Outcome Prediction for Estrogen Receptor–Positive Breast Cancer Based on Postneoadjuvant Endocrine Therapy Tumor Characteristics


Background
Understanding how tumor response is related to relapse risk would help clinicians make decisions about additional treatment options for patients who have received neoadjuvant endocrine treatment for estrogen receptor–positive (ER+) breast cancer.

Methods
Tumors from 228 postmenopausal women with confirmed ER+ stage 2 and 3 breast cancers in the P024 neoadjuvant endocrine therapy trial, which compared letrozole and tamoxifen for 4 months before surgery, were analyzed for posttreatment ER status, Ki67 proliferation index, histological grade, pathological tumor size, node status, and treatment response. Cox proportional hazards were used to identify factors associated with relapse-free survival (RFS) and breast cancer–specific survival (BCSS) in 158 women. A preoperative endocrine prognostic index (PEPI) for RFS was developed from these data and validated in an independent study of 203 postmenopausal women in the IMPACT trial, which compared treatment with anastrozole, tamoxifen, or the combination 3 months before surgery. Statistical tests were two-sided.

Results
Median follow-up in P024 was 61.2 months. Patients with confirmed baseline ER+ clinical stage 2 and 3 tumors that were downstaged to stage 1 or 0 at surgery had 100% RFS (compared with higher stages, $P < .001$). Multivariable testing of posttreatment tumor characteristics revealed that pathological tumor size, node status, Ki67 level, and ER status were independently associated with both RFS and BCSS. The PEPI model based on these factors predicted RFS in the IMPACT trial ($P = .002$).

Conclusions
Breast cancer patients with pathological stage 1 or 0 disease after neoadjuvant endocrine therapy and a low-risk biomarker profile in the surgical specimen (PEPI score 0) have an extremely low risk of relapse and are therefore unlikely to benefit from adjuvant chemotherapy.

J Natl Cancer Inst 2008;100:1380–1388

An accurate test to predict the effectiveness of adjuvant endocrine therapy for hormone receptor–positive breast cancer on an individual basis would be an important advance (1). Current approaches focus on biomarker analysis of the diagnostic specimen. An alternative is to treat patients with an endocrine agent for several months before surgery to identify tumors that are responsive to treatment, with the assumption that responsiveness indicates a lower risk of relapse. However, compared with neoadjuvant chemotherapy studies (2), fewer neoadjuvant endocrine therapy trials have been conducted; thus, fewer data are available to link postneoadjuvant therapy tumor characteristics and survival.

The P024 neoadjuvant endocrine therapy trial, which compared 4 months of letrozole and tamoxifen before surgery (3,4), now has sufficient follow-up (median >60 months) to address the relationships between postneoadjuvant endocrine therapy tumor characteristics and risk of early relapse. In this study, we used data from P024 to examine pathological stage posttreatment, histological grade posttreatment, response to treatment, and the biomarker status of the surgical specimen to develop a prognostic model that incorporates standard pathological staging variables and “on-treatment” biomarker values. We validated the model internally though bootstrap analysis and subsequently validated it externally using data from an independent neoadjuvant
endocrine therapy study that compared anastrozole, tamoxifen, or the combination for 3 months before surgery (the IMPACT trial) (5,6).

Subjects and Methods

Study Population and Tumor Bank

Both clinical trials described in this manuscript were approved by the local institutional review boards that enrolled patients into the studies (3,5). The P024 protocol compared 4 months of neoadjuvant letrozole therapy with 4 months of neoadjuvant tamoxifen therapy in postmenopausal women with clinical stage 2 and 3 hormone receptor–positive breast cancers (classified as at least 10% nuclear staining for estrogen receptor [ER] and/or progesterone receptor [PgR]) who were ineligible for breast conservative surgery (3). The clinical findings, tumor bank characteristics, and biomarker measurements have been described previously (3,4,7). The cut point for ER positivity for central laboratory analysis was an Allred score of 3 (8). Information on tumor grade, clinical response by caliper measurements, definitive pathological staging at surgery, and chemotherapy administration was collected prospectively. Patients in P024 were recommended to receive adjuvant tamoxifen for 5 years. The IMPACT study design, short- and long-term outcomes, and biomarker methodology have also been described previously (5,6,9). For the validation analysis, we compiled information on surgical stage, surgical specimen Ki67 proliferation antigen levels, ER data, duration of follow-up, and relapse dates. The IMPACT study used the H-score (10) to assess ER status. We converted the H-score ER cutoff to an Allred score ER cutoff for the analyses. An Allred score of 2 can be derived in only one way, i.e., less than 1% of cells staining weakly, which equates to an H-score of less than 1. Thus, it is valid to use an H-score of at least 1 as the equivalent of an Allred score of at least 3 as the threshold for ER positivity.

