

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

Matthew J. Ellis, Yu Tao, Jingqin Luo, Roger A'Hern, Dean B. Evans, Ajay S. Bhatnagar, Hilary A. Chaudri Ross, Alexander von Kameke, William R. Miller, Ian Smith, Wolfgang Eiermann, Mitch Dowsett

- 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 through bootstrap analysis and subsequently validated it externally using data from an independent neoadjuvant

Affiliations of authors: Siteman Cancer Center, Washington University, St Louis, MO (MJE, YT, JL); Clinical Trials and Statistics Unit, Institute of Cancer Research, Sutton, UK (RAH); Novartis Pharma AG, Basel, Switzerland (DBE, ASB, HACR, AvK); Edinburgh Breast Unit, Edinburgh University, Edinburgh, UK (WRM); Royal Marsden Hospital, London, UK (IS, MD); Red Cross Women's Hospital, Munich, Germany (WE).

Correspondence to: Matthew J. Ellis, MB, BChir, PhD, Siteman Cancer Center, Washington University School of Medicine, 660 South Euclid Ave, St Louis, MO 63119 (e-mail: mellis@wustl.edu).

See “Notes” following “References.”

DOI: 10.1093/jnci/djn309

© 2008 The Author(s).

This is an Open Access article distributed under the terms of the Creative Commons Attribution Non-Commercial License (<http://creativecommons.org/licenses/by-nc/2.0/uk/>), which permits unrestricted non-commercial use, distribution, and reproduction in any medium, provided the original work is properly cited.

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, ie, 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 ≥ 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

CONTEXT AND CAVEATS

Prior knowledge

Endocrine therapy before surgery increases the rate of breast conservation surgery for patients with hormone receptor–positive breast cancer, but factors to predict risk of relapse after treatment have not been identified.

Study design

Posttreatment prognostic factors from patients enrolled in clinical trials of neoadjuvant endocrine therapy were used to develop and validate a model to predict risk of relapse.

Contribution

A model that includes information on standard surgical staging parameters after neoadjuvant endocrine treatment, estrogen receptor status, and levels of Ki67 proliferation antigen can define broad relapse risk groups.

Implications

Data from this model may prove useful with respect to other decisions that must be made after a patient is treated with neoadjuvant endocrine therapy, such as the use of adjuvant chemotherapy.

Limitations

The studies used for developing and validating the model were small, used different treatments, and had relatively short median follow-up data available (just more than 5 years).

From the Editors

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 scaled Schoenfeld residuals. Hazard ratio (HR) estimates of each factor in the final multivariable model were used to construct a score, 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 overfitting (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

