Clinical Cancer Research

A Prognostic Model That Makes Quantitative Estimates of Probability of Relapse for Breast Cancer Patients

Michelino De Laurentiis, Sabino De Placido, Angelo R. Bianco, et al.

Updated Version	Access the most recent version of this article at: http://clincancerres.aacrjournals.org/content/5/12/4133

Cited Articles	This article cites 33 articles, 4 of which you can access for free at: http://clincancerres.aacrjournals.org/content/5/12/4133.full.html#ref-list-1
Citing Articles	This article has been cited by 3 HighWire-hosted articles. Access the articles at: http://clincancerres.aacrjournals.org/content/5/12/4133.full.html#related-urls

E-mail alerts	Sign up to receive free email-alerts related to this article or journal.
Reprints and Subscriptions	To order reprints of this article or to subscribe to the journal, contact the AACR Publications Department at pubs@aacr.org.
Permissions	To request permission to re-use all or part of this article, contact the AACR Publications Department at permissions@aacr.org.

A Prognostic Model That Makes Quantitative Estimates of Probability of Relapse for Breast Cancer Patients[1]

Michelino De Laurentiis,[2] Sabino De Placido, Angelo R. Bianco, Gary M. Clark, and Peter M. Ravdin

Cattedra di Oncologia Medica, Dipartimento di Endocrinologia e Oncologia Molecolare e Clinica, Università 'Federico II', 80131 Napoli, Italy [M. D. L., S. D. P., A. R. B.]; Breast Center at Baylor College of Medicine, Houston, Texas 77030 [G. M. C.]; and Division of Medical Oncology, Department of Medicine, University of Texas Health Science Center at San Antonio, San Antonio, Texas 78229-3900 [P. M. R.]

ABSTRACT

Tumor-node-metastasis (TNM) staging is the standard system for the estimation of prognosis of breast cancer patients. However, this system does not exploit information yielded by markers of the biological aggressiveness of breast cancer and is clearly unsatisfactory for optimal-treatment decision-making and for patient counseling. We have developed a prognostic model, based on a few routinely evaluated prognostic variables, that produces quantitative estimates for risk of relapse of individual breast cancer patients. We used data concerning 2441 of 2990 consecutive breast cancer patients to develop an artificial neural network (ANN) for the prediction of the probability of relapse over 5 years. The prognostic variables used were: patient age, tumor size, number of axillary metastases, estrogen and progesterone receptor levels, S-phase fraction, and tumor ploidy. Performances of the model were evaluated in terms of discrimination ability and quantitative precision. Predictions were validated on an independent series of 310 patients from an institution in another country. The ANN discriminated patients according to their risk of relapse better than the TNM classification ($P = 0.0015$). The quantitative precision of the model's estimates was accurate and was confirmed on the series from the second institution. The 5-year relapse risk yielded by the model varied greatly within the same TNM class, particularly for patients with four or more nodal metastases. The model discriminates prognosis better than the TNM classification and is able to identify patients with strikingly different risks of relapse within each TNM class.

INTRODUCTION

The best cure rates for early breast cancer are achieved by a combination of loco-regional and systemic treatments (usually referred to as "adjuvant treatments"; Refs. 1–3). However, cure rates are unrelated to the extent of loco-regional intervention (4). This has prompted a tendency toward less extensive loco-regional treatment (breast-conserving surgery) and a large body of research aimed at developing new adjuvant regimens.

Clinicians have now a wide choice of adjuvant regimens, ranging from low-toxicity hormonal regimens to very aggressive high-dose myeloablative chemotherapy with hematopoietic stem cell support, which is still under investigation. However, the wider the choice of adjuvant regimens, the harder it is to determine which treatment is best for a given patient. Decision-making about adjuvant treatment involves three issues. For each patient, one must first determine the likelihood of disease recurrence if no further therapy is administered and then the efficacy of each therapy available for that patient. Thirdly, the predicted therapeutic effect of each specific regimen must be weighted against potential side effects so as to determine the probable net benefit. Therefore, prognostic estimation remains one of the basic issues of treatment decision-making, not simply to identify patients who, because of their excellent prognosis, should be spared any adjuvant treatment-induced toxicity but also to identify those high-risk patients who should be given the opportunity of very aggressive regimens.

At present, TNM[3] staging is used to classify breast cancer patients according to prognoses (5, 6). This type of categorization, based on the extent of the disease at diagnosis, does not exploit the many markers of biological aggressiveness of breast cancer that are now widely available, and it is not satisfactory for optimal-treatment decision-making and patient counseling. On the other hand, it has been demonstrated that practicing oncologists are inaccurate when making prognostic estimates based on information obtained from multiple prognostic variables (7). Hence, it is widely recognized that a new prognostic system is needed for breast cancer patients, and the development of such a system is a major goal of cancer research.

Using data from a large breast cancer database, we generated a prognostic model based on an ANN. This model yielded numerical estimates for risk of relapse of individual breast cancer patients that we validated on an independent data set of patients treated and followed at another institution.

