

Tree-Based Model for Breast Cancer Prognostication

Mousumi Banerjee, Julie George, Eun Young Song, Anuradha Roy, and William Hryniuk

From the Department of Biostatistics, University of Michigan, Ann Arbor; Barbara Ann Karmanos Cancer Institute and the Center for Healthcare Effectiveness Research, Wayne State University, Detroit, MI; and the Department of Management Science and Statistics, The University of Texas at San Antonio, San Antonio, TX.

Submitted November 27, 2002; accepted April 7, 2004.

Supported by grants from the National Science Foundation (DMS 9973410; M.B.) and the Ford Motor Company Foundation (W.H. and E.Y.S.).

Authors' disclosures of potential conflicts of interest are found at the end of this article.

Address reprint requests to Mousumi Banerjee, PhD, Department of Biostatistics, University of Michigan, 1420 Washington Heights, Ann Arbor, MI 48109-2029; e-mail: mousumib@umich.edu.

© 2004 by American Society of Clinical Oncology

0732-183X/04/2213-2567/\$20.00

DOI: 10.1200/JCO.2004.11.141

A B S T R A C T

Purpose

To define prognostic groups for recurrence-free survival in breast cancer, assess relative effects of prognostic factors, and examine the influence of treatment variations on recurrence-free survival in patients with similar prognostic-factor profiles.

Patients and Methods

We analyzed 1,055 patients diagnosed with stage I-III breast cancer between 1990 and 1996. Variables studied included socioeconomic factors, tumor characteristics, concurrent medical conditions, and treatment. The primary end point was recurrence-free survival (RFS). Multivariable analyses were performed using recursive partitioning and Cox proportional hazards regression.

Results

The most significant difference in prognosis was between patients with fewer than four and those with at least four positive nodes ($P < .0001$). Four distinct prognostic groups (5-year RFS, 97%, 78%, 58%, and 27%) were developed, defined by the number of positive nodes, tumor size, progesterone receptor (PR) status, differentiation, race, and marital status. Patients with fewer than four positive nodes and tumor ≤ 2 cm, PR positive, and well or moderately differentiated had the best prognosis. RFS in this group was unaffected by type of adjuvant therapy ($P = .38$). Patients with at least four positive nodes and PR-negative tumors had the worst prognosis, and those treated with tamoxifen plus chemotherapy had the best outcome in this group ($P = .0001$). Among patients in the two intermediate-risk groups, those treated with tamoxifen or a combination of tamoxifen and chemotherapy had the best outcome.

Conclusion

Lymph node status, PR status, tumor size, differentiation, race, and marital status are valuable for prognostication in breast cancer. The prognostic groups derived can provide guidance for clinical trial design, patient management, and future treatment policy.

J Clin Oncol 22:2567-2575. © 2004 by American Society of Clinical Oncology

INTRODUCTION

Breast cancer is the most common malignancy diagnosed among women and the second leading cause of cancer mortality in the United States.¹ In 2004, approximately 215,990 new cases of female invasive breast cancer will be diagnosed in the United States, and 40,110 women will die of the disease.¹ A decline in breast cancer mortality rates has been observed since 1990.^{2,3} Results from two large population-based studies suggest the decline is due to adjuvant tamoxifen treatment in women older than 50 years and adjuvant chemotherapy in women younger than 50 years.^{4,5} Still, patients treated with adjuvant chemotherapy

or tamoxifen present great heterogeneity in terms of recurrence-free survival, suggesting that patient and/or tumor characteristics may influence outcome more strongly than modifications in standard therapeutic approaches. An analysis of the association of recurrence-free survival (RFS) with patient and tumor characteristics, as well as treatment-related variables, is necessary to assure reliable evaluation of new approaches for treatment of breast cancer.

The goals of the current study, which uses a nonparametric statistical technique known as recursive partitioning (RP), are to (1) analyze the relative contributions of patient and tumor-related prognostic factors to the RFS of patients with stage I-III breast