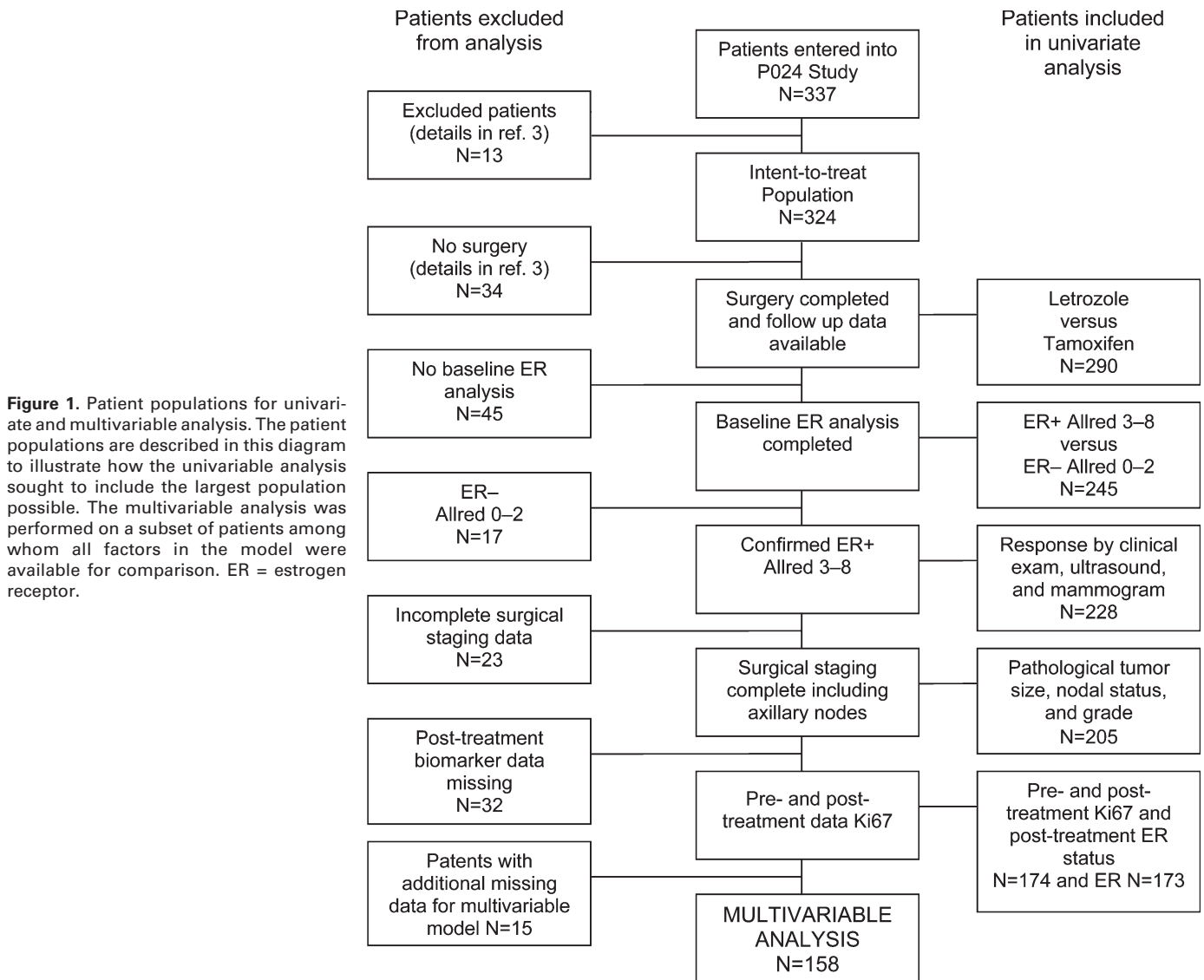


Figure 1. Patient populations for univariate and multivariable analysis. The patient populations are described in this diagram to illustrate how the univariable analysis sought to include the largest population possible. The multivariable analysis was performed on a subset of patients among whom all factors in the model were available for comparison. ER = estrogen receptor.