[3] The abbreviations used are: TNM, tumor-node-metastasis; ANN, artificial neural network; TS, tumor size; ER, estrogen receptor; PgR, progesterone receptor; DFS, disease-free survival; ROC, receiver operating characteristic; fmol, femtomole(s).

Fig. 1 Data profile for the San Antonio and the Naples databases. *SPF*, S-phase fraction.
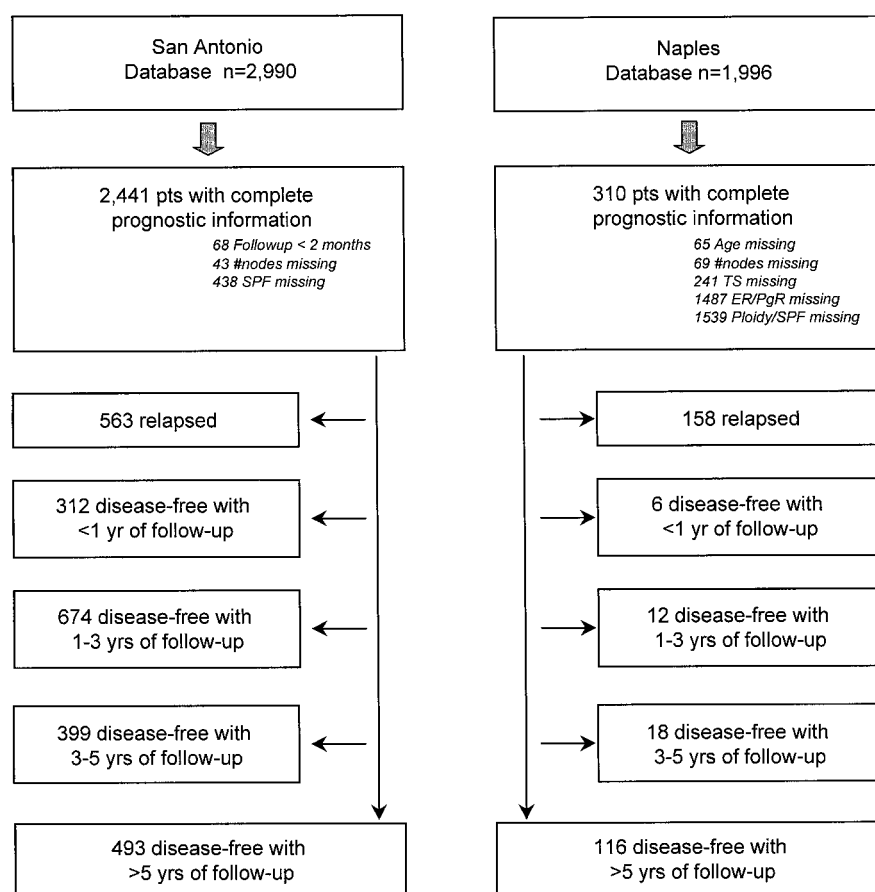
PATIENTS AND METHODS

Data. The model described here was constructed using data from 2990 patients stored in the "San Antonio database." These data pertain to a cohort of breast cancer patients whose tumor specimens were consecutively received at the Division of Medical Oncology, University of Texas Health Science Center at San Antonio (San Antonio, TX) for prognostic factors evaluation. Patients had undergone surgical resection between 1974 and 1992 at a broad range of sites across the United States. Data were collected in a prospective fashion and partly verified by audit visits at the single treatment sites with direct check of the original source documents. After computerization, database information was verified with the referring physicians to minimize data entry errors. Recurrences were actively searched for by systematic patient follow-up. The follow-up schedule varied according to each center policy.

The following information was available for these cases: (*a*) age, (*b*) TS; (*c*) number of axillary lymph nodes affected by tumor (#Nodes), (*d*) ER levels; (*e*) PgR levels; (*f*) proportion of tumor cells in S-phase (S-phase fraction); and (*g*) tumor ploidy. The analysis was restricted to patients who fulfilled the following criteria: (*a*)invasive cancer; (*b*) follow-up of at least 2 months; and (*c*)no missing values for prognostic factors data. Data profile is reported in Fig. 1.

Age was recorded at the time of surgery. TS was defined as the larger diameter of the invasive tumor at pathological examination. The minimum number of axillary lymph nodes examined was equal to 6. ER and PgR were assayed with the charcoal-coated dextrane method as described elsewhere (8) and are expressed as femtomoles per milligram of cytoplasmic proteins (fmol/mg protein). Ploidy and S-phase fraction were estimated with DNA flow cytometry of fresh or frozen tumor specimens as previously reported (9). Prognostic variables from (*b*) to (*f*) were logarithmically transformed to reduce the asymmetry of the distribution of the values and to obtain approximate normality. This stabilized the variance and reduced the impact of the small numbers of cases with very high values for given prognostic variables (10, 11). Treatment information was not available at the time of the analysis and was not used for modeling. See Table 1 for a list of the prognostic variables used for the construction of the model and their respective transformations.

