Predicting Response to Adjuvant and Radiation Therapy in Patients with Early Stage Breast Carcinoma

Harry B. Burke, M.D., Ph.D.[1]
Albert Hoang, Ph.D.[1]
J. Dirk Iglehart, M.D.[2]
Jeffrey R. Marks, Ph.D.[2]

[1] Department of Medicine, New York Medical College, Valhalla, New York.

[2] Departments of Surgery, Pathology, and Cell Biology, Duke University, Durham, North Carolina.

Supported in part by a research grant from the U.S. Army Medical Research and Development Command Breast Cancer Research Program (DAMD 17-94-J-4383) and by the Duke Specialized Program of Research (SPORE) of the National Cancer Institute.

Address for reprints: Harry B. Burke, M.D., Ph.D., Department of Medicine, New York Medical College, Valhalla, NY 10595.

Received March 14, 1997; revision received July 31, 1997; accepted September 8, 1997.

BACKGROUND. Screening and surveillance is increasing the detection of early stage breast carcinoma. The ability to predict accurately the response to adjuvant therapy (chemotherapy or tamoxifen therapy) or postlumpectomy radiation therapy in these patients can be vital to their survival, because this prediction determines the best postsurgical therapy for each patient.

METHODS. This study evaluated data from 226 patients with TNM Stage I and early Stage II breast carcinoma and included the variables p53 and c-*erb*B-2 (HER-2/*neu*). The area under the receiver operating characteristic curve (Az) was the measure of predictive accuracy. The prediction endpoints were 5- and 10-year overall survival.

RESULTS. For Stage I and early Stage II patients, the 5- and 10-year predictive accuracy of the TNM staging system were at chance level, i.e., no better than flipping a coin. Both the 5- and 10-year artificial neural networks (ANNs) were very accurate—significantly more so than the TNM staging system (Az 5-year survival, TNM = 0.567, ANN = 0.758; $P < 0.001$; Az 10-year survival, TNM = 0.508, ANN = 0.894; $P < 0.0001$). For patients not receiving postsurgical therapy and for either chemotherapy or tamoxifen therapy, the ANNs containing p53 and c-*erb*B-2 and the number of positive lymph nodes were accurate predictors of survival (Az 5-year survival, 0.781, 0.789, and 0.720, respectively).

CONCLUSIONS. The molecular genetic variables p53 and c-*erb*B-2 and the number of positive lymph nodes are powerful predictors of survival, and using ANN statistical models is a powerful method for predicting responses to adjuvant therapy or radiation therapy in patients with breast carcinoma. ANNs with molecular genetic prognostic factors may improve therapy selection for women with early stage breast carcinoma. *Cancer* **1998;82:874–7.** © *1998 American Cancer Society.*

KEYWORDS: TNM staging system, artificial neural networks, prognostic factors, breast carcinoma, tamoxifen therapy, chemotherapy, radiation therapy, outcomes, c-*erb*B-2, p53.

Screening and surveillance is increasing the prevalence of early stage breast carcinoma. The ability to predict accurately the responses to adjuvant therapy (chemotherapy or tamoxifen therapy) or postlumpectomy radiation therapy in these patients can be vital to their survival, because this prediction determines the best postsurgical therapy for each patient. The pathologic TNM staging system is the current cancer prognostic system. Its predictions are based on three variables: 1) location, size, and depth of tumor; 2) existence and location of involved lymph nodes; and 3) existence of distant metastases.[1] We have shown that artificial neural networks (ANNs)

are more accurate at predicting survival than the TNM staging system for all stages of breast carcinoma.[2] It is not known how accurate the TNM staging system is in predicting the survival of patients with early stage breast carcinoma. It is also not known whether ANNs with molecular genetic prognostic factors, i.e., p53 and c-*erb*B-2 (HER-2/*neu*), can improve prognostic accuracy in early stage breast carcinoma across postsurgical therapies and for specific therapies. This article compares the survival prediction accuracy of the TNM staging system with ANN models across all postsurgical therapies. In addition, it presents a method for properly assessing putative therapy-dependent prognostic factors and examines the accuracy of ANNs in terms of specific therapies. Because the TNM staging system does not predict response to adjuvant or radiation therapy, it is not included in the individual therapy analyses.