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

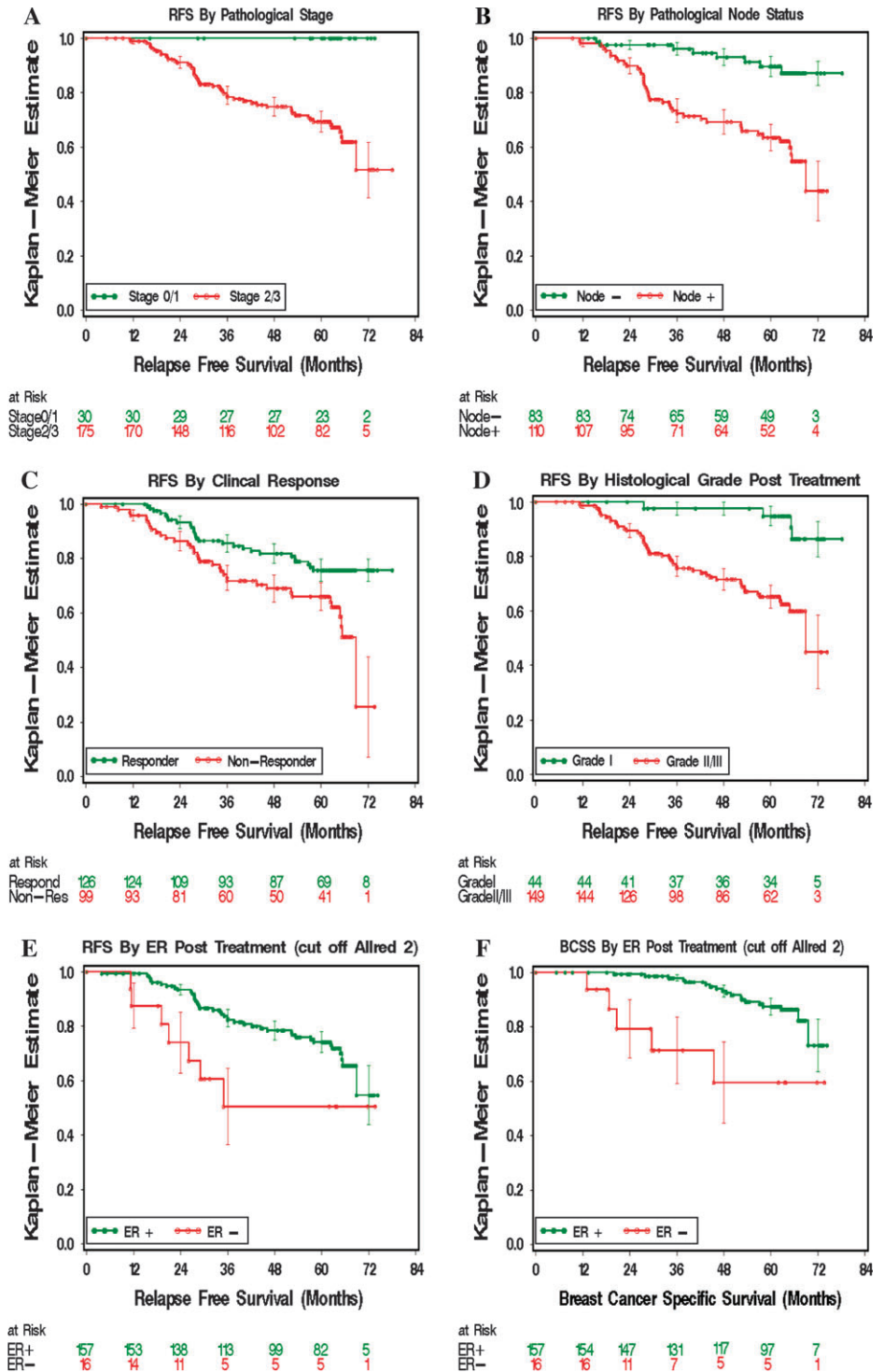


Figure 2. Kaplan–Meier analysis of relapse-free survival (RFS) and breast cancer–specific survival (BCSS) among women whose tumors were established to be estrogen receptor positive (ER+) at baseline by central laboratory analysis. **A**) RFS for posttreatment pathological stage 1 or 0 (T1N0 or T0N0, **green**) vs higher stages (**red**), $P < .001$; **B**) RFS for posttreatment pathological node negative (**green**) vs node positive (**red**), $P < .001$; **C**) RFS for posttreatment clinical responders (**green**) vs no clinical response (**red**), $P = .002$; **D**) RFS for posttreatment histological grade (Grade I, **green**, vs Grade II/III, **red**), $P < .001$. **E**) RFS for patients with ER+ tumors posttreatment (**green**) vs ER negative (–) tumors posttreatment (**red**), $P = .03$; **F**) BCSS for patients with ER+ tumors posttreatment (**green**) vs ER– tumors posttreatment (**red**), $P = .002$; censorship marks are provided as **open circles**. All P values (two-sided) were calculated using the log-rank test.

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

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*

Factor definitions	No. of patients in each group	No. of events/no. of patients	Relapse-free survival			
			Univariate analysis		Multivariable analysis	
			HR (95% CI)	P	HR (95% CI)	P
Pathological tumor size†						
T1/2 vs T3/4	138/33	47/171	2.7 (1.4 to 5.0)	.002	3.0 (1.54 to 5.91)	.001
T1 vs T2–4	53/118	47/171	2.01 (1.0 to 4.1)	.05	—	
Node status (positive vs negative)	90/69	44/159	3.9 (1.8 to 8.4)	<.001	2.8 (1.31 to 6.19)	.009
Ki67 level, per 2.7-fold increase‡	NA	48/174	1.4 (1.2 to 1.6)	<.001	1.3 (1.05 to 1.50)	.01
ER, Allred score (0 or 2 vs 3–8)§	16/157	48/173	2.4 (1.0 to 5.3)	.04	2.6 (1.1 to 6.0)	.03
Clinical response (yes vs no)	70/104	49/174	2.8 (1.6 to 4.9)	<.001	1.72 (0.96 to 3.09)	.07
Grade (I vs II/III)	33/126	46/159	3.8 (1.4 to 10.8)	.011	2.72 (0.95 to 7.8)	.06

* 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 point 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+.

