ORIGINAL ARTICLE

A Multigene Assay to Predict Recurrence of Tamoxifen-Treated, Node-Negative Breast Cancer

Soonmyung Paik, M.D., Steven Shak, M.D., Gong Tang, Ph.D., Chungyeul Kim, M.D., Joffre Baker, Ph.D., Maureen Cronin, Ph.D., Frederick L. Baehner, M.D., Michael G. Walker, Ph.D., Drew Watson, Ph.D., Taesung Park, Ph.D., William Hiller, H.T., Edwin R. Fisher, M.D., D. Lawrence Wickerham, M.D., John Bryant, Ph.D., and Norman Wolmark, M.D.

ABSTRACT

BACKGROUND

The likelihood of distant recurrence in patients with breast cancer who have no involved lymph nodes and estrogen-receptor–positive tumors is poorly defined by clinical and histopathological measures.

METHODS

We tested whether the results of a reverse-transcriptase–polymerase-chain-reaction (RT-PCR) assay of 21 prospectively selected genes in paraffin-embedded tumor tissue would correlate with the likelihood of distant recurrence in patients with node-negative, tamoxifen-treated breast cancer who were enrolled in the National Surgical Adjuvant Breast and Bowel Project clinical trial B-14. The levels of expression of 16 cancer-related genes and 5 reference genes were used in a prospectively defined algorithm to calculate a recurrence score and to determine a risk group (low, intermediate, or high) for each patient.

RESULTS

Adequate RT-PCR profiles were obtained in 668 of 675 tumor blocks. The proportions of patients categorized as having a low, intermediate, or high risk by the RT-PCR assay were 51, 22, and 27 percent, respectively. The Kaplan–Meier estimates of the rates of distant recurrence at 10 years in the low-risk, intermediate-risk, and high-risk groups were 6.8 percent (95 percent confidence interval, 4.0 to 9.6), 14.3 percent (95 percent confidence interval, 8.3 to 20.3), and 30.5 percent (95 percent confidence interval, 23.6 to 37.4). The rate in the low-risk group was significantly lower than that in the high-risk group ($P<0.001$). In a multivariate Cox model, the recurrence score provided significant predictive power that was independent of age and tumor size ($P<0.001$). The recurrence score was also predictive of overall survival ($P<0.001$) and could be used as a continuous function to predict distant recurrence in individual patients.

CONCLUSIONS

The recurrence score has been validated as quantifying the likelihood of distant recurrence in tamoxifen-treated patients with node-negative, estrogen-receptor–positive breast cancer.

From the Division of Pathology, Operation Center, and the Biostatistics Center, National Surgical Adjuvant Breast and Bowel Project, Pittsburgh (S.P., G.T., C.K., T.P., W.H., E.R.F., D.L.W., J.B., N.W.); Genomic Health, Redwood City, Calif. (S.S., J.B., M.C., M.G.W., D.W.); the Department of Statistics, University of Pittsburgh, Pittsburgh (G.T., J.B.); and the University of California, San Francisco, San Francisco (F.L.B.). Address reprint requests to Dr. Paik at the Division of Pathology, NSABP, 4 Allegheny Center, 5th Fl., East Commons Professional Bldg., Pittsburgh, PA 15212, or at spaik.nejm@nsabp.org.

N Engl J Med 2004;351:2817-26.

Over the past two decades, the molecular dissection of cancer has increased our understanding of the pathways that are altered in neoplastic cells.[1,2] Nevertheless, the diagnosis of cancer and decisions about its treatment still rely largely on classic histopathological and immunohistochemical techniques. A more quantitative approach to diagnosis and rational individualization of treatment are needed.

Large clinical trials, such as National Surgical Adjuvant Breast and Bowel Project (NSABP) trials B-14 and B-20, have demonstrated the benefit of tamoxifen and chemotherapy in women who have node-negative, estrogen-receptor–positive breast cancer.[3-5] However, since the likelihood of distant recurrence in patients treated with tamoxifen alone after surgery is about 15 percent at 10 years, at least 85 percent of patients would be overtreated with chemotherapy if it were offered to everyone. Numerous attempts have been made to identify biomarkers of residual risk,[6-9] but none of them have been recommended for guiding treatment.[10-15] Molecular signatures of gene expression in tumor tissue that correlate with recurrence of breast cancer have been identified by methods based on the use of DNA arrays.[16-21] However, the requirement for fresh or snap-frozen tissue and uncertainties about the reproducibility of such methods have limited their clinical application.

We used a multistep approach to develop an assay of the expression of tumor-related genes for use with routinely prepared tumor blocks and to validate the assay clinically. First, a high-throughput, real-time, reverse-transcriptase–polymerase-chain-reaction (RT-PCR) method was developed to quantify gene expression with the use of sections of fixed, paraffin-embedded tumor tissue.[22] Second, we selected 250 candidate genes from the published literature, genomic databases, and experiments based on DNA arrays performed on fresh-frozen tissue.[17-19,23] Third, we analyzed data from three independent clinical studies of breast cancer involving a total of 447 patients, including the tamoxifen-only group of NSABP trial B-20, to test the relation between expression of the 250 candidate genes and the recurrence of breast cancer.[24-26] Fourth, we used the results of the three studies to select a panel of 16 cancer-related genes and 5 reference genes and designed an algorithm, based on the levels of expression of these genes, to compute a recurrence score for each tumor sample. The study reported here was performed to validate the ability of the prospectively defined, 21-gene RT-PCR assay and recurrence-score algorithm to quantify the likelihood of distant recurrence in patients with node-negative, estrogen-receptor–positive breast cancer who had been treated with tamoxifen in the large, multicenter NSABP trial B-14.

