PgR, HER2, p53 and Ki-67 biological markers between primary and recurrent breast cancer: discordance rates and prognosis. World journal of surgical oncology **9(1)**, 131.

[18] Idirisinghe PK, Thike AA, Cheok PY, et al. (2010) Hormone receptor and c-ERBB2 status in distant metastatic and locally recurrent breast cancer pathologic correlations and clinical significance. American Journal of clinical pathology **133(3)**, 416-429.

[19] Liu T, Tong Z, He L, Zhang, L. (2011) Clinicopathological characteristics and survival analysis of 87 male breast cancer cases. Breast Care **6(6)**, 446-451.

[20] Selcukbiricik F, Tural D, Aydogan F, et al. (2013) Male breast cancer: 37-year data study at a single experience center in Turkey. Journal of Breast Cancer **16(1)**, 60-65.

[21] Liukkonen S, Saarto T, Maenpaa H, Sjostrom-Mattson J. (2010) Male breast cancer: a survey at the Helsinki University Central Hospital during 1981-2006. Acta Oncologica **49(3)**, 322-327.

[22] Kim H, Choi D, Park W, et al. (2013) Prognostic factors for survivals from first relapse in breast cancer patients: analysis of decreased patients. Radiat Oncol J. **31(4)**, 222-227.

[23] Zagouri F, Sergentains TN, Koutoulidis V, et al. (2013) Aromatase inhibitors with or without gonadotropin-releasing hormone analogue in metastatic male breast cancer: a case series. British Journal of Cancer **108(11)**, 2259-2263.

[24] Agrawal A, Cheung KL, Robertson JF. (2007) Fulvestrant in advanced male breast cancer. Breast Cancer Research and Treatment **101**, 123.

[25] De leso PB, Potter AE, Le H, et al. (2012) Male breast cancer: a 30-year experience in South Australia. Asia-Pacific Journal of Clinical Oncology **8(2)**, 187-193.

# Predicting Breast Cancer Recurrence using Machine Learning Techniques: A Systematic Review

Pedro Henriques Abreu[a], Miriam Seoane Santos[a], Miguel Henriques Abreu[b], Bruno Andrade[a], Daniel Castro Silva[c]

[a]*Department of Informatics Engineering, Faculty of Sciences and Technology of Coimbra University and CISUC, Pólo II - Pinhal de Marrocos 3030-290, Coimbra Portugal*
[b]*Portuguese Institute of Oncology of Porto, Rua António Bernardino de Almeida s/n, 4200-465, Porto Portugal*
[c]*Faculty of Engineering of Porto University, Department of Informatics Engineering, Laboratory of Artificial Intelligence and Computer Science, Rua Dr. Roberto Frias, s/n 4200-465 Porto Portugal*

**Abstract**

**Background:** Recurrence is an important cornerstone in breast cancer behaviour, intrinsically related to mortality. In spite of its relevance, it is rarely recorded in the majority of breast cancer datasets, which makes research in its prediction more difficult.

**Objectives:** To evaluate the performance of machine learning techniques applied to the prediction of breast cancer recurrence. Material and Methods: Revision of published works that used machine learning techniques in local and open source databases between 1997 and 2014.

**Results:** The revision showed that it is difficult to obtain a representative dataset for breast cancer recurrence and there is no consensus on the best set of predictors for this disease. High accuracy results are often achieved, yet compromising sensitivity. The missing data and class imbalance problems are rarely addressed and most often the chosen performance metrics are inappropriate for the context.

**Discussion and Conclusions:** Although different techniques have been used, prediction of breast cancer recurrence is still an open problem. The combination of different machine learning techniques, along with the definition of standard predictors for breast cancer recurrence seem to be the main future directions to obtain better results.

*Keywords:* Breast Cancer Recurrence, Machine Learning Techniques

## 1 Introduction

Breast Cancer (BC) figures among the major causes of concern worldwide. According to the latest GLOBOCAN statistics [1], it was the second most frequently diagnosed cancer and the fifth cause of cancer mortality worldwide, responsible for 6.4% of all deaths.

The mortality associated to this pathology is mostly related to Metastization [2], the spread of cancer to other parts of the body remote from the breast, and Recurrence (or relapse), which describes cancer that reappears after treatment [3]. Being documented in 10-15% of all BC patients [4], recurrence assumes a pivotal importance in their prognosis. However, it is not as well studied as BC itself. Searching in Thomson Reuters [5] platform for research works with the expression “breast cancer” in the title yields more than 330.000 results. A similar search focused on recurrence yields only around 20.000 results (approximately 6%), obtained when the search terms are extended to “recurrence(s)”, “relapse(s)” and “metastasis (es)” (individually or in combination). These results can be partially explained by the fact that, for instance, none of the three major American cancer registries reports cancer recurrence information [6].

Besides the obvious implications of recurrence in mortality, BC patients also face serious treatment-related complications, which increases their risk of death from causes unrelated to breast cancer itself [7]. In this scenario, accurate prediction of breast cancer behaviour assumes an important role, since it aids clinicians in their decision-making process, enabling a more personalised treatment for patients. Some of the studies regarding cancer recurrence involve the use of statistical methodologies, or machine learning algorithms, which have a long history in cancer research[8][9][10]. This research work attempts to provide an overview of the prediction of BC recurrence using machine learning techniques. The challenge is to accurately predict recurrence events, within a binary outcome (yes/no). This challenge encompasses not only the choice of a good dataset (containing quality data) but also the selection of the most appropriate features, as well as the most advantageous algorithm.

The remainder of the paper is organized as follows: Section 2 covers the steps used by different authors to predict BC recurrence, highlighting the datasets, variables included in the reviewed studies, data mining algorithms, sampling strategies and eval-

uation metrics used. Section 3 depicts the analyzed works in more detail and Section 4 presents a discussion on the different works. Finally, some conclusions and future directions are discussed in Section 5.

## 2 Predicting Breast Cancer Recurrence phases and tasks

The most common processes to develop a data mining approach are Knowledge Discovery in Databases (KDD) [11], Sample, Explore, Modify, Model and Assess (SEMMA) [12] and CRoss-Industry Standard Process for Data Mining (CRISP DM) [13]. The first two are composed by five steps each; despite the different designations, their steps are generally equivalent [14]. The third strategy, CRISP DM, presents two novel steps which consist in "Business Understanding", where, after the Evaluation phase, the results are interpreted from a business perspective; and the "Deployment" step, where the final process achievements are somehow incorporated in a product/service (more related to a business perspective). Table 1 highlights the difference between these processes. As all BC recurrence studies analyzed use a KDD strategy, its steps are used to highlight the methodology followed by each of the works. In the following subsections, we explore some of the steps in the KDD approach and how they were addressed in the reviewed works.

Table 1: A comparison between KDD, SEMMA and CRISP-DM knowledge discovery processes [14]

| **KDD** | **SEMMA** | **CRISP-DM** |
|---|---|---|
| Pre-KDD | ———— | Business understanding |
| Selection | Sample | Data Understanding |
| Preprocessing | Explore | |
| Transformation | Modify | Data preparation |
| Data mining | Model | Modeling |
| Interpretation/Evaluation | Assessment | Evaluation |
| Post KDD | ———— | Deployment |

### *2.1 Selection*

This step consists ins the selection of a dataset and an appropriate set of features for knowledge extraction. The datasets can be publicly available (e.g. online) or they

may result from a collaboration between institutions and research teams, not available for the general public. Feature selection may be performed manually, or using variable selection algorithms. In particular for BC recurrence, 7/17 of the studies used manual selection (5 with the help of medical experts), while some of the others took advantage of well-known feature selection algorithms; for instance, Jonsdottir et al. [15].

The datasets and the number of patients used in the analysed studies are summarised in Table 2. From the 17 reviewed works, the majority uses available datasets (9 works). Among those, 4 works use the Wisconsin Prognostic Breast Cancer (WPBC) dataset and 3 use the Breast Cancer dataset, both available from UC Irvine Machine Learning Repository (UCI Repository) [16]. The remaining 2 datasets are available from van't Veer's study [17] and the widely known SEER database (US National Cancer Institute). The unavailable databases are collaborations with specialised Breast Cancer centers, registers or teaching hospitals, in several different countries (Sweden, Spain, California, Iceland, South Korea and Ljubljana). The Institute of Oncology in Ljubljana was the greatest contributor for BC recurrence studies, providing the data for 5 of the reviewed works.

Although the end-point for predicting BC recurrence is not defined for some cases, most of the datasets are associated with a specific time period for recurrence prediction (e.g. 4-5 years after the diagnosis, 10 years after surgery). Moreover, the great majority of the datasets suffered from a considerable class imbalance, with uneven cases of "recurrence'" versus "no-recurrence"', following a 30%-70% distribution. The most affected works are Mani et al. [18] and Razavi et al. (2007) [19], with a class distribution (recurrence/no-recurrence) of 10%-90% and 20%-80%, respectively. Going against this trend, 3 works (Sun et al. [20], Strumbelj et al. [21] and Tomczak [22]) perform their experiments on balanced datasets with approximately 50%-50% class distribution. The distribution is unknown for 2 works, Razavi et al. (2005) [23] and Jerez-Aragonés [24].

Concerning the feature selection, and for some years, many factors were linked to BC recurrence, namely age at diagnosis, size, stage and grade of tumor, involvement of lymph nodes, menopausal status, estrogen (ER) and progesterone receptors (PR) and HER2 pattern (Human Epidermal growth factor Receptor 2) [3]. Frequently, some of them are associated, given that tumors in younger patients (pre menopausal) tend to

be high grade, with a triple negative phenotype: without expression of ER, PR, and also HER2. Variables used in the BC recurrence prediction in the previously analysed studies were compared using three groups: Patient Characteristics, Tumor Characteristics and Treatments.

From the analysis of Table 2, it is important to highlight that there are many datasets with different origins (local and open source) used to deal with this problematic. Also, different authors used dissimilar varieties with a weak attention for the treatment followed by the patient (only 7/17 studies focused on this factor). Attending to Patient Characteristics, the majority of studies (10/17) identified age as an important predictor, followed by menopausal status (5/17). This last factor could not be totally independent from age as very young patients are also pre menopausal and very old ones are always post menopausal. All the studies considered Size as the main predictor in the Tumor Characteristics group, followed by Lymph Nodes involvement (16/17).

### *2.2 Data Cleaning and Preprocessing*

Data cleaning and preprocessing tasks are performed to reduce noise and increase the consistency of data. The preprocessing steps most addressed in the reviewed research works were Normalisation/Standardisation of data and Missing Data handling. Two simple ways of data preprocessing are Normalisation (Min-Max transformation) and Standardisation (Z-Score transformation) [25]. Normalisation refers to the feature scaling between it minimum and maximum values, while Standardisation rescales the features so that they follow a standard normal distribution (zero mean and unitary standard deviation). The objective of normalisation/standardisation is to make features with different scales and ranges of measurement (e.g. age, hemoglobin values) comparable, so that none has more influence than the others on classification task [26].

Missing Data (MD) can result from a huge variety of events and represents a common challenge in healthcare contexts[27].

Table 2: Datasets and Feature Selection used in the analyzed studies (NU - not used, ER -Estrogen Receptor, PR-Progesterone Receptor, ICD-International Classification of Diseases)

| Publications | Dataset (recurrence/no recurrence | Patient Characteristics | Tumor Characteristics | Treatments |
|---|---|---|---|---|
| Mani et al., 1997 [18] | Breast cancer center Orange, California (85/802) | Lymphedema | Tumor presence and its invasive nature, Size, Lymph nodes involvement, Stage | NU |
| Jerez-Aragonés et al., 2003 [24] | Hospital Clínico Universitario Malaga, Spain (1035 patients, not effective, distribution unknown) | Age, Menarchy age, Menopausal age, First pregnancy age/pregnancies number, Number of miscarriages | Size, Grade, Lymph nodes involvement, Expression of ER, PR, p53 accumulation, Ploidy, S-Phase | NU |
| Razavi et al., 2005, 2007 [23][19] | Swedish Regional BC Register (**2005**: 3949 patients, distribution unknown; **2007**: 3699, 664/3035) | Age | Size, Perigland growth, Lymph nodes involvement, Expression of ER, PR, S-Phase | NU |
| Sun et al., 2007 [20] | Publicly available microarray data (van't Veer et al. 2002 [17]) (46/51) | Age | Size, Vascular invasion, Lymphocytic infiltration, Expression of ER, PR, 70-gene profile | NU |
| Ryu, 2007 [28] | Breast Cancer Dataset (Available from UCI Repository) (85/201) | Age, Menopausal status | Location, Size, Grade, Lymph nodes involvement | Radiotherapy |
| Jonsdottir et al., 2008 [15] | Rose dataset University Hospital in Iceland (73/184) | Age, Comorbidities, CEA/CA-15.3 values | Clinical detectable, Histological type, Size, Local invasion, Inflammatory component, Lymphovascular invasion, Lymph nodes involvement, Metastasis (lung, bone), Expression of ER, PR, S-Phase | Chemotherapy (pre and post surgery), Hormonotherapy, Radiotherapy |
| Fan et al., 2010 [29] | SEER Public-Use Data 2005 (46.996.113 patients, not effective, distribution unknown) | Race, Age, Marital status | Behaviour defined by ICD-03, Location (laterality, breast regions), Size, Grade, Local invasion, Lymph nodes involvement | Surgery, Radiotherapy |

*Continued on next page*

Table 2 – *Datasets and Feature Selection used in the analyzed studies*

| Publications | Dataset (recurrence/no recurrence) | Patient Characteristics | Tumor Characteristics | Treatments |
|---|---|---|---|---|
| Belciug et al., 2010 [30] | Wisconsin Prognostic BC (Available from UCI Repository) (47/151) | NU | Area, perimeter, compactness, Texture, Concavity, concave points, Size, Lymph nodes involvement | NU |
| Strumbelj et al., 2010 [21] | oncologyBCR (Institute of Oncology, LjublJana) (449/432) | Age, Menopausal status, Personal or Familiar previous malignancies | Histological type, Size, Grade, Local invasion, Lymphovascular invasion, Lymph nodes involvement, Expression of ER, PR | Chemotherapy, Hormonotherapy |
| Kim et al., 2012 [31] | Tertiary Teaching Hospital, South Korea (195/484) | NU | Number of tumors, Size, Grade, Local invasion, Lymphovascular invasion, Lymph nodes involvement, Expression of ER | NU |
| Salama et al., 2012 [32] | Wisconsin Prognostic BC (Available from UCI Repository) (47/151) | NU | Area, perimeter, compactness, Texture, Concavity, concave points, Size, Lymph nodes involvement | NU |
| Murti, 2012 [33] | Breast Cancer Dataset (Available from UCI Repository) (81/196) | Age, Menopausal status | Location, Size, Grade, Lymph nodes involvement | Radiotherapy |
| Tomczak, 2013 [22] | Institute of Oncology, Ljubljana (follow up from Strumbelj et al.) (949 patients, distribution unknown but assumed the same as the one from Strumbelj's: 51%/49%) | Age, Menopausal status, Personal or Familiar previous malignancies | Histological type, Size, Grade, Local invasion, Lymph/vascular invasion, Lymph nodes involvement, Stage, Expression of ER, PR | Chemotherapy, Hormonotherapy |
| Pawlowsky and Nagahashi, 2014 [34] | Wisconsin Prognostic BC (Available from UCI Repository) (46/148) | NU | Area, perimeter, compactness, Texture, Concavity, concave points, Size, Lymph nodes involvement | NU |
| Beheshti et al., 2014 [35] | Wisconsin Prognostic BC (Available from UCI Repository) (47/151) | NU | Area, perimeter, compactness, Texture, Concavity, concave points, Size, Lymph nodes involvement | NU |

*Continued on next page*

Table 2 – *Datasets and Feature Selection used in the analyzed studies*

| **Publications** | **Dataset (recurrence/no recurrence)** | **Patient Characteristics** | **Tumor Characteristics** | **Treatments** |
|---|---|---|---|---|
| Vikas Chaurasia et al., 2014 [36] | Breast Cancer Dataset (Available from UCI Repository) (85/201) | Age, Menopausal status | Location, Size, Grade, Lymph nodes involvement | Radiotherapy |

In brief, MD can be produced at random (MAR), completely at random (MCAR) or completely not at random (MNAR) [37]. Over the years, several strategies have been studied to handle this issue. The most simple one is Listwise Deletion, where records with missing data simply discarded. This approach may be inappropriate, especially in environments like healthcare, where most often patients are characterised with a large number of variables with high probability of missing observations. According to the literature, imputation is a more appropriate strategy to deal with MD: using the available complete data, the MD are estimated and filled with plausible values [38] [39]. Mean/Mode is one of the simplest imputation strategies, where continuous variables are imputed according to their mean, and categorial variables using their mode. Despite its simplicity, this strategy causes the data to loose some variability, which constitutes its major drawback.

A more sophisticated strategy is using mixture models trained with the Expectation-Maximization (EM) approach, which consists of two steps [40]: the Expectation step ("E-step") and the Maximisation step ("M-step"). Basically, the EM algorithm is based on finding the Maximum Likelihood of data in order to find the best estimates for missing observations. For the algorithm to start, the E-step makes an initial guess of the model parameters. Using those parameters, and according to the observed (complete) data, it produces estimates for missing observations. The M-step is then responsible for computing new model parameters using the current MD estimations. This process continues repeatedly until the algorithm converges. More specific details can be found in [40] [41][42].

Multiple Imputation (MI), substitutes every missing observation $M$ times ($M > 1$), using $M$ different estimators (e.g. EM, Markov Chain Monte Carlo methods) [43]. As

the name implies, multiple complete datasets are generated, each with different estimates for the absent observations. Then, the $M$ complete datasets have to be analysed using standard methods – for instance, classification models – in order to combine the different estimates and obtain a single set of results (a discussion on combination rules is given in [37]). On one hand, MI is able to reflect the data variability due to missing values. On the other hand, it is computationally expensive, given the generation of different MD estimations, and the required time to further analyse its results.

### *2.3 Machine Learning Methods*

Throughout the years, many Machine Learning (ML) algorithms have been used to predict BC recurrence. These methods can be categorised as "black-box" or "white-box". Black-box algorithms work on the basis of "input stimulus" and "output reactions", without any knowledge on their internal procedures. From the user perspective, this type of algorithms rises a wide range of questions that will always remain unclear, such as how the results are generated and how can the results be explained by the internal methods, given a specific input, among others. This issue becomes especially critical when the user (e.g. a clinician) considers interpretability as a key requirement, in order to use this kind of approaches and benefit from them in his daily decision-making activities. Contrary to black-box algorithms, white-box algorithms allow the inspection and explanation of their internal rules; that is, the results of a white-box algorithm may be analytically (mathematically) derived from a given set of inputs. This section presents a review on the algorithms used to predict BC recurrence in the studied research works, starting with the white-box algorithms.

#### *2.3.1 Decision Trees*

Decision Trees (DT) are defined by recursively partitioning the input space from a root node to multiple branch nodes [44][45]. The root node is the "first division" of a DT, from which outgoing edges create several other nodes. Nodes with outgoing edges (with the exception of the root node) are known as internal (or test) nodes, while the remaining (that only have incoming edges) are called leaves, each one assigned to a class. The test nodes divide the input feature space into $p \geq 2$ sub-spaces according to a condition test of the input features values. Typically, a single feature is considered in

each test node and the feature space is divided according to that feature's values. For continuous features, each outgoing edge represents a certain range. An input vector is classified in the DT by sorting it from the root to a leaf, according to the results of the conditions tested along the path. DT are computationally efficient, can easily handle mixed variables (continuous and discrete) and the rules generated by them are relatively easy to interpret and understand, particularly in healthcare contexts. Some of the most popular DT approaches are the C4.5 algorithm [45][46] and CART (Classification And Regression Tree) [47], which use entropy-based measures (typically, the gain ratio) as splitting criterion during the tree construction process. C5.0 is an extension of C4.5 [48]. The main advantage of DT is its interpretability; however, noise and missing data can contribute to drastically decrease the accuracy of these algorithms.