NA (not applicable) because Ki67 divided into five risk groups (see Table 4).

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).

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

Table 2. Univariate and multivariable analysis of breast cancer–specific survival according to pathological tumor size, node status, posttreatment Ki67, posttreatment ER status, and posttreatment tumor grade in the P024 trial*

Factor definition	No. of patients in each group	No. of events/no. of patients	Breast cancer–specific mortality			
			Univariate analysis		Multivariable analysis	
			HR (95% CI)	P	HR (95% CI)	P
Pathological tumor size†						
T1/2 vs T3/4	138/33	24/171	3.5 (1.5 to 8.3)	.004	4.4 (1.7 to 11.3)	.002
T1 vs T2–4	53/118	24/171	4.1 (1.2 to 13.8)	.025	—	
Node status (positive vs negative)	90/69	22/159	4.6 (1.4 to 15.8)	.01	3.2 (0.9 to 11.2)	.07
Ki67 level, per 2.7-fold increase‡	NA	25/174	1.4 (1.1 to 1.7)	.009	1.4 (1.1 to 1.8)	.02
ER, Allred score (0 or 2 vs 3–8)§	16/157	25/173	4.3 (1.6 to 11.7)	.005	6.3 (2.1 to 18.7)	<.001
Clinical response (yes vs no)	70/104	25/174	2.2 (0.97 to 4.9)	.06	1.1 (0.5 to 2.5)	.78
Grade (I vs II/III)	33/126	24/159	7.1 (0.96 to 53)	.05	4.62 (0.6 to 35.0)	.1

* The total number in each univariate analysis varied depending on the number of tumors in which information on the individual factor was available. Cox proportional hazards model was used to calculate hazard ratios (HRs) and their confidence intervals (CIs) of relapse, comparing tumors with the adverse factor relative to those without the adverse factor. Two-sided *P* values are provided throughout.

† Tumor size was examined by two cutoffs, 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 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 the tumors in this dataset were ER+.

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 3. Multivariable Cox proportional hazards analysis of relapse-free survival (RFS) and breast cancer–specific survival (BCSS) from the P024 trial*

Pathology biomarker and response factors	RFS		BCSS	
	HR (95% CI)	P	HR (95% CI)	P
Pathological tumor size (T1/2 vs T3/4)	2.8 (1.4 to 5.4)	.003	4.4 (1.7 to 11.2)	.002
Node status (positive vs negative)	3.2 (1.5 to 6.9)	.004	3.9 (1.1 to 13.7)	.04
Ki67 level per 2.7-fold increase	1.3 (1.1 to 1.6)	.003	1.4 (1.07 to 1.9)	.01
ER, Allred score (0 or 2 vs 3–8)	2.8 (1.2 to 6.4)	.02	7.0 (2.4 to 20.9)	<.001

* 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.

($P = .03$; Figure 3, D). Of particular note, there were no recorded relapses in either study for the p-stage 1 or 0 group with a PEPI risk score of 0, with a combined median follow-up of 60.3 months (range = 4.5–86.5 months).

To determine the potential influence of adjuvant chemotherapy on the RFS estimates in the PEPI score, the use of