To reduce the risk of overfitting of the model and to test its generalizing ability, we randomly split the entire data sample into three subsets of approximately equal size: (*a*) a training Set; (*b*) a testing Set; and (*c*) an internal validation set (12, 13). The training set was used to develop the model, the testing set to select the optimal duration of the training process, and the internal validation set to assess performance.

Table 1 Prognostic variables used for the construction of the model and their transformation

Original prognostic variable	After transformation	Type of variable
Age (years)	Age (years)	Continuous
TS (mm)	log (TS + 1)	Continuous
#Nodes[a]	log (#Nodes + 1)	Continuous
ER (fmol/mg protein)	log (ER + 1)	Continuous
PgR (fmol/mg protein)	log (PgR + 1)	Continuous
S-phase fraction	log (S phase + 1)	Continuous
Ploidy	1 if diploid, 0 if nondiploid	Categorical

[a] #Nodes, number of axillary lymph node metastases.

External Validation. After the model was constructed, a data set from a different institution ("external validation set") became available. This database contains information regarding 1996 consecutive cases of breast cancer treated between 1977 and 1993 and followed at the University Federico II of Naples, Italy. Complete prognostic data for the variables in the model were available for only 310 patients; these were used to check the robustness of the estimates of the model across an independent series. See Fig. 1 for the profile of the data.

ANN Modeling. DFS was defined as the time elapsing between surgery and any of the following events, whichever occurred first: local recurrence, distant relapse, second breast cancer, or death. Patients without any event were censored at the time of the last follow-up.

The ANN model was constructed using a variation of the technique for censored data that we described previously (14–17). It is a multilayer perceptron with one hidden layer of five processing elements. The hyperbolic tangent was chosen as activation function of the processing elements. We used data in the training set for the learning process of the ANN (training). During this process, prognostic variables pertaining to each patient in this data set were iteratively presented to the ANN, the target value being each patient's relapse status at 60 months. For patients whose status at 60 months was unknown (patients censored before 60 months) the conditional probability of relapse from the Kaplan Meier curve was used as surrogate target value. This conditional probability is expressed as the ratio between the average probability of relapse at 60 months and the average probability of relapse at the time the patient was censored. The back-propagation-of-error algorithm was used to adjust the connection weights during the learning process (18).

The testing set was then used to check when, in the training process, the ANN had reached optimal predictive power (curtailed training; 13) and to adjust the model's calibration. The validation set was finally used to assess the predictive power of the model.

ANNs were implemented on a Sun Sparkstation IPX (SUN Microsystems, Inc.) using the Nworks Professional II plus software package (NeuralWare, Inc., Pittsburgh, PA).

Evaluation Criteria. Performances of the ANN model were assessed in terms of refinement and calibration. The refinement is defined as the ability of a model to rank patients on the basis of their true probability of relapse, *i.e.,* to discriminate between patients with different prognoses. The calibration is defined as numerical precision of the predicted probability of relapse, *i.e.,* how close are the model's predictions to the true probability of relapse (19, 20).

The ANN model was used to make predictions of risk of relapse at 5 years for patients in the internal validation set and in the Naples database. The refinement of the model was evaluated as area-under-the-ROC-curve (Az) for predicting relapses on both data sets. This area is an expression of the probability that, for two randomly chosen patients whose status at 5 years is known, one of whom is relapsed and the other is not relapsed, the relapsed one had the highest predicted risk (21). This method tests the model's sensitivity and specificity. The values of the area-under-the-ROC-curve were calculated by a maximum likelihood estimation, and the statistical significance of the differences between these values was evaluated by a z statistic (22, 23). These estimations were made with the software "Indroc" for independent data and "Corroc2" for correlated data.[4]

The calibration of the ANN model was tested by comparing the average predicted probability with the observed relapse rate at 5 years of follow-up within different prediction ranges. Observed relapse rates were estimated according to the product-limit method (24).

The relationships between the ANN projections and the current TNM prognostic system were analyzed by examining the distribution of the ANN projections for each TNM class. Individual probabilities of relapse, as predicted by the ANN model, were clustered into 10 (0.1-wide) intervals, and a smoothed plot of the percentage of patients falling into each interval was reported for all of the TNM classes.

All *P*s were two-sided.

RESULTS

The distribution of the prognostic characteristics of the patients in the four data sets used is shown in Table 2. Apart from the longer follow-up and the higher proportion of node-positive patients in the Naples database, there was no other significant difference in the variable distribution among the data sets.