### *2.3.2 Naive Bayes*

Naive Bayes (NB) classifier takes into account the probability distribution of the patterns in each class to make a decision, assuming that there is a probabilistic relationship between predictors (features) and the output (class) [49]. Bayesian classification determines the probability of a given pattern represented by **x** to belong to class $\omega_i$, $P(\omega_i \mid x)$, called *posteriori* probability. Considering a binary classification problem, where two *posteriori* probabilities exist, $P(\omega_1 \mid x)$ and $P(\omega_2 \mid x)$, NB decision rule considers that:

If $P(\omega_1 \mid x) > P(\omega_2 \mid x)$, then **x** belongs to $\omega_1$;

If $P(\omega_1 \mid x) < P(\omega_2 \mid x)$, then **x** belongs to $\omega_2$;

Alternatively, if $P(\omega_1 \mid x) = P(\omega_2 \mid x)$, then the choice is arbitrary. The *posteriori* probabilities are calculated according to the well-known Bayes' law (equation 1).

$$P(\omega_i \mid \mathbf{x}) = \frac{p(\mathbf{x} \mid \omega_i) P(\omega_i)}{p(\mathbf{x})} \tag{1}$$

where $P(\omega_i)$ is the *prior* probability of class $\omega_i$, i.e., an estimate of the probability of pattern **x** to belong to $\omega_i$; $p(\mathbf{x} \mid \omega_i)$ is the *likelihood* of **x**, that can be estimated through the probability density function (*pdf*) of **x**; and $p(\mathbf{x})$ is the total probability of **x**, which can be determined using equation 2.

$$p(\mathbf{x}) = \sum_{i=1}^{c} p(\mathbf{x} \mid \omega_i) P(\omega_i) \quad (2)$$

Due to the fact that NB uses probability rules, it inherits somehow the strong results of the statistics. Also, another advantage of this method is allowing the researcher to include is domain experience in the modelling process of NB classifiers. Moreover, being a white-box method, it can be more easily understood, for instance, by clinicians. However, its computational complexity, especially when a large dataset is used, constitutes its main drawback [50].

### *2.3.3 Logistic Regression*

Logistic Regression (LR) is a mathematical method which aims to describe the relation between a group of independent variables and a dichotomous dependent variable. To achieve that, LR tries to estimate a set of unknown parameters using a maximum likelihood method [51]. The term "Logistic Regression" may be slightly misleading, since regression is mostly used to build models where the target feature is continuous. However, LR is used for classification, not regression. In brief, LR involves a probabilistic view of classification. It maps a point of a multidimensional feature space to a value in the range [0,1], using a logistic function. The logistic model can therefore be interpreted as a probability of class membership by applying a certain threshold to such probability. In conclusion LR gives the class probability for each considered feature vector. The class assignment depends on the chosen threshold. One of the main advantages of this method is that it clearly illustrates how the inputs justify the outputs through the final generated equation. However, its performance drops when the dataset contains MD.

### *2.3.4 K-Means Algorithm*

K-Means is one of the most well-known clustering algorithms, due to its easy implementation, efficiency and success over a wide range of pattern recognition applications [52]. K-Means is a partitional clustering algorithm, which means that it does not impose a hierarchical structure and finds clusters through the recursive partitioning of data, according to a similarity criteria between data points [53]. In brief, K-Means algorithm

works as follows. First, the desired number of clusters, $k$, needs to be specified. Then, $k$ randomly chosen *centroids* (which are simply pseudo-data points with the same dimensionality as the ones intended to cluster) are initialised. The distances (e.g. euclidean distance) of each point to those $k$ centroids is calculated, and each point is assigned to its closer centroid. The initial partitions of data are defined at this point. However, the objective of K-Means is to find a partition such as the sum of the squared error over all $k$ clusters is minimised [52]. For that reason, the *centroids* of each partition are updated at each iteration of the algorithm. The *new centroids* are given by the mean vectors of the point belonging to each cluster. Again, new distances are calculated for each point, now considering the new centroids. This is successively repeated until there are no changes in cluster membership (none of the points changes cluster).

It is worth mentioning that, unlike the previously discussed algorithms, K-Means is an unsupervised learning algorithm. Moreover, despite its simplicity and low computational cost, K-Means has some drawbacks that relate to the number of clusters, the initialisation of centroids, and the presence of noisy data [52] [54]. The number of centroids $k$ needs to be specified *a priori*, which sometimes is not trivial, especially without sufficient knowledge domain. Also, since the initialisation of centroids is random, different runs of the algorithm may return different results. Finally, K-Means is not robust to noisy data, what may skew the update of cluster centers in some cases.

### *2.3.5 Bagging*

Created by Breiman, Bagging ("bootstrap aggregating") uses several bootstrap samples to train different classifiers, that are afterwards combined to achieve the final classification results. Each bootstrap sample is created by randomly selecting examples (with replacement) from a training set of size $m$. From $n$ bootstrap samples, $n$ classifiers $C_1$, $C_2$,...,$C_n$ are built [55], each one using a different training set. The final classifier $C_b$ is built from all the $C_n$ classifiers, by combining them through majority voting, where ties are broken arbitrarily (for more details please refer to [56] and [57]).

### *2.3.6 Boosting*

Boosting was created to improve the accuracy of a specific algorithms' family called "weak learning algorithms", which are typically slightly correlated to the true classifi-

cation. On the contrary, strong learners are algorithms well correlated with the true labels, providing good classification results. One of the advantages of weak learners is that they are usually much faster than strong ones. The first Boosting procedure was introduced by Schapire [58] and worked somehow similarly to bagging. An subset of $n$ examples ($n <$ the total number of training examples $N$) was taken randomly without replacement from a initial training set (considered to be the training subset $Z_1$). $Z_1$ was the used to train a weak classifier $C_1$. Afterwards, a training subset $Z_2$ (with $n < N$) was built, containing half the samples misclassified by $C_1$, and another weak classifier $C_2$ was trained. Finally, all the samples of the initial training set $N$ for whose $C_1$ and $C_2$ predictions disagreed were trained against a third weak classifier, $C_3$. The final classifier was obtained by a voting scheme of $C_1$, $C_2$ and $C_3$ [59] .

In 1995, Freund and Schapire [60] introduced the most well-known boosting algorithm, called Adaptive Boosting (AdaBoost). In AdaBoost, the idea is to consider a weighting scheme to select the training subsets. This algorithm starts by considering a maximum number of classifiers $M$ and weighting each training example equally. The misclassified examples get their weights increased for the next classification stages, while the correctly classified examples get their weights decreased. These "weights" simply determine their probability of being chosen for the training set in the next stages and therefore, wrongly labelled examples have a higher probability of being used (learned) again. Moreover, each classifier will get a specific weight attending to its performance in the training set and the final classifier is defined by a linear combination all the considered $M$ classifiers, each one contributing with its associated weight [59].

Formulated by Friedman et al., LogitBoost is a probabilistic interpretation of AdaBoost. It fits an additive logistic regression model using Newton steps to find estimates for its parameters via the maximum likelihood [61] [59]. Instead of using an exponential loss function (which AdaBoost does), LogitBoost minimises the logistic loss, which makes it less sensitive to outliers (known to be a bad feature of AdaBoost).

### *2.3.7 Linear Discriminant Analysis*

Linear Discriminant Analysis (LDA) is a linear transformation technique, generally used to reduce the dimensionality of a dataset in the preprocessing phase, in order to

decrease the computational cost of classification and avoid overfitting. Nevertheless, it is also used for classification alone, based on the concept of searching for a linear combination of features that allows the maximisation of between-class variance, while minimising the within-class variance. In other words, the optimisation criterion of LDA is to maximise the ratio of between-class and within-class scatter. LDA, also called FDA (Fisher Discriminant Analysis), was first developed by Fisher to deal with only two classes [62]. However, after more than 10 years, this method was extended to deal with multiple classes [63].

### *2.3.8 Support Vector Machines*

Support Vector Machines (SVM) were first introduced by Vladimir Vapnik for two-class classification [64]. Basically, this algorithm tries to find the optimal decision hyperplane that maximises the separation margin between data points of distinct classes [65]. The middle of the separation margin defines the decision boundary (optimal hyperplane) and the data points that are closest to it are the support vectors. SVM belong to the general category of kernel methods. Kernel methods can operate in high-dimensional spaces, since they depend on the data only through dot-products. This has two main advantages: it allows the generation of non-linear decision boundaries and enables the classification of data that has no obvious fixed-dimensional vector space representation [66][67]. SVM are known for excellent classification performance, since they can handle high-dimensionality problems and have a good generalisation behaviour. They balance the model's complexity against its success at fitting the data, which translates into a successful trade-off between the model's flexibility and the error in training data [67]. However, and despite being a white-box algorithm, they require a comprehensive understanding of how they work. When training SVM, researchers have to face several decisions concerning the preprocessing stages of the input data and the SVM's hyperparameters (e.g. kernel function, regularisation constant).

### *2.3.9 k-Nearest Neighbours*

k-Nearest Neighbours (KNN) is a supervised classification algorithm in which the $k$ nearest neighbours of a point are chosen, found by minimising a similarity measure (e.g. euclidean distance, mahalanobis distance) [68]. To determine the class of an

unlabelled example, KNN computes its distance to the remaining (labeled) examples, and determines its $k$-nearest neighbours and respective labels. The unlabelled object is then classified either by majority voting – the predominant class in the neighbourhood – or by a weighted-majority, where a greater weight is given to points closer to the unlabelled object. The major drawback of KNN is related to the fact that it is a lazy learning algorithm. That means that there is no "model": the training data is not used to perform any generalisation. Therefore, whenever KNN searches for each instance's nearest neighbours, it needs to go through the entire dataset, which is especially problematic for large databases. Another issue is finding the optimal number of neighbours ($k$) and the most appropriate distance metric to use. This requires a careful study of the dataset and the development of several KNN models, in order to achieve the best results.

#### *2.3.10 Association Rule Learning*

Association Rule Learning allows to unveil the relationship among variables in a dataset. Proposed by Agrawal et al. in 1993 [69], this method assumes that all variables are categorical and because of that it is not a good algorithm to deal with numerical data. Each identified association rule follows two main concepts: support and confidence. Support identifies the percentage of the population that follows a specific rule. Confidence is the measure of certainty associated with each discovered rule. In a simple manner, association rules can be perceived as "if-then" rules which describe relations between the data. They are extremely advantageous due to their exhaustive exploration of the data [70]. Moreover, the final rules returned by this algorithm are usually simple enough to be understood by users. Nevertheless, some of their inconveniences are that they are affected by noisy data and have a slight tendency to overfit the data.

#### *2.3.11 Isotonic Separation*

Isotonic Separation, developed by Chandrasekaran et al. [71], is a linear programming model that follows the principles of isotonic consistency. The isotonic consistency constraint assumes an ordering relation of data points in the feature space, given by "$S = \{(i, j) : a_i \geq a_j\}$, where $a_i$ and $a_j$ are coordinate vectors" that represent the at-

tribute values of $i$ and $j$, in all $d$ dimensions (for more details please consult [72]). Therefore, $S$ consists of $(i, j)$ pairs of ordered data points such that, considering a two-classification problem:

If $i$ is classified as belonging to class $\omega_1$, then $j$ must be classified as belonging to class $\omega_1$ and conversely;

If $j$ is classified as belonging to class $\omega_2$, then $i$ must be classified as belonging to class $\omega_2$;

Transposing this scheme to the problem of BC recurrence; if, for instance, certain patients have registered values of age, size of tumour, expression of estrogen receptor, and so on, that cause them to be classified as "recurrent"; then all patients registering the same (or greater) values are also considered "recurrent". Isotonic separation also takes into account misclassification costs, where each misclassified data point receives a penalty, for instance, $\alpha > 0$ for each "recurrent" patient classified as "non-recurrent" and $\beta > 0$ for each "non-recurrent" patient classified as "recurrent". Isotonic Separation minimises the total cost of misclassification, $\alpha n_i + \beta n_j$, where $n_i$ is the number of wrongly classified recurrent patients and $n_j$ is the number of wrongly classified non-recurrent patients.

### *2.3.12 Random Forests*

Created by Breiman in 2001 [73] Random Forests (RF) instantly became a commonly used method, mainly due to its simplicity (in terms of training and tuning) and performance [74]. Similar to the bagging algorithm (in the sense that it uses individual decision trees as individual classifiers), RF construct correlated trees. However, in this case, for each tree node, *v* features out of the total $V$ input features are randomly selected (considering *v* $<< V$) and the best split of $v$ features divides the node. Finally, the "forest" picks the most voted class, over all the trees in the forest, either by considering the mode of the classes of the individual trees (classification) or their mean (regression). By designing a multitude of decision trees and later combining their predictions, RF decrease the risk of overfitting, usually associated with individual decision trees [75].

### *2.3.13 Neural Networks*

Artificial Neural Networks (ANN) [76] are mathematical-computational models inspired by neuronal cells' functioning, simulating human reasoning. A generic ANN model is composed by three layers: the input, output and processing layer (or hidden-layer) [77]. The input layer receives the data, while the output layer communicates the result. The hidden-layer is responsible for data processing and results' calculation. ANN analyse existing patterns in the information they receive and derive associations between input and output variables. These associations are used to produce the most correct output for each input, which is then compared to the correct output and, based on this comparison, the algorithm resets the associations between the input data and the previously determined output. This process continues iteratively until the correct result is determined or the maximum of iterations is achieved. Then, the system memorises the model of such association between inputs and outputs in order to classify new cases.

A Multi-Layer Perceptron (MLP) is a modification of the standard linear perceptron and can distinguish non-linearly separable data [78]. In is basic form, it is simply a type of feed-forward ANN. It consist of multiple nodes interconnected in a direct graph, where the input layer passes the input vectors to the network and the output layer communicates the response. A MLP can have one or more hidden layers, composed by neurones with non-linear activation functions (e.g. sigmoid, tangential), responsible for the computation of results.

The major advantage of ANN models is that they avoid the construction of "if-then" rules, and its definition by experts. They also do not need a very large set of data to produce estimates, though the larger the training set is, the more accurate the results are. On the other hand, the training phase can be time-consuming. However, the main disadvantage of this type of algorithms is their model's interpretation – these algorithms are black-box models, since the associations between data are complex and difficult to explain.

### *2.3.14 Self-Organizing Maps*

Self-Organizing Maps (SOM) are a type of artificial networks that use a form of unsupervised learning (competitive learning) to represent the input data in a low-dimensional space (a map), typically with one or two dimensions[79]. SOM network is built from a grid of neurons ("nodes"), where each node has a specific position in the grid and is completely connected to the input layer. Furthermore, each node is associated with a weight vector, which has the same dimension as the input feature space: feature vectors of $d$ dimensions will origin nodes with weight vectors of size $n$. Like most ANN, SOM performs training and testing, or in this case, training and mapping. In the training phase, SOM builds the map using the input examples, by placing each one next to the node with the most similar weights, known as "Best Matching Unit" (BMU) [38]. The BMU's and adjacent nodes' weights iteratively adapt every time a new training input is given to SOM. In the mapping phase, the test input vectors are classified according to their distance to the existing nodes in the map, constructed in the training phase.

### *2.3.15 Classification Restricted Boltzmann Machines*

Restricted Boltzmann Machines (RBM) are a variant of Boltzmann machines, with the constraint that there can be only inter-layer connections, i.e., there cannot be connections of nodes within a layer, only between layers. RBM can be seen as stochastic neural networks, given their neurones-like units whose activation has a probabilistic element [80]. They typically have one layer of visible units (inputs), and one layer of hidden units. They may or may not have a bias unit. Each visible unit is connected to all the hidden units, and the connections are symmetric, meaning that each hidden unit is also connected to all the visible units. As mentioned above, no hidden unit is connected to any other and no visible unit is connected to another visible unit.

Despite being mostly used as unsupervised learners, RBM can also be used as supervised, black-box algorithms for classification [81]. Classification Restricted Boltzmann Machines (ClassRBM) are a variant of RBM oriented to classification. In classification tasks, RBM are treated as parametric models (considering that the number of hidden layer is fixed) of the joint distribution between the layer of hidden units (neurones) and the visible layer of inputs. Based on this joint probability, ClassRBM can

compute the distribution $p(y \mid x)$, that is, the probability of $x$ belonging to $y$, which allows the determination of the most probable class label (see [82] for more details).

#### *2.3.16 Genetic Algorithms*

Genetic Algorithms (GA) are inspired in Darwin's Evolutionary Theory which explains the evolution of species through natural selection. As species evolve in order to adapt to their environment, a GA also uses a "survival of the fittest" philosophy in order to obtain the result that best fits the data from a population of individual potential solutions [83]. A fitness function determines which solutions should be kept and which should be eliminated. At each generation, a new population is generated and the fitness values of all individuals are evaluated based on their performance in the problem domain. Three main genetic operators can actuate over each selected population, so as to generate the next-generation population – copy, crossover and mutation. These mechanisms are repeated, and the population continues evolving, until the optimal solution (fitness value) is produced, or a stopping condition is reached (e.g. a maximum number of generations).

### *2.4 Sampling Strategies*

To evaluate a classifier, researchers need to find its true error rate, i.e. the classifier's error rate in the entire population. However, in real-world applications, it is not possible to access the entire population. Only a finite set of examples is available, from which an estimation of the true error rate must be calculated. A naive approach to this issue of finite datasets would be to use all the available data to train and test the models. However, this would return an overly optimistic error estimation and serious overfitting [54]. For that reason, another approach must be pursued: the division of the available (and labeled) examples into training and test sets. Following the principles of supervised learning, the training set is used to build the model, while the test set is used to perform its evaluation. The techniques to divide data into training and test sets are called sampling strategies and in this section we will review the ones used in the reviewed works: Holdout method, Random Subsampling, k-fold Crossvalidation and Leave-One-Out [54][84][85][86].

### 2.4.1 Holdout method

The holdout method simply divides the available examples into two disjoint sets, according to some percentage. Traditionally, train and test sets are divided in a 50%-50% partitioning scheme [54], although most authors consider a train/test division of 70%/ 30% or 80%-20%. Holdout is the simplest of the sampling strategies, but along with its simplicity come some limitations. For small datasets, the holdout method may be subjected to unfortunate splits, where the training data may not be representative of the population, which may lead to biased results.

### 2.4.2 Random Subsampling

In Random Subsampling, we consider several $p$ experiments (runs or splits). Each split considers a fixed number of random training and test examples, selected without replacement. Then, for each split, training and testing is performed individually. The individual error rates $e_i$ (determined using each test set) are averaged to form the final error estimate, according to equation 3. When choosing the samples for each split, the samples cannot be repeated. However, between splits, the same samples may be selected. Therefore, Random Subsampling does not guarantee that all samples are used for training and testing, which constitutes its major drawback.

$$error = \frac{1}{p}\sum_{i=1}^{p} e_i \tag{3}$$

### 2.4.3 k-fold Crossvalidation

k-fold Crossvalidation divides the data into $k$ subsets (folds) that rotate, in order to consider all folds for both training and testing. More specifically, k-fold Crossvalidation considers $k > 1$ distinct folds, where $k-1$ folds are used to train a classifier and the left-out fold is used for validation. This is performed $k$ times, so that every fold is considered in both training and test design. Similarly to Random Subsampling, the true error rate is estimated by averaging the each error $e_i$, obtained from each fold.

The choice of $k$ influences the bias-variance tradeoff in performance estimation through k-fold Crossvalidation. For small values of $k$, the bias increases, although the variance is low; for higher values of $k$, the error estimate is more close to the true error

(low bias), but the variance increases.

### *2.4.4 Leave-One-Out*

Leave-One-Out (LOO) is a particular case of k-fold Crossvalidation, when $k = N$, the total number of available examples. Therefore, in LOO, $N - 1$ examples are used for training, while the held out example is used to test the classifier. Thus, there are $N$ error estimates that need to be averaged to determine the final estimate of the error rate.

In LOO, only one sample is used for testing, which leads to a high variance in error estimation. On the other hand, since all $N-1$ are used in the training design, the bias is low. For that reason, the averaged test set error is a good estimate of the performance error. When the sample size is low, LOO is the best approach to provide an accurate estimate of the true error.

## *2.5 Evaluation Metrics Background*

The performance evaluation of a classifier is normally based on a confusion matrix (Table 3). This matrix illustrates the actual versus the predicted class in classification problems, where each column of the matrix represents the instances in an actual class and the rows represent the instances in a predicted class.

Table 3: Confusion Matrix

| | | Actual Class | |
|---|---|---|---|
| | | Negative | Positive |
| Predicted Class | Negative | True Negative (TN) | Fase Negative (FN) |
| | Positive | False Positive (FP) | True Positive (TP) |

Based on that matrix, many metrics can be derived. Precision shows the proportion of the correctly predicted positive cases relative to all the predicted positive ones (equation 4).

$$\text{Precision} = \frac{TP}{TP + FP} \tag{4}$$

Recall or Sensitivity represents how many positive examples the classifier was able to correctly identify (equation 5).

$$\text{Recall} = \frac{TP}{TP + FN} \tag{5}$$

Specificity represents how accurately the classifier behaves in terms of predicting the negative class (equation 6).

$$\text{Specificity} = \frac{TN}{TN + FP} \tag{6}$$

Mean Square Error (MSE) of a classifier represents the different between a vector of $n$ predictions ($\hat{Y}_i$) and the true observable vector ($Y_i$) for all $n$ examples (equation 7).

$$\text{MSE} = \frac{1}{N}\sum_{i=1}^{n}(\hat{Y}_i - Y_i)^2 \tag{7}$$

Accuracy represents how many predictions of the classifier were in fact correct (equation 8).

$$\text{Accuracy} = \frac{TP + TN}{TP + TN + FP + FN} \tag{8}$$

Other measures that can also be used are Area Under the Curve (AUC) and F-measure. On one hand, AUC is a measure of how well a classification model can distinguish between two classes. In practice, AUC is often used when a representative measure of discrimination is needed and it can even replace accuracy as a performance measure [87].