Statistical Analysis

Relapse-free survival (RFS) was defined as the interval between random assignment to treatment and the earliest subsequent breast cancer event. No new breast primary tumors were documented in either dataset, effectively excluding this class of event from the risk model that was developed. Breast cancer–specific survival (BCSS) was defined as the interval between random assignment and the date of death after breast cancer relapse. When we included all deaths, irrespective of cause, as RFS and overall survival events, there were no major changes in the P024 results (data not shown). Survival curves were estimated by the Kaplan–Meier method, and a two-sided \( P \) value of .05 from a log-rank test was considered a statistically significant difference. A multivariable Cox proportional hazard regression model was used to evaluate the independent prognostic relevance of each factor, namely, pathological tumor size (T1/2 vs T3/4 or T1 vs T2–4), pathological node status (negative vs positive), clinical response (complete plus partial clinical response vs stable disease plus progressive disease), surgical specimen ER status (Allred score \( \geq 3 \) vs 0 or 2), histological grade (grade 1 vs grade 2/3), and the Ki67 level, as natural log-transformed intervals (9), in both the pretreatment specimen and the surgical specimen. \( P \) values from Wald chi-square tests indicated the statistical significance of a factor after adjusting for the other factors. The proportionality assumption was checked by testing time-dependent covariates in the model and graphically examining the Schoenfeld residuals. Hazard ratio (HR) estimates of each factor in the final multivariable model were used to construct a risk index, the preoperative endocrine prognostic index (PEPI), for risk of relapse and breast cancer–specific death for each sample in the test cohort. The PEPI score was derived as an arithmetic sum of risk points weighted by the size of the HR assigned to each statistically significant factor (11). The overall discriminatory capacity of the Cox model on the P024 data was assessed using Harrell’s C index, which was then adjusted after a 1000 bootstrap assessment of the optimism portion. The 95% confidence interval (CI) for the adjusted C index was also reported (12). For the IMPACT trial validation analysis, we calculated the PEPI score for all patients who had data for the four factors and examined the association of PEPI score with RFS. SAS version 9.02 (SAS Institute, Cary, NC) and R 2.6 software (R Foundation for Statistical Computing) were used for all analyses.

Results

Definition of the Patient Populations for Analysis

Detailed information on the P024 population used in each of the analyses was compiled and is summarized in Figure 1. The median follow-up was 61.2 months for the patients who underwent surgery at the end of 4 months of endocrine treatment (\( n = 290 \)) and 62.0 months for the patients with a complete dataset for the multivariable analysis (\( n = 158 \)). Assignment to neoadjuvant letrozole
or tamoxifen had no impact on RFS or BCSS, but because all patients received adjuvant tamoxifen a difference was not anticipated. Thus, for the purposes of these analyses, the results from both arms were pooled together. As described in the original report on central laboratory ER testing (4), 12% of patients had ER− tumors at baseline, and of the 28 ER− and PgR− tumors, only one was reported to respond to treatment. Long-term follow-up confirmed that these tumors were likely to have been correctly assigned by the central laboratory because baseline ER− status was associated with poor overall survival ($P = .004$, data not shown). Subsequent analysis therefore focused on patients with ER+ tumors as assigned by the central laboratory ($n = 228$).