METHODS

Data

These data were described in detail in a previous article.[3] Briefly, all patients were pathologic TNM Stage I or early Stage II. Early stage breast carcinoma includes Stage I and limited Stage II. Limited Stage II included all the TNM Stage II patients except those with five or more positive lymph nodes. The variables were age, race, tumor size, lymph nodes positive, lymph node stage, nuclear grade, histologic grade, p53, c-*erb*B-2, estrogen receptor (ER) and progesterone receptor (PR) status, vascular invasion, adjuvant therapy (tamoxifen or chemotherapy), and radiation therapy. Patients who underwent a lumpectomy received radiation therapy. Patients who underwent a modified radical mastectomy did not receive radiation therapy. There were 229 cases, of which 226 had complete data for all variables except ER and PR status. Because of the number of cases missing, both ER and PR were removed from the data set. The survival rate was 70%. The prediction endpoints were 5-and 10-year overall survival.

Accuracy

The area under the receiver operating characteristic curve (Az) is a measure of prediction accuracy.[4] It can be used to assess and compare the adequacy of statistical models. Az can be directly calculated by Somer's D,[5] or it can be approximated by its trapezoidal area.[6] The area under the curve is a nonparametric measure of discrimination. It is independent of both the prior probability of each outcome and the threshold cutoff for category. Its computation requires only that the prediction method produce an ordinally scaled relative predictive score. In terms of mortality, the receiver operating characteristic area estimates the probability that the prediction method will assign a higher mortality score to the patient who died than to the patient who lived. The receiver operating characteristic area varies from 0 to 1. When the predictions are unrelated to survival, the score is 0.5, indicating chance accuracy. The farther the score is from 0.5, the better, on average, the prediction method is for predicting which of the two patients will be alive.

TABLE 1
Comparison of the Accuracy of the TNM Staging System and Artificial Neural Networks in Predicting the 5- and 10-year Survival of Patients with Early Stage Breast Carcinoma

Model	5-yr survival Az (SE)[a]	10-year survival Az (SE)[b]
TNM	0.567 (0.046)	0.508 (0.053)
ANN	0.758 (0.042)	0.894 (0.034)

ANN: artificial neural network; Az: area under the receiver operating characteristic curve; SE: standard error.
[a] TNM vs. ANN 5-year survival, $P < 0.001$.
[b] TNM vs. ANN 10-year survival, $P < 0.0001$.

Statistical Models

ANN models have been described in detail elsewhere.[2] Briefly, the three-layer backpropagation ANN was composed of an input layer, a hidden layer, and an output layer. Each layer of an ANN was composed of nodes. The number of input nodes was equal to the number of variables. The hidden layer was composed of three nodes. There was one output node. All the variables were entered into the three-layer ANN model. The two-layer ANN was identical to the three-layer network, except that it did not possess a hidden layer. After a sensitivity analysis to reduce the number of input variables to the three with the highest predictive accuracy, the three selected variables, namely, the number of positive lymph nodes, p53, and c-*erb*B-2, were entered into the two-layer ANN. Both the two- and three-layer ANNs employed the maximum likelihood loss function and weight decay. Model accuracy estimates and standard errors were calculated by the bootstrap resampling method.[7]

RESULTS

The predictive accuracies of the TNM staging system and the three-layer ANN models are shown in Table 1. For Stage I and early Stage II patients, the 5- and 10-year prediction accuracy of the TNM staging sys-

TABLE 2
Artificial Neural Network Accuracy in Predicting 5-Year Survival for Each Theory Combination

Strata	T	C	R	No. of cases	5-yr survival Az (SE)
1			–	48	0.781 (0.091)
2		–	+	23	
3		+	–	53	0.789 (0.049)
4		+	+	19	
5	+	–	–	43	0.720 (0.072)
6	+	–	+	7	
7	+	+	–	14	
8	+		+	19	

T: tamoxifen; C: chemotherapy; R: radiation; Az: area under the receiver operating characteristic curve; SE: standard error; +: patient received the therapy; –: patient did not receive the therapy.

tem was at chance level, i.e., no better than flipping a coin. Both the 5- and 10- year ANNs were very accurate and significantly more accurate than the TNM staging system (Az 5- year survival, TNM = 0.567, ANN = 0.758, $P < 0.001$; Az 10-year survival, TNM = 0.508, ANN = 0.894, $P < 0.0001$).

The evaluation of therapy-dependent prognostic factors requires the mutually exclusive and exhaustive partitioning of the adjuvant therapies and radiation therapy. Because the numbers of patients and outcomes were small in this and in the subsequent analyses, three variables (number of positive lymph nodes, p53, and c-*erb*B-2) and two-layer ANNs with a 5-year survival endpoint were employed. The stratification by postsurgical therapy into eight bins is shown in Table 2.

There was a no-therapy bin (Stratum 1) and there were bins representing all combinations of the three postsurgical therapies, i.e., tamoxifen, chemotherapy, and radiation therapy (Strata 2–8). Only Stratum 1 (no adjuvant therapy), Stratum 3 (only chemotherapy), and Stratum 5 (only tamoxifen) contained enough patients for analysis. The ANNs for these three strata were accurate predictors of survival (Az 5-year survival, 0.781, 0.789, and 0.720, respectively).