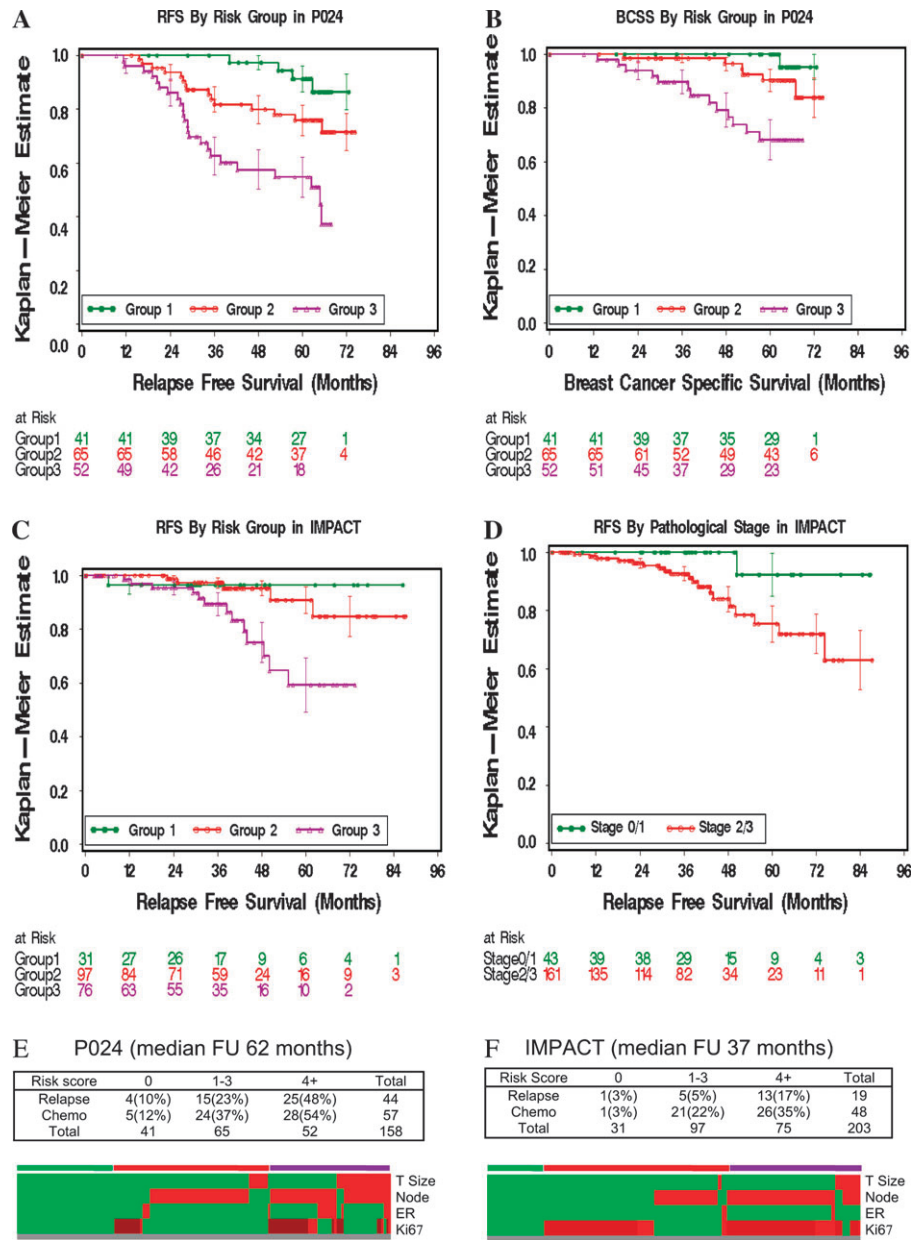
Table 4. The preoperative endocrine prognostic index*

Pathology, biomarker status	RFS		BCSS	
	HR	Points	HR	Points
Pathological tumor size				
T1/2	—	0	—	0
T3/4	2.8	3	4.4	3
Node status				
Negative	—	0	—	0
Positive	3.2	3	3.9	3
Ki67 level				
0%–2.7% (0–1†)	—	0	—	0
>2.7%–7.3% (1–2†)	1.3	1	1.4	1
>7.3%–19.7% (2–3†)	1.7	1	2.0	2
>19.7%–53.1% (3–4†)	2.2	2	2.7	3
>53.1% (>4†)	2.9	3	3.8	3
ER status, Allred score				
0–2	2.8	3	7.0	3
3–8	—	0	—	0