**Pathological Stage, Tumor Grade, and Response on Outcome**

All patients had clinical stage 2 or 3 tumor at diagnosis. To examine the impact of pathological (p) stage on outcome, patients whose tumors were downstaged at surgery to pT1N0 (node negative with a tumor size of ≤2 cm diameter with negative axillary lymph nodes—stage 1) or pT0N0 (pathological complete response—stage 0) were compared with patients with p-stage 2 or 3 disease at surgery. No relapses and only one death without known relapse (which was censored) were recorded in the group of 29 patients with p-stage 1 disease plus one patient with p-stage 0 disease vs 53 of the 175 patients (30% relapse incidence) with p-stage 2 and 3 disease (Figure 2, A, $P < .001$). The p-stage 1 or 0 status appeared to reflect downstaging of a group of endocrine therapy–responsive tumors because the clinical response rate was 79% in the p-stage 1 or 0 group vs 52% in the p-stage 2 or 3 group (response rate difference = 27%, 95% CI = 10.5% to 42.5%; $P = .006$). Furthermore, tumors in the p-stage 1 or 0 group had lower posttreatment geometric mean Ki67 levels than higher p-stage tumors (0.39 vs 0.98, ratio of geometric means = 0.40, 95% CI = 0.18 to 0.89; $P = .03$). The average baseline tumor size for the pT1 or pT0N0 tumors was predictably smaller than that of tumors with a higher pT stage at surgery; 3.9 vs 5.2 cm (longest diameter) by clinical caliper measurement ($P < .001$), 2.8 vs 4.0 by mammogram ($P < .001$), and 2.6 vs 3.4 cm by ultrasound ($P = .003$). However, given that these average measurements are all larger than 2 cm, it is unlikely that many tumors in the P024 study...
were p-stage 1 at diagnosis and therefore likely that p-stage 1 status usually reflects the effects of tumor regression induced by treatment. The excellent outcome in the p-stage 1 or 0 group was not strongly influenced by chemotherapy because only 2 of the 30 (7%) patients with p-stage 1 or 0 disease underwent adjuvant chemotherapy vs 61 of 175 (35%) patients with p-stage 2 or 3 disease ($P = .001$). Other factors that were associated with RFS in univariate analysis included posttreatment pathological node status ($P < .001$; Figure 2, B), clinical response to treatment ($P = .002$; Figure 2, C), pretreatment grade ($P = .002$) (data not shown), and posttreatment grade (grade I vs grade II or III, $P < .001$; Figure 2, D). The alternative grade dichotomy of grade I or II vs III did not have better prognostic characteristics (data not shown).

The Association of Posttreatment ER Status with Outcome
We had previously noted that a number of tumor samples that were ER+ before treatment had lost ER expression in the posttreatment sample [ie, converted to an Allred score of 0 or 2, or unequivocally...
ER (4)). We therefore compared RFS and BCSS between patients with tumors that converted from ER+ to ER− and patients with tumors that were persistently ER+ at surgery. The 16 patients with posttreatment ER− tumors had worse RFS (HR of relapse = 2.4, 95% CI = 1.0 to 5.3; \( P = .03 \)) and BCSS (HR of breast cancer death = 4.3, 95% CI = 1.6 to 11.7; \( P = .002 \) (Figure 2, E and F)) than patients with tumors that retained ER+ after treatment. The pretreatment Allred scores of tumors that were ER− after treatment was similar to that of tumors that retained ER expression (\( P = .2 \)).

The Association of Pre- and Posttreatment Ki67 Proliferation Index with Outcome

Ki67 is a commonly used proliferation antigen that identifies cells in the G1/S and M phases of the cell cycle (13). Ki67 was examined as a continuous variable after natural log transformation, as recommended by Dowsett et al. (9). This analysis approach examines the HR per 2.7-fold increase in the Ki67 value (referred to as “natural log intervals”). Pretreatment Ki67 natural log intervals were not associated with relapse, whereas there was a highly statistically significant association between RFS and posttreatment Ki67 natural log intervals (HR = 1.4, 95% CI = 1.2 to 1.6 per log unit increase; \( P < .001 \); Table 1), which also held for BCSS (HR = 1.4, CI = 1.1 to 1.7; \( P = .009 \); Table 2).