The number of cases in each bin could be increased by stratifying by therapy regardless of whether a patient received another therapy. This was not a mutually exclusive and exhaustive partitioning of the therapy variables. Thus, the results must be viewed as an approximation, because the variables were not being treated as purely therapy-dependent prognostic factors. Table 3 shows the accuracy of the ANN for each of the three therapies. With larger numbers in each strata, it is clear that the ANNs that contained the three variables lymph nodes positive, p53, and c-*erb*B-2 were excellent predictors of response to adjuvant therapy (Az 5-year survival, tamoxifen = 0.855, chemotherapy = 0.782, radiation = 0.861).

DISCUSSION

We have demonstrated that ANNs that contain p53 and c-*erb*B-2 are significantly more accurate than the TNM staging system at predicting 5- and 10-year survival in women with early stage breast carcinoma. We have also demonstrated that the molecular genetic variables p53 and c-*erb*B-2 and the number of positive lymph nodes can be used to accurately predict responses to surgery, chemotherapy, and tamoxifen therapy.

An understanding of therapy-dependent prognostic factors, and why there must be a mutually exclusive and exhaustive partitioning of the therapies prior to the assessment of therapy-dependent prognostic factors, requires a description of the types and functions of prognostic factors. Prognostic factor types are defined in terms of their function. There are three prognostic factor functions and therefore three types of prognostic factors: natural history, therapy-dependent, and posttherapy.[8] Natural history prognostic factors predict the course of the disease if no effective therapy exists or if an effective therapy is not administered. For example, clinically palpable lymph nodes may be a natural history prognostic factor. Therapy-dependent prognostic factors predict, prior to the patient's receiving the therapy, a change in the course of the disease caused by a change in the patient's

TABLE 3
Artificial Neural Network Accuracy in Predicting 5-Year Survival for Each Therapy

Treatment group	No. of cases	5-yr survival Az (SE)
Tamoxifen	83	0.855 (0.052)
Chemotherapy	105	0.782 (0.055)
Radiation	68	0.861 (0.047)

Az: area under the receiver operating characteristic curve; SE: standard error.

condition due to receipt of an effective therapy. For example, ER status may predict response to tamoxifen. Posttherapy prognostic factors predict, after the patient has received the therapy, whether there has been a change in the course of the disease due to the intervention. For example, the number of positive lymph nodes on axillary dissection may predict whether the patient will respond to the primary surgery. Posttherapy prognostic factors are important because we do not want to wait any longer than necessary to administer a second-line therapy to patients who do not respond to the primary therapy. All three prognostic factors are relative to therapy. For each therapy in a succession of therapies (for example, if a therapy is given and the patient does not respond to that therapy and another therapy is contemplated), all three types of prognostic factors can be analyzed.

Within the context of the small sample size of this study, the molecular genetic variables p53 and c-*erb*B-2 are powerful therapy-dependent prognostic factors for early stage breast carcinoma, and ANN models are an efficient statistical method for capturing their predictive power.

REFERENCES

1. Beahrs OH, Henson DE, Hutter RVP, Kennedy BJ, editors.. Manual for staging of cancer. American Joint Committee on Cancer. 4th edition. Philadelphia: J. B. Lippincott, 1992.
2. Burke HB, Goodman PH, Rosen DB, Henson DE, Weinstein JN, Harrell Jr. FE, et al. Artificial neural networks improve the accuracy of cancer survival prediction. *Cancer* 1997;79: 857–62.
3. Marks JR, Wu K, Berry D, Bandarenko N, Kerns BJ, Iglehart JD. Overexpression of p53 and HER-2/*neu* proteins as prognostic markers in early stage breast cancer. *Ann Surg* 1994; 219:332–41.
4. Swets JA. Signal detection theory and ROC analysis in psychology and diagnostics: collected papers. Mahwah, NJ: Lawrence Erlbaum Associates, 1996.
5. Somers RH. A new asymmetric measure of association for ordinal variables. *Am Sociological Rev* 1962;27:799–811.
6. Bamber D. The area above the ordinal dominance graph and the area below the receiver operating characteristic. *J Math Psy* 1975;12:387–415.
7. Efron B, Tibshirani RJ. An introduction to the bootstrap. New York: Chapman & Hall, 1993:45–57.
8. Burke HB. Increasing the power of surrogate endpoint biomarkers: the aggregation of predictive factors. *J Cell Biochem* 1994;19S:278–82.