After completion of the training process, the ANN model was used to project a quantitative estimate of the 5-year probability of recurrence for each patient in the internal validation set and in the Naples database (external validation set). The predictive accuracy of these estimates is shown in Table 3. The discrimination ability of the model, expressed in terms of Az values, is good and does not show any degradation across data sets (Az: 0.728 *versus* 0.732; $z = -1.60$; $P = 0.10$). These Az values represent the probability that random pairs of patients (one relapsed and one disease-free) are correctly ranked by the model on the basis of their predicted risk. The calibration of the ANN model, *i.e.,* how close the individual model predictions are to the actual probability of relapse, is also reported in Table 3. To generate this table, patients were ranked on the basis of their ANN predicted risk and then divided into eight groups of increasing risk (ANN1 to ANN8). For each group, the average model prediction, the observed relapse rate at 5 years and the number of patients falling in that prediction range were tabu-

---

[4] Charles E. Metz, Helen B. Kronman, Pu-Lan Wang and Jong-Her Shen, Department of Radiology, The University of Chicago, Chicago, IL.

Table 2 Distribution of the patients' main characteristics in the four data sets

	Training set (n = 811) median (range)	Testing set (n = 816) median (range)	Internal validation set (n = 814) median (range)	External validation set (n = 310) median (range)
Follow-Up (months)	33 (21–177.3)	35 (2.1–172.9)	36.7 (2–173)	86 (8–217)
Age (years)	60 (21–89)	59 (25–102)	60 (25–91)	59 (25–79)
TS (cm)	2.5 (0.3–22)	2.2 (0.3–13)	2.5 (0.4–12)	2.6 (0.3–15)
S phase (%)	7.6 (0.07–54.6)	7.4 (0.10–70)	6.75 (0.10–52)	9.1 (0.10–62)
ER (fmol/mg protein)	33.5 (0–1797)	30 (0–2824)	33 (0–1439)	25 (0–1100)
PgR (fmol/mg protein)	14.5 (0–2963)	9 (0–2824)	11.5 (0–3325)	7.1 (0–2990)
	n	n	n	n
Nodal status				
N−	463	475	471	140
N+	348	341	343	170
Ploidy				
Diploid	329	340	369	119
Nondiploid	482	476	445	191

Table 3 Predictive accuracy of the model's projections on the internal validation set and on the external validation set

Risk group	Prediction range	Internal validation set [Az^a 0.728 (SD = 0.03)]			External validation set [Az 0.732 (SD = 0.04)]		
		Average prediction	Observed relapse rate (SE)	n	Average prediction	Observed relapse rate (SE)	n
ANN1	0.00–0.15	0.133	0.110 (0.03)	172	0.130	0.121 (0.06)	16
ANN2	0.16–0.25	0.200	0.170 (0.03)	290	0.210	0.231 (0.05)	78
ANN3	0.26–0.35	0.290	0.245 (0.05)	122	0.296	0.294 (0.03)	88
ANN4	0.36–0.45	0.404	0.385 (0.07)	72	0.415	0.450 (0.08)	39
ANN5	0.46–0.55	0.494	0.465 (0.08)	64	0.510	0.530 (0.09)	36
ANN6	0.56–0.65	0.583	0.549 (0.08)	34	0.589	0.578 (0.10)	23
ANN7	0.66–0.80	0.712	0.698 (0.08)	39	0.725	0.865 (0.09)	15
ANN8	0.81–1.00	0.903	0.914 (0.09)	21	0.911	0.869 (0.09)	15

[a] Az, area-under-the-ROC-curve.

lated. Individual patients' predictions of 5-year probability of relapse ranged from 0.06 to 0.97 and were well distributed along the entire prognostic range. A comparison of predicted and observed values shows that the prediction of the model is consistent with the true relapse rate in each group of risk, except for a small discrepancy in the ANN7 group of the external validation set (0.725 predicted *versus* 0.865 observed).

To explore the clinical potential of the model, patients in the two validation sets were pooled (n = 1124). The predictive ability (refinement) of the model was then compared with that of the TNM staging. According to usual clinical practice, TNM was modified to account for the number of axillary lymph node metastases with resulting five TNM classes: T_1N_0, T_2N_0, T_3N_0, N_{1-3} (patients with 1–3 metastatic nodes), N_{4+} (patients with 4 or more metastatic nodes). As shown in Table 4, the predictive ability of the ANN model, expressed in terms of Az value, is significantly higher than that of the modified TNM classification (P = 0.0015). The clinical meaning of this difference is well depicted by the DFS curves according to the five TNM classes and the eight ANN-based prognostic groups (Fig. 2).

To reveal the disagreements between the ANN model and the TNM staging, in Fig. 3 is plotted the distribution of the ANN predictions for patients belonging to the five classes of the

Table 4 Discrimination ability of the ANN model *versus* the modified TNM system on patients of the two validation sets

	Az^a (SD)	z	P
ANN	0.726 (0.02)	−3.1778	0.0015
TNM	0.677 (0.02)		

[a] Az, area-under-the-ROC-curve; z, z statistic.