On the other hand, the F-measure is defined as the harmonic mean of Precision and Recall, providing a balance between both performance metrics (equation 9).

$$\text{F-measure} = \frac{2 * precision * recall}{precision + recall} \tag{9}$$

## 3 Application of Breast Cancer Recurrence

In 1997, Subramani Mani et al. [18] compared the performance of rule-based classifiers (DT and Association Rules) with a well-known probabilistic classifier (Naive

Bayes), in the identification of tumor features associated with BC recurrence. The data was collected from a Breast Care Center in California, where 887 patients were characterised by demographics and tumour-specific information, including diagnostic and treatment features. From the initial set of features, 6 were hand-picked by a medical expert to proceed with the study. Since only 10% of the collected patients suffered from a recurrence event (85 patients), the majority class (no-recurrence) was randomly divided into 6 datasets, in order to follow a 60%-40% class distribution (no-recurrence / recurrence) in each dataset. Naive Bayes proved to be the best approach with an average accuracy of 68.3%, overcoming all others in the majority of the tested datasets. A particular type of association rule, the First Order Combined Learner (FOCL) has also stood out with an average accuracy of 66.4%.

This work rises a common controversial topic within the bioinformatics community: the discussion between classification results and interpretability. The use of rule-based classifiers is generally encouraged in medical contexts [18][88][89], due to the additional information they provide. However, as shown by this work, they did not offer leading results in terms of classification accuracy. Moreover, most of the generated rules reflected a somewhat obvious knowledge domain, which does not constitute a meaningful contribution to medical experts. From a technical perspective, there are some points to be further discussed in this work, namely the feature selection phase, the sampling phase and the evaluation metrics used. Although some works make use of clinical guidance to select relevant variables to study, they should be clearly stated to allow a proper comparison with related works. Furthermore, the explanation of the division of patients into the considered 6 datasets is too vague. As it is stated, it seems that the division of the majority class (no-recurrence) followed a random subsampling method, which does not guarantee that all samples are used to model/test the classifiers. Therefore, important information might be unused due to the sampling phase design. Finally, in spite of the authors' efforts to design datasets with more appropriate class distributions, they still suffer from a considerable imbalance, which requires the evaluation of algorithms to go beyond the traditional accuracy measure, more efficiently applied to balanced classes. However, accuracy was the only metric used, which may hint on misleading conclusions.

In 2003, José Jerez-Aragonés et al. [24] employed a hybrid model, combining ANN and DT, to a database from a hospital in Malaga, Spain, in order to determine whether a patient will suffer a post-surgical relapse at any period during follow-up time, considering an end-point of 5 years. Out of 85 available features (including demographics, post-surgical and treatment information), a subset of 14 features was selected by medical experts as the most relevant for predicting outcome. The hybrid model was then used to predict BC recurrence for seven different time intervals from the surgical intervention. The first six intervals are equally spaced (10-month periods), while the remaining one considered a period of over 60 months. Initially, the dataset was constituted by 1035 patients, but records with Missing Data were discarded, resulting in a decrease of the sample size: from 845 patients for the first interval to 466 patients for the last interval. Using a holdout method (80% train / 20% test), the performance of the proposed hybrid approach was compared to a Cox statistical technique, commonly used by medical experts. In terms of accuracy, the proposed approach outperformed the Cox model in all intervals (with results ranging from 93.4% to 96%), except for the last interval ($>$60 months), lagging behind by just 0.3%. To complement the accuracy analysis of the proposed method, recall, precision and specificity measures are also included. Recall varied between 78.7% and 88.7%, while precision ranged from 64.8% to 77.2% and specificity between 94.5% and 97.2%.

The combination of DT with ANN is an interesting approach, since it does not discard the advantages of one in favour of the other. DT provides useful information for selecting the most relevant prognostic factors for each considered interval, while ANN are able to use that information to make an accurate prediction, using personalised topologies for different time intervals. By not choosing one algorithm over the other, the authors take advantage of each one's potential, achieving accurate, yet interpretable results, which is an improvement from the previous illustrated work. Also, in this work, the authors have in mind that accuracy is not always the best classification metric, and complement this information with additional metrics (such as sensitivity and specificity), allowing for a better evaluation of the power of the proposed method. Nevertheless, some topics remain for discussion. Although the authors mention that their approach is appropriate for data with a considerable number of features with missing values, this

is not supported by the work itself, since the MD perspective is ignored. Furthermore, its application for a high number of features is not discussed yet, since a medical team performed the feature selection phase beforehand.

Amir Razavi et al. followed the idea of combining DT with other algorithms to improve the prediction of BC relapse during the first five years after diagnosis [23]. In 2005, they applied Canonical Correlation Analysis (CCA) as a preprocessing step, prior to classification, to study the influence of dimensionality reduction in prediction performance. They used a dataset obtained from a Swedish regional center, with 3949 patients characterised by more than 150 features. Following the same methodology as the previous discussed works, the feature selection phase was performed by a team of medical experts, resulting in a decrease of the feature space to only 17 predictors. However, unlike previous works, values for MD fields were imputed using the Expectation Maximisation (EM) algorithm [23]. A 10-cross validation procedure was used to evaluate the performance of three different predictive models: (i) DT coupled with CCA, (ii) DT without any preprocessing step and (iii) DT with MD imputation as the only preprocessing step. The results showed that DT coupled with CCA overcame the other two approaches in terms of accuracy (67%) and specificity (63%); yet lagging behind both in terms of sensitivity, which is generally not a good indicator. However, it is important to state that this solution yields trees with only 10% the size of those without preprocessing, resulting in a simpler system, and improving interpretability. Still, it would be interesting to make a comparison between rule-based models and other types of classifiers, and their behaviour when coupled with the mentioned preprocessing strategies. In 2007, the same authors applied the previously developed combined model (CCA + DT) to predict BC recurrence within four years after diagnosis [19]. The used dataset consisted in 3699 patients (with absent observations), where 664 (18%) suffered recurrence in the first four years of follow-up. MD imputation was performed using Multiple Imputation (MI), and 10-fold cross validation was used to estimate performance error. A hundred cases were previously separated from the initial dataset (by stratified random sampling) to validate the developed model against the predictions of two medical experts (Oncologist 1 and 2).

Although the comparison of the AUC (area under the ROC curve) values between

the three approaches (DT and two oncologists) did not significantly differ, a more detailed analysis on the performance results is required. The authors present the confusion matrix for the validation set of 100 patients (81 without and 19 with recurrence) but at no point do they make a critic comparison of the results. In fact, despite a higher accuracy (82%) and precision (57.1%), DT is overpowered by one of the two medical doctors collaborating in the study (Oncologist 2) in terms of AUC values, specificity and F-measure. With a training and validation set presenting such a high imbalance (80% without recurrence - 20% with recurrence), the sensitivity should assume a pivotal importance. Sensitivity results are poor for DT (21.1%); however, Oncologist 1 had an even lower rate of sensitivity (5.56%), despite having a higher percentage of specificity (97.5%).

The poor results of DT in this work might be explained by the dominance of "no-recurrence" cases in the training model. Although the authors discuss the usage of sampling strategies to balance the dataset, they do not apply them, presenting arguments such as the small sample size or the lack of representativeness they would generate. However, this should have been considered since it is not clear that the proposed approach is the most suitable to predict BC recurrence: Oncologist 2 had a sensitivity rate of 57.8% without severely compromising precision (50%), by analyzing patient records with missing data.

In the same year, Yijun Sun et al.[20] combined clinical and genetic information to create a "hybrid signature", capable of predicting BC recurrence in the first five years after diagnosis. This work makes use of microarray data, publicly available in the Nature website [90]. The dataset includes 97 patients, 46 of which suffered recurrence, while 51 remained recurrence-free. According to previous works using this dataset [91], it also contains MD, although this perspective is not addressed throughout the work. Preprocessing steps before classification include data standardisation (Min-Max) and a feature selection algorithm (I-RELIEF), developed earlier by Sun and Li [92]. Four different approaches are tested: one using only genetic markers; another using only clinical markers; a hybrid signature (including genetic and clinical information); and St. Gallen's criterion [93], a consensus criterion to determine recurrence used in oncology guidelines. To compare the performance between the approaches, the authors specified a

threshold for each one, in a way the sensitivity is 90% for all. Then, the comparison was done by analysing the corresponding specificity values: 47%, 48%, 67% and 12% for genetic-only, clinical-only, hybrid signature and St. Gallen approaches, respectively. The ROC curves for the first three approaches were also compared, with the hybrid signature outperforming the other two (which in turn showed a similar behaviour).

It must be noted that this is the work that includes the smallest number of patients. With a small sample size, there is a higher danger of overfitting the training data. To avoid this problem, a nested Leave-One-Out Cross Validation (LOOCV) is adopted [20]. In a nested crossvalidation, an inner loop is responsible for the selection of the optimal classification parameters (for I-RELIEF in this case), while in the outer loop the classification of the held-out sample is performed. Linear Discriminant Analysis (LDA) is used in the classification task (outer loop), since it does not require the estimation of hyperparameters, and thus makes the experiments computationally less expensive.

Before Sun, some authors had previously attempted to combine genetic and clinical information, but rather unsuccessfully [94][95], which reinforces the achievements of this work. On one hand, one may argue that 67% is far from an optimal specificity result. On the other hand, it must be noted that the proposed hybrid signature improved the specificity of the remaining approaches by nearly 20% to 60%. This proves that the combination between genetic and clinical information is a suitable approach to determine the prognosis of BC patients, even though combination strategies are difficult to design.

Also in 2007, Young Ryu [28] used the Ljubljana Breast Cancer Dataset (Breast Cancer Dataset available in UCI Repository [16]) to compare several methods to predict BC recurrence in the first five years after removal of tumour: isotonic separation, robust linear programming (LP) and three variants of DT (C4.5, OC1 and QUEST), SVM, AdaBoost and learning vector quantisation. The Breast Cancer Dataset contains 286 patients, where 201 (70.3%) did not suffer recurrence and the remaining 85 (29.7%) had recurrence events. Each patient is characterised by 9 features, and there are some missing values. In particular, “node-caps” and “breast-quad” are responsible for the 9 missing observations present in this dataset. From those 9 missing values, 5 belong to the “no-recurrence” class while the other 4 belong to the “recurrence”. All

the methods were evaluated according to a holdout method (70% for train and 30% for test), except for QUEST decision three, which has its own sampling scheme (3-fold crossvalidaton). The results are presented in terms of error rate; however, to make them comparable with the remaining works, we translated them to accuracy values ($1 - error$). Isotonic separation outperformed all others with 80% of accuracy. Only the accuracy (conversely, the error rate) was determined for each classifier, which may be considered a limitation, since the Breast Cancer Dataset is known to have a 29.7% / 70.3% of recurrence/no-recurrence distribution [16]. Moreover, a backward sequential elimination process for feature selection showed that age, menopause status, node capsules, tumor grade and irradiation were the most relevant features for recurrence.

The most comprehensive study in terms of tested classifiers, feature selection algorithms and performance measures was performed by Thora Jonsdottir et al. [15], in 2008. They implemented 17 classification algorithms, including NB, several variants of DT and other rule base classifiers (OneR, PART, Jrip), Logistic Regression and some meta-classifiers, including boosting, bagging and ensemble schemes. Moreover, this work also uses a wide range of feature selection algorithms, such as OneR, correlation based feature selection method (CFS), consistency subset evaluator (LVF), RELIEF, information gain, and C4.5 decision tree. Furthermore, the existing knowledge domain (from previously published works in BC, medical experts and authors' experience) is also explored to select the most relevant features. However, the results from these feature selection methods are not discussed. The authors do not state if one returned better results than the other, or, alternatively, if each one's results were combined to select a final subset of features. The classification results were evaluated in terms of accuracy, kappa values, AUC, sensitivity and specificity. The algorithms were run on a relatively small dataset (257 patients) with high dimensionality (400 features), obtained from the University Hospital of Iceland (Rose dataset). Jonsdottir's study focused on two main goals: (i) predicting BC recurrence within five years after diagnosis and (ii) predicting recurrence risk (low, intermediate or high) within the same time period. The latter is out of scope of this review, although the recurrence risk, not as outcome but as predictor, was also included in (i), in order to determine if the inclusion of this feature had any influence in predicting recurrence events. In order to fulfil objective (i), the

authors conducted a feature selection phase that resulted in 3 different datasets:

1. Base-DS, including 98 features selected according to the experience of a medical expert and the results of the feature selection methods;
2. Med-DS, with 22 features selected from Base-DS by a medical doctor;
3. Small-DS, where only 5 features were manually selected from Base-DS.

The distribution of recurrence/no-recurrence events was 28.4% / 71.6%, exhibiting a considerable imbalance between classes, which the authors have counteracted by random subsampling of patients. The MD perspective was not directly addressed (there is no information on absent observations in the data); however, all the used classification algorithms can handle MD directly.

A 10-fold cross validation scheme was used across all classifiers for evaluation. Only the results for the algorithms with the best results were discussed in this work, namely NB, DT and PART. However, the incorporation of results for the remaining ones would have been interesting to allow for a cross-sectional evaluation among different works. In terms of accuracy, and considering the "risk of recurrence" as an extra feature, DT overcomes the other two approaches in Base-DS (76%) and Med-DS (77%), lagging behind PART in Small-DS by just 1% (PART achieves 80% of accuracy). When the "risk of recurrence" is added, DT still maintains its superiority in both Base-DS (75%) and Med-DS (76%), although lagging behind NB in Small-DS by just 1% (NB obtains a 78% accuracy). Regarding AUC values, NB outperforms the other two approaches with and without considering "recurrence risk" as an extra feature, although its superiority is not highly pronounced. DT obtains sensitivities of 48%, 45% and 37% for Base-DS, Med-DS and Small-DS, rivalling with PART, whose results were 48%, 33% and 37% for the same datasets, without considering "recurrence risk". When the extra feature is added, both approaches still obtain similar results: 48%, 40% and 30% for DT versus the 51%, 40% and 32% of PART. Finally, in terms of specificity, all three approaches obtain very similar (and high) results – from 87%-96% for NB, 86%-96% for DT and 78%-97% for PART – which makes it harder to asses the best one. The authors are not very conclusive in assessing the best approach, suggesting either NB or C4.5 decision tree are both suitable to be selected as the best classifier. It can be discussed that DT

achieves the best performance for Base-DS and Med-DS while Small-DS benefits the most by using PART. However, in a general view, DT seems to be the best approach, since it always achieves higher sensitivity results, and higher or comparable specificity. Furthermore, they have the advantage of producing interpretable rules, and, as the authors mention, may be clearly visualised when the dataset is small. In addition, the risk of recurrence did not improve the classification results; nor did the presence of a high number of features: the results are similar across all datasets. Therefore, it can be discussed that small dimensional spaces are suitable to address the BC recurrence context, having as main advantage the reduced complexity of the classification models.

The study developed by Qi Fan et al. [29] in 2010 targeted the internationally available SEER dataset [96], where 46.996.113 patients are described by a set of 117 features. The SEER Public-Use Data used in this work includes patients diagnosed with breast cancer from 1973 to 2005; however the end-point for determining recurrence is not specified. Records with Missing Data were ignored, but the number of patients kept in the final dataset is not mentioned. The feature selection phase was performed according to the validation of medical experts, with 13 features being selected as final inputs. A holdout method (80% train / 20% test) was used to evaluate the performance of five different algorithms, namely ANN and other four variants of DT. The results show that all the DT variants outperformed ANN in terms of accuracy, with the best accuracy results being achieved by C5.0 algorithm (71.17%). However, ANN had the highest precision rate for recurrence events (77.79%), while CART had the highest precision rate for no-recurrence events (73.75%).

Although nearly all accuracy and precision results are above 70%, a more detailed discussion should have been presented. Since the final number of patients included in the study (and of those, how many did or did not recur) is unknown, the readers do not have the necessary information to evaluate the results. If the class distribution of recurrence and no-recurrence is not balanced, the accuracy results are not reliable. Furthermore, although precision results are a common finding in most machine learning study, they are so in combination with recall results, which translate the sensitivity of the classifiers. This information is not presented in the paper, and is of much importance when the objective is to accurately predict a particular class of interest, in this

case, the “recurrence” class. Furthermore, no single classifier overcomes all others in the three considered metrics (accuracy, precision for “recurrence” and precision for “no-recurrence”). Therefore, it is not possible to determine which is the most suitable approach.

Smaranda Belciug et al. (2010)[30] compared the performance of k-means, SOM and cluster network in the detection of BC recurrence events, using the Wisconsin Prognostic Breast Cancer (WPBC) dataset [16]. Each record of WPBC dataset represents the follow-up data for one patient. This dataset includes invasive breast cancer cases with no evidence of distance metastases at the time of diagnosis, from 1984 to 1995. Therefore, prior to any classification study, the dataset should first be filtered to translate a defined end-point (e.g. recurrence within 2-years after diagnosis). The authors do not specify such in their research, and thus the true BC problematic cannot be identified. The WPBC dataset is available from UCI Repository, and contains 198 (47 recurrent and 151 non-recurrent) patients characterised by 34 features. These features describe the characteristics of the cell nuclei observable in an image of the patient’s breast mass. According to the dataset description [16], it contains absent observations. Specifically, “lymph node status” is missing in 4 cases (3 “non-recurrent” and 1 “recurrent”). However, the MD perspective is never mentioned in this work. From the 34 features in the original dataset, the authors chose 12 to be used, discarding redundant information unnecessary to the clustering algorithms. The feature selection process is not mentioned: the features are selected according to the authors’ assumption on their relevance to the study, without further elaboration on the subject. The authors compare both training and testing performance between the used algorithms, except for k-means, where only the training performance is assessed. For SOM and cluster network, a 10-fold cross validation procedure is used, and the accuracy results are then averaged to achieve the final classification results. Cluster network obtained the highest accuracy results in both training (83%) and testing (78%), versus the 72% and 67% obtained for SOM and the 62% (training) for k-means. On one hand, cluster network has shown an efficient behaviour in predicting BC relapse, achieving accuracy results higher than the majority of the discussed works. On the other hand, this paper fails by disregarding the clear class imbalance between “recurrence” and “no-recurrence” events (76.3%/23.7%). In

such cases, as previously mentioned, accuracy is not an appropriate metric, and other metrics, such as sensitivity and specificity, should be presented.

Furthermore, the fact that the MD perspective is not addressed is quite intriguing, since clustering methods cannot generally analyse data points with missing data, without further constraints. Usually, if no specifications are provided to the algorithm, data points with MD are discarded, and thus not clustered. Therefore, if MD has not been handled in any way, the results may be somewhat biased. Also, the 10-fold cross validation procedure could have been extended to k-means as well, for testing purposes. Using a LOO approach, each sample would be held out while the remaining were used for training. Then, the held out sample would be assigned to a class according to its proximity to the cluster centers, by majority voting, for instance. This would provide the testing performance for k-means, making it comparable with the remaining approaches. Furthermore, the authors raise a question that is not further discussed in the paper: "how many clusters are needed so that the clustering process is optimal?" [30]. SOM and cluster networks do not require an a priori specification of the number of clusters, and thus the data points are labeled according to a majority voting of the points belonging to the cluster they are assigned. This could also be achieved by k-means, where different numbers of clusters had to be tested. The fact that k-means was pre-defined for $k = 2$ may help explain its poor results, when compared to the other two algorithms, that do not have that limitation. In conclusion, despite proving good results, this work raises some technical questions.

In the same year, Strumbelj et al. (2010) [21] addressed the problem of BC recurrence in two directions: as a classification problem, predicting recurrence/no-recurrence events within 10 years after surgery, and as a regression problem, determining how many years would it take until cancer reappears. The latter is out of scope of this review.

Regarding the prediction of recurrence events, Strumbelj compared the performance of several well-known classifiers, namely NB, DT, SVM, RF and MLP with the evaluation of two oncologists. A bagging procedure coupled with NB was also considered. The initial data was provided by Ljubljana Institute of Oncology (not to be confused with the Breast Cancer Dataset), consisting of 1035 patients characterised by 32 features. After

removing some features (due to their redundancy) and some patients (whose follow-up was inferior to 10 years), the final dataset included 881 patients and 13 features, all categorical. Although the authors state that some of the collected features are redundant, and that not all features are considered for classification, the criteria to select the most relevant features is not depicted. Furthermore, no information was given regarding MD.

