

# Chapter 27

## Exploratory Data Analysis on Breast Cancer Prognosis

**Mohammad Mehdi Owrang O.**  
American University, USA

**Yasmine M. Kanaan**  
Howard University, USA

**Robert L. Copeland Jr.**  
Howard University, USA

**Melvin Gaskins**  
Howard University Hospital, USA

**Robert L. DeWitty Jr.**  
Providence Hospital, USA

### ABSTRACT

*Breast cancer prognosis is a vital element of providing effective treatment for breast cancer patients. Breast cancer prediction survivability has mainly been studied based on pathological factors such as tumor size, tumor grade, number of positive lymph nodes, and hormone receptors among others. This chapter looks at the significance of the non-clinical prognostic factors of age, ethnicity, and marital status in finding the prognosis for breast cancer patients. The National Cancer Institute's SEER data and the Howard University Cancer Center Tumor Registry data are analyzed. Prognostic tool NPI (Nottingham Prognostic Index) and survival analysis tools of Cox proportional hazards and Kaplan-Meier survival curve are used in analyzing the experiments. The results suggest that age, ethnicity, and marital status have some influence on the survivability rate of breast cancer patients.*

### INTRODUCTION

Breast cancer is the most common female cancer in the US, the second most common cause of cancer death in women ("American Cancer Society", 2016), and the main cause of death in women ages 40 to 59 (Siegel et al., 2012).

In 2016, it is estimated that 249,260 new cases of breast cancer will be diagnosed and estimated 40,890 breast cancer deaths; and an invasive breast cancer will be diagnosed in about 246,660 women and 2,600

DOI: 10.4018/978-1-5225-7598-6.ch027

men. An additional 61,000 new cases of in situ breast cancer will be diagnosed in women (“American Cancer Society”, 2016). The lifetime probability of developing breast cancer is one in six overall (one in eight for invasive disease) (“American Cancer Society”, 2013; Siegel et al., 2012).

Worldwide, breast cancer is the most frequently diagnosed cancer among women in 140 of 184 countries, according to the World Cancer Research Fund International (“BCRF”, n.d.). Since 2008, breast cancer incidence has increased by more than 20 percent and mortality has increased by 14 percent (“BCRF”, n.d.). Nearly 1.7 million new breast cancer cases were diagnosed in the last report of 2012.

Medical prognosis is an evaluative component of medicine that encompasses the science of estimating the complication and recurrence of disease and predictive survival of patients (Ohno-Machado, 2001). Medical prognosis plays an increasing role in health care outcome. Reliable prognostic models that are based on survival analysis statistics and techniques have been applied to a variety of domains with varying degrees of success (i.e., APACHE IV (Zimmerman, 2006)).

Breast cancer prognosis is a vital element of providing effective treatment for patients. It has become increasingly important that clinicians are provided with accurate prognostic information on which to base therapeutic decision as the range of options for the treatment of patients with breast cancer widens. A large number of factors, including tumor grade, tumor size, and lymph node status including other aspects may influence or correlate with prognosis for breast cancer patients.

Breast cancer prediction survivability has mainly been studied through clinical approaches, based on pathological factors such as tumor grade, tumor size and number of the positive lymph nodes, estrogen (ER), progesterone (PR), and human epidermal growth factor receptor 2 (Her2) receptors, etc. Most studies are carried out in an effort to find factors that clarify the large unexplained variation in prognosis of the breast cancer patients. There is still uncertainty about the importance of most prognostic factors. There are other non-clinical factors such as age, ethnicity, obesity, and marital status that may have prognostic impact but are not used routinely in clinical practice.

This chapter is a survey of the significance of the non-clinical prognostic factors (i.e., age, ethnicity) in finding the prognosis for breast cancer patients. The National Cancer Institute’s Surveillance, Epidemiology, and End Results (SEER) Public-Use Data (years 1977-2013; 421,056 records) (“SEER,” n.d.) and the Howard University Cancer Center Tumor Registry (1995-2013; 1599 records) data were analysed. NPI (Nottingham Prognostic Index (Galea, 1992)) is a prognostic tool that enables grouping of patients based on calculated prognosis (i.e., excellent, good, moderate, poor) and the impact of each non-clinical prognostic factor on these subgroups. In addition, survival analysis tools such as Cox proportional hazards and Kaplan-Meier survival curve (“Cox proportional-hazards regression”, 2013) were used in analyzing the data.

Our data analysis suggested that age, ethnicity, and marital status have some influence on the survivability rate of breast cancer patients. Whether such influence is significant enough in relationship with the clinical factors such as tumor size and grade remains to be further studied.

## **BACKGROUND**

Several studies have been carried out on the survivability prediction of breast cancer using Naïve Bayes and Classification Trees, Artificial Neural Networks and statistical techniques of regression (Delen et al., 2005; Gupta et al., 2011). Delen and colleagues (Delen et al., 2005) have used data mining algorithms Artificial Neural Networks, decision trees, and logistic regression to develop the breast cancer predic-

12 more pages are available in the full version of this document, which may be purchased using the "Add to Cart" button on the publisher's webpage:

[www.igi-global.com/chapter/exploratory-data-analysis-on-breast-cancer-prognosis/214628](http://www.igi-global.com/chapter/exploratory-data-analysis-on-breast-cancer-prognosis/214628)

## Related Content

---

### An End-to-End Network Evaluation Method for Differentiated Multi-Service Bearing in VPP

Wanqiao Wang, Jian Su, Hui Zhang, Luyao Guan, Qingrong Zheng, Zhuofan Tang and Huixia Ding (2024). *International Journal of Mobile Computing and Multimedia Communications* (pp. 1-16).

[www.irma-international.org/article/an-end-to-end-network-evaluation-method-for-differentiated-multi-service-bearing-in-vpp/340381](http://www.irma-international.org/article/an-end-to-end-network-evaluation-method-for-differentiated-multi-service-bearing-in-vpp/340381)

### Headache App: Usability Assessment and Criterion Validity

Tânia Dantas, Milton Rodrigues dos Santos, Alexandra Queirós and Anabela G. Silva (2018). *International Journal of Mobile Computing and Multimedia Communications* (pp. 1-11).

[www.irma-international.org/article/headache-app/205676](http://www.irma-international.org/article/headache-app/205676)

### A J2ME Mobile Application for Normal and Abnormal ECG Rhythm Analysis

Qiang Fang, Xiaoyun Huang and Shuenn-Yuh Lee (2010). *Handheld Computing for Mobile Commerce: Applications, Concepts and Technologies* (pp. 86-108).

[www.irma-international.org/chapter/j2me-mobile-application-normal-abnormal/41629](http://www.irma-international.org/chapter/j2me-mobile-application-normal-abnormal/41629)

### Corporate Disclosure Measurement

Md. Salah Uddin Rajib and Md. Qutub Uddin Sajib (2019). *Advanced Methodologies and Technologies in Network Architecture, Mobile Computing, and Data Analytics* (pp. 489-501).

[www.irma-international.org/chapter/corporate-disclosure-measurement/214638](http://www.irma-international.org/chapter/corporate-disclosure-measurement/214638)

### Adaptive Dynamic Path Planning Algorithm for Interception of a Moving Target

H. H. Triharminto, A.S. Prabuono, T. B. Adjani and N. A. Setiawan (2013). *International Journal of Mobile Computing and Multimedia Communications* (pp. 19-33).

[www.irma-international.org/article/adaptive-dynamic-path-planning-algorithm/80425](http://www.irma-international.org/article/adaptive-dynamic-path-planning-algorithm/80425)