modified TNM system. The average ANN-projected 5-year probability of relapse was consistent with the observed relapse rate for each TNM class, but individual predictions ranged widely within each single TNM category, with a wide overlap across different classes. In other words, according to the projections of the ANN model, the probability of relapse in each modified TNM class is markedly heterogeneous. This is particularly true for the N_{4+} patients, some of whom have an ANN projected risk lower than that of some patients in each of the lower TNM classes (ANN-prediction range for N_{4+} patients, 0.19–0.97). To investigate whether or not this variability of estimates for N_{4+} could be related to an interaction with the adjuvant treatment, the analysis was repeated for patients who were both ER- and PgR-negative (Fig. 3). ANN estimates for
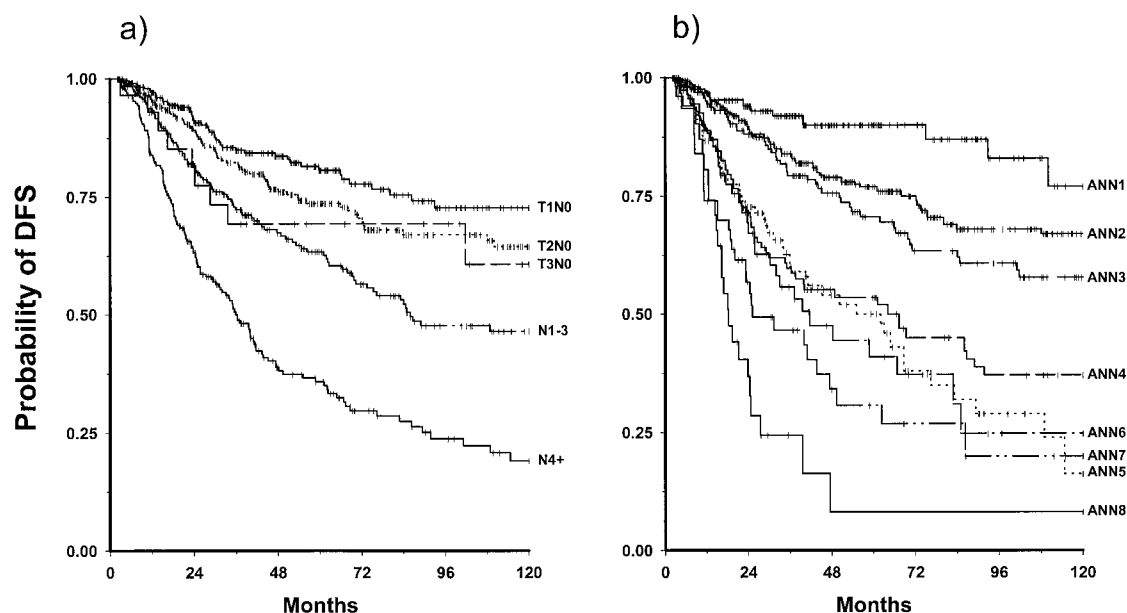
Fig. 2 Kaplan-Meier curves of DFS for patients of the two pooled validation sets ($n = 1124$), according to: *a*, modified TNM system; *b*, ANN prognostic groups.
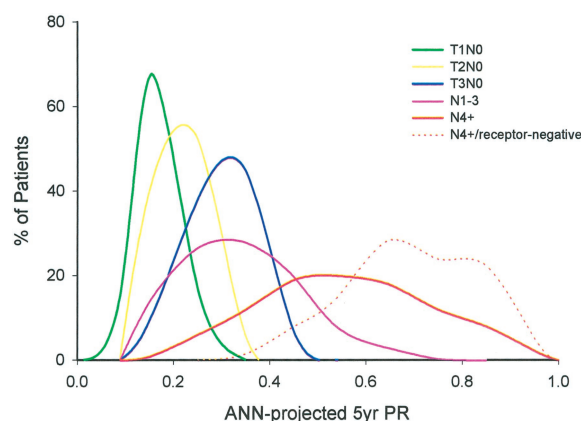

Fig. 3 Distribution of the ANN estimates of 5-year probability of relapse (*PR*) for patients classified according to the modified TNM system.

this subset of patients were less widely distributed, as expected for the known interaction of steroid receptors with endocrine treatments (2, 3). Nevertheless, a substantial heterogeneity of the AAN estimates was still present (ANN-prediction range for N4+/receptor-negative patients, 0.35–0.97).

DISCUSSION

Because of attempts to obtain a more reliable prognosis of breast cancer patients compared with the classical TNM staging, practicing oncologists are now facing a growing number of prognostic parameters. Paradoxically, this wealth of prognostic information has not improved clinicians' skill in predicting the outcome for breast cancer patients but has rather led to confusion. A survey of medical oncologists showed a surprisingly

wide variation in estimates of risk of relapse— provided by different expert clinicians—for a given patient (7). The conflicting estimates arise from the difficulty in taking into account complex, and sometimes conflicting, prognostic information. The subjective weighting of each prognostic variable does not yield a consistent quantitative evaluation of recurrence risk, and there is no standard procedure that takes into account various prognostic factors in an integrated fashion.