METHODS

PATIENTS

NSABP trial B-14 (entitled "A Clinical Trial to Assess Tamoxifen in Patients with Primary Breast Cancer and Negative Axillary Nodes Whose Tumors Are Positive for Estrogen Receptors") enrolled 2892 patients who were randomly assigned to receive placebo or tamoxifen between January 4, 1982, and January 25, 1988, and enrolled 1235 additional patients, all treated with tamoxifen, between January 26, 1988, and October 17, 1988. The current study of the recurrence score was approved by the Essex Institutional Review Board (Lebanon, N.J.) and by the institutional review boards of Allegheny General Hospital and the University of Pittsburgh (both in Pittsburgh). The need for additional informed consent was waived by the institutional review boards.

SAMPLE PREPARATION

Paraffin blocks with cancer cells occupying less than 5 percent of the section area were excluded from the study. Macrodissection was performed with the use of a safety blade for cases involving nontumor elements that were amenable to macrodissection and that constituted more than 50 percent of the overall area of the tissue section. RNA was extracted from three 10-μm sections when macrodissection had not been performed or from six 10-μm sections when macrodissection had been performed.

ASSAY METHODS, GENE SELECTION, AND RECURRENCE-SCORE ALGORITHM

Gene expression in fixed, paraffin-embedded tumor tissue was measured as described by Cronin et al.[22] The Oncotype DX assay (Genomic Health) was used. In brief, after RNA extraction and DNase I treatment, total RNA content was measured and the absence of DNA contamination was verified (as described in the Supplementary Appendix, available with the full text of this article at www.nejm.org). Reverse transcription was performed and was followed by quantitative TaqMan RT-PCR reactions in 384-well plates, performed with the use of Prism

7900HT instruments (Applied Biosystems). The expression of each gene was measured in triplicate and then normalized relative to a set of five reference genes (*ACTB* [the gene encoding β-actin], *GAPDH*, *GUS*, *RPLPO*, and *TFRC*). Reference-normalized expression measurements ranged from 0 to 15, where a 1-unit increase reflected an approximate doubling of RNA.

The list of 21 genes and the recurrence-score algorithm (Fig. 1) were designed by analyzing the results of the three independent preliminary studies involving 447 patients and 250 candidate genes[24-26] (as described in the Supplementary Appendix). The selection of the final 16 cancer-related genes was based primarily on the strength of their performance in all three studies and the consistency of primer or probe performance in the assay. The range of possible recurrence scores was 0 to 100 (where higher scores indicated a greater likelihood of recurrence) and was derived from the reference-normalized expression measurements for the 16 cancer-related genes.

Cutoff points were prespecified to classify patients into the following categories: low risk (recurrence score, less than 18), intermediate risk (recurrence score, 18 or higher but less than 31), and high risk (recurrence score, 31 or higher). The cutoff points were chosen on the basis of the results of NSABP trial B-20.

Reproducibility within and between blocks was assessed by performing the 21-gene assay in five serial sections from six blocks in two patients. The within-block standard deviation for the recurrence score was 0.72 recurrence-score unit (95 percent confidence interval, 0.55 to 1.04). The total within-patient standard deviation (including between-block and within-block standard deviations) was 2.2 recurrence-score units.

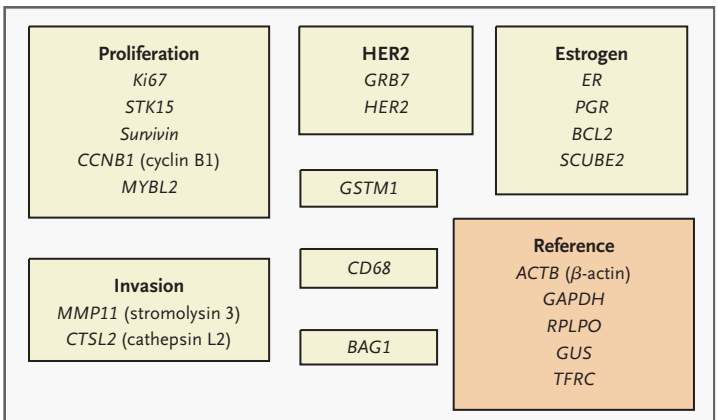


Figure 1. Panel of 21 Genes and the Recurrence-Score Algorithm.

The recurrence score on a scale from 0 to 100 is derived from the reference-normalized expression measurements in four steps. First, expression for each gene is normalized relative to the expression of the five reference genes (*ACTB* [the gene encoding β-actin], *GAPDH*, *GUS*, *RPLPO*, and *TFRC*). Reference-normalized expression measurements range from 0 to 15, with a 1-unit increase reflecting approximately a doubling of RNA. Genes are grouped on the basis of function, correlated expression, or both. Second, the *GRB7*, *ER*, proliferation, and invasion group scores are calculated from individual gene-expression measurements, as follows: *GRB7* group score = 0.9 × *GRB7*+0.1×*HER2* (if the result is less than 8, then the *GRB7* group score is considered 8); *ER* group score = (0.8×*ER*+1.2×*PGR*+*BCL2*+*SCUBE2*) ÷4; proliferation group score = (*Survivin*+*KI67*+*MYBL2*+*CCNB1* [the gene encoding cyclin B1]+*STK15*) ÷5 (if the result is less than 6.5, then the proliferation group score is considered 6.5); and invasion group score = (*CTSL2* [the gene encoding cathepsin L2] +*MMP11* [the gene encoding stromolysin 3]) ÷2. The unscaled recurrence score (RS_U) is calculated with the use of coefficients that are predefined on the basis of regression analysis of gene expression and recurrence in the three training studies[24-26]: RS_U=+0.47×*GRB7* group score−0.34×*ER* group score +1.04×proliferation group score+0.10×invasion group score+0.05×*CD68* −0.08×*GSTM1*−0.07×*BAG1*. A plus sign indicates that increased expression is associated with an increased risk of recurrence, and a minus sign indicates that increased expression is associated with a decreased risk of recurrence. Fourth, the recurrence score (RS) is rescaled from the unscaled recurrence score, as follows: RS=0 if RS_U<0; RS=20×(RS_U−6.7) if 0≤RS_U≤100; and RS=100 if RS_U>100.