Table 1. Univariate and multivariable analysis of relapse-free survival according to posttreatment pathological tumor size, posttreatment node status, posttreatment Ki67 level, posttreatment ER status, and posttreatment tumor grade in the P024 trial*

<table>
<thead>
<tr>
<th>Factor definitions</th>
<th>No. of patients in each group</th>
<th>No. of events/no. of patients</th>
<th>Univariate analysis</th>
<th>Multivariable analysis</th>
</tr>
</thead>
<tbody>
<tr>
<td>Pathological tumor size†</td>
<td>T1/2 vs T3/4</td>
<td>138/33</td>
<td>47/171</td>
<td>2.7 (1.4 to 5.0)</td>
</tr>
<tr>
<td></td>
<td>T1 vs T2–4</td>
<td>53/118</td>
<td>47/171</td>
<td>2.01 (1.0 to 4.1)</td>
</tr>
<tr>
<td>Node status (positive vs negative)</td>
<td>90/69</td>
<td>44/159</td>
<td>3.9 (1.8 to 8.4)</td>
<td>.01</td>
</tr>
<tr>
<td>Ki67 level, per 2.7-fold increase‡</td>
<td>NA</td>
<td>48/174</td>
<td>2.4 (1.0 to 5.3)</td>
<td>.04</td>
</tr>
<tr>
<td>Clinical response (yes vs no)</td>
<td>70/104</td>
<td>49/174</td>
<td>2.8 (1.6 to 4.9)</td>
<td>.01</td>
</tr>
<tr>
<td>Grade (I vs II/III)</td>
<td>33/126</td>
<td>46/159</td>
<td>3.8 (1.4 to 10.3)</td>
<td>.01</td>
</tr>
</tbody>
</table>

* The total number in each univariate analysis varied depending on the number of cases in which information on the individual factor was available. Cox proportional hazards models were used to calculate hazard ratios (HRs) and their 95% confidence intervals (CIs) of relapse, comparing tumors with the adverse factor relative to tumors without the adverse factor. Two-sided \( P \) values are provided throughout.
† Tumor size was examined with two cutoff points, pT1/2 vs pT3/4 and pT1 vs pT2–4. A cutoff of pT1/2 provided the smallest univariate \( P \) value and was used in the multivariable analysis (which excluded factors with a univariate \( P \) value of >.05).
‡ Ki67 was analyzed as the natural logarithm values, or per 2.7-fold increase according to the original scale of percentage values.
§ The estrogen receptor (ER) analysis refers to the posttreatment values; before treatment all tumors in this dataset were ER+.

Development of a Multivariable Cox Model

Pathological tumor size (T1/2 vs T3/4 or T1 vs T2–4), pathological node status (negative vs positive), clinical response (complete plus partial clinical response vs stable disease plus...
progressive disease), surgical specimen ER status (Allred score ≥3 vs 0 or 2), histological grade (grade I vs grade II/III), and Ki67 natural log intervals were used in a multivariable Cox proportional hazard model to determine their association with RFS (Table 1) and BCSS (Table 2). For Ki67, the posttreatment natural log intervals were statistically significantly associated with both RFS (per log unit increase, HR of relapse = 1.3, 95% CI = 1.05 to 1.50; \( P = .01 \)) and BCSS (per log unit increase, HR of breast cancer death = 1.4, 95% CI = 1.1 to 1.8; \( P = .02 \)). Pathological tumor size (pT1/2 vs pT3/4) was strongly associated with RFS (HR = 3.0, 95% CI = 1.54 to 5.91; \( P = .001 \)) and BCSS (HR = 4.4, 95% CI = 1.7 to 11.3; \( P = .002 \)). Pathological node status was also associated with RFS (positive vs negative, HR = 2.8, 95% CI = 1.31 to 6.19; \( P = .009 \)). ER status posttreatment (Allred score 0 or 2 vs 3–8) was also independently associated with RFS (HR = 2.6, 95% CI = 1.1 to 6.0; \( P = .03 \)). For BCSS, ER loss stands out as being particularly strongly associated with poor outcome (HR = 6.3, 95% CI = 2.1 to 18.7; \( P < .001 \)). Clinical response and grade were not independently associated with RFS or BCSS.