* To obtain the preoperative endocrine prognostic index (PEPI) score, risk points for relapse-free survival (RFS) and breast cancer–specific survival (BCSS) were assigned depending on the hazard ratio (HR) given in Table 3. The points scale was adapted from the cardiovascular literature on predicting outcomes for myocardial infarction (11). The total PEPI score assigned to each patient is the sum of the risk points derived from the pT stage, pN stage, Ki67 level, and estrogen receptor (ER) status of the surgical specimen. An HR in the range of 1–2 receives one risk point; a HR in the 2–2.5 range, two risk points; a HR greater than 2.5, three risk points. The total risk point score for each patient is the sum of all the risk points accumulated from the four factors in the model. For example, a patient with a T1N0 tumor, a Ki67 staining percentage of 1 and an ER Allred score of 6 will have no risk points assigned. In contrast, a patient with a T3N1 tumor, a Ki67 staining percentage of 25, and an ER Allred score of 2 will have a total relapse score of 3 + 3 + 2 + 3 = 11.

† The natural logarithm interval corresponding to the percent Ki67 values on the original percentage scale.

Figure 3. Development and validation of the Preoperative Endocrine Prognostic Index (PEPI). **A)** Relapse-free survival (RFS) for the three PEPI risk groups identified in the P024 model with a log-rank statistic to test the overall trend ($P < .001$). The **green line** represents group 1, patients with a PEPI risk score of 0; the **red line** group 2, a PEPI risk score of 1–3; and the **purple line** group 3, a PEPI risk score of 4 or more. The three groups have distinct risks of relapse. **B)** PEPI groups 1, 2, and 3 also have distinct risks of breast cancer death, with similar statistical significance as the RFS data ($P < .001$). **C)** The PEPI model was validated in the IMPACT trial for RFS, with a statistically significant association between relapse risk and risk score ($P = .002$). **D)** Pathological stage (stage 1 or 0 [**green line**] vs stage 2 or 3 [**red line**]) has a distinctly favorable outcome in the IMPACT trial ($P = .03$). Of 43 patients in the stage 1 or 0 group, only one experienced relapse. This patient's tumor had the highest Ki67 level in the stage 1 or 0 group, and had therefore been correctly assigned to PEPI group 2. **E)** Top, relationships among risk score, relapse events, and adjuvant chemotherapy administration (Chemo) in patients in the P024 trial. Bottom, heat map summarizing the distribution of the individual components of the risk score. **F)** Top, relationships among risk score, relapse events, and adjuvant chemotherapy administration (Chemo) in patients in the IMPACT trial. Bottom, heat map summarizing the distribution of the individual components of the risk score. The heat maps indicate the presence of a favorable factor (**green**) or an adverse factor (**red**) for large tumor size, node-positive status, or estrogen receptor (ER) negativity. The color coding in the Ki67 line of the heat map indicates Ki67 with a risk point of 0 as **green**, a risk point of 1 as **dark red**, and risk point of 2 as **red**. The **bar** over the heat map indicates the three risk groups generated by the risk point assignments (**green**, group 1; **red**, group 2; and **purple**, group 3).



adjuvant chemotherapy was tabulated by PEPI risk group for both studies (Figure 3, E [P024] and F [IMPACT]). The percentages of patients who were treated with chemotherapy in PEPI group 1 (no risk points) were 12% (P024) and 3% (IMPACT), too low for chemotherapy to have had a major role in producing the favorable outcomes observed in this group. The heat maps in Figure 3, E and F, serve to display the distribution of adverse factors in the two studies. These data illustrate the mixed nature of the intermediate group of tumors (PEPI score 1–3), which consisted of either low-stage tumors with adverse biomarkers or higher-stage tumors with favorable biomarkers. Ultimately, the clinical significance of the PEPI model lies in its ability to identify patients at low risk of relapse in the absence of adjuvant chemotherapy (group 1) and patients at very high relapse risk that should mandate all appropriate adjuvant treatments (group 3). More confidence around the estimates of

relapse risk assigned to PEPI group 2 will require studies with larger sample sizes and longer follow-up. Finally, because of the shorter follow-up, there were too few deaths in the IMPACT trial to validate the PEPI model for the prediction of BCSS.