STUDY DESIGN AND END POINTS

Patients were eligible if they had been randomly assigned to receive tamoxifen or had received tamoxifen as a member of the registration group of NSABP trial B-14 and if a tumor block was available in the NSABP Tissue Bank. Exclusion criteria were insufficient tumor tissue (less than 5 percent of the overall tissue sample) as assessed by histopathological analysis, insufficient RNA (less than 0.5 μg), or a weak RT-PCR signal (average cycle threshold for the reference genes, greater than 35).

The first prespecified primary objective was to determine whether the proportion of patients who were free of a distant recurrence for more than 10 years after surgery was significantly greater in the low-risk group than in the high-risk group. The second prespecified primary objective was to determine whether there was a statistically significant relation between the recurrence score and the risk of distant recurrence — one that went beyond the relation between recurrence and the standard measures of the patient's age and the size of the tumor. Contralateral disease, other second primary cancers, and death before distant recurrence were considered censoring events. Recurrence in the ipsilateral breast, local recurrence, and regional recurrence were not considered events or censoring events.

Prespecified secondary objectives included determination of the relapse-free interval (the time from surgery to any recurrence) over a 10-year period and the 10-year overall mortality from any cause in the low-risk and high-risk groups; the degree of agreement in the assignment of tumor grade among three pathologists; and the performance of the recurrence score in the context of the interobserver variability in tumor grading.

No samples from trial B-14 were used for prior testing or training. The prospectively defined assay methods and end points were finalized in a protocol signed on August 27, 2003. RT-PCR analysis was initiated on September 5, 2003, and RT-PCR data were transferred to the NSABP for analysis on September 29, 2003.

Estrogen- and progesterone-receptor proteins were measured by ligand-binding assays. *HER2* DNA was measured by a fluorescence in situ hybridization assay (PathVysion, Vysis). Tumor grade was determined independently by three pathologists from the NSABP, Stanford University Medical Center, and the University of California, San Francisco, School of Medicine with use of a modification of the Bloom–Richardson grading criteria.[27]

STATISTICAL ANALYSIS

We tested the hypothesis that the proportion of patients who are free of a distant recurrence at 10 years would be significantly higher in the low-risk group (recurrence score, less than 18) than in the high-risk group (recurrence score, 31 or higher). The test statistic was derived by adjusting the difference between the Kaplan–Meier estimates of the 10-year rate of distant recurrence in the two groups by the corresponding Greenwood variance estimates. A P value of less than 0.05 (two-sided) was considered to indicate a significant result. We also tested the hypothesis that there would be a significant difference between a (reduced) Cox proportional-hazards model for distant recurrence based only on age and clinical tumor size and a (full) proportional-hazards model based on age, clinical tumor size, and recurrence score. A P value of less than 0.05 (two-sided) in the likelihood-ratio test was considered to indicate a significant result. To define the continuous relation between the recurrence score and the 10-year risk of distant recurrence, the data were fitted by a time-varying, piecewise, log-hazard ratio model with the recurrence score and its quadratic term included as covariates.[28] The 10-year rate of distant recurrence was then estimated by a Breslow-type function.[29] The NSABP designed the study, collected the clinical data, and analyzed the results. The assay was carried out by Genomic Health. The NSABP held the combined clinical and laboratory data (after the removal of identifying information) and performed the data analyses. The manuscript was written by the NSABP, with input from Genomic Health.

RESULTS

CHARACTERISTICS OF THE PATIENTS

Paraffin blocks containing sufficient specimens of tissue involved by invasive breast cancer were available from 675 of 2617 tamoxifen-treated patients in trial B-14. RT-PCR was successful in 668 of the 675 blocks. The 668 patients who corresponded to these blocks were similar in terms of age distribution and the distribution of tumor size to the overall group of 2617 tamoxifen-treated patients (Table 1 of the Supplementary Appendix). For the group of 668 patients whose tumor sample could be evaluated, the Kaplan–Meier estimate for the proportion who had no distant recurrence 10 years after surgery was 85 percent.

RECURRENCE RATES IN THE LOW-RISK AND HIGH-RISK GROUPS

The Kaplan–Meier estimate for the proportion of patients in the low-risk group who were free of a distant recurrence at 10 years (93.2 percent) was significantly greater than the proportion in the high-risk category (69.5 percent) (P<0.001) (Table 1 and

Table 1. Kaplan–Meier Estimates of the Rate of Distant Recurrence at 10 Years, According to Recurrence-Score Risk Categories.*

Risk Category	Percentage of Patients	Rate of Distant Recurrence at 10 Yr (95% CI)†
		percent
Low	51	6.8 (4.0–9.6)
Intermediate	22	14.3 (8.3–20.3)
High	27	30.5 (23.6–37.4)‡

* A low risk was defined as a recurrence score of less than 18, an intermediate risk as a score of 18 or higher but less than 31, and a high risk as a score of 31 or higher.
† CI denotes confidence interval.
‡ P<0.001 for the comparison with the low-risk category.