The aim of this study was to generate a prognostic model in which numerical information from several prognostic variables are integrated so as to yield quantitative predictions for risk of relapse of individual breast cancer patients. The advantages and disadvantages of the diverse multivariate analysis techniques that can be used to generate such a model are amply discussed elsewhere (17, 25). We used an approach based on ANN. ANNs are a form of artificial intelligence devised to solve complex pattern recognition and prediction problems in many fields including biomedicine (26–30). Feed-forward ANNs like the one that we used are trained with historical data to learn the correlation between input variable patterns and a specified output. The training consists of an iterative process of presentation of known data to the network. During training, the network changes its internal structure by updating connection weights thanks to a complex mathematical algorithm to minimize the global error in predicting the output. Once the training is terminated, an ANN can be presented with new patterns of input variables to make predictions about the output. In theory, given a large enough historical database, a properly constructed and trained ANN can learn any kind of functional or statistical relationship between the input variables and the output, no matter how complex and nonlinear it is. Because prognostic estimation is essentially a problem of pattern recognition and prediction, ANN analysis seems particularly suited to perform this task.

The application of ANN to survival analysis was initially

viewed with skepticism by biostatisticians and with exaggerated enthusiasm by the "neural net community," followed by a progressive convergence of the two views. Many groups are trying to merge classical statistical techniques and ANNs, and it has been shown that the back-propagation-of-error algorithm can be used to obtain maximum likelihood estimates in standard regression models for survival data (31, 32). Although many clinicians view ANNs as magic boxes able to solve complex problems, feed-forward networks, like the one that we used, share some features and limitations with classical statistical techniques (33, 34). Indeed, ANNs should be viewed as flexible nonlinear multiple-regression models that, by searching in a multidimensional variable space, are particularly good in detecting complex interactions among variables in large data sets (16, 17).

Using simulation with synthetic data sets, we previously demonstrated that there are three conditions in which ANN should be better than standard statistical models for survival analysis: (*a*) when the prognosis of patients is a complex nonlinear unknown function of the prognostic variables; (*b*) when this function varies over time; and (*c*) when there are complex interactions among prognostic variables (16, 17). Complexity is a common feature in biological systems; thus, with the advent of new biological prognostic variables, these three situations are very likely to be present, and we venture to suggest that ANN-based models will be better than standard statistical modeling in this setting.

Major drawbacks of the ANN analysis are the complexity of the training process and the lack of standardization in selecting optimal starting parameters (such as the number of hidden layers and processing elements) for the network being trained. However, subtleties of training are irrelevant to a network whose predictions have been validated. Moreover, although the training process may appear cumbersome, once trained, the ANN model can be easily implemented as an easy-to-use prediction program on any desktop, notebook, or palmtop computer. Prototypes of this kind of software have already been developed at our institution and will be made available to the interested reader.

Here we demonstrate that several prognostic factors may be integrated as numerical variables into an ANN model able to generate survival predictions for individual patients. In our model, we used only well-established prognostic variables that are routinely used and whose measurement is well standardized. The ANN-based model produced very accurate predictions of probability of relapse that are robust across different series of patients. The ANN predictions were better than those obtained with the commonly used TNM variant in terms of ability to discriminate the 5-year risk of disease recurrence among breast cancer patients. Furthermore, the calibration of the model, *i.e.,* the precision of the quantitative estimation of this risk, was very accurate and fitted surprisingly well to a series of patients treated and followed in a different country.

Our model, however, has some limitations. First, detailed treatment information was not available when we constructed the model and is not taken into account by the model. Predictions of the model may, thus, underestimate the natural patient-prognosis. This bias, however, is likely to be modest for node-negative patients because the majority of them in the San Antonio database did not receive any adjuvant therapy (percentage of treated patients: $T_1N_0 = 39\%$; $T_2N_0 = 39\%$; and $T_3N_0 =$ 40%). On the other hand, about 89% of node-positive patients whose data were used for model generation were treated. This may raise concerns about the usefulness of our ANN model. Nonetheless, because no patient whose data were entered into the model received high-dose chemotherapy, the model predictions can be viewed as estimates of prognosis of the "average-treated" node-positive subject. Therefore, because all of the new node-positive patients are being treated, the model predictions are still useful because they provide information that may be functional both for the decision to treat with very aggressive regimens and for counseling. A second limitation of the ANN model deserve a comment: in the database used for model construction, only 6% of the patients have tumors less than 1 cm in size. This is because we required prognostic information that was not easily obtainable on small tumors. Thus the model should be used with caution for projections of relapse rates for patients with such small tumors.