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

1. Bast RC Jr, Hortobagyi GN. Individualized care for patients with cancer—a work in progress. *N Engl J Med*. 2004;351(27):2865–2867.
2. Mazouni C, Peintinger F, Wan-Kau S, et al. Residual ductal carcinoma in situ in patients with complete eradication of invasive breast cancer after neoadjuvant chemotherapy does not adversely affect patient outcome. *J Clin Oncol*. 2007;25(19):2650–2655.
3. Eiermann W, Paepke S, Appfelstaedt J, et al. Preoperative treatment of postmenopausal breast cancer patients with letrozole: a randomized double-blind multicenter study. *Ann Oncol*. 2001;12(11):1527–1532.
4. Ellis MJ, Coop A, Singh B, et al. Letrozole is more effective neoadjuvant endocrine therapy than tamoxifen for ErbB-1- and/or ErbB-2-positive, estrogen receptor-positive primary breast cancer: evidence from a phase III randomized trial. *J Clin Oncol*. 2001;19(18):3808–3816.
5. Smith IE, Dowsett M, Ebbs SR, et al. Neoadjuvant treatment of postmenopausal breast cancer with anastrozole, tamoxifen, or both in combination: the Immediate Preoperative Anastrozole, Tamoxifen, or Combined with Tamoxifen (IMPACT) multicenter double-blind randomized trial. *J Clin Oncol*. 2005;23(22):5108–5116.
6. Dowsett M, Ebbs SR, Dixon JM, et al. Biomarker changes during neoadjuvant anastrozole, tamoxifen, or the combination: influence of hormonal status and HER-2 in breast cancer—a study from the IMPACT trialists. *J Clin Oncol*. 2005;23(11):2477–2492.
7. Ellis MJ, Coop A, Singh B, et al. Letrozole inhibits tumor proliferation more effectively than tamoxifen independent of HER1/2 expression status. *Cancer Res*. 2003;63(19):6523–6531.
8. Allred DC, Harvey JM, Berardo M, Clark GM. Prognostic and predictive factors in breast cancer by immunohistochemical analysis. *Mod Pathol*. 1998;11(2):155–168.
9. Dowsett M, Smith IE, Ebbs SR, et al. Prognostic value of Ki67 expression after short-term presurgical endocrine therapy for primary breast cancer. *J Natl Cancer Inst*. 2007;99(2):167–170.
10. Katz RL, Patel S, Sneige N, et al. Comparison of immunocytochemical and biochemical assays for estrogen receptor in fine needle aspirates and histologic sections from breast carcinomas. *Breast Cancer Res Treat*. 1990;15(3):191–203.
11. Morrow DA, Antman EM, Charlesworth A, et al. TIMI risk score for ST-elevation myocardial infarction: a convenient, bedside, clinical score for risk assessment at presentation: an intravenous nPA for treatment of infarcting myocardium early II trial substudy. *Circulation*. 2000;102(17):2031–2037.
12. Harrell FE Jr, Lee KL, Mark DB. Multivariable prognostic models: issues in developing models, evaluating assumptions and adequacy, and measuring and reducing errors. *Stat Med*. 1996;15(4):361–387.
13. Urruticoechea A, Smith IE, Dowsett M. Proliferation marker Ki-67 in early breast cancer. *J Clin Oncol*. 2005;23(28):7212–7220.
14. Ma CX, Ellis MJ. Neoadjuvant endocrine therapy for locally advanced breast cancer. *Semin Oncol*. 2006;33(6):650–656.
15. Hughes KS, Schnaper LA, Berry D, et al. Lumpectomy plus tamoxifen with or without irradiation in women 70 years of age or older with early breast cancer. *N Engl J Med*. 2004;351(10):971–977.
16. Paik S, Shak S, Tang G, et al. A multigene assay to predict recurrence of tamoxifen-treated, node-negative breast cancer. *N Engl J Med*. 2004;351(27):2817–2826.
17. Hu Z, Fan C, Oh DS, et al. The molecular portraits of breast tumors are conserved across microarray platforms. *BMC Genomics*. 2006;7:96–108.

18. Ellis MJ, Tao Y, Luo J, et al. A poor prognosis ER and HER2-negative, nonbasal, breast cancer subtype identified through postneoadjuvant endocrine therapy tumor profiling. *J Clin Oncol*. 2008;26:Abstract 502.
19. Ellis MJ. Neoadjuvant endocrine therapy as a drug development strategy. *Clin Cancer Res*. 2004;10(1 pt 2):391S-395S.

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.