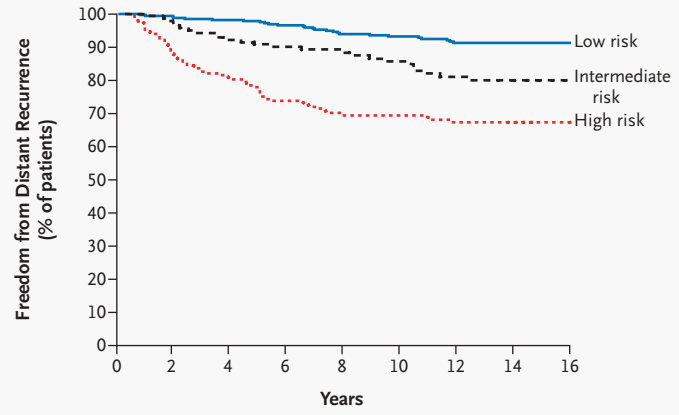
Fig. 2). The recurrence score was also significantly correlated with two secondary end points: the relapse-free interval and overall survival (P<0.001 for both) (Fig. 2B and 2C of the Supplementary Appendix).

RECURRENCE SCORE, AGE, TUMOR SIZE, AND RISK OF DISTANT RECURRENCE

As expected, younger patients (those less than 50 years of age) had higher rates of distant recurrence at 10 years than older patients (21.1 percent [95 percent confidence interval, 15.1 to 26.8 percent] vs. 12.3 percent [95 percent confidence interval, 9.1 to 15.3 percent]), whereas patients with smaller tumors (diameter, 2 cm or less) had lower estimated rates of distant recurrence at 10 years than those with larger tumors (13.3 percent [95 percent confidence interval, 9.9 to 16.8 percent] vs. 17.5 percent [95 percent confidence interval, 12.6 to 22.3 percent]). In a multivariate Cox model in which distant recurrence was evaluated in relation to both age and tumor size, age alone was significantly correlated with distant recurrence (P=0.004, with younger patients more likely to have recurrence), whereas tumor size trended toward significance (P=0.06, with larger tumors more likely to recur) (Table 2). In a multivariate Cox model in which distant recurrence was evaluated in relation to the recurrence score, age, and tumor size, the recurrence score provided significant predictive power that was independent of age and tumor size (P<0.001) (Table 2). When recurrence score was added to the model, age and tumor size were no longer statistically significant. Similar results were observed when more than two categories of age and tumor size were used in the model (data not shown).

ESTROGEN- AND PROGESTERONE-RECEPTOR PROTEINS AND AMPLIFICATION OF *HER2*

No relation was observed between the levels of estrogen- or progesterone-receptor proteins and the risk of distant recurrence (Fig. 1 of the Supplementary Appendix). *HER2* was amplified in 55 of the 668 tumors (8.2 percent) and not amplified in 605 tumors (90.6 percent); the result was indeterminate in 8 (1.2 percent). The Kaplan–Meier estimate of the proportion of patients free of distant recurrence at 10 years among those with tumors in which *HER2* was amplified was 75.0 percent (95 percent confidence interval, 63.2 to 86.9 percent), and 86.0 percent (95 percent confidence interval,



No. at Risk									
Low risk	338	328	313	298	276	258	231	170	38
Intermediate risk	149	139	128	116	104	96	80	66	16
High risk	181	154	137	119	105	91	83	63	13

Figure 2. Likelihood of Distant Recurrence, According to Recurrence-Score Categories.

A low risk was defined as a recurrence score of less than 18, an intermediate risk as a score of 18 or higher but less than 31, and a high risk as a score of 31 or higher. There were 28 recurrences in the low-risk group, 25 in the intermediate-risk group, and 56 in the high-risk group. The difference among the groups is significant (P<0.001).

Table 2. Multivariate Cox Proportional Analysis of Age, Tumor Size, and Recurrence Score in Relation to the Likelihood of Distant Recurrence.*

Variable	P Value	Hazard Ratio (95% CI)†
Analysis without recurrence score		
Age at surgery	0.004	0.57 (0.39–0.83)
Clinical tumor size	0.06	1.44 (0.99–2.11)
Analysis with recurrence score‡		
Age at surgery	0.08	0.71 (0.48–1.05)
Clinical tumor size	0.23	1.26 (0.86–1.86)
Recurrence score	<0.001	3.21 (2.23–4.61)

* Age at surgery was a binary variable (0 for an age of less than 50 years and 1 for an age of 50 years or more); clinical tumor size was a binary variable (0 for a diameter of 2 cm or less and 1 for a diameter greater than 2 cm); and the recurrence score was a continuous variable, with the hazard ratio for distant recurrence calculated relative to an increment of 50 units (chosen to dichotomize the recurrence score and thus improve comparability of the hazard ratio with the hazard ratios based on the clinical covariates).
† CI denotes confidence interval.
‡ P<0.001 and chi-square=33.7 for the comparison with the analysis without the recurrence score (by the likelihood-ratio test).

83.1 to 88.9 percent) among patients with tumors in which *HER2* was not amplified (P=0.08) (Fig. 2A of the Supplementary Appendix). In Cox models that included the recurrence score and traditional measures (estrogen receptor, progesterone receptor, or DNA amplification of *HER2*), only the recurrence score was a significant predictor of distant recurrence (data not shown).

Table 3. Multivariate Cox Proportional Analysis of Age, Tumor Size, Tumor Grade, and Recurrence Score in Relation to the Likelihood of Distant Recurrence.*

Variable	P Value	Hazard Ratio (95% CI)†
Analysis without recurrence score		
Age at surgery	0.10	0.70 (0.45–1.07)
Clinical tumor size	0.13	1.35 (0.92–1.98)
Tumor grade		
Moderately differentiated	0.04	1.87 (1.04–3.37)
Poorly differentiated	<0.001	5.14 (2.89–9.15)
HER2 amplification	0.89	1.04 (0.57–1.90)
Estrogen-receptor protein		
50–99 fmol/mg	0.23	0.71 (0.41–1.24)
100–199 fmol/mg	0.38	0.78 (0.45–1.35)
≥200 fmol/mg	0.90	0.97 (0.55–1.69)
Analysis with recurrence score‡		
Age at surgery	0.22	0.76 (0.50–1.18)
Clinical tumor size	0.38	1.19 (0.81–1.76)
Tumor grade		
Moderately differentiated	0.15	1.55 (0.85–2.81)
Poorly differentiated	<0.001	3.34 (1.79–6.26)
HER2 amplification	0.06	0.51 (0.26–1.02)
Estrogen-receptor protein		
50–99 fmol/mg	0.32	0.75 (0.43–1.31)
100–199 fmol/mg	0.72	0.90 (0.52–1.58)
≥200 fmol/mg	0.94	1.02 (0.58–1.70)
Recurrence score	<0.001	2.81 (1.70–4.64)