Another finding of our study is worthy of discussion. The ANN quantitative estimates of probability of relapse, when applied to patients grouped by the modified TNM classification, highlights the prognostic heterogeneity of each TNM group. This coincides with clinical experience, in which the outlook of patients belonging to the same TNM class can be quite different. In our study, the heterogeneity was particularly striking for N_{4+} patients whose projected relapse risks overlapped those of patients belonging to other TNM classes, including T_1N_0 patients. As shown, this heterogeneity is partially due to the inclusion in the model of ER and PgR information and, thus, related to some extent to the predictive effect of these variables on the adjuvant tamoxifen efficacy, rather than to a pure prognostic effect. Nonetheless, even among receptor-negative patients a substantial heterogeneity of the ANN estimates is observed. Because the effect of endocrine treatment is poor for these patients, and given that chemotherapy efficacy is not actually related to any other variable in the model (1), this variability of the ANN estimates may be linked to only a genuine prognostic effect of the covariates. This observation may have important clinical implications. Indeed, N_{4+} patients are generally considered at high risk of relapse and are candidates for very aggressive experimental regimens, including new drug combinations or high-dose myeloablative regimens with autologous stem cell reinfusion, which are usually very toxic and have a high incidence of side effects and a nontrivial rate of toxic deaths. According to our model, some N_{4+} patients are being over-treated, whereas patients in lower TNM classes but with a worse estimated prognosis are denied potentially more effective treatment. Should the estimates of our model be confirmed, selection of patients for aggressive regimens based only on the number of axillary metastases should be abandoned.

Various multivariate prognostic indices based on biological markers have been proposed (35–37). These indices have been used to classify patients in a few groups with different prognoses. Some of these indices have good performances, and one of them—the Nottingham Prognostic Index—has been validated prospectively (37). They all, however, suffer, in theory, from the same shortcoming as the TNM system, *i.e.,* heterogeneous prognoses for patients assigned to the same prognostic class. By contrast, our model, rather than classifying patients in

different prognostic groups, produces quantitative predictions of the probability of relapse on a continuous scale.

In conclusion, three findings of this study are clinically relevant: (*a*) using a few prognostic variables, we have generated an ANN model that produces quantitative estimates that represent an improvement on the classical TNM staging of breast cancer patients; (*b*) the model's estimates confirm that the system used presently is inadequate for optimal decision-making regarding the treatment; and (*c*) the accuracy of the model's projections is maintained on a series of patients from a different country. This accuracy, however, is amenable to further improvement: (*a*) using detailed treatment information; (*b*) using new biological variables; and (*c*) generating models on subsets of patients who pose specific therapeutic questions.

ACKNOWLEDGMENTS

REFERENCES

1. Early Breast Cancer Trialists' Collaborative Group. Polychemotherapy for early breast cancer: an overview of the randomised trials. Lancet, *352:* 930–942, 1998.

2. Early Breast Cancer Trialists' Collaborative Group. Tamoxifen for early breast cancer: an overview of the randomised trials. Lancet, *351:* 1451–1467, 1998.

3. Early Breast Cancer Trialists' Collaborative Group. Ovarian ablation in early breast cancer: overview of the randomised trials. Lancet, *348:* 1189–1196, 1998.

4. Early Breast Cancer Trialists' Collaborative Group. Effects of radiotherapy and surgery in early breast cancer: an overview of the randomized trials. N. Engl. J. Med., *333:* 1444–1455, 1995.

5. Fleming, I. D., Cooper, J. S., Henson, D. E., *et al.* (eds.), American Joint Committee on Cancer Cancer Staging Manual, Ed. 5. Philadelphia: Lippincott-Raven, 1995.

6. Yarnold, J., Dixon, J., and Greco, M. Consensus document on staging for breast cancer. Breast, *3:* 238, 1994.

7. Loprinzi, C. L., Ravdin, P. M., De Laurentiis, M., and Novotny, P. Do American oncologists know how to use prognostic variables for patients with newly diagnosed primary breast cancer? J. Clin. Oncol., *12:* 1422–1426, 1994.

8. McGuire, W. L., De La Garza, M., and Ghamness, G. C. Evaluation of estrogen receptor assay in human breast cancer tissue. Cancer Res., *37:* 637–639, 1977.

9. Dressler, L. G., Seamer, L. C., Owens, M. A., Clark, G. M., and McGuire, W. L. DNA flow cytometry and prognostic factors in 1331 frozen breast cancer specimens. Cancer (Phila.), *61:* 420–427, 1988.

10. Gamel, J. W., and McLean, I. W. A method for determining the optimum transform for covariates of the Cox Model with the application to 3680 cases of intraocular melanoma. Comput. Biomed. Res., *21:* 471–477, 1988.

11. Chapman, J. W., Murray, D., McCready, D. R., Hanna, W., Kahn, H. J., Lickley, H. L. A., Trudeau, M. E., Mobbs, B. G., Sawka, C. A., Fish, E. B., and Pritchard, K. I. An improved statistical approach: can it clarify the role of new prognostic factors for breast cancer? Eur. J. Cancer, *32A:* 1949–1956, 1996.

12. Miller, M. E., Hui, S. L., and Tierney, W. M. Validation techniques for logistic regression models. Stat. Med., *10:* 1213–1226, 1991.

13. Astion, M. L., Wener, M. H., Thomas, R. G., Hunder, G. G., and Bloch, D. A. Overtraining in neural networks that interpret clinical data. Clin. Chem. *38:* 1998–2004, 1993.

