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.

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.
OVER THE PAST TWO DECADES, THE MOLECULAR DISSECTION OF CANCER HAS INCREASED OUR UNDERSTANDING OF THE PATHWAYS THAT ARE ALTERED IN NEOPLASTIC CELLS. 

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. 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, but none of them have been recommended for guiding treatment. 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. 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. Second, we selected 250 candidate genes from the published literature, genomic databases, and experiments based on DNA arrays performed on fresh-frozen tissue. 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. Fourth, we used the results of the three studies to select a panel of 16 cancer-related genes and five 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 prospective-
Involving 447 patients and 250 candidate genes, the results of the three independent preliminary studies were designed by analyzing the algorithm (Fig. 1) were derived from the 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 (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.

**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.

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 (RSU) 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): RSU = +0.47 × ER group score + 0.34 × GRB7 group score + 1.04 × proliferation group score + 0.10 × invasion group score + 0.05 × CD68 + 0.07 × GSTM1. 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 RSU < 0; RS = 20 × (RSU − 6.7) if 0 ≤ RSU ≤ 100; and RS = 100 if RSU > 100.

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.

**Multigene Assay to Predict Recurrence of Breast Cancer**

**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 (RSU) 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): RSU = +0.47 × ER group score + 0.34 × GRB7 group score + 1.04 × proliferation group score + 0.10 × invasion group score + 0.05 × CD68 + 0.07 × GSTM1. 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 RSU < 0; RS = 20 × (RSU − 6.7) if 0 ≤ RSU ≤ 100; and RS = 100 if RSU > 100.
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

**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

<p>| Table 1. Kaplan–Meier Estimates of the Rate of Distant Recurrence at 10 Years, According to Recurrence-Score Risk Categories.* |
|---|---|---|</p>
<table>
<thead>
<tr>
<th>Risk Category</th>
<th>Percentage of Patients</th>
<th>Rate of Distant Recurrence at 10 Yr (95% CI)†‡</th>
</tr>
</thead>
<tbody>
<tr>
<td>Low</td>
<td>51</td>
<td>6.8 (4.0–9.6)</td>
</tr>
<tr>
<td>Intermediate</td>
<td>22</td>
<td>14.3 (8.3–20.3)</td>
</tr>
<tr>
<td>High</td>
<td>27</td>
<td>30.5 (23.6–37.4)</td>
</tr>
</tbody>
</table>

* 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.
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).

**RECURRENT 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, 74.6 to 94.4 percent) among those with tumors in which HER2 was not amplified. 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). 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).

**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 differences 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.**

<table>
<thead>
<tr>
<th>Variable</th>
<th>P Value</th>
<th>Hazard Ratio (95% CI)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Analysis without score</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Age at surgery</td>
<td>0.004</td>
<td>0.57 (0.39–0.83)</td>
</tr>
<tr>
<td>Clinical tumor size</td>
<td>0.06</td>
<td>1.44 (0.99–2.11)</td>
</tr>
<tr>
<td>Analysis with score</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Age at surgery</td>
<td>0.08</td>
<td>0.71 (0.48–1.05)</td>
</tr>
<tr>
<td>Clinical tumor size</td>
<td>0.23</td>
<td>1.26 (0.86–1.86)</td>
</tr>
<tr>
<td>Recurrence score</td>
<td>&lt;0.001</td>
<td>3.21 (2.23–4.61)</td>
</tr>
</tbody>
</table>

* 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).
The New England Journal of Medicine

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).

**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).

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

<table>
<thead>
<tr>
<th>Variable</th>
<th>P Value</th>
<th>Hazard Ratio (95% CI)†</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Analysis without recurrence score</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Age at surgery</td>
<td>0.10</td>
<td>0.70 (0.45–1.07)</td>
</tr>
<tr>
<td>Clinical tumor size</td>
<td>0.13</td>
<td>1.35 (0.92–1.98)</td>
</tr>
<tr>
<td>Tumor grade</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Moderately differentiated</td>
<td>0.04</td>
<td>1.28 (1.04–3.37)</td>
</tr>
<tr>
<td>Poorly differentiated</td>
<td>&lt;0.001</td>
<td>5.14 (2.89–9.15)</td>
</tr>
<tr>
<td>HER2 amplification</td>
<td>0.89</td>
<td>1.04 (0.57–1.90)</td>
</tr>
<tr>
<td>Estrogen-receptor protein</td>
<td></td>
<td></td>
</tr>
<tr>
<td>50–99 fmol/mg</td>
<td>0.23</td>
<td>0.71 (0.41–1.24)</td>
</tr>
<tr>
<td>100–199 fmol/mg</td>
<td>0.38</td>
<td>0.78 (0.45–1.35)</td>
</tr>
<tr>
<td>≥200 fmol/mg</td>
<td>0.90</td>
<td>0.97 (0.55–1.69)</td>
</tr>
<tr>
<td><strong>Analysis with recurrence score</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Age at surgery</td>
<td>0.22</td>
<td>0.76 (0.50–1.18)</td>
</tr>
<tr>
<td>Clinical tumor size</td>
<td>0.38</td>
<td>1.19 (0.81–1.76)</td>
</tr>
<tr>
<td>Tumor grade</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Moderately differentiated</td>
<td>0.15</td>
<td>1.55 (0.85–2.81)</td>
</tr>
<tr>
<td>Poorly differentiated</td>
<td>&lt;0.001</td>
<td>3.34 (1.79–6.26)</td>
</tr>
<tr>
<td>HER2 amplification</td>
<td>0.06</td>
<td>0.51 (0.26–1.02)</td>
</tr>
<tr>
<td>Estrogen-receptor protein</td>
<td></td>
<td></td>
</tr>
<tr>
<td>50–99 fmol/mg</td>
<td>0.32</td>
<td>0.75 (0.43–1.31)</td>
</tr>
<tr>
<td>100–199 fmol/mg</td>
<td>0.72</td>
<td>0.90 (0.52–1.58)</td>
</tr>
<tr>
<td>≥200 fmol/mg</td>
<td>0.94</td>
<td>1.02 (0.68–1.57)</td>
</tr>
<tr>
<td>Recurrence score</td>
<td>&lt;0.001</td>
<td>2.81 (1.70–4.64)</td>
</tr>
</tbody>
</table>

* 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.
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).

The use of the recurrence score as a continuous predictor provides an accurate estimate of the risk of distant recurrence in individual patients.
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.\textsuperscript{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.\textsuperscript{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.\textsuperscript{33} compared the interobserver reproducibility in their study to the published results of four other groups.\textsuperscript{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.\textsuperscript{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.\textsuperscript{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 Stinemann, 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


32. Anderson WF, Chatterjee N, Ershler WB, Brawley OW. Estrogen receptor breast cancer phenotypes in the Surveillance, Epi-

Copyright © 2004 Massachusetts Medical Society.