* The tumor grades were those of one of the three pathologists. Age at surgery was a binary variable (0 for an age of less than 50 years and 1 for an age of 50 years or more); clinical tumor size was a binary variable (0 for a diameter of 2 cm or less and 1 for a diameter greater than 2 cm); grade was a binary variable (poorly differentiated relative to well differentiated and moderately differentiated relative to well differentiated); *HER2* amplification was a binary variable (0 for no amplification on fluorescence in situ hybridization and 1 for amplification); the amount of estrogen-receptor protein was an ordinal variable, with the baseline level being 10 to 49 fmol per milligram; and recurrence score was a continuous variable, with the hazard ratio for distant recurrence calculated relative to an increment of 50 units.

† CI denotes confidence interval.

‡ P<0.001 and chi-square=15.2 for the comparison with the analysis without the recurrence score.

RECURRENCE SCORE, TUMOR GRADE, AND RISK OF DISTANT RECURRENCE

The assessment of tumor grade by each of the three pathologists correlated with the risk of distant recurrence (Tables 2A, 2B, and 2C of the Supplementary Appendix). The recurrence score provided significant information beyond tumor grade for each of the three pathologists (P<0.001). The concordance in assessment of grade between any two pathologists was 59 to 65 percent, and the overall concordance among all the three pathologists was 43 percent (Table 3 of the Supplementary Appendix). Agreement among the three pathologists was lowest for well-differentiated and moderately differentiated tumor grades (kappa, 0.36 and 0.23, respectively) and highest for a poorly differentiated grade (kappa, 0.61).

Finally, multivariate Cox proportional-hazards analyses were performed to explore the relation between distant recurrence and age, tumor size, tumor grade, *HER2* amplification, amounts of estrogen- and progesterone-receptor protein, and recurrence score (Table 3). The recurrence score and poor tumor grade were significant predictors of distant recurrence.

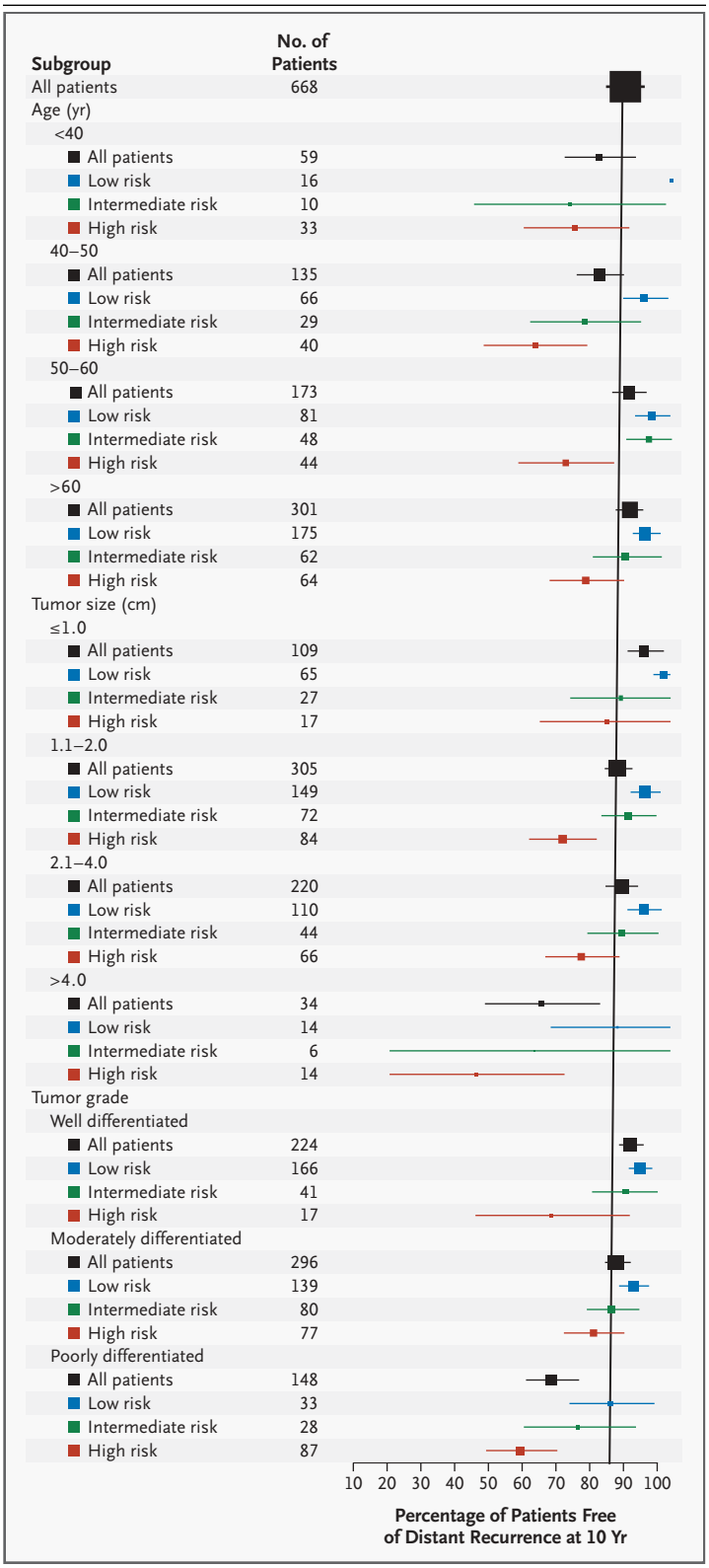
RISK OF DISTANT RECURRENCE IN SUBGROUPS OF PATIENTS

The recurrence score predicted distant recurrence for all age categories and all categories of tumor size (Fig. 3). Patients with a low-risk recurrence score (less than 18) had less frequent distant recurrences at 10 years than patients with a high-risk score (31 or higher). Moreover, not all patients with small tumors (109 patients with a tumor 1 cm in diameter or smaller) were at low risk; the recurrence score identified 44 of those patients as having an intermediate or high risk and a 15 to 20 percent risk of distant recurrence at 10 years.