The distribution of recurrence/no-recurrence events is 51%/49%, thus class imbalance is not a constraint and accuracy is considered an appropriate evaluation metric. A 10-fold cross validation scheme was used to access the performance of the chosen ML algorithms, where DT, NB (both as a standard formulation and coupled with a bagging scheme) and RF performed similarly with accuracies ranging from 67.4% to 68%, outperforming SVM (59.9%) and MLP (60.8%). Therefore, the best approaches (DT, NB and RF) were further compared with the predictions of two medical experts (two oncologists), using a validation test of 100 randomly chosen patients. NB classifier achieved the best results (both standard and considering bagging) with an accuracy of 70%. However, overall, the accuracy results were very similar, with DT and RF obtaining accuracies of 67% and 68%. Both oncologists lagged slightly behind the ML algorithms, with accuracies of 63% and 65%. In fact, ML results did not prove to be significantly higher than the predictions of medical experts. The fact that the final dataset contains only categorical variables is a topic for discussion in this work. The authors state that their "preliminary analysis" has not shown significant differences between numerical or discretised versions of some features in prediction results, although these results are not presented in the work. This may explain the poor results achieved by SVM and MLP, which generally tend to deal better with continuous variables [97][98]. However, it has to be stated that the main objective of this work was not to achieve optimal classification results. More than building a successful model to predict BC recurrence, the aim of this work is to improve the interpretability of machine learning models and develop a method to assess their reliability. This topic is also out of scope of this study; however, it highlights the increasing interest of machine learning experts in developing accurate, yet still easy to use and interpretable strategies to be used by non-machine learning experts, in particular medical doctors, when dealing with a healthcare context.

In 2012, Woojae Kim et al. [31] studied the application of SVM, ANN and Cox-

regression model to the prediction of BC recurrence within 5 years after surgery. To assess the performance of the proposed approaches, three well-known BC prognostic models were also selected: St. Gallen's guidelines [93], NPI (Nottingham Prognostic Index) [99] and Adjuvant! Online [100]. The initial dataset was composed by 1541 patients from a tertiary hospital in South Korea. However, after discarding patients with incomplete follow-up, late-stage and male breast cancer patients, as well as patients suffering from other types of cancers (besides BC), the study population consisted of 679 patients, with 195 recurrence cases (28.7%) and 484 no-recurrence cases (71.3%). Out of 193 available features, 7 were chosen to be included in the prediction models, namely "histological grade, tumor size, number of metastatic lymph nodes, estrogen receptor (ER status), lymphovascular invasion (LVI), local invasion of tumour, and number of tumours" [31]. They were selected beforehand by the authors in collaboration with medical experts, and further refined based on Kaplan-Meier and Cox-regression analysis. The results were evaluated in terms of accuracy, sensitivity, specificity, precision, AUC, and negative predictive value (NPV), using a holdout method (70% - 30%). Regarding the computational models, SVM and ANN performed similarly, outperforming the Cox model except in terms of specificity: 73%, 52% and 94% for SVM, ANN and Cox model, respectively. ANN achieved the best sensitivity (95%) and precision (80%) results, followed by SVM with 89% and 75%, respectively. However SVM proved to be the best approach, outperforming the others in terms of NPV (89%), accuracy (84.58%) and AUC (0.85). The authors further compared the performance of SVM with the previous mentioned prognostic models: St. Gallen's, NPI and Adjuvant!. St. Gallen's achieved the highest sensitivity and NPV (100%); however, it had poor results in the remaining metrics. Similarly, Adjuvant! also returned high sensitivity and NPV results (95% and 83%), although its superiority was not verified for the other metrics. The same may be said of Cox model, which had the highest specificity (94%), but failed to keep its advantage over the other performance metrics. Thus, SVM proved a superior performance over the "classic" models for the prognosis of BC recurrence. The authors highlight that although machine learning algorithms generally achieve higher performances, their use in clinical practice is still very limited "because they cannot be easily calculated with a traditional calculator". In our opinion, they are right in that ML

are currently not used in practice, despite their undoubtedly higher performance. However, we don't agree on the reason. The real reason boils down to the interpretability again. Even if tools to calculate ML predictions are made available, medical doctor will not "trust" models they cannot fully understand and interpret. Another point mentioned by the authors is that ML algorithms can be adjusted to data. For instance, SVM hyperparameters may adjusted to different subject populations. This may bring an important advantage over traditional prognostic models that impose a universal prediction model for all races or countries.

In that same year, Gouda Salama et al. [32] compared the performance of DT, MLP, SVM, NB and KNN in the prediction of BC recurrence using the WPBC dataset (similarly to Belciug et al. the end-point is not defined). The fusion between classifiers was also explored, to assess if a multi-classifier approach could bring some benefit in terms of classification performance. The comparison between classifiers was performed using a 10-fold crossvalidation sampling scheme, and evaluating their accuracy. Among the five considered classifiers, SVM and DT outperformed all others, with an accuracy of 76.3%, followed by MLP, KNN and NB with 66.5%, 64.4% and 50.5%, respectively. Moreover, a fusion analysis of two, three and four classifiers was conducted. The first fusion considered SVM coupled with the remaining: SVM-NB, SVM-MLP, SVM-DT and SVM-KNN. All the combinations have achieved the same accuracy results: 76.3%. The fusion of three classifiers considered SVM and DT coupled with the remaining: SVM-DT-NB, SVM-DT-MLP and SVM-DT-KNN. Once again, all the combinations showed the exact same accuracy: 76.3%. Finally, the third fusion considered the coupling of SVM, DT and MLP with the remaining: SVM-DT-MLP-KNN and SVM-DT-MLP-NB. The combination of SVM-DT-MLP-KNN resulted in an improvement of accuracy, 77.3%, while SVM-DT-MLP-NB did not improve the previous results, achieving an accuracy of 74.2%. In conclusion, the fusion of SVM, DT, MLP and KNN proved to be superior when compared to the remaining combinations of classifiers and the other setups of standalone classifiers. This work shows that the combination of classifiers may be beneficial to the classification performance. However, the authors do not mention what type of combination was used (using probability results, majority voting or combination rules, for instance). Also, and as previously mentioned, WPBC is an imbalanced dataset, and

therefore more appropriate performance metrics would be required, namely sensitivity and specificity. Finally, the MD perspective is also ignored in this work, which constitutes another of its limitations.

Also in 2012, Mahantappa Murti [33] used three rule-based classifiers to predict BC recurrence within five years after surgery, namely RIPPER (Repeated Incremental Pruning to Produce Error Reduction), Decision Table and DTNB (Decision Table with Naive Bayes). To conduct the experiments, the database from the Oncology Institute of Ljubljana (Breast Cancer Dataset) was used. The initial dataset was preprocessed to remove missing values, and although the final number of patients included in the study is not mentioned, we assume that all the 9 records with missing data are eliminated, thus resulting in a final dataset of $(286 - 9)$ 277 patients. The algorithms are compared in terms of precision, recall, F-measure and AUC, for both "no-recurrence" and "recurrence" events. As previously mentioned for the case of Fan et al., when the objective is to predict recurrence (and considering that this class has always a lower number of cases), it should be defined as the positive class, and the metrics should be analysed having that in mind. Only such analysis would provide a meaningful and suitable comparison with the other studied works. For that reason, only the performance results of the "recurrence" class is analysed and compared within this review. Accordingly, RIPPER obtained 72.3%, 36.5%, 84.6%, 50%, 0.4 and 0.58 in terms of accuracy, sensitivity, specificity, precision, F-measure and AUC, respectively. In turn, Decision Table achieved 72.7%, 23.5%, 91%, 53%, 0.33 and 0.64 for the same metrics. Finally, outperforming these approaches, DTNB returned and accuracy of 75.2%, sensitivity, specificity and precision of 36.5%, 89.6% and 59.6%, respectively, while achieving a F-measure of 0.45 and AUC of 0.68. Although DTNB outperformed the other approaches, achieving a good accuracy (75.2%) and specificity (89.6%), the sensitivity results are very poor, being among the worst approaches reviewed. Similarly, the F-measure and particularly the AUC results show that this is not a feasible approach to predict BC recurrence, being only slightly better than random guessing.

In 2013, Tomczak [22] used the Classification Restricted Boltzmann Machine (Class-RBM) to predict BC recurrence within 10 years after surgery and determine input features (symptoms) relevant for disease reappearance. Several methods for learning

ClassRBM are discussed, namely DropOut, Drop Connect and DropPart [22]. These algorithms are compared to classical approaches such as NB, SVM, RF and CART (coupled with AdaBoost, Bagging, and LogitBoost). This work also counted with the collaboration of two oncologists in order to provide a comparison of machine learning techniques with the opinion of medical experts. A holdout method (70% train - 30% test) was used across all computational methods while predictions from oncologists were obtained using 100 cases of the test set. Overall, the computational approaches achieved better accuracy results than the medical experts, except for SVM, which performed poorly. ClassRBM and ensemble approaches had very similar results; however, from all the considered algorithms, the ensemble LogitBoost + CART outperformed all others with an accuracy of 75%. This work is somewhat of a follow-up of Strumbelj's study, using the same dataset provided by the Oncology Institute of Ljubljana (not Breast Cancer Dataset). However, in Tomczak's study, the final dataset is composed by 949 patients (there are more patients with a minimum follow-up of 10 years) and 15 features (the feature selection process is not discussed). These 15 features include all 13 used by Strumbelj and two more, regarding the application of two different types of therapy (cTherapy and hTherapy). All input features were binarized, resulting in a dataset composed by 55 binary features. As discussed in Strumbelj's study, the binarization of all input features could explain the poor performance of SVM. The distribution of recurrence/no-recurrence events is not depicted; however, we assume it is very similar to Strumbelj's estimates (51% - 49%). The authors do not perform a thorough discussion on the best approach to predict BC recurrence. Nevertheless, they highlight the ClassRBM's ability to retrieve relevant information regarding the most important input features while also achieving a high classification performance.

Alberto Pawlovsky and Mai Nagahashi (2014) [34] proposed a method based on scoring to select the best configuration to be used in KNN classification of WPBC dataset (the end-point was not defined). In their approach, patients with missing data are removed from the study (4 patients), and only 32 features are kept. After discussing the effects of different combinations of training size and number of neighbours and runs considered, the authors present their scoring scheme and perform its validation by addressing the BC recurrence problem. The best classification setting is chosen ac-

cording to the preprocessing method used (raw data, standardisation or normalisation), number of $k$ neighbours, number of runs, sample size for classification, average, maximum, minimum and standard deviation of the accuracy results. Their strategy provided the best results for a configuration using raw data, 19 neighbours, 80% of samples in classification and 100 simulation runs. These configurations achieved a mean accuracy of 76%, and minimum and maximum values of 62% and 90%, respectively. It is also important to note that, overall, the preprocessing method used does not significantly affect the final classification results.

Although discussing an interesting topic, this work is more focused in finding a strategy to select appropriate KNN configurations than addressing the particular problem of BC recurrence: no feature selection is performed and the class imbalance problem is not addressed (again, only accuracy results are presented). Nevertheless, it takes into account the existence of absent observations, removing them. The generalisation of this work could be a topic for further research, and its extension to include sensitivity/specificity results could possibly by a more suitable approach to the BC recurrence problem. However, as discussed in Section 2.3, it must be noted that KNN is a lazy learner, as it makes local approximations, without further generalisation, and thus the classification task for this algorithm is very time consuming. With a considerable amount of data, given the number of different combinations to be tested, it could become infeasible for real-time applications.

In the same year, Zahra Behesti et al. [35] tackled the principles of Genetic Programming, by comparing the performance of several genetic approaches when coupled with MLP: Centripetal Accelerated Particle Swarm Optimisation (CAPSO), Particle Swarm Optimisation (PSO), Gravitational Search Algorithm (GSA) and Imperialist Competitive Algorithm (ICA). These four hybrid approaches (CAPSO-MLP, PSO-MLP, GSA-MLP and ICA-MLP) were applied to nine medical datasets targeting different diseases. Among them is the WPBC dataset [16], previously presented. Before running the simulations, the dataset was normalised and absent observations handled with mean imputation. All approaches were evaluated in terms of Mean Square Error (MSE), AUC, accuracy, sensitivity and specificity, following a holdout scheme (80% train - 20% test). GSA-MLP achieved 0.167 of MSE, 0.55 of AUC, and 79.3%, 7.86% and 80.23%

of accuracy, sensitivity and specificity. In turn, ICA-MLP and PSO-MLP obtained MSE results of 0.177 and 0.173 and AUC results of 0.57 and 0.6, respectively. In terms of accuracy, sensitivity and specificity, these approaches have returned the same results: 78.3%, 43% and 83%. CAPSO-MLP achieved an MSE of 0.170 and an AUC of 0.63 while returning accuracy, sensitivity and specificity results of 80.3%, 52.3% and 83.4%, clearly outperforming all others and being considered the most suitable approach for unseen data.

According to the authors, the adjustment of the parameters of PSO algorithm is time-consuming. The CAPSO approach was created to solve this problem, by using less *a priori* parameters, resulting in a simplified tuning process (more automated). The inclusion of the original PSO approach was an important step, to evaluate if the new technique (CAPSO) improves the results by comparison. However, applying only this type of algorithms does not provide a real assessment of their performance. There should have been a setup including a more traditional approach as a baseline for comparison, e.g. backpropagation, in accordance with other authors cited in this work [101][102],[103] [104] [105]. This would present the opportunity to compare the two different methodologies, verifying whether the proposed algorithms generate better results. Moreover, the chosen algorithms do not agree with the literature review of this article: GSA is not referred in any of the cited articles, while others were inexplicably left out (e.g. Artificial Immune System, Ant Colony Optimisation, and Artificial Bee Colony). ICA and GSA happen to be the two most recent approaches mentioned, but there was no explicit indication of the reason to choose them. Nevertheless, the used algorithms are thoroughly explained, which is especially important in modern techniques.

Still in 2014, Vikas Chaurasia et al. [36] investigated the performance of DT, ANN and LR in BC recurrence within five years after surgical intervention, using the previously described Breast Cancer Dataset to conduct their experiments. The results were evaluated through a 10-fold crossvalidation procedure, by determining the accuracy, true positive rate, false positive rate, precision and recall for both “recurrence” and “no-recurrence” classes. As discussed in previous works (Fan et al., Murti), the class of interest is “recurrence”. For that reason, and to allow an appropriate comparison between all the research works, we strict our analysis to accuracy, sensitivity, specificity

and precision results considering “recurrence” as the positive class. In terms of accuracy, specificity and precision, LR performed the best, with 74.5%, 92.5% and 64.3%, respectively, versus the 71.3%, 92% and 54.3% of DT and 73..8%, 88.6% and 58.9% of ANN. Regarding sensitivity, ANN was the best approach, with 38.8%, over the 31.8% and 22.4% obtained by LR and DT, respectively. Although ANN achieves the best sensitivity results (is the best classifier in identifying recurrence events), LR is overall the best approach, outperforming all others in terms of accuracy, specificity and precision. The authors have also analysed the impact of the chosen feature to recurrence prediction, which revealed that tumour grade is the most explanatory features, followed by lymph nodes involvement, node capsules, tumour size, irradiation, age, breast quad, breast and menopause. As many of the works using WPBC dataset, the fact that this works neglects the MD perspective is its main weakness.

Table 4 presents a resume of the ML algorithms used in each research work and the performance results of the best approach (highlighted in bold). The results are measured in terms of accuracy (Acc), sensitivity (Sen), specificity (Spe) and AUC values. The strategies used for data sampling and handling MD are also depicted.

Table 4: Comparison of ML algorithms (bold indicates the one that presented the best performance), achieved results and sampling strategies used in the analyzed studies.

| Publications | Algorithms | Acc | Sen | Spe | AUC | MD | Sampling Strategy |
|---|---|---|---|---|---|---|---|
| Mani et al., 1997 [18] | **NB**, CART, C4.5, C4.5 rules, FOCL | 68.30% | – | – | – | – | Stratified Random Subsampling Construction of 6 datasets (40%-60%) |
| Jerez-Aragonés et al., 2003 [24] | **ANN+DT** | 93,4%-96% | 78,7%-88.7% | 94,5%-97.2% | – | Removed | 10-fold Cross-validation |
| Razavi et al., 2005 [23] | **C4.5+CCA**, C4.5 + EM, C4.5 | 67% | 80% | 63% | – | EM | 10-fold Cross-validation |
| Razavi et al., 2007 [19] | **C4.5+CCA**, Two oncologists | 82% | 21.10% | 96.30% | 0.76 | MI | 10-fold Cross-validation Stratified Random Sampling for validation |
| Sun et al., 2007 [20] | **LDA+hybrid**, LDA+genetic, LDA+clinical, St.Gallen criterion | – | 90% | 67% | – | Unknown | Leave-One-Out |
| Ryu et al., 2007 [28] | **Isotonic Separation**, Robust Linear Programming, DT (C4.5, OC1, QUEST), SVM, AdaBoost, Learning Vector Quantisation | 80% | – | – | – | – | Holdout (70% - 30%) |
| Jonsdottir et al., 2008 [15] | **C4.5**, NB, LMT, REP tree, RF, SVM, Log, Slog, MetaClass1, MetaClass2, MetaClass3, Bag +REP tree, DT, OneR, PART, Jrip, VFI | 79% (Small-DS) | 48% (Base-DS) | 96% (Small-DS) | 0,70 (Base-DS) | Not mentioned, but handled by algorithms | 10-fold Cross-validation |
| Fan et al., 2010 [29] | **C5.0**, ANN, CHAID, CART, QUEST | 71.20% | – | – | – | Removed | Holdout (80% - 20%) |
| Belciug et al., 2010 [30] | **Cluster network**, k-means, SOM | 78% | – | – | – | Unknown | 10-fold Cross-validation |

*Continued on next page*

Table 4 – *Comparison of the DM algorithms, achieved results and sampling strategy used in the analyzed studies*

| Publications | Algorithms | Acc | Sen | Spe | AUC | MD | Sampling Strategy |
|---|---|---|---|---|---|---|---|
| Strumbelj et al., 2010 [21] | **NB**, **NB+Bagging** DT, SVM, RF | 70% | – | – | – | Considered as a separate feature value. | 10-fold Cross-validation |
| Kim et al., 2012 [31] | **SVM**, ANN, Cox model, St. Gallen, NPI, Adjuvant! | 84.58% | 89% | 73% | 0.85 | Removed | Holdout (70% - 30%) |
| Salama et al., 2012 [32] | **SVM-DT-MLP-KNN**, DT, MLP, SVM, NB, KNN, SVM-NB, SVM-MLP, SVM-DT, SVM-KNN, SVM-DT-NB, SVM-DT-MLP, SVM-DT-KNN, SVM-DT-MLP-NB | 77.30% | – | – | – | Unknown | 10-fold Cross-validation |
| Murti, 2012 [33] | **DTNB**, RIPPER, Decision Table | 75.17% | 37% | 90% | 0.676 | Removed | Unknown |
| Tomczak, 2013 [22] | **CART+LogitBoost**, ClassRBM, ClassRBM+DropOut, ClassRBM+DropConnect, ClassRBM+DropConnect, CART+Bagging, CART+AdaBoost, NB, SVM, RF, Two oncologists | 75% | – | – | – | Unknown | Holdout (70% - 30%) |
| Pawlowsky and Nagahashi, 2014 [34] | **KNN** | 76% | – | – | – | Removed | n.a |
| Beheshti et al., 2014 [35] | **CAPSO-MLP**, PSO-MLP, GSA-MLP, ICA-MLP | 80.25% | 52.33% | 83.38% | 0.63 | Mean | Holdout (80% - 20%) |
| Vikas C. et al., 2014 [36] | **LR**, C4.5, ANN | 74.50% | 31.80% | 92.50% | – | Unknown | 10-fold Cross-validation |

## 4 Discussion

Predicting BC recurrence is a very important challenge for oncological clinicians because it has direct influence in their daily practice e.g. in choosing the most beneficial treatment for a patient. Over the past decade, several works have tried to propose suitable approaches to model BC behaviour; however, after performing this revision, it is clear that this is still an open problem. This observation is based on five problems detected in the Reviewed Works (RW):

*Lack of data.* The majority of RW used local datasets (datasets that contain only data from a local/regional center), which complicates the replication and further comparison of results by other researchers. Also, the number of patients enrolled in most of these studies can be considered small (less than 1000 patients), especially for a common pathology like BC. The reduced size of the datasets becomes even more critical when most of the works does not deal with MD, either at all (more than 80%) or with proper thoroughness. Only three research works have addressed this issue (Razavi et al., 2005, 2007 [23][19]; Beheshti et al. [35]).