An Integrated Biomarker Model to Predict Long-term Outcome for Patients Treated with Neoadjuvant Endocrine Therapy

The four factors that were associated with \( P \) values of less than .05 in the multivariable analysis (pathological tumor size, pathological node status, ER status, and Ki67 natural log intervals—all derived from the surgical specimen analysis) were promoted to the next analysis phase, which focused on developing a prognostic classification schema. For the first step in model building, the four statistically significant risk factors were reanalyzed by Cox proportional hazards to derive a “final” HR associated with each factor (Table 3). Risk points were then applied using an approach used for the development of prognostic models in cardiovascular disease to weight the different prognostic strengths associated with the final HRs assigned to each factor (Table 4) (11). The scoring process outlined in Table 4 produced three groups (risk score 0, 1–3, and ≥4) that were associated with relapse risks of 10%, 23%, and 48% (\( P < .001 \); Figure 3, A). For breast cancer death, the risk was 2%, 11%, and 17% (\( P < .001 \); Figure 3, B). For RFS, the C index for the four-component model after adjustment for overfitting was 0.723 (95% CI = 0.67 to 0.82). For BCSS, the adjusted C index was 0.78 (95% CI = 0.71 to 0.91). These C indices are within the range that defines a clinically valuable prognostic model, which we termed “PEPI” for preoperative endocrine prognostic index.

Independent Validation of the PEPI Model

We sought to validate the PEPI model in the IMPACT study, a neoadjuvant endocrine therapy trial with short-term endpoints similar to those of the P024 trial (6). Data on pathological stage, ER status, and Ki67 intervals were available from 203 surgical specimens. The model produced a statistically significant separation of the three PEPI risk groups (risk score 0, 1–3, and ≥4), indicating that the model is valid (Figure 3, C) (\( P = .002 \)). As with the P024 study, IMPACT patients with p-stage 1 or 0 disease had a more favorable outcome than patients with p-stage 2 or 3 disease

<table>
<thead>
<tr>
<th>Pathology, biomarker status</th>
<th>RFS (HR 95% CI)</th>
<th>BCSS (HR 95% CI)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Pathological tumor size</td>
<td></td>
<td></td>
</tr>
<tr>
<td>T1/2</td>
<td>2.8 (1.4 to 5.4)</td>
<td>4.4 (1.7 to 11.2)</td>
</tr>
<tr>
<td>T3/4</td>
<td>3.2 (1.5 to 6.9)</td>
<td>3.9 (1.1 to 13.7)</td>
</tr>
<tr>
<td>Node status (positive vs negative)</td>
<td>1.3 (1.1 to 1.6)</td>
<td>1.4 (1.07 to 1.9)</td>
</tr>
<tr>
<td>Ki67 level per 2.7-fold increase</td>
<td>2.8 (1.2 to 6.4)</td>
<td>7.0 (2.4 to 20.9)</td>
</tr>
<tr>
<td>ER, Allred score (0 or 2 vs 3–8)</td>
<td>3 – 8</td>
<td>&lt;0.001</td>
</tr>
</tbody>
</table>

* The four factors associated with a \( P \) value of .05 or less for RFS in Table 1 were reanalyzed in a Cox model to assign final hazard ratios for risk of relapse (RFS) and breast cancer mortality (BCSS). ER = estrogen receptor; HR = hazards ratio; CI = confidence interval.

To determine the potential influence of adjuvant chemotherapy on the RFS estimates in the PEPI score, the use of...
postmenopausal women with ER+ stage 2 and 3 breast cancer (14), but little was known about how the post–neoadjuvant endocrine therapy pathological stage and biomarker status could be used to make decisions regarding other adjuvant treatments. To address this question, we integrated information on standard pathological staging parameters after neoadjuvant endocrine therapy with measurements of ER status and levels of the Ki67 proliferation antigen in the surgical

Discussion

Neoadjuvant endocrine therapy has been widely adopted as a practical means to improve surgical outcomes for postmenopausal women with ER+ stage 2 and 3 breast cancer (14), but little was known about how the post–neoadjuvant endocrine therapy pathological stage and biomarker status could be used to make decisions regarding other adjuvant treatments. To address this question, we integrated information on standard pathological staging parameters after neoadjuvant endocrine therapy with measurements of ER status and levels of the Ki67 proliferation antigen in the surgical
specimen to create the PEPI score that weights these factors according to the magnitude of the HR. Of particular note, patients with low pathological stage (stage 1 or 0) and a favorable biomarker profile (PEPI score 0) at surgery had such a low rate of relapse that further adjuvant systemic therapy beyond continuation of an endocrine agent appears unnecessary. In striking contrast, patients with high pathological stage disease at surgery and a poor biomarker profile (PEPI group 3) had a statistically significant higher risk of early relapse, more typical of ER− disease, and therefore should be offered all appropriate adjuvant treatments available.