The subgroup of patients with moderately differentiated tumors (the most common grade) could be distinguished to be at low or high risk by the recurrence score (Fig. 3). A subgroup of patients with well-differentiated tumors had high recurrence scores and high rates of distant recurrence. For two of the three pathologists, a subgroup of patients with poorly differentiated tumors had low recurrence scores and low rates of distant recurrence (Fig. 3A and 3B of the Supplementary Appendix).

Figure 3. Kaplan–Meier Estimates of the Proportion of Patients Free of Distant Recurrences at 10 Years, According to Age, Tumor Size, and Tumor Grade.

For each group of patients, the results for low-, intermediate-, and high-risk recurrence-score categories (scores of less than 18, 18 or higher but less than 31, and 31 or higher, respectively) are shown. The tumor grades are those of one of the three pathologists. The size of each square corresponds to the size of the subgroup; the horizontal lines represent the 95 percent confidence interval.



RECURRENCE SCORE AS A CONTINUOUS PREDICTOR OF DISTANT RECURRENCE

The likelihood of distant recurrence at 10 years increased continuously as the recurrence score increased (Fig. 4). Two-sided confidence intervals for the likelihood of distant recurrence are generally ±2 to 3 percent for recurrence scores of less than 30 and ±3 to 5 percent for recurrence scores of 30 to 50. For recurrence scores greater than 50, the likelihood of distant recurrence increases only slightly as the score increases. On average, patients with recurrence scores greater than 50 (12 percent of the 668 patients) had a risk of distant recurrence at 10 years of 33.8 percent (95 percent confidence interval, 23.4 to 44.2 percent).

DISCUSSION

Using a prospectively defined gene-expression assay and an algorithm for calculating recurrence scores, we were able to quantify the likelihood of distant recurrence in patients with node-negative, estrogen-receptor–positive breast cancer who had been treated with tamoxifen. The difference in the risk of distant recurrence between patients with low recurrence scores and those with high recurrence scores was large and statistically significant. Many patients (51 percent of the patients in the study) were categorized as having a low risk, and their rate of distant recurrence at 10 years was 6.8 percent. A smaller group of patients (27 percent) was categorized as having a high risk; their rate of distant recurrence at 10 years was 30.5 percent — a risk similar to that observed among patients with node-positive disease.[30] The use of the recurrence score as a continuous predictor provides an accurate estimate of the risk of distant recurrence in individual patients.

The recurrence score can also predict overall sur-

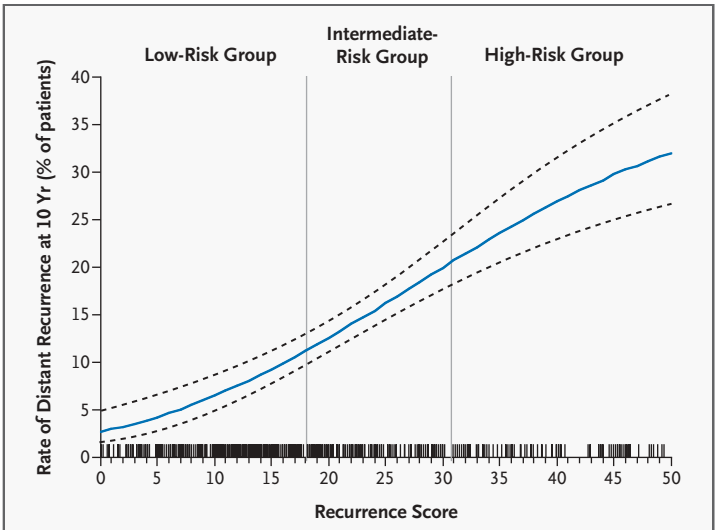


Figure 4. Rate of Distant Recurrence as a Continuous Function of the Recurrence Score.

The continuous function was generated with use of a piecewise log-hazard-ratio model.[28] The dashed curves indicate the 95 percent confidence interval. The rug plot on top of the x axis shows the recurrence score for individual patients in the study.

vival. This feature is notable, since approximately 50 percent of the deaths occurred in the absence of recurrent breast cancer. In addition, the recurrence score predicts the relapse-free interval (including the interval free of local and regional recurrences). Thus, the recurrence score correlates in a statistically significant manner with all the end points we examined.

The patient's age and the size of the tumor are routinely used as predictors of recurrence in breast cancer and are incorporated into current treatment guidelines.[13-15] When the recurrence score was combined with data pertaining to age and tumor size to predict the risk of distant recurrence, only the recurrence score remained statistically significant in a multivariate analysis. It is likely that the decreased risk of recurrence in older patients is not related to age itself but instead, at least in part, to the higher amount of estrogen-receptor protein in older patients' tumors.[31,32] The contribution of *ER* expression to the recurrence score captures this factor.