*Imbalanced Binary Decision Problem.* The second problem, as mentioned in the Introduction, is that the prediction of BC recurrence is a binary classification problem where the goal is to accurately predict whether a BC patient will or will not recur. To achieve that, these two classes should be balanced (have similar proportions in the dataset); otherwise, the algorithms could predict one class better than the other. From our analysis, it can be noted that the majority of the RW presented imbalanced datasets, which will somehow degrade the performance of ML techniques. This point could be easily overcome by using appropriate sampling strategies to balance data, such as Synthetic Minority Oversampling Technique (SMOTE) [106].

*Feature Selection.* The third problem concerns feature selection. Only a small number of works used computational feature selection techniques. Most of the RW use a manual feature selection process, in which medical doctors are consulted to select the variables to use in the prediction studies. However, this process has one great disadvantage: the information that the algorithms are able to find is exactly what they

were expected to find: the doctors select the variables using previously established knowledge or informed intuition, which may prevent potentially useful variables from being used in the models, and new relationships between variables and recurrence to be found. Also, it is important to note that only one study tried to mix clinical markers with genetic information [20]. Many of the variables selected in the RW were not mutually exclusive (e.g. BC stage is the conjunction of tumour size, lymph nodes involved and presence or absence of metastasis) and some were not routinely described as important recurrence factors (e.g. tumour location included breast regions and laterality), which could return somewhat misleading conclusions. Moreover, important factors that must be present in daily clinical practice are missing, such as HER2 expression. The determination of HER2 expression is mandatory for the definition of intrinsic subgroups that define BC behaviour. Even in the most frequent BC subgroup (that express hormonal receptors - HR), patients with HER2 enriched tumors are associated with a high rate of brain, liver and lung metastasis [107]. These tumours display also different patterns of relapse and metastatic spread depending on HR status, with a median relapse free survival of 19.5 months after surgery in HR negative patients compared to 32.0 months in HR positive. Patients with HER2+/HR- disease have significantly increased hazard of early (0-2 and 2-5 years), but not late death ($>$5 years), when compared to HER2+/HR+ [4]. The non-standardization of such set also hampers direct comparisons between studies and compromises future investigations in the field. A consensus in the definition of important variables to study and its validation over appropriate datasets is still a current challenge.

*Interpretability.* The fourth problem is interpretability, an important concept in the healthcare area. If the expert/clinician cannot validate the approach, it will never be accepted by the community as a valid one. This sometimes leads to a scenario where researchers try to find a trade-off between interpretability and performance for their approaches. Accordingly, it is not surprising that 13 out of the 17 RW used ML techniques that are well known for their interpretability, like Decision Trees. However, other techniques that are in the opposite side (traditionally achieving higher performance, although less interpretable) have not been neglected, such as ANN. The comparison between these different methods is impossible due to a number of factors: the used

datasets are different, the selected set of features and algorithms do not always match, and finally the evaluation metrics used are not always the same. Some studies even use clinicians to validate their approaches. However, hybrid or combination algorithms generally seem to be among the best approaches.

*Evaluation Metrics.* Finally, regarding the metrics used in the evaluation phase, it is quite surprising how 8 of the 17 RW only used accuracy to measure classification performance, specially considering the class imbalance present in the associated datasets. Accuracy is not the most appropriate metric for imbalanced datasets, since it does not properly identify the true positive and true negative rates (i.e sensitivity and specificity). When considering other studies that present both accuracy and sensitivity results, it can be noted that it is easier to achieve a good accuracy performance than sensitivity results (only three of the RW presented good sensitivity). This may also be explained by the imbalanced distribution between "recurrence" and "no-recurrence" cases.

## 5 Conclusions and Future Work Challenges

As discussed in this revision, predicting recurrence is a key point in the BC context. However, and in spite of the fact that researchers have tried to address this topic in the past decade, it remains an open challenge. Based on the previously analysed works, this can be justified by several factors: it is difficult to obtain a representative dataset for BC recurrence; there is no standard characterisation of patients (a predefined set of variables) for this disease and there is still a long path to follow regarding the use of appropriate ML algorithms, especially when it comes to high performance results and generalisation ability. Regarding some future directions, a multi-center study can be created to include the highest number of patients as possible. Also, with the help of clinicians, a consensus regarding the most important predictors for BC recurrence could be achieved. From the MD perspective, the use of previously implemented approaches in other oncology centers, trying to obtain a generalisation of results, can constitute a good contribution. Another possible direction is to try to combine different techniques for performance improvement, following to work proposed by Jerez-Aragonés [24]. Finally, the development of new ML algorithms or the exploration of ML algorithms never used

in this context may also constitute a valid perspective.

## References


[1] Organization, W. H. Globocan 2012: Estimated cancer incidence, mortality and prevalence worldwide in 2012 (2012).

[2] Moody, S. E., Perez, D., Pan, T. C., Sarkisian, C., Portocarrero, C. P., Sterner, C., Notorfrancesco, K., Cardiff, R., and Chodosh, L. (2005) *Cancer Cell* **8(3)**, 197–209.

[3] Mendonza, E. (2013) *Journal of cancer research and clinical oncology* **139(4)**, 645–652.

[4] denHurk, C. V. (2011) *Breast cancer research and treatment* **128(3)**, 795–805.

[5] Reuters, T. Web of science (2015).

[6] In, H., Bilimoria, K. Y., Stewart, A. K., Wroblewski, K. E., Posner, M. C., Talamonti, M. S., and Winchester, D. P. (2014) *Annals of Surgical Oncology* **21**, 1520–1529.

[7] Farr, A., Wuerstlein, R., Heiduschka, A., Singer, C. F., and Harbeck, N. (2013) *Reviews in Obstetrics and Gynecology* **6(3)**, 165–173.

[8] Kononenko, I. (2001) *Artificial Intelligence in Medicine* **23**, 89–109.

[9] Cruz, J. A. and Wishart, D. S. (2006) *Cancer Informatics* **2**, 59–77.

[10] Kouroua, K., Exarchosa, T. P., Exarchosa, K. P., Karamouzisc, M. V., and Fotiadisa, D. I. (2015) *Computational and Structural Biotechnology Journal* **13**, 8–17.

[11] Fayyad, U., Piatetsky-Shapiro, G., and Smyth, P. (1996) *Artificial Intelligence Magazine* **17(3)**, 37–54.

[12] Institute, S. Sas enterprise miner semma (2015).

[13] Chapman, P., Clinton, J., Kerber, R., Khabaza, T., Reinartz, T., Shearer, C., and Wirth, R. Crisp-dm 1.0: Step-by-step data mining guide. the crisp-dm consortium (2000).

[14] Azevedo, A. and Santos, M. F. (2008) In In Proceedings of Informatics and Data Mining : pp. 182–185.

[15] Jonsdottir, T., Hvannberg, E. T., Sigurdsson, H., and Sigurdsson, S. (2008) *Expert Systems with Applications* **34(1)**, 108–118.

[16] Lichman, M. Uci machine learning repository (2015).

[17] vanťVeer, L. J., Dai2, H., van deVijver, M. J., He, Y. D., Hart, A. A. M., Mao, M., Peterse, H. L., van derKooy, K., nadA. T. Witteveen, M. J. M., Schreiber, G. J., Kerkhoven, R. M., Roberts, C., Linsley, P. S., Bernards, R., and Friend, S. H. (2002) *Nature* **415**, 530–536.

[18] Mani, S., Pazzani, M. J., and West, J. (1997) *Artificial Intelligence in Medicine* **1211**, 130–133.

[19] Razavi, A. R., Gill, H., Ahlfeldt, H., and Shahsavar, N. (2007) *Journal of Medical Systems* **31(4)**, 263–273.

[20] Sun, Y., Goodison, S., Li, J., Liu, L., and Farmerie, W. (2007) *Bioinformatics* **23(1)**, 30–37.

[21] trumbelj, E., Bosni, Z., Kononenko, I., Zakotnik, B., and Kuhar, C. G. (2010) *Knowledge Information System* **24**, 305–324.

[22] Tomczak, J. M. (2013) *CoRR* **abs/1308.6324**, 9 pages.

[23] Razavi, A. R., Gill, H., Ahlfeldt, H., and Shahsavar, N. (2005) A data pre-processing method to increase efficiency and accuracy in data mining In Artificial Intelligence in Medicine volume **3581**, pp. 434–443 Springer Berlin Heidelberg.

[24] Jerez-Aragonés, J. M., Gomez-Ruiz, J. A., Ramos-Jimenez, G., Munoz-Perez, J., and Alba-Conejo, E. (2003) *Artificial Intelligence in Medicine* **27(1)**, 45–63.

[25] Suarez-Alvarez, M. M., Pham, D.-T., Mikhail, Y., and Prostov, Y. I. (2012) *Proceedings of the Royal Society of London A: Mathematical, Physical and Engineering Sciences* **468(2145)**, 2630–2652.

[26] Shalabi, L. A. and Shaaban, Z. (2006) In Proceedings of the International Conference on Dependability of Computer Systems : pp. 207–214.

[27] García-Laencina, P. J., Abreu, P. H., Abreu, M. H., and Afonso, N. (2015) *Computers in Biology and Medicine* **59(2015)**, 125–133.

[28] Ryu, Y. U., Chandrasekaran, R., and Jacob, V. S. (2007) *European Journal of Operational Research* **181(2)**, 842–854.

[29] Fan, Q., Zhu, C. J., and Yin, L. (2010) In In Proceedings of International Conference on Bioinformatics and Biomedical Technology : pp. 310–311.

[30] Belciug, S., Gorunescu, F., Salem, A. B., and Gorunescu, M. (2010) In International Conference on Intelligent Systems Design and Applications (ISDA) : pp. 533–538.

[31] Kim, W., Kim, K. S., Lee, J. E., Noh, D. Y., Kim, S. W., Jung, Y. S., Park, M. Y., and Park, R. W. (2012) *Journal of Breast Cancer* **15(2)**, 230–238.

[32] Salama, G. I., Abdelhalim, M. B., and Zeid, M. A. E. (2012) In In Proceedings of International Conference on Computer Engineering & Systems (ICCES) : pp. 180–185.

[33] Murti, M. S. (2012) *Journal of Imformation Engineering and Applications* **2(2)**, 12–19.

[34] Pawlovsky, A. P. and Nagahashi, M. (2014) In In Proceedings of IEEE-EMBS International Conference on Biomedical and Health Informatics : pp. 189–192.

[35] Beheshti, Z., Shamsuddin, S. M. H., Beheshti, E., and Yuhaniz, S. S. (2014) *Soft Computing* **18(11)**, 2253–2270.

[36] Chaurasia, V. and Pal, S. (2014) *International Journal of Computer Science and Mobile Computing* **3(1)**, 10–22.

[37] Little, R. J. A. and Rubin, D. B. (2002) Statistical Analysis with Missing Data, Wiley, 2 edition.

[38] García-Laencina, P. J., Sancho-Gómez, J. L., and Figueiras-Vidal, A. (2010) *Neural Computing & Applications* **19(2010)**, 263–282.

[39] Cismondi, F., Fialho, A. S., Vieira, S. M., Reti, S. R., Sousa, J. M., and Finkelstein, S. N. (2013) *Artificial Intelligence in Medicine* **58(1)**, 63–72.

[40] Dempster, A. P., Laird, N. M., and Rubin, D. B. (1977) *Journal of the Royal Statistical Society* **39(1)**, 1–38.

[41] Bishop, C. M. (2006) Pattern Recognition and Machine Learning, Springer, .

[42] Zio, M. D., Guarnera, U., and Luzi, O. (2007) *Computational Statistics and Data Analysis* **51(11)**, 5305–5316.

[43] Rubin, D. B. (2004) Multiple imputation for nonresponse in surveys, John Wiley & Sons, .

[44] Mitchell, T. M. (1997) Machine Learning, McGraw-Hill, .

[45] Quinlan, J. R. (1993) C4.5:Programs for Machine Learning, Morgan Kaufmann, .

[46] Kantardzic, M. (2011) Data Mining: Concepts, Models, Methods, and Algorithms, Wiley-IEEE Press, 2 edition.

[47] Breiman, L., Friedman, J., Olshen, R., and Stone, C. (1984) Classification and Regression Trees, Wadsworth & Brooks, Monterey, CA, .

[48] Patel, B. R. and Rana, K. K. (2014) *Journal of Engineering Development and Research* **2(1)**, 5 pages.

[49] Luttrell, S. P. (1994) *IEE Proc Vision, Image Signal Process* **141(4)**, 251–260.

[50] Lee, S.-M. and Abbott, P. A. (2003) *Journal of Biomedical Informatics* **36(2003)**, 389–399.

[51] Kleinbaum, D., Klein, M., and Pryor, E. (2002) Logistic Regression: A Self-Learning Text. Statistics for Biology and Health Series, Springer-Verlag, .

[52] Jain, A. and Dubes, R. (1988) Algorithms for clustering data, Prentice-Hall, Inc, Upper Saddle River, NJ, USA.

[53] Jain, A. K. (2010) *Pattern Recognition Letters Journal* **31(8)**, 651–666.

[54] deSá, J. M. (2001) Pattern Recognition: concepts, methods and applications,, Springer-Verlag, .

[55] Efron, B. and Tibshirani, R. (1994) An Introduction to the Bootstrap, Chapman and Hall/CRC, 1 edition.

[56] Breiman, L. (1996) *In Machine Learning* **24(2)**, 123–140.

[57] Breiman, L. (1998) *The Annals of Statistics Journal* **26(3)**, 801–849.

[58] Schapire, R. E. (1990) *Machine Learning* **5(2)**, 197–227.

[59] Kumar, B. S. (2012) *International Journal of Advanced Research in Computer Science and Software Engineering* **2(10)**, 27–35.

[60] Freund, Y. and Schapire, R. E. (1995) In In Proceedings of the Second European Conference on Computational Learning Theory : pp. 23–37.

[61] Friedman, J., Hastie, T., and Tibshirani, R. (2000) *Annals of Statistics* **28(2)**, 337–407.

[62] Fisher, R. A. (1936) *Annals of Eugenics* **7(2)**, 179–188.

[63] Rao, C. R. (1948) *Journal of the Royal Statistical Society. Series B (Methodological)* **10(2)**, 159–203.

[64] Vapnik, V. 11 1999 The Nature of Statistical Learning Theory (Information Science and Statistics), Springer, 2 edition.

[65] Boser, B., Guyon, I., and Vapnik, V. (1992) In In Proceedings of the annual workshop on Computational learning theory : pp. 144–152.

[66] Shawe-Taylor, J. and Cristianini, N. (2004) Kernel Methods for Pattern Analysis, Cambridge University Press, .

[67] Scholkopf, B. and Smola, A. (2002) Learning with Kernels, MIT Press, Cambridge MA, .

[68] Altman, N. S. (1992) *The American Statistician* **46(3)**, 175–185.

[69] Agrawal, R., Imielinski, T., and Swami, A. N. (1993) In In Proceedings of the 1993 ACM SIGMOD International Conference on Management of Data : pp. 207–216.

[70] Molina, C., Prados-Suarez, B., Prados, D. R. M., and Pena, Y. C. (2013) *Studies in health technology and informatics* **197**, 91–95.

[71] Chandrasekaran, R., Ryu, Y. U., Jacob, V. S., and Hong, S. (2005) *INFORMS Journal on Computing* **17(4)**, 462–474.

[72] Ryu, Y. U., Chandrasekaran, R., and Jacob, V. S. (2007) *European Journal of Operational Research* **181**, 842–854.

[73] Breiman, L. (2001) *Machine Learning Journal* **45**, 5–32.

[74] Trevor, H., Tibshirani, R., and Friedman, J. (2009) The elements of Statistical Learning: Data Mining, Inference, and Prediction, Springer Series in Statistics, .

[75] Verikas, A., Gelzinis, A., and Bacauskiene, M. (2011) *Pattern Recognition* **44(2)**, 330–349.

[76] McCulloch, W. and Pitts, W. (1943) *Bulletin of Mathematical Biophysics* **5(4)**, 115–133.

[77] Minsky, M. and Papert, S. (1969) An Introduction to Computational Geometry, MIT Press - ISBN 0-262-63022-2, .

[78] García-Laencina, P. J., Sancho-Gómez, J. L., and Figueiras-Vidal, A. R. (2013) *Expert Systems with Applications* **40(4)**, 1333–1341.

[79] Kohonen, T. (1995) Self-Organizing Maps, Springer, Berlin, Heidelberg, .

[80] Fischer, A. and Igel, C. (2014) *Pattern Recognition* **47(1)**, 25–39.

[81] Larochelle, H. and Bengio, Y. (2008) In Proceedings of the 25th international conference on machine learning : pp. 536–543.

[82] Larochelle, H., Mandel, M., Pascanu, R., and Bengio, Y. (2012) *Journal of Machine Learning Research* **13(1)**, 643–669.

[83] Mitchell, M. (1996) An introduction to genetic algorithms, MIT Press, .

[84] Duda, R. O., Hart, P. E., and Stork, D. G. (2012) Pattern classification, John Wiley & Sons, 2 edition.

[85] Han, J., Kamber, M., and Pei, J. (2011) Data mining: concepts and techniques: concepts and techniques, Morgan Kaufmann, 3 edition.

[86] Arlot, S. and Celisse, A. (2010) *Statistics surveys* **4**, 40–79.

[87] J.Huang (2005) *IEEE Transactions on Knowledge and Data Engineering* **17(3)**, 290–310.

[88] Intrator, O. and Intrator, N. (2001) *Computational statistics & data analysis* **37(3)**, 373–393.

[89] Zhou, Z. H. and Jiang, Y. (2003) *IEEE Transactions on Information Technology in Biomedicine* **7(1)**, 37–42.

[90] Nature Nature journal (2015).

[91] Guo-Zheng, L. (2011) Machine learning for clinical data processing In Machine Learning: concepts, methodologies, tools and applications pp. 875–897 IGI Global.

[92] Sun, Y. (2007) *IEEE Transactions on Pattern Analysis and Machine Intelligence* **29(6)**, 1035–1051.

[93] Harbeck, N., Thomssen, C., and Gnant, M. (2013) *Breast Care* **8(2)**, 102–109.

[94] Dettling, M. and Buhlmann, P. (2004) *Journal of Multivariate Analysis* **90(1)**, 106–131.

[95] Gevaert, O., Smet, F. D., Timmerman, D., Moreau, Y., and Moor, B. D. (2006) *Bioinformatics* **22(14)**, 184–190.

[96] Research, S. Surveillance, epidemiology, and end results (seer) program (2015).

[97] Irshad, H., Gouaillardd, A., Rouxa, L., and Racoceanub, D. (2014) *Computerized Medical Imaging and Graphics* **38(5)**, 390–402.

[98] Kotsiantis, S. B. (2007) *Informatica* **31**, 249–268.

[99] Galea, M. H., Blamey, R. W., Elston, C. E., and Ellis, I. O. (1992) *Breast cancer research and treatment* **22(3)**, 207–219.

[100] Olivotto, I. A., Bajdik, C. D., Ravdin, P. M., Speers, C., Coldman, A. J., Norris, B. D., and Gelmon, K. A. (2005) *Journal of Clinical Oncology* **23(22)**, 2716–2735.

[101] Chau, K. W. (2007) *Automation in Construction* **16(5)**, 642–646.

[102] Socha, K. and Blum, C. (2007) *Neural Computing and Applications* **16(3)**, 235–247.

[103] Ozkan, C., Kisi, O., and Akay, B. (2011) *Irrigation Science* **29(6)**, 431–441.

[104] Ahmadi, M. A., Ahmadi, M. R., and Shadizadeh, S. R. (2013) *Neural Computing & Applications* **13(2)**, 1–9.

[105] Mahmoudi, M. T., Taghiyareh, F., Forouzideh, N., and Lucas, C. (2013) *Neural Computing and Applications* **22(1)**, 1–16.

[106] Chawla, N. V., Bowyer, K. W., Hall, L. O., and Kegelmeyer, W. P. (2002) *Journal of Artificial Intelligence Research* **16(2002)**, 321–357.

[107] Beca, F., Santos, R., Vieira, D., Zeferino, L., Dufloth, R., and Schmitt, F. (2014) *Pathology - Research and Practice* **210(9)**, 571–575.

# Chapter 3

# Predictive Factors of Response to Therapy

Treatment approaches in MBC patients have been extrapolated from female, in the absence of trials in MBC [1].

Actually, there are already some differences regarding endocrine therapy between genders. In MBC patients with tumours that express hormonal receptors, the standard endocrine therapy in early and advanced disease is tamoxifen [2]. Aromatase inhibitors (AI), namely anastrozole did not appear to have as a complete estrogen suppression as in women, and also raise testosterone levels by 58% [3], leading to an increase of androgens that become available for conversion to estrogen [4]. In adjuvant setting there was a retrospective study that compared tamoxifen and AI in MBC patients [5]. For a total of 257 men with BC in stage I or II, 50 were treated with AI and 207 with tamoxifen with a higher risk of death for AI group (HR: 1.55, 95% IC:1.13-2.13).