14. Ravdin, P. M., and Clark, G. M. A practical application of neural network analysis for predicting outcome of individual breast cancer patients. Breast Cancer Res. Treat., *22:* 285–293, 1992.

15. Ravdin, P. M., Clark, G. M., Hilsenbeck, S. G., Owens, M. A., Vendely, P., Pandian, M. R., and McGuire, W. L. A demonstration that breast cancer recurrence can be predicted by neural network analysis. Breast Cancer Res. Treat., *21:* 47–53, 1992.

16. De Laurentiis, M., and Ravdin, P. M. A technique for using neural network analysis to perform survival analysis of censored data. Cancer Lett., *77:* 127–138, 1994.

17. De Laurentiis, M., and Ravdin, P. M. Survival analysis of censored data: neural network analysis detection of complex interactions between variables. Breast Cancer Res. Treat., *32:* 113–118, 1994.

18. Rumelhart, D. E., Hinton, G. E., and Williams, R. J. Learning representation by back propagating errors. Nature (Lond.), *323:* 533–536, 1986.

19. Van Houwelingen, J. C., and Le Cessie, S. Predictive value of statistical models. Stat. Med., *9:* 1303–1325, 1990.

20. Harrel, F. E., Califf, R. M., Pryor, D. B., Kerry, L. L., and Rosati, R. A. Evaluating the yield of medical tests. JAMA, *247:* 2543–2546, 1982.

21. Hanley, J. A., and McNeil, B. J. The meaning and use of the area under a receiver operating characteristic (ROC) curve. Radiology, *143:* 29–36, 1982.

22. Hanley, J. A., and McNeil, B. J. A method of comparing the areas under receiver operating characteristic curves derived from the same cases. Radiology, *148:* 839–843, 1983.

23. McNeil, B. J., and Hanley, J. A. Statistical approaches to the analysis of receiver operating characteristic (ROC) curves. Med. Decis. Making, *4:* 137–150, 1984.

24. Kaplan, E. L., and Meier, P. Nonparametric estimation from incomplete observation. J. Am. Stat. Assoc., *53:* 457–481, 1958.

25. Clark, G. M., Hilsenbeck, S. G., Ravdin, P. M., De Laurentiis, M., and Osborne, C. K. Prognostic factors: rationale and methods of analysis and integration. Breast Cancer Res. Treat., *32:* 105–112, 1994.

26. Ravdin, P. M., Clark, G. M., Hough, J. J., Owens, M. A., and McGuire, W. L. Neural network analysis of DNA flow cytometry histograms. Cytometry, *14:* 74–80, 1993.

27. Bottaci, L., Drew, P. J., Hartley, J. E., Hadfield, M. B., Farouk, R., Lee, P. W., Macintyre, I. M., Duthie, G. S., and Monson, J. R. Artificial neural networks applied to outcome prediction for colorectal cancer patients in separate institutions. Lancet, *350:* 469–472, 1997.

28. Dybowski, R., Weller, P., Chang, R., and Gant, V. Prediction of outcome in critically ill patients using artificial neural network synthesised by genetic algorithm. Lancet, *347:* 1146–1150, 1996.

29. Baxt, W. G., and Skora, J. Prospective validation of artificial neural network trained to identify acute myocardial infarction. Lancet, *347:* 12–15, 1996.

30. Cross, S. S., Harrison, R. F., and Kennedy, R. L. Introduction to neural networks. Lancet, *346:* 1075–1079, 1995.

31. Liestol, K., Andersen, P. K., and Andersen, U. Survival analysis and neural nets. Stat. Med., *13:* 1189–1200, 1994.

32. Faraggi, D., and Simon, R. A neural network model for survival data. Stat. Med., *14:* 73–82, 1995.

33. White, H. Learning in artificial neural networks: a statistical approach. Neural Comput., *1:* 425–464, 1989.

34. Ripley, B. D. Statistical aspects of neural networks. *In:* O. E. Barndorff-Nielsen, J. L. Jensen, and W. S. Kendall (eds.), Networks and Chaos: Statistical and Probabilistic Aspects. London: Chapman & Hall, 1993.

35. Menard, S., Bufalino, R., Rilke, F., Cascinelli, N., Veronesi, U., and Colnaghi, M. I. Prognosis based on primary breast carcinoma instead of pathological nodal status. Br. J. Cancer, *70:* 709–712, 1994.

36. Balslev, I., Axelsson, C. K., Zedeler, K., Rasmussen, B. B., Carstensen, B., and Mouridsen, H. T. The Nottingham Prognostic Index applied to 9,149 patients from the studies of the Danish Breast Cancer Cooperative Group (DBCG). Breast Cancer Res. Treat., *32:* 281–290, 1994.

37. Galea, M. H., Blamey, R. W., Elston, C. E., and Ellis, I. O. The Nottingham Prognostic Index in primary breast cancer. Breast Cancer Res. Treat., *22:* 207–219, 1992.