The subgroup analysis of patients according to age and tumor size was exploratory, and the results should be interpreted cautiously. Nevertheless, the recurrence score was a consistent predictor of distant recurrence in patients of all age categories and all tumor-size categories. For example, more than a third of the patients with small tumors (1 cm in diameter or smaller) had intermediate-risk or high-risk recurrence scores and a 15 to 20 percent risk of distant recurrence.

We evaluated the recurrence score in the context of the interobserver variability in tumor grading that is typical in oncology practice. Tumor grade correlates with the likelihood of recurrence when analyzed in large populations of patients. However, previous studies have also documented that the grading of breast cancer entails a degree of subjective judgment, leading to low concordance among pathologists. Robbins et al.[33] compared the interobserver reproducibility in their study to the published results of four other groups.[34-36] Complete agreement in those five studies ranged from 54 percent to 83 percent (kappa, 0.17 to 0.73). We found that the concordance among pathologists for the poorly differentiated grade is moderate (kappa, 0.61) and for the well-differentiated and moderately differentiated grades is low (kappa, 0.23 and 0.36, respectively). Recently, a Breast Task Force serving the American Joint Committee on Cancer did not add tumor grade to its staging criteria because of the sparseness and variability of the data.[37]

Traditional measures of estrogen-receptor protein (by ligand-binding assay) and *HER2* (by fluorescence in situ hybridization) in this study were only weakly predictive of the risk of distant recurrence. The quantitative information that the RT-PCR assay provides for *ER*, *HER2*, and the other 14 cancer-related genes is clearly important.

It is important to emphasize that we do not know whether the genes used in the calculation of the recurrence score correlate with recurrence in the population we studied because they show a relation with the natural history of breast cancer, because they predict responsiveness to tamoxifen, or both. Esteva et al.[38] found no correlation between the recurrence score and the rate of distant recurrence in 149 selected patients with node-negative breast cancer who did not receive adjuvant systemic therapy. However, in that cohort, patients with well-differentiated tumors (i.e., those with a low nuclear grade) had a surprisingly worse survival rate than patients with moderately differentiated or poorly differentiated tumors. The current data cannot be used to select women for tamoxifen therapy.

Few assays have been rigorously validated for use as prognostic or predictive tests in oncology. We conducted a prospectively designed validation

study of a multigene-expression assay in a large, multicenter clinical trial. It is of practical importance that this assay involves the use of very small amounts of the tumor tissue that is routinely prepared after surgery.

Supported by the National Surgical Adjuvant Breast and Bowel Project and Genomic Health. Genomic Health paid the costs of shipping the paraffin-embedded tissue sections and performing all RT-PCR assays.

Drs. Paik, Shak, Baker, Cronin, Walker, and Bryant report holding a patent for the RT-PCR assay used in this study. Drs. Shak, Baker, Cronin, and Watson report holding equity ownership or stock options in Genomic Health and being employed by Genomic Health, the commercial entity that sponsored the study. Dr. Walker reports having received consulting fees from Genomic Health and owning stock options. Dr. Baehner reports having received consulting fees from Genomic Health; Dr. Paik, lecture fees from Genomic Health; and Dr. Wickerham, consulting fees from AstraZeneca.

We are indebted to Tracy George (Stanford University); to Terry Mamounas (NSABP) for his comments; to Randy Scott, Debjani Dutta, Daniel Klaus, Mylan Pho, Anhthu Nguyen, Jennie Jeong, Stephanie Butler, Joel Robertson, Ken Stineman, Marti Haskins, and Claire Alexander (all of Genomic Health); and to Clifford Hudis, Tom Fleming, David Botstein, David Agus, and Fred Cohen for their helpful advice and suggestions.