The evidence to tamoxifen use in adjuvant setting in MBC patients come from retrospective studies [6–8] that demonstrated survival benefit (17% improvement in 5-year overall survival) [9] compared to controls that did not perform any endocrine therapy. In that time, the duration of tamoxifen was less than 2 years, and so these results could probably underestimate the real impact of this drug in male, which may improve with longer duration of therapy. The longer against shorter (ATLAS) trial in women [10], than supported the administration of 10 rather that 5 years of tamoxifen, did not include MBC patients, and so the decision on whether to continue treatment beyond five years in these patients should be based on an individual consideration having into account the side effects. These adverse symptoms may include: decreased libido (22-29%), weight gain (22-25%), hot flashes (13-21%), mood alterations (3-21%) and depression symptoms (3-17%) [11, 12], and are responsible for a large proportion of treatment discontinuations before 5 years with a negative impact in outcomes [11–13] . The adherence of MBC patients falls from

65% at first year to 18% at the fifth year [13] and low adherence was associated with lower 10-year rate of DFS (42 vs 73%) and OS (50 vs 80%) [11].

Comprehensive analysis of tamoxifen and 22 of its metabolites confirm that endoxifen is the most abundant active metabolite of tamoxifen [14–16]. Bioactivation of tamoxifen to endoxifen is mediated by a multitude of cytochrome P450 (CYP) enzymes, with CYP2D6 being central to metabolic activation [17, 18]. There are some geographic variations described in this expression [19] and over the years more than 25 reports investigating whether or not CYP2D6 genotype influences the efficacy of tamoxifen treatment in FBC patients have been published [20]. The principal studies about this issue are summarized in Table 3.1.

The conclusions of these studies are contradictory and range from a possible longer DFS interval to a substantially shorter recurrence-free survival time for patients carrying CYP2D6 genotypes conferring diminished tamoxifen metabolism. The inconsistency in these results is likely attributable to heterogeneity in study designs and patients populations regarding to: DNA sources (tumour, blood), phenotypes evaluated (efficacy, toxicity, pharmacokinetics), study designs (retrospective, prospective), different treatment settings (adjuvant and palliative), different tumours (with and without expression of hormonal receptors), and different concomitant medications (other anticancer medications and CYP2D6 inhibitors) [43]. These findings have led to confusion among clinicians regarding whether or not CYP2D6 genotyping should be performed in daily practice, and Peppercorn et al. [44] reported that in USA, 31% of oncologists ordered this test outside a clinical trial and 14% reported routine use of the test in their clinical practice.

In MBC there was no study that addressed the role of CYP2D6 polymorphism in the efficacy of tamoxifen despite of the importance of this drug in these patients' approach.

Table 3.1: Studies that addressed the impact of CYP2D6 genotypes in tamoxifen efficacy in FBC patients

| Reference | n | FU, years | Age, years | Not-Menop. | BC stage | ER neg | Tam. dose (duration) | CYP2D6 | Outcome | Impact |
|---|---|---|---|---|---|---|---|---|---|---|
| Nowel et al. 2005 [21] | 162 | 5.4 | >50 (59%) | NR | I-IV | 14% | NR | *3,4,6 | PFS, OS | No |
| Wegman et al. 2007 [22] | 238 | 7.1 | 69 | 0% | II/III | 0% | 20 or 40 (2 or 5) | *4 | RecFS | No |
| Okishiro et al. 2009 [23] | 173 | 4.7 | 47 | 78% | NR | 9% | 20 (4.3) | *10 | RecFS | No |
| Toyama et al. 2009 [24] | 154 | 7.9 | 59 | NR | NR | 3% | 20 (3.2) | *10 | DFS, DDFS, BCSS, OS | No |
| Stingl et al. 2010 [25] | 493 | 7.0 | 59 | NR | NR | 0% | 20 (3.4) | *4 | TTP, PFS | No |
| Lash et al. 2011 [26] | 1682 | 1-10 | NR | 6% | I-III | 5% | 20 (1,2 or 5) | *4 | BCRec | No |
| Park et al. 2012 [27] | 716 | 5.6 | 45 | NR | I-II | 3% | 20 (4.4) | *5,10,41 | RecFS | No |
| Rae et al 2012 [28] | 588 | 10 | NR | 0% | I-IIIA | 0.3% | 20 (5) | *1,2,3,4,6, 10,41 | Rec rates | No |
| Regan et al 2012 [29] | 2193 | 6 | 61 | 0% | NR | 1% | 20 (5) | *2,3,4,6,7,10,17,41 | BCFI | No |
| Wegman et al. 2005 [30] | 112 | 10.7 | NR | 0% | NR | 0% | 40 (2) | *4 | DRecFS | Yes |
| Goetz et al. 2007 [31] | 223 | 11 | 68 | 0% | NR | 0% | 20 (5) | *4 | TTRec, RecFS, DFS, OS | Yes |
| Schroth et al. 2007 [32] | 486 | 6.4 | 60 | NR | NR | 0% | NR | *4,5,10,41 | RFT, EFS, OS | Yes |
| Xu et al. 2008 [33] | 293 | 5.3 | NR | NR | 0-III | 9% | 20 (5) | *10 | DFS, DSS | Yes |
| Newman et al. 2008 [34] | 205 | 10 | 41 | NR | NR | 23% | 20 (>4) | *3,4,5,41 | TTRec, RecFS, OS | Yes |
| Bijl et al. 2009 [35] | 85 | NR | 76 | NR | NR | 0% | 20 or 40 (2.1) | *4 | BCM, CM, M | Yes |
| Schroth et al. 2009 [36] | 1325 | 6.3 | 66 | 4% | I-III | 3% | NR (5) | *3,4,5,10,41 | TTRec, DFS, EFS, OS | Yes |
| Ramon et al. 2010 [37] | 91 | 9.0 | NR | 43% | NR | 0% | NR | 27 alleles | DFS | Yes |
| Abraham et al. 2010 [38] | 3155 | 6.0 | 53 | NR | I-IV | 7% | 20 (NR) | *4,5,6,9, 10,41 | BCSS, OS | Yes |
| Lammers et al. 2010 [39] | 99 | NR | 51.8 | NR | NR | 0% | 20 (NR) | *3,4,5,6, 10,41 | TTP, OS | Yes |
| Kyiotani et al. 2010 [40] | 282 | 7.1 | 51 | 44% | NR | 8.9% | 20 (5) | *4,6,10,14,18,21,36,41 | RecFS | Yes |
| Thompson et al. 2011 [41] | 618 | 4.9<br>9.4 | 60.5<br>63.1 | 19% | I-III | 0% | 20 (5) | 33 alleles | RFS | Yes |
| Teh et al. 2011 [42] | 95 | NR | 51 | 34.7% | I-IV | 17.9% | 20 (NR) | *4,5,10,14 | Rec. risk | Yes |

BC: breast cancer; BCFI: breast cancer free interval; BCM: breast cancer mortality; BCRec: breast cancer recurrence; BCSS: breast cancer specific-survival; CM: cancer mortality; DFS: disease-free survival; DDFS: distant disease-free survival; DRecFS: distant recurrence-free survival; DSS: disease-specific survival; EFS: event-free survival; ER: estrogen receptor; FU: follow-up, median values in years; M: mortality; Not-Menop: pre-menopausal; NR: Not Reported; OS: overall survival; PFS: progression-free survival; PR: progesterone receptor; RecFS: recurrence-free survival; Rec Rates: recurrence rates; RFT: relapse-free time; Tam.dose: Tamoxifen dose (mg); TTP: time to progression; TTRec: time to recurrence.

# References


[1] Kornegoor R, Verschuur-Maes AH, Buerger H, et al. (2012) Molecular subtyping of male breast cancer by immunohistochemistry. Modern Pathology, **25(3)**, 398-404.

[2] Dietz JR, Partridge AH, Gemignani ML, et al. (2015) Breast cancer management updates. Young and older, pregnant, or male. Annals of Surgical Oncology, **22(10)**, 3219-3224.

[3] Mauras N, O'Brien KO, Klein KO, Hayes V. (2000) Estrogen suppression in males: metabolic effects. Journal of Clinical Endocrinology and Metabolism, **85(7)**, 2370-2377.

[4] Czene K, Bergqvist J, Hall P, Bergh J. (2007) How to treat male breast cancer. Breast, **16**, S147-S154 (suppl 2).

[5] Eggermann H, Ignatov A, Smith BJ, et al. (2013) Adjuvant therapy with tamoxifen compared to aromatase inhibitors for 257 male breast cancer patients. Breast Cancer Res Treat, **137(2)**, 465-470.

[6] Ribeiro GG, Swindell R, Harris M, et al. (1996) A review of the management of the male breast carcinoma based on an analysis of 420 treated cases. Breast, **5(3)**, 141-146.

[7] Ribeiro G, Swindell R. (1992) Adjuvant tamoxifen for male breast cancer (MBC). British Journal of Cancer, **65(2)**, 252-254.

[8] Goss PE, Reid C, Pintilie M, et al. (1999) Male breast carcinoma: a review of 229 patients who presented to the Princess Margaret Hospital during 40 years. 1995-1996. Cancer, **85(3)**, 629-639.

[9] Young I, Kurian K, Mackenzie M, et al. (2000) The CAG repeat within the androgen receptor in male breast cancer patients. Journal of Medical Genetics, **37(2)**, 139-140.

[10] Davies C, Pan H, Godwin J, et al. (2013) Long-term effects of continuing adjuvant tamoxifen to 10 years versus stopping at 5 years after diagnosis of oestrogen receptor-positive breast cancer. ATLAS, a randomised trial. Lancet, **381(9869)**, 805-816.

[11] Anelli TF, Anelli A, Tran KN, et al. (1994) Tamoxifen administration is associated with a high rate of treatment-limiting symptoms in male breast cancer patients. Cancer, **74(1)**, 74-77.

[12] Pemmaraju N, Munsell MF, Hortobagyi GN, Giordano SH. (2012) Retrospective review of male breast cancer patients: analysis of tamoxifen-related side-effects. Annals of Oncology, **23**, 1471-1474.

[13] Xu S, Yang Y, Tao W, et al. (2012) Tamoxifen adherence and its relationship to mortality in 116 men with breast cancer. Breast Cancer Research and Treatment, **136(2)**, 495-502.

[14] Hertz DL, McLeod HL, Irvin Jr WJ. (2012) Tamoxifen and CYP2D6: a contradiction of data. The oncologist, **17(5)**, 620-630.

[15] Jordan VC, Collins MM, Rowsby L, Prestwich GI. (1977) A monohidroxylated metabolite of tamoxifen with potent antiestrogenic activity. Journal of Endocrinology, **75(2)**, 305-316.

[16] Murdter TE, Schroth W, Bacchus-Gerybadze L, et al. (2011) Activity levels of tamoxifen metabolites at the estrogen receptor and the impact of genetic polymorphisms of phase I and II enzymes on their concentration levels in plasma. Clinical Pharmacology and Therapeutics, **89(5)**, 708-717.

[17] Desta Z, Ward BA, Soukhova NV, Flockhart, DA. (2004) Comprehensive evaluation of tamoxifen sequential biotransformation by the human cytochrome P450 system in vitro: prominent roles for CYP3A and CYP2D6. Journal of Pharmacology and Experimental Therapeutics, **310(3)**, 1062-1075.

[18] Crewe HK, Notley LM, Wunsch RM, et al. (2002) Metabolism of tamoxifen by recombinant human cytochrome P450 enzymes: formation of the 4-hydroxy, 4'-hydroxy and N-desmethyl metabolites and isomerization of trans 4-hydroxytamoxifen. Drug Metabolism and Disposition, **30(8)**, 869-874.

[19] Zhou SF. (2009) Polymorphism of human cytochrome P450 2D6 and its clinical significance: Part I. Clinical Pharmacokinetics, **48(11)**, 689-723.

[20] Lum DW, Perel P, Hingorani AD, Holmes MV. (2013) CYP2D6 Genotype and tamoxifen response for breast cancer: a systematic review and meta-analysis. PLoS One, **8(10)**, e76648.

[21] Nowell SA, Ahn J, Rae JM, et al. (2005) Association of genetic variation in tamoxifen-metabolizing enzymes with overall survival and recurrence of disease in breast cancer patients. Breast Cancer Research and Treatment, **91(3)**, 249-258.

[22] Wegman P, Elingarami S, Carstensen J, et al. (2007) Genetic variants of CYP3A5, CYP2D6, SULT1A1, UGT2B15 and tamoxifen response in postmenopausal patients with breast cancer. Breast cancer research, **9(1)**, R7.

[23] Okishiro M, Taguchi T, Jin Kim S, et al. (2009) Genetic polymorphisms of CYP2D6*10 and CYP2C19*2,*3 are not associated with prognosis, endometrial thickness, or bone mineral density in Japanese breast cancer patients treated with adjuvant tamoxifen. Cancer **115(5)**, 952-961.

[24] Toyama T, Yamashita H, Sugiura H et al. (2009) No association between CYP2D6*10 genotype and survival of node-negative Japanese breast cancer patients receiving adjuvant tamoxifen treatment. Jpn J Clin Oncol, **39(10)**, 651-656.

[25] Stingl JC, Parmar S, Huber-Wechselberger A, et al. (2010) Impact of CYP2D6*4 genotype on progression free survival in tamoxifen breast cancer treatment. Current Medical Research and Opinion **26(11)**, 2535-2542.

[26] Lash TL, Cronin-Fenton D, Ahern TP, et al. (2011) CYP2D6 inhibition and breast cancer recurrence in a population-based study in Denmark. Journal of the National Cancer Institute **103(6)**, 489-500.

[27] Park IH, Ro J, Park S, et al. (2012) Lack of any association between functionally significant CYP2D6 polymorphisms and clinical outcomes in early breast cancer patients receiving adjuvant tamoxifen treatment. Breast Cancer Research and Treatment **131(2)**, 455-461.

[28] Rae JM, Drury S, Hayes DF, et al. (2012) CYP2D6 and UGT2B7 Genotype and Risk of Recurrence in Tamoxifen- Treated Breast Cancer Patients. Journal of the National Cancer Institute **104(6)**, 452-460.

[29] Regan MM, Leyland-Jones B, Bouzyk M, et al. (2012) CYP2D6 Genotype and Tamoxifen Response in Postmenopausal Women with Endocrine-Responsive Breast Cancer: The Breast International Group 1-98 Trial. Journal of the National Cancer Institute **104(6)**, 441-451.

[30] Wegman P, Vainikka L, Stal O et al. (2005) Genotype of metabolic enzymes and the benefit of tamoxifen in post- menopausal breast cancer patients. Breast Cancer Res **7(3)**, R284-R290.

[31] Goetz MP, Knox SK, Suman VJ, et al. (2007) The impact of cytochrome P450 2D6 metabolism in women receiving adjuvant tamoxifen. Breast Cancer Research and Treatment **101(1)**, 113-121.

[32] Schroth W, Antoniadou L, Fritz P, et al. (2007) Breast cancer treatment outcome with adjuvant tamoxifen relative to patient CYP2D6 and CYP2C19 genotypes. Journal of Clinical Oncology **25(33)**, 5187-5193.

[33] Xu Y, Sun Y, Yao L, et al. (2008) Association between CYP2D6 *10 genotype and survival of breast cancer patients receiving tamoxifen treatment. Annals of Oncology **19(8)**, 1423-1429.

[34] Newman WG, Hadfield KD, Latif A, et al. (2008) Impaired tamoxifen metabolism reduces survival in familial breast cancer patients. Clinical Cancer Research **14(18)**, 5913-5918.

[35] Bijl MJ, van Schaik RH, Lammers LA, et al. (2009) The CYP2D6*4 polymorphism affects breast cancer survival in tamoxifen users. Breast Cancer Research and Treatment **118(1)**, 125-130.

[36] Schroth W, Goetz MP, Hamann U, et al. (2009) Association between CYP2D6 polymorphisms and outcomes among women with early stage breast cancer treated with tamoxifen. Journal of the American Medical Association **302(13)**, 1429-1436.

[37] Ramón y Cajal T, Altés A, Paré L, et al. (2010) Impact of CYP2D6 polymorphisms in tamoxifen adjuvant breast cancer treatment. Breast Cancer Research and Treatment **119(1)**, 33-38.

[38] Abraham JE, Maranian MJ, Driver KE, et al. (2010) CYP2D6 gene variants: Association with breast cancer specific survival in a cohort of breast cancer patients from the United Kingdom treated with adjuvant tamoxifen. Breast Cancer Research **12(4)**, R64.

[39] Lammers LA, Mathijssen RH, Van Gelder T, et al. (2010) The impact of CYP2D6-predicted phenotype on tamoxifen treatment outcome in patients with metastatic breast cancer. British Journal of Cancer **103(6)**, 765-771.

[40] Kiyotani K, Mushiroda T, Imamura CK, et al. (2010) Significant effect of polymorphisms in CYP2D6 and ABCC2 on clinical outcomes of adjuvant tamoxifen therapy for breast cancer patients. Journal of Clinical Oncology **28(8)**, 1287-1293.

[41] Thompson AM, Johnson A, Quinlan P, et al. (2011) Comprehensive CYP2D6 genotype and adherence affect outcome in breast cancer patients treated with tamoxifen monotherapy. Breast Cancer Research and Treatment **125(1)**, 279-287.

[42] Teh LK, Mohamed NI, Salleh MZ, et al. (2011) The Risk of Recurrence in Breast Cancer Patients Treated with Tamoxifen: Polymorphisms of CYP2D6 and ABCB1. American Association of Pharmaceutical Scientists Journal **14(1)**, 52-59.

[43] Giordano SH, Valero V, Budzar AV, Hortobagyi GN. (2002) Efficacy of anastrozole in male breast cancer. American Journal of Clinical Oncology, **25(3)**, 235-237.

[44] Peppercorn J, Hamilton E, Marcom PK, et al. (2013) Pharmacogenetic testing in the face of unclear clinical efficacy. Cancer, **119(20)**, 3703-3709.

## Related Articles

Article 5 - CYP2D6*4 polymorphism: A new marker of response to hormonotherapy in male breast cancer?. (Published) 3.1.

# CYP2D6*4 polymorphism: A new marker of response to hormonotherapy in male breast cancer?

Miguel Henriques Abreu[a], Mónica Gomes[b], Francisco Menezes[c], Noémia Afonso[a], Pedro Henriques Abreu[d], Rui Medeiros[b], Deolinda Pereira[a], Carlos Lopes[e]

[a]*Department of Medical Oncology, Portuguese Institute of Oncology of Porto, Rua António Bernardino de Almeida s/n, 4200-465, Porto Portugal*
[b]*Molecular Oncology Group, Portuguese Institute of Oncology of Porto, Rua António Bernardino de Almeida s/n, 4200-465, Porto Portugal*
[c]*Department of Pathology, Portuguese Institute of Oncology of Porto, Rua António Bernardino de Almeida s/n, 4200-465, Porto Portugal*
[d]*Department of Informatics Engineering, Faculty of Sciences and Technology of Coimbra University and CISUC, Pólo II - Pinhal de Marrocos 3030-290, Coimbra Portugal*
[e]*Department of Pathology and Molecular Immunology, Biomedical Sciences Institute of Abel Salazar, University of Porto, Rua Jorge Viterbo Ferreira, 228, 4050-313 Porto, Portugal*

---

**Abstract**

**Background:** Tamoxifen remains the standard hormonotherapy for Male breast cancer patients (MBC). Previous studies, in women, tried to evaluate the impact of CYP2D6 polymorphisms in tamoxifen efficacy with conflicting results. Herein we analyze the relation between CYP2D6*4 polymorphism and survival in MBC patients.

**Patients and Methods:** Fifty-three patients, proposed to tamoxifen in adjuvant setting, were enrolled. Clinical information was collected from records and histological revision with additional immunohistochemical analysis was done to better characterize the tumours. Comprehensive CYP2D6*4 genotyping from blood or tumour tissue was performed and translated into two predictive metabolic activity groups.

**Results:** Patients included in the two CYP2D6*4 groups did not differ concerning to age, histological characteristics, and primary treatments performed. Median age at diagnosis was 63 years-old and patients were submitted at least to mastectomy and adjuvant hormonotherapy. Recurrence was observed in 7 patients (13.2%) and 13 patients (25.5%) died with a 5-year disease-free survival of 86.2%. The poorer metabolizer group had a high risk for recurrence ($p = 0.034$) and this outcome effect remains in different subgroups: in tumours larger than 2 cm ($p < 0.001$), nodal status, N0 vs N+

($p$= 0.04) and in advanced stage, stage III ($p$< 0.001). Poorer metabolizer patients had also a worse overall survival when tumours were larger than 2 cm ($p$= 0.03).
**Conclusions:** In our series, there was an association between CYP2D6*4 polymorphism and a probability of recurrence, with a consistent effect in risk groups defined by classic prognostic factors. Multicentric studies with larger samples are needed to validate these results.
*Keywords:* Male breast cancer, Tamoxifen, CYP2D6*4 polymorphism, Disease-free survival, Overall-survival

## 1 Introduction

Male Breast Cancer (MBC) is a rare disease, accounting for approximately 1% of all breast carcinomas [1, 2, 3]. Its incidence has limited the development of research studies and treatment recommendations are presently derived from the standards for female breast cancer, even though there are already many established differences between them [4, 5, 6].