The biomarker analysis we have presented here would clearly benefit from additional retrospective studies of tumor samples from other patients who were treated in a similar fashion, with longer follow-up. Prospective validation studies are also warranted to confirm the PEPI as a tool for individualization of adjuvant chemotherapy treatment. As part of these investigations, further study of the PEPI group 2 to more clearly define relapse risk categories would greatly improve the model. For example, a larger sample size could further define the risk of relapse associated with modest elevations of Ki67 above 2.7% in pathological stage 1 and 2A disease. An ideal prospective validation study would include an immediate-surgery control arm to directly compare a standard adjuvant chemotherapy decision-making approach vs pathological staging and tumor biomarker profiling after neoadjuvant endocrine therapy. Pending the results from such a trial, we can state with confidence that analysis of post–neoadjuvant endocrine therapy surgical samples with Ki67 and ER provides additional useful prognostic information that is not provided by an analysis of the baseline sample alone. Repeat ER analysis reveals the presence of a subset of tumors with unstable ER expression and very aggressive clinical course. “On-treatment” Ki67 levels more accurately predict relapse risk than baseline values, whether measured at 3–4 months, as discussed in this study, or as early as 2 weeks after initiation of endocrine treatment, as shown earlier by the IMPACT investigators (9). A standardized ER and Ki67 analysis approach should be adopted as part of any prospective plan to validate the PEPI model as a clinical decision tool.

The data presented may prove useful with respect to other decisions that must be made when a patient is treated with neoadjuvant endocrine therapy. For example, nodal stage after neoadjuvant endocrine therapy was powerfully predictive, suggesting that upfront staging of the axillary nodes may not be necessary before initiation of neoadjuvant endocrine treatment because the prognostic impact of this factor is preserved after therapy. In addition, the PEPI score might be used to tailor radiotherapy decisions. That is, in the setting of a favorable PEPI score and low pathological stage, it might be appropriate to consider partial breast irradiation, or even no radiation, as part of a prospective study (15). The ER and Ki67 data in the present study suggests that prognostic tests that contain signatures for proliferation, ER and ER regulated genes, such as the 21 gene recurrence score (16) or the distinction between Luminal A and luminal B tumors (17) may have additional valuable qualities when applied to samples that have been exposed to endocrine treatment. Specifically, tumors that lose the proliferation gene expression signature in response to endocrine therapy would be expected to have a better prognosis than tumors in which the proliferation signature persisted despite treatment. In contrast, tumors that lose aspects of the luminal signature, particularly ER, would be expected to have a worse prognosis because these tumors have not maintained the pathway required to respond to endocrine therapy after treatment had been initiated (18).

In addition to consideration of the PEPI approach as a tool for treatment individualization, the data presented also strongly support clinical trials in the neoadjuvant setting that study the effects of novel endocrine therapy agents and combinations on short-term endpoints, such as Ki67 and pathological downstaging, as a prelude to large adjuvant trials (19).

References


**Notes**

M. J. Ellis is a consultant for, on the speaker’s bureau of, and currently doing research sponsored by, Novartis Pharmaceuticals Corp and AstraZeneca. D. B. Evans, H. A. Chaudri Ross, and A. von Kameke are employees of Novartis Pharmaceuticals Corp and hold stock in the company. A. S. Bhatnagar is on the speaker’s bureau for Novartis Pharmaceuticals Corp. Ian Smith received honoraria for lecturing at and attending advisory board meetings for Novartis Pharmaceuticals Corp, Pfizer, Inc, and AstraZeneca. M. Dowsett receives research grant funds and has received honoraria for consulting and advisory board work from AstraZeneca and Novartis Pharmaceuticals Corp. Representatives from Novartis Pharmaceuticals Corp have coauthored, edited, reviewed, and approved parts of this manuscript.

Present address: World Wide Services Group Ltd, Geispelgasse 13, CH-4132 Muttenz, Switzerland (A. S. Bhatnagar).

Manuscript received March 10, 2008; revised July 14, 2008; accepted July 29, 2008.