REFERENCES

1. Buetow KH, Klausner RD, Fine H, Kaplan R, Singer DS, Strausberg RL. Cancer Molecular Analysis Project: weaving a rich cancer research tapestry. Cancer Cell 2002; 1:315-8.
2. Hahn WC, Weinberg RA. Rules for making human tumor cells. N Engl J Med 2002; 347:1593-603. [Erratum, N Engl J Med 2003; 348:674.]
3. Fisher B, Costantino J, Redmond C, et al. A randomized clinical trial evaluating tamoxifen in the treatment of patients with node-negative breast cancer who have estrogen-receptor-positive tumors. N Engl J Med 1989;320:479-84.
4. Fisher B, Jeong JH, Bryant J, et al. Treatment of lymph-node-negative, oestrogen-receptor-positive breast cancer: long-term findings from National Surgical Adjuvant Breast and Bowel Project randomised clinical trials. Lancet 2004;364:858-68.
5. Fisher B, Dignam J, Wolmark N, et al. Tamoxifen and chemotherapy for lymph node-negative, estrogen receptor-positive breast cancer. J Natl Cancer Inst 1997;89: 1673-82.
6. Bryant J, Fisher B, Gunduz N, Costantino JP, Emir B. S-phase fraction combined with other patient and tumor characteristics for the prognosis of node-negative, estrogen-receptor-positive breast cancer. Breast Cancer Res Treat 1998;51:239-53.
7. Henderson IC, Patek AJ. The relationship between prognostic and predictive factors in the management of breast cancer. Breast Cancer Res Treat 1998;52:261-88.
8. Hayes DF. Do we need prognostic factors in nodal-negative breast cancer? Arbiter. Eur J Cancer 2000;36:302-6.
9. Esteva FJ, Hortobagyi GN. Prognostic molecular markers in early breast cancer. Breast Cancer Res 2004;6:109-18.
10. Fitzgibbons PL, Page DL, Weaver D, et al. Prognostic factors in breast cancer: College of American Pathologists Consensus Statement 1999. Arch Pathol Lab Med 2000;124: 966-78.
11. Clinical practice guidelines for the use of tumor markers in breast and colorectal cancer: adopted on May 17, 1996 by the American Society of Clinical Oncology. J Clin Oncol 1996;14:2843-77.
12. Bast RC Jr, Ravdin P, Hayes DF, et al. 2000 Update of recommendations for the use of tumor markers in breast and colorectal cancer: clinical practice guidelines of the American Society of Clinical Oncology. J Clin Oncol 2001;19:1865-78. [Errata, J Clin Oncol 2001;19:4185-8, 2002;20:2213.]
13. Goldhirsch A, Glick JH, Gelber RD, Coates AS, Senn HJ. Meeting highlights: International Consensus Panel on the Treatment of Primary Breast Cancer: Seventh International Conference on Adjuvant Therapy of Primary Breast Cancer. J Clin Oncol 2001;19:3817-27.
14. Eifel P, Axelson JA, Costa J, et al. National Institutes of Health Consensus Development Conference Statement: adjuvant therapy for breast cancer, November 1-3, 2000. J Natl Cancer Inst 2001;93:979-89.
15. Carlson RW, Edge SB, Theriault RL. NCCN: breast cancer. Cancer Control 2001; 8:Suppl 2:54-61.
16. Davis RE, Staudt LM. Molecular diagnosis of lymphoid malignancies by gene expression profiling. Curr Opin Hematol 2002;9:333-8.
17. Perou CM, Sorlie T, Eisen MB, et al. Molecular portraits of human breast tumours. Nature 2000;406:747-52.
18. Golub TR, Slonim DK, Tamayo P, et al. Molecular classification of cancer: class discovery and class prediction by gene expression monitoring. Science 1999;286:531-7.
19. van 't Veer LJ, Dai H, van de Vijver MJ, et al. Gene expression profiling predicts clinical outcome of breast cancer. Nature 2002; 415:530-6.
20. van de Vijver MJ, He YD, van 't Veer LJ, et al. A gene-expression signature as a predictor of survival in breast cancer. N Engl J Med 2002;347:1999-2009.
21. Schena M, Shalon D, Davis RW, Brown PO. Quantitative monitoring of gene expression patterns with a complementary DNA microarray. Science 1995;270:467-70.
22. Cronin M, Pho M, Dutta D, et al. Measurement of gene expression in archival paraffin-embedded tissues: development and performance of a 92-gene reverse transcriptase-polymerase chain reaction assay. Am J Pathol 2004;164:35-42.
23. Sorlie T, Perou CM, Tibshirani R, et al. Gene expression patterns of breast carcinomas distinguish tumor subclasses with clinical implications. Proc Natl Acad Sci U S A 2001;98:10869-74.
24. Esteban J, Baker J, Cronin M, et al. Tumor gene expression and prognosis in breast cancer: multi-gene RT-PCR assay of paraffin-embedded tissue. Prog Proc Am Soc Clin Oncol 2003;22:850. abstract.
25. Cobleigh MA, Bitterman P, Baker J, et al. Tumor gene expression predicts distant disease-free survival (DDFS) in breast cancer patients with 10 or more positive nodes: high throughout RT-PCR assay of paraffin-embedded tumor tissues. Prog Proc Am Soc Clin Oncol 2003;22:850. abstract.
26. Paik S, Shak S, Tang G, et al. Multi-gene RT-PCR assay for predicting recurrence in node negative breast cancer patients — NSABP studies B-20 and B-14. Breast Cancer Res Treat 2003;82:A16. abstract.
27. Elston CW, Ellis IO. Pathological prognostic factors in breast cancer. I. The value of histological grade in breast cancer: experience from a large study with long-term follow-up. Histopathology 1991;19:403-10.
28. Gray RJ. Flexible methods for analyzing survival data using splines, with applications to breast cancer prognosis. J Am Stat Assoc 1992;87:942-51.
29. Valenta Z, Weissfeld L. Estimation of the survival function for Gray's piecewise-constant time-varying coefficients model. Stat Med 2002;21:717-27.
30. Fisher B, Redmond C, Legault-Poisson S, et al. Postoperative chemotherapy and tamoxifen compared with tamoxifen alone in the treatment of positive-node breast cancer patients aged 50 years and older with tumors responsive to tamoxifen: results from the National Surgical Adjuvant Breast and Bowel Project B-16. J Clin Oncol 1990;8: 1005-18.
31. Fisher B, Wickerham DL, Brown A, Redmond CK. Breast cancer estrogen and progesterone receptor values: their distribution, degree of concordance, and relation to number of positive axillary nodes. J Clin Oncol 1983;1:349-58.
32. Anderson WF, Chatterjee N, Ershler WB, Brawley OW. Estrogen receptor breast cancer phenotypes in the Surveillance, Epi-

demiology, and End Results database. Breast Cancer Res Treat 2002;76:27-36.
33. Robbins P, Pinder S, de Klerk N, et al. Histological grading of breast carcinomas: a study of interobserver agreement. Hum Pathol 1995;26:873-9.
34. Hopton DS, Thorogood J, Clayden AD, MacKinnon D. Observer variation in histological grading of breast cancer. Eur J Surg Oncol 1989;15:21-3.
35. Davis BW, Gelber RD, Goldhirsch A, et al. Prognostic significance of tumor grade in clinical trials of adjuvant therapy for breast cancer with axillary lymph node metastasis. Cancer 1986;58:2662-70.
36. Theissig F, Kunze KD, Haroske G, Meyer W. Histological grading of breast cancer: interobserver, reproducibility and prognostic significance. Pathol Res Pract 1990;186:732-6.
37. Singletary SE, Allred C, Ashley P, et al. Revision of the American Joint Committee on Cancer staging system for breast cancer. J Clin Oncol 2002;20:3628-36.
38. Esteva FJ, Sahin AA, Coombes K, et al. Multi-gene RT-PCR assay for predicting recurrence in node negative breast cancer patients — MD Anderson Clinical Validation Study. Breast Cancer Res Treat 2003;82:A16. abstract.