Tamoxifen remains the standard adjuvant hormonotherapy in the treatment of MBC with expression of hormonal receptors (the vast majority of the cases) [7, 8], as the use of aromatase inhibitors without concomitant suppression of testicular steroidogenesis is not ineffective and did not appear to have a complete estrogen suppression compared to that observed in postmenopausal women (with only a 50% decrease in estradiol level). In addition, therapy with anastrozole raise testosterone levels by 58% with a secondary elevation of estrogen levels (loop effect) [9].

The growth inhibitory effect of tamoxifen is mediated by its active metabolites (4-hydroxytamoxifen and endoxifen), and this conversion to active formulas is mediated by polymorphic cytochrome P450 2D6 (CYP2D6) enzyme [9, 10].

The wide variety of responses and side effects of tamoxifen in different patients, questioned whether these differences result from an inter-individual genetic variability [11, 12]. Since 2003, more than 20 studies tried to define the role of CYP2D6 polymorphisms in the efficacy of tamoxifen in females with breast cancer [13, 14]. These studies lead to conflicting results and several reasons are given to justify the inconsistency, namely the designs used and the types of patients enrolled [11, 13].

In the absence of studies regarding male breast cancer, the present study was performed to determine whether CYP2D6 variation is associated with clinical outcomes in patients receiving adjuvant tamoxifen.

## 2 Material and methods

### *2.1 Study population*

The study population comprised 53 males with breast cancer treated in the same tertiary cancer Institution (Portuguese Institute of Oncology of Porto), from 1992 to 2012, with a median follow-up of 5 years (1-19 years). The patients' inclusion criteria included: male gender, histological confirmation of breast carcinoma, tumour expression of hormonal receptors, stage I, II or III and adjuvant treatment with tamoxifen 20 mg/day. From the clinical records, information about patients and tumour characteristics, treatments and survival was collected by the same Oncologist. This study was approved by the local ethics committee.

### *2.2 Study design*

The association between CYP2D6 variations and clinical outcomes was assessed based in two end points: disease-free and overall survival, defined as the time from surgery to first occurrence of a breast event and as the time from surgery to death from any cause, respectively.

### *2.3 Immunohistochemical characterization of the tumours*

The histological specimens available were reviewed by two independent breast Pathologists and additional immunohistochemical analysis was performed to better characterize the tumours. This process was done in a blinded way from clinical information. The histological grade was assessed with the Nottingham score. From the immunohistochemical analysis, estrogen and progesterone receptors were considered to be positive if $\geq$ 1% cells showed nuclear staining [15], Ki67 staining was interpreted as low or high by the use of a 14% threshold [8], p53 staining was considered to be positive if $\geq$ 5% cells showed accumulation [8] and androgen receptors were considered positive when at least 10% of nuclei were stained [8]. Cases were considered to be HER2-positive

when they were 3+ by immunohistochemically test, according to the Dako score, or FISH-amplified.

### *2.4 DNA source, genotype and definition of phenotype*

Genomic DNA was prospectively collected and extracted from whole blood in live patients (n = 38, 71.7%) and retrospectively from paraffin-embedded tumour tissues in dead patients (n = 15, 28.3%). Peripheral blood was sampled in K2EDTA plastic vacutainer tubes and after centrifugation, germ line DNA was extracted from the precipitated leucocyte cell fraction according to the commercial Kit "QIAmp DNA Blood Mini Kit" (Qiagen©), following the manufacturer's instructions. For paraffin-embedded tumour tissues, extraction of DNA was performed using a manual extraction method.

Genotyping for the CYP2D6*4 polymorphism (rs 3892097 1846G > A) was analyzed by allelic discrimination using 7300 real-time polymerase chain reaction system (real-time PCR) (Applied Biosystems, Foster City, CA, USA). Real-time PCR was performed in a 6 ul reaction mixture, containing 1 × Master Mix (Applied Bio- systems, Foster City, Ca, EUA), with 1 × probes (TaqMan assay C_27102431_D0, Applied Biosystems, Foster City, CA, EUA) and 90 ng of the DNA sample.

Quality control procedures implemented for genotyping included double sampling in about 10% of the samples to assess reliability and the use of negative controls to step-away false-positives. The genotype frequencies from tissue samples are in Hardy-Weinberg equilibrium.

To analyze the proposals, patients who have the A allele (in a homozygous or heterozygous form) were considered the poorer tamoxifen metabolizers, and were referred in the text as CYP2D6*4 A+ versus those that do not have the A allele mentioned as CYP2D6*4 A−. Pharmacogenetic analysis was performed in a blinded way from clinical information.

### *2.5 Statistical Analysis*

To reduce the effect of time, time-dependent variables were adjusted or grouped in main categories (stage was adjusted to $7^{th}$ edition of AJCC Staging, and treatment was grouped in categories). $X^2$ test was used to compare categorical variables and t-student test was used to compare parametric continuous variables.

Survival curves were calculated by the Kaplan-Meier method and compared using the log-rank test. Statistical significance was defined as $p < 0.05$. SPSS for Windows version 19.0 (SPSS Inc., Chicago, IL, USA) was used for statistical calculations.

## 3 Results

All patients were Caucasians, with a median age at diagnosis of 63 years-old (range from 35 to 95 years-old).

The majority (51.0%), had at least one risk factor for cardiovascular diseases, with 34 patients (64.2%) presenting excessive weight or obesity (body-mass index superior to 25 or 30 kg/m2 respectively). In 46 patients (86.8%) it was possible to evaluate the family history. Twenty-eight patients (52.8%) had a positive history of cancer and 37.5% of these patients had a first degree member (women) with breast cancer. The genetic test was performed in 23 patients (43.4%) and 3 of them (13.0%) presented mutation of BRCA genes (BRCA2 in 67%). During the follow-up, 8 patients (15.1%) developed a second primary tumour (prostate cancer in 75% of the cases).

All patients were evaluated for CYP2D6*4 mutations and 36 patients (67.9%) were considered CYP2D6*4 A$-$. The principle characteristics of patients and tumours classically related to recurrence and/or death were well balanced between groups defined by CYP2D6*4 genotypes (see Table 1).

All patients underwent, at least, mastectomy and adjuvant hormonotherapy with tamoxifen. There were no major problems registered concerning therapeutic adherence. Chronic patients' medication did not include paroxetine or fluoxetine. The only variable related to adjuvant treatment choice was stage ($p = 0.001$).

### *3.1 Survival Analysis*

Recurrence was observed in seven patients (13.2%), with a 5- year disease-free survival of 86.2%. In half of the cases (50%), there was only visceral metastasis (lung, 33.3% or liver, 16.7%), and in 16.7% bone was the only site affected. In metastatic setting, five patients (71.4%), received palliative treatment, with chemotherapy in 80%.

CYP2D6*4 genotypes were associated with recurrence ($p = 0.03$), with the CYP2D6*4 A$-$ group having an estimate disease-free survival of 17.9 years (95% CI: 16.4-19.3

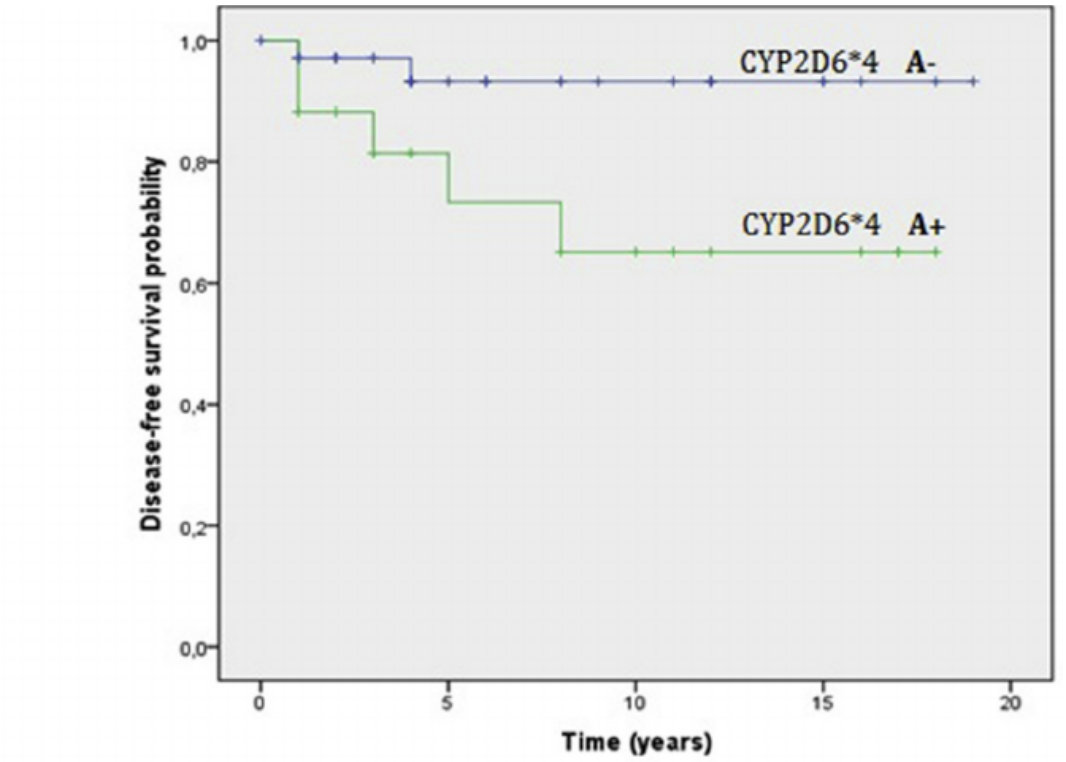
Table 1: Patients and tumours characteristics by CYP2D6*4 subgroups.

| **Characteristics** | **Overall (N=53)** | **CYP2D6*4 A− (N=36)** | **CYP2D6*4 A+ (N=17)** | **P value** |
|---|---|---|---|---|
| Age at diagnosis, median (range), years, n = 53 | 63 (35-95) | 63.5 (44-85) | 61 (35-95) | 0.25 |
| Follow-up, median (range), months, n = 53 | 57 (9-228) | 54.5 (9-228) | 99 (20-225) | 0.12 |
| Tumor laterality, left, No. (%), n = 53 | 28 (52.8) | 20 (55.6) | 8 (47.1) | 0.56 |
| Tumour size, cm, No. (%), n = 53 | | | | |
| ≤2 | 27 (50.9) | 15 (44.1) | 12 (70.6) | 0.07 |
| >2 | 24 (45.3) | 19 (55.9) | 5 (29.4) | |
| Nodal status, No. (%), n = 53 | | | | |
| N0 | 22 (41.5) | 14 (38.9) | 8 (47.1) | 0.57 |
| N+ | 31 (58.5) | 22 (61.1) | 9 (52.9) | |
| Histological grade, No. (%), n = 51 | | | | |
| 1 | 11 (21.6) | 7 (20.6) | 4 (23.5) | 0.42 |
| 2 | 25 (49) | 15 (44.1) | 10 (58.8) | |
| 3 | 15 (29.4) | 12 (35.3) | 3 (17.6) | |
| Stage, No. (%), n = 53 | | | | |
| I | 16 (30.2) | 11 (30.6) | 5 (29.4) | 0.98 |
| II | 16 (30.2) | 11 (30.6) | 5 (29.4) | |
| III | 21 (39.6) | 14 (38.9) | 7 (41.2) | |
| Hormonal receptor positive (ER and/or PR), No. (%), n = 53 | 53 (100) | 36 (100) | 17 (100) | 1 |
| HER2, No. (%), n = 47 | | | | |
| Negative | 43 (91.5) | 29 (87.9) | 14 (100) | (a) |
| Positive | 4 (8.5) | 4 (12.1) | 0 (0) | |
| AR, No. (%), n = 41 | | | | |
| Negative | 7 (17.1) | 5 (16.1) | 2 (20) | 0.77 |
| Positive | 34 (82.9) | 26 (83.9) | 8 (80) | |
| Ki67, No. (%), n = 41 | | | | |
| ≤14% | 25 (61) | 19 (61.3) | 6 (60) | 0.94 |
| >14% | 16 (39) | 12 (38.7) | 4 (40) | |
| p53, No. (%), n = 42 | | | | |
| Low | 32 (76.2) | 23 (74.2) | 9 (81.8) | 0.61 |
| High | 10 (23.8) | 8 (28.8) | 2 (18.2) | |
| Primary treatment, No. (%), n = 53 | | | | |
| Surgery + hormonotherapy | 18 (34) | 11 (30.6) | 7 (41.2) | 0.62 |
| Surgery + hormonotherapy + radiotherapy | 15 (28.3) | 11 (30.6) | 4 (23.5) | |
| Surgery + hormonotherapy + radiotherapy + chemotherapy ± trastuzumab (b) | 20 (37.7) | 14 (38.9) | 6 (35.3) | |

ER, estrogen receptor; PR, progesterone receptor; AR, androgen receptor.

(a) Not possible to compare (there were no tumors in CYP2D6*4 A+ that express HER2).

(b) The use of trastuzumab was done according to HER2 expression.




| No, at risk | | | | | |
|---|---|---|---|---|---|
| CYP2D6*4 A- | 36 | 17 | 10 | 3 | 0 |
| CYP2D6*4 A+ | 17 | 9 | 6 | 2 | 0 |

Figure 1: Disease-free survival by CYP2D6*4 genotype.

years) in contrast with the 13.1 years (95% CI: 9.5-16.6 years) in the CYP2D6*4 A+ group, Figure 1.

Thirteen patients (24.5%) died with a 5-year overall survival of 78.5%. In general, CYP2D6*4 genotypes were not related to overall survival ($p = 0.95$), with an estimate survival for CYP2D6*4 A− group of 13.7 years (95% CI: 10.6-16.8 years) and for CYP2D6*4 A+ group of 13.2 years (95% CI: 9.8-16.6 years).

The survival effect of CYP2D6*4 genotypes was also evaluated according to different groups, in classical predictive factors of recurrence and death: age, tumour size, nodal status, histological grade, stage, expression of HER2, androgen recep-

tors, Ki67, p53 and primary treatments. The groups with sufficient cases that allow survival comparisons, between CYP2D6*4 genotypes, can be seen in Table 2. In recurrence, CYP2D6*4 genotypes maintain their discriminatory capacity in tumours larger than 2 cm (mean disease-free survival for CYP2D6*4 A$-$ :17.9 years vs 3.6 years for CYP2D6*4 A$+$, $p < 0.001$), nodal status, N0 and N$+$ (mean disease-free survival for N0 tumours CYP2D6*4 A$-$: 17.7 years vs 11.7 years for CYP2D6*4 A$+$, in N$+$ tumours 14.3 years vs 13.1 years, $p = 0.04$) and in advanced stage, stage III (mean disease-free survival for CYP2D6*4 A$-$ :17.8 years vs 9.1 years for CYP2D6*4 A$+$, $p < 0.001$).

In tumours with more than 2 cm, CYP2D6*4 genotypes predict a different overall survival (mean of 13.9 years for CYP2D6*4 A$-$ vs 4.6 years for CYP2D6*4 A$+$, $p = 0.02$).

## 4 Discussion

For our knowledge this is the first study that addresses the role of CYP2D6*4 polymorphism in survival outcome of receptor positive MBC patients treated with tamoxifen in adjuvant setting. In our series, the poorer metabolizer patients, defined by the presence of A allelle (CYP2D6*4 A$+$ group) had an early recurrence and CYP2D6*4 genotypes remain their discriminatory capacity for recurrence in classical risk groups for metastization: in tumours with more than 2 cm, in N$+$ disease (versus N0) and in advanced stage (stage III). Although we did not find a global impact in overall survival, the association between CYP2D6*4 genotypes and death in patients with tumours larger than 2 cm, could guide future investigation in this particular group of worse prognosis.

The rarity of MBC hinders the development of randomized controlled trials. However and particularly in the controversy between CYP2D6 genotype and tamoxifen response in postmenopausal women, there are already many authors that agree that purely observational studies may have equally good standardization of the population and the outcomes of interest as randomized controlled trials [16]. Still in women domain, many possible factors have been proposed to justify the variations among the studies: many DNA sources (tumour, blood), different phenotypes evaluated (efficacy, toxicity, pharmacokinetics), several studies designs (retrospective, prospective), with different populations (adjuvant and palliative settings), different tumours (with and with-

Table 2: Stratification of patients and tumours' characteristics by CYP2D6*4 genotypes, according to survival outcomes.

| Characteristics | Disease Free Survival | | | Overall Survival | | |
|---|---|---|---|---|---|---|
| | **CYP2D6*4 A− mean, years (CI 95%)** | **CYP2D6*4 A+ mean, years (CI 95%)** | **P value** | **CYP2D6*4 A− mean, years (CI 95%)** | **CYP2D6*4 A+ mean, years (CI 95%)** | **P value** |
| Age at diagnosis, years | | | | | | |
| ≤65 | 16.8 (13.9-19.6) | 13.2 (9.4-17.1) | 0.35 | 16.5 (13.3-19.7) | 13.5 (9.9-17.0) | 0.61 |
| >65 | - | - | - | 7.1 (4.9-9.4) | 8.5 (4.2-12.7) | |
| Tumor size, cm | | | | | | |
| ≤2 | 14.1 (12.2-15.8) | 16.7 (14.4-19.0) | 0.87 | 8.1 (5.8-10.4) | 16.8 (14.7-18.9) | **0.08** |
| >2 | 17.9 (15.9-19.9) | 3.6 (1.2-5.9) | <**0.001** | 13.9 (10.0-17.9) | 4.6 (2.2-6.9) | **0.03** |
| Nodal status | | | | | | |
| N0 | 17.7 (15.2-20.1) | 11.71 (6.7-16.6) | **0.04** | 14.8 (10.6-18.9) | 11.6 (6.6-16.6) | 0.91 |
| N+ | 14.3 (13.1-15.5) | 13.1 (8.6-17.6) | | 8.9 (6.8-10.9) | 13.3 (9.1-17.6) | 0.97 |
| Grade | | | | | | |
| 2 | - | - | - | 9.7 (7.4-11.9) | 12.7 (7.9-17.6) | 0.76 |
| Stage | | | | | | |
| III | 17.8 (15.6-20.0) | 9.1 (3.3-14.9) | **0.01** | 12.6 (7.9-17.3) | 8.8 (4.0-13.6) | 0.26 |
| HER2 | | | | | | |
| Negative | 17.6 (15.7-19.4) | 13.1 (9.1-17.0) | 0.11 | 13.9 (10.3-17.4) | 12.8 (8.8-16.8) | 0.87 |
| AR | | | | | | |
| Positive | - | - | - | 15.0 (11.5-18.5) | 9.25 (4.0-14.4) | 0.09 |
| Ki67 | | | | | | |
| ≤14% | 14.3 (12.3-16.4) | 12.4 (6.2-18.6) | 0.25 | 9.6 (5.6-13.6) | 12.4 (6.2-18.6) | 0.74 |
| p53 | | | | | | |
| Low | 17.2 (14.9-19.5) | 11.6 (6.0-17.3) | 0.09 | 13.2 (9.46-17.0) | 11.6 (5.9-17.2) | 0.61 |
| Primary treatment | | | | | | |
| Surgery + hormonotherapy | - | - | - | 12.3 (7.8-16.7) | 10.4 (7.5-13.2) | 0.80 |
| Surgery + hormonotherapy + radiotherapy | 15.6 (11.5-19.8) | 13.5 (7.5-19.4) | 0.84 | 9.6 (4.7-14.5) | 13.0 (8.8-17.1) | |
| Surgery + hormonotherapy + radiotherapy ± trastuzumab (a) | - | - | - | 11.3 (10.0-12.5) | 12.6(6.6-18.5) | |

AR, androgen receptor.

The *P* values that are in bold are the statistical significant ones.

– Not possible to compare (few cases in the groups).

(a) The use of trastuzumab was done according to HER2 expression.

out expression of hormonal receptors), and different concomitant medications (other anticancer medications and CYP2D6 inhibitors) [9]. This heterogeneity weakens the comparisons between studies as demonstrated in the published revisions and meta-analysis [13, 14, 17] and compromises future works.

In relation to our population study, all patients were Caucasian, all tumours express hormonal receptors and all patients were treated with the same intention (adjuvant) with the same drug (tamoxifen). As previously described, CYP2D6 polymorphisms have geographic and ethnic variations (Caucasian vs Asiatic), and some studies (71%, according to Lum et al. [14]) enrolled patients with a mixture of tumours with and without expression of hormonal receptors, who do not benefit from tamoxifen, and therefore would not have differential benefit based on CYP2D6 activity. Other possible confounding factors consisted in the adherence to tamoxifen, the used dose, and the administration of CYP2D6 inhibitors that could influence the endoxifen concentrations and potentially tamoxifen efficacy [18, 19, 20]. Contrarily to Bijl et al. [21] and Lammers LA et al. [22], the used dosage in our series was the same (20 mg/day) and any patient reported concomitant administration of paroxetine or fluoxetine. In compliance domain, the published studies [23, 24] show that the rates of tamoxifen discontinuation in males are near 20.3%, fewer than in women (range from 30 to 50%) [25] and are secondary to side effects, mainly weight gain and sexual dysfunctional/loss of libido. In our series there was no major compliance problems registered. The free delivery of hormonotherapy in our Institution could have contributed to this fact.

Another major limitation of the new studies [26, 27, 28] is the use of tumour DNA to determine germline genotypes. As the CYP2D6 phenotype results from germline predisposition, it must be assessed from DNA that is derived from normal cells, because tumour cores contain insufficient numbers of normal epithelial cells for detection [29]. The DNA analyzed in our series was extracted from blood in the majority of the cases (71.7%) and the genotype frequencies from the tumour DNA are consistent with those expected by Hardy-Weinberg equilibrium, which are compatible with genotyping quality [30]. We choose the CYP2D6*4 to genotype, to be the most prevalent in the Caucasian race, the more frequent analyzed in the adjuvant setting and to have also a pharmacogenetic model developed based on it [13, 14, 17, 31, 32]. The possible limitation

of this approach, namely the misclassification phenomenon [33, 34], was not probably relevant attending the rarity of MBC, with less impact in the evaluated endpoints.

With the knowledge from the recent studies, we could state that MBC is a different entity from female breast cancer [5], with a more advanced age at diagnosis (median of 63 years-old in our report) and with some distinct risk factors proposed, like imbalance between androgen and estrogen and constitutional variables (63.6% of patients had excessive weight or obesity) [5, 35, 36].

The tumour size ($>$2 cm) and lymph nodes involvement are documented prognostic factors, and others like age and stage were already proposed with some controversy results [37, 38]. This could justify different approaches even in metastatic setting, where there were only few publications described the role of aromatase inhibitors in MBC patients [39, 40, 41], remaining doubts about the best hormonotherapy sequence in tamoxifen-refractory patients.

Having into account the differences between genders, the absence of studies achieving the role of CYP2D6 polymorphisms in MBC patients and due to the variety of methodologies used and populations enrolled in women studies, comparisons between our study and the published data must be done with precaution. When looking for studies that have been performed in Caucasian women with sporadic breast cancer, treated with tamoxifen in adjuvant setting and with a genotyping comprehensive analysis that include the CYP2D6*4, we find controversial results. The two studies published by Wegman et al. in 2005 [42] and 2007 [43] did not find a relation between CYP2D6*4 polymorphism and disease-free survival. The first study included pre and post-menopausal women that presented only large tumours with lymph node involvement and all of them were treated with 40 mg of tamoxifen. In the second study there are no patients in stage I (only stage II and III) and the subgroup treated before 1994 also received 40 mg of tamoxifen. These could be possible reasons for their negative findings. In another direction Schroth et al. [10], described a relation between CYP2D6*4 variants and recurrence rates in patients treated with adjuvant tamoxifen and no chemotherapy. However this finding was not replicated in other studies that enrolled only Caucasian patients [44].

Contrarily to our findings, Bijl et al. [22], stated a relation between CYP2D6*4 geno-

types and breast cancer mortality in a population-based cohort study. However in this study there was no reference to the type of adjuvant treatments performed by the patients beyond hormonotherapy and we do not know the immunohistochemical profile of the tumours related to the expression of hormonal receptors.

In conclusion, our study demonstrates a possible role in recurrence prediction of CYP2D6*4 genotypes in MBC patients which could mean that some of them benefit less from adjuvant tamoxifen, starting the debate about the need of changing the systemic treatment in some subgroups. The relation between these genotypes and other recurrence and prognostic factors could signify that a longer follow-up and the enrollment of a large number of patients in future studies are necessary to provide definitive evidence for the value of this biomarker in clinical management.

**Acknowledgments**

The authors want to thank Dr. Fatima Cardoso and Dr. Rosário Couto for the review of this paper and to Dr. Paula Lopes for the preparation of samples for immunohistochemical analysis.

**References**

[1] Fentiman IS,Fourquet A, Hortobagyi CN. (2006) Male breast cancer. Lancet **367(9510)**, 595-604.

[2] Giordano SH. (2005) A review of the diagnosis and management of male breast cancer. The Oncologist **10(7)**, 471-479.

[3] Jemal A, Murray T, Ward E, Samuels A, Tiwari RC, Ghafoor A, et al. (2005) Cancer statistics. CA: A Cancer Journal for Clinicians, **55(1)**, 10-30.

[4] Giordano SH, Cohen DS, Buzdar AU, et al. (2004) Breast carcinoma in men: a population-based study. Cancer **101(1)**, 51-57.

[5] Nahleh ZA, Srikantiah R, Safa M, et al. (2007) Male breast cancer in the veterans affairs population: a comparative analysis. Cancer **109(8)**, 1471-1477.

[6] Anderson WF, Jatoi I, Tse J, Rosenberg PS. (2010) Male breast cancer: a population based comparison with female breast cancer. Journal of Clinical Oncology **28(2)**, 232-239.

[7] Muir D, Kanthan R, Kanthan SC. (2003) Male versus female breast cancers. A population-based comparative immunohistochemical analysis. Archives of Pathology & Laboratory Medicine **127(1)**, 36-41.

[8] Kornegoor R, Verschuur-Maes A, Buerger H, Hogenes MC, de Bruin PC, Oudejans JJ, et al. (2012) Immunophenotyping of male breast cancer. Histopathology **61(6)**, 1145-1155.

[9] Giordano SH, Valero V, Budzar AV, Hortobagyi GN. (2002) Efficacy of anastrozole in male breast cancer. American Journal of Clinical Oncology **25(3)**, 235-237.

[10] Schroth W, Goetz MP, Hamann U, Fasching PA, Schmidt M, Winter S, et al. (2009) Association between CYP2D6 polymorphism and outcomes among women with early stage breast cancer treated with tamoxifen. Journal of the American Medical Association, **302(13)**, 1429-1436.

[11] Antunes MV, Linden R, Santos TV, Wallemacq P, Haufroid V, Classen JF, et al. (2012) Endoxifen levels and its association with CYP2D6 genotype and phenotype: evaluation of a southern Brazilian population under tamoxifen pharmacology. Therapeutic Drug Monitoring **34(4)**, 422-431.

[12] Fleeman N, Martin Saborido C, Payne K, Boland A, Dickson R, Dundar Y, et al. (2011) The clinical effectiveness and cost-effective of genotyping for CYP2D6 for the management of women with breast cancer treated with tamoxifen: a systematic review. Health Technology Assessment **15(33)**, 1-102.

[13] Zhou SF. (2009) Polymorphism of human cytochrome P450 2D6 and its clinical significance: Part I. Clinical Pharmacokinetics **48(11)**, 689-723.

[14] Hertz DL, McLeod HL, Irvin Jr WJ. (2012) Tamoxifen and CYP2D6: a contradiction of data. The Oncologist **17(5)**, 620-30.

[15] Lum DWK, Perel P, Hingorani AD, Holmes MV. (2013) CYP2D6 genotype and tamoxifen response for breast cancer:a systematic review and meta-analysis. PLoS One **8(10)**, e76648.

[16] Harvey JM, Clark GM, Osborne CK, Alfred DC. (1999) Estrogen Receptor Status by immunohistochemistry is superior to the ligand-binding assay for predicting response to adjuvant endocrine therapy in breast cancer. Journal of Clinical Oncology **17(5)**, 1474-1481.

[17] Pharoah PDP, Abraham J, Caldas C. (2012) Re: CYP2D6 genotype and tamoxifen-responsive breast cancer: the Breast International Group 1-98 trial and Re: CYP2D6 and UGT2B7 genotype and risk of recurrence in tamoxifen-treated breast cancer patients. Journal of the National Cancer Institute **104(16)**, 1263-1264.

[18] Singh MS, Francis PA, Michael M. (2011) Tamoxifen, cytochrome P450 genes and breast cancer clinical outcomes. Breast **20(2)**, 111-118.

[19] Stearns V, Johnson MD, Rae JM, Morocho A, Novielli A, Bhargava P, et al. (2003) Active tamoxifen metabolite plasma concentrations after coadministration of tamoxifen and the selective serotonin reuptake inhibitor paroxetine. Journal of the National Cancer Institute **95(23)**, 1758-1764.

[20] Kelly CM, Juurlink DN, Gomes T, Duong-Hua M, Pritchard KI, Austin PC, et al. (2010) Selective serotonin reuptake inhibitors and breast cancer mortality in women receiving tamoxifen: a population based cohort study. BMJ **340**, c693.

[21] Irvin Jr WJ, Walko CM, Weck KE, Ibrahim JG, Chiu WK, Dees EC, et al. (2011) Genotype-guided tamoxifen dosing increases active metabolite exposure in women with reduced CYP2D6 metabolism: a multicenter study. Journal of Clinical Oncology **29(24)**, 3232-3239.

[22] Bijl MJ, van Schaik RH, Lammers LA, Hofman A, Vulto AG, van Gelder T, et al. (2009) The CYP2D6*4 polymorphism affects breast cancer survival in tamoxifen users. Breast Cancer Research and Treatment **118(1)**, 125-130.

[23] Lammers LA, Mathijssen RG, van Gelder T, Bijl MJ, de Graan AJ, Seynaeve C, et al. (2010) The impact of CYP2D6-predicted phenotype on tamoxifen treatment outcome in patients with metastatic breast cancer. British Journal of Cancer **103(6)**, 765-71.

[24] Anelli TF, Anelli A, Tran KN, Lebwohl DE, Borgen PI. (1994) Tamoxifen administration is associated with a high rate of treatment-limiting symptoms in male breast cancer patients. Cancer **74(1)**, 74-77.

[25] Pemmaraju N, Munsell MF, Hortobagyi GN, Giordano SH. (2012) Retrospective analysis of male breast cancer patients: analysis of tamoxifen-related side-effects. Annals of Oncology **23(6)**, 1471-1474.

[26] Hershman DL, Kushi LH, Shao T, Buono D, Kershenbaum A, Tsai WY, et al. (2010) Early discontinuation and non-adherence to adjuvant hormonal therapy in a cohort of 8,769 early-stage breast cancer patients. Journal of Clinical Oncology **28(27)**, 4120-4128.

[27] Regan MM, Leyland-Jones B, Bouzk M, Pagani O, Tang W, Kammler R, et al. (2012) CYP2D6 genotype and tamoxifen and tamoxifen response in postmenopausal women with endocrine-responsive breast cancer: the Breast International Group 1e98 Trial. Journal of National Cancer Institute **104(6)**, 441-51.

[28] Rae JM, Drury S, Hayes DF, Steams V, Thibert JN, Haynes BP, et al. (2012) CYP2D6 and UGT2B7 genotype and risk of recurrence in tamoxifen-treated breast cancer patients. Journal of the National Cancer Institute **104(6)**, 452-460.

[29] Lash TL, Cronin-Fenton D, Ahern TP, Rosenberg CL, Lunetta KL, Silliman RA, et al. (2011) CYP2D6 inhibition and breast cancer recurrence in a population-based study in Denmark. Journal of the National Cancer Institute **103(6)**, 489-500.

[30] Brauch H, Schroth W. (2013) Tamoxifen use in postmenopausal breast cancer: CYP2D6 Matters. Journal of Clinical Oncology **31(2)**,176-180.

[31] Gomes I, Collins A, Lonjou C, Thomas NS, Wilkinson J, Watson M, et al. (1999) Hardy-Weinberg quality control. Annals of Human Genetics **63(6)**, 535-538.

[32] Dieudonee AS, Lambrechts D, Claes B, Vandorpe T, Wildiers H, Timmerman D, et al. (2009) Prevalent breast cancer patients with a homozygous mutant status for CYP2D6*4: response and biomarkers in tamoxifen users. Breast Cancer Research and Treatment **118(3)**, 531-538.

[33] Punglia RS, Burstein HJ, Winer EP, Weeks J. (2008) Pharmacogenomic variation of CYP2D6 and the choice of optimal adjuvant endocrine therapy for postmenopausal breast cancer: a modeling analysis. Journal of the National Cancer Institute **100(9)**, 642-648.

[34] Schroth W, Hamann U, Fasching PA, Dauser S, Winter S, Eichelbaum M, et al. (2010) CYP2D6 polymorphism as predictors of outcome in breast cancer patients treated with tamoxifen: expanded polymorphism coverage improves risk stratification. Clinical Cancer Research **16(17)**, 4468-4477.

[35] Thompson A, Johnson A, Quinlan P, Hillman G, Fontecha M, Bray SE, et al. (2011) Comprehensive CYP2D6 genotype and adherence affect outcome in breast cancer patients treated with tamoxifen monotherapy. Breast Cancer Research and Treatment **125(1)**, 279-287.

[36] Sasco AJ, Lowenfels AB, Pasker-de-Jong P. (1993) Review article: epidemiology of male breast cancer. A meta- analysis of published case-control studies and discussion of selected etiological factors. International Journal of Cancer **53(4)**, 538-549.

[37] Sorensen HT, Friis S, Olsen JH, Thulstrup AM, Mellemkjaer L, Linet M, et al. (1998) Risk of breast cancer in men with liver cirrhosis. American Journal of Gastroenterology **93(2)**, 231-233.

[38] Herman K, Lobaziewicz W, Skotmicki P, Fortuna J, Kusy T, Lesniak T. (2000) Male breast cancer. Does the prognosis differ compared to female? Neoplasma **47(3)**, 191-195.

[39] Zagouri F, Sergentanis TN, Koutoulidis V, Sparber C, Steger GG, Dubsky P, et al. (2013) Aromatase inhibitors with or without gonadotropin-releasing hormone analogue in metastatic male breast cancer: a case series. British Journal of Cancer **108(11)**, 2259-2263.

[40] Doyen J, Italiano A, Largillier R, Ferrero JM, Fontana X, Thyss A. (2010) Aromatase inhibition in male breast cancer patients: biological and clinical implications. Annals of Oncology **21(6)**, 1243-1245.

[41] Zagouri F, Sergentanis TN, Chrysikos D, Dimopoulos MA, Psaltopoulou T. (2015) Fulvestrant and male breast cancer: a pooled analysis. Breast Cancer Research and Treatment **149(1)**, 269-275.

[42] Wegman P, Vainikka L, Stal O, Nordenskjold B, Skoog L, Rutqvist LE, et al. (2005) Genotype of metabolic enzymes and the benefit of tamoxifen in post-menopausal breast cancer patients. Breast Cancer Research **7(3)**, R284-R290.

[43] Wegman P, Elingarami S, Carstensen J, Stal O, Nordenskjold B, Wingren S. (2007) Genetic variants of CYP3A5, CYP2D6, SULT1A1, UGT2B15 and tamoxifen response in postmenopausal patients with breast cancer. Breast Cancer Research **9(1)**, R7.

[44] Stingl JC, Parmar S, Huber-Wechselberger A, Kainz A, Renner W, Seeringer A, et al. (2010) Impact of CYP2D6*4 genotype on progression free survival in tamoxifen breast cancer treatment. Current Medical Research Opinion **26(11)**, 2535-2542.

# Chapter 4

# Final Remarks

Due to its rarity and different biology, MBC patients are seldom included in BC trials from which standard treatment recommendations are made, and because of that, are still managed as FBC patients. Data related to MBC come from retrospective analysis that cover long time periods, different regions, distinct therapeutic approaches and a large range of missing variables from cancer registries that could definitively be sources of potential confounders and must be noticed when analysing the results. However, retrospective studies have much relevance in rare diseases, as they may equally standardize the population and the outcomes of interest as randomized controlled trials namely when they are performed in a high-volume centers, where bias related for example with treatment standards could be diluted.

The present work addressed two main important issues when defining BC patients' approach: prognostic and predictive factors to response to adjuvant endocrine therapy (tamoxifen). In prognostic domain, there are already many doubts about the factors that influence MBC patients' survival and if these patients present or not poorer survival than female, even when some important factors are controlled. The developed research work tried to find specifically subgroups that translate MBC patients' behaviour and also addressed an important issue never specifically studied in these patients: recurrence. Worst prognosis was independently associated with tumours larger than 2 cm that did not express ER and presented distant metastasis at diagnosis. N+ disease showed borderline relation with prognosis. FBC subgroups, based on three or four IHC markers and the most recent one with new IHC cut-offs for PR, had prognostic implications also in MBC patients, but lost their relevance when analysed only luminal- like tumours (the majority of them), which means that FBC subgroups did not provide the same information in male, as the others types of BC (triple negative and HER2-enriched) are rarely documented. This was the first time that the new classification of FBC subgroups (2014 classification) was tested in MBC. Based on a routine IHC panel of six markers, cluster analysis identified two new MBC subgroups; one related to good prognosis (patients that express ER and PR without expression of AR, HER2

and ki67 and p53 low) and other related to bad one (patients without expression of PR). The high frequency of luminal-like tumours and the preferential use of tamoxifen in these patients, make it important to find clinical markers to response to this therapy. According to our findings, MBC patients treated with adjuvant tamoxifen that were considered poor metabolizers according to CYP2D6*4 polymorphism, had higher risk for relapse, and this effect was independent for classic prognostic factors (size>2cm, N+ disease and advanced stage). In patients with tumour larger than 2 cm, CYP2D6*4 polymorphisms confer also a negative impact on prognosis.

Recurrence was studied to clarify the doubts regarding survival differences between female and male BC patients based on two well-balanced cohorts. Male presented poorer prognosis than female and recurred more frequently to lung, although the local of relapse was not related to survival differences between groups. MBC patients were proposed more often to symptomatic treatment at first recurrence, which probably justify the distinct outcomes, as it was also proved in this work that palliative systemic treatment had favourable impact in survival in male patients like in female.

In summary, with this work clinicians must look with major concerns for MBC patients that present with tumours with more than 2cm, or have lymph nodes involvement or do not express ER or PR or present with distant metastasis at diagnosis (stage IV). Poor tamoxifen metabolizers based on CYP2D6*4 polymorphism seem to have higher risk for relapse and probably need a more aggressive adjuvant approach. In recurrence time, palliative systemic treatment had favourable implications in these patients survival.

## Future Perspectives

With this work, many interesting aspects about MBC disease were clarified defining new research directions that started with replication of these findings in large and, ideally, prospective series. A multidisciplinary meeting on MBC topic [1] stated that for rare diseases, the key to understanding them could be the pooling of data from a wide range of sources and so international consortia are essential to moving research forward. Breast International Group (BIG) and North American Breast Cancer Group (NABCG) have joined efforts to develop an International MBC Program to pool epidemiologic data, clinical information and tumour specimens. This program has two essential parts that will determine, at the end of the process, the feasibility of a randomized clinical trial in these patients. The first one, also called retrospective part, has the objective of performing a meta-analysis based on clinical data retrospectively collected and a central pathological review of tumour specimens from MBC patients diagnosed and treated at participating institutions

[1] Korde LA, Zujewski JA, Kamin L, et al. (2010) Multidisciplinary meeting on male breast cancer. Summary and research recommendations. Journal of Clinical Oncology, 28(12), 2114- 2122.

over the last 20 years. The second one, or prospective part, has the coordination of European Organization for Research and Treatment of Cancer (EORTC) and will registry all MBC treated for a period of 2 years (remaining the follow-up ideally for a period of 10 years) and is linked to biologic material collection (paraffin-embedded and frozen tumour samples and blood/serum) and evaluation of quality of life (QOL) in these patients. This study is already registered by the names EORTC-10085, BIG 2-07, TBCRC 029, and there are actually two centers in Portugal that are recruiting. One of them is IPOPFG, and I was granted with the Principle Investigator position in this study, based on my interest on the topic. Nowadays we have recruited 7 patients that constituted 5% of the series already enrolled. This could be also the opportunity to confirm some of our findings as the collected biological samples could be used for additional tests.

With the present findings, the principal research directions with a larger series could be:

- Evaluate specifically the group of PR negative patients and compare it with the others groups;
- Reanalyse recurrence, collecting patients' comorbidities information and causes of death to compare BC specific survival between genders;
- Define relapse model prediction for MBC patients;
- Test gene dose effect based on the three MBC metabolizers groups defined by CYP2D6*4;
- Perform comprehensive genotyping of CYP2D6, evaluating the effect of multiple CYP2D6 alleles in MBC outcome.